Fibre Array Storage Technology
A FASTT Introduction

FASTT200 and FASTT500 described in detail

Version 7.10 Storage Manager software explained

Storage solutions to meet your business requirements

Barry Mellish
Jure Arzensek
Christian Demmer
Noam Rosen

ibm.com/redbooks
First Edition (May 2001)

This edition applies to the IBM FASTT200 and FASTT500 Storage Servers and the FASTT Storage Manager software Version 7.10. This book was written using pre-GA versions of the Storage Manager software. Some of the screen detail may be different than the GA version of the code.

This document was updated on August 31, 2001.

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. 471F 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 2001. 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 .................................................. xi

Preface .................................................. xiii
The team that wrote this redbook .......................... xiii
Comments welcome ..................................... xiv

Chapter 1. Introduction ................................... 1
1.1 Managing the FAST Storage Servers .................. 3
1.2 Arrays and logical drives ............................. 5
1.3 Storage partitioning and heterogeneous hosts .......... 6
1.4 Building blocks and sample configurations ............ 7
1.5 What is new in FAST software V7.10 .................. 10

Chapter 2. Hardware details ............................... 15
2.1 The FAST200 Storage Server and FAST200 HA Storage Server ........ 15
   2.1.1 Front view .................................. 15
   2.1.2 Rear view .................................. 16
   2.1.3 The RAID controller ......................... 16
   2.1.4 Replacing the cache battery .................. 18
2.2 The FAST500 Storage Server ......................... 19
   2.2.1 Front view .................................. 20
   2.2.2 Rear view .................................. 21
   2.2.3 Host side connections ....................... 22
   2.2.4 Drive side connections ..................... 22
   2.2.5 FAST500 diagnostic LEDs .................. 23
   2.2.6 FAST200 and FAST500 comparison .......... 25
2.3 EXP500 storage expansion enclosure .................. 26
   2.3.1 Front view .................................. 26
   2.3.2 Rear view .................................. 26
   2.3.3 EXP500 diagram ............................ 28
   2.3.4 Incoming and outgoing GBIC ports .......... 28
   2.3.5 Tray ID switches ........................... 28
   2.3.6 FAST200 and EXP500 cabling ............... 29
   2.3.7 FAST500 and EXP500 cabling ............... 29
2.4 FAST host adapter .................................. 30
   2.4.1 Fast!UTIL .................................. 31
2.5 Hubs and switches .................................. 35
   2.5.1 IBM SAN Fibre Channel Managed Hub (3534) .... 35
   2.5.2 IBM SAN Fibre Channel 8 and 16-port Switch (2109) .... 36

Chapter 3. The FAST Storage Manager software ............ 37
3.1 Direct-attached and host-attached management ....... 38
3.2 The FAST Storage Manager Client ................... 40
3.3 Event monitor ..................................... 41
3.4 Basic management .................................. 42
   3.4.1 Creating arrays and logical drives ............ 43
   3.4.2 Expanding arrays ........................... 44
   3.4.3 Migrating RAID levels ...................... 45
   3.4.4 Changing the segment size .................. 47
3.4.5 Cache parameters ................................................. 48
3.4.6 Media scan ......................................................... 50
3.5 Advanced management ............................................... 50
  3.5.1 Storage partitioning .............................................. 50
  3.5.2 Heterogeneous hosts ........................................... 53
  3.5.3 Auto Volume Transfer (AVT) and RDAC: providing redundancy .... 55
  3.5.4 FASTT Utilities ............................................... 59
  3.5.5 Diagnostics and troubleshooting ................................. 59
  3.5.6 Command line support ....................................... 60
3.6 Performance ......................................................... 61

Chapter 4. Step by step procedures for the FASTT Storage Server ............. 63
4.1 Driver and host software installation .................................. 63
  4.1.1 Microsoft Windows NT 4.0 and Windows 2000 .................. 64
  4.1.2 Linux Redhat 6.2 .............................................. 65
  4.1.3 Novell Netware 5.1 ............................................ 67
  4.1.4 Solaris ......................................................... 68
  4.1.5 HP-UX ......................................................... 70
  4.1.6 Network attached management .................................. 73
4.2 Configuring the FASTT ............................................... 74
  4.2.1 Planning the configuration ..................................... 74
  4.2.2 Starting the FASTT Storage Manager Client ................... 75
  4.2.3 Updating the controller microcode ............................. 77
  4.2.4 Updating the drive microcode .................................. 82
  4.2.5 Initial configuration steps ..................................... 83
  4.2.6 Creating arrays and logical drives ............................ 85
  4.2.7 Define hot spare drives ....................................... 88
  4.2.8 Configuring storage partitioning .............................. 89
  4.2.9 Configuring, monitoring and alerting .......................... 95
4.3 Installing a Microsoft Cluster Server with FASTT Storage Server ........ 97
  4.3.1 Preparing the installation of the cluster server ............... 97
  4.3.2 Preparing Node A .............................................. 97
  4.3.3 Preparing Node B ............................................. 98
  4.3.4 Installing the cluster service on Node A ....................... 99
  4.3.5 Installing the cluster service on Node B ....................... 99
  4.3.6 Apply service pack .......................................... 99
4.4 Maintenance, diagnostics and troubleshooting .............................. 100
  4.4.1 The storage subsystem profile .................................. 100
  4.4.2 Cache settings and media scan ................................ 102
  4.4.3 Logical drive properties, cache settings and RAID migration .... 103
  4.4.4 The Major Event Log (MEL) .................................. 108
  4.4.5 The Recovery Guru .......................................... 109
  4.4.6 Handling premium features .................................... 111
  4.4.7 The performance monitor ..................................... 113
  4.4.8 Load scripts to the FASTT Storage Server ..................... 114
  4.4.9 Load and save configuration of the FASTT Storage Server .... 115
  4.4.10 The FASTT Check Application ................................ 116
  4.4.11 Obtaining the World Wide Name of the host bus adapter ....... 119
  4.4.12 Common problems .......................................... 123
  4.4.13 Network setup .............................................. 126
  4.4.14 Resetting the controllers .................................... 128
Chapter 5. FASTT solution design .......................................................... 131
5.1 Infrastructure deployment framework ............................................... 131
   5.1.1 Design and planning ............................................................. 132
5.2 Information gathering ................................................................. 132
5.3 Design guidelines ............................................................................ 133
   5.3.1 Disk drive performance characteristics ........................................ 134
   5.3.2 RAID level and performance ..................................................... 136
   5.3.3 Fibre Channel versus Ethernet (SAN versus NAS) ....................... 145
   5.3.4 FASTT Storage Manager — Performance Monitor ......................... 146
   5.3.5 FASTT best practices based on observations .................................. 151
5.4 Clustering ......................................................................................... 156
   5.4.1 Concepts of Microsoft Cluster Server ............................................ 156
   5.4.2 Cluster Server configuration ....................................................... 162
   5.4.3 Design Microsoft Cluster Server — The storage perspective ............ 165
   5.4.4 Design Exchange 2000 cluster ..................................................... 167
   5.4.5 SQL2000 Cluster recommendation — The storage perspective ......... 173
5.5 Case studies ..................................................................................... 174
   5.5.1 Extending your file server storage — direct attached (FASTT 200) .. 175
   5.5.2 Extending your file server storage — direct attached (FASTT 500) .. 177
   5.5.3 Two servers attached directly to one FASTT 200 ......................... 178
   5.5.4 Low cost Microsoft Cluster Server .............................................. 180
   5.5.5 Highly available Microsoft Cluster Server on FASTT 200 ............... 182
   5.5.6 No single point of failure MSCS direct attached — FASTT500 ....... 183
   5.5.7 Two or more Microsoft Clusters using the same shared storage ....... 185
   5.5.8 Two MSCS with NSPOF direct attached ....................................... 186
   5.5.9 More than two MSCS with NSPOF ............................................. 188
   5.5.10 Small SAN solution ............................................................... 190
   5.5.11 Large SAN solution ............................................................... 191

Appendix A. Critical events ................................................................. 195

Appendix B. Installing Windows 2000 Cluster service .......................... 203
B.1 Install Cluster service on Node A ..................................................... 203
B.2 Install Cluster service on Node B ..................................................... 209
B.3 Examining the system event log ...................................................... 211
B.4 View changes made to each node of the cluster ............................... 212

Appendix C. Special notices ............................................................. 215

Appendix D. Related publications ..................................................... 217
D.1 IBM Redbooks ................................................................................. 217
D.2 IBM Redbooks collections ............................................................ 217
D.3 Referenced Web sites ...................................................................... 217

How to get IBM Redbooks ................................................................. 219
IBM Redbooks fax order form ............................................................ 220

Glossary ......................................................................................... 221

Abbreviations and acronyms ............................................................. 231

Index .............................................................................................. 233

IBM Redbooks review ...................................................................... 235
Tables

1. FASTT200 and FASTT500 comparison ........................................... 25
2. Failover options by operating systems ...................................... 58
3. Primary and secondary paths for HP-UX ..................................... 72
4. Mapping table for our example configuration .............................. 74
5. Scripts included ........................................................................... 115
6. Performance with RAID levels .................................................... 144
7. RAID-5 versus RAID-10 ............................................................... 144
8. Monitor data fields ....................................................................... 147
9. Best practices for Exchange database components .................... 172
10. Case studies reference table ....................................................... 175
11. FASTT200 direct attached ......................................................... 176
12. FASTT500 direct attached ........................................................... 178
13. Two servers direct attached to FASTT200 .................................. 180
14. FASTT200 low cost MSCS .......................................................... 182
15. SPOF MSCS with FASTT200 ...................................................... 183
16. Solution components no SPOF MSCS ......................................... 185
17. FASTT Hosting Multi Microsoft Clusters ..................................... 186
18. Two clusters on FASTT500 .......................................................... 188
19. Multi Microsoft Clusters on FASTT500 ....................................... 189
20. Small FASTT200 SAN ................................................................. 191
21. Large FASTT500 SAN solution .................................................. 193
22. Critical events in the major event log ......................................... 195
Preface

This IBM Redbook gives a detailed introduction to the FASTT range of storage products with particular emphasis on the FASTT200. Chapter 1 introduces the FASTT concepts. Chapter 2 discusses the hardware in detail. Chapter 3 discusses the FASTT Storage Manager in detail with emphasis on the Version 7.10 release. This release of the software introduces heterogeneous host support. Chapter 4 discusses operating system considerations. In Chapter 4 we assume that the reader is familiar with the operating system, so although we discuss command syntax and other considerations, we do not discuss them beyond the level required to use a FASTT product. Chapter 5 details storage solutions for various business scenarios.

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.

Barry Mellish is a Project Leader at the International Technical Support Organization, San Jose Center. He has coauthored six previous redbooks and has taught many classes on storage subsystems. He joined IBM in the United Kingdom 17 years ago, and before joining the ITSO, he worked as a Senior Storage Specialist on the Disk Expert Team in EMEA.

Jure Arzensek is an IT Specialist and PC Institute Instructor for IBM in Slovenia. He has six years of experience in IBM @server xSeries, Netfinity and PC servers and is certified as an IBM @server xSeries Systems Expert. Jure has delivered the PC Institute courses in more than 15 EMEA countries. He holds a degree in Computer Sciences from the University of Ljubljana. His areas of expertise include Microsoft Windows NT and 2000, MSCS and Novell NetWare.

Christian Demmer is an IT Specialist and Pre-sales Consultant in the IBM Server Group in Germany. He provides technical support to other IBM Professionals, Business Partners and Customers in xSeries solutions. He holds a degree in Mathematics from the University of Dortmund. His areas of expertise include Linux, Microsoft Windows NT and 2000 and high availability solutions.

Noam Rosen is an IT Specialist and Sales Specialist in the IBM Server Group in Israel. He provides technical and solution pre-sales support to IBM Specialists, Business Partners and Customers. He holds a degree in Economics from Bar-Ilan University in Israel. Noam is a Microsoft Certified Trainer and Microsoft Certified Systems Engineer for Windows 2000, and he is also certified as an IBM @server xSeries System Expert. His broad range of technical support activities include consultation on clustering and storage implementation as well as presentations and training for Large Customers and Business Partners.

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

Bob Reiland, IBM Raleigh
Mick Watkins, IBM Raleigh
Charlie Stephan, IBM Raleigh
Dave Worley, LSI Logic
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 “IBM Redbooks review” on page 235 to the fax number shown on the form.
- Use the online evaluation form found at ibm.com/redbooks
- Send your comments in an Internet note to redbook@us.ibm.com
Chapter 1. Introduction

The FAST Storage Server is a RAID controller device that contains Fibre Channel (FC) interfaces to connect the host systems and the disk drive enclosures. The Storage Server provides high system availability through use of hot-swappable and redundant components. We will discuss the following two products:

- The FASTT200 Storage Server
  This solution is suitable for entry and medium level environments.

- The FASTT500 Storage Server
  This Storage Server can support high-end configurations with massive storage capacities and a large number of heterogeneous host systems. It offers a higher level of availability, performance and expandability than FASTT200.

The FASTT200 Storage Server

The FASTT200 Storage Server is a 3U rack-mountable Fibre Channel RAID controller and disk drive enclosure. It targets the entry and midrange segment of the FC storage market. A typical use of the FASTT200 would be in a 2-node cluster environment with up to 30 Fibre Channel disk drives attached to the Storage Server. It is shown in Figure 1.

Two models are available:

- The FASTT200 Storage Server, with a single RAID controller.

- The FASTT200 High Availability (HA) Storage Server which contains two RAID controllers and can therefore provide higher availability.

Both models feature hot-swap and redundant power supplies and fans and you can install up to ten slim-line or half-high FC disk drives. If you need to connect more than ten disks, you can use the EXP500 FC storage expansion enclosures. Each EXP500 can accommodate ten additional disk drives, and up to five EXP500s are supported on the FASTT200. This means that the maximum supported number of disk drives is 60.
The use of hot-swappable and redundant components provides high availability for the FASTT200 Storage Server. A fan or a power supply failure will not cause downtime and such faults can be fixed while the system remains operational. The same is true for a disk drive failure if fault-tolerant RAID levels are used. With two RAID controller units and proper cabling, a RAID controller or path failure will not cause loss of access to data.

Each RAID controller has one host and one drive FC connection. The FASTT200 HA model can use the two host and drive connections to provide redundant connection to the host adapters and to EXP500 enclosures. Each RAID controller unit also contains 128MB of battery-backup cache. Hardware components of the FASTT200 are explained in detail in 2.1, “The FASTT200 Storage Server and FASTT200 HA Storage Server” on page 15.

**FASTT500 Storage Server**

The FASTT500 Storage Server, shown in Figure 2, is a 4U rack-mountable Fibre Channel RAID controller device. It provides the levels of performance, availability and expandability needed to satisfy high-end storage requirements. You would typically use the FASTT500 Storage Server in advanced cluster environments and possibly with heterogeneous operating systems running on the host systems.

![Figure 2. FASTT500 Storage Server](image)

The Storage Server features two RAID controller units, redundant power supplies and fans. All these components are hot-swappable, which assures excellent system availability. You use the EXP500 external storage expansion enclosures to install the FC disk drives and you can connect up to 22 EXP500 enclosures to the FASTT500. This means a total of up to 220 disk drives. The enclosures can be connected in a fully redundant manner, which provides a very high level of availability. On the host side FC connections, you can use up to four mini-hubs.

---

**Please note**

Up to 126 FC devices may be configured in a FC loop. But arbitrated loop physical addresses (AL_PA) are needed also for controller ports and expansion enclosure monitoring. Therefore, you could connect up to ten EXP500 with 110 disks total (10 disks in FASTT200, 100 disks in additional EXP500 units). This is not recommended from a performance perspective. Performance of FASTT200 Storage Server will be best with (up to) 30 disk drives.
This allows you to establish up to eight host connections without needing an external hub or a switch.

For performance and availability, each RAID controller unit contains 256MB of battery-backed cache and this amount can be further expanded.

We discuss the hardware details of FAStT500 in 2.2, “The FAStT500 Storage Server” on page 19.

1.1 Managing the FAStT Storage Servers

To configure and manage the FAStT Storage Servers, use the FAStT Storage Manager software. The current version at the time of writing is v7.10. This utility allows you to configure arrays and logical drives, assign your logical drives into storage partitions, replace and rebuild failed disk drives, expand the size of arrays and convert from one RAID level to another. It also allows you to perform troubleshooting and management tasks, like checking the status of FAStT Storage Server components, update the firmware of RAID controllers and similar tasks.

Another possibility is to use the serial interface and a terminal emulation utility. However, this is only meant for advanced troubleshooting and management and should only be used when other management methods fail.

FAStT Storage Manager components

FAStT Storage Manager components include:

- **FAStT Storage Manager Client**
  
  This is the GUI utility used to configure, manage and troubleshoot the Storage Server. It can be installed either on the host system or on a managing workstation.

- **FAStT Storage Manager Agent**
  
  This component must be installed on the host system to enable host-attached in-band management.

  The client may be installed on either the host server which is connected to FAStT storage and/or it may be installed on a networked client or management station. Also, the FAStT storage may be managed in-band (through fibre) or out-of-band (through direct network, Ethernet). In-band management requires that the host server has the agent installed. Out-of-band management requires that the Ethernet connection and IP addresses have been provided to the FAStT controllers. Both management methods may be used simultaneously. If both connections are used, out-of-band management is the default connection with in-band as the alternate (backup) method.

- **Redundant Disk Array Controller (RDAC)**

  The RDAC component contains a multipath driver and hot-add support. It must be installed on the host system and it will provide redundant path to the Storage Server when both RAID controllers are installed. If a RAID controller fails or becomes inaccessible due to connectivity problems, RDAC will reroute the I/O requests through another RAID controller. The hot-add part of RDAC allows you to register new logical drives to the operating system dynamically.
Some operating systems do not use RDAC as they have their own multi-path drivers.

- **FAStT Utilities**

  The FAStT Utilities package contains two command line tools: hot_add and sm7devices.

  The hot_add utility allows the operating system to detect new logical drives without rebooting the host system. When you run the utility, it will rescan the host bus adapters and handle the operating system assignments of all new devices found.

  The SM7devices utility will list all logical drives, World Wide Names (WWNs) and the storage subsystem that it can access. We mainly use this utility for troubleshooting as it provides a basic check of Storage Server setup and fibre channel (FC) connectivity.

**Management methods**

You can manage your FAStT Storage Server using either one of these two methods:

- **Direct-attached management**

  If you want to use this method, you must establish an Ethernet link to each RAID controller unit. For security reasons, it is usually a good idea to connect the Ethernet ports on the RAID controllers and the managing workstation to a dedicated LAN segment, separate from production network. You will need to install FAStT Storage Manager Client on the managing machine. Be aware that it will require some effort to set the network access to FAStT up. One possibility is to use a DHCP/BOOTP server that will provide IP address for each RAID controller. Alternatively, you can also set fixed IP address to each controller, but you must use the serial interface to do this. Figure 3 shows an example of direct-attached management scenario.

![Figure 3. Direct-attached management](image)

- **Host-attached management**

  As mentioned, you must install the agent code on the host system in order to be able to manage the Storage Server this way. Here you do not have to deal with networking configuration of FAStT. But this method requires a special logical drive for communication between the RAID controller and the management utility — it is called the Access logical drive. A LUN must be
assigned to the Access logical drive and this means you can configure one LUN less than with the direct-attached method. An example of host-attached management scenario is shown in Figure 4.

![Figure 4. Host-attached management](image)

For more information about FASiT Storage Manager, Chapter 3, “The FASiT Storage Manager software” on page 37 explains the concepts, and 4.2, “Configuring the FASiT” on page 74 lists all the important details.

### 1.2 Arrays and logical drives

FASiT Storage Server is basically a RAID controller. It supports RAID levels 0, 1, 3, and 5 (RAID level 1 is actually implemented as RAID 10). You should always select the RAID level which is most appropriate for your particular data access pattern and which suits your availability requirements. Various RAID levels differ in performance, disk space efficiency and level of fault-tolerance. They are discussed in details in 5.3.2, “RAID level and performance” on page 136.

To configure the Storage Server, use the FASiT Storage Manager to group your physical disks into arrays and then create one or more logical drives inside those arrays. Logical drives are the entities which appear to the operating system as physical disk drives. RAID level is specified per each array. This means all logical drives inside an array will use the same RAID level. You can configure up to 128 logical drives per FASiT Storage Server.
Figure 5 shows a sample configuration that consists of 12 physical disk drives. We have divided the physical disk drives into four arrays and created a total of eight logical drives in these arrays. The operating system sees these logical drives as physical disks.

Several array and logical drive operations can be done dynamically using the FASTT Storage Manager:

- Adding new disk drives to arrays
- Changing the RAID level for an array
- Changing the segment size of a logical drive
- Defragmentation

Detailed information on working with arrays and logical drives is available in 3.4, "Basic management" on page 42.

1.3 Storage partitioning and heterogeneous hosts

Storage partitioning allows you to connect multiple host systems to the same Storage Server. It is a way of assigning logical drives to specific host systems or groups of hosts — this is known as LUN masking. Logical drives in a storage partition will only be visible and accessible by their assigned group or individual hosts. Heterogeneous hosts support means that the host systems can run different operating systems. But be aware that all the host systems within a particular storage partition must run the same operating system. As all host systems within a particular storage partition have unlimited access to all logical drives in this partition. Therefore, file systems on these logical drives must be compatible with host systems. The best way to ensure this is running the same operating system on all hosts within the same partition. Some operating systems (for example, Linux) may be able to mount foreign file systems. Additionally, please keep in mind that FASTT Storage Manager does not provide any control of concurrent access to the same disk blocks. Access control must be provided by cluster or file sharing software.

Storage partitioning gives you a high level of flexibility when attaching multiple hosts to the same Storage Server. For example, if you attach multiple non-clustered hosts, your storage space can be divided into storage partitions and each host will only be able to access its own logical drives. Another

Fibre Array Storage Technology: A FASTT Introduction
would be to use the storage space for a mixture of several clusters and individual hosts. For this to work, each cluster and each host must have access to its own logical drives and must not gain access to other logical drives. With storage partitioning, you would create several storage partitions, typically one per each cluster or host.

You can assign up to 32 logical drives into each storage partition and each logical drive can belong to exactly one storage partition.

Figure 6 shows an example of storage partitioning. Five logical drives are divided into two storage partitions:

- Logical drives 1 and 5 are assigned to Host A.
- Logical drives 2, 3 and 4 are assigned to Cluster B.

**Figure 6. Storage partitioning**

You can find detailed information on storage partitioning in 3.5.1, “Storage partitioning” on page 50.

### 1.4 Building blocks and sample configurations

Fibre components required to build a Storage Server-based solution include:

- FASiT Storage Server
- FASiT Host adapters
- EXP500 Fibre Storage Expansion Enclosures
- Fibre Channel hubs and switches

The best way to explain the role of each component is through use of sample configurations.

A minimum FASiT200 configuration consists of a FASiT200 Storage Server and a host system with one FASiT host bus adapter. The Storage Server is directly attached to the host system. This is shown in Figure 7. We have also attached an EXP500 to the FASiT200 to allow more expandability for the disk drives. We use only one Host Adapter and one RAID controller, so this is obviously not a redundant solution. This solution is only recommended for non-critical environments as failure of a single component can result in loss of access to data.
For more complex and highly available configurations, you have to use a FC hub or a switch. A hub would be a more probable choice with FASTT200, because FC switches are usually not cost-efficient in small and medium sized Fibre storage solutions. Figure 8 shows an example of a 2-node cluster with two RAID controllers installed in FASTT200HA Storage Server. We use a FC Managed Hub to connect the two host systems to both RAID controllers. This solution does not provide full redundancy. A fully redundant one would require duplicate managed hubs and also duplicate host adapters installed in each node.
A final example in Figure 9 shows four servers, implemented as two 2-node clusters. We use FAST500 Storage Server and three EXP500 expansion enclosures. The use of storage partitioning is required here to create an exclusive set of logical drives for each cluster.

This is a fully redundant setup. Each cluster node contains a pair of host adapters and each adapter in the pair is connected to a different RAID controller. As you can see, there is no need to use an external hub or a switch: the mini-hubs on the host side of the FAST500 Storage Server provide very flexible connectivity options. The EXP500 enclosures are also connected through a fully redundant FC loop.

In this example, flexibility and expandability of FAST500 Storage Server is much higher than that of FASTT200.
1.5 What is new in FASTT software V7.10

FASTT software includes the FASTT Storage Manager, operating system drivers and FASTT firmware. The new V7.10 release of FASTT software brings several major enhancements to the FASTT200 and FASTT500 Storage Server products.

**Heterogeneous hosts**

A new feature of the latest version of FASTT code V7.10 is support for heterogeneous hosts connected to the same FASTT Storage Server. Various storage partitions on a single FASTT Storage Server can now be accessed by servers running a combination of these operating systems:

- Windows NT
- Windows 2000
- Linux
- Novell NetWare
- Sun Solaris
- HP-UX

You must use at least one dedicated storage partition for each of the operating systems that appear in your network, as all host systems within a particular storage partition have unlimited access to all logical drives in this partition. Therefore, file systems on these logical drives must be compatible with host systems. Additionally, please keep in mind that FASTT Storage Manager does not
provide any control of concurrent access to the same disk blocks. Access control must be provided by cluster or file sharing software.

**Figure 10. FASTT200HA in heterogeneous hosts environment**

An example of heterogeneous hosts environment is shown in Figure 10. In this particular example, we would have to create four storage partitions on FASTT200: one for each host system and its logical drives.

**Configuration replication**

You can save configuration of FASTT Storage Server and use it later to import it to another Storage Server with identical hardware setup. This allows you to replicate your configuration settings to multiple Storage Servers. Being able to save configuration is useful in its own right — it allows you to make and keep a backup of your Storage Server configuration.

**Event monitoring**

In previous versions of Storage Manager software, you had to keep the Enterprise Management Window open at all times if you wanted to constantly monitor your Storage Server. When the Enterprise Management Window was closed, there was no event monitoring and alerting anymore. The new release V7.10 includes the Event Monitor, a new program that is installed as part of the Storage Manager Client package. You should run the Event Monitor on a managing workstation that is connected to the Storage Server 24 hours a day. This utility runs in the background and constantly checks the Storage Server for possible critical events, like disk drive failures, RAID controller faults and similar...
checks. If a fault is detected, Event Monitor can alert the remote system by sending an e-mail or an SNMP trap. If there are multiple Event Monitors running and checking a specific Storage Server, each of them will send an alert — so duplicate alert notifications will be received.

More storage partitions
FASiT software V7.10 now supports up to 16 storage partitions. You will only need so many storage partitions when the number of host systems and disk drives is very high. For performance reasons, the FASiT200 Storage Server will usually not be used in such large configurations. But the FASiT500 offers a much higher throughput and therefore handles a large number of hosts much easier.

Auto Volume Transfer (AVT)
This feature offers a lot more flexibility to logical drives ownership. In previous releases, a RAID controller could only own a complete array with all its logical drives. It was impossible to transfer a single logical drive to another controller: ownership was determined on array level.

The new release allows individual logical drives within an array to be owned by different RAID controllers. This obviously gives you more flexibility. You can now use a fewer number of physical disk drives and arrays to create more logical drives for your storage space. But be aware that there might be performance-related implications: a lower number of physical disks will usually decrease performance and so will a large number of logical drives within your arrays. For performance, it is usually best to configure a single logical drive within an array and spread your array across many physical disks.

With AVT enabled, the logical drive ownership will automatically be transferred to the alternate RAID controller if there is a problem with the preferred RAID controller. As soon as the problem is corrected, the preferred controller re-establishes the ownership. AVT works in conjunction with the multipath I/O driver on the host system.

Command line support
You can now execute the script engine commands using your operating system command line interface. This can be very useful and it also allows you to create batch files. Such capability can be effectively used together with various management applications, like IBM Director. For example, you can create automated responses to certain alerts received from the Storage Server.

It is possible to run either individual script engine commands or you can group multiple commands into a sequence; and you can also run script files.

Controller diagnostics
You can perform certain built-in diagnostic tests when troubleshooting the RAID controllers. FC loop connectivity on both the host and drive side can be verified.

Access logical drive mapping
The Access Logical Drive is a special drive which uses none of the physical disk drives and should be assigned to the last (highest) available LUN number. Typically that LUN number will be LUN 31. On older versions of the Storage Manager software this LUN number may be LUN 7. The Access Logical Drive allows the Storage Manager Agent to communicate to the Fibre Channel RAID controllers through the fibre connection for storage management services. These
services include monitoring, configuring, and maintaining the RAID storage device.

The Storage Manager Agent is used in conjunction with the Access Logical Drive. If either the agent or the access drive is not used then the other component is not required.

The access drive and agent can be restricted to specific management stations (servers) so that all servers do not have to have or use the agent and access drive. You should keep in mind that having redundant accesses to the RAID Storage from multiple servers allows the maintaining of storage management in situations where a server fails.

**Note:** To provide redundant management capability it is good practice to supply both in-band (fibre) and out-of-band (Ethernet) connections to the RAID Storage device.
Chapter 2. Hardware details

In this chapter we describe the hardware components of a Fibre channel storage solution. We will cover the FASTT200 and FASTT200 HA Storage Servers first. The FASTT500 Storage Server is also discussed, because it is very important to differentiate between the FASTT200 and FASTT500 solutions and position them properly. We also mention additional hardware components, essential for a complete storage solution. These include the EXP500 external expansion enclosure, the FAST Host Adapter, the HF hub and the FC switch.

2.1 The FASTT200 Storage Server and FASTT200 HA Storage Server

The FASTT200 Storage Server is a 3U rack-mountable device containing a single RAID controller and space for up to 10 Fibre Channel (FC) hard disk drives. This is a cost-effective Fibre RAID solution for environments where the highest level of availability and performance is not essential. It contains hot-swappable and redundant power supplies and fans, however there is no RAID controller redundancy.

The FASTT200 HA Storage Server is in most ways identical to FASTT200 Storage Server, except that it contains two hot-pluggable RAID controllers and can therefore provide a higher level of availability.

You can upgrade the FASTT200 Storage Server by adding the hot-pluggable FASTT200 Redundant RAID Controller, effectively turning it into the FASTT200 HA Storage Server.

2.1.1 Front view

From the front, you can access up to ten hot-swappable FC hard disk drives and view the two status LEDs on top (Figure 11):

- Green Power LED
  This LED indicates the DC power status is OK.
- Amber General-System-Fault LED
  When a storage server component fails (such as a disk drive, fan or power supply), this LED will be on.

Each disk drive also has its own status LEDs:

- Green Activity LED
When this LED is flashing, it indicates disk drive activity. When it is steadily ON, it means the disk drive is OK and properly installed.

- Amber Disk-Drive-Fault LED

If a disk drive fails, this LED will be ON. While the drive is being identified or rebuilt, it will flash.

### 2.1.2 Rear view

![Figure 12. FASiT200 Storage Server, rear view](image)

You access the RAID controllers, power supplies and fans from the rear side of the FASiT200 unit (Figure 12). Two hot-swappable power supplies and fan units provide redundancy and therefore offer a higher availability level. If a fan unit fails, you should not remove it from the storage server until the replacement is available. The cooling airflow will not be at optimum when a fan unit is missing. The same is true for power supplies: do not remove a failed power supply unit before you have a replacement available. If you do, the cooling will not be efficient enough anymore.

RAID controller redundancy is only available on the FASiT200HA Storage Server. RAID controllers are hot-swappable as well, but obviously hot-swap functionality only makes sense in dual-controller configurations.

---

**Important**

If you configure a new storage subsystem with only one RAID controller in the FASiT200 Storage Server, you must install the controller into position A. It will not be recognized in slot B if slot A is not populated. This restriction does not apply to storage subsystems that were originally configured with two controllers.

### 2.1.3 The RAID controller

Each RAID controller contains an Intel 80969RN processor, running at 100MHz and using 128MB of ECC cache. The cache memory is protected by a backup battery. If power fails, the battery will maintain the data in cache for at least 72 hours. Life expectancy of the battery is three years.
**RAID controller connectors**

The RAID controller unit, shown in Figure 13, has the following connectors:

- **Host port GBIC slot**
  
  It is used to attach the FASTT200 to the host system and you may install either a short or long wave GBIC into the slot.

- **Expansion port GBIC slot**
  
  This connector provides storage expandability. Up to ten additional EXP500 storage expansion enclosures can be physically connected, allowing a total of up to 110 disk drives. But, be aware that the maximum configuration will not provide the best performance. Again, you can use either short or long wave GBICs.

- **RJ-45 Ethernet connector**
  
  You use the Ethernet connector for direct-attached management of FASTT200 with FASTT Storage Manager V7.10.

- **TJ-6 RS232 connector**
  
  This connector is used for management and diagnostics through serial connection, using a terminal emulation utility. A dongle is shipped with each FASTT200 system to convert from TJ-6 to DB-9 serial connector.

---

**Please note**

Working with serial interface is potentially risky. With certain commands you can delete your RAID configuration and therefore all data. You should only use it when instructed to do so by IBM level 3 technical support. In this book, we provide only limited instructions on how to use the serial interface. In case of need, you will receive all instructions from technical support.

---

**Figure 13. FASTT200 RAID module**

**RAID controller LEDs**

Several diagnostic LEDs are present on the RAID controller unit:

- **Controller Fault LED (amber)**
  
  This LED is ON during the power-up sequence or when the controller fails. During normal operation it should be OFF.
- Host Loop Link LED (green)
  This LED indicates the host connection loop is operational.
- Cache Active LED (green)
  This LED is ON when the data is in cache.
- Battery Charged LED (green)
  Normally this LED should be ON. If it is OFF, it indicates a battery fault. While the battery is charging, the LED blinks.
- Expansion Port Bypass LED (amber)
  When no additional EXP500 enclosures are connected, this LED will be ON.
- Expansion Loop Link LED (green)
  When you connect additional storage enclosures, this LED will indicate the expansion loop is operational.

Power supply and fan units have additional diagnostic LEDs:
- Fan Fault LED (amber) on fan unit
- Power Fault LED (amber) on power supply unit
- Power Good LED (green) on power supply unit

### 2.1.4 Replacing the cache battery

Each RAID controller unit contains its own cache battery. When the battery expiration date is near, you should replace it. The battery is installed inside RAID controller, so in order to replace it, you will have to remove the RAID controller unit from the FASTT200. If you use only one RAID controller in FASTT200, you will obviously have to shut the host system down. The same will be needed if you use two RAID controllers in non-redundant configuration. Only when you use the RAID controllers in redundant configuration, you can avoid downtime. FASTT200 and FASTT500 are different in this aspect: FASTT500 uses a separate battery module, which can be replaced while both controllers remain operational.

If you use redundant RAID configuration and keep your FASTT200 up and running during the battery replacement, be aware that you will have to complete the replacement procedure within 30 minutes. This is a lot of time and should always suffice. But be aware that when the procedure takes more than 30 minutes, overheating might appear.

Follow these steps to replace the cache battery:

1. If you use two RAID controllers, use the FASTT Storage Manager to locate which controller battery needs to be replaced.
2. If you use redundant configuration, use FASTT Storage Manager to take the controller with the failed battery offline.
   - If two controllers operate in non-redundant mode or only one controller is used, you will have to shut down the host systems connected to the controller with the failed battery.
3. Disconnect FC cables, GBICs, Ethernet and serial cable from the controller.
4. Remove the RAID controller from FASTT200.
5. Undo eight screws that hold the battery cover on the bottom side of the FASTT controller module, then replace the battery and the battery cover.

6. Reinstall the controller module and reconnect GBICs and all cables.

7. Use the label provided with the replacement battery to note the date of replacement and expiration of the new battery.

8. Use the FASTT Storage Manager to bring the controller back online, if you are using two redundant controllers; otherwise, restart the host systems connected to the controller.

9. Use FASTT Storage Manager to reset the battery age. If you are using two controllers, be careful to perform this step for the correct controller. This step is important: you will receive an alert that the battery is nearing its expiration date unless you perform the battery age reset. You may also receive an alert that the battery has failed; this alert will disappear once the battery is fully charged.

2.2 The FASTT500 Storage Server

The FASTT500 Storage Server, displayed in Figure 14, provides a higher level of performance, availability and expandability than FASTT200. It is a 4U rack-mountable device that is suitable for high-end storage needs.

![Figure 14. FASTT500 Storage Server](image)

**Better performance**

FASTT500 uses an AMD K6 processor running at 300MHz in each RAID controller unit. It also supports a larger amount of cache memory than FASTT200. These factors contribute to a much higher throughput than FASTT200 can offer. Additionally, the ability to connect a higher number of hosts directly, without needing a hub or switch, also increases performance. Hubs and switches add their latency delay to FC communication.

**Higher availability**

The following high availability features are standard on FASTT500 Storage Server:

- It contains two hot-swappable RAID controller units.
- Each RAID controller unit can be connected to two host loops through two mini-hubs.
• Each RAID controller unit contains 256MB of battery backed cache. The cache can be expanded to 512MB.
• The drive side FC loops can be connected in a fully redundant way.
• It contains hot-swappable and redundant power supplies and fan units.
• Cache backup battery is hot-swappable.
• With proper FC cabling, you can avoid any single point of failure.

**Better scalability**
You can connect up to 11 EXP500 expansion enclosures to each redundant FC loop on the drive side, which means 110 FC disk drives. Two such fully redundant loops can be used; therefore, you can connect up to 22 EXP500 expansion enclosures. This means you can attach up to 220 disk drives to the FAST500 without a single point of failure.

### 2.2.1 Front view

After removing the front bezel, you can access the following components:

- **Two RAID controller units**
  Each RAID controller unit or blade contains diagnostic LEDs and a reset switch.

- **Controller fan module (which also contains status LEDs)**
  Two fans are integrated in this module. Their role is to provide cooling to both RAID controller blades. FAST500 diagnostic LEDs are also present on the module and there is an alarm switch as well.

- **Battery module**
  It is possible to hot-replace the battery module. During this operation, the cache will temporarily lack battery protection. If write-back cache policy is used and the FAST500 experiences a fault, this might lead to data loss. However, the process of hot-swapping the battery does not take much time and such an event is highly unlikely.

---

**Figure 15. FAST500 front view**
2.2.2 Rear view

The rear view is shown in Figure 16. Components accessible on this side of the FASiT500 include:

- Mini-hubs with GBIC ports for host connectivity
- Mini-hubs with GBIC ports for FC expansion enclosure connectivity
- Fan and communications module
- Two hot-pluggable and redundant power supplies

Management ports

Similar to FASiT200, FASiT500 also features a serial and Ethernet port per each RAID controller unit for management and troubleshooting purposes. These ports are located on the fan and communications module. Besides management ports, this module has two fans that provide redundant cooling to both power supplies.

A diagram of FASiT500 is shown in Figure 17. It explains clearly how the host side and the drive side ports are connected to both RAID controller units.
2.2.3 Host side connections

While FASiT200 only offers one host port and one drive port per RAID controller unit, FASiT500 gives you a much higher flexibility level.

On the host side, each RAID controller is connected to two FC loops utilizing two mini-hubs. You can have a total of four mini-hubs and each mini-hub can take two GBICs. This means you can establish up to eight host connections without needing an external hub or switch. To assure redundancy, the usual way is to install a pair of host adapters inside each server and connect each of the two adapters to a different RAID controller. For example, if one host adapter is connected to mini-hub 1 or 3 (these are connected to RAID controller 1), the other host adapter must be connected to either mini-hub 2 or 4 (RAID controller 2). This way you can connect up to four servers directly to FASiT500. For more servers you will have to resort to using external FC hubs or switches.

2.2.4 Drive side connections

On the drive side, each RAID controller is connected to four FC loops. These four FC loops are normally used as two redundant loops. We use mini-hubs 1 and 2 to form one redundant loop and mini-hubs 3 and 4 to form another. You can connect each redundant loop to up to 11 EXP500 expansion enclosures, which gives you a total of up to 22 enclosures (220 disk drives) on both loop pairs. For more detailed information on drive side connectivity, refer to 2.3.7, “FASiT500 and EXP500 cabling” on page 29.
2.2.5 FASTT500 diagnostic LEDs

As you can see in Figure 18, FASTT500 has three sets of diagnostic LEDs:

- RAID Controller module status LEDs
- Fan module LEDs
- Battery module LEDs

**RAID controller module status LEDs**

The following status LEDs are present on each RAID controller module:

- **Power LED (green)**
- **Fault LED (amber)**
- **Heartbeat LED (green)**
  This LED blinks during normal operation. If it is steady ON, it indicates controller problem
- **Status LEDs (green)**
  These display a binary value of the status code. Heartbeat LED is part of the status code.

**Figure 19. RAID controller module status LEDs**

**FAN module LEDs**

Figure 20 displays the role of status LEDs on the fan module.
**Battery module LEDs**

Battery status LEDs are shown in Figure 21.

**Mini-hub LEDs**

On the rear side of FASTT500, you can install up to eight mini-hubs (four on the host side and four on the drive side) and they have their status LEDs as well (Figure 22).
2.2.6 FASTT200 and FASTT500 comparison

FASTT200 and FASTT500 Storage Servers address different market segments. Table 1 compares their features and capabilities.

Table 1. FASTT200 and FASTT500 comparison

<table>
<thead>
<tr>
<th></th>
<th>FASTT200</th>
<th>FASTT500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host connection</td>
<td>Single GBIC per controller</td>
<td>Two mini-hubs with up to four GBICs per controller</td>
</tr>
<tr>
<td>Drive connection</td>
<td>Single GBIC per controller</td>
<td>Mini-hubs with GBICs</td>
</tr>
<tr>
<td>Battery</td>
<td>Embedded in RAID controller unit</td>
<td>Separate battery module</td>
</tr>
<tr>
<td>Controller status LEDs</td>
<td>None</td>
<td>Status code LEDs on each RAID controller module</td>
</tr>
<tr>
<td>Reset button</td>
<td>None</td>
<td>One per each RAID controller</td>
</tr>
<tr>
<td>Serial port</td>
<td>TJ-6 connector</td>
<td>DB-9 connector</td>
</tr>
<tr>
<td>Maximum performance</td>
<td>Up to 30 disk drives</td>
<td>Up to 90 disk drives</td>
</tr>
<tr>
<td>Maximum supported drives</td>
<td>60 drives (physically 110 possible)</td>
<td>220 drives</td>
</tr>
</tbody>
</table>

Fault LED (amber): indicates mini-hub or GBIC fault

Bypass upper GBIC port (amber): this LED should normally be off. If GBIC is not present, it will be on

Loop Good LED (green): this LED should normally be on

Bypass lower GBIC port (amber), this LED should normally be off, except when GBIC is not present

Figure 22. Mini-hub LEDs
2.3 EXP500 storage expansion enclosure

EXP500 is a rack-mountable storage expansion enclosure that contains ten bays for HH or slim-line hot-swappable FC disk drives. It occupies 3U inside a rack and features hot-pluggable and redundant power supplies and fans. Additionally, it contains two Enclosure Service Monitor (ESM) boards and can be connected to FASTT200 or FASTT500 in a fully redundant FC loop.

---

**Important**
You should never disconnect and remove any hot-swap component (power supply, fan, disk drive or ESM board) from EXP500 for a long period of time. This will reduce cooling efficiency, because it will affect the airflow inside the enclosure. Should any of the above components fail, leave it connected and only replace it when the replacement part is available.

---

2.3.1 Front view

EXP500 is physically very similar to FASTT200. The difference is it contains two ESM (Enclosure Services Monitor) boards instead of RAID controller units.

From the front, the EXP500 expansion enclosure looks identical to FASTT200 (Figure 23). All the diagnostic LEDs at the front have the same meaning as on FASTT200. See 2.1.1, “Front view” on page 15 for details.

![Figure 23. EXP500, front view](image)

2.3.2 Rear view

The rear view shows some similarities and some differences to FASTT200 (Figure 24).

There is a pair of redundant and hot-swappable power supplies and fans. They function in the same way as in FASTT200. As mentioned, instead of RAID controller units, the EXP500 contains a pair of ESM boards. You can also see an incoming and outgoing GBIC port on each ESM board.
Figure 25 shows diagnostic LEDs present on each ESM board. They have the following roles:

- **Input Bypass LED (amber)**
  The LED indicates no valid signal is detected on input port. When FC cables are connected to the input port, the LED should be off. If it is on, check your loop connection.

- **Power LED (green)**
  This LED should be on and it indicates power is supplied to the ESM board.

- **Fault LED (amber)**
  It indicates ESM board failure and should remain off during normal operation.

- **Over-temperature LED (amber)**
  This LED should be off during normal operation. When on, it indicates the EXP500 is overheating. A likely reason would be failure of both fan units, operating the EXP500 without a power supply or a fan module in place, or using it in an improper environment that does not meet required thermal specifications.

- **ID Conflict LED (amber)**
  If tray IDs are set differently on the two ESM boards, this LED will be on. In such a case, the tray ID of the left ESM board is used. But be aware that this will cause a problem if this ESM board fails: the other ESM board will not be able to provide access to the drives due to tray ID mismatch. You should therefore avoid this situation. Tray ID setting is explained in 2.3.5, “Tray ID switches” on page 28.

- **Output Bypass LED (amber)**
  If no valid signal is detected on output port, this LED will be on.
2.3.3 EXP500 diagram

As you can see in Figure 26, both ESM boards are connected to all ten disk drive bays in the EXP500. This provides redundancy in case of a failed ESM board or faulty loop connection: all disk drives remain connected through the other ESM board or FC loop. Each ESM board has an incoming and outgoing GBIC port.

2.3.4 Incoming and outgoing GBIC ports

One GBIC port is marked as IN, the other one OUT. A close look at Figure 26 reveals both ports actually function identically, so it is not mandatory to stick to their incoming or outgoing role. However, we recommend that you use them in the suggested manner. If you always connect outgoing ports on FASTT200 or FASTT500 to incoming ports on EXP500, you will introduce clarity and consistency to your cabling implementation and this will allow for much easier, quicker and more efficient troubleshooting.

2.3.5 Tray ID switches

It is very important to correctly set the tray ID switches on ESM boards. They are used to differentiate between multiple EXP500 enclosures that are connected to the same FASTT200 or FASTT500. Both ESM boards in the EXP500 must be set to the same tray ID value, but each EXP500 within the FC loop must use a unique
value. The FASTT Storage Manager utility uses the tray IDs to identify each EXP500 enclosure. Additionally, the FC loop ID for each disk drive is automatically set according to:

- The EXP500 bay where the disk drive is inserted
- Tray ID setting

Two rotary switches are available to set the tray ID:

- A switch for tens (x10)
- A switch for ones (x1)

You can therefore set any number between 0 and 99.

### 2.3.6 FASTT200 and EXP500 cabling

FASTT200 HA Storage Server with two RAID controller units allows you to connect up to 10 EXP500 enclosures in a redundant loop. It is possible to install up to 110 disk drives: 10 in each enclosure, including FASTT200HA. If you use FASTT200 Storage Server with only one RAID controller unit, you can still connect the same number of EXP500 enclosures, but there is no loop redundancy. An example in Figure 27 shows redundant loop configuration with two additional EXP500 storage enclosures connected to FASTT200HA.

![Figure 27. FASTT200HA and EXP500 cabling](image)

---

**Please note**

Although the paragraph above lists 110 disk drives as maximum, officially up to 60 are supported on FASTT200. It is true that you can physically connect up to 110 drives, but for performance reasons it is not sensible to use more than the supported number. With higher number, the RAID controller becomes a performance bottleneck.

### 2.3.7 FASTT500 and EXP500 cabling

FASTT500 RAID Controller offers more flexibility and scalability. You can connect up to 11 EXP500 enclosures in each redundant loop, so that gives you a maximum of 22 enclosures.

Figure 28 shows four EXP500 enclosures connected to FASTT500 through two fully redundant FC loops. The two EXP500s in the upper part are connected to mini-hubs 1 and 2 and the lower two EXP500s are connected to mini-hubs 3 and 4.
2.4 FASTT host adapter

This is a 64-bit PCI adapter, capable of running at 66MHz. However, you can also use it in a 32-bit PCI slot and it will also run at 33MHz. This is important as there are many server systems on the market that only support 33MHz PCI bus speeds.

Figure 29. FASTT host adapter

It is supported in the following operating systems:
- Windows NT 4.0

Important!

Please note that on FASTT500 you only use one GBIC port on each drive side mini-hub: either the incoming or outgoing one (whichever is consistent with your cabling strategy), but not both.

Figure 28. FASTT500 and EXP500 cabling
• Windows 2000 Server and Advanced Server
• Novell NetWare 4.2 and 5.1

For higher level of availability, the adapter supports Active PCI functionality. In a Windows 2000 environment, you can configure two adapters as a fault-tolerant pair. When one fails and another takes over, the failed adapter can be replaced without bringing the server down. Hot-add functionality is supported as well; it can further increase system availability. Using hot-add, you can avoid system downtime when adding new adapters.

The adapter supports short-wave fibre cabling. If you wish to use long-wave cables in your FC network, you have to use a FC hub or a switch with long-wave GBIC. The connection between the adapter and the hub or switch must be a short-wave one.

2.4.1 Fast!UTIL

At boot time, you can press Alt+Q to enter the Fast!UTIL configuration utility. The utility is stored in BIOS of the adapter and allows you to set various adapter parameters, scan for devices on the FC loop and format the disk drives. You should understand the parameters very well before trying to change them, because improper values can cause the adapter to not function properly. Be aware that these parameters and their default values might change with future firmware releases.

Host adapter settings
This selection appears in the Configuration settings menu and allows you to check and change some basic host adapter settings.

• Host adapter BIOS
  By default, the Host adapter BIOS is disabled. In most cases, the adapter BIOS is not needed at all. If it is disabled, this will free some memory space in upper memory and make it available to other adapters installed in the host system. You do need to enable the BIOS if you wish to boot from FC disk drives attached to this particular host adapter.

• Frame size
  This parameter specifies the maximum frame length the adapter will support. The value can be 512, 1024 or 2048 and the default size is 2048.

• Loop reset delay
  With this setting, you specify the period of loop inactivity after a loop reset. The range is 0 to 15 seconds and 5 seconds is the default value.

• Adapter hard loop ID
  You can enable or disable the hard loop ID with this parameter. It is enabled by default.

• Hard loop ID value
  When you enable the hard loop ID (see parameter above), this parameter sets the value. It can range from 0 to 125, and 125 is the default.

Selectable boot settings
This option is available in the Configuration settings menu. When you enable selectable boot, you can then use the Boot ID and Boot LUN settings to identify
the bootable LUN. It is not very usual to boot from the Storage Server, therefore, these settings will rarely be used.

**Adapter hard ID settings**

If you need to set the hard ID for the host adapter, use this option from the Configuration settings menu. Available range is between 0 and 125.

**Restore default settings**

The meaning of this option is obvious: it resets the host adapter settings to their default values. This would typically be used when performing a new installation of previously used adapter and possibly in troubleshooting scenarios.

**Raw NVRAM data**

For troubleshooting purposes, it can be useful to see the NVRAM settings of the adapter. This option displays the host adapter NVRAM contents in hexadecimal format. Note: you can only read the values, but not modify them.

**Advanced adapter settings**

This is another item on the Configuration settings menu. With these parameters, you can tune the behavior and performance of host adapter.

- Execution throttle
  
  This parameter sets the maximum number of commands that can execute on any port. When the maximum is reached, no new commands are being issued until the current one finishes. The valid range is from 1 to 256 and default value is 16.

- Fast command posting
  
  When this parameter is enabled, it will minimize the number of interrupts and therefore speed up command execution. The default is enabled.

- >4GB addressing
  
  This parameter has to be enabled when the host system uses more than 4GB of memory. Be aware it is disabled by default; do not forget to change its value when required.

- LUNs per target
  
  Multiple LUN support is usually needed for a RAID environment. When you use FAST Storage Server with more than eight LUNs, set this parameter to 0.

- Enable LIP reset
  
  This parameter sets the type of LIP (loop initialization process) reset. If it is set to Yes, then a global LIP reset occurs to clear the target device reservations. When it is set to No, a global LIP reset occurs with full login. The default value is No.

- Enable LIP full login
  
  This parameter is enabled by default and it causes login to all ports after each LIP.

- Enable target reset
  
  With this setting you can enable or disable a Target Reset command to all devices on the loop when a SCSI Bus Reset command is encountered. This parameter is enabled by default.
• Login retry count
  This setting can have a value between 0 and 255, and the default is 30.

• Port down retry count
  This retry count can also have a value from 0 to 255, and the default is 30 as well.

• Drivers load RISC code
  If this parameter is enabled, the host adapter uses the RISC firmware embedded in the device driver. This ensures a valid and certified combination of device driver and RISC firmware. But the driver must obviously support this setting, otherwise it works as if this option was disabled.

• Enable database updates
  You can use this option to save the loop configuration information in the flash memory at each power down. The default value is No.

• Disable database load
  This setting can be used in a Windows NT and 2000 environment. When enabled, the device database will be read from registry every time the device driver initializes. If disabled, the database will always be created dynamically during driver initialization. It is disabled by default.

• IOCB allocation
  This setting controls the maximum number of buffers from the firmware buffer pool, allocated to any port. It can range from 1 to 512, and 256 is the default.

• Extended error logging
  This is another parameter that you can use in a Windows NT or 2000 environment. It provides additional error logging to the operating system. If enabled, the Windows NT or 2000 Event Viewer will receive events from host adapter. The parameter is disabled by default.

**Extended firmware settings**

You can select this menu option from the Configuration settings menu. It contains additional firmware-related parameters.

• Extended control block
  With this parameter you either enable or disable all other extended firmware settings, as they are described below. They are enabled by default.

• RIO operation mode
  This setting controls the use of Reduced Interrupt Operation (RIO) mode. RIO mode allows the posting of multiple command completions in a single interrupt and it must be supported by the device driver in order to be functional. There are five possible values for this parameter, 0 to 4:
  • 0 — no multiple responses
  • 1 — multiple responses, 16-bit handles, interrupt host
  • 2 — multiple responses, 32-bit handles, interrupt host
  • 3 — multiple responses, 16-bit handles, delay host interrupt
  • 4 — multiple responses, 32-bit handles, delay host interrupt
  The default value for this parameter is 0.
• Connection options
With this parameter, you can define the connection type and preference. The following values are allowed:
• 0 — loop connection only
• 1 — point-to-point connection only
• 2 — loop connection is preferred; point-to-point allowed
• 3 — point-to-point connection is preferred, loop allowed
3 is the default value.

• Nonparticipating hard ID
Leave this setting at default, which is disabled. It is a reserved setting.

• Class 2 service
If you enable this option, Class 2 service parameters are provided during loop port automatic logins. It is disabled by default.

• ACK0
This parameter will determine the type of acknowledgements. When enabled, the sequence acknowledgement is used; when disabled, it is frame acknowledgement. The default value is disabled and this option will only be used when the Class 2 service option is enabled.

• Fibre Channel tape support
The default value for this parameter is disabled and it is reserved for Fibre Channel tape support.

• Fibre Channel confirm
Same as above, this option is reserved for FC tape support. It is disabled by default.

• Command reference number
This is another setting reserved for tape support. It is also disabled by default.

• Read transfer ready
Do not modify this setting. It is reserved and should be set to disabled.

• Response timer
With this setting, you can limit the wait time for accumulating multiple responses. The wait time is the value multiplied by 100 microseconds. 0 is the default value for this parameter.

• Interrupt delay timer
This parameter is used to specify the wait time between accessing a set of handles and generating an interrupt. The default is 0. Again, the wait time is the value multiplied by 100 microseconds.

Scan Fibre Channel devices
This selection will scan the FC loop and it will list all the connected devices, providing information about each device found. You can use this option to verify FC loop connectivity at installation time or when performing troubleshooting.
2.5 Hubs and switches

Using redundant cabling, you can connect only a single host system to the FASTI200 Storage Server directly. For a higher number of hosts and for more complex SAN configurations, you have to use a FC hub or a switch. For entry level environments, a hub is usually sufficient and much more cost-effective than a switch. However, for complex SAN configurations, a switch may be necessary, because it provides more flexibility and higher throughput.

2.5.1 IBM SAN Fibre Channel Managed Hub (3534)

The IBM SAN Fibre Channel Managed Hub can be used in various FC configurations, such as:

- High-availability clustering
- Storage consolidation
- LAN-free backup
- Remote mirroring

![IBM SAN Fibre Channel Managed Hub](image)

It contains seven fixed short-wave FC ports and one slot for additional short-wave or long-wave GBIC. The hub utilizes non-blocking switch technology and provides simultaneous 100 MB/s throughput for all ports. Arbitrated loop can be formed by connecting all eight ports on the hub into a single loop, or the ports can be zoned into several arbitrated loops.

You can cascade two managed hubs and this will give you up to 14 ports on your FC loop. It is also possible to interconnect the managed hub with IBM SAN Fibre Channel Switch. When using the managed hub together with a switch, be aware that all host systems must be connected to the switch. Only storage devices may be connected to the hub in such a scenario.

The managed hub supports industry standard MIBs and therefore allows standard SNMP management.

To configure, manage and troubleshoot the managed hub, use the StorWatch FC Managed Hub Specialist software. This utility works via a Web browser from a workstation over an Ethernet network connection to the managed hub. StorWatch utility also provides SNMP messages, traps, and MIBs that can be integrated into an existing enterprise management structure.
2.5.2 IBM SAN Fibre Channel 8 and 16-port Switch (2109)

You can use the IBM SAN Fibre Channel Switches to interconnect multiple host systems with many storage servers and other SAN devices. They can provide the throughput required for large SAN configurations. Two versions exist: with 8 and 16 ports. Each switch port provides bandwidth of up to 100MB/s with maximum latency of 2 microseconds and non-blocking architecture allows multiple simultaneous connections. Switch ports can operate in either of these modes: F, FL or E.

The switch comes with four short-wave GBICs: you can install either short or long-wave GBICs into additional switch ports.

If you cascade several switches, you can achieve complex SAN connections with a massive number of switch ports. You can also increase the distance by connecting the switches in series. Up to seven hops are allowed and this means up to 70km if you use long-wave GBICs and cables.

The switches are self-learning: this feature allows the Fabric to automatically discover and register host and storage devices. Another important capability is self-healing which enables the Fabric to isolate a problem port and reroute traffic onto alternate paths.

For performance and flexibility, an internal processor provides services such as name serving, zoning and routing. For availability, you can add an optional second hot-plug power supply to provide redundant power, so that you can avoid downtime when a power supply fails. You can also perform dynamic microcode upgrades.

To configure the FC switch, use the StorWatch SAN Fibre channel Switch Specialist management tool. The switch features embedded Web browser interface for configuration, management and troubleshooting.
Chapter 3. The FAStT Storage Manager software

The FAStT Storage Manager software is used to configure, manage and troubleshoot the FAStT Storage Servers. It is a Java-based GUI utility that is available for various operating systems:

- Windows NT and 2000
- Linux
- Novell NetWare
- HP-UX
- Sun Solaris

You can use the FAStT Storage Manager software to create new arrays and logical drives, expand existing arrays, migrate to a different RAID level, configure storage partitioning and also perform diagnostic and troubleshooting tasks.

What is new?
The latest revision of FAStT Storage Manager software at the time of writing is V7.10 and it introduces several important enhancements:

- Event monitoring
  This is a very useful addition to the Storage Manager software. It allows you to receive alert notifications on critical Storage Server events even when the FAStT Storage Manager Client is not running.

- Heterogeneous hosts support
  In previous versions, the FAStT Storage Server could only be accessed by host systems which all ran the same operating system. This is no longer the case. Now hosts running different operating systems can use the same Storage Server. Currently supported operating systems include:
    - Windows NT 4.0 and 2000
    - Novell NetWare
    - Linux
    - Sun Solaris
    - HP-UX

  AIX support is scheduled for 2001.

  Heterogeneous hosts support works in conjunction with storage partitioning. All the host systems in a particular storage partition must run the same operating system. But the operating systems in different storage partitions can vary.

- Configuration replication
  You can save configuration of a FAStT Storage Server into a file and then import it on another Storage Server with identical hardware setup. This way, you can quickly create replicated configuration of similar systems.

- Auto-Volume Transfer
  Previous versions of software only allowed the controllers to own the whole arrays. Now ownership is based on a logical drive level, which introduces a much higher level of flexibility. It also means that now also individual logical drives can be transferred from one controller to another, not just entire arrays.
• Command line support

Command Line Interface support is based on script engine commands. Up to now, you had to use the FASiT Storage Manager Client to run scripts. With V7.10, you can run scripts from the operating system command prompt and you can also use these commands in batch files.

• Controller diagnostics

The Storage Manager Client contains a Controller menu on the menu bar. Certain built-in controller diagnostic tests are available, like testing the FC loops on host and drive side.

• Access logical drive mapping allowed

Access logical drive is required to allow host-attached management. The default LUN assigned to this logical drive is 7 and it is now possible to change the LUN assignment to another value. This is usually not needed, but you might still want to do it in certain cases, for example to achieve consistent LUN numbering.

3.1 Direct-attached and host-attached management

The managing machine onto which you install the FASiT Storage Manager can be a dedicated workstation or possibly a host system itself. We usually use the dedicated management workstation for this purpose. To be able to use the FASiT Storage Manager, you must establish a connection between that machine and the FASiT Storage Server. Two options are available:

• Direct-attached management

The managing workstation is attached to a dedicated Ethernet port on each RAID controller unit.

• Host-attached management

We manage the Storage Server through the FC host connection.

These two management methods are not mutually exclusive. You can be connected to a particular Storage Server directly through Ethernet and also through FC host connection at the same time.

Direct-attached management

Each RAID controller unit has its own Ethernet port. In order to manage both controllers, you must obviously provide connections to both Ethernet ports. The usual way is to connect the management workstation and both Ethernet ports to a hub. It is not a bad idea to use a dedicated hub, separate from production network segments — this will increase security. Figure 32 shows an example of direct-attached management.

The only component you must install on the managing workstation in this case is the FASiT Storage Manager Client.

The communication protocol is TCP/IP and each RAID controller must have a valid IP address. The addresses can be provided either by a DHCP/BOOTP server or you can manually set static IP addresses. To set static IP addresses, you will have to connect to the serial management port and use the terminal emulation session. The procedure is described in 4.4.13.2, “Assigning fixed IP addresses” on page 126.
The direct-attached method does not require an Access logical drive; therefore, all LUNs can be assigned to data logical drives. Another very useful benefit is that you can manage the FASTT Storage Server connected to the host systems which run operating systems not supported by the FASTT Storage Manager.

**Host-attached management**

In this case, we utilize the FC connection between the host system and the Storage Server for management. You need the FASTT Storage Manager Agent to enable management through the FC connection and you must install it to the host system. There are two options where to run the Storage Manager Client:

- On the host system itself.
- On another computer, connected to the host system through the LAN. This is usually the preferred option.

Figure 33 displays an example of host-attached management.

Be aware that in this management scenario, the Storage Manager software requires an Access logical drive to communicate with RAID controllers. A LUN
has to be assigned to the Access logical drive and this means you have one LUN less available for the data logical drive mappings.

With this method, there is obviously no need to configure the network access to the RAID controller Ethernet ports: you do not need to set the IP address manually nor use the DHCP/BOOTP server.

3.2 The FASiT Storage Manager Client

The FASiT Storage Manager Client uses two main window types to give you control over your Storage Servers:

- The Enterprise Management Window
- The Subsystem Management Window

The Enterprise Management window (EMW)

This window appears when you start the FASiT Storage Manager Client. It displays a list of all FASiT Storage Servers that the client can access either directly or through the host agents. If you can access a certain FASiT in both ways and possibly through several host agents, you will see it listed not just once, but many times in the EMW. The utility can automatically detect new Storage Servers or you can add them to the EMW manually. The name, status and management type (through the Ethernet or through the host agent) are displayed for each listed Storage Server and you can perform various tasks, like executing scripts, configuring alert notification destinations or selecting a particular Storage Server (or Subsystem) you wish to manage.

Figure 34 displays the Enterprise Management window.

![Figure 34. The Enterprise Management window](image)

The Subsystem Management window

Once you select a system you wish to manage in the EMW, the Subsystem Management window for that particular system opens. As you can see in Figure 35, this window has two parts:
• Logical view

This is a tree-like structure that shows all arrays and logical drives configured on the Storage Server. You can select any array or logical drive object in the tree and perform various tasks on it.

• Physical view

This view shows the RAID controllers installed in the Storage Server. It also displays the physical disk drives in all attached storage enclosures. The controllers and disk drives are selectable as well which enables you to perform the tasks on them.

Figure 35. The Subsystem Management window

A specific Subsystem Management window allows you to manage one particular Storage Server, but you can have multiple such windows open at any time.

For a detailed explanation of FAStT Storage Manager Client usage, refer to 4.2, “Configuring the FAStT” on page 74.

3.3 Event monitor

Whenever the Enterprise Management Window (EMW) is running, it performs monitoring of Storage Servers. In FAStT Storage Manager, versions up to V7.02, it was necessary to keep the EMW operational at all times for monitoring the function. If EMW was closed, the monitoring of Storage Servers would stop, too.

The Storage Manager V7.10 introduces a new component: the Event monitor, which provides Storage Server monitoring functions when the EMW is not active. The Event monitor performs Storage Server monitoring and in case of any critical events it sends alerts via e-mail or SNMP. Monitoring is now performed either by EMW or, when EMW does not run, by the Event monitor.

You cannot install the Event monitor separately: it can only be installed on a managing workstation as a part of the whole FAStT Storage Manager Client package. Once installed, it runs in the background and checks for possible critical problems. If it detects a problem, it will notify a remote system through e-mail or
SNMP or both. Use the EMW (Edit -> Alert Destinations) to configure alert actions of the Event monitor. Be aware that when you monitor a FASiT Storage Server from multiple management systems, each running its own instance of the Storage Manager Client or Event monitor, then each Event monitor will send its own alert to the configured destination when a critical event occurs. In such a case, the destination system will receive several alert notifications for a single FASiT Storage Server event.

Event monitor settings are stored in the EMWDATA.BIN file on the managing system. This file contains the monitored Storage Server name and the e-mail and SNMP alert settings. You must configure a mail server if you wish to use the e-mail alerting. You can do this in the EMW by selecting Edit -> Configure Mail Server.

If you wish to use identical alert settings on multiple managing systems, you can simply copy the EMWDATA.BIN file from one system to another. Remember to restart the Event monitor and EMW after copying the file in order to activate the changes. The location of EMWDATA.BIN differs depending on the operating system.

The changes you make in the EMW (like adding new Storage Servers or altering monitoring settings for existing ones) are not automatically reflected in the Event monitor. You have to manually synchronize the EMW and Event monitor. Use the Event monitor button in EMW to do the synchronization. The button, shown in Figure 36, is only active when synchronization is needed, it will be unavailable at all other times.

For a detailed information on Event Monitor, refer to 4.2.9, “Configuring, monitoring and alerting” on page 95.

### 3.4 Basic management

In this section we describe some common administration concepts like configuring new storage space, adding more capacity to the existing arrays, changing the RAID levels, altering the segment size and tuning the cache settings.
3.4.1 Creating arrays and logical drives

This is one of the most basic steps. Before you can start using the physical disk space, you must configure it: divide your disk drives into RAID arrays and create one or more logical drives inside each array. An example is shown in Figure 37.

As you can see, we have grouped 12 physical disk drives into four arrays:

- Array 1 contains four physical disks and we have divided it into two logical drives. We use the RAID level 5.
- Array 2 uses RAID 1 and it consists of two physical drives. It is divided into two logical drives.
- Array 3 uses two physical disk drives as well and it contains one single logical drive. We use RAID 0, which does not provide any redundancy.
- Array 4 spans across four physical disk drives. We have selected RAID 1 for this array; be aware that this is actually RAID 10. It combines mirroring and striping. We use three logical drives in this array.

RAID levels

When creating arrays, you have to specify the RAID level you plan to use. Each RAID level has its benefits and limitations. You should select the level that best meets requirements for data availability and performance in your particular case.

For example, RAID 0 offers the highest performance, but on the other hand does not provide any protection against a disk drive failure. So you should only use RAID 0 to store the data that requires fast access and can be restored easily if a disk drive fails. If you cannot tolerate downtime, do not use RAID 0.

RAID 5 on the other hand offers good fault tolerance and efficient use of disk space, but generally provides lower performance. Again, this depends on your particular scenario: in read-intensive environments, RAID 5 will usually perform better than in write-intensive cases. Write performance of RAID 5 arrays is usually not the best, due to the fact that the controller must issue a couple of reads before it can write the data. But this can be improved by efficient caching.

It is also important whether the workload is random or sequential. If you need good RAID 5 performance, it is a good idea to spread the array across many disk drives.
RAID 5 uses parity for protection of the data. When a disk drive fails, the data within RAID 5 array is obviously still accessible. But reading performance decreases, as the data that resided on the failed disk drive must now be calculated from corresponding data blocks on all other disks. Instead of reading just one data block, the controller needs to read a block from each surviving disk drive.

The FAST Storage Manager also allows you to configure RAID 1. Traditionally, RAID 1 means mirroring of all data between exactly two disk drives. But on the FAST Storage Server, you are not limited to just two disk drives for RAID 1. You can use any even number up to 30. This implementation is actually called RAID 10 and it combines data striping across mirrored pairs of disks.

RAID 10 offers both fault tolerance and very good performance, but the disk space efficiency is lower than with RAID 5. When the efficient usage of physical disk space is not an issue, RAID 10 is usually the preferred selection. It also performs well with a failed disk drive, because it uses mirroring for data protection. The data blocks that were stored on the failed disk drive can simply be read from the mirrored location. This is generally faster than RAID 5, where the controller must read from all surviving disk drives and then calculate the missing data.

Another RAID level is supported: RAID 3. It is similar to RAID 5 as it uses parity to protect the data, but it is more suitable for large data transfers than RAID 5.

You may want to configure several arrays that use different RAID levels on the same FAST Storage Server to accommodate various types of data with different performance and availability requirements.

For a detailed explanation of all supported RAID levels, refer to 5.3.2, “RAID level and performance” on page 136.

**Hot-spare drives**

Hot-spare disk drives provide additional protection that might prove to be essential in case of a disk drive fault. Hot-spare drive is like a replacement drive installed in advance. The data from the failed disk drive will automatically be rebuilt to the hot-spare when one exists. Hot-spare drives are global: they can replace a disk drive in any array on the Storage Server. When assigning disks as hot-spare, make sure they have enough storage capacity. If the failed disk drive is larger than the hot-spare, reconstruction will obviously not be possible.

4.2.7, “Define hot spare drives” on page 88, explains how to work with the hot-spare disk drives.

### 3.4.2 Expanding arrays

The ability to increase array size without needing to restart the host system is a very important feature. In today's IT environment, the need for storage space grows constantly. Many customers will exhaust their existing space sooner or later and will have to expand their storage capacity. So it is essential that this process is non-disruptive and does not cause any downtime.

With FAST Storage Manager, you can simply add new disk drives to the Storage Server and start the expansion procedure while the system remains fully operational. Be aware that you may see some performance impact as the
expansion process will compete with normal disk access. You can use this free capacity to create additional logical drives. Existing logical drives in the array do not increase in size as a result of this operation. Once logical drives have been created it is not possible to increase their capacity.

When expanding RAID 1 arrays, you need to add disk drives in pairs since a RAID 1 array must have an even number of drives.

It is not possible to use more than 30 disk drives in one array. Once the maximum number of drives is reached, you obviously cannot add new drives anymore.

The procedure will reorganize the data segments in the array and spread them across all drives now contained in the array, as you can see in Figure 38. As this requires a huge amount of I/O to be performed there will be an impact on performance while the expansion procedure runs. You can influence the performance impact by changing the value of the modification priority. This parameter is set per logical drive and you must change it for all logical drives in the array. The higher modification priority means the shorter migration time, but the performance impact will obviously be higher. You may want to change the modification priority to a low value to minimize performance degradation. When the migration finishes you should change the value to a higher one again to reduce the time for a rebuild in the case of a drive failure. This will minimize the critical time of non-redundant operation caused by the disk drive fault. Section 4.4.3.1, “Logical drive properties” on page 103, explains how to change the modification priority.

Once the procedure starts you cannot stop it anymore.

In Figure 38 a drive is added to convert a three-drive RAID 5 array to a four-drive one. The capacity of the new disk drive will increase the free capacity within the array.

![Figure 38. Adding new drives to an array](image)

### 3.4.3 Migrating RAID levels

Changing the RAID level of an array can also be performed in a non-disruptive manner. The system remains fully operational while the process takes place. A few possible reasons why customers may want to do this operation are listed here:
The storage requirements have changed over time and existing RAID levels are no longer optimal for a particular environment.

The performance tuning process has indicated that a different RAID level would be more appropriate than the existing one.

You can change any RAID level to any other one. Be aware there are some restrictions that apply also to the new arrays:

- RAID 1 or 10 requires an even number of disk drives.
- RAID 3 and 5 require at least three drives.
- There is a limit of 30 drives per array.

There are limitations if you do not have enough free space in the array. For example, a RAID 5 array of four disk drives with no free space cannot be migrated directly to RAID level 1. If you start this migration you will receive an error message stating that you do not have enough free space. You need to add new drives to the array first to increase the free capacity and then you can change the RAID level. Also if the array has an odd number of drives and you want to migrate to RAID level 1, you must add a drive first to have an even number.

By doing the opposite, changing from RAID level 1 to 5, you will gain free space in the array which can be used to define new logical drives.

Starting this procedure is simple. Use FASIT Storage Manager to select your array and perform the RAID level migration. For details, refer to 4.4.3.3, “Migrating RAID level” on page 106.

Once the procedure starts it will reorganize the data segments in the array according to the new RAID level. As this requires a huge amount of I/O to be performed there will be an impact on performance while the migration lasts. You can influence the performance impact by changing the value of the modification priority. This parameter is set on a logical drive basis and you should change it for all logical drives in the array. The higher modification priority means the shorter migration time, but the performance impact will be higher. You may change the modification priority to a low value during the migration process to minimize performance degradation. When the migration finishes you should change the value to a higher one again to reduce the time for a rebuild in the case of a drive failure. This will minimize the critical time of non-redundant operation caused by the disk drive fault. Section 4.4.3.1, “Logical drive properties” on page 103, explains how to change the modification priority.

Once the migration starts it cannot be stopped anymore.

Figure 39 shows the process of changing the RAID level of an array from 1 to 5. The data segments were mirrored between the disk drives in the RAID level 1 array. After the migration, the data and parity segments are striped across the drives. Since RAID 5 has lower overhead than RAID 1, the amount of free space within the array will increase.
3.4.4 Changing the segment size

The RAID controllers read and write data to and from the physical disk drives in fixed-sized segments, before writing data on the next physical drive. A segment is the amount of data, in kilobytes, that the controller writes on a single drive in a logical drive before writing data on the next drive. Data blocks store 512 bytes of data and are the smallest units of storage. The size of a segment determines how many data blocks it contains. For example, an 8 KB segment holds 16 data blocks and a 64 KB segment holds 128 data blocks. You can use the following values for the segment size: 8K, 16K, 32K, 64K, 128K or 256K. The segment size is specified per each logical drive.

**Note:** The segment size was expressed in number of data blocks in previous versions of this storage management software. It is now expressed in kilobytes.

You should choose the segment size very carefully, because it can have a large impact on performance.

A large segment size (relative to the average request size) increases the request rate by allowing multiple disk drives to respond to multiple requests. If one disk drive contains all of the data for one request, the other disk drives in the storage set are available to handle other requests. Thus, in principle, separate I/O requests can be handled in parallel, increasing the request rate.

A small segment size (relative to the average request size) increases the data transfer rate by allowing multiple disk drives to participate in one I/O request. Sequential write and read requests should use a small segment size relative to the I/O size to increase performance.

To determine the optimal segment size, you will often have to go through performance measurement and tuning tasks. This might mean the Storage Server will have to run in production for some time to give you an idea of data access patterns. Once the best value for the segment size is determined, you can use the FASTT Storage Manager to change it to this value. This process is again non-disruptive. No downtime is needed.
3.4.5 Cache parameters

Efficient use of the RAID controller cache is essential for good performance of the FASTT Storage Server. FASTT Storage Manager utility enables you to configure various cache settings:

- Start and stop cache flushing levels
- Cache block size
- Enabling or disabling write cache mirroring
- Cache Read-Ahead multiplier
- Write-Back and Write-Through mode

These settings have a large impact not only on performance of the FASTT Storage Server but also on the availability of data. Be aware that performance and availability are often in conflict with each other. If you want to achieve maximum performance, you will in most cases have to sacrifice system availability and vice versa. Most customers are usually more motivated in the highest availability they can get, rather than the maximum performance.

Write-Back and Write-Through

If you configure Write-Through caching, the cache will only increase performance of reading operations. Writing operations will not use cache at all, the data is always going to be written directly to the disk drives.

Write-Back on the other hand can also increase performance of write operations. The data is not written straight to the disk drives, it is only changed in the cache. From the application perspective, this is much faster than waiting for the disk write operation to complete. You can therefore expect significant gain in application writing performance. It is the responsibility of the cache controller to eventually flush the unwritten cache entries to the disk drives. When using Write-Back mode, you can expect to have some unwritten data in the cache anytime. If FASTT loses power, this data may be lost. Therefore, the backup battery for cache is essential in this case. FASTT Storage Servers do have cache batteries, but it is vital that you keep track of battery expiration dates and replace them in time. Be aware of the following limitations during the battery replacement procedure:

- The cache memory in FASTT500 is not protected during the battery replacement procedure.
- On FASTT200, you must remove the whole RAID controller to replace its battery. In this case, the FASTT200 only operates on one RAID controller while replacing the battery and therefore temporarily loses controller redundancy. If the FASTT200 only uses one RAID controller, the downtime will be unavoidable.

Write-Back mode appears to be faster than Write-Through, since it does not only increase performance of reads, but also writes. But this way of thinking can be very wrong! It is very dependant on disk access pattern and workload. A lightly loaded disk subsystem will usually work faster in Write-back mode. But when the workload is high, the write cache may become inefficient. As soon as the data is written to the cache, it has to be flushed to the disks to make room for new data arriving into cache. The controller would actually perform faster if the data went directly to the disks. In this case, writing the data to the cache is an unneeded step that actually decreases throughput.
You can use the FASTT Storage Manager Client to set the appropriate caching mode for arrays and logical drives.

**Start and stop cache flushing levels**
These two settings affect the way how cache controller handles unwritten cache entries. They are obviously only effective when you configure Write-Back cache policy. Writing the unwritten cache entries to the disk drives is called flushing. You can configure the start and stop flushing level values. They are expressed as percentages of the entire cache capacity. When the number of unwritten cache entries reaches the start flushing value, the controller begins to flush the cache (write the entries to the disk drives). The flushing stops when the number of unwritten entries drops below the stop flush value. The controller always flushes the oldest cache entries first. Unwritten cache entries older than 20 seconds will be flushed automatically.

A typical start flushing level would be 80%. Very often the stop flushing level is set to 80%, too. This means the cache controller will not allow more than 80% of the entire cache size for Write-Back cache, but it will also try to keep as much of it as possible for this purpose. If you use such settings, you can expect a high amount of unwritten entries in the cache. This will be good for writing performance, but be aware that it offers less data protection.

If you are concerned for data protection, you may want to use lower start and stop values. With these two parameters, you can actually tune your cache for either reading or writing performance.

Performance tests have shown that it is a good idea to use similar values for start and stop flushing levels. If stop level value was significantly lower than the start value, this would cause a high amount of disk traffic when flushing the cache. This would interfere with normal disk access and would usually not even be needed. If the values are similar, then the controller only flushes the amount needed to stay within limits.

**Cache block size**
This is the size of cache memory allocation unit and it can be either 4K or 16K. By selecting the proper value for your particular situation, you can significantly improve the caching efficiency and performance. For example, if applications mostly access the data in small blocks up to 8K, but you use 16K for cache block size, then each cache entry block will only be partially populated. You will always occupy 16K in cache to store 8K (or less) of data. This means only up to 50% of cache capacity will effectively be used to store the data. You can obviously expect lower performance. For random workloads and small data transfer sizes, 4K is better.

On the other hand, if the workload is sequential and you use large segment size, it is a good idea to use larger cache block size, 16K. Larger block size means lower number of cache blocks and reduces cache overhead delays. Additionally, larger cache block size will require fewer cache data transfers to handle the same amount of data.

**Write cache mirroring**
FASTT cache mirroring can only be used when the RAID controllers operate in active-active configuration. In such a case, mirroring provides the integrity of cached data if a RAID controller fails. This is obviously excellent from high
availability perspective. But it decreases performance. The data is mirrored between controllers across the drive-side FC loop and this competes with normal data transfers on the loop.

**Read-ahead multiplier**
This parameter affects the reading performance and an incorrect setting can have a large negative impact. It controls how many additional sequential data blocks will be stored into cache after a read request.

Obviously, if the workload is random, this value should be 0. Otherwise each read request will unnecessarily prefetch additional data blocks. Since these data blocks will rarely be needed, the performance is going to be impacted.

For sequential workloads, a good value would be between 1 and 4, depending on the particular environment. When using such setting, a read request will cause prefetching of several sequential data blocks into the cache and this will speed up consequent disk access. This leads to a fewer number of I/O transfers required to handle the same amount of data, which is good for performance in a sequential environment.

### 3.4.6 Media scan

Magnetic media errors can cause various problems in RAID arrays, including complete data loss. Media scan is a background process that checks all physical disk drives in the array for possible media errors. The FAStT Storage Manager allows you to run media scan and you can also configure how to run this process. For example, you can enable redundancy check.

Fault-tolerant RAID levels 1, 3 and 5 provide protection against a single disk drive failure. If the disk drive fails, its data will be rebuilt to a replacement drive, using information from other drives in the array. Obviously, this will only work if all data and either mirrored or parity blocks are correct on the surviving disk drives.

An error on either of those drives will cause data loss, because it will not be possible to rebuild the data that resided on the failed disk drive. It is therefore vital to ensure that there are no magnetic media errors on the disk drives in the array and that all redundancy information corresponds to the data blocks.

### 3.5 Advanced management

In this section, we describe some advanced management tasks which you can perform with the FAStT Storage Manager. These include storage partitioning and heterogeneous hosts support, diagnostics and troubleshooting. We also cover the command line support.

#### 3.5.1 Storage partitioning

Storage partitioning adds a high level of flexibility to the FAStT Storage Server. It allows you to connect a much higher number of host systems, either in standalone or clustered mode. Intel processor-based host systems connected to the FAStT Storage Server usually run an operating system with limited storage handling capabilities. Most of these operating systems can only treat the storage as if it was locally attached to the host system. Two or more individual host systems or clusters cannot access the same storage space, at least not without
disastrous results. This is in conflict with the idea of SAN, where the storage is supposed to be globally accessible to many host systems.

Without storage partitioning, the logical drives configured on a FASit Storage Server could only be accessed by a single host system or by a single cluster. This can surely lead to inefficient use of Storage Server hardware. Storage partitioning, on the other hand, allows you to create sets, containing the hosts with their host bus adapters and the logical drives. We call these sets storage partitions. Now the host systems can only access their assigned logical drives, just as if these logical drives were locally attached to them. You can therefore connect multiple hosts or clusters running the popular Intel-platform operating systems to the same FASit Storage Server. Storage partitioning adapts the SAN idea of globally-accessible storage to the local-storage-minded operating systems.

A storage partition contains several components: logical drive mappings, hosts or host groups and host ports.

A host group is a collection of hosts that are allowed to access the same logical drives, for example a cluster of two systems.

A host is a single system that can be contained in a host group.

A host port is the FC port of the host bus adapter in the host system. The host port is identified by its World Wide Name (WWN). A single host can contain more than one host port. If you attach the servers in a redundant way, each server will need two host bus adapters which means two host ports within the same host system.

The FASit Storage Server only communicates through the use of WWN. So the storage subsystem is not aware which host bus adapters are in the same server or in servers that have a certain relationship, for example, cluster. The host groups, the hosts and their host ports actually reflect a logical view of the physical connections of your SAN as well as the logical connection between servers, for example, clusters.

With the logical setup defined above, mappings are specific assignments of logical drives to particular host groups or hosts.

The storage partition is the combination of all these components and it ensures proper access to the different logical drives even if there are several hosts and/or clusters connected.

The Default Host Group contains everything that is not defined or assigned to any specific host or group. Every unassigned logical drive is mapped to this default host group. This means every host (or host port, to be exact), which is not assigned to a storage partition, will have access to all such logical drives and this can cause severe problems if the operating systems of the attached servers cannot handle this.

The FASit Storage Servers may support up to 16 storage partitions. Every mapping of a logical drive to a new host or host group will create a new storage partition. If you map additional logical drives to the same host or host group, this will not count as a new storage partition. For example, a cluster with two nodes with redundant I/O paths would be configured as one host group with two hosts.
Each host would have two host ports for redundancy. Several logical drives would be mapped to this host group. All these components represent one storage partition. If you attach another single host system to the same storage subsystem and map a logical drive to this host, you create another storage partition. If you then define a new logical drive and map it to either the cluster or the single host you will still be using two storage partitions.

Figure 40 displays a single host and a cluster, and they are attached to the same Storage Server. Host A, its host ports and its assigned logical drives 1 and 5 represent one storage partition. Cluster B with two hosts, their host ports and logical drives 2, 3 and 4 represent another storage partition.

**Mapping logical drives to the host systems**

If you do not use storage partitioning, then all host systems attached to the Storage Server will access the logical drives in the Default Host Group. In this case, each host system has equal access to all defined logical drives. Since there is hardly any operating system available for the Intel platform that will allow such connectivity of multiple hosts to the same logical drives, you must create storage partitions. You do this by mapping specific logical drives to the host ports of host systems. You have to identify the host ports by the World Wide Names (WWN) of the host bus adapters.

So, we create a host group first. The next step is to define a new host and its host ports. This is where you must know the WWN of your host bus adapters. Then we map one or more logical drives to this host group. You must also select a LUN for each logical drive. Be careful when changing the mapping and LUN of the Access logical drive. If you use host-attached management through the host agent component, you may lose the management connection when altering these Access logical drive settings. The complete storage partitioning procedure is...
described in 4.2.8, “Configuring storage partitioning” on page 89. This section explains clearly how to navigate through the FASTt Storage Manager to create storage partitions. Please also see 3.5.2, “Heterogeneous hosts” on page 53, for more information on storage partitioning in conjunction with different host operating systems.

**Storage partitioning considerations**

There are several reasons to use storage partitioning and keep the default host group empty except for the access logical drive.

Alternatively, you can also map the access logical drive but keep in mind that in this case you will only be able to manage through the host systems that belong to the storage partition which contains the access logical drive. This applies only to host-attached management, not the direct-attached management.

Even if only one host system is attached to the storage subsystem you should define one partition for this host. If you want to attach other host systems at a later time, you only need to define new host groups and map the logical drives to those new host groups. This process will not interfere with the existing host system at all. On the other hand, if you kept the original host system in the default host group, you would definitely have to change this when adding additional hosts. This is obviously much more disruptive.

If you ever have to replace a host bus adapter, the WWN of the new adapter will be different. Be aware of this, because storage partitioning assignments are based on the WWN. The new WWN does not appear in any of the storage partitioning assignments: after replacement this host system will therefore have access to the default host group logical drives. This can cause severe problems if there is another host system in the default host group: two hosts will try to access the same logical drives. This is another good reason why you should not use the default host group for any logical drives.

In a security-sensitive environment you can also assign the Access logical drive to a particular storage partition and ensure host-based management access only through the servers in this storage partition. In this environment you probably will assign a password to the storage subsystem as well. For more details, refer to 4.2.5, “Initial configuration steps” on page 83.

### 3.5.2 Heterogeneous hosts

Heterogeneous hosts support is a new feature of FASTt software V7.10, and it is tightly related to storage partitioning. In previous releases, it was possible to use storage partitioning and attach multiple host systems to a FASTt Storage Server, but they all had to run the same operating system. This limitation no longer applies: now the host systems can run a mixture of various operating systems. FASTt software currently supports:

- Windows NT
- Windows 2000
- Novell NetWare
- Linux
- HP-UX

AIX support is scheduled for the 2nd half of 2001.
You can use a mixture of different operating systems and also clustered and non-clustered variants of the same operating systems.

Figure 41 shows an example of heterogeneous hosts attached to the same Storage Server. Each host system runs a different type of operating system and they are all assigned to their own storage partitions.

As each operating system requires slightly different NVSRAM settings of the FASTT Storage Server, it is very important that you set the host port type in the storage partitioning to the correct value.

You specify the operating system type when defining new host ports. FASTT Storage Manager V7.10 contains an additional field in the Define New Host Port dialog box: the host operating system type. Please see 4.2.8, “Configuring storage partitioning” on page 89 for details.

The NVSRAM settings of a FASTT Storage Server are preset for hosts running Windows NT 4.0. These settings also affect the access logical drive. So without adjusting the host operating system you may encounter problems while accessing logical drives.

This basic host type setting can be changed to support any other operating system. You will need to disable storage partitioning to change the basic host type. If you want to use partitioning afterwards you have to contact the IBM Help.
Center to receive a feature enable key again. See also 4.4.6, “Handling premium features” on page 111. We strongly recommend that you use the storage partitioning approach to support other host systems rather than to change the basic host type.

Nevertheless it is also possible to attach hosts running operating systems other than Windows NT 4.0 to the FASIT Storage Server and manage the storage subsystem inband through Fibre Channel with the host agent. This is especially interesting for the initial setup.

Only the settings for HP-UX conflict with the default ones. If you attach a HP-UX system to the FASIT Storage Server the initial setup must be either performed by direct-attached management or through another host not running HP-UX by host-attached management.

After you have defined storage partitions for the different host groups with the correct operating system you may map the access logical drive in a host group with HP-UX systems, for example. Now you can manage the FASIT Storage Server also inband through the host agent of HP-UX.

### 3.5.3 Auto Volume Transfer (AVT) and RDAC: providing redundancy

If you have a FASIT Storage Server with two controllers you can provide redundant I/O paths between the host systems and the storage subsystem.

There are two different components that provide redundancy in the I/O data paths: the Auto Volume Transfer (AVT) feature of the storage subsystem and a host multi-path driver, for example, Redundant Disk Array Controller (RDAC).

AVT is a feature introduced with the new microcode release and resides in the FASIT Storage Server. It is enabled by default and provides redundant I/O paths in conjunction with a multi-path driver installed on the host system.

When a logical volume is defined, it is assigned to one of the two controllers, called the preferred owner. Normally, this controller will manage all I/O requests for this logical drive. If there is a problem in this path, the multipath driver in the host system starts using the other path. Now the alternate, non-preferred, controller is receiving the I/O requests for the logical drive. The logical drive will now be moved from the preferred controller to the alternate one and I/O requests are now handled through the alternate I/O path. Contrary to the old version of the storage management software, only this particular logical drive will be moved.

**Multi-path driver and AVT enabled**

If the host operating system does have its own mechanism for handling multi-path I/O, for example, HP-UX, you do not need to install any other components on the host system to provide protected I/O paths. If there is a problem on the preferred path the operating system will redirect the I/O to the alternate path and AVT will guarantee that the logical drive is accessible through the alternate controller.

If the host operating system does not include support for multi-path I/O, you can use the RDAC driver, which is part of the storage manager software, to enable the multi-path I/O. The RDAC package is available for Windows NT 4.0, Windows 2000 and Solaris. It will provide redundant I/O paths. In conjunction with the AVT feature, it will behave exactly as explained above.
After the I/O path recover from the failure, the preferred controller will re-establish ownership of the logical drive and control the I/O along this path.

There is a situation where you may wish to disable the AVT feature. If you have defined a host group with more than one host concurrently accessing the same logical drive, you can experience a performance problem with AVT enabled, and you may want to disable AVT.

Two hosts, both with redundant paths, share access to a logical drive. If only host B has a path failure (Figure 42) in the blue path, AVT will move the logical drive from the preferred to the alternate controller (red path).

Host A continues to send I/O requests to the logical drive through the preferred path (blue) unaware of the path problems of host B. If the I/O of host A reaches the preferred controller, AVT will now move the logical drive back to the preferred controller (red path).

As long as the path problem exists, host B will continue to send I/O through the alternate controller and host A through the preferred controller. The logical drive will change ownership with every I/O sent, which will significantly degrade the performance of the storage subsystem.

In a cluster environment you often have disks that are accessible by different hosts at the same time. It now depends on the architecture of the cluster if this situation can be accepted or not. A Netware Cluster Service for example uses the shared disks as a heartbeat, so all nodes access the disks in short intervals. In this case you need to disable AVT. In a Microsoft Cluster Server, concurrent access to the disks only occurs during the arbitration process for the quorum.
resource and only for the logical drive holding the quorum. In this environment it may be acceptable to use AVT.

If AVT is disabled only host systems with an installed RDAC can be attached to the FASiT Storage Server. If a host system running RDAC has a path failure, RDAC will move all logical drives to the other controller. Other host systems with a multi-path I/O driver different to RDAC will survive this move of their logical drives to the other controller. But if a host system without RDAC but with its own multi-path I/O driver has a path failure, the FASiT Storage Server would not move the logical drives to the other controller as AVT is disabled. And the host cannot issue a failover of the logical drives as no RDAC is installed. The consequence is that the host loses connection to its logical drives.

**Note**

In a cluster environment you may disable the AVT feature to ensure proper handling of the logical drives even in the case of a path failure depending on the cluster in use.

Be sure that all attached hosts are using RDAC when you disable AVT. Other systems will not failover properly as their multi-path I/O driver is based on AVT. Also do not install single attached systems in such an environment. As all logical drives will be moved in the case of a path failure single attached servers lose connection to their logical drives.

**Multi-Path driver and AVT disabled**

Even if AVT is disabled the RDAC will still provide redundancy. Each logical drive still has a preferred controller which controls the I/O along the path. If a component fails in this path the RDAC will cause a move of all logical drives to the other controller of the FASiT Storage Server, as opposed to transfer only a single logical drive with AVT.

Ensure that all attached hosts have a multi-path driver installed in this situation, because all logical drives will be moved to the other controller. This also includes logical drives from other host groups with different mappings in the storage partitioning.

Once the I/O path problem is fixed and both controllers are active again, you have to manually move the logical drives back to the preferred owner. Refer to 4.4.12.4, “Logical drives not on preferred path” on page 126.

Check Table 2 on page 58 for the operating systems that support failover when AVT is disabled.

**No multi-path driver with AVT enabled**

If the hosts do not support multi-path driver you can still attach them to a FASiT Storage Server with two active controllers. Each logical volume will also have a preferred controller. However, as there is no multi-path driver there will not be a failover to the other controller in the case of a path failure.

Again there is one situation were you may experience problems with the AVT. If two host systems without multi-path drivers are each connected to one of the controllers, you should use storage partitioning to divide the logical drives between the host systems. Otherwise, it may happen that a logical drive of host A
is moved to the alternate controller by AVT, because host B tries to access this logical drive. Host B is sending I/O to the alternate controller of the logical drive, AVT will move the drive, and host A loses connection to his logical drive. Host A needs to gain access to his logical drive again by sending I/O to the preferred controller. The logical drive will move back again. As in the case of the concurrent access, the performance will be degraded significantly. But opposed to the situation above you do not have to disable AVT but use storage partitioning.

**Operating system specific failover**

The supported failover environments depend on the operating system and the combination of multi-path driver and AVT. In Table 2 we list all currently supported combinations.

<table>
<thead>
<tr>
<th>Operating system</th>
<th>Multi-path driver name</th>
<th>Supported failover options</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Multi-path driver and AVT enabled</td>
</tr>
<tr>
<td>Windows NT 4.0</td>
<td>FAStT RDAC</td>
<td>X</td>
</tr>
<tr>
<td>Windows 2000</td>
<td>FAStT RDAC</td>
<td>X</td>
</tr>
<tr>
<td>Solaris</td>
<td>FAStT RDAC</td>
<td>X</td>
</tr>
<tr>
<td>HP-UX</td>
<td>HP PV-Links</td>
<td>X</td>
</tr>
<tr>
<td>Netware</td>
<td>FAStT Management a</td>
<td></td>
</tr>
<tr>
<td>Linux</td>
<td>FAStT Management a</td>
<td></td>
</tr>
</tbody>
</table>

a Multi-path driver support is pending. It is scheduled for the 2nd half 2001.

The FAStT Management Suite Java will be part of the FAStT Storage Manager package. It will provide multi-path I/O drivers for Netware and Linux. Based on a driver which resides on top of the hardware driver, the FAStT Management Suite Java can be configured with a preferred and an alternate path for every logical drive. In case of a failure along the I/O path, the driver will send the I/O requests through the alternate path, and AVT will move the logical drive within the FAStT Storage Server. On the Netware and Linux system, the FAStT Management Suite Java agent will be installed which then can be configured from a central management workstation running also the FAStT Management Suite Java.

**Disable and enable AVT**

The AVT feature can be disabled and enabled with two scripts included in the software package. Refer to 4.4.8, “Load scripts to the FAStT Storage Server” on page 114.

To disable AVT run the script `avtdisable.scr`; to enable AVT, again run the script `avtenable.scr`. Be sure to power cycle the FAStT Storage Server after the update.
3.5.4 FASTT Utilities

For each operating system there is a package available called the FASTT Utilities. It contains two command line tools, `hot_add` and `SM7devices`.

**hot_add utility**

The hot_add utility provides a mechanism to allow the detection of new disk drives within the operating system without the need to reboot the server. After creating new logical drives, just execute `hot_add` on the particular server. The available host bus adapters will be rescanned for new devices and the new devices are assigned within the operating system. Now you can start with the steps needed to prepare the disks for access in the operating system. For example, on Linux systems you could now run `fdisk` to define partitions, format the partitions and mount the partitions.

**SM7devices**

The other command line tool, `SM7devices`, is mainly used for troubleshooting. It lists all logical drives, their name, WWN and the storage subsystem accessible through the host bus adapters of the system. This is very useful to provide a basic check of connectivity and correct storage partitions.

3.5.5 Diagnostics and troubleshooting

The FASTT Storage Server provides a high level of system availability thanks to the built-in features and components used in the system. But even now it may happen that a failure occurs. It is very important that the failure can be identified and fixed as fast as possible.

The FASTT Storage Server logs all error conditions in its own event log stored in the controller. With the FASTT Storage Manager Software this event log can be read and also saved to a local system. In this event log, the error will be logged including time, exact location, and cause of the error. Please refer to Appendix A, “Critical events” on page 195.

The Recovery Guru, another part of the FASTT Storage Manager, can interpret the event log and will present you a detailed step by step procedure of how to recover from a particular failure. An example of an error condition would be a failed power supply in one of the drive enclosures. We show this error in 4.4.5, “The Recovery Guru” on page 109.

Also the controllers itself have a built-in test that can check the basic connectivity within the Fibre Channel paths to identify, for example, cable problems or defect GBICs.

We guide you through the most common failures in section 4.4.12, “Common problems” on page 123.

The diagnostics of the controller run from the FASTT Storage Server side. For Windows and Netware hosts, the FASTT Check application is available which provides similar tests for the host bus adapter side. This can help to detect...
problems when switches or hubs are used. Refer to 4.4.10, “The FAStT Check Application” on page 116.

### 3.5.6 Command line support

FAStT software V7.10 includes a SM7CLI executable file, that allows you to run Script Engine commands from an operating system command prompt. This gives you a lot of flexibility in FAStT Storage Server management. You are not only limited to Script Engine inside FAStT Storage Manager anymore. For example, you can use the Script commands in batch files and you can run these batch files as automated responses to certain FAStT events.

Several examples of SM7CLI usage follow.

**Executing one or more Script Engine commands**

To run one or more Script Engine commands, use this syntax:

```bash
SM7cli <hostname or IP address>
```

```bash
[<hostname or IP address>]
-c "<command>;[<command2>;...]" [-n <name>]
[-o <outputfile>] [-p <password>] [-e]
```

The role of the first parameter `<hostname or IP address>` differs for direct-attached and host-agent-attached management:

- **When managing directly**, specify the hostname or IP address of Ethernet port on the RAID controller.
- **When managing through host agent**, specify the hostname or IP address of the host system which runs the agent and is connected to the FAStT. If several FAStT Storage Servers are attached to this particular host system, you must also use the `-n` parameter to select the appropriate Storage Subsystem.

The next parameter, `-c`, indicates that a command or a list of commands to be executed follows. These are the Script Engine commands. Use a semicolon (`;`) to separate the commands and do not forget to enclose them in double quotes ("").

Several other parameters appear above:

- `-o` specifies the output file, where the command response and error messages will be redirected. If this is not specified, stdout is used.
- `-p` allows you to provide the password for accessing the Storage Server.
- `-e` indicates that we want to execute the commands without performing the syntax check.

**Executing a Script file**

The syntax for running a Script file is slightly different and it looks like this:

```bash
SM7cli <hostname or IP address>
```

```bash
[<hostname or IP address>]
-f <scriptfile> [-n <name>] [-o <outputfile>]
[-p <password>] [-e]
```
Most of the parameters are the same as the ones listed in the first example, except \(-c\). Now we use \(-f\) instead and the script filename to be executed follows.

**Referencing the Storage Server by name**

```
SM7cli -n <name>
   -c "<command>;<command2>;..."
   [-o <outputfile>] [-p <password>] [-e]
```

In this example, we do not reference the Storage Server by the hostname or IP address. Instead, we use the parameter \(-n\) and then we provide the Storage Subsystem name which is defined in the Enterprise Management Window.

**Displaying the configuration file**

```
SM7cli -d [ -i ]
```

The \(-d\) parameter displays the configuration file contents: it lists all Storage Subsystems that appear in the Enterprise Management Window. If you use option \(-i\), then the IP addresses are shown for each subsystem, otherwise you will see their hostnames.

**Usage information**

```
SM7cli -?
```

This syntax displays usage information for SM7CLI. It explains all the parameters that you can use.

For additional explanation of the SM7CLI syntax, refer to the SM7CLI.TXT file, found on the FASTT Storage Manager Installation CD. A Command Reference for Script Engine commands is available in the Enterprise Management Window online help.

### 3.6 Performance

Performance is one of the key requirements that customers are looking for. Through understanding of the subsystem limitations and scale options, storage performance guidelines and application characteristics will help you drive ahead the performance of the FASTT solution.

The FASTT Storage Manager software which lets you manage the FASTT Storage Servers includes its own performance monitoring tool, called the Subsystem Management Performance Monitor. This tool gives you information about the performance aspects of your Fibre Channel subsystem.

**Note:** This performance monitor tool is not related to the Windows NT Performance Monitor tool.

Use the Subsystem Management Performance Monitor to monitor storage subsystem performance in real-time and save performance data to a file for later analysis. You can specify the logical drives and/or controllers to monitor and the polling interval. Also, you can receive storage subsystem totals, which is data that combines the statistics for both controllers in an active-active controller pair.

To learn about disk drive characteristics and performance issues, please refer to 5.3.1, “Disk drive performance characteristics” on page 134.
To find out about RAID levels performance recommendations and guidelines, please refer to 5.3.2, “RAID level and performance” on page 136.

To understand how to use the data from the Subsystem Management Performance Monitor and what tuning options are available in the Storage Manager for optimizing the Fibre Channel subsystem’s performance, please refer to 5.3.4, “FAStT Storage Manager — Performance Monitor” on page 146.
Chapter 4. Step by step procedures for the FASTt Storage Server

This chapter provides a step-by-step procedure about how to setup and configure the FASTt Storage Server.

We presume:

- You understand the basic architecture of the FASTt Storage Server and the different ways to manage it with the FASTt Storage Manager.

- You have installed your FASTt Storage Server according to Chapter 2, “Hardware details” on page 15 and cabled the FASTt Storage Server in a way that a connection to each of the two controllers can be established from any system, either via Fibre Channel or Ethernet connection.

- The host systems attached are installed with the operating system and configured according to your needs, except for the driver for the FASTt Host Adapter and the FASTt Storage Manager software.

As the configuration of the FASTt Storage Server is done with the FASTt Storage Manager software, which is very similar for all supported operating systems, only the installation of the drivers and the software is operating system-dependent. This will be explained in the first part of this chapter.

Then we will show you how to configure the FASTt Storage Server and define arrays, volumes and enable storage partitioning. Also the different ways to monitor the FASTt Storage Server are explained.

If you want to install a high availability solution, we explain this in 4.3, “Installing a Microsoft Cluster Server with FASTt Storage Server” on page 97.

A SAN is a very complex construction, it is very useful to have diagnostics tools to check the SAN for problems. Even though the installation of the FASTt Storage Server using the GUI is very user-friendly, there can be situations, where you have to do some troubleshooting on the product. We will explain the available tools in the last part of this chapter.

4.1 Driver and host software installation

The host software package consists of different packages as explained in Chapter 3, “The FASTt Storage Manager software” on page 37. The two packages available for all operating systems are the FASTt Host Agent package and the FASTt Host Client package. For Windows NT 4.0 and Windows 2000 and Solaris, the FASTt RDAC packages provide redundancy within in Fibre Channel paths. For Netware and Linux, you need third-party software to enable multipath I/O to your storage system, if you are using the FASTt Host Adapter you can use the QLogic High Availability Package.

Because the installation of the host software varies with the operating system used, we will explain the procedure for each supported operating system.

The package names used in this documentation should be considered generic. Perhaps you have to replace them with the correct package names as found on the Internet or CD.
4.1.1 Microsoft Windows NT 4.0 and Windows 2000

Because the installation is the same for Windows NT 4.0 and Windows 2000, both will be covered in this section.

The host software for Windows consists of four packages, FASiT Client, FASiT RDAC, FASiT Host Agent and the FASiT utilities.

The driver for the host bus adapter is delivered in a separate package.

If two host bus adapters are installed in the system, redundancy is achieved with the FASiT RDAC package.

Because you are installing new software, including new drivers, you need to logon as Administrator.

4.1.1.1 Installing drivers for the host bus adapter

The driver for the FASiT host bus adapter is available for download at http://www.pc.ibm.com/support. The packages differ for Windows NT 4.0 and Windows 2000.

**Installing the driver for Windows NT 4.0**

Open **SCSI Adapters** from the **Control Panel** and choose the **Drivers** tab. You need to install a new driver, click **Add** to open the dialog, choose **Have Disk** and point to the driver location. After the installation of the driver you have to reboot the system.

**Installing the driver for Windows 2000**

Open the **Device Manager** and look for unknown devices, the FASiT Host Bus Adapter is not recognized by Windows 2000. Highlight the unknown adapter, right-click and choose **Properties**. On the driver tab select **Update Driver**, in the following dialog choose other source and point to the location of the driver. You must follow this procedure for each adapter installed in the system. After the installation you need to reboot the system.

4.1.1.2 Installing the host software

The following order should be kept while installing the different parts of the host software:

1. FASiT Client for Windows (same for Windows NT 4.0 and Windows 2000)
2. Java Virtual Machine (only for Windows NT 4.0)
3. FASiT RDAC for Windows (different for Windows NT 4.0 and Windows 2000)
4. FASiT Host Agent for Windows (same for Windows NT 4.0 and Windows 2000)
5. FASiT Utilities for Windows (same for Windows NT 4.0 and Windows 2000)

**Installing the FASiT Client**

Invoke setup.exe to install the software, you will be prompted if you want to install the Event Monitor Service, see Figure 43.
This service provides stand-alone alerting of the FASIT Storage Server through SNMP traps or email alerts. You should install and configure this service at least on a server operational 24 hours and directly attached to FASIT Storage Server, via Fibre Channel or Ethernet. The service requires the server to have access to the Access Logical Drive.

If you install the client software on a management workstation, which is not directly attached to the FASIT Storage Server, either through Fibre Channel or Ethernet, you don’t need to install the service.

If you install the storage software on a Windows NT 4.0 host, you have to install the Java Virtual Machine now.

**Installing the FASIT RDAC and the FASIT Host Agent**

The FASIT RDAC package provides failover and load balancing to the FASIT Storage Server. To install the RDAC only call setup and install the product, it will require a reboot of the system.

The FASIT Host Agent allows inband management of the FASIT Storage Server through Fibre Channel. It will also allow other management workstations access the FASIT Storage Server via Fibre Channel. To install the package, just invoke the setup, for Windows NT 4.0 it will require a reboot.

**Installing the FASIT Utilities**

The FASIT Utilities package delivers two command tools. The `hot_add` allows the addition of new LUNs while the operating system is up, `SM7devices` dumps a list of devices (controllers, LUNs) that are attached to the server; the package is installed by invoking setup.

**4.1.1.3 Enabling multi-path I/O with RDAC**

Because you installed the RDAC package, you do not need to configure anything to provide redundancy in the fibre paths.

**4.1.2 Linux Redhat 6.2**

The Linux host software consists of two packages, the FASIT Client and the Utilities, so you cannot manage the FASIT inband via Fibre Channel.

The driver is delivered in a separate package.
The FAStT Management Suite Java which will provide redundancy in the Fibre Channel paths is still under development at the time of writing this book, so unfortunately we cannot cover the installation and configuration here. The FAStT Management Suite will be included in the host software package. After the installation, you need to define a preferred and an alternate path for each logical drive. The AVT feature of the FAStT Storage Server will move the logical drive from the preferred to the alternate controller if a failure occurs along the I/O path. Updated information will be available on http://www.pc.ibm.com/support.

Because you are installing new software, including new kernel modules, and changing the bootup procedure, you need to logon as root.

4.1.2.1 Installing the kernel modules for the host bus adapter

The kernel module for the host bus adapter is available at the QLogic Web site, http://www.qlogic.com, look for the QLogic QLA22xx adapter. After downloading the rpm-package, install the package by rpm -iv qla2x00.i386.rpm. This will install the kernel module, now you have to configure the startup procedures of your system to load the module by default.

Add the following to the file /etc/modules.conf:

```
alias scsi_hostadapter qla2x00
```

Then you need to rebuild the ramdisk to include the module, for a single processor kernel perform:

```
/sbin/mkinitrd /boot/newinitrd-image 2.2.14-5.0
```

For a multiprocessor kernel the command is:

```
/sbin/mkinitrd /boot/newinitrd-image 2.2.14-5.0smp
```

The name `2.2.14-5.0` or `2.2.14-5.0smp` represents the directory name under `/lib/modules` where the kernel module was installed.

The next step is to modify the configuration of the lilo boot loader to load the new ramdisk image. Modify the line:

```
initrd=/boot/initrd-2.2.14-5.0
```

or

```
initrd=/boot/initrd-2.2.14-5.0smp
```

in the file /etc/lilo.conf to:

```
initrd=/boot/newinitrd-image
```

Apply the new configuration of lilo by executing `/sbin/lilo`. After a system reboot the driver will be loaded automatically.

4.1.2.2 Installing the host software

Now you are ready to install the FAStT Storage Manger software. You must have installed the Java Runtime environment on your system. The following order of installation should be kept:

1. FAStT Client for Redhat 6.2
2. FAStT Utilities for Redhat 6.2
All packages are added to the path and can be invoked just after installing the software.

**Installing the FASiT Client**
In the first step the client package is installed. Change to the directory where the rpm-package of the software is located, then issue:

```
rpm -iv SM7client-Linux.i386.rpm
```

The client requires X-Windows running to be started by typing `SM7client`.

**Installing the FASiT Utilities**
The last package to install is the utility package. To install it, enter:

```
rpm -iv SM7utils.i386-Linux.rpm
```

### 4.1.2.3 Enabling multi-path I/O with the FASiT Management Suite Java
After you have defined the logical drives and the storage partitioning you need to configure the FASiT Management Suite Java if you want to ensure redundancy in the I/O paths. For each logical drive, you need to declare a preferred and alternate path. Because this software was not yet available at the time of writing this book, please refer to the help included with the software.

### 4.1.3 Novell Netware 5.1
The host software consists of three packages, the FASiT client, the FASiT Host agent and the FASiT Utilities.

The driver is available as a separate package for download.

The FASiT Management Suite Java which will provide redundancy in the Fibre Channel paths is still under development at the time of writing this book, so unfortunately we cannot cover the installation and configuration here. The FASiT Management Suite will be included in the host software package. After the installation, you need to define a preferred and an alternate path for each logical drive. The AVT feature of the FASiT Storage Server will move the logical drive from the preferred to the alternate controller if a failure occurs along the I/O path. Updated information will be available on [http://www.pc.ibm.com/support](http://www.pc.ibm.com/support).

#### 4.1.3.1 Installing the driver for the host bus adapter
The driver for the FASiT host bus adapter is available for download at:


Start the Netware configuration program by executing `load nwconfig` and select **Driver options** from the menu. Under **Configure disk and storage device drivers**, choose **Select and additional driver**. Press **Insert** to install an unlisted driver and point to location, where the driver is stored. The QLogic HAM driver will now be installed.

If you have multiple adapters installed in the system, you may need to add another load command to the file `STARTUP.NCF`. You need one line per installed adapter:

```
LOAD QL2x00.HAM SLOT=n
```

The slot number `n` specifies which adapter to configure for each load instance.
If you want to use more than eight LUNs, including the access volume of the FASTT Storage Server add the /LUN option in the above load line. Be sure to have the operating system on a level, that does support more than eight LUNs.

4.1.3.2 Installing the host software
After the installation of the driver for the host bus adapter you can proceed and install the host software.

Installing the FASTT Client
Open the Novell Install Products dialog and select Add to install the host software. Point the installer to the directory containing the FASTT Client package and select the file product.ni. You will have to accept the license agreement during installation.

Installing the FASTT Utilities
Open the Novell Install Products dialog and select Add to install the utilities software. Point the installer to the directory containing the FASTT Utilities package and select the file product.ni.

4.1.3.3 Enabling multi-path I/O with the FASTT Management Suite Java
After you have defined the logical drives and the storage partitioning you need to configure the FASTT Management Suite Java if you want to ensure redundancy in the I/O paths. For each logical drive you need to declare a preferred and alternate path. Because this software was not yet available at the time of writing this book, please refer to the help included with the software.

4.1.4 Solaris

The host software for Solaris consists of four packages, FASTT Client, FASTT RDAC, FASTT Host Agent and the FASTT utilities.

The driver for the host bus adapter is delivered in a separate package.

If two host bus adapters are installed in the system, redundancy is achieved with the FASTT RDAC package.

Because you are installing new software, including new kernel modules, and changing the bootup procedure, you need to logon as root.

4.1.4.1 Installing the drivers for the host bus adapter
The driver for the most commonly used JNI host bus adapters is available for download at http://www.jni.com. Download the appropriate package for the adapter you are using and install it by executing:

```
pkgadd -d jni.pkg
```

You need to reboot the machine to have the driver loaded. Since this requires a configuration change and to ensure proper reconfiguration of your system, reboot with the command:

```
reboot -- -r
```

To allow proper failover in conjunction with the FASTT RDAC software you need to adjust a parameter in the configuration file jnic.conf for SBUS host bus adapter, or fca-pci.conf for PCI host bus adapter. Search for the line:

```
Failover Delay 0
```

68  Fibre Array Storage Technology: A FASTT Introduction
4.1.4.2 Installing the host software

Now you are ready to install the FASTT Storage Manager software. The following order of installation should be kept:

1. FASTT Client for Solaris
2. FASTT RDAC for Solaris
3. FASTT Host Agent for Solaris
4. FASTT Utilities for Solaris

All packages are added to the path and can be invoked just after installing the software.

**Installing the FASTT Client**

In the first step the client package is installed. Change to the directory where the software package located, then issue:

```bash
pkgadd -d ./SM7client-SPARC.pkg
```

Choose the default to install all packages and accept the execution of scripts; they are part of the installation.

The client requires X-Windows running to be started by typing `SM7client`.

**Installing the FASTT RDAC**

After the installation of the client package you can install the RDAC which provides redundant paths to the storage subsystem if you have two adapters in the system. This package requires a reboot of the system after installation. Execute:

```bash
pkgadd -d ./SM7rdac-SPARC.pkg
```

To ensure a proper reconfiguration during the bootup of the system reboot with the command:

```bash
reboot -- -r
```

**Installing the FASTT Host Agent**

After the RDAC is integrated in your system, you can install the agent:

```bash
pkgadd -d ./SM7agent-SPARC.pkg
```

The agent installs a daemon which is started directly after the installation; also the daemon will be included in the startup scripts and will start with every system reboot.

**Installing the FASTT Utilities**

The last package to install is the utility package. To install it, enter:

```bash
pkgadd -d ./SM7util-SPARC.pkg
```

4.1.4.3 Enabling multi-path I/O with RDAC

As you installed the RDAC package, you do not need to configure anything to provide redundancy in the fibre paths.
4.1.5 HP-UX

The host software for HP-UX consists of three packages, FASiT Client, FASiT Host Agent and the FASiT utilities.

The driver for the host bus adapter is included in the latest revision of the HP-UX.

If two host bus adapters are installed in the system, redundancy is achieved with the PV-Links package, which is part of the operating system. See 4.1.5.3, “Enabling multipath I/O with PV-Links” on page 71.

Because you are installing new software and changing the bootup procedure, you need to logon as root.

4.1.5.1 Installing the drivers for the host bus adapter

As the driver for the host bus adapter is included in the HP-UX kernel, there is no need to install a separate package.

After you have installed to adapters, the system recognizes the adapters automatically.

4.1.5.2 Installing the host software

Now you are ready to install the FASiT Storage Manager software. The following order of installation should be kept:
1. FASiT Client for HP-UX
2. FASiT Host Agent for HP-UX
3. FASiT Utilities for HP-UX

All packages are added to the path and can be invoked just after installing the software.

Installing the FASiT Client

In the first step the client package is installed. Change to the directory where the software package located, then issue:

```
swinstall -s <complete path>/SM7client-HP.pkg
```

Choose to install all packages. If you did not copy the software package in the HP-UX software depot, you must enter the full path where the package is located.

Verify the software installation:

```
swverify -v SM7client
```

If there are any error messages, check the file `/var/adm/sw/swagent.log` and follow the instructions outlined there.

The client requires X-Windows running to be started by typing `SM7client`.

Installing the FASiT Host Agent

After the client package is integrated in your system, you can install the agent:

```
swinstall -s <complete path>/SM7agent-HP.pkg
```

The agent installs a daemon which is started directly after the installation, also the daemon will be included in the startup scripts and will start with every system reboot. The event monitor will also be installed with the functionality matching the
system configuration. If you have SNMP and SMTP installed, it will be installed with both options; if you only have one, it will only install support for this.

Verify the software installation:

```
swverify -v SM7agent
```

If there are any error messages, check the file `/var/adm/sw/swagent.log` and follow the instructions outlined there.

### Note

As HP-UX conflicts with the default NVSRAM settings of the FASTT Storage Server, it is not possible to carry out the initial setup through a host-attached HP-UX system. You can either use the direct-attached method or another host-attached operating system to configure the FASTT Storage Server.

Once configured for HP-UX you can assign the access logical drive to the hostgroup for HP-UX and manage the storage subsystem inband through the host agent of HP-UX.

#### Installing the FASTT Utilities

The last package to install is the utility package. To install it, enter:

```
pkgadd -d ./SM7util-HP.pkg
```

Verify the software installation:

```
swverify -v SM7util
```

If there are any error messages, check the file `/var/adm/sw/swagent.log` and follow the instructions outlined there.

#### 4.1.5.3 Enabling multipath I/O with PV-Links

If the HP-UX system is attached with two host bus adapters to the FASTT Storage Server you will take advantage of this layout to provide redundant access to the storage. This can be realized with PV-Links, which is part of the HP-UX operating system. The redundancy is achieved by using volumes with a primary and secondary path to the same device.

You must logon as root, because new devices are added in this procedure.

The first step is to determine the primary and alternate path for each logical drive. Make sure, that all logical drives are on the preferred path in the subsystem management window. Then execute the command `SM7devices`, a dump similar to the following will be listed (the output was edited to fit the page). Notice that every logical drive, HP_1 to HP_4, and the access volume, is listed twice as you have two paths to each logical drive.
Fibre Array Storage Technology: A FASiT Introduction

A part of the World Wide Name of each logical drive is unique for each controller in the FASiT Storage Server. If you examine the World Wide Names, above, for the access volume you notice that they only differ in five digits, 75a54 and 75e60. Also the World Wide Names for the logical drives differ only in these five digits.

In our example, the devices are seen via the adapters c28 and c30. To determine the primary path of a specific logical drive, for example, HP_1, compare those five digits of /dev/rdsk/c28t1d2 with the one of the access volume, /dev/rdsk/c28t1d7. In our example, the digits do not match, so adapter c28 is not the primary path for the logical drive HP_1. If we compare the World Wide Names for the other logical drives, we will see the primary and secondary paths for the four logical drives as in Table 3:

<table>
<thead>
<tr>
<th>Logical drive</th>
<th>primary path</th>
<th>secondary path</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP_1</td>
<td>/dev/rdsk/c30t1d2</td>
<td>/dev/rdsk/c28t1d2</td>
</tr>
<tr>
<td>HP_2</td>
<td>/dev/rdsk/c30t1d3</td>
<td>/dev/rdsk/c28t1d3</td>
</tr>
<tr>
<td>HP_3</td>
<td>/dev/rdsk/c28t1d4</td>
<td>/dev/rdsk/c30t1d4</td>
</tr>
<tr>
<td>HP_4</td>
<td>/dev/rdsk/c28t1d5</td>
<td>/dev/rdsk/c30t1d5</td>
</tr>
</tbody>
</table>

Now we can define the primary path for the logical drive HP_1 by issuing:

```
# pvcreate /dev/rdsk/c30t1d2
Physical volume "/dev/rdsk/c30t1d2" has been successfully created.
```

The system will confirm the creation of the new physical volume /dev/rdsk/c30t1d2. Change to the directory /dev now, create a new directory for the volume group and create a new device for this volume group:

```
# cd /dev
# mkdir vgibm1
# cd vgibm1
# mknod group c 64 0x010000
```

Verify the successful creation with the command `Is -l`.

Now we can create the volume group and extend it with the secondary path.
The volume group is now successfully created with redundant paths by using both host bus adapters. You can verify the configuration of the volume group:

```
# vgcreate /dev/vgibm /dev/dsk/c30t1d2
Increased the number of physical extents per physical volume to 4331.
Volume group "/dev/vgibm" has been successfully created.
Volume Group configuration for /dev/vgibm has been saved in /etc/lvmconf/vgibm.
#
# vgextend /dev/vgibm1 /dev/dsk/c28t1d2
Volume group "/dev/vgibm1" has been successfully extended.
Volume Group configuration for /dev/vgibm1 has been saved in /etc/lvmconf/vgibm1.
#
```

The volume group is now successfully created with redundant paths by using both host bus adapters. 

```
# vgdisplay -v /dev/vgibm1
Information in "/etc/lvmtab" differs from that in "/etc/lvmpvg".
Physical volume "/dev/dsk/c26t0d0" does not belong to volume group "/dev/vglsi4".
--- Volume groups ---
 VG Name /dev/vgibm1
 VG Write Access read/write
 VG Status available
 Max LV 255
 Cur LV 0
 Open LV 0
 Max PV 16
 Cur PV 1
 Act PV 1
 Max PE per PV 4331
 VGDA 2
 PE Size (Mbytes) 4
 Total PE 4330
 Alloc PE 0
 Free PE 4330
 Total PVG 0
 Total Spare PVs 0
 Total Spare PVs in use 0
--- Physical volumes ---
 PV Name /dev/dsk/c30t1d2
 PV Name /dev/dsk/c28t1d2 Alternate Link
 PV Status available
 Total PE 4330
 Free PE 4330
 Autoswitch On
#
```

The volume group may now be expanded by adding the primary and secondary path to another logical drive. The volume group is now ready to be used.

### 4.1.6 Network attached management

Only the FAST Client has to be installed on a management workstation. The management workstation needs to be connected to either the FAST Storage Server or to other hosts running the FAST Host Agent through Ethernet. The installation is the same as for the host running the whole FAST software, see the appropriate installation for the operating system you need.
4.2 Configuring the FAStT

For the following discussion, we presume this configuration has to be set up (see also Table 4 on page 74):

- One host running Microsoft Windows NT 4.0 with two logical drives
- One host running Linux Redhat 6.2 with two logical drives
- A two-node Windows 2000 Advanced Server Cluster with four logical drives
- One host running Sun Solaris with two logical drives
- The access logical drive is available to all hosts

The FAStT Storage Manager is divided into two main parts, the Enterprise Management Window, which will show you all available FAStT Storage Servers, and the Storage Subsystem Management Window. In the Enterprise Window you choose the storage subsystem you want to manage and define global settings, such as the alert settings for the Event Monitor service. We presume that you are already working in the Storage Subsystem Management window on the FAStT Storage Server, if not stated otherwise.

4.2.1 Planning the configuration

Since a configuration of a FAStT Storage Server can be complex, especially when different operating systems and storage partitioning is involved, you should plan the configuration you want to apply in advance.

On one side, you need to define the arrays and the logical drives you need, including considerations such as number of drives in the arrays, size of the logical drives, RAID level and segment size. To plan the disk layout you need to know the attached hosts, their operating system and also the application using the storage of the FAStT Storage Server. Refer to 5.3, “Design guidelines” on page 133 to achieve an optimal configuration.

On the other side, you also need to define the layout of the attached hosts with their host bus adapter and the mappings of the logical drives to specific host groups or hosts. You should prepare a mapping table as shown in Table 4 where you keep all necessary information regarding the storage partitioning.

<table>
<thead>
<tr>
<th>Host group</th>
<th>Host</th>
<th>Host port</th>
<th>WWN</th>
<th>OS</th>
<th>Log. drive</th>
<th>LUN #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows</td>
<td>NT40</td>
<td>NT40_A</td>
<td>210000e08b017e3d</td>
<td>Win NT 4.0 nonclustered</td>
<td>NT40_01</td>
<td>00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NT40_B</td>
<td>210000e08b01ef62</td>
<td></td>
<td>NT40_02</td>
<td>01</td>
</tr>
<tr>
<td>Linux</td>
<td>REDHAT</td>
<td>REDHAT_A</td>
<td>210000e08b027386</td>
<td>Linux</td>
<td>Linux_01</td>
<td>00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>REDHAT_B</td>
<td>210000e08b027286</td>
<td></td>
<td>Linux_02</td>
<td>01</td>
</tr>
<tr>
<td>Win2000Cluster</td>
<td>NODE-A</td>
<td>NODE-A_A</td>
<td>200000e069ef5464</td>
<td>Win 2000 clustered</td>
<td>W2K_01</td>
<td>00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NODE-A_B</td>
<td>200000e06934e08f</td>
<td></td>
<td>W2K_02</td>
<td>01</td>
</tr>
<tr>
<td></td>
<td>NODE-B</td>
<td>NODE-B_A</td>
<td>200000e0695e4320</td>
<td></td>
<td>W2K_03</td>
<td>02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NODE-B_B</td>
<td>200000e06956a6a3</td>
<td></td>
<td>W2K_04</td>
<td>03</td>
</tr>
</tbody>
</table>
We will refer to this mapping table throughout the whole configuration of the FASiT Storage Server.

4.2.2 Starting the FASiT Storage Manager Client

Start the FASiT Storage Manager Client. It will launch the Enterprise Management Window. When you start the client for the first time, it will ask for an initial discovery of available storage subsystems (Figure 44).

The client software will send out broadcasts via Fibre Channel and the IP network, if it can find directly attached storage subsystems or other hosts running the FASiT Storage Manager Host Agent with an attached storage subsystem.

You have to invoke the Initial Automatic Discovery every time you add new FASiT Storage Server in your network or install new host agents on already attached systems. To see them in your Enterprise Management Window, choose Edit >> Rescan as shown in Figure 45.
If the FASiT Storage Server is connected via Ethernet or you want to manage through the host agent of another host, which is not in the same broadcast segment as your management station, you have to add the devices manually. Choose **Edit >> Add device** to enter the host name or the IP address you want to attach (Figure 46). If you add a FASiT Storage Server which is directly managed be sure to enter both IP addresses if two controllers are used in the subsystem.

An overall status is already included in the Enterprise Management window. You can see all FASiT Storage Servers and how they are managed: either direct or host-agent attached. There is a status column, usually the status will be *Optimal* with a green icon next to it. But if there are any problems, the status changes to *Needs attention* and a red icon appears.

If the FASiT Storage Server is not found by the discovery process and you cannot attach to the device manually, please refer to 4.4, “Maintenance, diagnostics and troubleshooting” on page 100.
Open the Subsystem Management window for the attached storage subsystem (Figure 47).

![Subsystem Management Window](image)

**Figure 47. First launch of Subsystem Management window**

Verify that the enclosures in the right half of the window reflect your actual physical layout. If the enclosures are listed in an incorrect order, choose **Storage Subsystem >> Enclosure Order** and sort the enclosures according to your site setup (Figure 48).

![Enclosure Order](image)

**Figure 48. Adapting the enclosure order**

### 4.2.3 Updating the controller microcode

The microcode of the FASiT Storage Server consists of two packages: the actual firmware and the NVSRAM package, which include the settings for booting the FASiT Storage Server. The NVSRAM is similar to the settings in the BIOS of a host system. The firmware and the NVSRAM are not independent! Ensure that you install the correct combination of the two packages.

The upgrade procedure needs two independent connections to the FASiT Storage Server, one for each controller. It is not possible to perform a microcode update with only one controller connected, so both controllers must be accessible either via Fibre Channel or Ethernet. Both controllers must also be in the active state.
If you plan to upgrade via Fibre Channel, make sure that you have a multipath driver installed on your management host, for example the FAStT RDAC package. This is necessary since access to the logical drive will move from one controller to the other during this procedure and the FAStT Storage Server must be manageable during the whole time.

Open the Subsystem Management Window for the FAStT Storage Server you want to upgrade, highlight the storage subsystem and from the Storage Subsystem menu choose Download >> Firmware. See Figure 49.

![Subsystem Management: Download >> Firmware](image)

You will be prompted for the file containing the firmware you want to download.

In the first part of the window you can see the actual versions of the firmware and NVSRAM installed on the controllers. The second part will tell you which firmware revision is included in the file you have chosen beneath (Figure 50).
Chapter 4. Step by step procedures for the FASIT Storage Server

79

Figure 50. Firmware Download

After clicking OK, a dialog box appears, where you have to confirm that all the above requirements are met, as shown in Figure 51.

Figure 51. Confirm Download of new firmware

The firmware upgrade will then proceed. It can take several minutes to finish (Figure 52).
After upgrading the firmware you have to upgrade the NVSRAM as well. Highlight the storage subsystem again and choose **Storage Subsystem >> Download >> NVSRAM** (Figure 53).

As for the firmware download, you have to choose the correct file you want to download to the controllers. See Figure 54.

The window has the same structure as the firmware download dialog (Figure 50), so you will see the firmware revision and the NVSRAM revision installed on the storage subsystem.
The confirmation dialog will show up again to ensure that you are downloading a version compatible to the firmware you just downloaded. Also you have to meet the above requirements again (Figure 55).

Since the NVSRAM is much smaller than the firmware package, it won’t take as long as the firmware download (Figure 56).
After the upgrade procedure, close the FAStT Storage Manager and reboot the whole FAStT Storage Server by power cycling the controller unit. Be sure to switch off both power supplies.

If the FAStT Storage Server is not recognized or unresponsive after the upgrade you have to initiate a new discovery in the Enterprise Management Window. If the FAStT Storage Server is still unresponsive you also have to reboot the host system and initiate a discovery when the system is up again.

### 4.2.4 Updating the drive microcode

The physical disk drives attached to the storage subsystem also include a microcode. To be able to download firmware to the drives you need a special tool, called the FAStT Field Tool. This tool is included in the download package containing the firmware updates.

Install the Field Tool on your system. The installation is similar to the installation of the FAStT Client package.

Start the FAStT Field Tool, as with the client you will be asked for an initial discovery of the storage subsystems. When the discovery is finished open the Subsystem Management Window for the FAStT Storage Server in question.

Ensure the proper configuration and status of the system.

---

**Note**

All I/O to the drives must be stopped during the microcode update.

---

Before you can download drive firmware you must enable this download by applying the script `EnableDriveFWDownload.scr` to the storage subsystem. The script is included in the package. To execute the script, refer to 4.4.8, “Load scripts to the FAStT Storage Server” on page 114.

Select **Diagnostics >> Download Drive Firmware** to start the download procedure (Figure 57).
Chapter 4. Step by step procedures for the FAStT Storage Server

4.2.5 Initial configuration steps

Before defining any arrays or logical drives you should perform some basic configuration steps. This also applies when you just reset the configuration of
your FASiT Storage Server, either with the Storage Manager software or through the serial connection.

Especially if you have installed more than one FASiT Storage Server, it is very useful to give it a literal name. To rename the FASiT open the Subsystem Management window and highlight the subsystem itself. From the menu bar choose **Storage Subsystem >> Rename** (Figure 59).

![Figure 59. Rename the storage subsystem](image)

Enter a new name for the subsystem in the dialog box (Figure 60).

![Figure 60. Enter the new name of the subsystem](image)

For security reasons, especially if the FASiT Storage Server is directly attached to the network, you can set a password. This password is required for all actions on the FASiT Storage Server that change or update the configuration. To set a password, highlight the storage subsystem, right-click and select **Change Password** from the context menu (Figure 61).
If the FAST Storage Server is booted with a reset configuration or without a configuration it will always define one array with a small logical drive. This is the Access logical drive and is needed to allow operating systems such as Windows to assign a disk resource to the host bus adapter driver. This array can be surely deleted if you define at least one new logical drive in the same step without booting any of the attached host systems. Refer to Figure 47 on page 77.

Highlight the array, right-click and choose Delete. A confirmation dialog will appear. You have to enter Yes in the text box to ensure you really want to delete the array and all logical drives.

**4.2.6 Creating arrays and logical drives**

The storage subsystem is installed and upgraded to the newest microcode level. Now the arrays and logical drives can be configured. If you are not sure how to divide the available drives into arrays or logical drives, and which restrictions apply, see 3.4, “Basic management” on page 42 and 5.3, “Design guidelines” on page 133, to avoid improper or inefficient configurations of the FAST Storage Server.

In the Subsystem Management window highlight the Unconfigured Capacity and choose **Configure >> Create Array/Logical Drive** from the menu bar (Figure 62).
You need to define the desired RAID level first, before choosing the number of drives you want to use.

There are two possibilities for the drive selection. You can choose each drive manually. For this, you need to switch to manual Drive Capacity Selection. Be sure that you consider performance and availability concerns when you choose the drives. We recommend that you keep the default automatic drive selection. Then you only need to choose the number of drives and the capacity you want to use. The storage software shows all possible array combinations for the RAID level you selected above. There will only be a selectable drive combination, if all drives are of the same size. If you decided for a specific array, the software will choose the best combination of available drives to keep up with performance and availability (Figure 63).
After you selected the number of drives and the RAID level of the new array, you need to define at least one logical drive in the array. All logical drives in one array will have the same RAID level.

By default all the available space in the array will be configured as one logical drive. If you want to define more than one logical drive, enter the size of the first logical drive (Figure 64).

Assign a name to the logical drive, which complies to your mapping table (Table 4 on page 74), if you plan to use storage partitioning, see 4.2.8, “Configuring storage partitioning” on page 89. Also you should not map the logical drive automatically but leave the default Map later with storage partitioning. Otherwise, the drive will be immediately seen by the attached hosts. Then if you change the mapping later, the logical drive, which appears as a physical drive to the operating system, will be removed without notifying the hosts and this can cause severe problems.

There are some possibilities to tune the logical drive for performance already in this early stage. You can define the segment size used for the logical drive according to the latter use in two ways. You can directly assign a fixed segment size or define the estimated drive usage. In this case the storage software will choose the right value for the segment size. As each logical drive has a preferred controller you can select one of the controllers for this logical drive. The software will choose a different controller for each array you define, so first array to Controller A, second array to Controller B and third array again to Controller A. As this distribution is just based on a round robin algorithm it may not always be the best choice. The logical drives should be spread across the controllers according to their load. As this is difficult to predict, you can change the preferred ownership
later on the fly, to reflect changing workloads. See 4.4.3.2, “Change ownership of a logical drive” on page 106.

![Create Logical Drive - Specify Logical Drive Parameters](image)

The array and a logical drive is now defined. It will be initialized now, but it is already accessible.

If you defined a logical drive smaller than the capacity of the array, there will be unconfigured capacity inside the array. To define another logical drive in this array, highlight this capacity, right-click and choose Define Logical Drive. The same steps as above are necessary, except the selection of drives and RAID level.

### 4.2.7 Define hot spare drives

You can define hot spare drives to keep the critical state of an array, after a drive failure, as short as possible. Hot spares are defined globally and protect all arrays, except RAID 0 arrays. Depending on the number of arrays defined and the number of drives attached to the FASiT Storage Server it may be useful to define more than one hot spare drive for the case that another drive fails while the first failed drive is still not replaced and the first hot spare is in use.

The hot spare drive must have at least the same capacity as the drives used in the arrays. It can be larger; then only a part of the capacity will be used to hold the data from the failed drive.

To define a hot spare drive, highlight the drive you want to use and select Drive >> Assign Hot Spare from the menu. The icon of the drive will change to a drive with a small cross (Figure 65).
4.2.8 Configuring storage partitioning

Since heterogeneous hosts can be attached to the FASIT Storage Server, you need to configure storage partitioning for two reasons. Each host operating system requires slightly different settings on the FASIT Storage Server, so you need to tell the storage subsystem the host type, which is attached. The other reason is the interference between the hosts if every host is able to access every logical drive. By using storage partitioning and LUN masking, you ensure that each host or host group only has access to its assigned logical drives.

Start the Mappings window from the Subsystem Management window by choosing **Configure >> Storage Partitioning**. In the Mappings window all the storage partitioning is defined. In the first step the host groups need to be defined.

Highlight **Default Host Group**, right-click and choose **Define New Host Group**, (Figure 66).
In the small popup window enter the names of the host groups you need. When you are finished, click cancel to exit the dialog (Figure 67). Note that if there is only going to be one server accessing logical disks in a storage partition, then it is not necessary to define a host group. However, as requirements are constantly changing we recommend that a host group is defined.

After the definition of the required host groups you add the hosts in the host group according to your mapping table. Highlight the group in which you want to add a new host, right-click and choose Define New Host (Figure 68).
Chapter 4. Step by step procedures for the FASIT Storage Server

Just enter the name of the host you want to define in the selected group (Figure 69).

If you assign a host to the wrong group, just highlight the host, right-click and choose **Move Host**. In a popup window you can select the new host group (Figure 70).

If you have defined all hosts that you plan to attach to the FASIT storage server, you have to define the host port and the operating system. The host port is identified by the World Wide Name. Highlight the host, right-click and choose **Define New Host Port** (Figure 71).
In the popup window, enter the port name for this adapter and choose the correct operating system. The host port identifier corresponds to the World Wide Name of the adapter port. In the drop down box you will only see the World Wide Names, that are currently active. If you want to enter a host port, which is currently not active, just type the World Wide Name in the field, but check for typing errors (Figure 72).

Because the FASiT only works with the World Wide Names of the host ports, the definitions for the host groups and the hosts itself only represent a view of the physical and logical setup. It is much easier to identify which host ports are allowed to see the same logical drives, and which are in different storage partitions, when this structure is available.

Choose the correct operating system from the list of available operating systems. This is the part of the configuration where you configure the heterogeneous host support. Because each operating system expects slightly different settings and handles SCSI commands a little different, it is very important to select the correct value here. Otherwise, your operating systems may not even boot anymore if connected to the storage subsystem (Figure 73).
Chapter 4. Step by step procedures for the FASIT Storage Server

Figure 73. Define the operating system of the host port

Now we start defining the mapping for each of our logical drives created in 4.2.6, “Creating arrays and logical drives” on page 85. Highlight the host group which you want to assign a new logical drive, right-click and choose Define New Mapping (Figure 74).

Figure 74. Define a new mapping for a selected host group

In the popup window, select the logical drive you want to map to this host group and assign the correct LUN number (Figure 75), according to the mapping table (Table 4 on page 74).
If you change the mapping of a logical drive the change will happen immediately, so make sure that this logical drive is not in use or even assigned by any of the machine attached to the storage subsystem.

**Note**

If you change the mapping of the access logical drive, even if you only change the LUN, you may loose the management connection to the FASiT Storage Server.

Now the first storage partition is used, because you just mapped a logical drive to another group and then to the default host group (Figure 76).

*Figure 75. Select the logical drive to be mapped and the LUN*

*Figure 76. The first storage partition is configured*

Define all other logical drive mappings that are needed according to your mapping table.
To make the logical drives available to the host systems without rebooting, the FASTT Utilities package provides a command line tool called **hot_add**. You only need to execute **hot_add**: all host bus adapters will be rescanned for new devices and the devices will be assigned within the operating system.

Now the drives are available to the hosts and their operating systems. You will probably have to take appropriate steps to enable the use of the storage inside the operating system, like running **fdisk** and formatting the disks.

### 4.2.9 Configuring, monitoring and alerting

Included in the FASTT Client package is the event monitor service. It enables the host running this monitor to send out alerts via email (SMTP) or traps (SNMP). The event monitor can be used for alerting in case of a problem with any of the FASTT Storage Servers. It should be installed and configured on at least two systems that are attached to the storage subsystem and allow inband management running 24 hours a day. This will ensure proper alerting, even if one server is down.

Depending on the setup you choose, different storage subsystems will be monitored by the event monitor. If you right-click on your local system in the Enterprise Management window (at the top of the tree) and choose **Alert Destinations** this will apply to all storage subsystems you see in the Enterprise Management window (Figure 77). Also if you see the same storage subsystem through different paths, directly attached and through different hosts running the host agent, you will receive multiple alerts. If you right-click on a specific storage subsystem you only define the alerting through this management connection.

When you remove or add new devices in the Enterprise Management window, the list of devices will not be automatically synchronized with the event monitor. If there is a mismatch between the devices listed in the Enterprise Management window and the devices known to the event monitor, a small icon will appear in the lower left corner of the Enterprise Management window next to the status icon for the subsystems. You need to synchronize the event monitor with the Enterprise Management window by clicking this icon.

![Figure 77. Alert Destinations](image-url)
In Figure 77 only alerts from host system aae48 through the host-agent would be processed.

If you want to send e-mail alerts you have to define an SMTP server before. Choose Edit >> Configure Mail Server. Enter the IP address or the name of your mail server and the sender address.

If you open the Alert Destination dialog, you define the e-mail addresses to which alerts will be sent. If you do not define an address, no SMTP alerts will be sent out. You also can validate the e-mail addresses to ensure a correct delivery.

If you choose the SNMP tab, you can define the settings for SNMP alerts: the IP address of your SNMP console and the community name. As with the e-mail addresses, you can define several trap destinations (Figure 78).

You need an SNMP console available for receiving and handling the traps sent by the service. There is a MIB file included in the Storage Manager software, which should be compiled into the SNMP console to allow proper display of the traps.

If you are using the IBM Director v2.2 as the SNMP console, open the console of the IBM Director. In the tasks list, right-click the SNMP Browser entry and choose Compile new MIB. Point the browse dialog to the MIB file of the storage manager software and click OK. Now the IBM Director can receive and display SNMP traps from the event monitor properly.

You will notice a green check mark in front of the system. The location of the check mark shows on which level the alerting is defined. All FAST Storage Servers beneath this level will be monitored.
4.3 Installing a Microsoft Cluster Server with FASTT Storage Server

As the installation of a Microsoft Cluster Server (MSCS) is different to the installation of a stand alone server, we will now explain the installation of the FASTT software in a cluster environment.

We presume that:

1. You are familiar with MSCS, because we will not explain the installation of MSCS here, but only explain the steps that are specific for the FASTT Storage Server to prepare it for the MSCS installation.

2. The systems are installed with the base operating systems (without the cluster software), driver for the host bus adapter, and the storage software.

The installation of the Microsoft Cluster Service is explained in Appendix B, “Installing Windows 2000 Cluster service” on page 203. We also go through the prerequisites there.

The Microsoft Cluster is based on a shared disk architecture. All systems need the ability to access all logical disks in the storage subsystem. The cluster service as part of the operating system is responsible to allow each server access to the disks or lock the disks if another server in the cluster is already using them. This architecture must be considered when you plan the storage partitioning and the mapping of the logical drives. All host systems must be in the same host group and all logical drives used as physical disk resource for the cluster must be mapped to the host group and not to a specific host. See Table 4 on page 74, the mapping table we used in our example.

If you are using dual paths to attach the FASTT Storage Server you must use them in all servers of the cluster. It is not possible to use a single path in one server and dual paths in the other one.

We will speak of Node A and Node B to differ between the two servers. The basic procedure is the same for both Windows NT 4.0 Enterprise Edition and Windows 2000 Advanced Server, and we will cover both.

4.3.1 Preparing the installation of the cluster server

Collect all the information you need to install the cluster, including node names, IP addresses and account and password for the cluster service login. Plan the disk layout and the logical drives you will use as cluster disks. Prepare a mapping table including the WWN of the host bus adapter installed in the systems.

4.3.2 Preparing Node A

Power on Node B, but stop the boot process by pressing the space bar during the operating system startup.

Power on Node A and boot to the operating system. Install the drivers and the storage software according to 4.1.1, “Microsoft Windows NT 4.0 and Windows 2000” on page 64. Because you are installing a high availability solution, we recommend that you install the event monitor of the FASTT Client package and configure alerting in the case of any problems in the storage subsystem. Refer to 4.2.9, “Configuring, monitoring and alerting” on page 95.
After the reboot of the system, prepare the FAStT Storage Server and update the microcode of the controllers if needed. See 4.2.3, “Updating the controller microcode” on page 77.

Now you can start to configure the logical drives and the storage partitioning for the cluster disks and the attached host systems. Define the arrays and logical drives according to your planning. Refer to 4.2.6, “Creating arrays and logical drives” on page 85.

The last step in the configuration is the storage partitioning. As mentioned above, both systems need access to all logical drives. We have to define a host group containing both cluster systems. Ensure that you always choose the same operating system (Windows 2000 clustered or Windows NT clustered) when defining the host ports. Map the logical drives to the host group.

Even if you can setup a cluster without using storage partitioning, we still recommend that you use a dedicated partition other than the default host group for the cluster, because you may add more hosts to the storage subsystem at a later time which can cause severe problems. See 3.5.1, “Storage partitioning” on page 50.

The configuration of the FAStT Storage Server is now finished and we can continue to configure the operating system. Use the hot_add utility to scan for the new drives just defined. It may be that Windows requires a reboot of the system. Use the Disk Manager to define partition per logical drive, assign a drive letter to the disk and format it using the NTFS file system. The operating system will ask to write a new signature on each logical drive. In Windows 2000 you will also be asked to upgrade the disks to dynamic disks, but you have to deselect this option.

--- Note ---

Dynamic disks are not supported by Microsoft as shared disk in a Windows 2000 cluster environment.

As the disk label is the only information that will be available, we recommend that you include the drive letter and the function of the disk in the label, for example, **Q-Quorum** for the quorum disk of the cluster using the drive letter **Q**:

The configuration of Node A is now finished. Reboot Node A, but do not allow the operating system to startup by pressing the space bar in the startup process.

### 4.3.3 Preparing Node B

Boot up Node B now. It is not necessary to install the whole FAStT host software on the second node, as all necessary configuration steps can be done on Node A. The minimum to install is the FAStT RDAC and the FAStT Host Agent package.

Nevertheless, we recommend that you install the complete host software also on Node B to allow access to the FAStT Storage Server, even if Node A is down.

Install the drivers and the storage software according to 4.1.1, “Microsoft Windows NT 4.0 and Windows 2000” on page 64. Because you are installing a high availability solution, we recommend that you install the event monitor of the FAStT Client package and configure alerting in the case of any problems in the
storage subsystem. If the event monitor is installed and configured on both servers, every alert will be sent twice, but it will provide redundancy if one server is down. Refer to 4.2.9, “Configuring, monitoring and alerting” on page 95.

Start the Disk Manager; all drives should be defined and labeled and there is no need to write signatures. Assign the same drive letters for each disk as on Node A. The volume labels you gave before, including the drive letter, will help you to identify the logical drives.

The preparation of Node B is now finished. Reboot the server, but do not allow the operating system to startup by pressing the space bar in the startup process.

### 4.3.4 Installing the cluster service on Node A

We can now install the cluster service. Invoke the setup for the cluster and choose to install a new cluster. During the setup, all drives defined on the storage subsystem should show up as shared drives.

Enter all needed information like the IP address and the name of the cluster and finish the installation. If you install on Windows NT 4.0, you will have to reboot the server.

Check that the cluster is running properly and all disks are online with the Cluster Administrator.

Also, refer to B.1, “Install Cluster service on Node A” on page 203, for a complete procedure to install the service on a Windows 2000 server.

### 4.3.5 Installing the cluster service on Node B

When the cluster service is running properly on Node A, we can install the cluster service on Node B. Call the setup routine and choose to join an existing cluster. Enter the name of the cluster and install the cluster software.

If you install on Windows NT 4.0 you will have to reboot the server.

Also, refer to B.2, “Install Cluster service on Node B” on page 209 for a complete procedure to install the service on a Windows 2000 server.

### 4.3.6 Apply service pack

Since we installed software from the original CD, we have to reapply the service pack installed before.

Start with Node B and pause this node in the cluster administrator, install the service pack and reboot the server. If the system is online again, resume it in the cluster administrator and move all resources from Node A to Node B.

Then pause Node A and reinstall the service pack on this system. After the reboot resume Node A in the cluster administrator.

The Microsoft Cluster Server is now installed and you can start to configure it and install resources in the cluster.
4.4 Maintenance, diagnostics and troubleshooting

In the above sections, we described how to install the drivers, the host software and how to setup a basic configuration of the FASTT Storage Server. We will now go to the different maintenance options available, for example, to tune the Storage Server for performance. We will also show some procedures for troubleshooting the storage subsystem.

4.4.1 The storage subsystem profile

To configure a FASTT Storage Server is a complex task. Therefore, the so called subsystem profile is a single location where all informations on the configuration is stored. The profile not only includes information on the controllers, attached drives and enclosures and their microcode levels. All informations on arrays, logical drives and storage partitioning is included as well.

We recommend that you save this profile each time the configuration or setup is changed even for minor changes.

To obtain the profile open the Subsystem Management window and select Storage Subsystem >> Profile.

All information in the profile is gathered form the various components when you request the profile. The profile can be saved locally and included in the documentation to maintain a change history for the storage subsystem. We recommend that you save a new revision of the profile and store it securely whenever a configuration change takes place. Even in the case of a complete configuration loss you can restore the array and logical drive configuration as well as the mappings for the storage partitioning (Figure 79).

![Figure 79. Profile for the storage subsystem](image)

The information collected here is also available for each component. Highlight the component you want to examine, right-click and choose Properties. The properties will be displayed for this component, for example, for one of the attached drive enclosures (Figure 80).
Drives can easily be located with the help of the storage manager software. Just highlight the component, for example, an array, and choose **Locate** from the context menu after right-clicking (Figure 81). All drives of this array will start flashing the yellow LED.

Sometimes it is very useful to see all components associated with one object. For example, you want to know which arrays, logical and physical drives are managed by Controller A. Just right-click Controller A and choose **List Associated Components** from the context menu. In the window you will see a list of all disk drives contained in arrays managed by this controller and also the logical drives and arrays (Figure 82).
4.4.2 Cache settings and media scan

The global settings for the cache usage can be modified to reflect special needs. Usually there is no need to adapt the values here. In most cases it is more efficient to change the values for individual logical drives which allows a much more granular optimization of the storage subsystem.

In the Subsystem Management window, select **Storage Subsystem >> Cache Settings**.

The values for Start and Stop flushing are discussed in 3.4.5, “Cache parameters” on page 48 as well as the cache block size value (Figure 83). The changes will be executed immediately.
4.4.3 Logical drive properties, cache settings and RAID migration

Except for the mapping, which was logical drive based, all configuration settings apply to the whole storage subsystem. We will now explain how you can change the parameters for each logical drive to tune the logical drive for performance, availability or even both. We will also explain how to change the properties of the array, for example, the RAID level or the number of drives.

4.4.3.1 Logical drive properties

To open the properties of a logical drive, right-click the drive and choose Properties (Figure 84).

![Logical Drive Properties](image)

**Figure 84. Base logical drive properties**

The first tab displays the information also contained in the storage profile. You can also define the priority to rebuild the drive in case of a drive failure. A higher
value will reduce the time needed for the rebuild but will have impact on the performance. A lower value will cause nearly no effect on the performance. As you are in a critical state while the rebuild lasts we recommend to keep the default or even use the highest value when the logical drive contains important data.

The Cache tab allows you to fine tune the usage of the cache for this individual logical drive (Figure 85).

![Figure 85. Cache settings for a logical drive](image)

The parameters you can configure here depend directly on the usage pattern of the logical drive as well as the importance of the contained data.

The Read Caching parameter can be safely enabled without having the risk of data loss. There are only rare conditions where it is useful to even disable this parameter, which then provides more cache for the other logical drives.

The Write Caching parameter allows the storage subsystem to cache write data instead of writing them directly to the disks. This can improve performance significantly especially for environments with random writes such as databases. For sequential writes the performance gain will vary with the size of the data written. If the logical drive is only used for read access it may improve overall performance to disable the write cache for this array. Then no cache memory will be reserved for this logical drive.

By default a write cache will always be mirrored to the other controller to ensure proper contents even if the logical drive moves to the other controller. Otherwise the data of the logical drive could be corrupted if the logical drive is shifted to the other controller and the cache still contains unwritten data. This option should only be turned off if the attached host does not use redundant paths to the storage subsystem.

The cache of the FAStT Storage Server is protected by battery against power loss. If the batteries are not online, for example, just after the power on, the controllers will automatically disable the write cache. If you enable the parameter the write cache will be used even if there is no battery backup available, resulting in a higher risk of data loss.
The Read Ahead Multiplier defines how many data blocks should be read ahead. The default value of zero will not read ahead any data. If the usage pattern for this logical drive is sequential reads, mainly it may increase performance to change the value up to four. A high value can cause an overall performance decrease as the cache is filled with read ahead data that may never be used. Use the performance monitor to watch the cache hit rate for this logical drive to find a proper value.

The Media Scan tab enables the background media scan which can provide a higher availability of the data (Figure 86).

The media scan will check, as a background operation, the physical disks for defects. It does this by reading the raw data from the disk and writing it back. This will detect possible problems caused by bad sectors of the physical disks. Depending on the global media scan rate this can have an impact on performance but improves data integrity. Before you can enable the media scan for a logical drive you must enable the global media scan. Choose Storage Subsystem >> Change Media Scan Settings and enable the media scan (Figure 87).
The duration is the time for one complete check of all logical drives where the media scan is enabled. The shorter the duration, the bigger the impact on performance will be.

4.4.3.2 Change ownership of a logical drive
If you found a performance bottleneck and want to change the preferred controller of a logical drive, just highlight the logical drive, right-click and choose Change Ownership and the new controller. Remember that this requires redundant paths from the host to the FASTT Storage Server since the change will happen immediately. It may be necessary to adapt the host operating system, for example, the primary and alternate path for the volume group in HP-UX (Figure 88).

![Figure 88. Change ownership of a logical drive](image)

4.4.3.3 Migrating RAID level
The RAID level is the same for all logical drives in one array as stated above. But, you can migrate the whole array to a new RAID level. This may be useful to adapt the array to a new usage pattern without the need to backup and restore the data since the migration can be done concurrently.

There are no restrictions for what RAID level can be migrated to another.

Depending on the RAID level you want to migrate to, you will need enough free space in the array to perform this operation. If there is not enough free space you will receive an error message. In this case, add more free capacity first, according to 4.4.3.4, “Adding more capacity to an array” on page 106.

To change the RAID level of an array, highlight the array, right-click and select Change RAID Level and the desired RAID level. The operation will start and cannot be stopped once started. The data will remain accessible during this operation which can take a long time.

4.4.3.4 Adding more capacity to an array
To increase the size of an array, you add new physical drives to it. This process can be done concurrently with access to the data on the logical disks. After the migration process, the new drives will be included in the array and provide new
free space within the array. This free space can then be used to define new logical drives.

To add new drives to an array, highlight the array, right-click and choose **Add free Capacity (Drives)**. You need at least one unassigned drive that can be added to the array (Figure 89).

![Add Free Capacity (Drives)](image)

*Figure 89. Add free capacity to an array*

For RAID levels 3 and 5, you must select at least one drive. For RAID level 1 and 10 you must choose an even number of drives.

Once the procedure is started it cannot be stopped. As the subsystem needs to redistribute the data contained in the array to all drives including the new ones, there will be a performance impact during this operation, but the logical drives of the array will remain available to the host systems.

**4.4.3.5 Defragment an array**

A logical drive can be deleted anytime to free the space on the disks. It may happen that the free space is fragmented within the array in different free space nodes.

As new logical drives cannot spread across several free space nodes the logical drive size is limited to the greatest free space node available, even if there is more free space in the logical drive. Compare with array four in Figure 90.
The array needs to be defragmented first to consolidate all free space nodes to one free space node for the array. Then new logical drives can use the whole available free space.

Open the Subsystem Management window, highlight the array to defragment, and choose **Array >> Defragment** to start the procedure. The defragmentation can run concurrently to normal I/O, but it will impact performance as data of the logical drives must be moved within the array. Depending on the array configuration, this process will continue to run for a longer period of time. Once the procedure is started it cannot be stopped again. During this time no configuration changes can be performed on the array.

### 4.4.4 The Major Event Log (MEL)

The FASiT Storage Server logs all major events in the storage subsystem itself. Even when no server is attached or attached but down, the FASiT will still monitor itself and log errors. To view the log file, open the Subsystem Management window and choose **Storage Subsystem >> View Event Log**. The event log will be transferred from the storage subsystem and displayed. If you select to view the details, you can directly see the exact error description and location, as well as a procedure to recover from the situation (Figure 91).
You can save the entries from the event log. Select the entries you want to save or use the Select All button and then save the log file locally on the host system.

After the initial setup or an error situation, it may be useful to clear the complete event log to have fixed starting point.

4.4.5 The Recovery Guru

The Storage Manager software contains the built-in Recovery Guru, a program that will help you identify problems as well as guide you through the necessary recovery procedure.

If you open the Enterprise Management window and there is a problem with one of the managed storage subsystems, you will see a red mark and the status Needs attention. In the Subsystem Management window, the icon for the Recovery Guru will flash and there is also a red mark next to the name of the storage subsystem.

4.4.5.1 Failed physical drive

In the following example a physical drive which is part of array 1 has failed (Figure 92).
To narrow the error, click on the Recovery Guru icon or choose Storage Subsystem >> Recovery Guru from the menu (Figure 93).

In the Recovery Guru window, the error is described, together with the exact location and a recovery procedure.

As you can see the actual error, a failed drive, caused the logical drive to change the status to degraded. But since we defined hot spare drives, the logical drive is already recovering from the failed drive.
The procedure is now to identify the failed drive and replace it. When the logical drive reaches the online status again, the replaced drive will be rebuilt and the hot spare will turn back to be hot spare again.

4.4.5.2 Failed power supply in expansion unit

Another example for an error is a failed power supply in one of the expansion units (Figure 94).

![Figure 94. Failed power supply](image)

Again as in the example above, you get information on the exact location of the error and also hints regarding the severity of the error. As there are two power supplies in the expansion units, there is no data loss yet, but the storage subsystem is in a critical state. If the other power supply of the same expansion unit loses power as well we could already have data loss depending on our array configuration.

4.4.6 Handling premium features

The support for storage partitioning is not part of the base firmware of the FASIT Storage Server. It is realized as a so-called premium feature. The storage partitioning premium feature is activated by default. In some circumstances it may be necessary to disable this feature and as a consequence also the heterogeneous host support. There is also one error condition that requires you to disable and then enable the premium features.

4.4.6.1 List premium features

To list the currently enabled premium features, select Storage Subsystem >> Premium Features >> List in the Subsystem Management window. A small popup window will appear (Figure 95).
In our example the storage partitioning for four hosts is enabled. The Feature Enable Identifier is needed if you request a new key to re-enable storage partitioning since the key is unique for each FASiT Storage Server.

4.4.6.2 Disable premium features
To disable a premium feature, select Storage Subsystem >> Premium Features >> List in the Subsystem Management window. Choose the feature you want to disable from the list and confirm (Figure 96).

Keep in mind that the change will happen immediately. If you are using storage partitioning and disable the premium feature, then you will not be able to create new partitions. Any existing partitions remain.

4.4.6.3 Enable premium features
To enable a premium feature, you need a key file which can be generated by the IBM Help Center. To generate a key for your FASiT Storage Server, the Feature Enable Identifier is needed. When you have the key available, choose Storage Subsystem >> Premium Features >> Enable in the Subsystem Management window. In the dialog window, point to the location where the key file is stored. You need to confirm to enable the premium feature selected. The change will happen immediately (Figure 97).
### 4.4.7 The performance monitor

The FASTT Storage Server provides a built-in performance monitor which can be used to identify bottlenecks. It is also very useful to collect information of the load of each controller in the FASTT Storage Server and to decide which logical drive to move to another controller so that the load is spread across both controllers.

Start the performance monitor from the menu with **Storage Subsystem >> Performance Monitor**. By default all logical drives and both controllers will be monitored. Pressing the Settings button will allow you to choose specific logical drives and controllers, for example, if you only want to analyze one host or host group as well as the update interval (Figure 98).

The performance monitor must be started manually. If you plan to analyze the performance figure, make sure that you leave the monitor running a considerable time to achieve representative figures (Figure 99).

The results can be saved locally to the host system since the data will be lost when you close the monitor again.

Also, see 5.3.2, “RAID level and performance” on page 136 for hints to tune the FASTT Storage Server.
4.4.8 Load scripts to the FASiT Storage Server

The FASiT Storage Server stores its configuration data in a reserved memory area called NVSRAM. The NVSRAM contains all configuration information regarding the behavior of the controllers. For example, there is one value responsible for the networking interface. By modifying this value you can enable or disable the network interfaces of the controllers. Any changes that may be necessary for your environment are provided as predefined scripts.

To execute a script, open the Enterprise Management window and highlight the storage subsystem. To open the script editor, choose **Tools >> Execute Script** (Figure 100).

In the script editor, you can now load the script you want to execute.
When you load the script it is ready to be executed. We recommend that you always use the Verify and Execute option instead of executing the script directly, to avoid errors caused by typing errors. Choose **Tools >> Verify and Execute** from the menu (Figure 101).

![Figure 101. Verified and executed script](image)

The scripts included in the software package are listed in Table 5.

<table>
<thead>
<tr>
<th>Script name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>avtenable.scr</td>
<td>Enable AVT feature</td>
</tr>
<tr>
<td>avtdisable.scr</td>
<td>Disable AVT feature</td>
</tr>
<tr>
<td>networkon.scr</td>
<td>Enable the network interface of the FASiT</td>
</tr>
<tr>
<td>networkoff.scr</td>
<td>Disable the network interface of the FASiT</td>
</tr>
<tr>
<td>lun0on.scr</td>
<td>Enable creation of LUN 0, default 1MB LUN</td>
</tr>
<tr>
<td>lun0off.scr</td>
<td>Disable creation of LUN 0, default 1 MB LUN</td>
</tr>
<tr>
<td>nw_mpio.scr</td>
<td>Enable settings for multi-path driver for Netware 5.1</td>
</tr>
<tr>
<td>factorydefault.scr</td>
<td>Reset the FASiT to factory default settings</td>
</tr>
</tbody>
</table>

We do not recommend that you use the scripts for changing the host type of the FASiT Storage Server. You should use the storage partitioning approach instead.

### 4.4.9 Load and save configuration of the FASiT Storage Server

If you want to setup multiple storage subsystems with the same configuration or if you want to store the configuration of one storage subsystem, you can save the
configuration to disk and apply it to another subsystem, which must have the same hardware layout, as number of enclosures and drives, and same drive capacities. The saved configuration only includes the array and logical drive configuration, the name of the subsystem, its cache settings and media scan rate. It does not include the information regarding storage partitioning. This information is included in the storage subsystem profile.

All information will be stored in a file that contains a script for the script editor.

To save the configuration of the subsystem open the Subsystem Management window, highlight the subsystem and choose **Configure >> Save Configuration**. The configuration is saved to a file.

To load the configuration on a subsystem, open the Enterprise Management window and select the subsystem. Choose **Tools >> Load Configuration** from the menu. Point to the file containing the configuration and load it. The script editor with the configuration script will appear. Choose **Tools >> Verify and Execute** to load the script to the FASTT Storage Server. The procedure can take a long time depending on the number of arrays and logical drives defined. When the procedure finishes the subsystem contains the same configuration as the source subsystem.

4.4.10 The FASTT Check Application

The FASTT Check Application is a tool for Microsoft Windows NT 4.0, Windows 2000 and Novell Netware 5.1, which allows basic management of the FASTT Host Adapter. You can check and also change values in the NVSRAM of the adapter, receive the WWN, and check attached devices.

4.4.10.1 Installation of the FASTT Check Application

To install the application just invoke the application, it will extract and install itself. The FASTT Check Application consists of three parts: the application itself, agent for Windows NT 4.0 and Windows 2000, and an agent for Netware (Figure 102).

![Figure 102. Installation of the FASTT Check Application](image)
You need the application on every machine from where you plan to manage, for example, the system itself for stand-alone management or an administration workstation for remote management. On each server you plan to manage, you have to install the agent.

For Windows, start the setup on the machine and choose the components to install. During the installation process, you have to install the Intel DMI 2.0 SDK. In the Component Selection window of the DMI installation choose to install both packages (Figure 103).

![Figure 103. Installation of Intel DMI SDK during FASiT Check Installation](image)

The installation will place an icon on your desktop or Start menu.

The NetWare Agent is installed through a Windows system by mapping the NetWare SYS volume. If you select to install the NetWare agent, you are prompted for a destination location. Point to the \SYSTEM directory on the NetWare SYS volume, the drive letter must be mapped to the root of the SYS volume. The NetWare agent will be added automatically to the AUTOEXEC.NCF file, and the next time the system is restarted, the agent and all components will start. If you want to start the agent manually, type the following at the NetWare prompt:

```
RPCSTART.NCF
LOAD NWSL -r
LOAD NWDMI2
LOAD NWCDMI
LOAD ONCSP
LOAD QLFVAGENT
```

### 4.4.10.2 Using the FASiT Check Application

Start the application and connect to the machine you want to work with.

As this tool establishes a direct connection to the adapter and bypasses the RDAC driver, you will see every attached LUN on each adapter, for example, LUN 3 is the same on both of the adapters.
If you highlight an adapter, you will see basic information about the adapter, like
the NOVRAM level, driver revision, WWN and also the PCI slot. The Information
and Statistics window contains statistics about the I/O and occurred errors.
Included is also a port database, where you can check which devices are seen by
the adapter. If you are using switches or hubs and have zoning configured, you
can check your connectivity in this window.

In the utilities, there are additional statistics regarding the single disks attached to
the adapter. You also have the possibility to check and change the NVRAM
settings of the adapter and even upgrade firmware and NVRAM through this
connection (Figure 104).

Figure 104. FASTT Check Application: NVRAM settings

Per default you are not allowed to change the values of the adapter. With the
agent, there is a small utility installed, the FASTT Check Security. If you start this
tool, you can enable updates of the NVRAM. For security reasons, you have to
assign a password to grant write access to the adapter (Figure 105).
Figure 105. FASIT Check Security

You should not change any values in the NVRAM unless you are told to do so, because it can cause severe problems in the communication between the adapter and FASIT Storage Server.

With the last tool, the diagnostics, you can run two basic tests, a loopback test, which is internal to the adapters and checks for frame checksums, disparity and length errors. With a wrap plug you can identify cable problems with tests, while it is mainly for testing the adapter. The read/write test uses SCSI commands to write data to the attached devices and reads them back in and compares the result. This test can be used to test the integrity and stability of the link in use. You should stop all I/O when you plan to use the diagnostics tests.

4.4.11 Obtaining the World Wide Name of the host bus adapter

The World Wide Name (WWN) of the Fibre Channel host bus adapter is a unique identifier of the adapter, similar to the MAC address of a network interface card. Sometimes it is necessary or useful to have this WWN available. The WWN is needed as soon as you implement the storage partitioning feature of the FASIT Storage Server. It is also used in the Fibre Channel switch and hub for zoning the fabric. On a switch or hub level it is also useful for problem determination as most switches and hubs allow monitoring of their ports and the attached hosts are only identified by the adapter in use.

In general, you can figure out the WWN of the adapter by entering the BIOS of the adapter. For the QLogic QLA2x00 adapter, you have to press <Alt><Q>, and enter the BIOS, then choose the host adapter. In the options windows, select Configuration Settings >> Host Adapter Settings. The WWN is the value of the entry adapter node name.

4.4.11.1 Windows NT 4.0, Windows 2000 and Netware 5.1

To obtain the WWN within Microsoft Windows NT 4.0, Windows 2000 or Netware 5.1, you need to install and use the FASIT Check Application. Refer to 4.4.10, “The FASIT Check Application” on page 116. Start the application and connect to
the server you want to examine. If the connection is established, highlight the adapter and the WWN will be displayed on the right side (Figure 106).

![Figure 106. FASTT Check Application](image)

### 4.4.11.2 Linux Redhat 6.2

The Linux operating system provides the information in the loaded modules and drivers in the system log file, which is normally located in the `/var/log` directory and called `messages`. After a certain period of time, a new system log file will be created, also called `messages`, and the old one will be renamed to `messages.#`, where # is an increasing number.

Log on as root and open the system log file containing the last bootup of the server, probably `/var/log/messages`. Open the file for reading by issuing `less /var/log/messages`.

Search for the last system bootup and the initialization of the Fibre Channel adapters; for each adapter you will see the startup procedure. While the adapter is opened, the WWN will be displayed.
You find for each adapter a line `scsi-qla?-adapter-port`, the value displayed there is the WWN of the adapter.

### 4.4.11.3 Sun Solaris
The Solaris operating system provides the information in the loaded modules and drivers in the system log file, which is normally located in the `/var/adm` directory and called `messages`. After a certain period of time a new system log file will be created, also called `messages`, and the old one will be renamed to `messages.#`, where `#` is an increasing number.

Log on as root and open the system log file containing the last bootup of the server, probably `/var/adm/messages`. Open the file for reading by issuing `less /var/adm/messages`.

Search for the last system bootup and the initialization of the Fibre Channel adapters. For each adapter you will see the startup procedure. While the adapter is opened, the WWN will be displayed.
The WWN is displayed for each installed adapter.

### 4.4.11.4 HP-UX

HP-UX delivers a small utility which displays the important information about the Fibre Channel adapter. To first determine the device name for the Fibre Channel adapters, logon as root and issue ioscan -fn. This will display a list of all I/O related information. You will find some entries regarding Fibre Channel adapter.

<table>
<thead>
<tr>
<th>Class</th>
<th>I</th>
<th>H/W Path</th>
<th>Driver</th>
<th>S/W State</th>
<th>H/W Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ba</td>
<td>3</td>
<td>0/3</td>
<td>lba</td>
<td>CLAIMED</td>
<td>BUS_NEXUS Local PCI Bus Adapter</td>
<td></td>
</tr>
<tr>
<td>fc</td>
<td>0</td>
<td>0/3/0/0</td>
<td>td</td>
<td>CLAIMED</td>
<td>INTERFACE HP Tachyon TL/TS Fibre /dev/td0</td>
<td></td>
</tr>
<tr>
<td>fcp</td>
<td>4</td>
<td>0/3/0/0.2</td>
<td>fcp</td>
<td>CLAIMED</td>
<td>INTERFACE FCP Domain</td>
<td></td>
</tr>
<tr>
<td>ext_bus</td>
<td>28</td>
<td>0/3/0/0.2.16.0.0</td>
<td>fcparray</td>
<td>CLAIMED</td>
<td>INTERFACE FCP Array Interface</td>
<td></td>
</tr>
<tr>
<td>target</td>
<td>6</td>
<td>0/3/0.2.16.0.0.0</td>
<td>tgt</td>
<td>CLAIMED</td>
<td>DEVICE</td>
<td></td>
</tr>
</tbody>
</table>

In our example the device name of the first Fibre Channel adapter would be /dev/td0. Now use the fcsmutil to display the information on the specific adapter by calling fcsmutil /dev/td0:
The WWN is displayed in the output. You can also see the link status and the kind of link, and here the fabric protocol is used.

4.4.12 Common problems

There are problems that occur sometime during the installation and configuration of the FAStT Storage Server.

4.4.12.1 Out of compliance error message in Recovery Guru

There are cases where the Recovery Guru reports the error Out of Compliance. This can be caused by an improper firmware and NVSRAM upgrade. The procedure to recover from this error condition is to disable the premium feature storage partitioning and enable it with a new key.

The key file must be generated for your FAStT Storage Server, because it is unique for each storage subsystem. Contact the IBM Help Center to have them generate the key for you. They will need the Feature Enable Identifier of your FAStT Storage Server to do so. Refer to 4.4.6, “Handling premium features” on page 111 for the exact procedure.

You can use the `fcsmdutil /dev/td0` to get a list of all logical devices including their properties that can be accessed through this server. If all logical drives are listed but not the access logical drive, you probably have a problem with storage partitioning.

4.4.12.2 FAStT Host Agent does not see FAStT Storage Server

When the FAStT Host Agent is started by the operating system, it will scan the fibre paths to find the attached storage subsystem. When the service (or daemon) cannot find a storage subsystem, it will stop itself. On Windows systems, you will receive a message that the host agent stopped with error code 100.

The reason is that the host system has no connection to the access logical drive. There can be several reasons for this error, which we will go through.

No physical connection
No connection between the host bus adapter and the FAStT Storage Server can be established, because the physical connection is broken or not connected.
properly. All cables must be seated properly in the GBIC ports and the link status must be online. Check the link status LEDs on switches, hubs and the storage subsystem. Also the cable itself may be broken. Swap cables and watch if the error moves.

Within the operating system, check if the host bus adapter itself does see any devices of the FASTT Storage Server. For the host agent the access logical drive must be visible.

You can use the controller built-in diagnostics to test basic connectivity. Highlight the controller in the Subsystem Management window you want to check, right-click and choose **Run Diagnostics**.

---

**Note**

Using the diagnostics will make the controller unavailable for I/O. If you want to run the tests concurrently to I/O, make sure all attached systems are using a multi-path I/O driver.

The controller may be placed offline by diagnostics. you need to activate it manually after diagnostics finish.

We recommend that you stop all I/O to the controller while diagnostics are running.

---

When diagnostics finish, the results are listed in the status window; refer to Figure 107.

![Figure 107. Built-in diagnostics of the controller](image)
**Zoning in the fabric**

Even if the connections are physically online, it may be still impossible for the operating system to establish a link if there are zones defined in the fabric.

You need one connection between the host bus adapter and the controller in the FAStT Storage Server. There are different combinations possible depending on the number of host bus adapters, controllers and if the operating systems use redundant paths.

In general each host bus adapter should see only one controller of the FAStT Storage Server. If you are attaching one server with two host bus adapters to one switch and connect the storage subsystem with two paths, you have to define two zones.

If the host is attached with one host bus adapter and the FAStT Storage Server is connected with two paths to the switch, all three ports should be in the same zone to allow the host agent to access both controllers. In this case, you still have multiple paths to the same logical drives in the storage subsystem.

In general for a host with redundant paths, there must be only one path per host bus adapter to one of the controllers. This gives you two completely separate paths to the storage subsystem.

If the host does not support redundant paths, you should ensure that the host systems only see one controllers of the storage subsystem. If the host sees both controllers it depends on the operating system to access only one path.

**Mapping does not allow to attach to the Access logical drive**

Another reason for this error is a mapping of the Access logical drive to another group other than the default host group. If there is a mapping defined for the Access logical drive, you only can manage the storage subsystem through the fibre connection if you use one of the hosts inside this host group.

If you mapped the access logical drive to a host group containing only dummy entries for the host ports, you can only manage the FAStT Storage Server directly through the network ports. If the network cannot be configured or is disabled, there is no way to recover from the situation than to reset both controllers through the serial connection. Refer to 4.4.14, “Resetting the controllers” on page 128.

**4.4.12.3 Subsystem contains a controller that has not been added**

When you want to open the Subsystem Management window from the Enterprise Management window, an error message will popup, stating that the subsystem contains a controller that has not been added. There are two reasons for this error.

If you manage the FAStT Storage Subsystem directly through a network connection you need access to both controllers. Check that both IP addresses, one per controller, are known to the storage software. Use Tools >> Add Device to enter the IP address of the missing controller.

After a reset of the controllers or an exchange of one of the controllers, you may also receive this message. Then there is a mismatch between the NVSRAM configuration settings of the two controllers in the same storage subsystem. You have to download the NVSRAM to both controllers even if it has the same version. Refer to 4.2.3, “Updating the controller microcode” on page 77.
4.4.12.4 Logical drives not on preferred path
A very common but not severe problem is the message that not all logical drives are managed by their preferred controller. This happens, for example, if a host with redundant paths has a failure in one of the paths. The AVT will move the affected logical drives to the intact path. Even if the failed path returns online the logical drive will not be moved back automatically.

To distribute the logical drive to their preferred controller, select **Storage Subsystem >> Redistribute Logical Drives** and all logical drives will be moved (Figure 108). Be sure that all hosts support path failover before distributing the logical drives.

![Figure 108. Redistribute logical drives](image)

4.4.13 Network setup
If you plan to manage the FASiT Storage Server via network, you have two possibilities, DHCP/BOOTP and fixed IP addresses. In both cases you will need one IP address per controller. The default is the DHCP/BOOTP method.

You will need to attach both controllers to an Ethernet switch or hub. The built-in Ethernet controller supports either 100Mbps or 10Mbps.

4.4.13.1 Assigning IP addresses through DHCP/BOOTP
You need to setup a DHCP/BOOTP server on the same physical network as the two controllers. The protocol used by the FASiT Storage Server is the BOOTP protocol. Each time the FASiT Storage Server is powered on, the controllers request an IP address via BOOTP. Configure the DHCP/BOOTP server to assign the controllers an IP address according to the server you are using. The MAC addresses of the controllers are printed on each controller: on a FASiT 500 controller you can find label with the MAC address just beneath the lever on the front, and on a FASiT 200 you find the label address at the back.

4.4.13.2 Assigning fixed IP addresses
You need a serial connection to the controllers in the FASiT Storage Server. Because you have to reboot the FASiT Storage Server during this procedure, you have to stop all I/O activity.
Chapter 4. Step by step procedures for the FASIT Storage Server

Connect to the FASIT Storage Server and send a break signal to the controller. This will vary depending on the terminal emulation. For most terminal emulations, as HyperTerm, which is included in the Microsoft Windows products, type <Ctrl><Break>. There is a known issue with early version of HyperTerm, you need at least V3 to work properly. Visit http://www.hilgraeve.com to download the latest version.

If you only receive unreadable characters, send a break signal again, until this message appears:

Press <SPACE> for baud rate within 5 seconds.

Then press <Space> to ensure the correct baud rate setting. Send a break signal and this message will appear:

Press within 5 seconds: <ESC> for SHELL, <BREAK> for baud rate.

Press <Escape> to access the shell of the controller.

The password, you are to be prompted for, is infiniti. The command netCfgShow will show you the current network configuration. The controller will dump a list similar to the following:

```
-> netCfgShow

==== NETWORK CONFIGURATION ====
Interface Name : dse0
My MAC Address : 00:a0:b8:0b:11:b2
My Host Name : fastt_a
My IP Address : 100.100.100.236
Server Host Name : host
Server IP Address : 0.0.0.0
Gateway IP Address : 100.100.100.9
Subnet Mask : 255.255.255.0
Network Init Flags : 0x01
Network Mgmt Timeout : 30
Shell Password : ************
User Name : guest
User Password : ************
NFS Root Path : (null)
NFS Group ID Number : 0
NFS User ID Number : 0n.
value = 27 = 0x1b

->
```

For changing the above values, use the netCfgSet command. You will be asked for each entry to keep, clear or change the value.

---

Note
Please follow the procedure outlined here exactly, because some commands that can be issued from the serial console can cause data loss.
After you have assigned a fixed IP address to Controller A, disconnect from Controller A and repeat the procedure for Controller B. Remember to assign a different IP address.

After this procedure, reboot the whole FAStT Storage Server by power cycling the controller unit. Be sure to switch off both power supplies.

### 4.4.14 Resetting the controllers

This procedure explains how to reset the controllers in the FAStT Storage Server. You will lose all configuration data, including array and logical drive configuration. Backup any data before continuing the procedure.

There are two possibilities to reset the controllers in the FAStT Storage Server. The preferred way is to reset the controllers through the storage manager software, either directly or host-agent attached.

Normal problem determination does not require this procedure to be done. Because you have to reboot the FAStT Storage Server during this procedure, you need to stop all I/O activity.

Note

Both procedures will cause loss of all data, including array and logical drive configurations. Backup all data before proceeding.

### 4.4.14.1 Reset controllers with the FAStT Client

The FAStT Storage Server can be completely reset through the client software. Open the Subsystem Management window, highlight the storage subsystem and choose **Configure >> Reset Configuration** from the menu.

Before you can confirm the start of the procedure you must enter `yes` in the popup window as an extra confirmation since you lose all data.

Once the storage subsystem is reset you will need to rescan for the device in the Enterprise Management window. It may be necessary to reboot the system running the host agent to recognize the storage subsystem.
4.4.14.2 Reset controllers through serial connection

If you cannot reset the controllers with the client software because you lost connection to the Access logical drive and there is no network connection, the controllers can be reset through the serial connection.

**Note**

You only should enter the serial console under the advice of IBM Help Center support, because some commands that can be issued from there can cause data loss and hardware malfunction.

Connect Controller A of the FAStT Storage Server with a null modem cable. Start a terminal emulation on the host system with the following settings:

- 9600 baud
- 8 data bits
- no parity
- one stop bit
- XON/XOFF flow control

Connect to the FAStT Storage Server and send a break signal to the controller. This will vary depending on the terminal emulation. For most terminal emulations, like HyperTerm, which is included in the Microsoft Windows products, type <Ctrl><Break>. There is a known issue with an early version of HyperTerm, you need at least V3 to work properly. Visit [http://www.hilgraeve.com](http://www.hilgraeve.com) to download the latest version.

If you only receive unreadable characters, send a break signal again, until this message appears:

Press <SPACE> for baud rate within 5 seconds.

Then press <Space> to ensure the correct baud rate setting. Send a break signal and this message will appear:

Press within 5 seconds: <ESC> for SHELL, <BREAK> for baud rate.

Press <Escape> to access the shell of the controller.

The password, you will be prompted for, is **infiniti**. At the shell enter `sysWipe`. Wait until the procedure finishes. Disconnect the serial cable.

If there is a second controller installed, connect to the serial port of Controller B and repeat the procedure to ensure that you reset both controllers.

After this procedure reboot the whole FAStT Storage Server by power cycling the controller unit. Be sure to switch off both power supplies.

Because all configuration data is destroyed on the storage subsystem, you have to initiate a *Discovery* in the *Enterprise Management* window. The FAStT Storage Server will appear as a new storage subsystem.
Chapter 5. FAStT solution design

In this chapter we discuss various aspects of the FAStT solution design:

- How to approach the customer needs and how to identify the right solution according to their business needs
- Guidelines for building a storage solution with the FAStT product
- Useful customer cases and the solutions we implemented for them

5.1 Infrastructure deployment framework

Because implementing a new storage solution within a company can impose a change process to three organizational basic factors — people, technologies and processes — it is useful to outline the key phases of an infrastructure deployment project, when approaching that kind of activity.

Figure 109. Deploy cycle

**Phase 1: Envisioning phase**
During the envisioning phase in an infrastructure deployment project you and the customer develop business goals and determine the scope of the project. In addition, you gather users’ needs, develop a solution concept, begin analyzing risk, and determine the project structure.

**Phase 2: Planning phase**
Key deliverables during the infrastructure deployment planning phase are a draft functional specification, draft master project plan, draft project schedule, and a development environment. The infrastructure deployment functional specification will include information, such as design, usability, and deployment goals, the solution design, component specification, project risks, and project standards. The master project plan includes the approach, dependencies, and assumptions of the project. The master project plan also includes other plans, such as a deployment plan, a pilot plan, a purchasing and facilities plan, a test plan, a capacity plan, a training plan, and a security plan.

**Phase 3: Testing phase**
Key activities in this phase include validating the technology, developing the proof of concept, testing, performing a pilot, and incorporating feedback from the pilot.
Phase 4: Deploying phase

The infrastructure deployment deploying phase is an active phase rather than an analytical one. During this phase, you deploy the core technology, site components, stabilize the deployment, transition the project to operations and support, and obtain customer sign-off on the project. After the deployment, you conduct a project review and survey of customer satisfaction.

5.1.1 Design and planning

In this chapter our goal is to help you to deliver the right solution to your customers. Here we emphasize the second phase, which is the planning phase. We focus on the planning phase rather than on each of the other phases.

There are few steps involved when designing a solution:

Conceptual design
The goal in conceptual design is to identify business needs and to understand what the customers do and what they require. It is not the approach taken or the technologies used to build a solution. Conceptual design is analogous to the rough sketches and scenarios created when designing a house. These are easily understood models jointly created by the customer and the architect.

Logical design
Logical design organizes the details of the solution that the team builds to fulfill business needs and user requirements. Logical design is created by the architect's team and lays out the structure of the solution and the communication paths among elements. Logical design corresponds to a floor plan and elevation, where elements such as spatial relationships are organized.

Physical design
Physical design addresses the technology that will be used by the customer. The goal is to apply real-world technology constraints to the logical design, such as implementation and performance considerations. Physical design corresponds to a contractor's blueprints for the physical elements of a structure — wiring, plumbing, heating, and ventilation. The contractor's plans add detail to the architect's plans and reflect real-world construction constraints.

5.2 Information gathering

Understanding the customer's wants and needs is a very important part of obtaining the final goal: to have a system installed and tuned to meet the customer's requirements.

An important part of understanding the customer's wants and needs is gathering all the possible data from the customer before you design the solution. Some of this data may have been collected earlier by you or by other individuals, but you will require all of it to be successful. The data that you gather will help you determine what type of solution should to be deployed and how the hardware and software parameters need to be set. If you do not understand what the guidelines are before you begin, how can you be successful? In fact, you may find that you have designed it incorrectly and now need to redesign the entire solution.

Remember, if you do not satisfy the customer's requirements, the customer will probably not ask you back, no matter what the price is!
To meet your customer needs with your design, three steps are needed:

1. Know your customer goals and expectations.
2. Gather the information by interviewing your customer.

**Know the customer's goal**

What is the goal?

1. To solve a specific problem?
2. To implement new technology?

What are the expectations?

1. Performance
2. Availability
3. Return on investment
4. Extend legacy products/devices

We need to prioritize and set expectations levels.

**Gathering the information**

1. Is the intent to add shared storage to the network, clustering, etc.?
2. What is the performance characteristics required to support the various applications on the SAN?
3. How do you measure this? Throughput or I/O per second or response time?
4. What is the maximum capacity or bandwidth required for peak loads?
5. What percentage of the SAN capacity will be used on average, and at what level of utilization would it become saturated?
6. What happens to performance in the event of failure of SAN components?
7. Can sufficient spare bandwidth be provided to continue to deliver acceptable performance?
8. How much growth is expected?
9. Will the SAN be required to support additional applications?
10. If so, in what time scale, for instance within the next year or two years?
11. How fast is data growing, and will you need to expand storage resources, or add more servers?
12. Do you need to support legacy SCSI hardware?
13. What are the distances between the server and storage resources, and will this need to expand to include other departments and locations?

---

**5.3 Design guidelines**

These design guidelines are based on performance measurements in the lab and best practices.
5.3.1 Disk drive performance characteristics

When configuring your system you must take special care to properly design and configure the I/O subsystem. The configuration of the I/O subsystem can either enhance or degrade the performance of your system. By understanding the limitations of the I/O subsystem and configuring your system/server to function optimally within those limitations, you will help ensure that it will provide the level of performance that you need.

The disk drive is one of the fundamental components of the computer system. Amazingly enough the mechanics of the disk drives have not changed much in the last 20 years. Disk drives are much more reliable and faster then they originally were, but fundamentally, they are the same. From a performance standpoint, disk drives are still one of the most important hardware components to tune. Properly speaking, you can not really tune a disk drive; however, by knowing its performance characteristics and limitations and configuring your system with those limitations in mind, you are, in effect tuning the I/O subsystem.

5.3.1.1 Disk drive characteristics

Here are the descriptions of a disk drive’s characteristics: rotational latency, disk seeks, and average seek time.

Rotational latency

Most of the high performance disk drives out on the market spin at 10,000 revolutions per minute (rpm). If a request for data caused the disk to have to rotate completely before it was able to read the data, this spin would take approximately 6 milliseconds (ms), or 0.006 seconds. For the disk head to read a sector of data, that sector must be underneath the head. Because the disk drive is always rotating, the head simply waits for that sector to rotate to the position underneath it. The time it takes for the disk to rotate to where the data is under the head is called the rotational latency. The rotational latency can be as long as 6 ms (if the disk has to rotate completely), but on average, it is around 3 ms.

The rotational latency is added to the response time of a disk access. So, when you are choosing disk drives for your system, it is extremely important from a performance standpoint that you take into consideration the length of the disks rotational latency. As you have just seen, for 10,000 rpm the rotational latency is around 3 ms.

Disk seeks

When retrieving data, not only must the disk rotate under the heads that will read the data, but also the head must move to the track where the data resides. The disk armature moves in and out of the disk stack to move the head to the cylinder that holds the desired data. The time it takes the head to move to where the requested data resides called the seek time.

The time it takes for a seek to occur depends mainly on how far the disk heads need to move. When the disk drives are accessing data sequentially, the heads need to move only a small distance, which can occur quickly. When disk access are occurring all over the disk drive, the seek time can get much longer. In either case by minimizing the seek time, you improve your system’s performance.

Seek time and rotational latency both add to the time it takes for an I/O operation to occur, and therefore they worsen the performance of the disk drive. Rotational latency is usually around 3 ms for 10,000 rpm disks. The seek time of a disk
varies depending on the size and speed of the disk drive and the type of seek being performed.

*Track to track* seek time is the time the heads take to move between adjacent tracks. This type of seek is used when performing sequential I/O operations. A typical 10,000-rpm disk drive has a track to track seek time of around 0.8 ms. As you can see, for disks with track to track seek time of only 0.8 ms, the rotational latency of approximately 3 ms is the larger factor in the disk drive performance. If the I/O operations are submitted to the disk drive fast enough, the disk drive will be able to access adjacent tracks or even read or write an entire track at a time. However, this is not always the case. In some cases, the I/O operations are not requested fast enough, and disk rotation occurs between each sequential access. Whether this happens typically depends on the design and the speed of the disk controller.

**Average seek time**

The average seek time is the time the heads take on average to seek between random tracks on the disk. According to the specification sheet of an average 10,000 rpm disk drive, the seek time of such a disk is around 6 ms. Because almost all of the I/O operations a database (SQL, Oracle, DB2, and so on) generates are random, your disk drives will be performing a lot of random I/O.

The maximum seek time of this type of disk can be as long as 13 ms. The maximum seek occurs from the innermost track of the platter to the outermost track, or vice versa. This is referred to as a full-disk seek. But normally, the seeks will not be full-disk seeks, especially if the disk is not full.

### 5.3.1.2 Disk drive performance

As discussed several factors determine the amount of time it takes for an I/O operation to occur. The factors are as follow:

- The seek time
- The rotational latency
- Bus transfer time

So the times it takes for an I/O operation to occur is the sum of the times needed to complete the steps described here plus the time added by the overhead incurred in the device driver and in the operating system. Remember the total time for an I/O operation depends mainly on whether the operation in question is sequential or random. Sequential I/O performance depends on track to track seeks. Random I/O performance depends on the average seek time.

**Sequential I/O**

Sequential I/O consists of accessing adjacent data in disk drives. Because track to track seeks are much faster than random seeks, it is possible to achieve much higher throughput from a disk when performing sequential I/O. To get an idea of how fast sequential I/O can occur let's look at an example.

It takes approximately 0.8 ms to seek between tracks on a typical disk drive. If you add the seek time to the rotational latency of 3 ms, you can conclude that each I/O operation will take approximately 3.8 ms. This would theoretically allow us to perform 264 sequential I/O operations per second (because each second contain approximately 264 intervals of 3.8 ms). But, with other overhead factors, such as the OS and drivers, the maximum rate which a disk can sustain is around 250 operations per second (depending on how big the operations are). But if you
run a disk drive at more then 85 percent of its I/O capacity, queuing will occur, therefore the maximum recommended I/O rate is 225 operations per second.

**Random I/O**

Random I/O occurs when the disk heads must read data from various parts of the disk. These random head movements result in reduced performance. Again, let's look at a sample disk we covered earlier. Now instead of taking approximately 0.8 ms to seek between adjacent tracks on the disk, the heads must seek random tracks on disks. This random seeking takes approximately 6 ms (on average) to complete, which is 7.5 times longer than the track to track seeks. A typical random I/O operation requires approximately 6 ms (on average) for the heads to move to the track where the data is held and 3 ms in rotational latency, for a total of 9 ms, giving a theoretical maximum of approximately 111 I/O per second (each second contains approximately 111 intervals of 9 ms). Therefore, using the same rule that you saw earlier, if you run a disk drive at more than 85 percent of its capacity, queuing will occur. Therefore, the maximum recommended I/O rate is 94 I/O operations per second. After taking into account the overhead in the OS, driver and controller you would want to drive that disk at no more the 85 I/O per second.

Remember, when a disk drive is accessed faster than it can handle, queuing will occur, and the latency will increase. In fact if you get to 100 percent, queuing will certainly occur and performance will degrade dramatically, which can cause blocking and deadlocks in the database application.

**Solution to the disk performance limitation problem**

So how do we solve the problem of disk performance limitations? By following these guidelines, you should be able to design an I/O subsystem that performs optimally:

- Isolate sequential I/O operations. By isolating components that are sequential in nature on their own disk, you can maintain that sequential nature. The transaction log is an example of a sequentially accessed file. If you place more than one sequentially accessed file on the same disk, the I/O operations will become random, because the disk must seek between the various sequential components.
- Distribute random I/O operations. Because the I/O operations are random in nature, you can alleviate the load by adding disk drives. If you build a system with enough disk drives to handle the random I/O load, you should not experience any problems.

### 5.3.2 RAID level and performance

Most of us have heard of redundant array of independent disks (RAID) technology. Unfortunately, there is still significant confusion about how RAID actually works and the performance implications of each RAID strategy. Therefore, this section presents a brief overview of RAID and the performance issues as they relate to commercial server environments. RAID was created by computer scientists at the University of California at Berkeley, to address the huge gap between computer I/O requirements and single disk drive latency and throughput. RAID is a collection of techniques that treat multiple, independent disk drives as a unit, with the object of improving performance and/or reliability.
5.3.2.1 Overview of RAID levels

The RAID strategies that are supported by FASiT are:

- RAID-0
- RAID-1
- RAID-3
- RAID-5
- RAID-10

**RAID-0**

RAID-0 is a technique that stripes data evenly across all disk drives in the array. Strictly, it is not a RAID level, as no redundancy is provided. On average, accesses will be random, therefore keeping each drive equally busy. SCSI has the ability to process multiple, simultaneous I/O requests, and I/O performance is improved, because all drives can contribute to system I/O throughout. Because RAID-0 has no fault tolerance, when a single drive fails, the entire array becomes unavailable. RAID-0 offers the fastest performance of any RAID strategy for random commercial workloads. RAID-0 also has the lowest cost of implementation, because no parity information must be stored.

![Diagram of RAID-0](image)

RAID levels greater than 5 like RAID-10, RAID-0+1 or RAID-50 offer additional fault tolerance or performance enhancements. These tend to be proprietary systems, where each vendor has a different method of implementation.

**RAID-0 (not normally recommended)**

RAID-0 is normally not recommended for storing server data files. Because the data, most of the time, is so important to your business, losing that data could be devastating. Because a RAID-0 array does not protect you against a disk failure, you should not use it for any critical system component, such as the operating system.
system, files, database files, or transaction log files. It may be used for data that needs very fast access but can easily be restored.

**RAID-1**
RAID-1 provides fault tolerance by mirroring one drive to another drive. The mirror drive ensures access to data should a drive fail. RAID-1 also has good I/O throughput performance compared to single-drive configurations, because read operations can be performed on any data record on any drive contained within the array. Most array controllers (including the IBM ServeRAID family) do not attempt to optimize read latency by issuing the same read request to both drives in the mirrored pair. The drive in the pair that is least busy is issued the read command, leaving the other drive to perform another read operation. This technique ensures maximum read throughput. Write performance is somewhat reduced because both drives in the mirrored pair must complete the write operation. For example, two physical write operations must occur for each write command generated by the operating system.

![Figure 111. RAID-1](image)

**RAID-1 recommendations**
RAID-1 offers a high degree of fault tolerance and high performance. RAID-1 is a great solution when one disk drive can hold all of the data. Some recommendations for using RAID-1 are:

- Use RAID-1 for the disks that contain your operating system. It is a good choice because the operating system can usually fit on one disk.

- Use RAID-1 for the transaction log. Typically, the database server transaction log can fit on one disk drive. In addition the transaction log performs mostly sequential writes. Only rollback operations cause reads from the transaction logs. Therefore, you can achieve a high rate of performance by isolating the transaction log on its own RAID-1 array. RAID-5 can also be used for transaction logs provided that the RAID implementation uses a write cache
and has advanced destaging algorithms. These are found in systems such as the Enterprise Storage Server (ESS).

- Use write caching on RAID-1 arrays. Because a RAID-1 write will not complete until both writes have been done (two disks), performance of writes can be improved through the use of a write cache. When using a write cache be sure it is battery-backed up.

**RAID-3**
Originally, the Berkeley definitions of RAID levels specified that RAID-3 arrays stripe bytes across the disks. More recently, the RAID Advisory Board definitions allow for block striping across the drives, which is the way RAID-3 is implemented in the FAST Storage Server. Therefore, there is essentially no difference between RAID-3 and RAID-5.

**RAID-5**
RAID-5 offers an optimal balance between price and performance for most commercial server workloads. RAID-5 provides single-drive fault tolerance by implementing a technique called single equation single unknown. This technique says that if any single term in an equation is unknown, the equation can be solved to exactly one solution. The RAID-5 controller calculates a checksum using a logic function known as an exclusive-or (XOR) operation. The checksum is the XOR of all data elements in a row. The XOR result can be performed quickly by the RAID controller hardware and is used to solve for the unknown data element. A significant benefit of RAID-5 is the low cost of implementation, especially for configurations requiring a large number of disk drives. To achieve fault tolerance, only one additional disk is required. The checksum information is evenly distributed over all drives, and checksum update operations are evenly balanced within the array. However, RAID-5 yields lower I/O throughput than RAID-0 and RAID-1. This is due to the additional checksum calculation and write operations required. In general, uncached RAID-5 performance is 30-50% lower than with RAID-1 (write caching and advanced destaging algorithms mitigate the effects of RAID-5 write penalties in disk systems such as the Enterprise Storage Server).

RAID-5 will provide IO throughput performance similar to RAID-0 when the workload does not require many write operations or when the RAID-5 implementation has a write cache. In advanced storage subsystems, such as the Enterprise Storage Server, that have a write cache and advanced destaging algorithms, the performance is as good as a RAID-10 system.
**RAID-5 recommendations**
Because of the additional I/O operations incurred by RAID-5 writes, uncached RAID-5 level is recommended for disk arrays that are used mostly for reading. Because the parity is distributed among the various disks in the array, all disk are used for read operations. Because of this characteristic, the following is recommended:

- Use uncached RAID-5 on read only arrays. Any disk array with more than 10 percent writes is not a good candidate for uncached RAID-5.
- Use write caching on RAID-5 arrays, because RAID-5 writes will not be completed until at least two reads and two writes have occurred. The response time of writes can be improved through the use of write cache (be sure it is battery-backed up). RAID-5 arrays with caching can give as good as performance as any other RAID level and with some workloads the striping affect gives better performance than RAID-1.

**RAID-10**
RAID-10 is a new RAID level offered by the FASiT Storage Server. It provides striping of data across several RAID-1 sub-arrays. This way, you achieve the performance of RAID-0 and also fault tolerance of RAID-1. Another benefit is that the number of physical disk drives can be larger than with RAID-1 logical drives, which are limited to two physical disks. Because the sub-arrays use mirroring, you can only use 50% of the total disk capacity for data. The FASiT storage manager will configure each RAID-1 with more than two drives as a RAID-10 array, you cannot define it explicitly. A RAID-10 array is shown in Figure 113.
Chapter 5. FASTT solution design

Figure 113. RAID-10

**RAID-10 recommendations**

RAID-10 offers high performance and a high degree of fault tolerance. RAID-10 should be used when a large volume is required and more than 10 percent of the I/O operations are writes. RAID-10 recommendations include:

- Use RAID-10 whenever the array experiences more than 10 percent writes. RAID-5 without write cache, does not perform well as RAID-10 with a large number of writes. If the RAID-5 implementation uses write caching and features, such as advanced destage algorithms, and sequential prefetch found in storage subsystems, such as the ESS, then the RAID-5 performance can surpass that of RAID-10.

- Use RAID-10 when performance is critical.

- Use write caching on RAID-10. Because RAID-10 write will not be completed until both writes have been done, performance of the writes can be improved through the use of write cache (be sure it is battery backed up).

RAID-10 is the best fault tolerant solution in terms of protection and performance, but it comes at a cost. You must purchase twice the number of disks that are necessary with RAID-0. If your array is mostly read, RAID-5 might be acceptable.

5.3.2.2 **RAID levels performance comparison**

To properly configure and tune your RAID system, you must understand the performance differences between the various RAID levels, which the previous section outlined.

**Read performance**

The RAID level you choose will not significantly affect read performance. When read operations are performed on a RAID array, each drive contributes to the array's performance. Because random I/O operations are typically the most
problematic, they will be covered here. You can maximize sequential performance by isolating the sequential I/O operations on their own array. Let's look at random-read performance under the various RAID levels:

- RAID-0 volumes spread data evenly among all the disks in the array. Therefore, random I/O operations should be spread equally among all the disk drives in the system. If we estimate that a particular disk drive can handle 85 I/O operations per second, a RAID-0 array of 10 disk drives should be able to handle 850 I/O operations per second.

- RAID-1 volumes supports split seeks so both disk drives perform read operations. Therefore, RAID-1 array can support twice the number of reads that a single disk can, or 170 I/O operations per second. If reads occur more frequently than that, performance will suffer.

- RAID-5 arrays spread the data evenly among all the disk drives in the array. Even though one disk drive is used for parity in each stripe, because the I/O operations are random in nature, all drives are typically used. Therefore, as with RAID-0 array the read capacity of RAID-5 is 85 I/O operations per second times the number of disk drives in the array. An array running at more than that will reduce system performance.

- RAID-10 arrays, like RAID-1 arrays, support split seeks. The maximum read performance is therefore equivalent to the number of disk drives times 85 I/O operations per second. You may be able to initiate I/O operations more frequently, but they will not be completed as you request them.

As you can see, by adding enough disk drives to support your I/O requirements and staying within these limitations, you will optimize your systems performance.

**Write performance**

The type of RAID controller you use will dramatically affect write performance. Again, because random I/O operations are typically the most problematic, they will be covered here. You can maximize sequential performance by isolating the sequential I/O operations on their own array. Let's look at random-write performance under the various RAID levels:

- RAID-0 is the level most capable of handling writes without performance degradation, but you loose fault tolerance. Because RAID-0 does not mirror data or use parity, the performance of RAID-0 is simply the sum of the performance of the individual disk drive. Therefore, a RAID-0 array of 10 disk drives can handle 850 random writes per second.

- RAID-1 arrays must mirror any data that is written to the array. Therefore, a single write to the array will generate two I/O operations to the disk drive. So a RAID-1 array has the capacity of a single disk drive, or 85 I/O operations per second.

- RAID-5 arrays are even slower for write operations. A write to a basic RAID-5 array actually generates two reads from the disks and two writes to the disks. A write to a RAID-5 array generates four physical I/O operations to the disks. Therefore, the write capacity of a RAID-5 array is equivalent to the capacity of a quarter of the disk drives in the array. That is unless the RAID-5 implementation uses write caching and advanced destaging algorithms found in storage subsystems such as the ESS. In this case the striping effect can result in superior performance.
• RAID-10 has the same write characteristics as the RAID-1 array does. Each write to the RAID-10 volume generates two physical writes. Therefore, the capacity of the RAID-10 array is equivalent to the capacity of half of the disk drives in the array.

As you can see, calculating the write capacity of a RAID array is a fairly complex operation. By adding enough disk drives to support your I/O requirements and staying within these limitations, you will optimize your system performance.

The above discussion only deals with basic RAID levels. Advanced storage subsystems such as the ESS which has a large cache and uses advanced destaging algorithms and techniques such as sequential prefetch have performance which is far superior to that offered by controller based RAID systems. Although the ESS is usually configured with RAID-5 arrays its sophisticated features enable it to have performance equivalent or better to that of other systems that use different RAID levels.

5.3.2.3 Disk calculation

To determine how much load is being placed on the individual disk drive in the system, you must perform some calculations. While using the FAStT Storage Server the number of I/O operations per second that performance monitor (Windows 2000) displays is the number of I/O operations that are going to the array. Additional I/O operations that are generated by the controller for fault tolerance are not shown. In fact Windows 2000 doesn’t register that they are occurring, but you must be aware of them for determining the necessary number of disk drives required for optimal performance. The formulas in the following section can help you determine how many I/O operations are actually going to each disk in the array.

**RAID-0**

The rate of I/O operations per disk drive in a RAID-0 array is calculated by adding up all the reads and writes to the array and dividing by the number of disk drives in the array. RAID-0 requires only the simple equation:

\[
\text{operations per disk} = \frac{\text{read} + \text{writes}}{\text{number of disks}}
\]

**RAID-1**

With RAID-1 the calculation become a little more complicated. Because the number of writes is doubled, the number of I/O operations per disk per second is equal to the number of reads plus two times the number of writes, divided by the number of disk drives in the array (two for RAID-1). The equation is as follows:

\[
\text{operations per disk} = \frac{\text{reads} + (2 \times \text{writes})}{2}
\]

RAID-1 is slower on writes but offers a high degree of fault tolerance.

**RAID-5**

RAID-5 offers fault tolerance but has a high level of overhead on writes. RAID-5 reads are distributed equally among the various disk drives in the array, but writes actually cause four physical I/O operations to occur (two reads — one data and one parity, one data to the memory; and two writes — one parity and one data back to disk). To calculate the number of I/O operations occurring on the individual disk drives, you must add the reads to four times the number of writes before dividing by the number of disk drives. Therefore, the equation for RAID-5 is as follows:
However, the effect of caching the writes until there is enough data to do a full-stride write, that is write out the data across all the disks in the array, dramatically improves RAID-5 write performance.

**RAID-10**

RAID-10 is slow on writes, as is RAID-1, but RAID-10 offers a high degree of fault tolerance. The calculation for RAID-10 is the same as that for RAID-1. Because writes are doubled, the number of I/O operation per disk is equal to the number of read plus two times the number of writes divided by the number of disk drives in the array. The equation is as follows:

\[
\text{operations per disk} = \frac{\text{read} + (2 \times \text{writes})}{\text{number of disks}}
\]

**RAID comparison**

Let's compare the RAID levels directly. This might better help you to determine which RAID level is best for your system. When you compare I/O performance across RAID levels one of the most important factors is the read to write ratio. The various RAID levels perform comparably when performing reads, only writes rates differ. You should also consider whether your system needs to be fault tolerant. Finally, you should be aware of the various cost/space ratios. Table 6 shows a summary of the performance characteristics of the four RAID levels commonly used in array controllers. A comparison is also made between small and large I/O data transfers.

Table 6. Performance with RAID levels

<table>
<thead>
<tr>
<th>RAID levels</th>
<th>Data capacity</th>
<th>Sequential I/O performance</th>
<th>Random I/O performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Read</td>
<td>Write</td>
</tr>
<tr>
<td>Single Disk</td>
<td>n</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>RAID-0</td>
<td>n</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>RAID-1</td>
<td>n/2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>RAID-5</td>
<td>n-1</td>
<td>7</td>
<td>7³</td>
</tr>
<tr>
<td>RAID-10</td>
<td>n/2</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>

**Notes:**

1. In the data capacity, n refers to the number of equally sized disks in the array.
2. 10 = best, 1 = worst. You should only compare values within each column. Comparisons between columns is not valid for this table.
3. With the write back setting enabled

As you can see, your best choice really depends on your requirements. To see the difference between RAID-5 and RAID-10 at a different read/write ratios, look at the following table. Table 7 represents 500 I/O operations per second across 10 disk drives with varying read/write ratios.

Table 7. RAID-5 versus RAID-10

<table>
<thead>
<tr>
<th>Read/write ratio</th>
<th>RAID-5 I/O operations (reads + (4 x writes)) / disks</th>
<th>RAID-10 I/O operations (reads + (2 x writes)) / disks</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% reads, 0% writes</td>
<td>(500 + 0) / 10 50 I/O operations per disk</td>
<td>(500 + 0) / 10 50 I/O operations per disk</td>
</tr>
</tbody>
</table>
As you can see, at about 90 percent reads and 10 percent writes, the disk usage is approximately even. But for high percentages of writes, RAID-5 requires much more overhead. That is unless the RAID-5 implementation uses write caching and advanced features found in storage subsystems such as the ESS which mask this overhead.

### 5.3.3 Fibre Channel versus Ethernet (SAN versus NAS)

NAS, which stands for network attached storage, is not a supported method for clustering by Microsoft, but it is still a method which many customers consider.

SAN is a new approach to manage and access storage and is rapidly adopted in the data center: applying a networking model to storage environments. Known as Storage Area Networks (SANs), these network environments provide a scalable, reliable foundation that meets the high availability and performance requirements of today's storage and e-business applications.

The standard technology to enable SAN interconnection, Fibre Channel, was designed and optimized for server-to-storage data communications and high-speed server-to-server interconnection. Fibre Channel combines the flexibility of data communications networks with the specialized performance characteristics of channel technologies to create a scalable and reliable networked environment.

Why was a different approach to networking needed for storage, when Ethernet networks were already a ubiquitous infrastructure for local area networks and the basic infrastructure for the NAS solution?

The answer lies in the unique requirements of storage applications and the fundamentally different networking approaches and applications of Ethernet-based networks and Fibre Channel networks.

Fibre channel combines the best of data channel and networking characteristics. It was designed from the ground up to be optimized for storage environments and to meet the high availability and reliability needs of today's storage and e-business applications. Within a Fibre Channel network, data movement is optimized for the highest possible performance and throughput. Much of the performance penalty in a network is the cost of moving data between the host operating system, the network cards, and the distributed applications. When attempting to use the peak performance of an Ethernet connection, the TCP stack processing on a typical Ethernet network can swallow up an astonishing percentage of high-performance CPU cycles, consuming from 50 to 100 percent.
In Ethernet-based IP networks, operating system (context switching, data copying, etc.) and the TCP stack protocol overhead dominate the actual time spent sending bits on the physical media, regardless of its underlying speed. This performance suits client applications, such as Web browsing, but is not sufficient to support the performance intensive needs of storage applications serviced by Storage Area Networks. The fewer times that the data is copied within the data path, the better the throughput and the lower the cost of the moving data. Fibre Channel with its hardware-assisted flow control, buffer management and segmentation and re-assembly process, combined with direct memory access, allows fast and direct transfer of data from a host bus adapter into application buffers and vice versa. This structure optimizes the data throughput and reduces the memory requirement in the host memory hierarchy.

Another contributor to the performance penalty is the frequent interruption of the processor for handling small packets which often results in degraded data throughput, increased context switching and cache misses. Storage and SAN applications rely on the transfer of large amounts of data. The larger the block of data, the fewer the processor interrupts, and the less the packet processing time, the better the performance. Fibre channel was optimized to transfer large blocks of data with minimum processing overhead. Ethernet was designed to transfer large numbers of small blocks of data.

In Fibre Channel, the atomic data transfer unit is a frame, which can be as large as 2112 bytes. A hardware-based provision is made in the standard to concatenate up to 65536 frames into a single sequence. From a CPU processing perspective, the sequence is equivalent to a packet in networking. The sequencing allows large block transfers (block size can be as large as 128 Mbytes). This is critical during storage data transmission of large amounts of data where minimum processor interruption and negligible packet processing time are desired. Ethernet has a maximum packet size of 1518 bytes. The smaller packet size means that more packets are sent for a large transfer. This results in more processor interrupts and frequent packet-processing overhead, therefore reducing the overall throughput. This makes Fibre Channel a preferred transport mechanism for the deployment of SANs to achieve fast data access, retrieval and transfer of large blocks of data with high throughput and reliability. Just as Ethernet-based IP networks provide a standard communications infrastructure for the applications that companies use to run the business.

Fibre Channel networks provide the foundation for the information that supports it. A Fibre Channel-based Storage Area Network enables companies, for the first time, to use the information that resides on its data storage systems as a virtual resource that is available to the entire enterprise. Fibre Channel has made possible a networked environment for storage that is highly reliable, high performance, and manageable to help companies keep pace with today's rapidly increasing storage requirements.

5.3.4 FASTT Storage Manager — Performance Monitor

FASTT Storage Manager Version 7 is the software that lets you manage the FASTT Storage Server. It includes its own performance monitoring tool, the Subsystem Management Performance Monitor which gives you information about the performance aspects of your storage subsystem.
Note: The FAStT Performance Monitor is not related to the Windows NT Performance Monitor tool.

This section describes how to use data from the Subsystem Management Performance Monitor and what tuning options are available in the Storage Manager for optimizing the storage subsystem performance. Use the Performance Monitor to monitor storage subsystem performance in real-time and save performance data to a file for later analysis. You can specify the logical drives and/or controllers to monitor the polling interval. Also, you can receive storage subsystem totals, which is data that combines the statistics for both controllers in an active-active controller pair. Look at the performance monitor windows in Figure 114.

Table 8 describes the data that is displayed for selected devices.

<table>
<thead>
<tr>
<th>Data field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total I/Os</td>
<td>Total I/Os performed by this device since the beginning of the polling session. For more information, see “Balancing the I/O load” on page 148.</td>
</tr>
<tr>
<td>Read percentage</td>
<td>The percentage of Total I/Os that are read operations for this device. Write percentage can be calculated as 100 minus this value. For more information, see “Optimizing the I/O request rate” on page 149.</td>
</tr>
<tr>
<td>Cache hit percentage</td>
<td>The percentage of reads that are processed with data from the cache rather than requiring a read from disk. For more information, see “Optimizing the I/O request rate” on page 149.</td>
</tr>
<tr>
<td>Current K/B per second</td>
<td>Average transfer rate during the polling session. The transfer rate is the amount of data in kilobytes that can be moved through the I/O Data connection in a second (also called throughput). For more information, see “Optimizing the transfer rate” on page 148.</td>
</tr>
</tbody>
</table>
Balancing the I/O load
The total I/O data field is useful for monitoring the I/O activity to a specific controller and a specific logical drive. This field helps you identify possible I/O hot spots. Identify actual I/O patterns to the individual logical drives and compare those with the expectations based on the application. If a particular controller has considerably more I/O activity than expected, consider moving an array to the other controller in the storage subsystem using the Array >> Change Ownership option.

Since I/O loads are constantly changing, it can be difficult to perfectly balance I/O load across controllers and logical drives. The logical drives and data accessed during your polling session depend on which applications and users were active during that time period. It is important to monitor performance during different time periods and gather data at regular intervals so you can identify performance trends. The Performance Monitor allows you to save data to a comma-delimited file so you can import it to a spreadsheet for further analysis. If you notice that the workload across the storage subsystem (Storage Subsystem Totals Total I/O statistic) continues to increase over time while application performance decreases, this can indicate the need to add additional storage subsystems to your enterprise. By doing this, you can continue to meet application needs at an acceptable performance level.

Optimizing the transfer rate
The transfer rates of the controller are determined by the application I/O size and the I/O request rate. In general, a small application I/O request size results in a lower transfer rate, but provides a faster I/O request rate and a shorter response time. With larger application I/O request sizes, higher throughput rates are possible. Understanding your typical application I/O patterns can give you an idea of the maximum I/O transfer rates that are possible for a given storage subsystem. Because of the dependency on I/O size and transmission media, the only technique you can use to improve transfer rates is to improve the I/O request rate. Use the operating system performance monitor to gather I/O size data so you understand the maximum transfer rates possible. Then use tuning options available in Storage Manager to optimize the I/O request rate so you can reach the maximum possible transfer rate. See Figure 115 for how the I/O request size changes the values for I/O per second and MB/s.

<table>
<thead>
<tr>
<th>Data field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum K/B per second</td>
<td>The maximum transfer rate that was achieved during the Performance Monitor polling session. For more information, see “Optimizing the transfer rate” on page 148.</td>
</tr>
<tr>
<td>Current I/O per second</td>
<td>The average number of I/O requests serviced per second during the current polling interval (also called an I/O request rate). For more information, see “Optimizing the I/O request rate” on page 149.</td>
</tr>
<tr>
<td>Maximum I/O per second</td>
<td>The maximum number of I/O requests serviced during a one-second interval over the entire polling session. For more information, see “Optimizing the I/O request rate” on page 149.</td>
</tr>
</tbody>
</table>
Optimizing the I/O request rate

The factors that affect the I/O request rate include:

- I/O access pattern (random or sequential) and I/O size
- Use of write caching
- Cache hit rate
- RAID level
- Segment size
- Number of drives in the arrays or storage subsystem
- Fragmentation of files
- Logical drive modification priority

Note: Fragmentation affects logical drives with sequential I/O access patterns, not random I/O access patterns.

To determine if your I/O has sequential characteristics, try enabling a conservative cache read-ahead multiplier (four (4), for example) using the Logical Drive >> Properties option. Also, refer to 4.4.3.1, “Logical drive properties” on page 103. Then examine the logical drive cache hit percentage to see if it has improved. An improvement indicates your I/O has a sequential pattern. Use the operating system performance monitor to determine the typical I/O size for a logical drive. Higher write I/O rates are experienced with write-caching enabled compared to disabled, especially for sequential I/O access patterns. Regardless of your I/O pattern, we recommend that you enable write-caching to maximize I/O rate and shorten application response time.

Optimizing the cache hit percentage

A higher cache hit percentage is also desirable for optimal application performance and is positively correlated with I/O request rate. If the cache hit percentage of all logical drives is low or trending downward, and you do not have the maximum amount of controller cache memory installed, this could indicate the need to install more memory.
To prevent data loss or corruption, the controller periodically writes cache data to disk (flushes the cache). When the amount of unwritten data in the cache reaches a certain level, called a start percentage, the controller writes the data to disk. (The write can also be triggered by the amount of time the data has been in the cache.) The controller writes data to disk until the amount of data in the cache drops to a stop percentage level. Start and stop percentages are user configurable. For maximum data protection, you can choose low start and stop percentages. However, in both cases, this increases the chance that data requested by a read command will not be in the cache, decreasing the cache hit rate for writes, and the I/O request rate. Choosing low start and stop percentages also increases the number of disk writes necessary to maintain the cache level, increasing system overhead and further decreasing performance.

If an individual logical drive is experiencing a low cache hit percentage, consider enabling cache read-ahead (or pre-fetch) for that logical drive. Cache read-ahead can increase the cache hit percentage for a sequential I/O workload. If cache read-ahead is enabled, the cache reads the data from the disk. But in addition to the requested data, the cache also fetches more data, usually from adjacent data blocks on the drive. This feature increases the chance that a future request for data could be fulfilled from the cache rather than requiring a disk access. The cache read-ahead multiplier values specify the multiplier to use for determining how many additional data blocks are read into cache. Choosing a higher cache read-ahead multiplier can increase the cache hit percentage. If you have determined that your I/O has sequential characteristics, try enabling an aggressive cache read-ahead multiplier (8, for example) using the Logical drive >> Properties option. Then examine the logical drive cache hit rate to see if it has improved. Continue to customize logical drive cache read-ahead to arrive at the optimal multiplier. (In the case of a random I/O pattern, the optimal multiplier is zero.) Be aware that a great multiplier value uses a lot of the cache memory to store the read ahead data. This can lead to performance decrease for other logical drives. You must balance between the optimal value for a particular logical drive and for the whole storage subsystem.

**Choosing an appropriate RAID level**

Use the read percentage for a logical drive to determine actual application behavior. Applications with a high read percentage will do very well using RAID-5 logical drives because of the outstanding read performance of the RAID-5 configuration. However, applications with a low read percentage (write-intensive) do not perform as well on RAID-5 logical drives because of the way a controller writes data and redundancy data to the drives in a RAID-5 array. If there is a low percentage of read activity relative to write activity, you might consider changing the RAID level of an array from RAID-5 to RAID-1 for faster performance. You can find more information on RAID options in 5.3.1, “Disk drive performance characteristics” on page 134.

**Choose an optimal logical drive modification priority**

The modification priority defines how much processing time is allocated for logical drive modification operations versus system performance. The higher the priority, the faster logical drive modification operations complete, but the slower system I/O is serviced. Logical drive modification operations include reconstruction, copy back, initialization, media scan, defragmentation, change of RAID level, and change of segment size. The modification priority is set for each logical drive using a slider bar on the Logical drive >> Properties dialog. There are five relative settings on the reconstruction rate slider bar ranging from low to highest.
The actual speed of each setting is determined by the controller. Choose the low setting to maximize the I/O request rate. If the controller is idle (not servicing any I/O) it will ignore the individual logical drive rate settings and process logical drive modification operations as fast as possible. Be aware that a low modification priority prolongs the critical time after a drive failure until all data is rebuilt to a hot spare drive.

Choosing an optimal segment size

A segment is the amount of data, in kilobytes, that the controller writes on a single drive in a logical drive before writing data on the next drive. Data blocks store 512 bytes of data and are the smallest units of storage. The size of a segment determines how many data blocks it contains. For example, an 8 KB segment holds 16 data blocks and a 64 KB segment holds 128 data blocks.

Note: The segment size was expressed in number of data blocks in previous versions of this storage management software. It is now expressed in KB.

When you create a logical drive, the default segment size is a good choice for the expected logical drive usage. The default segment size can be changed using the Logical drive >> Change Segment Size option.

If your typical I/O size is larger than your segment size, increase your segment size in order to minimize the number of drives needed to satisfy an I/O request. This technique helps even more if you have random I/O access patterns. Using a single drive for a single request leaves other drives available to simultaneously service other requests.

If you are using the logical drive in a single-user, large I/O environment such as multimedia application storage, performance is optimized when a single I/O request can be serviced with a single array data stripe (the segment size multiplied by the number of drives in the array used for I/O). In this case, multiple disks are used for the same request, but each disk is only accessed once.

Minimize disk access by defragmentation

Each access of the drive to read or write a file results in spinning of the drive platters and movement of the read/write heads. Make sure the files on your array are defragmented. When the files are defragmented, the data blocks making up the files are next to each other, so the read/write heads do not have to travel all over the disk to retrieve the separate parts of the file. Fragmented files are detrimental to the performance of a logical drive with sequential I/O access patterns. It depends on the host operating system and the file system used on the logical drive if fragmentation can influence the performance.

5.3.5 FASTT best practices based on observations

The topics in this section deal with best practices based on observations from different environments with characteristics concerning size, performance, controllers, I/O operations, and so on.

FASTT200 Data Streaming — Read

With data streaming applications, like video-on-demand, it’s better use the dual controller version of the FASTT200 as long as you can drive both controllers. From 8KB I/O size and above to 512KB I/O size, you actually double the MB/s throughput when using a dual controller rather than only one. For example, with
256KB I/O transfer size, you can achieve around 80MB/s with one controller and around 160MB/s with two controllers (Figure 116).

**FASiT200 versus FAST500 in a specific OLTP environment**

Based on testing that was done by an IBM performance team with the configuration of FASiT with a 16KB segment size, 4KB FASiT cache block size and 60 disk drives, the FASiT500 achieved almost three times the I/O operations per second than FASiT200 in a workload of 8KB OLTP solution (online transaction processing). The main reason is because of the different RAID controllers which the FASiT500 has.

**Cache block size (4KB versus 16KB)**

The cache block size is the minimum memory allocation unit you can define for the controller cache. There are two options for sizes — 4KB and 16KB. If your I/O operations are random with less than 8KB transfer size it is best for you to use the 4KB cache block size. If your I/O operations are sequential or if your I/O operations are random with transfers 16KB or greater, use 16KB cache block (Figure 117).
If you are using RAID 5 arrays, then the comparative performance is as shown in Figure 118.

Figure 118. A 4KB cache block size versus 16KB for RAID 5 arrays
**FAStT500 two dual loops**

Arbitrated loop is a shared interconnect in which devices on the loop must arbitrate to gain access to the loop before frames can be sent. With this architecture, bandwidth and latency are not guaranteed when utilization is high. The FAStT500 has a host attachment side and a drive attachment side. On the drive side, we have two dual loops that we can utilize for EXP500 attachments. Each of the dual loops can hold up to 11 EXP500 boxes.

Since each loop can utilize 100MB/s, we recommend that you split drive enclosures (EXP500) over two loops in order to utilize both of the dual loops for performance, instead of only one dual loop even in a case you have only two EXP500. You definitely split disk drives over the loops when configuration requires 60 disk drives or more. See Figure 119 for how to connect the EXP500 to the FAStT500 to utilize the two dual loops.

![Figure 119. FAStT500 with two dual loops](image)

**FAStT read ahead multiplier (prefetch)**

As described earlier in 5.3.4, “FAStT Storage Manager — Performance Monitor” on page 146, the multiplier is actually:

\[
\text{application request size} \times \text{read ahead multiplier}
\]

This equals the amount of extra data from disk that cached after a read. With random I/O operations you should leave this with the value of 0. For a data streaming read I/O operations, you may benefit from choosing a value from one to four (1 - 4). But in any case, use the FAStT performance monitor to check cache hit rate.
**FAStT cache mirroring**

The reason to have cache mirroring is to provide cached data integrity in the event of a controller failure, mainly when using write cache. The mirroring in the FAStT functions only if the controllers are in active-active configuration.

Because mirroring occurs across drive-side FC loops at the FAStT, it can compete with normal I/O operations processing on the drive-side loops. With sequential writes the performance degradation is most extreme. Therefore, in case you do not need this feature, disable it from the FAStT Storage Manager. (Make sure you are not using the write cache or that the information there is not important to you, otherwise, you are not safe when it is turned off.) As you can see in Figure 120, the performance degrades with mirroring enabled especially when arriving to high throughput like 100MB/s.

![FAStT500 Throughput as a Function of Transfer Size](image)

**Figure 120. Cache mirroring**

**10km fiber optic cable direct attached versus fabric**

When you direct attached a host adapter to a RAID controller through long distance, like 10km, the acknowledgments take too long to return from one end to the other. And, because the host adapter and the RAID controllers have relatively small buffer capacity, the host adapter can only send a small number of 2k frames and then must wait for an acknowledge before sending the next set of frames. The result is that the bandwidth is not utilized.

Instead of using direct attached, you can use a fabric where you install two switches between the host adapter and the RAID controller. Because switches have a substantial internal buffer capacity and as a result a larger buffer, credit is established when one switch logs into the other and therefore throughput increases due to the fact that more frames can be stuffed into the fiber. You can see, in Figure 121, the performance that you can gain while using 25m cable, 10km fabric and 10 km direct attached.
5.4 Clustering

Because clustering can be a very large topic with many specific operating system solutions, we decided to cover only Microsoft clustering.

5.4.1 Concepts of Microsoft Cluster Server

To design storage solutions in a Microsoft clustering environment, you should first understand how the Microsoft Cluster Server works.

This section provides an explanation of server cluster terms and key concepts. We will also cover how resources and groups function within a server cluster, because this is key to storage design.

5.4.1.1 Introduction to Microsoft Cluster Server

A server cluster is a group of computers and storage devices that work together and can be accessed by clients as a single system. The individual computers in the cluster are referred to as nodes, and they act together to provide automatic recovery from failure of clustered services and applications.

The nodes communicate with each other over a high performance, reliable network, and share one or more common storage devices. The cluster service running on both nodes monitors all components of the cluster and the communication between the nodes; if a component, for example, a service fails, the cluster service will move all resources needed for this service, as shared drives or network names, to the surviving node and restart the service.
Clients communicate to logical servers, referred to as a virtual servers, to gain access to grouped resources, such as file or print shares, services like Internet Information Server, and application like Microsoft Exchange. When a client connects to a virtual server, the node controlling the requested resource, service or application will handle the request (Figure 122).

Figure 122. Microsoft Server Cluster

5.4.1.2 Key concepts of server cluster
Cluster server architecture consists of two parts, the logical cluster components and the physical cluster components. Clients interact with the logical cluster components. The physical cluster components manage the logical cluster components.

Microsoft Cluster service is the collection of software that manages all cluster-specific activity (Figure 123).

Logical resources
The logical cluster resources are groups of resources, such as virtual servers, file/print shares; and services, such as Internet Information Server (IIS). An administrator will either create or manage logical server cluster resources.

Groups
Groups are a collection of resources that are managed as a single unit for configuration purposes. Operations that are performed on groups, such as taking groups offline or moving them to another node, affect all resources contained within that group. Usually a group contains all of the elements that are needed to run a specific application, and for client systems to connect to the application.

Note: Cluster service moves entire groups, not single resources, from one node to another in a cluster if a single resource fails; the entire group to which it belongs will fail over.

Resources
Resources are the basic unit managed by the cluster service. Examples of resources are physical hardware devices, such as disk drives, or logical items such as Internet Protocol (IP) addresses, network names, applications, and
services. A cluster resource can only run on a single node at any time, and it is identified as *online* when it is available for use by a client. The resources are monitored by the resource monitor, that is part of the cluster service. This monitor is responsible for watching the status of each resource, if a failure occurs the resource monitor triggers the cluster service to handle the failure and bring the resource online again by moving it to another node.

**Virtual server**

Virtual server refers to a unique server name that appears as a physical server to clients. Cluster service uses physical servers to hold the virtual server. A virtual server has an IP address and a network name that are published to clients on the network. Users access an application or service on a virtual server in the same way that they would if the application or service were on a physical server.

**Physical resources**

Physical resources provide data storage and processing for the logical cluster server resources. You assign physical resources to a specific logical resource.

**Nodes**

Nodes are the unit of management for the cluster server. They are also referred to as systems and the terms are used interchangeably. A node can be an active or inactive member of a cluster, depending on whether or not it is currently online and in communication with the other cluster nodes. If the node is offline, the cluster service is not running or the node is turned off.

**Note:** Windows NT4 Enterprise Edition and Windows 2000 Advanced Server support a two-node cluster, and Windows 2000 Datacenter supports a four-node cluster.

**Cluster disk(s)**

The cluster disk is a shared disk device to which both server cluster nodes attach by means of a shared SCSI bus. All resources of the cluster must be stored on these shared disks, for example, data for file and print shares, applications, resources and services.

**Quorum resource**

The quorum resource plays a vital role in allowing a node to form a cluster and in maintaining consistency of the cluster configuration for all nodes. The quorum resource holds the cluster management data and recovery log. The nodes arbitrate for the quorum resource to determine which node controls the cluster. The quorum resource resides on one particular shared disk. It is best to use one of the shared disks only for the quorum resource, so that it will not be affected by the failover policies of other groups or resources in the cluster.

**Heartbeat**

Each node in a cluster periodically exchanges IP packets with the other node in the cluster to determine if the other nodes are operational. This process is known as heartbeat. If a node does not respond to a heartbeat, the other node assumes that the unresponsive node is offline and tries to take ownership of the resources owned by the non-responding node (Figure 123).
Figure 123. Key cluster concepts

**Application and services in a cluster server**

A cluster service interacts with applications and services that an administrator configures to run on the cluster. Applications and services are either cluster-aware or cluster-unaware. A cluster service can only manage applications properly that are cluster-aware.

Cluster-unaware applications that run on a cluster can cause severe problems if, for example, the application is moved to another node, but the application does store some configuration information on local disks instead of shared cluster disks. This information is not available anymore after the failover. A cluster-unaware application should never be used in a cluster environment, as it may seem to work, but collapse in certain, unpredictable circumstances.

Cluster-aware applications run on a cluster node and can be managed as cluster resources, they can use features that a cluster service offers through the cluster Application Programming Interface (API), for example:

- Report status upon request to the resource monitor.
- Respond to requests to be brought online or to be taken offline without data loss.

Some applications bring their own cluster resources, which allows a more granular monitoring and managing of the application. Other applications, even cluster-aware, run as a generic application or service and can only report a limited amount of information to the cluster service. The resource monitor can only report whether the application is running, not whether it is running properly. The resource monitor determines whether these generic resources are online or have failed. If they have failed the resource monitor triggers the cluster service to restart the resources on the other node.
As an administrator trying to obtain the best performance and reliability from your system, you should not only select cluster-aware applications, but cluster-aware applications that support:

- Data maintained in a configurable location
- Transaction processing
- The Cluster Service API

You can program most database applications, transaction processing applications, file and print server applications, and other groupware applications to be cluster-aware.

**Failover and failback**

Microsoft Cluster Service provides your system with the ability to reassign control of groups of logical resources in the event of a failure. If a resource fails, the cluster service will attempt to restart that service. You configure the failover and failback policies to determine when groups should transfer ownership from one node to another.

**Failover** occurs when one of the nodes in a cluster failed or is in the offline state. The resources and groups controlled by this node fail over to the other node in the cluster. For example, in a cluster where Microsoft Internet Information Services (IIS) is running on Node A, and Node A fails, IIS will fail over to Node B of the cluster.

**Failback** is the process of returning a resource or group of resources to the node on which it was running before a failover occurred. Using the preceding example, when Node A comes back online, IIS can fail back to Node A. This process can be performed automatically or manually. The administrator determines when and if a group should fail back to the preferred owner. For example, if Node A fails, the resources in a group could take five minutes or more to restart on Node B. To avoid additional delays in responding to client requests, the administrator can choose to fail back this group to Node A at a time when there will be no client requests, or to leave the ownership of the group with Node B and shift the groups manually.

**Groups and resources considerations**

A Microsoft Cluster solution can contain many resources. For administrative purposes, you can logically assign resources to groups. Some examples of resources are applications, services, disks, file shares, print shares, TCP/IP addresses, and network names. Multiple groups may be created within the cluster so that you can distribute resources among nodes in the cluster. The ability to distribute groups independently allows more than one cluster node to handle the workload.

A group usually contains all resources that are needed for one virtual server. For example a virtual file server group would contain the disks, a network name and IP address and the file shares as resources. Not all resources can be brought online before others are online. In a group you can define dependencies between resources, for example, the IP address must be online before the name can be brought online, or the file shares depend on the disk resource.

In the event of a failure within a group, the cluster software transfers the entire group of resources to a remaining node in the cluster. All resources will be moved with the group even if they did not fail. Therefore, clients on the network may still
access the same resources by the same network name and IP address (Figure 124).

**Resource states:** When a resource is offline, it is unavailable for use by a client or another resource. When a resource is online, it is available for use. The initial state of any resource is offline until the administrator brings it online. When a resource is in one of the pending states, it is in the process of either being brought online or taken offline.

You can set a resource to enter the failed state if it cannot be brought online or taken offline after a specified amount of time. You can then configure the time-out value to specify the amount of time that the cluster service should wait before failing the resource over to another node.

Resource state changes can occur either manually (when you use the administration tools to make a state transition) or automatically (during the failover process). When a group is failed over, the cluster service alters the states of each resource according to their dependencies to other resources in the group.

**Resource dependencies:** The administrator can establish resource dependencies within a group to ensure availability of specific resources before other resources attempt to go online. For example, a file share resource may require a physical disk to be online to provide data to clients. The only dependency relationships that the cluster service recognizes are relationships between resources.

![Diagram of groups, resources and virtual servers](image_url)
5.4.2 Cluster Server configuration

You can configure the cluster server to meet specific requirements. The configuration that you choose depends on the scalability features of your application and your availability objective for the resources.

5.4.2.1 Active/passive cluster configuration

This configuration provides high performance and failover capacity. One node of the cluster makes all of the resources available. The other node is a dedicated hot spare, that is ready to provide the identical performance if a failover occurs. The passive node should have the same specifications, such as CPU speed, as the original node that controlled the resource. This configuration is called active/passive, because only one node is actively providing resources to clients.

The disadvantage of this configuration is that it is an expensive allocation of hardware. One of the two servers will not be servicing any clients until a failure occurs.

In Figure 125, Node A has control of Group 1. The administrator has configured Node B as the hot spare with the capability to control Group 1. If Node A fails or goes offline, Node B will control Group 1. When Node A returns to an online state, Node A becomes the hot spare and Group 1 remains with Node B. Because failback does not occur, this configuration provides maximum availability by reducing the time that the service or application is unavailable.

Choose this configuration when you need to provide critical applications and resources with high availability. For example, an organization that is selling...
products on the World Wide Web could justify the expense of having an idle server by guaranteeing continuous high performance access to customers.

**Availability**

This configuration provides very high availability with the added benefit of no performance degradation during failover.

**Failover policy**

The hot spare node should provide identical performance to the failed node. Configure the failback policy for failover, but not for failback.

### 5.4.2.2 Active/active configuration

The active/active configuration with static load balancing provides optimum performance, because the configuration balances resources across both nodes in the cluster. Each node controls a different resource. Static load balancing refers to the type of failback policy that the administrator configures. If one node fails, the other node will temporarily take on all of the groups. When the failed node comes back online, the group fails back to the original node, allowing performance to return to normal. In general, this configuration is the one most often used in server clusters. This configuration is called the active/active configuration, because both nodes are actively providing resources to cluster clients. Depending on the resources and the capacity of the nodes, performance may suffer when a failover occurs and a single node must run all of the resources.

In Figure 126, Node A is the primary owner of Group 1, and Node B is the primary owner of Group 2. If Node B goes offline, Group 2 will fail over to Node A. When Node B goes back online, Group 2 will fail back to Node B. Performance is restored when the failed node comes back online and the group fails back to its original node.
Figure 126. Active/active configuration

Choose this configuration when you need to provide multiple resources simultaneously from a single cluster, provided that you can accept reduced performance during a failover.

**Availability**

This configuration provides high performance until failover. When a single node runs all of the resources, performance will be degraded.

**Failover policy**

Configure all of the groups to fail over, and then to fail back when the original node owner is back online.
5.4.3 Design Microsoft Cluster Server — The storage perspective

Here we describe how to design a Microsoft Cluster Server. While there are some different issues regarding cluster design, we only discuss the storage design for the cluster solution, and this is the most important when designing cluster solutions for your customer.

5.4.3.1 Storage architecture
A proper storage design is the key for the cluster deployment.

**SCSI RAID or Fibre Channel**
The first question you should ask yourself and your customer is what kind of shared storage technology you should use for your cluster.

Well you have two supported options from Microsoft and IBM, the first is the SCSI RAID, the second and most recommended is the Fibre Channel. While there are many SCSI RAID clusters installed, if you need to choose today, we strongly recommend Fibre Channel rather than SCSI RAID. There are many reasons why to choose fibre attached storage subsystems:

- You can scale your cluster storage based on Fibre Channel more than with SCSI.
- With Fibre Channel you can build fully redundant solutions, so the servers are attached to the disks with dual path and no single point of failure.
- Unlike SCSI, the Fibre Channel protocol was designed to be used with multiple controllers. ID conflicts which are common on SCSI are resolved automatically with intervention from the administrator.
- Fibre Channel storage solutions do not have the bus termination problems that you can experience in an SCSI cluster.
- With one Fibre Channel storage subsystem as the FASIT Storage Server, you can build multiple clusters, while a SCSI solution is limited to one cluster.

5.4.3.2 Cluster storage — Best practices
We will give some advice on the design of the disk layout for a cluster solution. This advice is based on practical experience as well as recommendations by Microsoft.
**Quorum**

As we described earlier, Microsoft Cluster Server needs one shared disk storing the quorum resource. We strongly recommend that you use one of the shared disks in the cluster only for the quorum resource, so that it will not be affected by the failover policies of other resources. While we speak on the shared disks, we mean a logical drive in our FASTT Storage Server that is presented as a shared physical disk to the cluster service. For example, a FASTT 200 Storage Server can hold a ten disk RAID level 5 array, where we define one logical drive of 1GB for the quorum disk instead of using a separate two drive RAID level 1 array, which can lead to a lot of wasted space if used only for the quorum. You also have to consider performance when designing the arrays. It may be better to use a small logical drive for the quorum resource on a two disk RAID level 1 array which also holds other logical drives storing the log files of a database. As the logical drives are presented as physical disks to the operating systems, you could define multiple partitions. Be aware that multiple partitions on a single physical disk are not supported by Microsoft for use as a shared cluster disk.

**Data**

The best practice guideline with storage design is to build a separate array for each data pattern. For example, assume you have ten 18.2GB disk drives in your FASTT 200 Storage Server and you already built a two drive RAID level 1 array for the quorum. If you now need to cluster a transactional based application, it is best to build a RAID-1 array of two 18.2GB disk drives for the transaction logs which have a sequential pattern. You may define a second logical drive in the first array to store the mirrored log file. For the data, build a five drive RAID-5 array since the data has a random access pattern. The remaining 18.2GB drive can be used as a global hot spare (Figure 127). If you have heavy writes it maybe better to use RAID-10 for the data, because you save the parity performance penalty but you will waste more disks.
A key factor when designing the storage for cluster service is to consider how many groups you are planning, because each group with an application needs at least one physical disk resource, and a particular resource cannot be a member of more than one group. Also you need to consider which cluster configuration you choose, active/passive or active/active. For an active/active configuration, you need at least one group per active node. To understand this issue, we present it through an example of clustering Exchange 2000 Enterprise.

5.4.4 Design Exchange 2000 cluster

Microsoft Exchange 2000 Server represents a significant technological evolution compared to previous versions. With Microsoft's focus on scalability, reliability, and availability in Exchange 2000, clustering support has been greatly enhanced. While Exchange Server version 5.5 supported clustering technology, functionality was limited and return on investment was often difficult to find. With Exchange 2000 and Windows 2000, clustering will become a viable option for organizations seeking to provide higher levels of availability for Exchange deployment projects and server consolidation efforts.

Exchange 2000 store architecture

Storage groups and stores in Microsoft Exchange 2000 provide the containers in which you store data. You have great flexibility in configuring these containers to fit your environment and to efficiently handle data (Figure 128).

Exchange 2000 supports multiple message databases on each server. Creating multiple databases enables greater scalability, efficient management, increased reliability and a reduction of backup and restore times.
The store
A store is a database that houses data. Exchange 2000 can support multiple stores on each server. Stores have no programmed size limit, so you can use multiple stores to enhance the flexibility of backup and restore tasks. There are two types of stores in Exchange 2000, mailbox stores and public information stores. A mailbox store contains user data and a public store contains a public folder (or shared) data. Each store is a logical database that has an associated streaming store file containing native Internet content.

Each store consists of the following database files:

- The streaming database file (.stm)
  The .stm file contains common Internet formatted content, such as native Multipurpose Internet Mail Extensions (MIME) content, that protocols other than the Messaging Application Programming Interface (MAPI) protocol places in the store.

- The rich text database file (.edb)
  The .edb file contains data placed in the store through MAPI, as well as the database tables that define mailboxes, messages, folders and attachments.

Because the .stm file only contains raw document content, which is referenced by the corresponding edb file, the streaming database and rich text database files that comprise a particular database are inseparable.

The benefits of multiple message databases include:

- Increased system reliability, because a failure in one database does not affect users in another database.
- Faster and more flexible backup scheduling is possible, because databases are typically smaller.
- Decreased recovery time in the event of hardware failure, because each database can be restored individually.

Storage group
A storage group is a set of stores that share the same set of transaction log files. A storage group contains up to five stores that use one set of transaction log files. Exchange 2000 uses storage groups to reduce the overhead of multiple sets of transaction log files. You can manage these stores as a group or independently.

Benefits of storage groups include:

- Support more users on each server, because multiple smaller stores can be created and managed more easily.
- Perform backup and restore activities on a single store while other stores in the storage group remain in operation.
- Host multiple businesses on a single server. Each company can have its own store or storage group. You can configure and maintain each storage group according to the requirements of the associated company.
- Provide individual support for critical mailboxes. For example, you may have one or more critical mailboxes that must be recovered individually as quickly as possible in the event of an emergency or disaster. You can configure each mailbox in its own dedicated store, enabling you to perform individual backup and recovery. The more stores and storage groups you create, the more
Exchange 2000 resources are required. For this reason, it is important to weigh the impact on resources against the business need for creating additional stores and storage groups.

- Use circular logging for a specific storage group. Circular logging enables Exchange 2000 to use and reuse a small set of transaction log files. For example, you may have a store that generates a volume of transactions that do not need to be recovered, such as a public information store that receives a news feed. If you place this store in its own storage group, it can use circular logging. You should disable circular logging for the other storage group(s) in this instance.

**Storage group limits**
You can create up to four storage groups for each Exchange 2000 server. Exchange 2000 creates an additional temporary storage group during restore operations. Each storage group can support five stores. Stores do not have a size limit, although administrators should limit their size so that they can easily backup or restore the stores.

**Transaction log files**
Transaction log files are history files recording server activity. These files are useful in restoring and backing up Exchange 2000 data. All Exchange 2000 transaction logs are 5 MB in size. Each storage group uses its own set of transaction log files. For example, if a storage group contains five stores, all transactions for all five stores are recorded in a single series of transaction log files. You can determine where to locate the transaction log files for each storage group. Transaction log files are the most important files for recovery, because they reflect all the transactions that have taken place up to the point of a system failure. ESE even saves those transactions that have not yet been written to the database file. To increase performance and reliability, place the log files on separate hard disk spindles for each storage group. Make sure that these log files are on separate hard disk spindles from the database files for the stores in the storage group.

**Extensible Storage Engine (ESE)**
Extensible Storage Engine (ESE) enables storage groups and stores to function. ESE manages the Exchange 2000 database. There are three basic aspects of database management: storing data; adding, deleting or changing data; and recovering data in the case of a system failure.
**Clustering support in Exchange 2000**

The goal for Exchange 2000 Enterprise was to build upon the initial clustering support provided in Exchange 5.5 Enterprise Edition and provide full application functionality in a clustered environment. Exchange 2000 Server supports active/active clustering. When any member resource of a resource group fails on a cluster node, the resource group will be failed over to another node in the cluster that will take over the services being provided by that resource group. One or more Exchange 2000 virtual servers can exist in the cluster, and each virtual server runs on one of the nodes in the cluster.

Exchange 2000 can support multiple virtual servers on a single node. From an administrative perspective, all components required to provide services and a unit of failover are grouped into an Exchange Virtual Server (EVS) in Exchange 2000. An EVS, at a minimum, will include a storage group, and required protocols. From the viewpoint of Microsoft Cluster Service, an Exchange Virtual Server exists as a resource in each cluster resource group.

If you have multiple Exchange Virtual Servers that shared the same physical disk resource (that is, the each has a storage group that resides on the same disk device), they must all exist within the same resource group and cannot be split into separate resource groups. This is done to ensure that the resources and virtual servers all failover as a single unit and is an administrative restriction that ensures that resource group integrity is maintained. Clients connect to the virtual servers the same way that they would connect to a stand-alone server. The cluster service monitors the virtual servers in the cluster.

In the event of a failure, the cluster service restarts or moves the affected virtual servers to a healthy node. For planned outages, the administrator can manually move the virtual servers to other nodes. In either event, the client will see an
interuption of service only during the brief time that the virtual server is in an online/offline pending state.

**Design the cluster storage**

As you understand from the overview above one of the most challenging but most important parts of deploying Exchange 2000 clusters is storage planning. With Exchange Server 5.5 clusters, only one virtual server technically existed in the cluster and storage allocation from the shared cluster storage was simplified. With Exchange 2000 Server, the support of multiple virtual servers and storage groups per node significantly complicates cluster deployment and management. Regardless of this challenge, storage design must be done right the first time in order for a successful implementation. The success and popularity of Storage Area Network (SAN) technology as a shared storage mechanism for Windows 2000 Cluster Server will facilitate more learning and a faster progression on the learning curve of storage design and allocation in a clustered environment.

When configuring Exchange 2000 in a clustered environment, you need to carefully plan the logical units that you want to share between the member nodes in the cluster. In fact, share is not the most appropriate word because Microsoft Clustering for Windows 2000 works in a shared-nothing model. This simply means that a logical unit can be owned, and therefore accessed, by only one member of the cluster at any point in time. The first step is to take a “backwards” approach to the hardware design and setup for a cluster. Start with the Exchange configuration and work backwards.

For example, if you plan on deploying a two-node cluster (four node Exchange cluster is also supported with Windows2000 Datacenter server) running Exchange 2000 Server, decide the user load requirements for the entire cluster first. As an example, suppose you want to support 4,000 users on a 2-node Exchange 2000 cluster. Evenly dividing these users across the cluster would yield 2,000 users per node. You could then design each cluster node to meet the performance and scalability requirement for 2,000 users. The next step would be to determine the failover scenarios required by the user and cluster configuration. As for now Exchange 2000 will limit the number of storage groups per server to four. This means that each cluster node can never have more than four Exchange Storage Groups running on it at any time. This limit is particularly important in a cluster failover condition. If a failure has occurred and an Exchange virtual server has moved to another node, the total number of storage groups is still limited to four. So clusters must be designed with this limitation in mind.

For our 4,000 user/2-node cluster example, failover rules must be configured in a manner that prevents any single node from exceeding the maximum four storage group per node limitation. Once you have considered the per-node storage group limitations, you can determine how many users per storage group will be configured. Again, since one Exchange virtual server can contain multiple storage groups, care must be taken to ensure the four-per-node is not exceeded during both normal and failover conditions. In the 4,000-user cluster example, let's keep it simple and plan for one virtual server per node and one storage group per virtual server (a ratio of 1:1). This means that one Exchange virtual server and one storage group would service all 2,000 users on each node. Continuing to work backward, we can now begin to plan storage requirements and configuration for each cluster node. Using well-known best practices for maximizing disk I/O is the best approach here. The general rule is to separate sequential from random I/O. An Exchange 2000 database (Store) actually consists of two files – the
properties store (*.EDB) and the streaming store (*.STM). The properties store is a B-Tree database structured file that is accessed in a random I/O fashion. The streaming store is structured in clusters of 4K pages and is typically accessed in a sequential manner. In addition, these files have different access characteristics depending upon the type of clients that will be supported. For MAPI protocol clients (Outlook), the streaming store is not utilized. For Internet protocol clients (such as IMAP (Outlook web), POP3, HTTP, and SMTP) the streaming store is the primary storage location with certain properties being stored in the properties store. Each Exchange storage group has one set of shared database transaction logs and can be configured with multiple database files (an .EDB and .STM pairing). Using our general rule each storage group should have a dedicated disk volume (preferable RAID-1 or 10) in which to store the transaction log files (*.LOG) since they are accessed in a strictly sequential manner.

Depending upon the clients supported, you may also choose to separate the streaming store and the property store on to separate physical arrays as well. However, based on the cost effectiveness of such a configuration, most deployments will typically choose to place both the property and the streaming store on the same volume (configured as RAID-5 or 10 for maximum performance and data protection).

Table 9 identifies each Exchange database component and the best practices that should be followed for optimal design.

<table>
<thead>
<tr>
<th>Database component</th>
<th>Storage design best practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Group Transaction Logs</td>
<td>Sequential I/O: dedicate a RAID-1 or 10 array to each storage group for transaction logs.</td>
</tr>
<tr>
<td>Property Store (*.EDB)</td>
<td>Random I/O: dedicate a RAID-1, 10, or 5 array to each storage group for the property store. Can be combined with streaming store if no or few Internet protocol clients are supported. For MAPI clients, combine with streaming store. For heavily I/O environments, a separate array for each property store in a storage group (up to 5 can be configured) may be necessary.</td>
</tr>
<tr>
<td>Streaming Store (*.STM)</td>
<td>Mostly Sequential I/O: dedicate a RAID-1, 10, or 5 array to each storage group for the streaming store. Can be combined with the properties store if no or few Internet protocol clients are supported. For MAPI clients, combine with property store. For Internet protocol clients in heavily I/O environments, a separate array for each streaming store in a storage group may be necessary. However, this will double storage requirements in a cluster.</td>
</tr>
</tbody>
</table>

When you have determined the number of users per node, the number of virtual servers per node, and the number of storage groups per virtual server, you can begin to design your cluster shared storage configuration using the storage design best practices and recommendations discussed above. In Figure 129 on page 173, consider a sample cluster design for a 2-node cluster supporting 4,000 users, where each virtual server configured with one storage group and four databases per storage group, each database holds 500 users.

From the Microsoft Cluster perspective, Node-A holds the Cluster group and group-2, which is Exchange virtual server-1; and for that Node-A needs three
physical disks one for the Quorum at the cluster group and two for the EVS1 — one for data and one for logs. Node-B holds group-2, which is EVS-2 that needs two physical disks — one for the data and one for logs.

From the storage perspective, in order to have the cluster configuration in our example, you need to define five arrays and five logical units before you start your Exchange cluster installation.

5.4.5 SQL2000 Cluster recommendation — The storage perspective

The same rules for a transaction based application apply also with SQL2000 server. What we would like to express here is one of the unique features of SQL2000 that you need to know when designing storage for cluster or stand alone. It is called filegroups.

**Placing tables on filegroups**

A table can be created on a specific filegroup rather than the default filegroup. If the filegroup comprises multiple files spread across various physical disk arrays, each with its own disk controller, then queries for data from the table will be spread across the disk arrays, thereby improving performance.

If the computer has multiple processors, Microsoft SQL Server 2000 can perform parallel scans of the data. Multiple parallel scans can be executed for a single table if the filegroup of the table contains multiple files. Whenever a table is accessed sequentially, a separate thread is created to read each file in parallel. For example, a full scan of a table created on a filegroup comprised of four files will use four separate threads to read the data in parallel. Therefore, creating
more files per filegroup can help increase performance because a separate thread is used to scan each file in parallel. Similarly, when a query joins tables on different filegroups, each table can be read in parallel, thereby improving query performance.

Additionally, any text, ntext, or image columns within a table can be created on a filegroup other than the one that contains the base table.

Eventually, there is a saturation point when there are too many files and therefore too many parallel threads causing bottlenecks in the disk I/O subsystem. These bottlenecks can be identified by using Windows NT Performance Monitor to monitor the Physical Disk object and Disk Queue Length counter. If the Disk Queue Length counter is greater than three, consider reducing the number of files.

It is advantageous to get as much data spread across as many physical disk arrays as possible in order to improve throughput through parallel data access using multiple files. To spread data evenly across all disk arrays, first set up hardware-based disk striping array (RAID-5 or RAID-10), and then use filegroups to spread data across multiple hardware stripe set arrays if needed. For more information of filegroups please refer to SQL 2000 books online.

**Optimizing transaction log performance for SQL 2000**

General recommendations for creating transaction log files include:

- Create the transaction log on a physically separate disk or RAID device (RAID-1). The transaction log file is written serially; therefore, using a separate, dedicated disk allows the disk heads to stay in place for the next write operation. (Do not put more the one transaction log on one array, because actually you will get a random I/O rather then sequential I/O.)

- Set the original size of the transaction log file to a reasonable size to prevent the file from automatically expanding as more transaction log space is needed. As the transaction log expands, a new virtual log file is created, and write operations to the transaction log wait while the transaction log is expanded. If the transaction log expands too frequently, performance can be affected.

- Set the file growth increment percentage to a reasonable size to prevent the file from growing by too small a value. If the file growth is too small compared to the number of log records being written to the transaction log, then the transaction log may need to expand constantly, affecting performance.

- Manually shrink the transaction log files rather than allowing Microsoft SQL Server 2000 to shrink the files automatically. Shrinking the transaction log can affect performance on a busy system due to the movement and locking of data pages.

### 5.5 Case studies

In this section, we will present some case studies related to FAStT solutions that we have already offered to our customers. Please note that for every situation there is usually more than one solution. The studies that we have presented are an indication of what is possible using the FAStT product set.
Table 10 summarizes the cases that we present:

<table>
<thead>
<tr>
<th>The solution</th>
<th>The product</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extending your file server storage - direct attached</td>
<td>FASiT200</td>
<td>Section 5.5.1, on page 175</td>
</tr>
<tr>
<td>Extending your file server storage - direct attached</td>
<td>FASiT500</td>
<td>Section 5.5.2, on page 177</td>
</tr>
<tr>
<td>Two servers attached directly to the same storage</td>
<td>FASiT200</td>
<td>Section 5.5.3, on page 178</td>
</tr>
<tr>
<td>Low cost Microsoft Cluster Server</td>
<td>FASiT200</td>
<td>Section 5.5.4, on page 180</td>
</tr>
<tr>
<td>No single point of failure MSCS</td>
<td>FASiT200</td>
<td>Section 5.5.5, on page 182</td>
</tr>
<tr>
<td>No single point of failure MSCS</td>
<td>FASiT500</td>
<td>Section 5.5.6, on page 183</td>
</tr>
<tr>
<td>Two or more MSCS using the same shared storage</td>
<td>FASiT200</td>
<td>Section 5.5.8, on page 186</td>
</tr>
<tr>
<td>Two MSCS with NOSPOF direct attached</td>
<td>FASiT500</td>
<td>Section 5.5.9, on page 188</td>
</tr>
<tr>
<td>Multi MSCS with NOSPOF</td>
<td>FASiT500</td>
<td>Section 5.5.9, on page 188</td>
</tr>
<tr>
<td>Small SAN solution</td>
<td>FASiT200</td>
<td>Section 5.5.10, on page 190</td>
</tr>
<tr>
<td>Large SAN solution</td>
<td>FASiT500</td>
<td>Section 5.5.11, on page 191</td>
</tr>
</tbody>
</table>

5.5.1 Extending your file server storage — direct attached (FASiT 200)

For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

Business situation
The customer had just purchased a new xSeries 340 server that he wanted to use as a file server. The maximum internal storage available on the xSeries 340 is 216GB.

The customer was asking for a solution that would allow him to migrate all his data from his other file servers to the new xSeries 340. The amount of capacity that he needs to migrate is 500GB. The customer also stated that he wanted to have up to 1TB of data capacity.

The solution
We choose the FASiT 200 Storage Server as an external direct attached storage solution. The xSeries 340 server will be directly attached using the FASiT Host Bus Adapter to the FASiT 200 Storage Server.

To have 1TB capacity, we used 30 drives of 36.4GB with two additional EXP500 (Figure 130). Up to five additional EXP500 with a total of 60 drives can be added...
to the FASTT200. Refer to 2.1, “The FASTT200 Storage Server and FASTT200 HA Storage Server” on page 15.

**The benefits**
- FASTT200 Storage server with one RAID controller is a relatively low cost box.
- The customer will have the option to grow his storage as the business grows, because each EXP500, which attaches to the storage server, can host ten 72GB drives which yield 720GB on a single box and can grow beyond this by using more drives housed in additional EXP500 attached to the FASTT 200.
- The FASTT 200 can be upgraded to a dual controller unit.

**Solution diagram**
Figure 130 presents a diagrammatic solution.

![Solution Diagram](image)

**Figure 130. FASTT200 direct attached**

**Solution components**
The solution components are shown in Table 11.

**Table 11. FASTT200 direct attached**

<table>
<thead>
<tr>
<th>Component name</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASTT200 Storage Server</td>
<td>1</td>
</tr>
<tr>
<td>Short Wave GBIC</td>
<td>4</td>
</tr>
<tr>
<td>Fibre Channel cable</td>
<td>22</td>
</tr>
<tr>
<td>FASTT EXP500</td>
<td>5</td>
</tr>
</tbody>
</table>
5.5.2 Extending your file server storage — direct attached (FAStT 500)

For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

**Business situation**
The customer had just purchased a new xSeries 350 server that he wanted to use as a file server. The maximum internal storage available on the xSeries 350 is 216GB.

The customer is asking for a solution that will let him migrate all his data from his other file servers to the new xSeries 350, the amount of capacity that he needs to migrate is 1TB. Your customer is also saying that he wants to have up to 3TB of data capacity and that he wants redundant connectivity from the server to the storage.

**The solution**
We chose the FAStT 500 storage server as an external direct attached storage solution. The xSeries 350 server will be direct attached with two FAStT Host Bus Adapters to the FAStT 500 Storage server. Each FAStT Host Bus Adapter will be routed to a different RAID controller on a different path by the mini Hub on the FAStT 500.

To have 3TB capacity, we used fifty 72GB drives with five additional EXP500 (Figure 131). Up to 22 EXP500 with a total of 220 drives can be added to the FAStT 500. See also 2.2, “The FAStT500 Storage Server” on page 19.

To get more performance from the drives, we also used two dual loops to spread the drives between the loops.

**Solution benefits**
- No external hub or switch is needed to connect the server with dual path to the FAStT 500, we can use the two default mini hubs on the host side.
- Since we are using several logical drives we can spread them across both controllers.
- Using the two dual loops on the drive side of the FAStT500 can increase I/O performance.

**Solution diagram**
Figure 131 presents a diagrammatic solution.
Figure 131. FASTT500 direct attached

**Solution components**
The solution components are listed in Table 12.

<table>
<thead>
<tr>
<th>Component name</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASTT500</td>
<td>1</td>
</tr>
<tr>
<td>Mini Hub</td>
<td>2</td>
</tr>
<tr>
<td>Fibre Channel cable</td>
<td>12</td>
</tr>
<tr>
<td>Fibre Channel GBIC</td>
<td>22</td>
</tr>
<tr>
<td>FASTT host bus adapter</td>
<td>2</td>
</tr>
</tbody>
</table>

5.5.3 Two servers attached directly to one FASTT 200
For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

**Business situation**
The customer runs a digital image processing business. For his business he uses one Windows 2000 server as a file server for internal users to process images.
and one Linux server which is used as FTP server for external users to put their images. The customer is looking for a storage solution for both servers.

His minimum requirements are:

1. Each server will have the option to store up to 500GB.
2. The servers should be 1U Rack mount size servers (no internal storage).
3. His service level agreement does not require high availability.
4. He wants to manage a single storage solution.
5. For future expansion he wants to be able to attach additional servers to the storage subsystem.
6. The customer has a limited budget.

**The solution**

Based on his needs and requirements, we offered the customer one FASiT 200 HA Storage Server, where each server will be attached directly from the FASiT Host Bus Adapter to one of the FASiT 200 controllers (Figure 132).

To ensure a proper setup of the FASiT 200 we used storage partitioning.

**The benefits**

- Low cost storage solution for two servers that can fit the limited customer budget.
- No hub or switch is required due to the ability of the FASiT200 to accept one host connection per controller.
- This solution can be upgraded in the future to accommodate more than two servers by adding a hub or switch between the hosts and FASiT 200.
- The FASiT 200 can scale with EXP500 storage expansions to support at least sixty drives of 73GB. (Up to 100 disk drives or 10 more EXP500 are supported from a technology standpoint, but you should consider subsystem performance.)

**Solution diagram**

Figure 132 presents a diagrammatic solution.
5.5.4 Low cost Microsoft Cluster Server

For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

Business situation
The customer wants a Microsoft Cluster Server (MSCS) for his application. Since most of his concern is with the application that can cause Windows 2000 to stop functioning, he just wants MSCS for the operating services to provide availability.

The data capacity is 500GB.

The solution
Since the customer only wants to address future problems with the stability of his application on Windows 2000, we can offer him a low cost MSCS on FASTT 200.
For this solution we cluster to nodes with a single path from each node to a single hub, using FASTT200 HA Storage Server with two controllers.

**The benefits**
- Relatively low-cost solution for a Microsoft Cluster Server, addressing the customer needs.
- Can be upgraded to a no single point of failure solution while disks are not needed to be configured again.
- Can scale with more EXP500 and disks.

**Note**

We must use at least one hub or switch in a Microsoft Cluster configuration due to the fact that if you will connect the servers directly to the FASTT200 you actually connect each server to a different controller on the FAStT 200, and in a case of a failover even though each controller is aware of the logical drives that are defined on the other controller, he can’t reach these logical drives; therefore, after failover you can end up with live online server that can’t reach disks that were assigned to the failed node. But if you use a hub or switch as in our solution, each server can communicate with each controller through the hub.

**Solution diagram**

Figure 133 presents a diagrammatic solution.

![Solution Diagram](image-url)
Solution components
The solution components are listed in Table 14.

Table 14. FASiT200 low cost MSCS

<table>
<thead>
<tr>
<th>Component name</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASiT 200 HA Storage Server</td>
<td>1</td>
</tr>
<tr>
<td>FASiT Host adapter</td>
<td>2</td>
</tr>
<tr>
<td>Fibre Channel cable</td>
<td>8</td>
</tr>
<tr>
<td>FC Short Wave GBIC</td>
<td>10</td>
</tr>
<tr>
<td>FC Managed Hub</td>
<td>1</td>
</tr>
<tr>
<td>EXP500</td>
<td>2</td>
</tr>
</tbody>
</table>

5.5.5 Highly available Microsoft Cluster Server on FASiT 200

For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

Business situation
The customer wants to deploy his new SQL2000 server. His main concern is high availability for his ASP business. He needs to deploy two different databases for his application solution. His current data capacity is 100GB and he wants to be able to scale in the next year to 300GB.

The solution
The solution we offer is based on MSCS with no single point of failure, where each server will be deployed with two FASiT host adapters. Each adapter will reach the FASiT200 HA Storage server through a different FC Managed Hub, to provide dual path from each server to the disk (Figure 134).

We also deploy an active/active MSCS solution, where each node hosts a different instance of SQL server each with its own database.

The benefits
- Dual paths from each server eliminate the single point of failure from the I/O paths. We use redundant adapters, cables and hubs which provide better total availability to the solution.
- Using an active/active solution we can utilize all the resources of the cluster solution where each node is running an SQL server independently.
5.5.6 No single point of failure MSCS direct attached — FASTT500

For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

**Business situation**
The customer is looking to implement an Exchange 2000 solution which will be highly available and highly scalable from a storage standpoint. The storage capacity the customer has is around 1TB now and the forecast for the next two years is that this will grow to around 4TB.
The solution
The customer will implement Microsoft Cluster Server with FAST500, where each server will have NSPOF connection to drives. He will use three EXP500 each with ten 36GB drives, yielding to a total of thirty drives (Figure 135).

The benefits
- FAST500 is a highly-scalable, highly-available storage solution, where you can scale up to 220 disk drives (from a performance standpoint you may not benefit from too many drives/storage on one box — depending on application and I/O operations.
- No need to buy extra hubs or switches since the FAST500 has as standard two mini hubs each with two links on the host side (and two more can be added on host side). See 2.2, “The FAST500 Storage Server” on page 19.
- When the need for extra storage will arise, the customer can split the extra storage between two dual loops for better performance.
- Even when using one dual loop, see Figure 135, you can benefit from the dual loop redundant capability, where each drive on the storage can talk to each loop in case that one of the loops fails.

Solution diagram
Figure 135 presents a diagrammatic solution.
**Solution components**
The solution components are listed in Table 16.

Table 16. Solution components no SPOF MSCS

<table>
<thead>
<tr>
<th>Component name</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASTT500</td>
<td>1</td>
</tr>
<tr>
<td>FASTT Host Adapter</td>
<td>4</td>
</tr>
<tr>
<td>FC Short wave GBIC</td>
<td>16</td>
</tr>
<tr>
<td>FC Cable</td>
<td>10</td>
</tr>
<tr>
<td>EXP500</td>
<td>3</td>
</tr>
</tbody>
</table>

5.5.7 **Two or more Microsoft Clusters using the same shared storage**
For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

**Business situation**
The customer has three file servers each serving a different department, where high availability for these services is a basic requirement. Each server has a storage requirement of 150GB currently and the customer wants to be able to grow to 400GB per server. The server and storage utilization are not so high.

**The solution**
The solution that will fit the customer needs is based on a clustering with no single point of failure. As each of the servers will be clustered, we therefore will get three independent clusters, where each of the clusters is connected to the same SAN infrastructure through two switches to one FASTT 200 HA storage server (Figure 136). Please note that we chose to use one FASTT 200 for the three clusters, because we know that the storage utilization is low. Otherwise, we would have recommended the use of a FASTT500 instead.

**Solution benefits**
- The main benefit for this solution is using one of the SAN infrastructures for three independent Fibre Channel attached clusters.
- Each of the clusters is attached with dual path to the same pair of switches to provide a redundant path for each of the switches.

**Solution diagram**
Figure 136 presents a diagrammatic solution.
**Solution components**

The solution components are listed in Table 16.

*Table 17. FASiT Hosting Multi Microsoft Clusters*

<table>
<thead>
<tr>
<th>Component name</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASiT200 HA storage server</td>
<td>3</td>
</tr>
<tr>
<td>Fibre Channel 8 port switches</td>
<td>2</td>
</tr>
<tr>
<td>FASiT Host Adapter</td>
<td>6</td>
</tr>
<tr>
<td>Fiber optic cable</td>
<td>18</td>
</tr>
<tr>
<td>Fibre Channel GBIC</td>
<td>16</td>
</tr>
<tr>
<td>EXP500</td>
<td>2</td>
</tr>
</tbody>
</table>

**5.5.8 Two MSCS with NSPOF direct attached**

For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.
Business situation
The customer wants to cluster two of his main infrastructure servers, the database and mail servers. The customer also wants a solution that will fit his budget and that will still provide him with good scalability options.

The solution
The solution we offered in this case is that the two clusters will share one FASTT500. With two extra mini hubs we can stay within the budget limit of the customer and still be able to connect the clusters with a dual path from each server to the FASTT500. The FASTT500 will be connected to three EXP500, one for each server cluster (Figure 137).

The benefits
• FASTT500 is a highly scalable unit that can hold up to 220 disk drives.
• To achieve a dual path from each server to the disks, we only need to purchase two additional mini hubs.
• Up to 16 clusters can be hosted on one FASTT500 with code 7.1. (You need to consider performance issues.) This is done by partitioning the FASTT for 16 partitions, one per cluster, because a cluster of two servers can use one partition.

Solution diagram
Figure 137 presents a diagrammatic solution.
5.5.9 More than two MSCS with NSPOF

For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

Business situation
The customer needs to cluster three servers and to stay within the budget limit. His demand is to have a Fibre Channel solution for the shared storage of the cluster.
**The solution**
To cluster his three servers and stay within his budget, we offered a FASiT500 for shared storage. We decided to connect each of the cluster servers with dual paths so we had to use external switches due to the amount of links that were needed (Figure 138).

**The benefits**
- FASiT500 can host up to 16 clusters (you need to consider performance issues).
- Since more I/O and more drives were involved in this configuration, the FASiT500 can split the disk drive connections to two different dual loops.

**Solution diagram**
Figure 138 presents a diagrammatically solution.

![Solution diagram](image)

**Figure 138. Multi clusters on FASiT500**

**Solution components**
The components are listed in Table 19.

<table>
<thead>
<tr>
<th>Component name on FASiT500</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASiT500</td>
<td>1</td>
</tr>
<tr>
<td>FASIT Host Adapter</td>
<td>6</td>
</tr>
<tr>
<td>FC 8 port switch</td>
<td>2</td>
</tr>
<tr>
<td>Fiber optic cable</td>
<td>22</td>
</tr>
</tbody>
</table>
5.5.10 Small SAN solution

For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

Business situation
The customer needs to consolidate his four servers onto one storage solution. The customer has one Windows 2000 server, one Linux server and two SUN Solaris servers. Since none of these servers have a large storage capacity, the customer is looking for a small SAN solution to fit his needs.

The solution
The solution we are offered is based on FASiT 200 with the new code 7.1, which gives us the heterogeneous host support. Each server will be connected through a Fibre Channel switch to the FASiT 200 HA Storage Server (Figure 139).

The benefits
- Even when using only one path from server to disks, there is a benefit to use the FASiT200 HA Storage Server versions (two controllers) due to the ability to split logical drives between controllers in order to obtain optimal performance.
- FASiT200 is a low-cost Fibre Channel solution that supports heterogeneous systems.
- You can build one large array (for example, ten drives of 18 GB with RAID-5) for all the servers that can be a good solution to form an economical standpoint, because you lose only one disk for fault tolerance, and then build logical drives and map them to each server.

Solution diagram
Figure 139 presents a diagrammatic solution.
Chapter 5. FASTT solution design

5.5.11 Large SAN solution

For this case study, we present the business situation, the solution, its benefits, a solution diagram, and the solution components.

**Business situation**
The customer infrastructure consists of three stand alone servers and two clusters with two nodes each. Each of the servers has its own storage device which is managed locally at the server. The customer is considering connecting all these servers to a SAN infrastructure. The customer demands high availability from this infrastructure. The customer wants a solution that can grow with his needs which are:

- **Component name**
  - FASTT200 HA storage server
  - FC Switch
  - Fiber optic cable
  - Fibre Channel GBIC
  - FASTT Host Adapter

<table>
<thead>
<tr>
<th>Component name</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASTT200 HA storage server</td>
<td>1</td>
</tr>
<tr>
<td>FC Switch</td>
<td>1</td>
</tr>
<tr>
<td>Fiber optic cable</td>
<td>10</td>
</tr>
<tr>
<td>Fibre Channel GBIC</td>
<td>12</td>
</tr>
<tr>
<td>FASTT Host Adapter</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 139. Small FASTT200 SAN

**Solution components**
The solution components are listed in Table 20.

Table 20. Small FASTT200 SAN

Component name | Quantity
---------------|----------
FASTT200 HA storage server | 1
FC Switch | 1
Fiber optic cable | 10
Fibre Channel GBIC | 12
FASTT Host Adapter | 4
1. Support for more servers on the storage unit
2. Highly scalable for I/O performance
3. Support for at least up to 6TB in the future

The solution
The solution we implemented for this customer was based on the FAStT500. Each server was connected with two paths to the FAStT500 via two Fibre Channel switches. On the FAStT500, we split the disk drives between the two dual loops by adding two more mini hubs (Figure 140).

The benefits
• Reducing storage management cost by using one tool (FAStT storage manager) to manage all the storage on the infrastructure.
• By implementing less arrays we can reduce the disk drive overhead for redundancy (each separate RAID-5 array has a penalty of one disk drive for parity), this can be achieved by building more than one logical unit per array, and partition each logical drive to a different server.
• By using dual paths from servers to disks we can achieve no single point of failure in the way from server to disks.
• By utilizing the two dual loops on the FAStT500, we can increase the performance of the system, which would occur on a single dual, due to loop saturation when increasing the number of devices attached to the loop.
• On the current configuration the partition consumption is five partitions, this is because each cluster consumes only one partition as both nodes must reside within the same host group. The maximum number of partitions that is supported with the 7.1 code is 16.
• With a unified SAN solution you can deploy some kind of unified backup over the SAN which is LAN Free or Server Less backup. (This topic is out of the scope of this book due to the various ways that this is implemented by the backup vendors.)
• The FAStT500 can support up to 220 drives of 73 GB each providing more than 15 TB of capacity.

Solution diagram
Figure 140 presents a diagrammatic solution.
Solution components

The solution components are listed in Table 21.

Table 21. Large FASiT500 SAN solution

<table>
<thead>
<tr>
<th>Component name</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASiT500</td>
<td>1</td>
</tr>
<tr>
<td>FC Switch</td>
<td>2</td>
</tr>
<tr>
<td>Fiber optic cable</td>
<td>26</td>
</tr>
<tr>
<td>Fibre Channel GBIC</td>
<td>30</td>
</tr>
<tr>
<td>FASiT Host Adapter</td>
<td>14</td>
</tr>
<tr>
<td>EXP500</td>
<td>5</td>
</tr>
</tbody>
</table>
Appendix A. Critical events

This list represents the critical errors that may occur on the FASTT Storage Server it will appear in the major event log.

<table>
<thead>
<tr>
<th>Critical Event Type and Sense Key/ASC/ASCQ</th>
<th>Description/Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event 1001 - Channel failed Sense key/ASC/ASCQ: 6/3F/C3</td>
<td><strong>Description:</strong> The controller failed a channel, and will not access drives on this channel any more. The Field Replaceable Unit (FRU) Group Qualifier (byte 26) in the sense data will indicate the 1-relative channel number of the failed channel. This condition is typically caused by a drive ignoring SCSI protocol on one of the controller's destination channels. The controller typically fails a channel if it issued a reset on a channel, and it continued to see drives ignore the SCSI Bus Reset on this channel. <strong>Action:</strong> Select the Recovery Guru to obtain the &quot;Failed Drive SCSI Channel&quot; recovery procedure. Contact your customer support representative to complete this procedure.</td>
</tr>
<tr>
<td>Event 1010 - Impending drive failure (PFA) detected Sense key/ASC/ASCQ: 6/5D/80</td>
<td><strong>Description:</strong> A drive has reported that a failure prediction threshold has been exceeded. This indicates that the drive may fail within 24 hours. <strong>Action:</strong> Select the Recovery Guru to obtain the &quot;Impending Drive Failure&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 1015 - Incorrect mode parameters set on drive Sense key/ASC/ASCQ: 6/3F/BD</td>
<td><strong>Description:</strong> The controller was unable to query the drive for its current critical mode page settings, or was unable to change these to the correct setting. Currently, this indicates the Qerr bit is set incorrectly on the drive specified in the FRU field of the Request Sense data. <strong>Action:</strong> The controller has not failed yet. Contact your customer support representative for instructions on recovering from this failure.</td>
</tr>
<tr>
<td>Event 1207 - Fibre Channel link errors - threshold exceeded Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> Invalid characters have been detected in the Fibre Channel signal. Possible causes for the error are a degraded laser in a gigabit interface converter (GBIC) or media interface adapter (MIA), damaged or faulty Fibre Channel cables, or poor cable connections between components on the loop. <strong>Action:</strong> In the main Array Management Window, select Help &gt;&gt; Recovery Procedures. Select &quot;Fibre Channel Link Errors - Threshold Exceeded&quot; for more information about recovering from this failure.</td>
</tr>
<tr>
<td>Event 150E - Controller loopback diagnostics failed Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> The controller cannot initialize the drive-side Fibre Channel loops. A diagnostic has been run identifying a controller problem and the controller has been placed offline. This event will only occur on certain controller models. <strong>Action:</strong> Select the Recovery Guru to obtain the &quot;Offline Controller&quot; recovery procedure. Use this procedure to replace the controller.</td>
</tr>
<tr>
<td>Event 202E - Read drive error during interrupted write Sense key/ASC/ASCQ: 3/11/8A</td>
<td><strong>Description:</strong> A media error has occurred on a read operation during interrupted write processing. <strong>Action:</strong> Select the Recovery Guru to obtain the &quot;Unrecovered Interrupted Write&quot; recovery procedure. Contact your customer support representative to complete this procedure.</td>
</tr>
<tr>
<td>Event 2109 - Controller cache not enabled - cache sizes do not match Sense key/ASC/ASCQ: 6/A1/00</td>
<td><strong>Description:</strong> The controller will not allow mirroring to be enabled if the alternate controller's cache size is different. Ensure that both controllers have the same cache size. <strong>Action:</strong> Contact your customer support representative for instructions on recovering from this failure.</td>
</tr>
<tr>
<td>Critical Event Type and Sense Key/ASC/ASCQ</td>
<td>Description/Action</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-------------------</td>
</tr>
</tbody>
</table>
| Event 210C - Controller cache battery failed  
Sense key/ASC/ASCQ: 6/0C/80 | **Description:** The controller has detected that the battery is (1) not physically present, (2) it is fully discharged, or (3) it has reached its expiration date.  
**Action:** Select the Recovery Guru to obtain the "Failed Battery Canister" recovery procedure. |
| Event 210E - Controller cache memory recovery failed after power cycle or reset  
Sense key/ASC/ASCQ: 6/0C/81 | **Description:** Recovery from a Data Cache error was unsuccessful. User data may have been lost.  
**Action:** Contact your customer support representative for instructions on recovering from this failure. |
| Event 2110 - Controller cache memory initialization failed  
Sense key/ASC/ASCQ: 6/40/81 | **Description:** The controller has detected the failure of an internal controller component (RAID Buffer). This failure may have been detected during operation as well as during an on-board diagnostic routine.  
**Action:** Contact your customer support representative for instructions on recovering from this failure. |
| Event 2113 - Controller cache battery nearing expiration  
Sense key/ASC/ASCQ: 6/3F/D9 | **Description:** The cache battery is within the specified number of weeks of failing.  
**Action:** Select the Recovery Guru to obtain the "Battery Nearing Expiration" recovery procedure. |
| Event 2229 - Drive failed by controller  
Sense key/ASC/ASCQ: None | **Description:** The controller has failed a drive because of a problem with the drive.  
**Action:** Select the Recovery Guru to obtain the recovery procedure and follow the procedures to correct the failure. |
| Event 222D - Drive manually failed  
Sense key/ASC/ASCQ: 6/3F/87 | **Description:** The drive was manually failed by a user.  
**Action:** Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure. |
| Event 2247 - Data lost on logical drive during unrecovered interrupted write  
Sense key/ASC/ASCQ: 6/3F/EB | **Description:** An error has occurred during interrupted write processing during the start-of-day routine causing the logical drive to transition to the failed state.  
**Action:** Select the Recovery Guru to obtain the "Unrecovered Interrupted Write" recovery procedure. Contact your customer support representative to complete this procedure. |
| Event 2248 - Drive failed - write failure  
Sense key/ASC/ASCQ: 6/3F/80 | **Description:** The drive failed a write command to it. The drive will be marked failed.  
**Action:** Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure. |
| Event 2249 - Drive capacity less than minimum  
Sense key/ASC/ASCQ: 6/3F/8B | **Description:** During drive replacement, the capacity of the drive is not large enough to support all the logical drives that must be reconstructed on it.  
**Action:** Replace the drive with a larger capacity drive. |
| Event 224A - Drive has wrong block size  
Sense key/ASC/ASCQ: 6/3F/8C | **Description:** The drive's block size does not match that of the other drives in the logical drive. The drive will be marked failed.  
**Action:** Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure. |
| Event 224B - Drive failed - initialization failure  
Sense key/ASC/ASCQ: 6/3F/86 | **Description:** The drive failed either a Format Unit command, or a Write operation (issued when a logical drive was initialized). The drive will be marked failed.  
**Action:** Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure. |
<table>
<thead>
<tr>
<th>Critical Event Type and Sense Key/ASC/ASCQ</th>
<th>Description/Action</th>
</tr>
</thead>
</table>
| Event 224D - Drive failed - no response at start of day Sense key/ASC/ASCQ: 6/3F/85 | **Description:** The drive failed a Read Capacity or Read command during the start-of-day routine. The controller was unable to read the configuration information stored on it. The drive will be marked failed.  
**Action:** Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure. |
| Event 224E - Drive failed - initialization/reconstruction failure Sense key/ASC/ASCQ: 6/3F/82 | **Description:** The (previously-failed) drive was marked failed because either (1) the drive failed a Format Unit command issued to it, or (2) the reconstruction on the drive failed due to the controller being unable to restore it (for example, an error occurring on another drive required for reconstruction).  
**Action:** Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure. |
| Event 2250 - Logical drive failure (3F E0) Sense key/ASC/ASCQ: 6/3F/E0 | **Description:** The controller has marked the logical drive failed. User data and/or redundancy (parity) can no longer be maintained to ensure availability. The most likely cause is the failure of a single drive in non-redundant configurations or a second drive in a configuration protected by one drive.  
**Action:** Select the Recovery Guru to obtain the “Failed Logical Drive - Drive Failure” recovery procedure. |
| Event 2251 - Drive failed - reconstruction failure Sense key/ASC/ASCQ: 6/3F/8E | **Description:** A drive failed due to a reconstruction failure during the start-of-day routine.  
**Action:** Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure. |
| Event 2252 - Drive marked offline during interrupted write Sense key/ASC/ASCQ: 6/3F/98 | **Description:** An error has occurred during interrupted write processing causing the logical drive to be marked failed. Drives in the array that did not experience the read error will transition to the Offline state and log this error.  
**Action:** Select the Recovery Guru to obtain the “Unrecovered Interrupted Write” recovery procedure. Contact your customer support representative to complete this procedure. |
| Event 2254 - Redundancy (parity) and data mismatch was detected Sense key/ASC/ASCQ: 6/8E/01 | **Description:** The controller detected inconsistent redundancy (parity)/data during a parity verification.  
**Action:** Contact your customer support representative for instructions on recovering from this failure. |
| Event 2255 - Logical drive definition incompatible with ALT mode - ALT disabled Sense key/ASC/ASCQ: 6/91/3B | **Description:** Auto-LUN Transfer (ALT) only works with arrays that have only one logical drive defined. There are currently arrays on the storage subsystem that have more than one logical drive defined; therefore, ALT mode has been disabled. The controller will operate in normal redundant controller mode, and if there is a problem, will transfer all logical drives on an array instead of transferring individual logical drives.  
**Action:** Contact your customer support representative for instructions on recovering from this failure. |
| Event 2602 - Automatic controller firmware synchronization failed Sense key/ASC/ASCQ: 02/04/81 | **Description:** The versions of firmware on the redundant controllers are not the same because the automatic controller firmware synchronization failed. Controllers with an incompatible version of firmware may cause unexpected results.  
**Action:** Try the firmware download again. If the problem persists, contact your Customer Support Representative. |
<table>
<thead>
<tr>
<th>Critical Event Type and Sense Key/ASC/ASCQ</th>
<th>Description/Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event 2801 - Storage array running on UPS battery</td>
<td>Description: The UPS has indicated that AC power is no longer present and the UPS has switched to standby power. While there is no immediate cause for concern, you should save your data frequently, in case the battery is suddenly depleted. <strong>Action:</strong> Select the Recovery Guru to obtain the &quot;Lost AC Power&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 2803 - UPS battery - two minutes to failure</td>
<td>Description: The Uninterruptible Power Source (UPS) has indicated that its standby power source is nearing depletion. <strong>Action:</strong> You should take actions to stop I/O activity to the controller. Normally, the controller will change from a write-back caching mode to a write-through mode.</td>
</tr>
<tr>
<td>Event 2804 - UPS battery failed</td>
<td>Description: The UPS battery has failed. <strong>Action:</strong> Contact your customer support representative for instructions on recovering from this failure.</td>
</tr>
<tr>
<td>Event 2807 - Environmental card failed</td>
<td>Description: An environmental card has failed. <strong>Action:</strong> Select the Recovery Guru to obtain the &quot;Failed Environmental Card Canister&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 2808 - Enclosure ID not unique</td>
<td>Description: The controller has determined that there are multiple drive trays with the same ID selected. Make sure that each drive tray has a unique ID setting. <strong>Action:</strong> Select the Recovery Guru to obtain the &quot;Enclosure ID Conflict&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 280A - Controller enclosure component missing</td>
<td>Description: A component other than a controller is missing in the controller tray (for example, fan, power supply, or battery). The FRU codes indicate the faulty component. <strong>Action:</strong> Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure.</td>
</tr>
<tr>
<td>Event 280B - Controller enclosure component failed</td>
<td>Description: A component other than a controller has failed in the controller tray (for example, fan, power supply, battery, and so on), or an over-temperature condition has occurred. The FRU codes indicate the faulty component. <strong>Action:</strong> Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure.</td>
</tr>
<tr>
<td>Event 280D - Drive enclosure component failed</td>
<td>Description: A component other than a drive has failed in the drive tray (for example, fan, power supply, battery, and so on), or an over-temperature condition has occurred. The FRU codes indicate the faulty component. <strong>Action:</strong> Select the Recovery Guru to obtain the recovery procedure and follow the instructions to correct the failure.</td>
</tr>
<tr>
<td>Event 280E - Standby power source not fully charged</td>
<td>Description: The UPS has indicated that its standby power source is not at full capacity.</td>
</tr>
<tr>
<td>Event 280F - Environmental card - loss of communication</td>
<td>Description: Communication has been lost to one of the dual environmental card canisters in a drive tray. The drive tray has only one I/O path available. <strong>Action:</strong> Select the Recovery Guru to obtain the &quot;Environmental Card Canister - Loss of Communication&quot; recovery procedure.</td>
</tr>
<tr>
<td>Critical Event Type and Sense Key/ASC/ASCQ</td>
<td>Description/Action</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Event 2813 - Mini-hub canister failed Sense key/ASC/ASCQ: 6/3F/C7</td>
<td><strong>Description</strong>: Communications with the mini-hub canister has been lost. This may be the result of a mini-hub canister failure, a controller failure, or a failure in an internal backplane communications board. If there is only one mini-hub failure, the storage subsystem is still operational but a second mini-hub failure could result in the failure of the affected subsystem. <strong>Action</strong>: Select the Recovery Guru to obtain the &quot;Failed Mini-Hub Canister&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 2815 - GBIC failed Sense key/ASC/ASCQ: None</td>
<td><strong>Description</strong>: A GBIC on either the controller tray or the drive tray has failed. If there is only one GBIC failure, the storage subsystem is still operational, but a second GBIC failure could result in the failure of the affected enclosure. <strong>Action</strong>: Select the Recovery Guru to obtain the &quot;Failed GBIC&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 2816 - Enclosure ID conflict - duplicate ID’s across drive enclosures Sense key/ASC/ASCQ: 6/98/01</td>
<td><strong>Description</strong>: Two or more drive trays are using the same enclosure identification number. <strong>Action</strong>: Select the Recovery Guru to obtain the &quot;Enclosure ID Conflict&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 2818 - Enclosure ID mismatch - duplicate ID’s in the same drive enclosure Sense key/ASC/ASCQ: 6/98/02</td>
<td><strong>Description</strong>: A drive tray in the storage subsystem contains environmental cards with different enclosure identification numbers. <strong>Action</strong>: Select the Recovery Guru to obtain the &quot;Enclosure ID Mismatch&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 281B - Nominal temperature exceeded Sense key/ASC/ASCQ: 6/98/03</td>
<td><strong>Description</strong>: The nominal temperature of the enclosure has been exceeded. Either a fan has failed or the temperature of the room is too high. If the temperature of the enclosure continues to rise, the affected enclosure may automatically shut down. Fix the problem immediately, before it becomes more serious. The automatic shutdown conditions depend on the model of the enclosure. <strong>Action</strong>: Select the Recovery Guru to obtain the &quot;Nominal Temperature Exceeded&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 281C - Maximum temperature exceeded Sense key/ASC/ASCQ: 6/3F/C6</td>
<td><strong>Description</strong>: The maximum temperature of the enclosure has been exceeded. Either a fan has failed or the temperature of the room is too high. This condition is critical and may cause the enclosure to shut down if you do not fix the problem immediately. The automatic shutdown conditions depend on the model of the enclosure. <strong>Action</strong>: Select the Recovery Guru to obtain the &quot;Maximum Temperature Exceeded&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 281D - Temperature sensor removed Sense key/ASC/ASCQ: 6/98/03</td>
<td><strong>Description</strong>: A canister containing a temperature sensor has been removed from the storage subsystem. <strong>Action</strong>: Replace the canister as soon as possible. Select the Recovery Guru to obtain the &quot;Failed or Removed Fan Canister&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 281E - Environmental card firmware mismatch Sense key/ASC/ASCQ: 6/98/03</td>
<td><strong>Description</strong>: A drive tray in the storage subsystem contains Environmental Cards with different versions of firmware. Environmental Cards in the same drive tray must have the same version firmware. If you do not have a replacement card, call your customer support representative to perform the firmware download. <strong>Action</strong>: Select the Recovery Guru to obtain the &quot;Environmental Card Firmware Version Mismatch&quot; recovery procedure.</td>
</tr>
<tr>
<td>Event 3019 - Logical drive ownership changed due to failover Sense key/ASC/ASCQ: None</td>
<td><strong>Description</strong>: The multi-path driver software has changed ownership of the logical drives to the other controller, because it could not access the logical drives on that particular path. <strong>Action</strong>: Select the Recovery Guru to obtain the &quot;Logical Drives Not on Preferred Path&quot; recovery procedure.</td>
</tr>
<tr>
<td>Critical Event Type and Sense Key/ASC/ASCQ</td>
<td>Description/Action</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Event 5005 - Place controller offline Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> The controller was placed offline. This could be caused by the controller failing a diagnostic test. (The diagnostics are initiated internally by the controller or by the Controller &gt;&gt; Run Diagnostics menu option.) Or the controller was manually placed Offline using the Controller &gt;&gt; Place Offline menu option.  <strong>Action:</strong> Select the Recovery Guru to obtain the “Offline Controller” recovery procedure. Use this procedure to replace the controller.</td>
</tr>
<tr>
<td>Event 5602 - This controller's alternate failed - timeout waiting for results Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> This controller initiated diagnostics on the alternate controller, but did not receive a reply indicating that the diagnostics completed. The alternate controller in this pair has been placed offline.  <strong>Action:</strong> Select the Recovery Guru to obtain the “Offline Controller” recovery procedure. Use this procedure to replace the controller.</td>
</tr>
<tr>
<td>Event 560B - ctrDiag task cannot obtain Mode Select lock Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> This controller was attempting to run diagnostics and could not secure the test area from other storage subsystem operations. The diagnostics were canceled.  <strong>Action:</strong> Contact your customer support representative for instructions on recovering from this failure.</td>
</tr>
<tr>
<td>Event 560C - ctrDiag task on controller's alternate cannot obtain Mode Select lock Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> The alternate controller in this pair was attempting to run diagnostics and could not secure the test area from other storage subsystem operations. The diagnostics were canceled.  <strong>Action:</strong> Contact your customer support representative for instructions on recovering from this failure.</td>
</tr>
<tr>
<td>Event 560D - Diagnostics read test failed on controller Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> While running diagnostics, the controller has detected that the information received does not match the expected return for the test. This could indicate that I/O is not completing or that there is a mismatch in the data being read. The controller was placed offline as result of this failure.  <strong>Action:</strong> Select the Recovery Guru to obtain the “Offline Controller” recovery procedure. Use this procedure to replace the controller.</td>
</tr>
<tr>
<td>Event 560E - This controller's alternate failed diagnostics read test Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> While running diagnostics, this controller’s alternate detected that the information received does not match the expected return for the test. This could indicate that I/O is not completing or that there is a mismatch in the data being read. The alternate controller in this pair was placed offline.  <strong>Action:</strong> Select the Recovery Guru to obtain the “Offline Controller” recovery procedure. Use this procedure to replace the controller.</td>
</tr>
<tr>
<td>Event 560F - Diagnostics write test failed on controller Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> While running diagnostics, the controller was unable to write data to the test area. This could indicate that I/O is not completing or that there is a mismatch in the data being written. The controller was placed offline.  <strong>Action:</strong> Select the Recovery Guru to obtain the “Offline Controller” recovery procedure. Use this procedure to replace the controller.</td>
</tr>
<tr>
<td>Event 5610 - This controller's alternate failed diagnostics write test Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> While running diagnostics, this controller’s alternate was unable to write data to the test area. This could indicate that I/O is not completing or that there is a mismatch in the data being written. The alternate controller in this pair was placed offline.  <strong>Action:</strong> Select the Recovery Guru to obtain the “Offline Controller” recovery procedure. Use this procedure to replace the controller.</td>
</tr>
<tr>
<td>Event 5616 - Diagnostics rejected - configuration error on controller Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> This controller was attempting to run diagnostics and could not create the test area necessary for the completion of the tests. The diagnostics were canceled.  <strong>Action:</strong> Contact your customer support representative for instructions on recovering from this failure.</td>
</tr>
<tr>
<td>Critical Event Type and Sense Key/ASC/ASCQ</td>
<td>Description/Action</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Event 5617 - Diagnostics rejected - configuration error on controller's alternate Sense key/ASC/ASCQ: None</td>
<td><strong>Description:</strong> This controller's alternate was attempting to run diagnostics and could not create the test area necessary for the completion of the tests. The diagnostics were canceled. <strong>Action:</strong> Contact your customer support representative for instructions on recovering from this failure.</td>
</tr>
</tbody>
</table>
Appendix B. Installing Windows 2000 Cluster service

Here are the objectives for installing the Windows 2000 Cluster Service. The following step by step guide will let you:

- Install Cluster service on the first node of the cluster.
- Add the second node to the cluster.
- Verify that Cluster service is installed and running on both nodes.
- Identify changes made to each node of the cluster.

**Prerequisite setup**
To complete this setup, you need the following:

- Two computers running Microsoft Windows 2000 Advanced Server, each with at least one FASTT Host based Adapter installed. Each computer is connected to the FASTT shared disk with fiber optic cables.
- Two network adapter cards in each node of the cluster. One for the interconnect between the nodes (Private link-heartbeat) and the other for the LAN (Public link)
- You will start the lab with a common drive letter on the disk shared by the two computers in the cluster.
- The public and private networks have been configured on each node in the cluster.

The preparation of the FASTT Storage Server is explained in 4.3, “Installing a Microsoft Cluster Server with FASTT Storage Server” on page 97.

---

**Note**
The following procedure will refer to your computers as Node A and Node B. Node A will be the first computer where Cluster service is installed.

---

**B.1 Install Cluster service on Node A**

On this stage, you will install Windows 2000 cluster service on Node A to create the first node in the cluster.

**To install Cluster service on the Node A computer**

1. Click **Start**, point to **Settings**, and then click **Control Panel**.
2. In Control Panel, double-click **Add/Remove Programs**.
3. In **Add/Remove Programs**, click **Add/Remove Windows Components**.
4. In the **Windows Components Wizard**, under **Components**, select **Cluster Service**, and then click **Next**.
5. If the Files Needed page appears, enter the path for the Windows 2000 source in the **Copy files from:** box, and then click **OK**.

6. On the **Cluster Service Configuration Wizard** page, click **Next**.

7. On the **Hardware Configuration** page, click **I Understand**, and then click **Next**.

   The Hardware Configuration page contains a warning that the Cluster service is only supported when installed on hardware configurations that have been tested and are on the Cluster service Hardware Compatibility List (HCL). To continue the setup process, click **I Understand** to accept the condition that Cluster service is only supported on tested hardware.
8. On the Create or Join a Cluster page, select The first node in the cluster, and then click Next.

![Create or Join a Cluster](image)

**Figure 143. Create or Join a Cluster**

9. On the Cluster Name page, type the **Cluster_Name**, and then click Next.

You will need to enter a unique network name (NetBIOS name) for the cluster when you select this option.

When you create a cluster, you must assign a cluster name which you will use for cluster management. The name is the first virtual server of the cluster. This virtual server is installed by default in a group called Cluster Group. When you have assigned the cluster name, click Next to continue the installation.

![Cluster Name](image)

**Figure 144. Cluster Name**
10. Use the following information to complete the **Select an Account** page:

   **User name**: (Choose a Name and type it)
   **Password**: (Choose a password and type it)
   **Domain**: (Choose the domain that already exists)

   Click **Next**.

   The Cluster service account is an account that only the operating system and integrated services use. Cluster service requires a Cluster service account to stop and start or take over resources. Enter the user name, password, and domain name for the user account under which the Cluster service will run. When you enter the information and click Next, the Setup Wizard validates the user account and password.

   **Important**: The Cluster service account must already exist in the domain, and a domain controller must be available, before you can complete the installation.

![Select an Account](image)

11. On the **Cluster Service Configuration** Wizard page, click **Yes** to add the service account to the Local Administrators group.

12. On the **Add or Remove Managed Disks** page, verify that Disks (Logical units) that you designated from your FAST appears under **Managed disks**.; and then click **Next**.

   You will need to identify which disks will be a part of the cluster. These managed disks will be available as a resource to Cluster service. Designating a disk as a managed disk allows Cluster service to use the ID reserve and the ID release command. The ID reserve command allows one node to gain access to a disk resource and denies access to other nodes that do not have that reservation. The ID release command makes that disk resource available so that another node can reserve it.

   You must specify which disks will be managed by Cluster service. By default, all of the disks on all of the buses other than the system bus will appear in the
Managed disks list. Therefore, you need to remove disks if you do not want Cluster service to manage them.

13. On the Cluster File Storage page, verify that the disk (logical unit) you designated for the quorum appears under Disks:, then click Next.

You need to identify a disk resource on which to store cluster files that help manage the cluster. These files include checkpoint files and the quorum log. We recommend that the disk where you store the cluster files have at least 100 megabytes (MB) of free space and is a dedicated disk. You place applications on separate shared disks so that if failover of the application occurs, the cluster file storage disk will not also fail over.
14. On the **Configure Cluster Networks** page, click **Next**.

The Configure Cluster Networks page reminds you that using a single network for private communications allows for a single point of failure. Whenever possible, use more than one network for node-to-node communications and another network for client-to-node communications. Click Next to start the scan of network adapters.

15. On the **Network Connections** page, click **Next**.

You will need to configure each of the network adapters on each node. The Network Connections page displays the network adapters on the node one at a time, starting with the first network adapter that is detected.

You must enter a network name and select one of the configuration settings for each network adapter. The network name is a name that will be used by the Cluster service for the network adapter and the network to which the adapter is connected. For example, if the network adapter is connected to a private network that is used only by the nodes to communicate, the entry for Network name could be Cluster Net, or Private cluster connection.

16. Depending on the binding of the network card, use the information below to fill out the Cluster Public and Cluster Private networks.

The **Enable the network for cluster use** check box is selected by default. It allows Cluster service to use the network adapter. Disable this option for an adapter that is used for a dedicated networking task, such as a Web site. Cluster service will not be able to use this adapter if this check box is cleared. When you accept the enable default for a network adapter, you will next assign that adapter a communications role according to your cluster communications plan.

You have three choices:

- **Client access only**: You select Client access only to use a network adapter for communication between the cluster and the clients. No node-to-node communication will take place by means of this network adapter. The Client access only option is for the public network.

- **Internal cluster communications only**: You select Internal cluster communications only to use a network adapter for node-to-node communication. No client communication will take place by means of this network adapter. The Internal cluster communication option is for the private network.

- **All communications or mixed network**: You select All communications or mixed network only to use a network adapter for a mixed network. You will use the adapter for both node-to-node communication and for communication with clients. This option is selected by default.
Appendix B. Installing Windows 2000 Cluster service

17. On the Internal Cluster Communication page, verify that in the Networks box, Cluster Private is at the top of this list; if it is not, click Cluster Private and click Up to move Cluster Private to the top of the list, and then click Next.

18. In the Cluster IP Address page, fill out the following:
   - **IP Address**: Enter the cluster IP address of Node A.
   - **Subnet mask**: Enter the Subnet mask
   - **Network**: Cluster Public

19. Click Next.

20. On the Completing the Cluster Service Configuration Wizard page, click Finish to finish the Cluster service installation.

21. On the Cluster Service Configuration Wizard page, click OK to verify that the service has started successfully.


23. Click Close to close the Add/Remove Programs box.


B.2 Install Cluster service on Node B

In this stage, you will join Node B to the cluster by installing Windows 2000 Cluster service on Node B.

**Install Cluster service on the Node B computer**

1. Log on as Administrator.

2. Click Start, point to Settings, and then click Control Panel.

3. In Control Panel, double-click Add/Remove Programs.

4. In Add/Remove Programs, click Add/Remove Windows Components.
5. On the Windows Components Wizard page, under Components, select Cluster Service, then click Next.

6. If the Files Needed page appears, enter the path for the Windows 2000 source files in the Copy files from: box, and then click OK.

7. On the Cluster Service Configuration Wizard page, click Next.

8. On the Hardware Configuration page, click I Understand, then click Next.

9. On the Create or Join a Cluster page, select The second or next node in the cluster, and then click Next.

![Cluster Service Configuration Wizard](image)

Figure 149. Create or Join a Cluster

10. On the Cluster Name page, type the Cluster_Name that you chose earlier for the cluster and then click Next.

11. On the Cluster Service Configuration Wizard page, click OK.

12. On the Select an Account page, in the Password box, type the cluster account password and then click Next.

13. On the Cluster Service Configuration Wizard page, click Yes to add the service account to the Local Administrators group.

On the **Cluster Service Configuration Wizard** page, click **OK**.
16. On the **Windows Components Wizard** page, click **Finish**.
17. Click **Close** to close the **Add/Remove Programs** box.
18. Close all of the windows.

**B.3 Examining the system event log**

After installing Cluster service, you need to verify that the service is properly installed and that the following changes have been made to your system:

- Cluster service is installed.
- Cluster service files and folders are installed in the correct directories.
- The new entries have been written to the registry.
- The event log indicates that a successful cluster has been formed.

You will need to repeat these verification steps each time that you add a new node to the cluster.

**Verify that Cluster service is installed on Node A and Node B**

1. Click **Start**, point to **Programs**, point to **Administrative Tools**, and then click **Event Viewer**.
2. Click the **System Log** from the console tree.
3. Scroll to the bottom of the details pane and locate the first entry of the ClusSvc Source. Read the descriptions of each event from source ClusSvc as you step through the System Log and make sure that there aren’t any errors.
4. Close all of the windows.
B.4 View changes made to each node of the cluster

You can now examine changes made to each node after the installation of Cluster service. If Cluster service fails to install, or there is an error in the system event log, you would typically check whether the service account has been correctly identified, cluster-specific files and folders were installed, and that the registry added the appropriate registry keys.

Verify the service account for Cluster service
1. Click Start, point to Programs, point to Administrative Tools, and then click Services.
2. The Services snap-in will open.
3. Double-click Cluster Service from the details pane.
4. Click the Log On tab.
5. Verify that the Logon as: account is what you defined Click OK to close the Cluster Service Properties.

Examine the files and folders installed by Cluster service
1. Click Start, click Run.
2. Type %Systemroot%\cluster and then click OK.
3. Windows Explorer opens showing the files installed by Cluster service to each node.
   New files and folders are found in:
   - %systemroot%\cluster
   - %systemroot%\system32
   - \mscs on the quorum disk
4. Some of the key files in this folder are the cluadmin.exe, clussvc.exe, and the cluster.log.

Examine the changes made to the registry by Cluster service
1. Click Start, and then click Run.
2. Type regedt32 and then click OK.
3. To prevent accidental changes to the registry, from the Registry Editor menu, click Options and then click Read Only Mode.
4. From the Registry Editor menu, click Windows and then click HKEY_LOCAL_MACHINE on Local Machine. Double-click on the Cluster subkey.
5. There are seven Cluster subkeys that were added to the local registry.
The proper subkeys for the Cluster Registry are as follows:
   - Groups
   - NetworkInterfaces
   - Networks
   - Nodes
   - Quorum
• Resources
• ResourcesTypes

6. Close all of the windows.
Appendix C. Special notices

This publication is intended to help IBM professionals, Business Partners and customers who are considering using one of the FASTT range of storage products. The information in this publication is not intended as the specification of any product or programming interfaces that are provided by the FASTT200 and FASTT500.

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, North Castle Drive, Armonk, NY 10504-1785.

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

- e (logo)®
- IBM®
- AIX
- AS/400
- 400
- AT
- CT
- CUA
- Current
- DB2
- DFSMS/MVS
- DFSMSdss
- Enterprise Storage Server
- Enterprise Systems Connection Architecture
- ESCON
- FICON
- IBM
- Lotus
- Redbooks
- Redbooks Logo
- MVS/DFP
- MVS/ESA
- Netfinity
- Notes
- RACF
- RAMAC
- RS/6000
- S/390
- Seaside
- ServeRAID
- SP
- StorWatch
- System/370
- System/390
- Wave
- Wizard
- XT

The following terms are trademarks of other companies:

- Tivoli, Manage. Anything. Anywhere., The Power To Manage., Anything. Anywhere., TME, NetView, Cross-Site, Tivoli Ready, Tivoli Certified, Planet Tivoli, and Tivoli Enterprise are trademarks or registered trademarks of Tivoli Systems Inc., an IBM company, in the United States, other countries, or both. In Denmark, Tivoli is a trademark licensed from Kjøbenhavns Sommer - Tivoli A/S.

- C-bus is a trademark of Corollary, Inc. 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.

- UNIX is a registered trademark in the United States and other countries licensed exclusively through The Open Group.

- SET, SET Secure Electronic Transaction, and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.

- 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 IBM Redbooks

For information on ordering these publications see “How to get IBM Redbooks” on page 219.

- *IBM e(logo)server xSeries Clustering Planning Guide*, SG24-5845
- *Netfinity Director - Integration and Tools*, SG24-5389
- *Tuning Netfinity Servers for Performance - Getting the most out of Windows 2000 and Windows NT 4.0*, SG24-5287
- *Netfinity Server Disk Subsystems*, SG24-2098

D.2 IBM Redbooks collections

Redbooks are also available on the following CD-ROMs. Click the CD-ROMs button at ibm.com/redbooks for information about all the CD-ROMs offered, updates and formats.

<table>
<thead>
<tr>
<th>CD-ROM Title</th>
<th>Collection Kit Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM System/390 Redbooks Collection</td>
<td>SK2T-2177</td>
</tr>
<tr>
<td>IBM Networking Redbooks Collection</td>
<td>SK2T-6022</td>
</tr>
<tr>
<td>IBM Transaction Processing and Data Management Redbooks Collection</td>
<td>SK2T-8038</td>
</tr>
<tr>
<td>IBM Lotus Redbooks Collection</td>
<td>SK2T-8039</td>
</tr>
<tr>
<td>Tivoli Redbooks Collection</td>
<td>SK2T-8044</td>
</tr>
<tr>
<td>IBM AS/400 Redbooks Collection</td>
<td>SK2T-2849</td>
</tr>
<tr>
<td>IBM Netfinity Hardware and Software Redbooks Collection</td>
<td>SK2T-8046</td>
</tr>
<tr>
<td>IBM RS/6000 Redbooks Collection</td>
<td>SK2T-8043</td>
</tr>
<tr>
<td>IBM Application Development Redbooks Collection</td>
<td>SK2T-8037</td>
</tr>
<tr>
<td>IBM Enterprise Storage and Systems Management Solutions</td>
<td>SK3T-3694</td>
</tr>
</tbody>
</table>

D.3 Referenced Web sites

These Web sites are also relevant as further information sources:

- http://www.pc.ibm.com/support IBM PC Support site
- http://www.qlogic.com Qlogic Web site
- http://www.jni.com JNI Web site
- http://www.hilgraeve.com Hilgraeve Web site
How to get IBM Redbooks

This section explains how both customers and IBM employees can find out about IBM Redbooks, redpieces, and CD-ROMs. A form for ordering books and CD-ROMs by fax or e-mail is also provided.

- **Redbooks Web Site** [ibm.com/redbooks](http://ibm.com/redbooks)
  
  Search for, view, download, or order hardcopy/CD-ROM Redbooks from the Redbooks Web site. Also read redpieces and download additional materials (code samples or diskette/CD-ROM images) from this Redbooks site.
  
  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.

- **E-mail Orders**
  
  Send orders by e-mail including information from the IBM Redbooks fax order form to:

<table>
<thead>
<tr>
<th>Location</th>
<th>e-mail address</th>
</tr>
</thead>
<tbody>
<tr>
<td>In United States or Canada</td>
<td><a href="mailto:pubscan@us.ibm.com">pubscan@us.ibm.com</a></td>
</tr>
<tr>
<td>Outside North America</td>
<td>Contact information is in the “How to Order” section at this site:</td>
</tr>
</tbody>
</table>

- **Telephone Orders**

<table>
<thead>
<tr>
<th>Location</th>
<th>Phone Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States (toll free)</td>
<td>1-800-879-2755</td>
</tr>
<tr>
<td>Canada (toll free)</td>
<td>1-800-IBM-4YOU</td>
</tr>
<tr>
<td>Outside North America</td>
<td>Country coordinator phone number is in the “How to Order” section at this site:</td>
</tr>
</tbody>
</table>

- **Fax Orders**

<table>
<thead>
<tr>
<th>Location</th>
<th>Phone Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States (toll free)</td>
<td>1-800-445-9269</td>
</tr>
<tr>
<td>Canada</td>
<td>1-403-267-4455</td>
</tr>
<tr>
<td>Outside North America</td>
<td>Fax phone number is in the “How to Order” section at this site:</td>
</tr>
</tbody>
</table>

This information was current at the time of publication, but is continually subject to change. The latest information may be found at the Redbooks Web site.

---

**IBM Intranet for Employees**

**IBM Redbooks fax order form**

Please send me the following:

<table>
<thead>
<tr>
<th>Title</th>
<th>Order Number</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

First name

Last name

Company

Address

City

Postal code

Country

Telephone number

Telefax number

VAT number

☐ Invoice to customer number

☐ 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.
This glossary contains a list of generally computing terms.

A

Access logical drive. The Access Logical Drive is a special drive which uses none of the physical disk drives and should be assigned to the last (highest) available LUN number. Typically that LUN number will be LUN 31. The Access Logical Drive allows the Storage Manager Agent to communicate to the Fibre Channel RAID controllers through the fibre connection for storage management services. These services include monitoring, configuring, and maintaining the RAID storage device.

application system. A system made up of one or more host systems that perform the main set of functions for an establishment. This is the system that updates the primary DASD volumes that are being copied by a copy services function.

AOM. Asynchronous operations manager.

APAR. Authorized program analysis report.

array. An arrangement of related disk drive modules that you have assigned to a group.

assigned storage. On the ESS, this is the space that you have allocated to volumes, and assigned to a port.

asynchronous operation. A type of operation in which the remote copy XRC function copies updates to the secondary volume of an XRC pair at some time after the primary volume is updated. Contrast with synchronous operation.

ATTIME. A keyword for requesting deletion or suspension at a specific target time.

availability. The degree to which a system or resource is capable of performing its normal function.

B

backup. The process of creating a copy of data to ensure against accidental loss.

C

cache. A random access electronic storage in selected storage controls used to retain frequently used data for faster access by the channel.

cache fast write. A form of fast write where the subsystem writes the data directly to cache, where it is available for later destaging.

CCA. Channel connection address.

CCW. Channel command word.

CEC. Central electronics complex.

channel. (1) A path along which signals can be sent; for example, data channel and output channel. (2) A functional unit, controlled by the processor, that handles the transfer of data between processor storage and local peripheral equipment.

channel connection address (CCA). The input/output (I/O) address that uniquely identifies an I/O device to the channel during an I/O operation.

channel interface. The circuitry in a storage control that attaches storage paths to a host channel.

channel path. The ESA/390 term for the interconnection between a channel and its associated controllers.

channel subsystem. The ESA/390 term for the part of host computer that manages I/O communication between the program and any attached controllers.

CKD. Count key data. An ES/390 architecture term for a device that specifies the format of and access mechanism for the logical data units on the device. The logical data unit is a track that can contain one or more records, each consisting of a count field, a key field (optional), and a data field (optional).

CLIST. TSO command list.

cluster. See storage cluster.

cluster processor complex (CPC). The unit within a cluster that provides the management function for the storage server. It consists of cluster processors, cluster memory, and related logic.

concurrent copy. A copy services function that produces a backup copy and allows concurrent access to data during the copy.

concurrent maintenance. The ability to service a unit while it is operational.

consistency group time. The time, expressed as a primary application system time-of-day (TOD) value, to which XRC secondary volumes have been updated. This term was previously referred to as "consistency time".

consistent copy. A copy of data entity (for example a logical volume) that contains the contents of the entire data entity from a single instant in time.

contingent allegiance. ESA/390 term for a relationship that is created in a controller between a device and a channel path when unit-check status is accepted by the channel. The allegiance causes the controller to guarantee access; the controller does not present the busy status to the device. This enables the controller to retrieve sense data that is associated with the unit-check status, on the channel path with which the allegiance is associated.
control unit address (CUA). The high order bits of the storage control address, used to identify the storage control to the host system.

Note: The control unit address bits are set to zeros for ESCON attachments.

CUA. Control unit address.

daisy chain. A method of device interconnection for determining interrupt priority by connecting the interrupt sources serially.

DA. Device adapter.

DASD. Direct access storage device. See disk drive module.

data availability. The degree to which data is available when needed. For better data availability when you attach multiple hosts that share the same data storage, configure the data paths so that data transfer rates are balanced among the hosts.

data sharing. The ability of homogenous or divergent host systems to concurrently utilize information that they store on one or more storage devices. The storage facility allows configured storage to be accessible to any attached host systems, or to all. To use this capability, you need to design the host program to support data that it is sharing.

DDM. Disk drive module

data compression. A technique or algorithm that you use to encode data such that you can store the encoded result in less space than the original data. This algorithm allows you to recover the original data from the encoded result through a reverse technique or reverse algorithm.

data field. The third (optional) field of a CKD record. You determine the field length by the data length that is specified in the count field. The data field contains data that the program writes.

data record. A subsystem stores data records on a track by following the track-descriptor record. The subsystem numbers the data records consecutively, starting with 1. A track can store a maximum of 255 data records. Each data record consists of a count field, a key field (optional), and a data field (optional).

DASD-Fast Write. A function of a storage controller that allows caching of active write data without exposure of data loss by journaling of the active write data in NVS.

DASD subsystem. A DASD storage control and its attached direct access storage devices.

data in transit. The update data on application system DASD volumes that is being sent to the recovery system for writing to DASD volumes on the recovery system.

data mover. See system data mover.

dedicated storage. Storage within a storage facility that is configured such that a single host system has exclusive access to the storage.

demote. The action of removing a logical data unit from cache memory. A subsystem demotes a data unit in order to make room for other logical data units in the cache. It could also demote a data unit because the logical data unit is not valid. A subsystem must destage logical data units with active write units before they are demoted.

destage. (1) The process of reading data from cache. (2) The action of storing a logical data unit in cache memory with active write data to the storage device. As a result, the logical data unit changes from cached active write data to cached read data.

device. The ESA/390 term for a disk drive.

device address. The ESA/390 term for the field of an ESCON device-level frame that selects a specific device on a control-unit image. The one or two leftmost digits are the address of the channel to which the device is attached. The two rightmost digits represent the unit address.

device adapter. A physical sub unit of a storage controller that provides the ability to attach to one or more interfaces used to communicate with the associated storage devices.

device ID. An 8-bit identifier that uniquely identifies a physical I/O device.

device interface card. A physical sub unit of a storage cluster that provides the communication with the attached DDMs.

device number. ESA/390 term for a four-hexadecimal-character identifier, for example 13A0, that you associate with a device to facilitate communication between the program and the host operator. The device number that you associate with a subchannel.

device sparing. Refers to when a subsystem automatically copies data from a failing DDM to a spare DDM. The subsystem maintains data access during the process.

Device Support Facilities program (ICKDSF). A program used to initialize DASD at installation and perform media maintenance.

DFDSS. Data Facility Data Set Services.

DFSMSdss. A functional component of DFSMS/MVS used to copy, dump, move, and restore data sets and volumes.

director. See storage director and ESCON Director.

disaster recovery. Recovery after a disaster, such as a fire, that destroys or otherwise disables a system. Disaster recovery techniques typically involve restoring
Data to a second (recovery) system, then using the recovery system in place of the destroyed or disabled application system. See also recovery, backup, and recovery system.

disk drive module. The primary nonvolatile storage medium that you use for any host data that is stored within a subsystem. Number and type of storage devices within a storage facility may vary.

drawer. A unit that contains multiple DDMs, and provides power, cooling, and related interconnection logic to make the DDMs accessible to attached host systems.

DRAIN. A keyword for requesting deletion or suspension when all existing record updates from the storage control cache have been cleared.

drawer. A unit that contains multiple DDMs, and provides power, cooling, and related interconnection logic to make the DDMs accessible to attached host systems.

dump. A capture of valuable storage information at the time of an error.

dual copy. A high availability function made possible by the nonvolatile storage in cached IBM storage controls. Dual copy maintains two functionally identical copies of designated DASD volumes in the logical storage subsystem, and automatically updates both copies every time a write operation is issued to the dual copy logical volume.

duplex pair. A volume comprised of two physical devices within the same or different storage subsystems that are defined as a pair by a dual copy, PPRC, or XRC operation, and are in neither suspended nor pending state. The operation records the same data onto each volume.

E

ECSA. Extended common service area.

EMIF. ESCON Multiple Image Facility. An ESA/390 function that allows LPARs to share an ESCON channel path by providing each LPAR with its own channel-subsystem image.

environmental data. Data that the storage control must report to the host; the data can be service information message (SIM) sense data, logging mode sense data, an error condition that prevents completion of an asynchronous operation, or a statistical counter overflow. The storage control reports the appropriate condition as unit check status to the host during a channel initiated selection. Sense byte 2, bit 3 (environmental data present) is set to 1.

Environmental Record Editing and Printing (EREP) program. The program that formats and prepares reports from the data contained in the error recording data set (ERDS).

EREP. Environmental Record Editing and Printing Program.

ERP. Error recovery procedure.

ESCD. ESCON Director.

ESCM. ESCON Manager.

ESCON. Enterprise Systems Connection Architecture. An ESA/390 computer peripheral interface. The I/O interface utilizes ESA/390 logical protocols over a serial interface that configures attached units to a communication fabric.

ESCON Director (ESCD). A device that provides connectivity capability and control for attaching any two ESCON links to each other.

ESCON Manager (ESCM). A licensed program that provides host control and intersystem communication capability for ESCON Director connectivity operations.

extended remote copy (XRC). A hardware- and software-based remote copy service option that provides an asynchronous volume copy across storage subsystems for disaster recovery, device migration, and workload migration.

F

F_Node Fabric Node - a fabric attached node.

F_Port Fabric Port - a port used to attach a NodePort (N_Port) to a switch fabric.

Fabric Fibre Channel employs a fabric to connect devices. A fabric can be as simple as a single cable connecting two devices. The term is most often used to describe a more complex network utilizing hubs, switches and gateways.

Fabric Login Fabric Login (FLOGI) is used by an N_Port to determine if a fabric is present and, if so, to initiate a session with the fabric by exchanging service parameters with the fabric. Fabric Login is performed by an N_Port following link initialization and before communication with other N_Ports is attempted.

FC Fibre Channel

FC-0 Lowest level of the Fibre Channel Physical standard, covering the physical characteristics of the interface and media

FC-1 Middle level of the Fibre Channel Physical standard, defining the 8B/10B encoding/decoding and transmission protocol.

FC-2 Highest level of the Fibre Channel Physical standard, defining the rules for signaling protocol and describing transfer of frame, sequence and exchanges.

FC-3 The hierarchical level in the Fibre Channel standard that provides common services such as striping definition.
FC-4 The hierarchical level in the Fibre Channel standard that specifies the mapping of upper-layer protocols to levels below.

FCA Fiber Channel Association.

FC-AL Fibre Channel Arbitrated Loop - A reference to the Fibre Channel Arbitrated Loop standard, a shared gigabit media for up to 127 nodes one of which may be attached to a switch fabric. FC-FP Fibre Channel HIPPI Framing Protocol - A reference to the document (ANSI X3.254-1994) defining how the HIPPI framing protocol is transported via the fibre channel

FC-GS Fibre Channel Generic Services - A reference to the document (ANSI X3.289-1996) describing a common transport protocol used to communicate with the server functions, a full X500 based directory service, mapping of the Simple Network Management Protocol (SNMP) directly to the Fibre Channel, a time server and an alias server.

FC-LE Fibre Channel Link Encapsulation - A reference to the document (ANSI X3.287-1996) which defines how IEEE 802.2 Logical Link Control (LLC) information is transported via the Fibre Channel.

FC-PH A reference to the Fibre Channel Physical and Signaling standard ANSI X3.230, containing the definition of the three lower levels (FC-0, FC-1, and FC-2) of the Fibre Channel.

FC-PLDA Fibre Channel Private Loop Direct Attach - See PLDA.

FC-SB Fibre Channel Single Byte Command Code Set - A reference to the document (ANSI X3.271-1996) which defines how the ESCON command set protocol is transported using the fibre channel.

FC-SW Fibre Channel Switch Fabric - A reference to the ANSI standard under development that further defines the fabric behavior described in FC-FG and defines the communications between different fabric elements required for those elements to coordinate their operations and management address assignment.

FBA. Fixed block address. An architecture for logical devices that specifies the format of and access mechanisms for the logical data units on the device. The logical data unit is a block. All blocks on the device are the same size (fixed size); the subsystem can access them independently.

FC-AL. Fibre Channel - Arbitrated Loop. An implementation of the fibre channel standard that uses a ring topology for the communication fabric.

FCS. See Fibre Channel standard.

Fibre Channel standard. An ANSI standard for a computer peripheral interface. The I/O interface defines a protocol for communication over a serial interface that configures attached units to a communication fabric. The protocol has two layers. The IP layer defines basic interconnection protocols. The upper layer supports one or more logical protocols (for example FCP for SCSI command protocols, SBCON for ESA/390 command protocols).

Fibre Channel ports. There are five basic kinds of ports defined in the Fibre Channel architecture, as well as some vendor-specific variations. The five basic ports are as follows:

1. Node Ports, N_ports. These ports are found in Fibre Channel Nodes, which are defined to be the source or destination of Information Units (IUs). I/O devices and host systems interconnected in point-to-point or switched topologies use N_ports for their connections. N_ports can only attach to other N_ports or to F_ports. The ESS Fibre Channel adapters support the N_port functionality when connected directly to a host or to a fabric.

Node-Loop Ports, NL_ports. These ports are just like the N_ports described above, except that they connect to a Fibre Channel Arbitrated Loop (FC-AL) topology. NL_ports can only attach to other NL_ports or to FL_ports. The ESS Fibre Channel adapters support the NL_port functionality when connected directly to a loop.

2. Fabric Ports, F_ports. These ports are found in Fibre Channel Switched Fabrics. They are not the source or destination of IUs, but instead function only as a "middleman" to relay the IUs from the sender to the receiver. F_ports can only attach to N_ports. The ESS Fibre Channel adapters do not support the F_port functionality, which is found only in fabrics.

3. Fabric-Loop Ports, FL_ports. These ports are just like the F_ports described above, except that they connect to an FC-AL topology. FL_ports can only attach to NL_ports. The ESS Fibre Channel adapters do not support the FL_port functionality, which is found only in fabrics or hubs.

4. Expansion Ports, E_ports. These ports are found in Fibre Channel Switched Fabrics and are used to interconnect the individual switch or routing elements. They are not the source or destination of IUs, but instead function like the F_ports and FL_ports to relay the IUs from one switch or routing element to another. E_ports can only attach to other E_ports. The ESS Fibre Channel adapters do not support the E_port functionality, which is found only in fabrics or hubs.

5. Combination E/F Ports, G_ports. These ports are sometimes found in Fibre Channel Switched Fabrics and are used either as E Ports, when the link is connected to another switch, or as F Ports, when the link is connected to an N Port for a host or device. This port automatically determines what mode to run in after determining what it is connected to.

fiber optic cable. A fiber, or bundle of fibers, in a structure built to meet optic, mechanical, and environmental specifications.
fixed utility volume. A simplex volume assigned by the storage administrator to a logical storage subsystem to serve as working storage for XRC functions on that storage subsystem.

FlashCopy. A point-in-time copy services function that can quickly copy data from a source location to a target location.

floating utility volume. Any volume of a pool of simplex volumes assigned by the storage administrator to a logical storage subsystem to serve as dynamic storage for XRC functions on that storage subsystem.

G
GB. Gigabyte.
gigabyte. 1 073 741 824 bytes.
group. A group consist of eight DDMs. Each DDM group is a raid array.
GTF. Generalized trace facility.

H
HA. Home address, host adapter.
hard drive. A storage medium within a storage server used to maintain information that the storage server requires.
HDA. Head and disk assembly. The portion of an HDD associated with the medium and the read/write head.
HDD. Head and disk drive.
home address. A nine-byte field at the beginning of a track that contains information that identifies the physical track and its association with a cylinder.
host adapter. A physical sub unit of a storage controller that provides the ability to attach to one or more host I/O interfaces.

I
ICKDSF. See Device Support Facilities program.
identifier (ID). A sequence of bits or characters that identifies a program, device, storage control, or system.
IML. Initial microcode load.
initial microcode load (IML). The act of loading microcode.
I/O device. An addressable input/output unit, such as a direct access storage device, magnetic tape device, or printer.
I/O interface. An interface that you define in order to allow a host to perform read and write operations with its associated peripheral devices.

implicit allegiance. ESA/390 term for a relationship that a controller creates between a device and a channel path, when the device accepts a read or write operation. The controller guarantees access to the channel program over the set of channel paths that it associates with the allegiance.

Internet Protocol (IP). A protocol used to route data from its source to its destination in an Internet environment.
invalidation. The action of removing a logical data unit from cache memory because it cannot support continued access to the logical data unit on the device. This removal may be the result of a failure within the storage controller or a storage device that is associated with the device.
IPL. Initial program load.
ITSO. International Technical Support Organization.

J
JCL. Job control language.
Job control language (JCL). A problem-oriented language used to identify the job or describe its requirements to an operating system.
journal. A checkpoint data set that contains work to be done. For XRC, the work to be done consists of all changed records from the primary volumes. Changed records are collected and formed into a "consistency group", and then the group of updates is applied to the secondary volumes.

K
KB. Kilobyte.
key field. The second (optional) field of a CKD record. The key length is specified in the count field. The key length determines the field length. The program writes the data in the key field. The subsystem uses this data to identify or locate a given record.
keyword. A symptom that describes one aspect of a program failure.
kilobyte (KB). 1,024 bytes.
km. Kilometer.

L
LAN. See local area network.
least recently used. The algorithm used to identify and make available the cache space that contains the least-recently used data.
licensed internal code (LIC).
(1) Microcode that IBM does not sell as part of a machine, but licenses to the customer. LIC is
implemented in a part of storage that is not addressable by user programs. Some IBM products use it to implement functions as an alternative to hard-wired circuitry.

(2) LIC is implemented in a part of storage that is not addressable by user programs. Some IBM products use it to implement functions as an alternative to hard-wired circuitry.

**link address.** On an ESCON interface, the portion of a source, or destination address in a frame that ESCON uses to route a frame through an ESCON director. ESCON associates the link address with a specific switch port that is on the ESCON director. Equivalently, it associates the link address with the channel-subsystem, or controller-link-level functions that are attached to the switch port.

**link-level facility.** ESCON term for the hardware and logical functions of a controller or channel subsystem that allows communication over an ESCON write interface and an ESCON read interface.

**local area network (LAN).** A computer network located on a user's premises within a limited geographical area.

**logical address.** On an ESCON interface, the portion of a source or destination address in a frame used to select a specific channel-subsystem or control-unit image.

**logical data unit.** A unit of storage which is accessible on a given device.

**logical device.** The functions of a logical subsystem with which the host communicates when performing I/O operations to a single addressable-unit over an I/O interface. The same device may be accessible over more than one I/O interface.

**logical disk drive.** See logical volume.

**logical subsystem.** The logical functions of a storage controller that allow one or more host I/O interfaces to access a set of devices. The controller aggregates the devices according to the addressing mechanisms of the associated I/O interfaces. One or more logical subsystems exist on a storage controller. In general, the controller associates a given set of devices with only one logical subsystem.

**logical unit.** The SCSI term for a logical disk drive.

**logical unit number.** The SCSI term for the field in an identifying message that is used to select a logical unit on a given target.

**logical partition (LPAR).** The ESA/390 term for a set of functions that create the programming environment that is defined by the ESA/390 architecture. ESA/390 architecture uses this term when more than one LPAR is established on a processor. An LPAR is conceptually similar to a virtual machine environment except that the LPAR is a function of the processor. Also the LPAR does not depend on an operating system to create the virtual machine environment.

**logical volume.** The storage medium associated with a logical disk drive. A logical volume typically resides on one or more storage devices. A logical volume is referred to on an AIX platform as an hdisk, an AIX term for storage space. A host system sees a logical volume as a physical volume.

**LRU.** See also least recently used. A policy for a caching algorithm which chooses to remove the item from cache which has the longest elapsed time since its last access.

**LSS.** See logical subsystem.

**LUN.** See logical unit number.

**M**

**MB.** Megabyte.

**megabyte (MB).** 1 048 576 bytes.

**metadata.** Internal control information used by microcode. It is stored in reserved area within disk array. The usable capacity of the array take care of the metadata.

**million instructions per second (MIPS).** A general measure of computing performance and, by implication, the amount of work a larger computer can do. The term is used by IBM and other computer manufacturers. For large servers or mainframes, it is also a way to measure the cost of computing: the more MIPS delivered for the money, the better the value.

**MTBF.** Mean time between failures. A projection of the time that an individual unit remains functional. The time is based on averaging the performance, or projected performance, of a population of statistically independent units. The units operate under a set of conditions or assumptions.

**Multiple Virtual Storage (MVS).** One of a family of IBM operating systems for the System/370 or System/390 processor, such as MVS/ESA.

**MVS.** Multiple Virtual Storage.

**N**

**nondisruptive.** The attribute of an action or activity that does not result in the loss of any existing capability or resource, from the customer's perspective.

**nonvolatile storage (NVS).** Random access electronic storage with a backup battery power source, used to retain data during a power failure. Nonvolatile storage, accessible from all cached IBM storage clusters, stores data during DASD fast write, dual copy, and remote copy operations.

**NVS.** Nonvolatile storage.
open system. A system whose characteristics comply with standards made available throughout the industry, and therefore can be connected to other systems that comply with the same standards.

operating system. Software that controls the execution of programs. An operating system may provide services such as resource allocation, scheduling, input/output control, and data management.

orphan data. Data that occurs between the last, safe backup for a recovery system and the time when the application system experiences a disaster. This data is lost when either the application system becomes available for use or when the recovery system is used in place of the application system.

PPRC dynamic address switching (P/DAS). A software function that provides the ability to dynamically redirect all application I/O from one PPRC volume to another PPRC volume.

predictable write. A write operation that can cache without knowledge of the existing formatting on the medium. All writes on FBA DASD devices are predictable. On CKD DASD devices, a write is predictable if it does a format write for the first record on the track.

primary device. One device of a dual copy or remote copy volume pair. All channel commands to the copy logical volume are directed to the primary device. The data on the primary device is duplicated on the secondary device. See also secondary device.

PTF. Program temporary fix.

path group. The ESA/390 term for a set of channel paths that are defined to a controller as being associated with a single LPAR. The channel paths are in a group state and are on-line to the host.

path-group identifier. The ESA/390 term for the identifier that uniquely identifies a given LPAR. The path-group identifier is used in communication between the LPAR program and a device to associate the path-group identifier with one or more channel paths. This identifier defines these paths to the control unit as being associated with the same LPAR.

partitioned data set extended (PDSE). A system-managed, page-formatted data set on direct access storage.

P/DAS. PPRC dynamic address switching.

PDSE. Partitioned data set extended.

peer-to-peer remote copy (PPRC). A hardware based remote copy option that provides a synchronous volume copy across storage subsystems for disaster recovery, device migration, and workload migration.

pending. The initial state of a defined volume pair, before it becomes a duplex pair. During this state, the contents of the primary volume are copied to the secondary volume.

pinned data. Data that is held in a cached storage control, because of a permanent error condition, until it can be destaged to DASD or until it is explicitly discarded by a host command. Pinned data exists only when using fast write, dual copy, or remote copy functions.

port. (1) An access point for data entry or exit. (2) A receptacle on a device to which a cable for another device is attached.

PPRC. Peer-to-peer remote copy.
access is for one or more channel programs, until the
allegiance ends.

**restore.** Synonym for recover.

**resynchronization.** A track image copy from the
primary volume to the secondary volume of only the
tracks which have changed since the volume was last
in duplex mode.

**RVA.** RAMAC Virtual Array Storage Subsystem.

**SAID.** System adapter identification.

**SAM.** Sequential access method.

**SCSI.** Small Computer System Interface. An ANSI
standard for a logical interface to computer peripherals
and for a computer peripheral interface. The interface
utilizes a SCSI logical protocol over an I/O interface
that configures attached targets and initiators in a
multi-drop bus topology.

**SCSI ID.** A unique identifier assigned to a SCSI device
that is used in protocols on the SCSI interface to
identify or select the device. The number of data bits
on the SCSI bus determines the number of available
SCSI IDs. A wide interface has 16 bits, with 16
possible IDs. A SCSI device is either an initiator or a
target.

**Seascape architecture.** A storage system
architecture developed by IBM for open system
servers and S/390 host systems. It provides storage
solutions that integrate software, storage
management, and technology for disk, tape, and
optical storage.

**secondary device.** One of the devices in a dual copy
or remote copy logical volume pair that contains a
duplicate of the data on the primary device. Unlike the
primary device, the secondary device may only accept
a limited subset of channel commands.

**segment.** A segment is the amount of data, in
kilobytes, that the controller writes on a single
drive in a logical drive before writing data on the
next drive.

**sequential access.** A mode of accessing data on a
medium in a manner that requires the storage device
to access consecutive storage locations on the
medium.

**server.** A type of host that provides certain services to
other hosts that are referred to as clients.

**service information message (SIM).** A message,
generated by a storage subsystem, that is the result of
error event collection and analysis. A SIM indicates
that some service action is required.

**sidefile.** A storage area used to maintain copies of
tracks within a concurrent copy domain. A concurrent
copy operation maintains a sidefile in storage control
cache and another in processor storage.

**SIM.** Service information message.

**simplex state.** A volume is in the simplex state if it is
not part of a dual copy or a remote copy volume pair.
Ending a volume pair returns the two devices to the
simplex state. In this case, there is no longer any
capability for either automatic updates of the
secondary device or for logging changes, as would be
the case in a suspended state.

**SMF.** System Management Facilities.

**SMS.** Storage Management Subsystem.

**SRM.** System resources manager.

**SnapShot copy.** A point-in-time copy services function
that can quickly copy data from a source location to a
target location.

**spare.** A disk drive that is used to receive data from a
device that has experienced a failure that requires
disruptive service. A spare can be pre-designated to
allow automatic dynamic sparing. Any data on a disk
drive that you use as a spare is destroyed by the
dynamic sparing copy process.

**SSA.** Serial Storage Architecture. An IBM standard for
a computer peripheral interface. The interface uses a
SCSI logical protocol over a serial interface that
configures attached targets and initiators in a ring
topology.

**SSID.** Subsystem identifier.

**stacked status.** An ESA/390 term used when the
control unit is holding for the channel; the channel
responded with the stack-status control the last time
the control unit attempted to present the status.

**stage.** The process of reading data into cache from a
disk drive module.

**storage cluster.** A power and service region that runs
channel commands and controls the storage devices.
Each storage cluster contains both channel and device
interfaces. Storage clusters also perform the DASD
control functions.

**storage control.** The component in a storage
subsystem that handles interaction between processor
channel and storage devices, runs channel
commands, and controls storage devices.

**STORAGE_CONTROL_DEFAULT.** A specification
used by several XRC commands and messages to
refer to the timeout value specified in the maintenance
panel of the associated storage control.

**storage device.** A physical unit which provides a
mechanism to store data on a given medium such
that it can be subsequently retrieved. Also see disk
drive module.
storage director. In an IBM storage control, a logical
entity consisting of one or more physical storage paths
in the same storage cluster. See also storage path.

storage facility. (1) A physical unit which consists of
a storage controller integrated with one or
more storage devices to provide storage
capability to a host computer. (2) A storage server
and its attached storage devices.

Storage Management Subsystem (SMS). A
component of MVS/DFP that is used to automate and
centralize the management of storage by providing the
storage administrator with control over data class,
storage class, management class, storage group,
aggregate group and automatic class selection routine
definitions.

storage server. A unit that manages attached
storage devices and provides access to the storage
or storage related functions for one or more
attached hosts.

storage path. The hardware within the IBM storage
control that transfers data between the DASD and a
channel. See also storage director.

storage subsystem. A storage control and its
attached storage devices.

string. A series of connected DASD units sharing the
same A-unit (or head of string).

striping. A technique that distributes data in bit,
byte, multibyte, record, or block increments
across multiple disk drives.

subchannel. A logical function of a channel
subsystem associated with the management of a
single device.

subsystem. See DASD subsystem or storage
subsystem.

subsystem identifier (SSID). A user-assigned
number that identifies a DASD subsystem. This
number is set by the service representative at the time
of installation and is included in the vital product data.

suspended state. When only one of the devices in a
dual copy or remote copy volume pair is being updated
because of either a permanent error condition or an
authorized user command. All writes to the remaining
functional device are logged. This allows for automatic
resynchronization of both volumes when the volume
pair is reset to the active duplex state.

synchronization. An initial volume copy. This is a
track image copy of each primary track on the volume
to the secondary volume.

synchronous operation. A type of operation in which
the remote copy PPRC function copies updates to the
secondary volume of a PPRC pair at the same time
that the primary volume is updated. Contrast with
asynchronous operation.

system data mover. A system that interacts with
storage controls that have attached XRC primary
volumes. The system data mover copies updates made to the XRC primary volumes to a set of
XRC-managed secondary volumes.

system-managed data set. A data set that has been
assigned a storage class.

T

TCP/IP. Transmission Control Protocol/Internet
Protocol.

TOD. Time of day.

Time Sharing Option (TSO). A System/370 operating
system option that provides interactive time sharing
from remote terminals.

timeout. The time in seconds that the storage control
remains in a “long busy” condition before physical
sessions are ended.

timestamp. The affixed value of the system
time-of-day clock at a common point of reference for all
write I/O operations directed to active XRC primary
volumes. The UTC format is yyyy.ddd
hh:mm:ss.thmiju.

track. A unit of storage on a CKD device that can be
formatted to contain a number of data records. Also
see home address, track-descriptor record, and data
record.

track-descriptor record. A special record on a track
that follows the home address. The control program
uses it to maintain certain information about the track.
The record has a count field with a key length of zero,
a data length of 8, and a record number of 0. This
record is sometimes referred to as R0.

TSO. Time Sharing Option.

U

Ultra-SCSI. An enhanced small computer system
interface.

unit address. The ESA/390 term for the address
associated with a device on a given controller. On
ESCON interfaces, the unit address is the same as the
device address. On OEMI interfaces, the unit address
specifies a controller and device pair on the interface.

Universal Time, Coordinated. Replaces Greenwich
Mean Time (GMT) as a global time reference. The
format is yyyy.ddd hh:mm:ss.thmiju.

utility volume. A volume that is available to be used
by the extended remote copy function to perform data
mover I/O for a primary site storage control’s
XRC-related data.

UTC. Universal Time, Coordinated.
v
vital product data (VPD). Nonvolatile data that is stored in various locations in the DASD subsystem. It includes configuration data, machine serial number, and machine features.

volume. An ESA/390 term for the information recorded on a single unit of recording medium. Indirectly, it can refer to the unit of recording medium itself. On a non-removable medium storage device, the terms may also refer, indirectly, to the storage device that you associate with the volume. When you store multiple volumes on a single storage medium transparently to the program, you may refer to the volumes as logical volumes.

VPD. See also vital product data. Information that uniquely defines the system, hardware, software, and microcode elements of a processing system.

VSAM. Virtual storage access method.

VTOC. Volume table of contents.

W
workload migration. The process of moving an application's data from one set of DASD to another for the purpose of balancing performance needs, moving to new hardware, or temporarily relocating data.

write hit. A write operation where the data requested is in the cache.

write miss. A write operation where the data requested is not in the cache.

write penalty. The term that describes the classical RAID write operation performance impact.

write update. A write operation that updates a direct access volume.

X
XDF. Extended distance feature (of ESCON).

XRC. Extended remote copy.

XRC planned-outage-capable. A storage subsystem with an LIC level that supports a software bitmap but not a hardware bitmap.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANSI</td>
<td>American National Standards Institute</td>
</tr>
<tr>
<td>API</td>
<td>application programming interface</td>
</tr>
<tr>
<td>ATA</td>
<td>advanced technology attachment</td>
</tr>
<tr>
<td>BIOS</td>
<td>basic input/output system</td>
</tr>
<tr>
<td>CA</td>
<td>Computer Associates</td>
</tr>
<tr>
<td>CAD/CAM</td>
<td>computer-aided design/computer-aided manufacturing</td>
</tr>
<tr>
<td>CD-ROM</td>
<td>compact disk read-only memory</td>
</tr>
<tr>
<td>CPU</td>
<td>central processing unit</td>
</tr>
<tr>
<td>CRC</td>
<td>cyclic redundancy check</td>
</tr>
<tr>
<td>DHCP</td>
<td>Dynamic Host Configuration Protocol</td>
</tr>
<tr>
<td>DLT</td>
<td>digital linear tape</td>
</tr>
<tr>
<td>DMA</td>
<td>direct memory access</td>
</tr>
<tr>
<td>DMI</td>
<td>Desktop Management Interface</td>
</tr>
<tr>
<td>DOS</td>
<td>disk operating system</td>
</tr>
<tr>
<td>ECC</td>
<td>error checking and correcting</td>
</tr>
<tr>
<td>EDO</td>
<td>extended data out</td>
</tr>
<tr>
<td>EEPROM</td>
<td>electrically erased programmable read-only memory</td>
</tr>
<tr>
<td>EIDE</td>
<td>enhanced integrated drive electronics</td>
</tr>
<tr>
<td>ESCON</td>
<td>enterprise systems connection</td>
</tr>
<tr>
<td>ESM</td>
<td>Enclosure Services Monitor</td>
</tr>
<tr>
<td>ESS</td>
<td>Enterprise Storage Server</td>
</tr>
<tr>
<td>FC</td>
<td>Fibre Channel</td>
</tr>
<tr>
<td>FC-AL</td>
<td>Fibre Channel Arbitrated Loop</td>
</tr>
<tr>
<td>FCP</td>
<td>Fibre Channel Protocol</td>
</tr>
<tr>
<td>FICON</td>
<td>fiber distributed data interface</td>
</tr>
<tr>
<td>FRU</td>
<td>field replaceable unit</td>
</tr>
<tr>
<td>GB</td>
<td>gigabytes</td>
</tr>
<tr>
<td>GBIC</td>
<td>gigabit interface converter</td>
</tr>
<tr>
<td>GUI</td>
<td>graphical user interface</td>
</tr>
<tr>
<td>HIPPI</td>
<td>high performance parallel interface</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>HVD</td>
<td>high voltage differential</td>
</tr>
<tr>
<td>I/O</td>
<td>input/output</td>
</tr>
<tr>
<td>IBM</td>
<td>International Business Machines Corporation</td>
</tr>
<tr>
<td>IDE</td>
<td>integrated drive electronics</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>IPI</td>
<td>intelligent peripheral interface</td>
</tr>
<tr>
<td>ISO</td>
<td>International Standards Organization</td>
</tr>
<tr>
<td>ISP</td>
<td>Internet Service Provider</td>
</tr>
<tr>
<td>ITSO</td>
<td>International Technical Support Organization</td>
</tr>
<tr>
<td>JBOD</td>
<td>“just a bunch of disks” - a set of disks attached to a controller</td>
</tr>
<tr>
<td>JDK</td>
<td>Java development kit</td>
</tr>
<tr>
<td>JRE</td>
<td>Java runtime environment</td>
</tr>
<tr>
<td>JVM</td>
<td>Java virtual machine</td>
</tr>
<tr>
<td>KB</td>
<td>kilobytes</td>
</tr>
<tr>
<td>LAN</td>
<td>local area network</td>
</tr>
<tr>
<td>LCD</td>
<td>liquid crystal display</td>
</tr>
<tr>
<td>LCT</td>
<td>Life Cycle Tools</td>
</tr>
<tr>
<td>LDM</td>
<td>logical drive migration</td>
</tr>
<tr>
<td>LED</td>
<td>light-emitting diode</td>
</tr>
<tr>
<td>LUN</td>
<td>logical unit number</td>
</tr>
<tr>
<td>LVDS</td>
<td>low-voltage differential SCSI</td>
</tr>
<tr>
<td>MAC</td>
<td>medium access control</td>
</tr>
<tr>
<td>MB</td>
<td>megabytes</td>
</tr>
<tr>
<td>Mbps</td>
<td>megabits per second</td>
</tr>
<tr>
<td>MBps</td>
<td>megabytes per second</td>
</tr>
<tr>
<td>MEL</td>
<td>Major Event Log</td>
</tr>
<tr>
<td>MIA</td>
<td>media interface adapter</td>
</tr>
<tr>
<td>MIB</td>
<td>management information base</td>
</tr>
<tr>
<td>MIME</td>
<td>multipurpose internet mail extensions</td>
</tr>
<tr>
<td>MSCS</td>
<td>Microsoft Cluster Server</td>
</tr>
<tr>
<td>NAE</td>
<td>Netfinity Availability Extensions for MSCS</td>
</tr>
<tr>
<td>NAS</td>
<td>network attached storage</td>
</tr>
<tr>
<td>NLM</td>
<td>NetWare loadable module</td>
</tr>
<tr>
<td>NVRAM</td>
<td>non-volatile random access memory</td>
</tr>
<tr>
<td>PCI</td>
<td>peripheral component interconnect</td>
</tr>
<tr>
<td>PDF</td>
<td>portable document format</td>
</tr>
<tr>
<td>PFA</td>
<td>predictive failure analysis</td>
</tr>
<tr>
<td>POST</td>
<td>power on self test</td>
</tr>
<tr>
<td>RAID</td>
<td>redundant array of independent disks</td>
</tr>
<tr>
<td>RAM</td>
<td>random access memory</td>
</tr>
<tr>
<td>RDAC</td>
<td>Redundant Disk Array Controller</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>ROM</strong></td>
<td>read only memory</td>
</tr>
<tr>
<td><strong>RPM</strong></td>
<td>revolutions per minute</td>
</tr>
<tr>
<td><strong>RPO</strong></td>
<td>rotational positioning optimization</td>
</tr>
<tr>
<td><strong>RSM</strong></td>
<td>Remote System Management</td>
</tr>
<tr>
<td><strong>SAN</strong></td>
<td>storage area network</td>
</tr>
<tr>
<td><strong>SAP</strong></td>
<td>Systeme, Anwendungen und Programme in der Datenverarbeitung (Systems, Products, and Programs in Data Processing)</td>
</tr>
<tr>
<td><strong>SCO</strong></td>
<td>Santa Cruz Operation, Inc.</td>
</tr>
<tr>
<td><strong>SCSI</strong></td>
<td>small computer system interface</td>
</tr>
<tr>
<td><strong>SDRAM</strong></td>
<td>static dynamic random access memory</td>
</tr>
<tr>
<td><strong>SIC</strong></td>
<td>serial interface chip</td>
</tr>
<tr>
<td><strong>SMP</strong></td>
<td>symmetric multiprocessing</td>
</tr>
<tr>
<td><strong>SNMP</strong></td>
<td>simple network management protocol</td>
</tr>
<tr>
<td><strong>SQL</strong></td>
<td>structured query language</td>
</tr>
<tr>
<td><strong>SRAM</strong></td>
<td>static random access memory</td>
</tr>
<tr>
<td><strong>SRN</strong></td>
<td>service request number</td>
</tr>
<tr>
<td><strong>SSA</strong></td>
<td>serial storage architecture</td>
</tr>
<tr>
<td><strong>TB</strong></td>
<td>terabytes</td>
</tr>
<tr>
<td><strong>TCP/IP</strong></td>
<td>Transmission Control Protocol/Internet Protocol</td>
</tr>
<tr>
<td><strong>TPC-C</strong></td>
<td>Transaction Processing Council - C benchmark</td>
</tr>
<tr>
<td><strong>UID</strong></td>
<td>unique identifier</td>
</tr>
<tr>
<td><strong>ULP</strong></td>
<td>Upper Layer Protocols</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td>Uniform Resource Locator</td>
</tr>
<tr>
<td><strong>VHDCI</strong></td>
<td>very high density connector interface</td>
</tr>
<tr>
<td><strong>WAN</strong></td>
<td>wide area network</td>
</tr>
</tbody>
</table>
Index

A
Access logical drive 38
Access logical drive mapping 12
Adapter settings 32
AIX 53
AL_PA 2
Arrays
 creation 43
 expansion 44
Auto Volume Transfer 12, 37, 55
AVT 12, 57, 58

B
Battery 16, 19, 20
 replacement 18
Boot settings 31

C
Cabling 29
Cache 16, 18, 20
 block size 49
 flushing levels 49
 read ahead 105
 read-ahead 150
 settings 102
 write mirroring 49
Cluster 1
Cluster disk 158
Cluster server
 configuration 162
 installation 97
Cluster storage 165
 design 171
Command line 12, 38, 60
Conceptual design 132
Configuration 11
 planning 74
 replication 37
Controller 2, 17
 diagnostics 38

D
Definitions
 GBIC 195
 MIA 195
Design
 guidelines 133
 Diagnostics 12, 59
 Disk calculation 143
 Disk drive 1
 character 134
 performance 134, 135
 Disk seeks 134

E
EMW 40, 95
Enterprise management window 40
ESM 26, 28
Ethernet 17
 port 21
Event monitor 41
Event monitoring 11, 37
Exchange 2000 167
EXP500 1, 2, 9, 22, 26, 28

F
Failback 160
Failover 160
 policy 163, 164
Fan 2
 module 23
FASIT
 client 64
 configuration 74
 Host agent 64
 RDAC 64
 utilities 64
FASIT Client 95
 installation 64, 67, 68, 70
 FASIT Utilities 4, 95
 installation 65, 67, 68, 69, 71
 FASIT200 xiii, 1, 9, 10, 15, 18, 240
 front/side view 1
 FASIT500 1, 2, 9, 10, 18, 19, 48
 FC 15, 18
 loop 28
 FC hub 8
 FC switch 8
 Fibre Channel 1, 145
 Firmware settings 33

G
GBIC 17, 19, 21, 28, 31
 defined 195

H
HA 15
Heartbeat 158
Heterogeneous hosts 10
 support 37
High availability 1, 2, 63
Host Adapter 8, 15, 30
Host connections 3
Host group 89
Host port, 92
Host software
 installation 64
Host-attached management 4
Hot spare 44, 88
hot_add 4
 59
HP-UX 10, 37, 53, 55, 70
Fibre Array Storage Technology: A FASTT Introduction

I
In-band 13

L
LED 15, 16, 17, 23, 27
Linux 10, 37, 53, 65
Logical design 132
Logical drive 43, 94
ownership 106
LUN 5, 12, 31, 38
masking 6

M
Management
direct attached 38
host attached 39
methods 4
MIB file 96
Microcode
upgrading 82
Mini-hub 22, 24
Monitoring 95
MSCS 97
Multi-path driver 57

N
NAS 145
Netware 63
Novell NetWare 10, 37, 53
Novell Netware 67
NVSRAM 54

O
Out-of-band 13

P
Performance 61
monitor 62
Performance Monitor 146
Physical design 132
Planning 131
Power supply 2
Pre-fetch 150
PV-Links 71

Q
QLogic 63
Quorum 158, 166

R
RAID array
defragment 107
RAID controller 7, 8, 19
RAID levels 5, 43
migration 45
RAID migration 103
RAID-0 143
RAID-1 143
RAID-5 143
Random IO 136
RDAC 3, 55, 57
installation 65, 69
Redundancy 16
Redundant 20
configuration 18
setup 9
Rotational latency 134

S
SAN 145
Script file
executing 60
Seek time 135
Segment 87
Segment size 47
Sequential IO 135
SM7devices 59
sm7devices 4
SNMP 35, 96
Software installation 63
Solaris 68
Solution design 131, 132
Storage group 168
limits 169
Storage management
Advanced 50
Storage Manager 4, 5, 11, 19, 37
client 40
Storage partition 12, 94
Storage partitioning 6
configuration 89
considerations 53
Storage Server 1
Sun Solaris 10, 37

T
TCP/IP 38
Testing 131
Transaction log performance 174
Troubleshooting 28

V
Virtual server 158

W
Windows 2000 10, 53, 64
Windows NT 10, 37, 53, 55, 64
World Wide Name WWN 4
# IBM Redbooks review

Your feedback is valued by the Redbook authors. In particular we are interested in situations where a Redbook “made the difference” in a task or problem you encountered. Using one of the following methods, please review the Redbook, addressing value, subject matter, structure, depth and quality as appropriate.

- Use the online Contact us review redbook form found at [ibm.com/redbooks](http://ibm.com/redbooks)
- Fax this form to: USA International Access Code + 1 845 432 8264
- Send your comments in an Internet note to redbook@us.ibm.com

<table>
<thead>
<tr>
<th>Document Number</th>
<th>SG24-6246-00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redbook Title</td>
<td>Fibre Array Storage Technology: A FASTT Introduction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What other subjects would you like to see IBM Redbooks address?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Please rate your overall satisfaction:</th>
<th>O Very Good</th>
<th>O Good</th>
<th>O Average</th>
<th>O Poor</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Please identify yourself as belonging to one of the following groups:</th>
<th>O Customer</th>
<th>O Business Partner</th>
<th>O Solution Developer</th>
<th>O IBM, Lotus or Tivoli Employee</th>
<th>O None of the above</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Your email address:</th>
<th>The data you provide here may be used to provide you with information from IBM or our business partners about our products, services or activities.</th>
<th>O Please do not use the information collected here for future marketing or promotional contacts or other communications beyond the scope of this transaction.</th>
</tr>
</thead>
</table>

| Questions about IBM’s privacy policy? | The following link explains how we protect your personal information. | [ibm.com/privacy/yourprivacy/](http://ibm.com/privacy/yourprivacy/) |
This IBM Redbook gives a detailed introduction to the FASTT range of storage products with particular emphasis on the FASTT200.

Chapter 1 introduces the FASTT concepts.

Chapter 2 discusses the hardware in detail.

Chapter 3 discusses the FASTT Storage Manager in detail with emphasis on the Version 7.10 release. This release of the software introduces heterogeneous host support.

Chapter 4 discusses operating system considerations. In Chapter 4, we assume that the reader is familiar with the operating system, so although we discuss command syntax and other considerations, we do not discuss them beyond the level required to use a FASTT product.

Chapter 5 details storage solutions for various business scenarios.