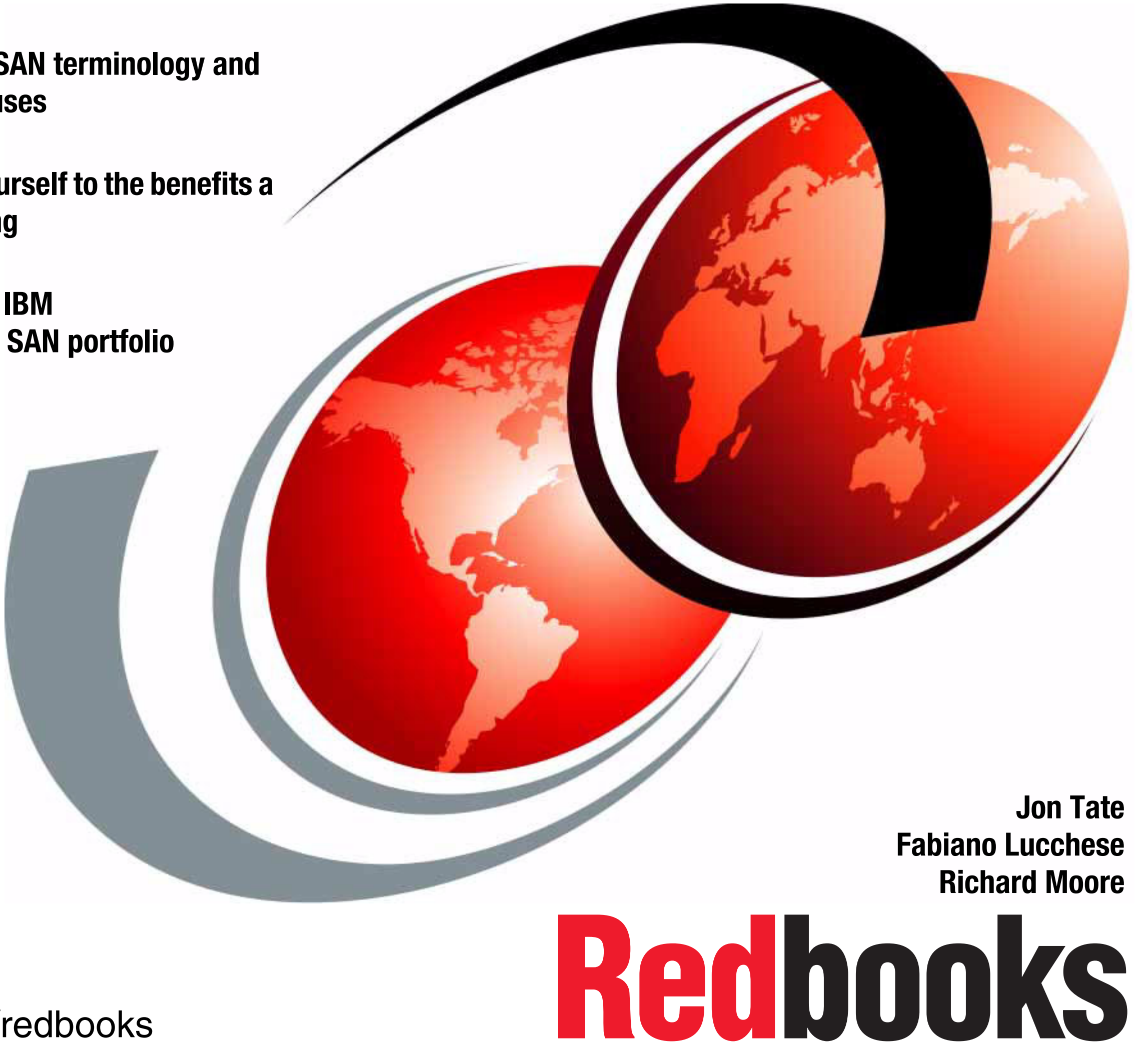
Introduction to Storage Area Networks

- Learn basic SAN terminology and component uses
- Introduce yourself to the benefits a SAN can bring
- Discover the IBM TotalStorage SAN portfolio



Jon Tate
Fabiano Lucchese
Richard Moore

Redbooks

ibm.com/redbooks

International Technical Support Organization

Introduction to Storage Area Networks

July 2006

SG24-5470-03

Note: Before using this information and the product it supports, read the information in "Notices" on page xv.

Fourth Edition (July 2006)

This edition applies to the IBM TotalStorage portfolio.

Contents

Figures

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:
This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Trademarks

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

eserver®
iSeries™
i5/OS®
pSeries®
xSeries®
z/Architecture™
z/OS®
z/VM®
zSeries®
z9™
AFS®
AIX 5L™
AIX®
AS/400®
BladeCenter®
Domino®
DB2®
DS4000™
DS6000™
DS8000™
Enterprise Storage Server®
Enterprise Systems Architecture/390®
ECKD™
ESCON®
FlashCopy®
FICON®
Geographically Dispersed Parallel Sysplex™
GDPS®
HACMP™
Informix®
IBM®
Lotus®
MVS™
Netfinity®
OS/390®
OS/400®
Parallel Sysplex®
POWER5™
PR/SM™
Redbooks™
Redbooks (logo) ™
RMF™
RS/6000®
S/360™
S/370™
S/390®
Storage Tank™
System z9™
System Storage™
System/360™
System/370™
Tivoli®
TotalStorage®
Virtualization Engine™
VM/ESA®
VSE/ESA™

The following terms are trademarks of other companies:

Java, Jini, Solaris, Streamline, Sun, Sun Microsystems, Ultra, and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Internet Explorer, Microsoft, Windows NT, Windows Server, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

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

Preface

The plethora of data created by the businesses of today is making storage a strategic investment priority for companies of all sizes. As storage takes precedence, three major initiatives have emerged:

- Infrastructure simplification: Consolidation, virtualization, and automated management with IBM® TotalStorage® can help simplify the infrastructure and ensure that an organization meets its business goals.
- Information lifecycle management: Managing business data through its lifecycle from conception until disposal in a manner that optimizes storage and access at the lowest cost.
- Business continuity: Maintaining access to data at all times, protecting critical business assets, and aligning recovery costs based on business risk and information value.

Storage is no longer an afterthought. Too much is at stake. Companies are searching for more ways to efficiently manage expanding volumes of data, and to make that data accessible throughout the enterprise; this is propelling the move of storage into the network. Also, the increasing complexity of managing large numbers of storage devices and vast amounts of data is driving greater business value into software and services.

With current estimates of the amount of data to be managed and made available increasing at 60 percent per annum, this is where a storage area network (SAN) enters the arena. Simply put, SANs are the leading storage infrastructure for the global economy of today. SANs offer simplified storage management, scalability, flexibility, availability, and improved data access, movement, and backup.

This IBM Redbook gives an introduction to the SAN. It illustrates where SANs are today, who are the main industry organizations and standard bodies active in the SAN world, and it positions an IBM comprehensive, best-of-breed approach of enabling SANs with its products and services. It introduces some of the most commonly encountered terminology and features present in a SAN.

For further reading, and a deeper dive into the SAN world, readers may find the following redbook especially useful to expand their SAN knowledge:

- *IBM TotalStorage: SAN Product, Design, and Optimization Guide*, SG24-6384

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.

Jon Tate is a Project Manager for IBM TotalStorage SAN Solutions at the International Technical Support Organization, San Jose Center. Before joining the ITSO in 1999, he worked in the IBM Technical Support Center, providing Level 2 support for IBM storage products. Jon has 20 years of experience in storage software and management, services, and support, and is both an IBM Certified IT Specialist, and an IBM SAN Certified Specialist.

Fabiano Lucchese is the Business Director of Sparsi Computing in Grid (`http://www.sparsi.com`) and works as a Grid computing consultant in a number of nation-wide projects. In 1994, Fabiano was admitted to the Computer Engineering undergraduate course of the State University of Campinas, Brazil, and in mid-1997, he moved to France to finish his undergraduate studies at the Central School of Lyon. Also in France, he pursued graduate-level studies in Industrial Automation. Back in Brazil, he joined Unisoma Mathematics for Productivity, where he worked as a Software Engineer on the development of image processing and optimization systems. From 2000 to 2002, he joined the Faculty of Electrical and Computer Engineering of the State University of Campinas as a graduate student and acquired a master's degree in Computer Engineering for developing a task scheduling algorithm for balancing processing loads on heterogeneous grids.

Richard Moore is an IT Architect who works in Technology Integration and Management (TIM) Competency, Integrated Technology Delivery. He joined IBM United Kingdom in 1979 as a Courier/Storeman. He helped to write the IBM Redbook *Implementing Snapshot*, SG24-2241, and the Redpaper *Virtualization in a SAN*, REDP-3633, and holds a software patent for a storage automation product.

Thanks to the previous authors of the first, second, and third editions of this redbook:

Angelo Bernasconi
Rajani Kanth
Ravi Kumar Khattar
Peter Mescher
Mark S. Murphy
Kjell E. Nyström
Fred Scholten
Giulio John Tarella
Andre Telles

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

Tom Cady
Deanna Polm
Sangam Racherla
Sokkieng Wang
International Technical Support Organization, San Jose Center

Lisa Dorr
David Nowlen
Michael Starling
Jeremy Stroup
IBM Storage Systems Group

Jim Baldyga
Brian Steffler
Brocade Communications Systems

John McKibben
Cisco Systems

Brent Anderson
McDATA Corporation

Tom and Jenny Chang
Garden Inn Hotel, Los Gatos, California

Become a published author

Join us for a two- to six-week residency program! Help write an IBM Redbook dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You'll team with IBM technical professionals, Business Partners and/or customers.

Your efforts will help increase product acceptance and customer satisfaction. As a bonus, you'll develop a network of contacts in IBM development labs, and increase your productivity and marketability.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our Redbooks™ to be as helpful as possible. Send us your comments about this or other Redbooks in one of the following ways:

- Use the online **Contact us** review redbook form found at:

  ibm.com/redbooks

- Send your comments in an Internet note to:

  redbook@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

1

Introduction

Computing is based on information. Information is the underlying resource on which all computing processes are based; it is a company asset. Information is stored on storage media, and is accessed by applications executing on a server. Often the information is a unique company asset. Information is created and acquired every second of every day. Information is the currency of business.

To ensure that any business delivers the expected results, they must have access to accurate information, and without delay. The management and protection of business information is vital for the availability of business processes.

This chapter introduces the concept of a Storage Area Network, which has been regarded as the ultimate response to all these needs.

1.1 What is a Storage Area Network?

The Storage Network Industry Association (SNIA) defines the SAN as a network whose primary purpose is the transfer of data between computer systems and storage elements. A SAN consists of a communication infrastructure, which provides physical connections; and a management layer, which organizes the connections, storage elements, and computer systems so that data transfer is secure and robust. The term SAN is usually (but not necessarily) identified with block I/O services rather than file access services.

A SAN can also be a storage system consisting of storage elements, storage devices, computer systems, and/or appliances, plus all control software, communicating over a network.

Note: The SNIA definition specifically does not identify the term SAN with Fibre Channel technology. When the term SAN is used in connection with Fibre Channel technology, use of a qualified phrase such as *Fibre Channel SAN* is encouraged. According to this definition, an Ethernet-based network whose primary purpose is to provide access to storage elements would be considered a SAN. SANs are sometimes also used for system interconnection in clusters.

Put in simple terms, a SAN is a specialized, high-speed network attaching servers and storage devices and, for this reason, It is sometimes referred to as “the network behind the servers.” A SAN allows “any-to-any” connection across the network, using interconnect elements such as routers, gateways, hubs, switches and directors. It eliminates the traditional dedicated connection between a server and storage, and the concept that the server effectively “owns and manages” the storage devices. It also eliminates any restriction to the amount of data that a server can access, currently limited by the number of storage devices attached to the individual server. Instead, a SAN introduces the flexibility of networking to enable one server or many heterogeneous servers to share a common storage utility, which may comprise many storage devices, including disk, tape, and optical storage. Additionally, the storage utility may be located far from the servers that use it.

The SAN can be viewed as an extension to the storage bus concept, which enables storage devices and servers to be interconnected using similar elements as in local area networks (LANs) and wide area networks (WANs): Routers, hubs, switches, directors, and gateways. A SAN can be shared between servers and/or dedicated to one server. It can be local, or can be extended over geographical distances.

The diagram in Figure 1-1 shows a tiered overview of a SAN connecting multiple servers to multiple storage systems.

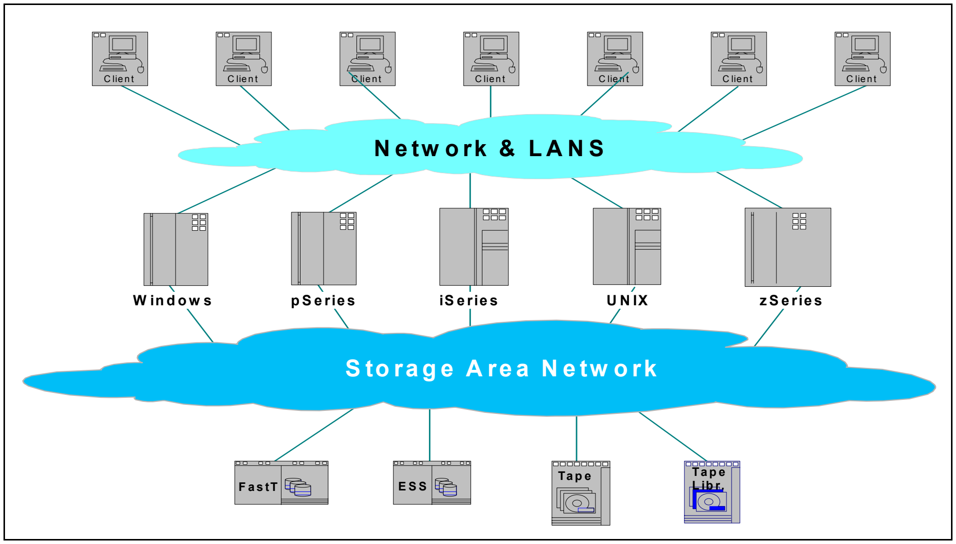


Figure 1-1 A SAN

SANs create new methods of attaching storage to servers. These new methods can enable great improvements in both availability and performance. Today's SANs are used to connect shared storage arrays and tape libraries to multiple servers, and are used by clustered servers for failover.

A SAN can be used to bypass traditional network bottlenecks. It facilitates direct, high-speed data transfers between servers and storage devices, potentially in any of the following three ways:

- Server to storage: This is the traditional model of interaction with storage devices. The advantage is that the same storage device may be accessed serially or concurrently by multiple servers.
- Server to server: A SAN may be used for high-speed, high-volume communications between servers.
- Storage to storage: This outboard data movement capability enables data to be moved without server intervention, thereby freeing up server processor cycles for other activities like application processing. Examples include a disk device backing up its data to a tape device without server intervention, or remote device mirroring across the SAN.

SANs allow applications that move data to perform better, for example, by having the data sent directly from the source to the target device with minimal server intervention. SANs also enable new network architectures where multiple hosts access multiple storage devices connected to the same network. Using a SAN can potentially offer the following benefits:

- Improvements to application availability: Storage is independent of applications and accessible through multiple data paths for better reliability, availability, and serviceability.
- Higher application performance: Storage processing is off-loaded from servers and moved onto a separate network.
- Centralized and consolidated storage: Simpler management, scalability, flexibility, and availability.
- Data transfer and vaulting to remote sites: Remote copy of data enabled for disaster protection and against malicious attacks.
- Simplified centralized management: Single image of storage media simplifies management.

1.2 SAN components

As stated previously, Fibre Channel is the predominant architecture upon which most SAN implementations are built, with FICON® as the standard protocol for z/OS® systems, and FCP as the standard protocol for open systems. The SAN components described in the following sections are Fibre Channel-based, and are shown in Figure 1-2 on page 5.

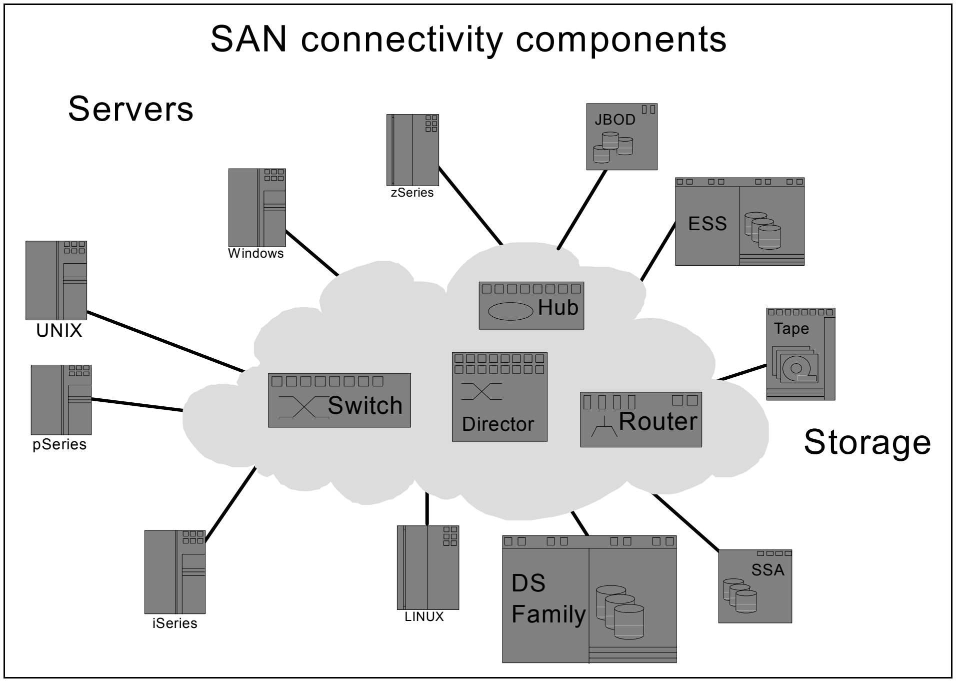


Figure 1-2 SAN components

1.2.1 SAN connectivity

The first element that must be considered in any SAN implementation is the connectivity of storage and server components typically using Fibre Channel. The components listed above have typically been used for LAN and WAN implementations. SANs, like LANs, interconnect the storage interfaces together into many network configurations and across longer distances.

Much of the terminology used for SAN has its origins in IP network terminology. In some cases, the industry and IBM use different terms that mean the same thing, and in some cases, mean different things.

1.2.2 SAN storage

The SAN liberates the storage device so it is not on a particular server bus, and attaches it directly to the network. In other words, storage is externalized and can be functionally distributed across the organization. The SAN also enables the centralization of storage devices and the clustering of servers, which has the potential to make for easier and less expensive centralized administration that lowers the total cost of ownership (TCO).

The storage infrastructure is the foundation on which information relies, and therefore must support a company's business objectives and business model. In this environment simply deploying more and faster storage devices is not enough. A SAN infrastructure provides enhanced network availability, data accessibility, and system manageability, and It is important to remember that a good SAN begins with a good design. This is not only a maxim, but must be a philosophy when we design or implement a SAN.

1.2.3 SAN servers

The server infrastructure is the underlying reason for all SAN solutions. This infrastructure includes a mix of server platforms such as Windows®, UNIX® (and its various flavors), and z/OS. With initiatives such as server consolidation and e-business, the need for SANs will increase, making the importance of storage in the network greater.

1.3 The importance of standards

Why do we care about standards? Standards are the starting point for the potential interoperability of devices and software from different vendors in the SAN marketplace. SNIA, among others, defined and ratified the standards for the SANs of today, and will keep defining the standards for tomorrow. All of the players in the SAN industry are using these standards now, as these are the basis for wide acceptance of SANs. Widely accepted standards potentially allow for heterogeneous, cross-platform, multivendor deployment of SAN solutions.

As all vendors have accepted these SAN standards, there *should* be no problem in connecting the different vendors into the same SAN network. However, nearly every vendor has an interoperability lab where it tests all kind of combinations between their products and those of other vendors. Some of the most important aspects in these tests are the reliability, error recovery, and performance. If a combination has passed the test, that vendor is going to certify or support this combination.

IBM participates in many industry standards organizations that work in the field of SANs. IBM believes that industry standards must be in place, and if necessary, re-defined for SANs to be a major part of the IT business mainstream.

Probably the most important industry standards organization for SANs is the Storage Networking Industry Association (SNIA). IBM is a founding member and board officer in SNIA. SNIA and other standards organizations and IBM participation are described in Appendix A, "SAN standards and organizations" on page 279.

1.4 Where are SANs heading?

Are SANs themselves evolving or are they likely to become extinct? Will they be overtaken by other technology? Certainly reports of the death of SANs have been greatly exaggerated. There has been far too much investment made for SANs to quietly lay down and go the way of the dinosaurs. There is no new “killer application” or technology in the immediate future that is threatening the SAN world. However, there is a gradual evolution that is beginning to pick up pace in the SAN world.

The evolution that is taking place is one of diversity. More and more we are seeing advances in technology find their way into the SAN chassis. What is quickly happening is that SANs are becoming multiprotocol capable. The industry recognizes that it is no longer acceptable to build a solution that will either create SAN islands (in much the same way as islands of information existed), or take an inordinate amount of cabling, support, power, and management.

Rather, the trend towards the simplification of the SAN infrastructure suddenly took a turn for the better. In a single footprint, multiple technologies that were once competing for floor space now happily sit alongside the “competition” in a single chassis. It is not uncommon to see FCIP, iFCP, and iSCSI together these days, and they are working together rather nicely. Most IT vendors also have virtualization solutions that present a single view of storage, and management solutions at the enterprise level. The SAN has quietly become an enabler for many technologies and protocols to share the same arena, without the somewhat tiresome arguments of which is “best.”

So, it is a case of evolution, not revolution, in the SAN world.

2

How, and why, can we use a SAN?

In the previous chapter, we introduced the basics by presenting a standard SAN definition, as well as a brief description of the underlying technologies and concepts that are behind a SAN implementation.

In this chapter, we extend this discussion by presenting real-life SAN issues, alongside well-known technologies and platforms used in SAN implementations. We also discuss some of the trends that are driving SAN evolution, and how they may affect the future of storage technology.

2.1 Why use a SAN?

In this section we describe the main motivators that drive SAN implementations, and present some of the key benefits that this technology might bring to data-dependent business.

2.1.1 The problem

As illustrated in Figure 2-1, the 1990's witnessed a major shift away from the traditional mainframe, host-centric model of computing to the client/server model. Today, many organizations have hundreds, even thousands, of distributed servers and client systems installed throughout its IT infrastructure. Many of these systems are powerful computers, with more processing capability than many mainframe computers had only a few years ago.

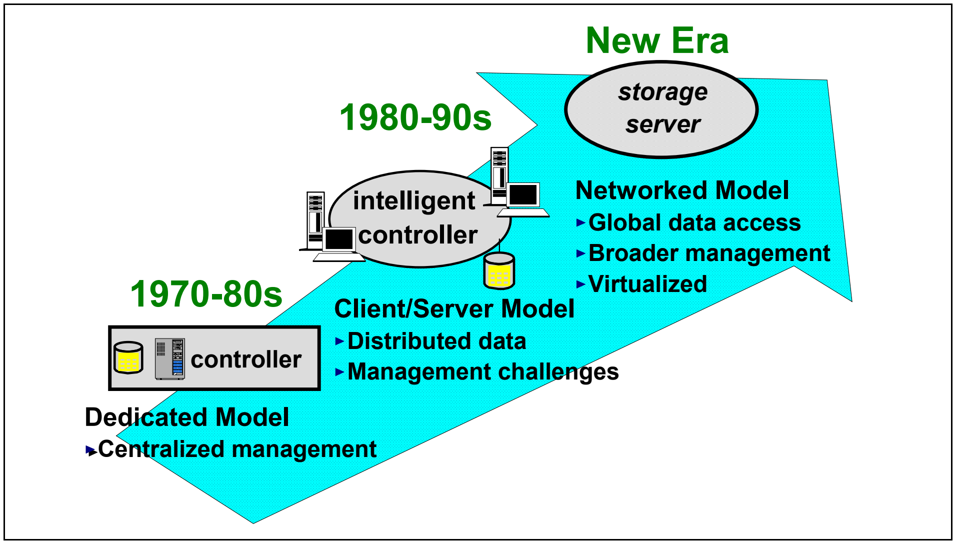


Figure 2-1 The evolution of storage architecture

Storage, for the most part, is directly connected by a dedicated channel to the server it supports. Frequently the servers are interconnected using local area networks (LAN) and wide area networks (WAN), to communicate and exchange data. The amount of disk storage capacity attached to such systems has grown exponentially in recent years. It is commonplace for a desktop personal computer or ThinkPad today to have storage in the order of tens of gigabytes. There has been a move to disk arrays, comprising a number of disk drives. The arrays may be "just a bunch of disks" (JBOD), or various implementations of redundant arrays of independent disks (RAID). The capacity of such arrays may be measured in tens or hundreds of gigabytes, but I/O bandwidth has not kept pace with the rapid growth in processor speeds and disk capacities.

Distributed clients and servers are frequently chosen to meet specific application needs. They may, therefore, run different operating systems (such as Windows NT®, UNIX of differing flavors, Novell NetWare, VMS, and so on), and different database software (for example, DB2®, Oracle, Informix®, SQL Server). Consequently, they have different file systems and different data formats.

Managing this multi-platform, multivendor, networked environment has become increasingly complex and costly. Multiple vendor's software tools, and appropriately skilled human resources must be maintained to handle data and storage resource management on the many differing systems in the enterprise. Surveys published by industry analysts consistently show that management costs associated with distributed storage are much greater, up to 10 times more, than the cost of managing consolidated or centralized storage. This includes costs of backup, recovery, space management, performance management, and disaster recovery planning.

Disk storage is often purchased from the processor vendor as an integral feature, and it is difficult to establish if the price you pay per gigabyte (GB) is competitive, compared to the market price of disk storage. Disks and tape drives, directly attached to one client or server, cannot be used by other systems, leading to inefficient use of hardware resources. Organizations often find that they have to purchase more storage capacity, even though free capacity is available in other platforms.

Additionally, it is difficult to scale capacity and performance to meet rapidly changing requirements, such as the explosive growth in e-business applications, and the need to manage information over its entire life cycle, from conception to intentional destruction.

Information stored on one system cannot readily be made available to other users, except by creating duplicate copies, and moving the copy to storage that is attached to another server. Movement of large files of data may result in significant degradation of performance of the LAN/WAN, causing conflicts with mission-critical applications. Multiple copies of the same data may lead to inconsistencies between one copy and another. Data spread on multiple small systems is difficult to coordinate and share for enterprise-wide applications, such as e-business, Enterprise Resource Planning (ERP), Data Warehouse, and Business Intelligence (BI).

Backup and recovery operations across a LAN may also cause serious disruption to normal application traffic. Even using fast Gigabit Ethernet transport, sustained throughput from a single server to tape is about 25 GB per hour. It would take approximately 12 hours to fully back up a relatively moderate departmental database of 300 GBs. This may exceed the available window of time in which this must be completed, and it may not be a practical solution if business operations span multiple time zones. It is increasingly evident to IT managers that these characteristics of client/server computing are too costly, and too inefficient. The islands of information resulting from the distributed model of computing do not match the needs of the enterprise.

New ways must be found to control costs, improve efficiency, and simplify the storage infrastructure to meet the requirements of the modern business world.

2.1.2 The requirements

With this scenario in mind, we can think of a number of requirements that today's storage infrastructures should meet. Some of the most important are:

- Unlimited and just-in-time scalability. Businesses require the capability to flexibly adapt to rapidly changing demands for storage resources without performance degradation.
- System simplification. Businesses require an easy-to-implement infrastructure with the minimum of management and maintenance. The more complex the enterprise environment, the more costs are involved in terms of management. Simplifying the infrastructure can save costs and provide a greater return on investment (ROI).
- Flexible and heterogeneous connectivity. The storage resource must be able to support whatever platforms are within the IT environment. This is essentially an investment protection requirement that allows you to configure a storage resource for one set of systems, and subsequently configure part of the capacity to other systems on an as-needed basis.
- Security. This requirement guarantees that data from one application or system does not become overlaid or corrupted by other applications or systems. Authorization also requires the ability to fence off one system's data from other systems.
- Availability. This is a requirement that implies both protection against media failure as well as ease of data migration between devices, without interrupting application processing. This certainly implies improvements to backup and recovery processes: attaching disk and tape devices to the same networked infrastructure allows for fast data movement between devices, which provides enhanced backup and recovery capabilities, such as:
  - Serverless backup. This is the ability to back up your data without using the computing processor of your servers.
  - Synchronous copy. This makes sure your data is at two or more places before your application goes to the next step.
  - Asynchronous copy. This makes sure your data is at two or more places within a short time. It is the disk subsystem that controls the data flow.

In the next section, we discuss the use of SANs as a response to these business requirements.

2.2 How can we use a SAN?

The key benefits that a SAN might bring to a highly data-dependent business infrastructure can be summarized into three rather simple concepts: infrastructure simplification, information lifecycle management and business continuity. They are an effective response to the requirements presented in the previous section, and are strong arguments for the adoption of SANs.

These three concepts are briefly described as follows.

2.2.1 Infrastructure simplification

There are four main methods by which infrastructure simplification can be achieved: ***consolidation, virtualization, automation*** and ***integration***:

- Consolidation

  Concentrating systems and resources into locations with fewer, but more powerful, servers and storage pools can help increase IT efficiency and simplify the infrastructure. Additionally, centralized storage management tools can help improve scalability, availability, and disaster tolerance.

- Virtualization

  Storage virtualization helps in making complexity nearly transparent and at the same time can offer a composite view of storage assets. This may help reduce capital and administrative costs, while giving users better service and availability. Virtualization is designed to help make the IT infrastructure more responsive, scalable, and available.

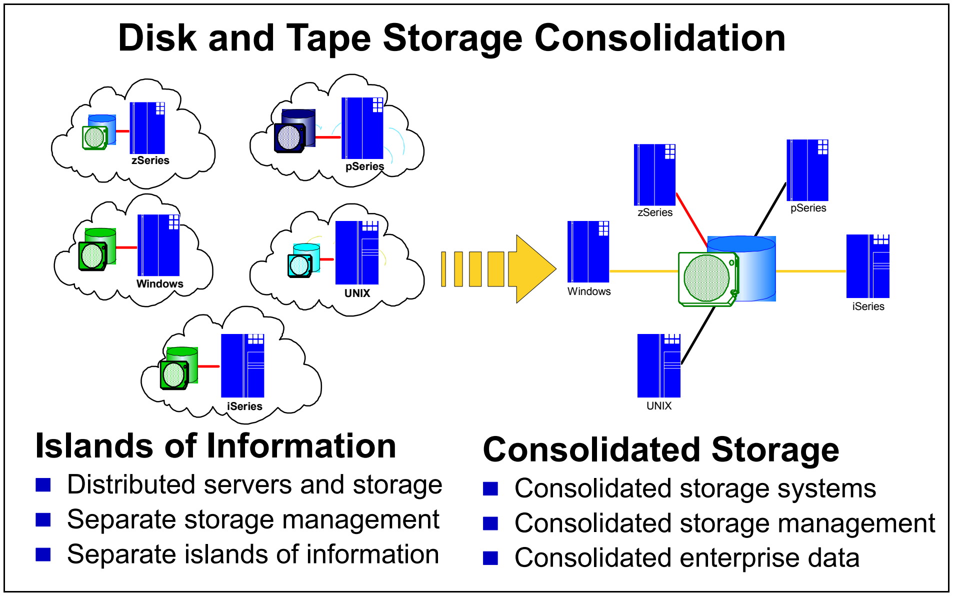
- Automation

  Choosing storage components with autonomic capabilities can improve availability and responsiveness—and help protect data as storage needs grow. As soon as day-to-day tasks are automated, storage administrators may be able to spend more time on critical, higher-level tasks unique to a company's business mission.

- Integration

  Integrated storage environments simplify system management tasks and improve security. When all servers have secure access to all data, your infrastructure may be better able to respond to your users information needs.

Figure 2-2 illustrates the consolidation movement from the distributed islands of information toward a single, and, most importantly, simplified infrastructure.



Disk and Tape Storage Consolidation

Islands of Information

- Distributed servers and storage
- Separate storage management
- Separate islands of information

Consolidated Storage

- Consolidated storage systems
- Consolidated storage management
- Consolidated enterprise data

Figure 2-2 Disk and tape storage consolidation

Simplified storage environments have fewer elements to manage, which leads to increased resource utilization, simplifies storage management, and can provide economies of scale for owning disk storage servers. These environments can be more resilient and provide an infrastructure for virtualization and automation.

2.2.2 Information lifecycle management

Information has become an increasingly valuable asset, but as the amount of information grows, it becomes increasingly costly and complex to store and manage it. Information lifecycle management (ILM) is a process for managing information through its life cycle, from conception until intentional disposal, in a manner that optimizes storage, and maintains a high level of access at the lowest cost.

A SAN implementation makes it easier to manage the information lifecycle as it integrates applications and data into a single-view system, in which information resides, and can be managed more efficiently.

2.2.3 Business continuity

It goes without saying that the business climate in today's on demand era is highly competitive. Customers, employees, suppliers, and business partners expect to be able to tap into their information at any hour of the day from any corner of the globe. Continuous business operations are no longer optional—they are a business imperative to becoming successful, and maintaining a competitive advantage. Businesses must also be increasingly sensitive to issues of customer privacy and data security, so that vital information assets are not compromised. Factor in those legal and regulatory requirements, and the inherent demands of participating in the global economy, and accountability, and all of a sudden the lot of an IT manager is not a happy one.

It is little wonder that a sound and comprehensive business continuity strategy has become a business imperative, and SANs play a key role in this. By deploying a consistent and safe infrastructure, they make it possible to meet any availability requirements.

2.3 Using the SAN components

The foundation that a SAN is built on is the interconnection of storage devices and servers. This section further discusses storage, interconnection components, and servers, and how different types of servers and storage are used in a typical SAN environment.

2.3.1 Storage

This section briefly describes the main types of storage devices that can be found in the market.

Disk systems

In brief a disk system is a device in which a number of physical storage disks sit side-by-side. By being contained within a single “box”, a disk system usually has a central control unit that manages all the I/O, simplifying the integration of the system with other devices, such as other disk systems or servers.

Depending on the “intelligence” with which this central control unit is able to manage the individual disks, a disk system can be a JBOD or a RAID.

Just A Bunch Of Disks (JBOD)

In this case, the disk system appears as a set of individual storage devices to the device they are attached to. The central control unit provides only basic functionality for writing and reading data from the disks.

Redundant Array of Independent Disks (RAID)

In this case, the central control unit provides additional functionality that makes it possible to utilize the individual disks in such a way to achieve higher fault-tolerance and/or performance. The disks themselves appear as a single storage unit to the devices to which they are connected.

Depending on the specific functionality offered by a particular disk system, it is possible to make it behave as a RAID and/or a JBOD; the decision as to which type of disk system is more suitable for a SAN implementation strongly depends on the performance and availability requirements for this particular SAN.

Tape systems

Tape systems, in much the same way as disk systems do, are devices that comprise all the necessary apparatus to manage the use of tapes for storage purposes. In this case, however, the serial nature of a tape makes it impossible for them to be treated in parallel, as RAID devices are leading to a somewhat simpler architecture to manage and use.

There are basically three types of systems: drives, autoloaders and libraries, that are described as follows.

Tape drives

As with disk drives, tape drives are the means by which tapes can be connected to other devices; they provide the physical and logical structure for reading from, and writing to tapes.

Tape autoloaders

Tape autoloaders are autonomous tape drives capable of managing tapes and performing automatic back-up operations. They are usually connected to high-throughput devices that require constant data back-up.

Tape libraries

Tape libraries are devices capable of managing multiple tapes simultaneously and, as such, can be viewed as a set of independent tape drives or autoloaders. They are usually deployed in systems that require massive storage capacity, or that need some kind of data separation that would result in multiple single-tape systems. As a tape is not a random-access media, tape libraries cannot provide parallel access to multiple tapes as a way to improve performance, but they can provide redundancy as a way to improve data availability and fault-tolerance.

Once more, the circumstances under which each of these systems, or even a disk system, should be used, strongly depend on the specific requirements that a particular SAN implementation has. However, we can say that disk systems are usually used for online storage due to their superior performance, whereas tape systems are ideal for offline, high-throughput storage, due to the lower cost of storage per byte.

In the next section we describe the prevalent connectivity interfaces, protocols and services for building a SAN.

2.3.2 SAN connectivity

SAN connectivity comprises all sorts of hardware and software components that make possible the interconnection of storage devices and servers. In this section, we have divided these components into three sections according to the level abstraction to which they belong: lower level layers, middle level layers, and higher level layers.

> **Note:** With respect to data throughput speeds, in this redbook we use the following representations:
>
> - 1 Gbps = 100 MBps
> - 2 Gbps = 200 MBps
> - 4 Gbps = 400 MBps
> - 8 Gbps = 800 MBps
> - 10 Gbps = 1000 MBps

Lower level layers

This section comprises the physical data-link, and the network layers of connectivity.

Ethernet interface

Ethernet adapters are typically used on conventional server-to-server or workstation-to-server network connections. They build up a common-bus topology by which every attached device can communicate with each other, using this common-bus for such. An Ethernet adapter can reach up to 10 Gbps of data transferred.

Fibre Channel

Fibre Channel (FC) is a serial interface (usually implemented with fiber-optic cable, and is the primary architecture for the vast majority of SANs. To support this there are many vendors in the marketplace producing Fibre Channel adapters, and other FC devices. One of the reasons that FC is so popular is that it allows the maximum SCSI cable length of 25 meters restriction to be overcome. Coupled with the increased speed that it supports, it quickly became the connection of choice.

SCSI

The Small Computer System Interface (SCSI) is a parallel interface. SCSI devices are connected to form a terminated bus (the bus is terminated using a terminator). The maximum cable length is 25 meters, and a maximum of 16 devices can be connected to a single SCSI bus. The SCSI interface has many configuration options for error handling and supports both disconnect and reconnect to devices and multiple initiator requests. Usually, a host computer is an initiator. Multiple initiator support allows multiple hosts to attach to the same devices and is used in support of clustered configurations. The Ultra3 SCSI adapter today can have a data transfer up to 160 MBps.

Middle level layers

This section comprises the transport protocol and session layers.

FCP

The Fibre Channel Protocol (FCP) is the interface protocol of SCSI on Fibre Channel. It is a gigabit speed network technology primarily used for Storage Networking. Fibre Channel is standardized in the T11 Technical Committee of the InterNational Committee for Information Technology Standards (INCITS), an American National Standard Institute (ANSI) accredited standards committee. It started for use primarily in the supercomputer field, but has become the standard connection type for storage area networks in enterprise storage. Despite its name, Fibre Channel signaling can run on both twisted-pair copper wire and fiber optic cables.

iSCSI

Internet SCSI (iSCSI) is a transport protocol that carries SCSI commands from an initiator to a target. It is a data storage networking protocol that transports standard Small Computer System Interface (SCSI) requests over the standard Transmission Control Protocol/Internet Protocol (TCP/IP) networking technology.

iSCSI enables the implementation of IP-based storage area networks (SANs), enabling customers to use the same networking technologies — for both storage and data networks. As it uses TCP/IP, iSCSI is also well suited to run over almost any physical network. By eliminating the need for a second network technology just for storage, iSCSI has the potential to lower the costs of deploying networked storage.

FCIP

Fibre Channel over IP (FCIP) is also known as Fibre Channel tunneling or storage tunneling. It is a method to allow the transmission of Fibre Channel information to be tunnelled through the IP network. Because most organizations already have an existing IP infrastructure, the attraction of being able to link geographically dispersed SANs, at a relatively low cost, is enormous.

FCIP encapsulates Fibre Channel block data and subsequently transports it over a TCP socket. TCP/IP services are utilized to establish connectivity between remote SANs. Any congestion control and management, as well as data error and data loss recovery, is handled by TCP/IP services, and does not affect FC fabric services.

The major point with FCIP is that is does not replace FC with IP, it simply allows deployments of FC fabrics using IP tunnelling. The assumption that this might lead to is that the “industry” has decided that FC-based SANs are more than appropriate, and that the only need for the IP connection is to facilitate any distance requirement that is beyond the current scope of an FCP SAN.

iFCP

Internet Fibre Channel Protocol (iFCP) is a mechanism for transmitting data to and from Fibre Channel storage devices in a SAN, or on the Internet using TCP/IP.

iFCP gives the ability to incorporate already existing SCSI and Fibre Channel networks into the Internet. iFCP is able to be used in tandem with existing Fibre Channel protocols, such as FCIP, or it can replace them. Whereas FCIP is a tunneled solution, iFCP is an FCP routed solution.

The appeal of iFCP is that for customers that have a wide range of FC devices, and who want to be able to connect these using the IP network, iFCP gives the ability to permit this. iFCP can interconnect FC SANs with IP networks, and also allows customers to use the TCP/IP network in place of the SAN.

iFCP is a gateway-to-gateway protocol, and does not simply encapsulate FC block data. Gateway devices are used as the medium between the FC initiators and targets. As these gateways can either replace or be used in tandem with existing FC fabrics, iFCP could be used to help migration from a Fibre Channel SAN to an IP SAN, or allow a combination of both.

FICON

FICON architecture is an enhancement of, rather than a replacement for, the now relatively old ESCON® architecture. As a SAN is Fibre Channel based, FICON is a prerequisite for z/OS systems to fully participate in a heterogeneous SAN, where the SAN switch devices allow the mixture of open systems and mainframe traffic.

FICON is a protocol that uses Fibre Channel as its physical medium. FICON channels are capable of data rates up to 200 MBps full duplex, they extend the channel distance (up to 100 km), increase the number of control unit images per link, increase the number of device addresses per control unit link, and retain the topology and switch management characteristics of ESCON.

Higher level layers

This section comprises of the presentation and application layers.

Server-attached storage

The earliest approach was to tightly couple the storage device with the server. This server-attached storage approach keeps performance overhead to a minimum. Storage is attached directly to the server bus using an adapter card, and the storage device is dedicated to a single server. The server itself controls the I/O to the device, issues the low-level device commands, and monitors device responses.

Initially, disk and tape storage devices had no on-board intelligence. They just executed the server's I/O requests. Subsequent evolution led to the introduction of control units. Control units are storage off-load servers that contain a limited level of intelligence, and are able to perform functions, such as I/O request caching for performance improvements, or dual copy of data (RAID 1) for availability. Many advanced storage functions have been developed and implemented inside the control unit.

Network Attached Storage

Network Attached Storage (NAS) is basically a LAN-attached file server that serves files using a network protocol such as Network File System (NFS). NAS is a term used to refer to storage elements that connect to a network and provide file access services to computer systems. A NAS storage element consists of an engine that implements the file services (using access protocols such as NFS or CIFS), and one or more devices, on which data is stored. NAS elements may be attached to any type of network. From a SAN perspective, a SAN-attached NAS engine is treated just like any other server, but a NAS does not provide any of the activities that a server in a server-centric system typically provides, such as e-mail, authentication, or file management.

NAS allows more hard disk storage space to be added to a network that already utilizes servers without shutting them down for maintenance and upgrades. With a NAS device, storage is not an integral part of the server. Instead, in this storage-centric design, the server still handles all of the processing of data, but a NAS device delivers the data to the user. A NAS device does not need to be located within the server but can exist anywhere in the LAN and can be made up of multiple networked NAS devices. These units communicate to a host using Ethernet and file-based protocols. This is in contrast to the disk units discussed earlier, which use Fibre Channel protocol and block-based protocols to communicate.

NAS storage provides acceptable performance and security, and it is often less expensive for servers to implement (for example, ethernet adapters are less expensive than Fibre Channel adapters).

In an effort to bridge the two worlds and to open up new configuration options for customers, some vendors, including IBM, sell NAS units that act as a gateway between IP-based users and SAN-attached storage. This allows for the connection of the storage device of choice (an ESS, for example) and share it between your high-performance database servers (attached directly through Fibre Channel) and your end users (attached through IP) who do not have performance requirements nearly as strict.

NAS is an ideal solution for serving files stored on the SAN to end users in cases where it would be impractical and expensive to equip end users with Fibre Channel adapters. NAS allows those users to access your storage through the IP-based network that they already have.

2.3.3 Servers

Each of the different server platforms (IBM @server zSeries®, UNIX, AIX®, HP, Sun, Linux, and others), OS/400®, and Windows (PC Servers) have implemented SAN solutions using various interconnects and storage technologies. The following sections review these solutions and the different implementations on each of the platforms.

Mainframe servers

In simple terms, a mainframe is a single, monolithic and possibly multi-processor high-performance computer system. Apart from the fact that IT evolution has been pointing toward a more distributed and loosely coupled infrastructure, mainframes still play an important role on businesses that depend on massive storage capabilities.

The IBM @server zSeries (formerly known as S/390®) is a processor(s) and operating system mainframe set. Historically, zSeries servers have supported many different operating systems, such as z/OS, OS/390®, VM, VSE, and TPF, which have been enhanced over the years. The processor to storage device interconnection has also evolved from a bus and tag interface to ESCON channels, and now to FICON channels. Figure 2-3 on page 22 shows the various processor-to-storage interfaces.

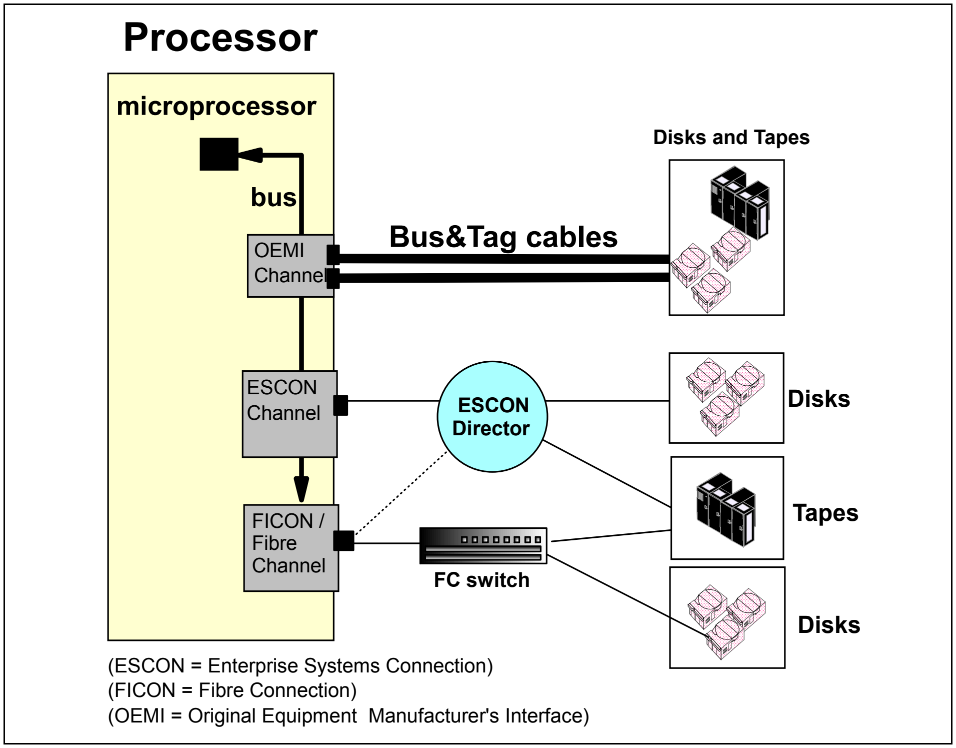


Figure 2-3 Processor-to-storage interface connections

Due to architectural differences, and extremely strict data integrity and management requirements, the implementation of FICON has been somewhat behind that of FCP on open systems. However, at the time of writing, FICON has now caught up with FCP SANs, and they coexist quite amicably.

For the latest news on zSeries TotalStorage products, refer to:

http://www-1.ibm.com/servers/storage/product/products_zseries.html

For the latest news on zSeries FICON connectivity, refer to:

http://www-1.ibm.com/servers/eserver/zseries/connectivity/

In addition to FICON for traditional zSeries operating systems, IBM has standard Fibre Channel adapters for use with zSeries servers that can implement Linux®.

UNIX-based servers

Originally designed for high-performance computer systems, such as mainframes, the UNIX operating systems is today present on a great variety of hardware platforms, ranging from Linux-based PCs to dedicated large-scale stations. Due to its popularity and maturity, it also plays an important role on both existing and legacy IT infrastructures.

The IBM @server pSeries® line of servers, running a UNIX operating system called AIX, offers various processor to storage interfaces, including SCSI, SSA, and Fibre Channel. The SSA interconnection has primarily been used for disk storage. Fibre Channel adapters are able to connect to tape and disk. Figure 2-4 shows the various processor-to-storage interconnect options for the pSeries family.

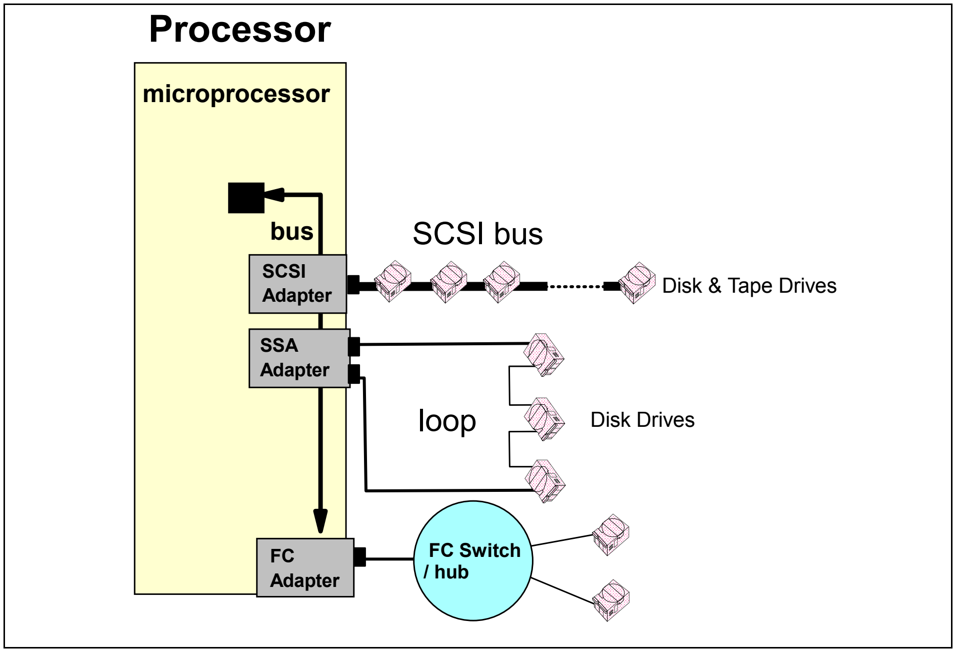


Figure 2-4 pSeries processor-to-storage interconnections

The various UNIX system vendors in the market deploy different variants of the UNIX operating system, each having some unique enhancements, and often supporting different file systems such as the Journal File System (JFS), Enhanced Journal File System (JFS2), and the Andrew File System (AFS®). The server-to-storage interconnect is similar to pSeries, as shown in Figure 2-4.

For the latest news on pSeries TotalStorage products, refer to:

http://www-1.ibm.com/servers/storage/product/products_pseries.html

Windows-based servers

Based on the reports of various analysts regarding growth in the Windows server market (both in the number and size of Windows servers), Windows will become the largest market for SAN solution deployment. More and more Windows servers will host mission-critical applications that will benefit from SAN solutions, such as disk and tape pooling, tape sharing, multipathing, and remote copy.

The processor-to-storage interfaces on xSeries® servers (IBM Intel®-based processors that support the Microsoft® Windows operating system) are similar to those supported on UNIX servers, including SCSI and Fibre Channel.

For more information, see the xSeries SAN Web site at:

http://www-1.ibm.com/servers/storage/product/products_xseries.html

Other servers

The iSeries™ system architecture is defined by a high-level machine interface, referred to as Technology Independent Machine Interface (TIMI), which isolates applications (and much of the operating system) from the actual underlying systems hardware.

The main processor and the I/O processors are linked using a system bus, including Systems Product Division (SPD), and also Peripheral Component Interconnect (PCI). Figure 2-5 summarizes the various modules of an iSeries hardware architecture.

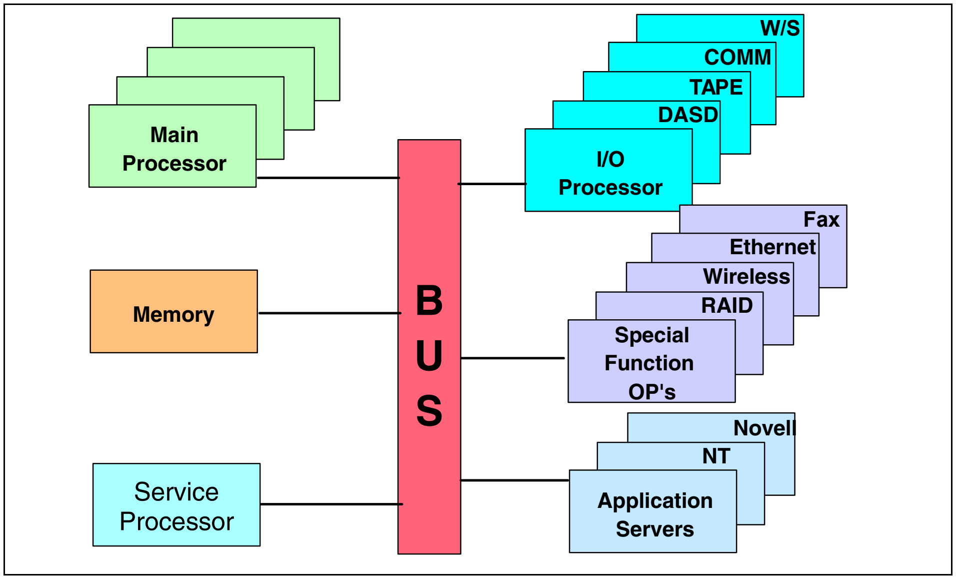


Figure 2-5 iSeries hardware design

Several architectural features of the iSeries server distinguish this system from other machines in the industry. These features include:

- Technology Independent Machine Interface
- Object-based systems design
- Single-level storage
- Integration of application programs into the operating system

For the latest news on iSeries TotalStorage products, refer to:

http://www-1.ibm.com/servers/storage/product/products_iseries.html

Single-level storage

Single-level storage (SLS) is probably the most significant differentiator in a SAN solution implementation on an iSeries server, as compared to other systems such as z/OS, UNIX, and Windows. In OS/400, both the main storage (memory) and the secondary storage (disks) are treated as a very large virtual address space known as SLS.

Figure 2-6 compares the OS/400 SLS addressing with the way Windows or UNIX systems work, using the processor local storage. With 32-bit addressing, each process (job) has 4 GB of addressable memory. With 64-bit SLS addressing, over 18 million terabytes (18 exabytes) of addressable storage is possible. Because a single page table maps all virtual addresses to physical addresses, task switching is very efficient. SLS further eliminates the need for address translation, thus speeding up data access.

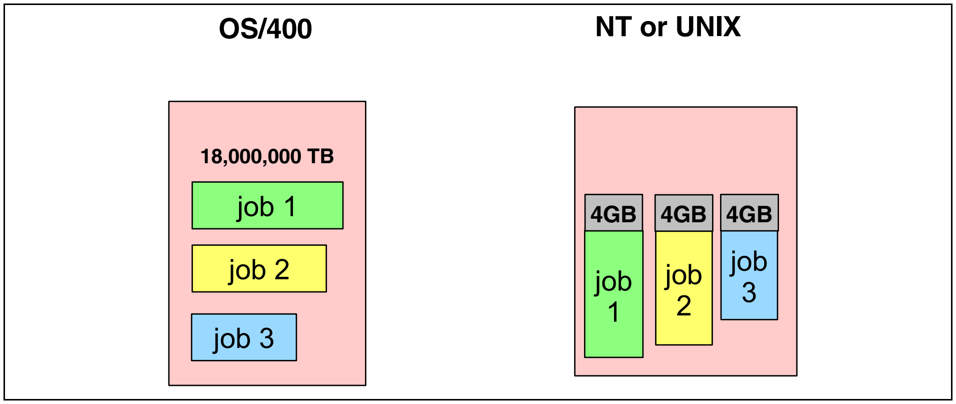


Figure 2-6 OS/400 versus NT or UNIX storage addressing

iSeries SAN support has rapidly expanded, and iSeries servers now support attachment to switched fabrics, and to most of IBM SAN-attached storage products.

For more information, see the iSeries SAN Web site:

http://www.ibm.com/servers/eserver/iseries/hardware/storage/san.html

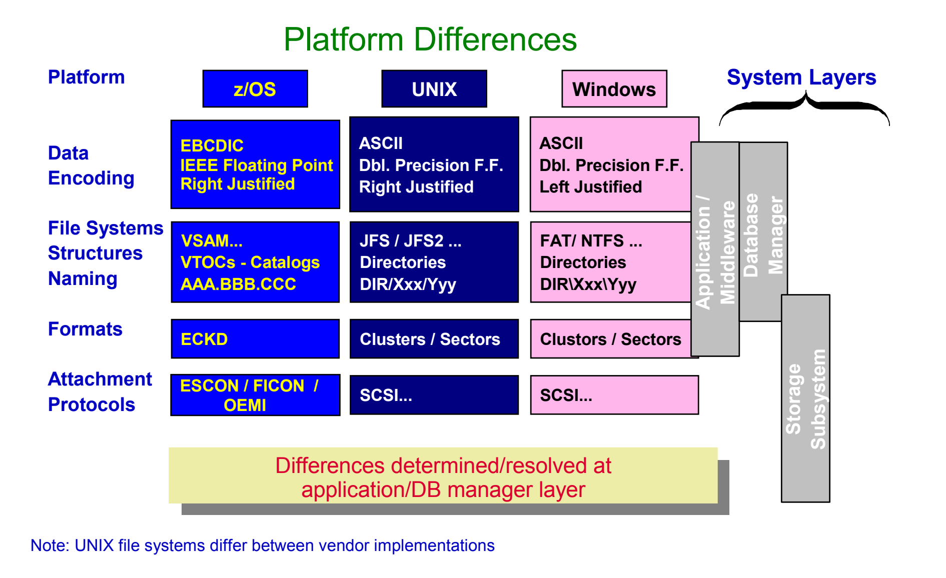
2.3.4 Putting the components together

After going through this myriad of technologies and platforms, we can easily understand why it is a challenge to implement true heterogeneous storage and data environments across different hardware and operating systems platforms; for example, disk and tape sharing across z/OS, OS/400, UNIX, and Windows.

One of the SAN principles, which is infrastructure simplification, cannot be easily achieved; each platform, along with its operating system, treats data differently at various levels in the system architecture, thus creating some of these many challenges:

- Different attachment interfaces and protocols, such as SCSI, ESCON and FICON.
- Different data formats, such as Extended Count Key Data (ECKD™), blocks, clusters, and sectors.
- Different file systems, such as Virtual Storage Access Method (VSAM), Journal File System (JFS), Enhanced Journal File System (JFS2), Andrew File System (AFS), and Windows NT File System (NTFS).
- OS/400, with the concept of single-level storage.
- Different file system structures, such as catalogs and directories.
- Different file naming conventions, such as AAA.BBB.CCC and DIR/Xxx/Yyy.
- Different data encoding techniques, such as EBCDIC, ASCII, floating point, and little or big endian.

In Figure 2-7 on page 27 is a brief summary of these differences for several different systems.



Platform Differences

Platform	z/OS	UNIX	Windows
Data Encoding	EBCDIC IEEE Floating Point Right Justified	ASCII Dbl. Precision F.F. Right Justified	ASCII Dbl. Precision F.F. Left Justified
File Systems Structures Naming	VSAM... VTOCs - Catalogs AAA.BBB.CCC	JFS / JFS2 ... Directories DIR/Xxx/Yyy	FAT/ NTFS ... Directories DIR\Xxx\Yyy
Formats	ECKD	Clusters / Sectors	Clustors / Sectors
Attachment Protocols	ESCON / FICON / OEMI	SCSI...	SCSI...

Differences determined/resolved at application/DB manager layer

Note: UNIX file systems differ between vendor implementations

Figure 2-7 Hardware and operating systems differences

3

Fibre Channel internals

Fibre Channel is the predominant architecture upon which SAN implementations are built. Fibre Channel is a technology standard that allows data to be transferred from one network node to another at extremely high speeds. Current implementations support data transfers at up to 10 Gbps or even more. The Fibre Channel standard is accredited by many standards bodies, technical associations, vendors, and industry-wide consortiums. There are many products on the market that take advantage of FC's high-speed, high-availability characteristics.

Fibre Channel was completely developed through industry cooperation, unlike SCSI, which was developed by a vendor, and submitted for standardization afterwards.

Fibre or Fiber?: Fibre Channel was originally designed to support fiber optic cabling only. When copper support was added, the committee decided to keep the name in principle, but to use the UK English spelling (Fibre) when referring to the standard. We retain the US English spelling when referring generically to fiber optics and cabling.

Some people refer to Fibre Channel architecture as the Fibre version of SCSI. Fibre Channel is an architecture used to carry IPI traffic, IP traffic, FICON traffic, FCP (SCSI) traffic, and possibly traffic using other protocols, all on the standard FC transport. An analogy could be Ethernet, where IP, NetBIOS, and SNA are all used simultaneously over a single Ethernet adapter, since these are all protocols with mappings to Ethernet. Similarly, there are many protocols mapped onto FC.

FICON is the standard protocol for z/OS, and will replace all ESCON environments over time. FCP is the standard protocol for open systems, both using Fibre Channel architecture to carry the traffic.

3.1 Firstly, why the Fibre Channel architecture?

Before we delve into the internals of Fibre Channel we will describe why Fibre Channel became the predominant SAN architecture.

3.1.1 The SCSI legacy

The Small Computer Systems Interface (SCSI) is the conventional, server centric method of connecting peripheral devices (disks, tapes and printers) in the open client/server environment. As its name indicates, it was designed for the PC and small computer environment. It is a bus architecture, with dedicated, parallel cabling between the host and storage devices, such as disk arrays. This is similar in implementation to the Original Equipment Manufacturer's Information (OEMI) bus and tag interface commonly used by mainframe computers until the early 1990's. SCSI shares a practical aspect with bus and tag, in that cables and connectors are bulky, relatively expensive, and are prone to failure.

The amount of data available to the server is determined by the number of devices which can attach to the bus, and by the number of buses attached to the server. Up to 15 devices can be attached to a server on a single SCSI bus. In practice, because of performance limitations due to arbitration, it is common for no more than four or five devices to be attached in this way, thus limiting capacity scalability.

Access to data is lost in the event of a failure of any of the SCSI connections to the disks. This also applies in the event of reconfiguration or servicing of a disk device attached to the SCSI bus, because all the devices in the string must be taken offline. In today's environment, when many applications need to be available continuously, this downtime is unacceptable.

The data rate of the SCSI bus is determined by the number of bits transferred, and the bus cycle time (measured in megahertz (MHz)). Decreasing the cycle time increases the transfer rate, but, due to limitations inherent in the bus architecture, it may also reduce the distance over which the data can be successfully transferred. The physical transport was originally a parallel cable comprising eight data lines, to transmit eight bits in parallel, plus control lines.

Later implementations widened the parallel data transfers to 16 bit paths (SCSI Wide), to achieve higher bandwidths.

Propagation delays in sending data in parallel along multiple lines lead to a well known phenomenon known as skew, meaning that all bits may not arrive at the target device at the same time. This is shown in Figure 3-1.

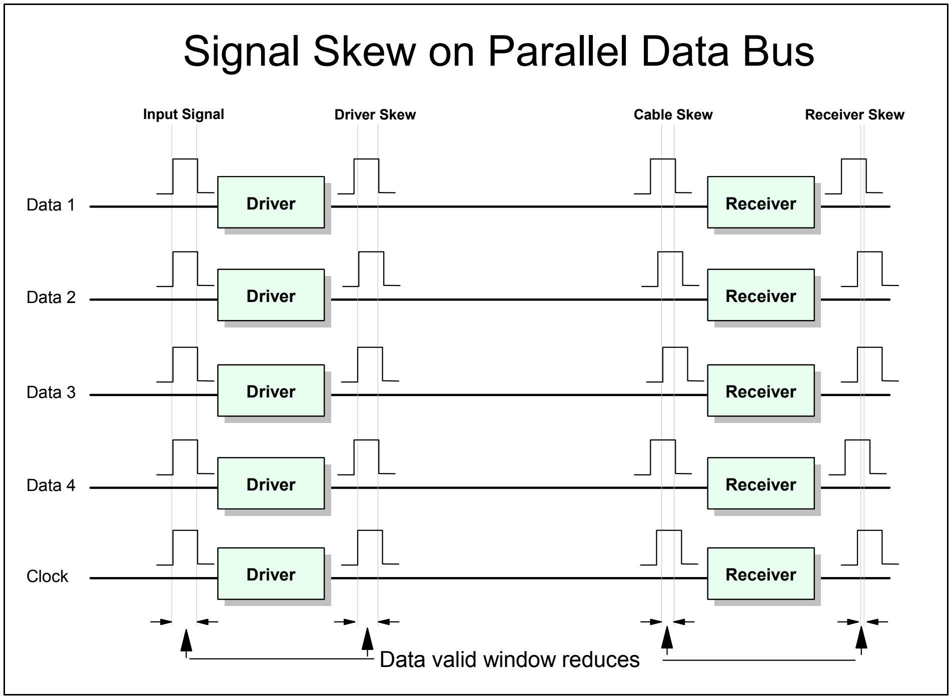


Figure 3-1 SCSI Propagation delay results in skew

Arrival occurs during a small window of time, depending on the transmission speed, and the physical length of the SCSI bus. The need to minimize the skew limits the distance that devices can be positioned away from the initiating server to between 2 to 25 meters, depending on the cycle time. Faster speed means shorter distance. The distances refer to the maximum length of the SCSI bus, including all attached devices. The SCSI distance limitations are shown in Figure 3-2 on page 32. These distance limitations may severely restrict the total GB capacity of the disk storage which can be attached to an individual server.

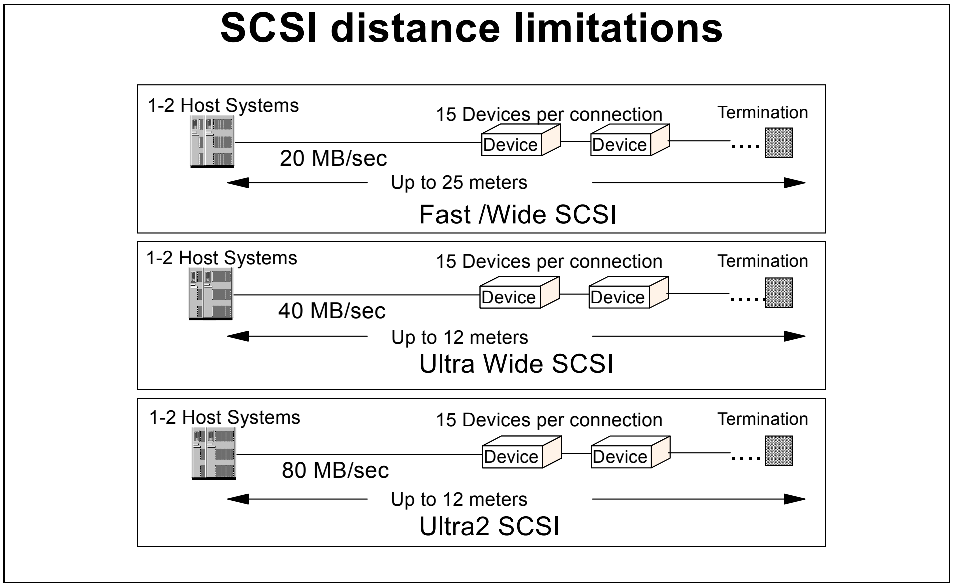


Figure 3-2 SCSI bus distance limitations

Many applications require the system to access several devices, or for several systems to share a single device. SCSI can enable this by attaching multiple servers or devices to the same bus. This is known as a multi-drop configuration. A multi-drop configuration is shown in Figure 3-3.

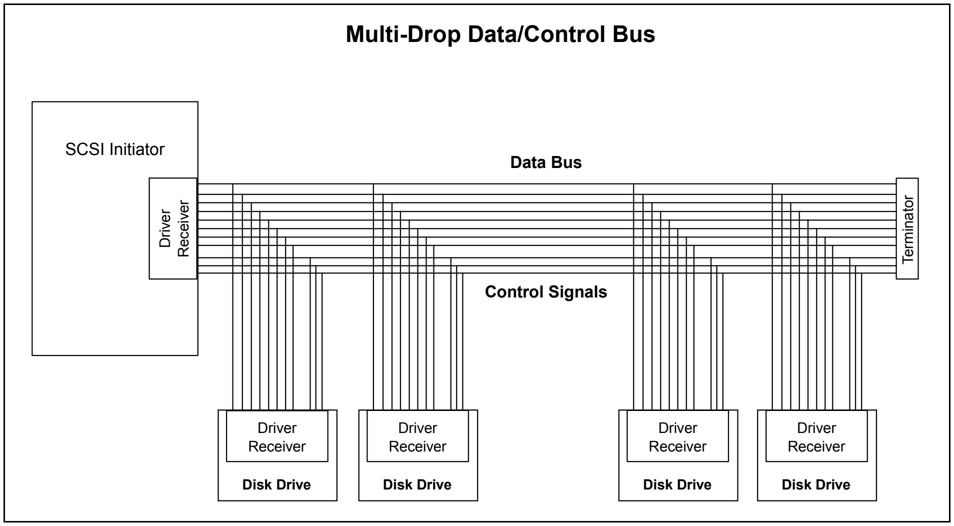


Figure 3-3 Multi-drop bus structure

To avoid signal interference, and therefore possible data corruption, all unused ports on a parallel SCSI bus must be properly terminated. Incorrect termination can result in transaction errors or failures.

Normally, only a single server can access data on a specific disk by means of a SCSI bus. In a shared bus environment, it is clear that all devices cannot transfer data at the same time. SCSI uses an arbitration protocol to determine which device can gain access to the bus. Arbitration occurs before and after every data transfer on the bus. While arbitration takes place, no data movement can occur. This represents an additional overhead which reduces bandwidth utilization, substantially reducing the effective data rate achievable on the bus. Actual rates are typically less than 50% of the rated speed of the SCSI bus.

In addition to being a physical transport, SCSI is also a protocol, which specifies commands and controls for sending blocks of data between the host and the attached devices. SCSI commands are issued by the host operating system, in response to user requests for data. Some operating systems, for example, Windows NT, treat all attached peripherals as SCSI devices, and issue SCSI commands to deal with all read and write operations.

It is clear that the physical parallel SCSI bus architecture has a number of significant speed, distance, and availability limitations, which make it increasingly less suitable for many applications in today's networked IT infrastructure. However, since the SCSI protocol is deeply embedded in the way that commonly encountered operating systems handle user requests for data, it would be a major inhibitor to progress if we were obliged to move to new protocols.

3.1.2 Why Fibre Channel?

Fibre Channel is an open, technical standard for networking which incorporates the "channel transport" characteristics of an I/O bus, with the flexible connectivity and distance characteristics of a traditional network.

Because of its channel-like qualities, hosts and applications see storage devices attached to the SAN as though they are locally attached storage. Because of its network characteristics it can support multiple protocols and a broad range of devices, and it can be managed as a network. Fibre Channel can use either optical fiber (for distance) or copper cable links (for short distance at low cost).

Fibre Channel is a multi-layered network, based on a series of American National Standards Institute (ANSI) standards which define characteristics and functions for moving data across the network. These include definitions of physical interfaces, such as cabling, distances and signaling; data encoding and link controls; data delivery in terms of frames, flow control and classes of service; common services; and protocol interfaces.

Like other networks, information is sent in structured packets or frames, and data is serialized before transmission. But, unlike other networks, the Fibre Channel architecture includes a significant amount of hardware processing to deliver high performance.

Fibre Channel uses a serial data transport scheme, similar to other computer networks, streaming packets, (frames) of bits one behind the other in a single data line to achieve high data rates.

Serial transfer by its very nature, of course, does not suffer from the problem of skew, so speed and distance is not restricted as with parallel data transfers as we show in Figure 3-4.

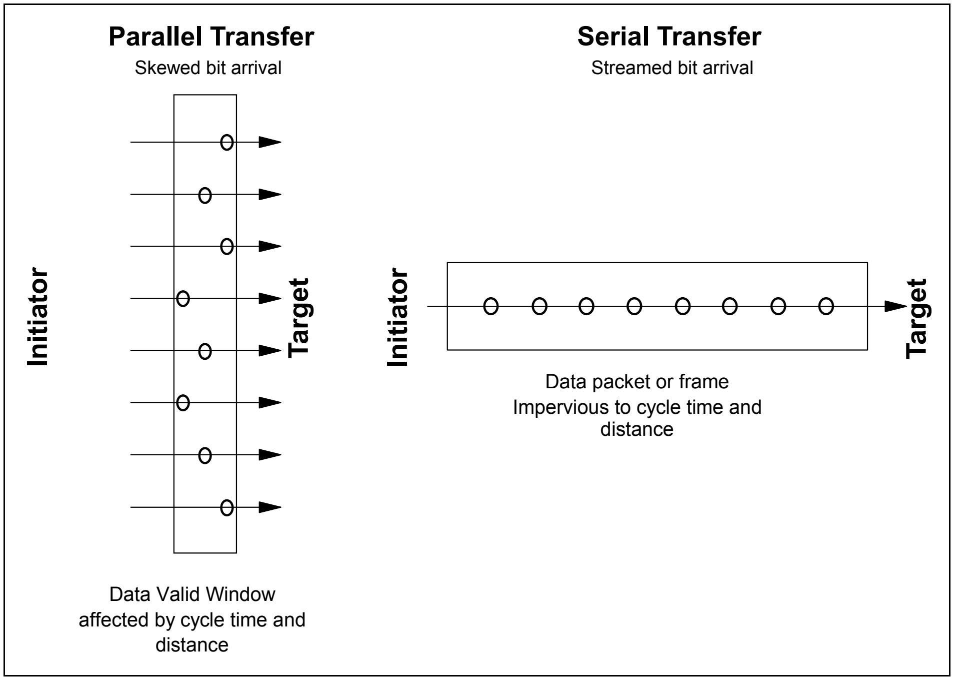


Figure 3-4 Parallel data transfers versus serial data transfers

Serial transfer enables simpler cabling and connectors, and also routing of information through switched networks. Fibre Channel can operate over longer distances, both natively and by implementing cascading, and longer with the introduction of repeaters. Just as LANs can be interlinked in WANs by using high speed gateways, so can campus SANs be interlinked to build enterprise wide SANs.

Whatever the topology, information is sent between two nodes, which are the source (transmitter or initiator) and destination (receiver or target). A node is a device, such as a server (personal computer, workstation, or mainframe), or peripheral device, such as disk or tape drive, or video camera. Frames of information are passed between nodes, and the structure of the frame is defined by a protocol. Logically, a source and target node must utilize the same protocol, but each node may support several different protocols or data types.

Therefore, Fibre Channel architecture is extremely flexible in its potential application. Fibre Channel transport layers are protocol independent, enabling the transmission of multiple protocols.

Using a credit based flow control methodology, Fibre Channel is able to deliver data as fast as the destination device buffer is able to receive it. And low transmission overheads enable high sustained utilization rates without loss of data.

Therefore, Fibre Channel combines the best characteristics of traditional I/O channels with those of computer networks:

- High performance for large data transfers by using simple transport protocols and extensive hardware assists
- Serial data transmission
- A physical interface with a low error rate definition
- Reliable transmission of data with the ability to guarantee or confirm error free delivery of the data
- Packaging data in packets (frames in Fibre Channel terminology)
- Flexibility in terms of the types of information which can be transported in frames (such as data, video and audio)
- Use of existing device oriented command sets, such as SCSI and FCP
- A vast expansion in the number of devices which can be addressed when compared to I/O interfaces — a theoretical maximum of more than 15 million ports

It is this high degree of flexibility, availability and scalability; the combination of multiple protocols at high speeds over long distances; and the broad acceptance of the Fibre Channel standards by vendors throughout the IT industry, which makes the Fibre Channel architecture ideal for the development of enterprise SANs.

In the topics that follow we describe some of the key concepts that we have touched upon in the previous pages and that are behind Fibre Channel SAN implementations. We also introduce some more Fibre Channel SAN terminology and jargon that the reader can expect to encounter.

3.2 Layers

Fibre Channel (FC) is broken up into a series of five layers. The concept of layers, starting with the ISO/OSI seven-layer model, allows the development of one layer to remain independent of the adjacent layers. Although, FC contains five layers, those layers follow the general principles stated in the ISO/OSI model.

The five layers can be categorized into these two:

- Physical and signaling layer
- Upper layer

Fibre Channel is a layered protocol. as shown in Figure 3-5.

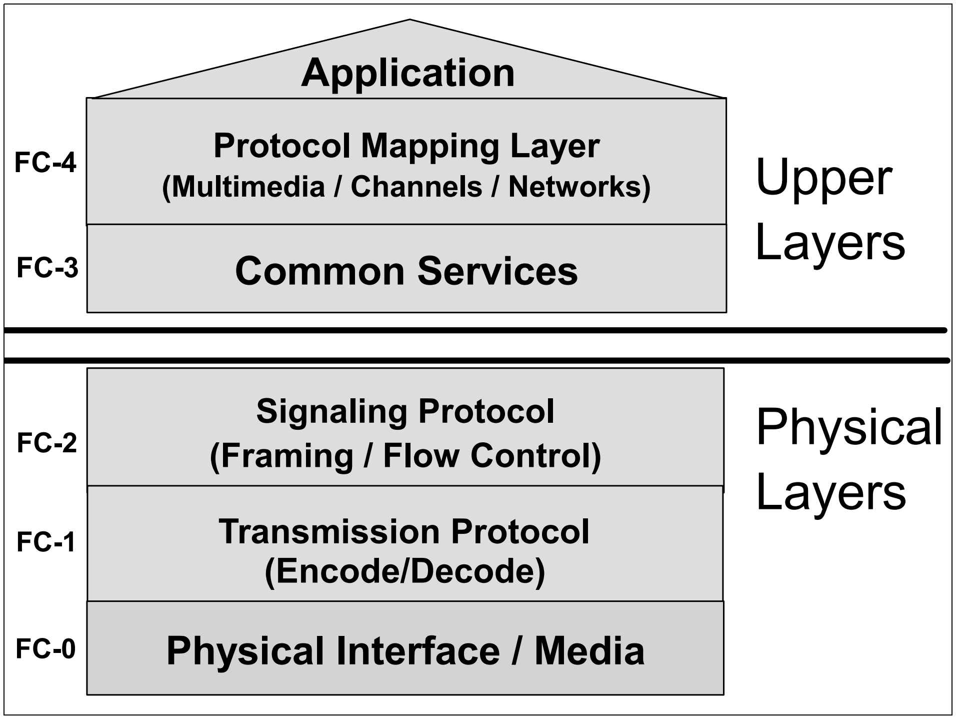


Figure 3-5 Upper and physical layers

The layers can be briefly described as follows:

Physical and signaling layers

The physical and signaling layers include the three lowest layers: FC-0, FC-1, and FC-2.

Physical interface and media: FC-0

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

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

Laser safety

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

Transmission protocol: FC-1

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

Framing and signaling protocol: FC-2

Reliable communications result from Fibre Channel's FC-2 framing and signaling protocol. FC-2 specifies a data transport mechanism that is independent of upper layer protocols. FC-2 is self-configuring and supports point-to-point, Arbitrated Loop, and switched environments.

FC-2, which is the third layer of the FC-PH, provides the transport methods to determine:

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

- Sequence and exchange identifiers
- Segmentation and reassembly

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

Together, FC-0, FC-1, and FC-2 form the Fibre Channel physical and signaling interface (FC-PH).

Upper layers

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

Common services: FC-3

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

- Hunt Groups
  - A *Hunt Group* is a set of associated N_Ports attached to a single node. This set is assigned an alias identifier that allows any frames containing the alias to be routed to any available N_Port within the set. This decreases latency in waiting for an N_Port to become available.
- Striping
  - *Striping* is used to multiply bandwidth, using multiple N_Ports in parallel to transmit a single information unit across multiple links.
- Multicast
  - *Multicast* delivers a single transmission to multiple destination ports. This includes the ability to broadcast to all nodes or a subset of nodes.

Upper layer protocol mapping (ULP): FC-4

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

Through mapping rules, a specific FC-4 describes how ULP processes of the same FC-4 type interoperate.

A channel example is Fibre Channel Protocol (FCP). This is used to transfer SCSI data over Fibre Channel. A networking example is sending IP (Internet Protocol) packets between nodes. FICON is another ULP in use today for mainframe systems. FICON is a contraction of *Fibre Connection* and refers to running ESCON traffic over Fibre Channel.

3.3 Optical cables

An optical fiber is a very thin strand of silica glass in geometry quite like a human hair. In reality it is a very narrow, very long glass cylinder with special characteristics. When light enters one end of the fiber it travels (confined within the fiber) until it leaves the fiber at the other end. Two critical factors stand out:

- Very little light is lost in its journey along the fiber.
- Fiber can bend around corners and the light will stay within it and be guided around the corners.

An optical fiber consists of two parts: the core and the cladding. See Figure 3-6 on page 42. The core is a narrow cylindrical strand of glass and the cladding is a tubular jacket surrounding it. The core has a (slightly) higher refractive index than the cladding. This means that the boundary (interface) between the core and the cladding acts as a perfect mirror. Light travelling along the core is confined by the mirror to stay within it — even when the fiber bends around a corner.

When light is transmitted on a fiber, the most important consideration is “what kind of light?” The electromagnetic radiation that we call light exists at many wavelengths. These wavelengths go from invisible infrared through all the colors of the visible spectrum to invisible ultraviolet. Because of the attenuation characteristics of fiber, we are only interested in infrared “light” for communication applications. This light is usually invisible, since the wavelengths used are usually longer than the visible limit of around 750 nanometers (nm).

If a short pulse of light from a source such as a laser or an LED is sent down a narrow fiber, it will be changed (degraded) by its passage down the fiber. It will emerge (depending on the distance) much weaker, lengthened in time (“smeared out”), and distorted in other ways. The reasons for this are as follows:

3.3.1 Attenuation

The pulse will be weaker because all glass absorbs light. More accurately, impurities in the glass can absorb light but the glass itself does not absorb light at the wavelengths of interest. In addition, variations in the uniformity of the glass cause scattering of the light. Both the rate of light absorption and the amount of scattering are dependent on the wavelength of the light and the characteristics of the particular glass. Most light loss in a modern fiber is caused by scattering. Typical attenuation characteristics of fiber for varying wavelengths of light are illustrated in Figure 13 on page 31.

3.3.2 Maximum power

There is a practical limit to the amount of power that can be sent on a fiber. This is about half a watt (in standard single-mode fiber) and is due to a number of non-linear effects that are caused by the intense electromagnetic field in the core when high power is present.

Polarization

Conventional communication optical fiber is cylindrically symmetric but contains imperfections. Light travelling down such a fiber is changed in polarization. (In current optical communication systems this does not matter but in future systems it may become a critical issue.)

Dispersion

Dispersion occurs when a pulse of light is spread out during transmission on the fiber. A short pulse becomes longer and ultimately joins with the pulse behind, making recovery of a reliable bit stream impossible. (In most communications systems bits of information are sent as pulses of light. 1 = light, 0 = dark. But even in analogue transmission systems where information is sent as a continuous series of changes in the signal, dispersion causes distortion.) There are many kinds of dispersion, each of which works in a different way, but the most important three are discussed below:

Material dispersion (chromatic dispersion)

Both lasers and LEDs produce a range of optical wavelengths (a band of light) rather than a single narrow wavelength. The fiber has different refractive index characteristics at different wavelengths and therefore each wavelength will travel at a different speed in the fiber. Thus, some wavelengths arrive before others and a signal pulse disperses (or smears out).

Modal dispersion

When using multimode fiber, the light is able to take many different paths or “modes” as it travels within the fiber. The distance traveled by light in each mode is different from the distance travelled in other modes. When a pulse is sent, parts of that pulse (rays or quanta) take many different modes (usually all available modes). Therefore, some components of the pulse will arrive before others. The difference between the arrival time of light taking the fastest mode versus the slowest obviously gets greater as the distance gets greater.

Waveguide dispersion

Waveguide dispersion is a very complex effect and is caused by the shape and index profile of the fiber core. However, this can be controlled by careful design and, in fact, waveguide dispersion can be used to counteract material dispersion.

Noise

One of the great benefits of fiber optical communications is that the fiber doesn't pick up noise from outside the system. However, there are various kinds of noise that can come from components within the system itself. Mode partition noise can be a problem in single-mode fiber and modal noise is a phenomenon in multimode fibre.

None of these effects are helpful to engineers, wishing to transmit information over long distances on a fiber. But much can be done, and is being done, about it.

And at this point it would be appropriate to say that it is not our intention to go any deeper into optical than this. We felt that some of this was fundamental to understanding why a SAN is the way that it is, why it also has limits, and without delving deeply into the laws of physics. And this is a good point to get back to basics!

3.3.3 Fiber in the SAN

Fibre Channel can be run over optical or copper media, but fiber-optic cables enjoys a major advantage in noise immunity as we mentioned previously. It is for this reason that fiber-optic cabling is preferred. However, copper is also used, and it is likely that in the short term a mixed environment will need to be tolerated and supported although this is less likely to be the case as SANs mature.

In addition to the noise immunity, fiber-optic cabling provides a number of distinct advantages over copper transmission lines that make it a very attractive medium for many applications. At the forefront of the advantages are:

- Greater distance capability than is generally possible with copper
- Insensitive to induced electro-magnetic interference (EMI)
- No emitted electro-magnetic radiation (RFI)
- No electrical connection between two ports
- Not susceptible to crosstalk
- Compact and lightweight cables and connectors

However, fiber-optic and optical links do have some drawbacks. Some of the considerations are:

- Optical links tend to be more expensive than copper links over short distances.
- Optical connections do not lend themselves to backplane printed circuit wiring.
- Optical connections may be affected by dirt and other contamination.

Overall, optical fibers have provided a very high-performance transmission medium, which has been refined and proven over many years.

Mixing fiber-optical and copper components in the same environment is supported, although not all products provide that flexibility, and this should be taken into consideration when planning a SAN. Copper cables tend to be used for short distances, up to 30 meters, and can be identified by their DB-9, 9 pin, connector.

Normally, fiber-optic cabling is referred to by mode or the frequencies of light waves that are carried by a particular cable type. Fiber cables come in two distinct types, as shown in Figure 3-6.

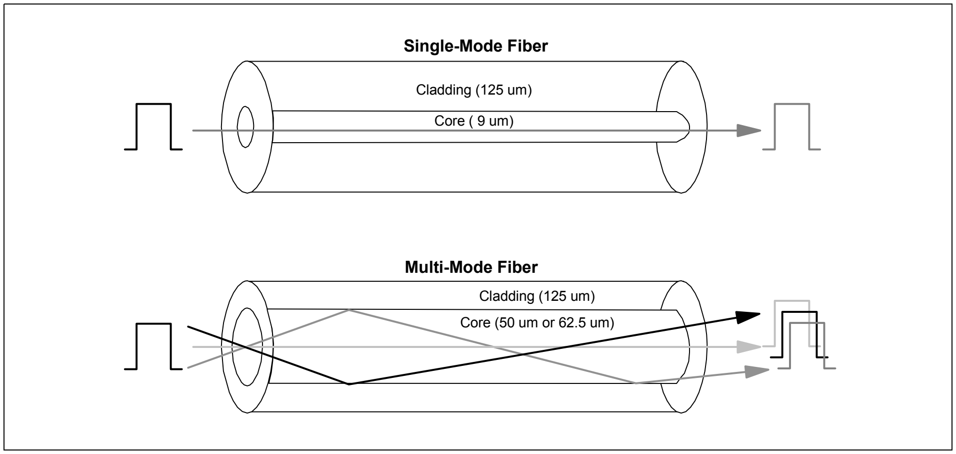


Figure 3-6 Cable types

- Multi-mode fiber (MMF) for shorter distances

  Multi-mode cabling is used with shortwave laser light and has either a 50 micron or a 62.5 micron core with a cladding of 125 micron. The 50 micron or 62.5 micron diameter is sufficiently large for injected light waves to be reflected off the core interior.

  Multi-mode fiber allows more than one mode of light. Common MM core sizes are 50 micron and 62.5 micron. Multi-mode fiber is better suited for shorter distance applications. Where costly electronics are heavily concentrated, the primary cost of the system does not lie with the cable. In such a case, MM fibre is more economical because it can be used with inexpensive connectors and laser devices, thereby reducing the total system cost.

- Single-mode fiber (SMF) for longer distances

  Single-mode (SM) fibre allows only one pathway, or mode, of light to travel within the fibre. The core size is typically 8.3 micron. Single-mode fibres are used in applications where low signal loss and high data rates are required, such as on long spans between two system or network devices, where repeater/amplifier spacing needs to be maximized.

Fibre Channel architecture supports both short wave and long wave optical transmitter technologies, as follows:

- Short wave laser

  This technology uses a wavelength of 780 nanometers and is only compatible with multi-mode fiber.

- Long wave laser

  This technology uses a wavelength of 1300 nanometers. It is compatible with both single-mode and multi-mode fiber.

3.3.4 Dark fiber

In order to connect one optical device to another, some form of fiber optic link is required. If the distance is short, then a standard fiber *cable* will suffice. Over a slightly longer distance, for example from one building to the next, then a fiber link may need to be laid. This may need to be laid underground or through a conduit, but it will not be as simple as connecting two switches together in a single rack.

If the two units which need to be connected are in different cities, then the problem is much larger. Larger, in this case, is typically associated with more expensive. As most businesses are not in the cable laying business they will lease fiber optic cables to meet their needs. When a company does this, the fiber optic cable that they lease is known as *dark fiber*.

Dark fiber generically refers to a long, dedicated fiber optic link that can be used without the need for any additional equipment. It can be as long as the particular technology supports.

Some forward thinking services companies have laid fiber optic links alongside their pipes and cables. For example, a water company might be digging up a road to lay a mains pipe; or an electric company might be taking a power cable across a mountain range using pylons, or a cable TV company might be laying cable to all of the buildings in a city. While carrying out the work to support their core business, they may also lay fiber optic links.

But these cables are simply cables. They are not used in anyway by the company who owns them. They remain dark until the user puts their own light down the fiber. Hence, the term *dark fiber*.

3.4 Classes of service

Applications may require different levels of service and guarantees with respect to delivery, connectivity, and bandwidth. Some applications will need to have bandwidth dedicated to them for the duration of the data exchange. An example of this would be a tape backup application. Other applications may be “bursty” in nature and not require a dedicated connection but they may insist that an acknowledgement is sent for each successful transfer.

The Fibre Channel standards provide different classes of service to accommodate the applications needs.

3.4.1 Class 1

In class 1 service, a dedicated connection source and destination is established through the fabric for the duration of the transmission. It provides acknowledged service. This class of service ensures that the frames are received by the destination device in the same order in which they are sent, and reserves full bandwidth for the connection between the two devices. It does not provide for a good utilization of the available bandwidth, since it is blocking another possible contender for the same device. Because of this blocking and necessary dedicated connections, class 1 is rarely used.

3.4.2 Class 2

Class 2 is a connectionless, acknowledged service. Class 2 makes better use of available bandwidth since it allows the fabric to multiplex several messages on a frame-by-frame basis. As frames travel through the fabric they can take different routes, so class 2 service does not guarantee in-order delivery. Class 2 relies on upper layer protocols to take care of frame sequence. The use of acknowledgments reduces available bandwidth, which needs to be considered in large-scale busy networks.

3.4.3 Class 3

There is no dedicated connection in class 3 and the received frames are not acknowledged. Class 3 is also called *datagram connectionless* service. It optimizes the use of fabric resources, but it is now upper layer protocol to ensure that all frames are received in the proper order, and to request to the source device the retransmission of missing frames. Class 3 is a commonly used class of service in Fibre Channel networks.

3.4.4 Class 4

Class 4 is a connection-oriented service like class 1, but the main difference is that it allocates only a fraction of available bandwidth of path through the fabric that connects two N_Ports. Virtual Circuits (VCs) are established between two N_Ports with guaranteed Quality of Service (QoS), including bandwidth and latency. Like class 1, class 4 guarantees in-order delivery frame delivery and provides acknowledgment of delivered frames, but now the fabric is responsible for multiplexing frames of different VCs. Class 4 service is mainly intended for multimedia applications such as video and for applications that allocate an established bandwidth by department within the enterprise. Class 4 was added in the FC-PH-2 standard.

3.4.5 Class 5

Class 5 is called isochronous service, and it is intended for applications that require immediate delivery of the data as it arrives, with no buffering. It is not clearly defined yet. It is not included in the FC-PH documents.

3.4.6 Class 6

Class 6 is a variant of class 1, known as multicast class of service. It provides dedicated connections for a reliable multicast. An N_Port may request a class 6 connection for one or more destinations. A multicast server in the fabric will establish the connections and get acknowledgment from the destination ports, and send it back to the originator. Once a connection is established, it should be retained and guaranteed by the fabric until the initiator ends the connection. Class 6 was designed for applications like audio and video requiring multicast functionality. It appears in the FC-PH-3 standard.

3.4.7 Class F

Class F service is defined in the FC-SW and FC-SW-2 standard for use by switches communicating through ISLs. It is a connectionless service with notification of non-delivery between E_Ports used for control, coordination, and configuration of the fabric. Class F is similar to class 2; the main difference is that Class 2 deals with N_Ports sending data frames, while Class F is used by E_ports for control and management of the fabric.

3.5 Fibre Channel data movement

To move data bits with integrity over a physical medium, there must be a mechanism to check that this has happened and integrity has not been compromised. This is provided by a reference clock, which ensures that each bit is received as it was transmitted. In parallel topologies this can be accomplished by using a separate clock or strobe line. As data bits are transmitted in parallel from the source, the strobe line alternates between high or low to signal the receiving end that a full byte has been sent. In the case of 16 and 32-bit wide parallel cable, it would indicate that multiple bytes have been sent.

The reflective differences in fiber-optic cabling mean that intermodal, or modal, dispersion (signal degradation) may occur.

This may result in frames arriving at different times. This bit error rate (BER) is referred to as the jitter budget. No products are entirely jitter free, and this is an important consideration when selecting the components of a SAN.

As serial data transports only have two leads, transmit and receive, clocking is not possible using a separate line. Serial data must carry the reference timing, which means that clocking is embedded in the bit stream.

Embedded clocking, though, can be accomplished by different means. Fibre Channel uses a byte-encoding scheme (which is covered in more detail in 3.5.1, “Byte encoding schemes” on page 46) and clock and data recovery (CDR) logic to recover the clock. From this, it determines the data bits that comprise bytes and words.

Gigabit speeds mean that maintaining valid signaling, and ultimately valid data recovery, is essential for data integrity. Fibre Channel standards allow for a single bit error to occur only once in a million, million bits (1 in 10^{12}). In the real IT world, this equates to a maximum of one bit error every 16 minutes; however, actual occurrence is a lot less frequent than this.

3.5.1 Byte encoding schemes

In order to transfer data over a high-speed serial interface, the data is encoded prior to transmission and decoded upon reception. The encoding process ensures that sufficient clock information is present in the serial data stream to allow the receiver to synchronize to the embedded clock information and successfully recover the data at the required error rate. This 8b/10b encoding will find errors that a parity check cannot. A parity check will not find even numbers of bit errors, only odd numbers. The 8b/10b encoding logic will find almost all errors.

First developed by IBM, the 8b/10b encoding process will convert each 8-bit byte into two possible 10-bit characters.

This scheme is called 8b/10b encoding, because it refers to the number of data bits input to the encoder and the number of bits output from the encoder.

This scheme is called 8b/10b encoding, because it refers to the number of data bits input to the encoder and the number of bits output from the encoder.

The format of the 8b/10b character is of the format Ann.m, where:

- A represents D for data or K for a special character.
- nn is the decimal value of the lower 5 bits (EDCBA).
- "." is a period.
- m is the decimal value of the upper 3 bits (HGF).

We illustrate an encoding example in Figure 3-7.

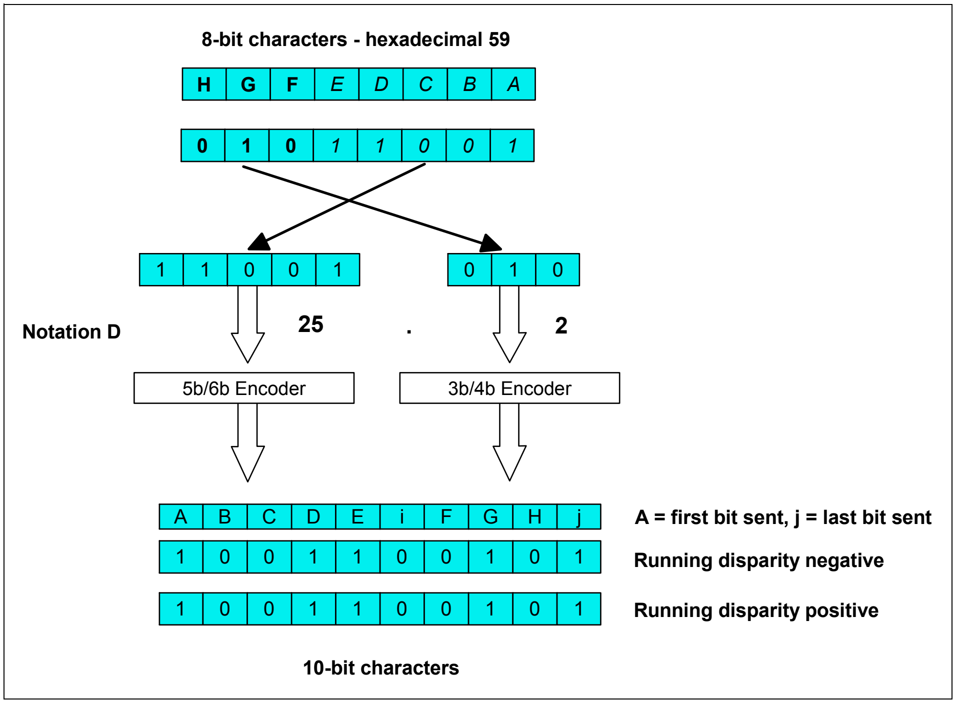


Figure 3-7 8b/10b encoding logic

In the encoding example the following occurs:

1. Hexadecimal representation x'59' is converted to binary: 01011001.
2. Upper three bits are separated from the lower 5 bits: 010 11001.

3. The order is reversed and each group is converted to decimal: 25 2.
4. Letter notation D (for data) is assigned and becomes: D25.2.

Running disparity

As we illustrate, the conversion of the 8-bit data bytes has resulted in two 10-bit results. The encoder needs to choose one of these results to use. This is achieved by monitoring the running disparity of the previously processed character. For example, if the previous character had a positive disparity, then the next character issued should have an encoded value that represents negative disparity.

You will notice that in our example the encoded value, when the running disparity is either positive or negative, is the same. This is legitimate. In some cases the encoded value will differ, and in others it will be the same.

It should be noticed that in the above example the encoded 10-bit byte has 5 bits that are set and 5 bits that are unset. The only possible results of the 8b/10b encoding are as follows:

- If 5 bits are set, then the byte is said to have neutral disparity.
- If 4 bits are set and 6 are unset, then the byte is said to have negative disparity.
- If 6 bits are set and four are unset, then the byte is said to have positive disparity.

The rules of Fibre Channel define that a byte that is sent cannot take the positive or negative disparity above one unit. Thus, if the current running disparity is negative, then the next byte that is sent must either have:

- Neutral disparity
  - Keeping the current running disparity negative.
  - The subsequent byte would need to have either neutral or positive disparity.
- Positive disparity
  - Making the new current running disparity neutral.
  - The subsequent byte could have either positive, negative, or neutral disparity.

Note: By this means, at any point in time, at the end of any byte, the number of set bits and unset bits that have passed over a Fibre Channel link will only differ by a maximum of two.

K28.5

As well as the fact that many 8-bit numbers encode to *two* 10-bit numbers under the 8b/10b encoding scheme, there are some other key features.

Some 10-bit numbers cannot be generated from any 8-bit number. Thus, it should not be possible to see these particular 10-bit numbers as part of a flow of data. This is really a useful fact, as it means that these particular 10-bit numbers can be used by the protocol for signaling or control purposes.

These characters are referred to as Comma characters, and rather than having the prefix D, have the prefix K.

The only one that actually gets used in Fibre Channel is the character known as K28.5, and it has a very special property.

The two 10-bit encoding of K28.5 are shown in Table 3-1.

Table 3-1 10-bit encoding of K28.5

Name of character	Encoding for current running disparity of	
	Negative	**Positive**
K28.5	001111 1010	110000 0101

It was stated above that all of the 10-bit bytes that are possible using the 8b/10b encoding scheme have either four, five, or six bits set. The K28.5 character is special in that it is the only character used in Fibre Channel that has five consecutive bits set or unset; all other characters have four or less consecutive bits of the same setting.

So, what is the significance? There are two things to note here:

- The first is that these ones and zeroes are actually representing light and dark on the fiber (assuming fiber optic medium). A 010 pattern would effectively be a light pulse between two periods of darkness. A 0110 would be the same, except that the pulse of light would last for twice the length of time.

  As the two devices have their own clocking circuitry, the number of consecutive set bits, or consecutive unset bits, becomes important. Let us say that device 1 is sending to device 2 and that the clock on device 2 is running 10 percent faster than that on device 1. If device 1 sent 20 clock cycles worth of set bits, then device 2 would count 22 set bits. (Note that this example is just given to illustrate the point.) The worst possible case that we can have in Fibre Channel is five consecutive bits of the same setting within one byte: The K28.5.

- The other key thing is that because this is the *only* character with five consecutive bits of the same setting, Fibre Channel hardware can look out for it specifically. As K28.5 is used for control purposes, this is very useful and allows the hardware to be designed for maximum efficiency.

3.6 Data transport

In order for Fibre Channel devices to be able to communicate with each other, there needs to be some strict definitions regarding the way that data is sent and received. To this end, some data structures have been defined. It is fundamental to understanding Fibre Channel that you have some knowledge of the way that data is moved around and the mechanisms that are used to accomplish this.

3.6.1 Ordered set

Fibre Channel uses a command syntax, known as an *ordered set*, to move the data across the network. The ordered sets are four-byte transmission words containing data and special characters which have a special meaning. Ordered sets provide the availability to obtain bit and word synchronization, which also establishes word boundary alignment. An ordered set always begins with the special character K28.5. Three major types of ordered sets are defined by the signaling protocol.

The frame delimiters, the Start Of Frame (SOF) and End Of Frame (EOF) ordered sets, establish the boundaries of a frame. They immediately precede or follow the contents of a frame. There are 11 types of SOF and eight types of EOF delimiters defined for the fabric and N_Port Sequence control.

The two primitive signals: idle and receiver ready (R_RDY) are ordered sets designated by the standard to have a special meaning. An Idle is a primitive signal transmitted on the link to indicate an operational port facility ready for frame transmission and reception. The R_RDY primitive signal indicates that the interface buffer is available for receiving further frames.

A primitive sequence is an ordered set that is transmitted and repeated continuously to indicate specific conditions within a port or conditions encountered by the receiver logic of a port. When a primitive sequence is received and recognized, a corresponding primitive sequence or Idle is transmitted in response. Recognition of a primitive sequence requires consecutive detection of three instances of the same ordered set. The primitive sequences supported by the standard are:

- Offline state (OLS)

  The offline primitive sequence is transmitted by a port to indicate one of the following conditions: The port is beginning the link initialization protocol, or the port has received and recognized the NOS protocol or the port is entering the offline status.

- Not operational (NOS)

  The not operational primitive sequence is transmitted by a port in a point-to-point or fabric environment to indicate that the transmitting port has detected a link failure or is in an offline condition, waiting for the OLS sequence to be received.

- Link reset (LR)

  The link reset primitive sequence is used to initiate a link reset.

- Link reset response (LRR)

  Link reset response is transmitted by a port to indicate that it has recognized a LR sequence and performed the appropriate link reset.

Data transfer

To send data over Fibre Channel, though, we need more than just the control mechanisms. Data is sent in frames. One or more related frames make up a sequence. One or more related sequences make up an exchange.

3.6.2 Frames

Fibre Channel places a restriction on the length of the data field of a frame at 528 transmission words, which is 2112 bytes. (See Table 3-2 on page 52.) Larger amounts of data must be transmitted in several frames. This larger unit that consists of multiple frames is called a sequence. An entire transaction between two ports is made up of sequences administered by an even larger unit called an exchange.

> **Note:** Some classes of Fibre Channel communication guarantee that the frames will arrive at the destination in the same order in which they were transmitted; other classes do not. If the frames do arrive in the same order in which they were sent, then we are said to have *in order* delivery of frames.

A frame consists of the following elements:

- SOF delimiter
- Frame header
- Optional headers and payload (data field)
- CRC field
- EOF delimiter

Figure 3-8 on page 52 shows the layout of a Fibre Channel frame.

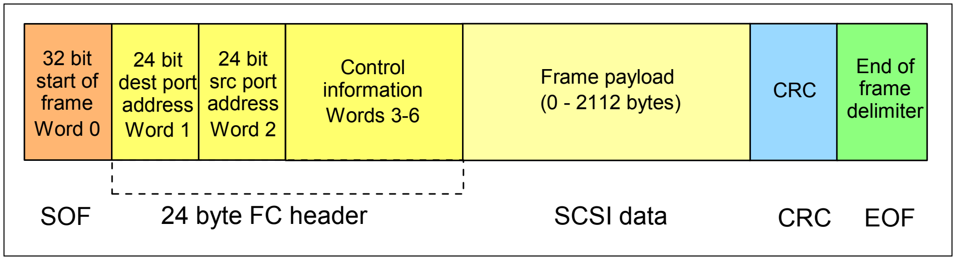

32 bit start of frame Word 0
24 bit dest port address Word 1
24 bit src port address Word 2
Control information Words 3-6
Frame payload (0 - 2112 bytes)
CRC
End of frame delimiter
SOF
24 byte FC header
SCSI data
CRC
EOF

Figure 3-8 Fibre Channel frame structure

Framing rules

The following rules apply to the framing protocol:

- A frame is the smallest unit of information transfer.
- A sequence has at least one frame.
- An exchange has at least one sequence.

Transmission word

A *transmission word* is the smallest transmission unit defined in Fibre Channel. This unit consists of four transmission characters, 4 x 10 or 40 bits. When information transferred is not an even multiple of four bytes, the framing protocol adds fill bytes. The fill bytes are stripped at the destination.

Frames are the building blocks of Fibre Channel. A *frame* is a string of transmission words prefixed by a Start Of Frame (SOF) delimiter and followed by an End Of Frame (EOF) delimiter. The way that transmission words make up a frame is shown in Table 3-2.

Table 3-2 Transmission words in a frame

SOF	Frame Header	Data Payload Transmission Words	CRC	EOF
1 TW	6 TW	0-528 TW	1 TW	1 TW

Frame header

Each frame includes a header that identifies the source and destination of the frame as well as control information that manages the frame as well as sequences and exchanges associated with that frame. The structure of the Frame header is shown in Table 3-3 on page 53. The abbreviations are explained below the table.

Table 3-3 The frame header

	Byte 0	Byte 1	Byte 2	Byte 3
Word 0	R_CTL	Destination_ID (D_ID)		
Word 1	Reserved	Source_ID (S_ID)		
Word 2	Type	Frame Control (F_CTL)		
Word 3	SEQ_ID	DF_CTL	SequenceCount (SEQ_CNT)	
Word 4	Originator X_ID (OX_ID)		Responder X_ID (RX_ID)	
Word 5	Parameter			

Routing control (R_CTL)

This field identifies the type of information contained in the payload and where in the destination node it should be routed.

Destination ID

This field contains the address of the frame destination and is referred to as the D_ID.

Source ID

This field contains the address of where the frame is coming from and is referred to as the S_ID.

Type

Type identifies the protocol of the frame content for data frames, such as SCSI, or a reason code for control frames.

F_CTL

This field contains control information that relates to the frame content.

SEQ_ID

The sequence ID is assigned by the sequence initiator and is unique for a specific D_ID and S_ID pair while the sequence is open.

DF_CTL

Data field control specifies whether there are optional headers present at the beginning of the data field.

SEQ_CNT

This count identifies the position of a frame within a sequence and is incremented by one for each subsequent frame transferred in the sequence.

OX_ID
This field identifies the exchange ID assigned by the originator.

RX_ID
This field identifies the exchange ID to the responder.

Parameter
Parameter specifies relative offset for data frames, or information specific to link control frames.

3.6.3 Sequences

The information in a sequence moves in one direction, from a source N_Port to a destination N_Port. Various fields in the frame header are used to identify the beginning, middle and end of a sequence, while other fields in the frame header are used to identify the order of frames, in case they arrive out of order at the destination.

3.6.4 Exchanges

Two other fields of the frame header identifies the exchange ID. An exchange is responsible for managing a single operation that may span several sequences, possibly in opposite directions. The source and destination can have multiple exchanges active at a time

Using SCSI as an example, a SCSI task is an exchange. The SCSI task is made up of one or more information units. The information units (IUs) would be:

- Command IU
- Transfer ready IU
- Data IU
- Response IU

Each IU is one sequence of the exchange. Only one participant sends a sequence at a time.

3.6.5 In order and out of order

When data is transmitted over Fibre Channel, it is sent in frames. These frames only carry a maximum of 2112 bytes of data, often not enough to hold the entire set of information to be communicated. In this case, more than one frame is needed. Some classes of Fibre Channel communication guarantee that the frames arrive at the destination in the same order that they were transmitted. Other classes do not. If the frames do arrive in the same order that they were sent, then we are said to have *in-order* delivery of frames.

In some cases, it is critical that the frames arrive in the correct order, and in others, it is not so important. In the latter case, *out of order*, the receiving port can reassemble the frames into the correct order before passing the data out to the application. It is, however, quite common for switches and directors to guarantee in-order delivery, even if the particular class of communication allows for the frames to be delivered out of sequence.

3.6.6 Latency

The term *latency* means the delay between an action requested and an action taking place.

Latency occurs almost everywhere. A simple fact is that it takes time and energy to perform an action. The areas where we particularly need to be aware of latency in a SAN are:

- Ports
- Switches, directors
- Interblade links in a core switch or director
- Long distance links
- Inter Switch links
- ASICs

3.6.7 Open Fiber Control

When dealing with lasers there is potential danger to the eyes. Generally, the lasers in use in Fibre Channel are low-powered devices designed for quality of light and signaling rather than for maximum power. However, they can still be dangerous.

> **Important:** Never look into a laser light source. Never look into the end of an fiber optic cable unless you know exactly where the other end is, and you also know that nobody could connect a light source to it.

To add a degree of safety, the concept of Open Fiber Control (OFC) was developed. The idea is as follows:

1. A device is turned on and it sends out low powered light.
2. If it does not receive light back, then it assumes that there is no fiber connected. This is a fail-safe option.
3. When it receives light, it assumes that there is a fiber connected and switches the laser to full power.
4. If one of the devices stops receiving light, then it will revert to the low power mode.

When a device is transmitting at low power, it is not able to send data. The device is just waiting for a completed optical loop.

OFC ensures that the laser does not emit light which would exceed the Class1 laser limit when no fiber is connected. Non-OFC devices are guaranteed to be below Class 1 limits at all times.

The key factor is that the devices at each end of a fiber link must either both be OFC or both be non-OFC.

All modern equipment uses Non-OFC optics, but it is possible that some legacy (or existing) equipment may be using OFC optics.

3.7 Flow control

Now that we know data is sent in frames, we also need to understand that devices need to temporarily store the frames as they arrive, and until they are assembled in sequence, and then delivered to the upper layer protocol. The reason for this is that due to the high bandwidth that Fibre Channel is capable of, it would be possible to inundate and overwhelm a target device with frames. There needs to be a mechanism to stop this happening. The ability of a device to accept a frame is called its credit. This credit is usually referred to as the number of buffers (its buffer credit) that a node maintains for accepting incoming data.

3.7.1 Buffer to buffer

During login, N_Ports and F_Ports (we describe these in 4.2, “Port types” on page 74) at both ends of a link establish its buffer to buffer credit (BB_Credit). Each port states the maximum BB_Credit that they can offer and the lower of the two is used.

3.7.2 End to end

At login all N_Ports establish end to end credit (EE_Credit) with each other.

During data transmission, a port should not send more frames than the buffer of the receiving port can handle before getting an indication from the receiving port that it has processed a previously sent frame.

3.7.3 Controlling the flow

Two counters are used to accomplish successful flow control: BB_Credit_CNT and EE_Credit_CNT, and both are initialized to 0 during login. Each time a port sends a frame it increments BB_Credit_CNT and EE_Credit_CNT by one. When it receives R_RDY from the adjacent port it decrements BB_Credit_CNT by one, and when it receives ACK from the destination port it decrements EE_Credit_CNT by one. If at any time BB_Credit_CNT becomes equal to the BB_Credit or EE_Credit_CNT equal to the EE_Credit of the receiving port, the transmitting port has to stop sending frames until the respective count is decremented.

The previous statements are true for Class 2 service. Class 1 is a dedicated connection, so it does not need to care about BB_Credit and only EE_Credit is used (EE Flow Control). Class 3 on the other hand is an unacknowledged service, so it only uses BB_Credit (BB Flow Control), but the mechanism is the same on all cases.

3.7.4 Performance

Here we can see the importance that the number of buffers has in overall performance. We need enough buffers to make sure the transmitting port can continue sending frames without stopping in order to use the full bandwidth. This is particularly true with distance. At 1 Gbps a frame occupies between about 75 m and 4 km of fiber depending on the size of the data payload. In a 100 km link we could send many frames before the first one reaches its destination. We need an acknowledgement (ACK) back to start replenishing EE_Credit or a receiver ready (R_RDY) indication to replenish BB_Credit.

For a moment, let us consider frames with 2 KB of data. These occupy approximately 4 km of fiber. We will be able to send about 25 frames before the first arrives at the far end of our 100 km link. We will be able to send another 25 before the first R_RDY or ACK is received, so we would need at least 50 buffers to allow for nonstop transmission at 100 km distance with frames of this size. If the frame size is reduced, more buffers would be required to allow nonstop transmission.

3.8 Addressing

All devices in a Fibre Channel environment have an identity. The way that the identity is assigned and used depends on the format of the Fibre Channel fabric. For example, there is a difference between the way that addressing is done in an arbitrated loop and a fabric.

3.8.1 World Wide Name

All Fibre Channel devices have a unique identity called the World Wide Name (WWN). This is similar to the way all Ethernet cards have a unique Media Access Control (MAC) address.

Each N_Port will have its own WWN, but it also possible for a device with more than one Fibre Channel adapter to have its own WWN as well. Thus, for example, an IBM TotalStorage Enterprise Storage Server® has its own WWN as well as incorporating the WWNs of the adapter within it. This means that a soft zone can be created using the entire array, or individual zones could be created using particular adapters. In the future, this will be the case of the servers as well.

This WWN is a 64-bit address, and if two WWN addresses are put into the frame header, this leaves 16 bytes of data just for identifying destination and source address. So 64-bit addresses can impact routing performance.

Each device in the SAN is identified by a unique world wide name (WWN). The WWN contains a vendor identifier field, which is defined and maintained by the IEEE, and a vendor-specific information field.

Currently, there are two formats of the WWN as defined by the IEEE. The original format contains either a hex 10 or hex 20 in the first two bytes of the address. This is then followed by the vendor-specific information.

Both the old and new WWN formats are shown in Figure 3-9 on page 59.

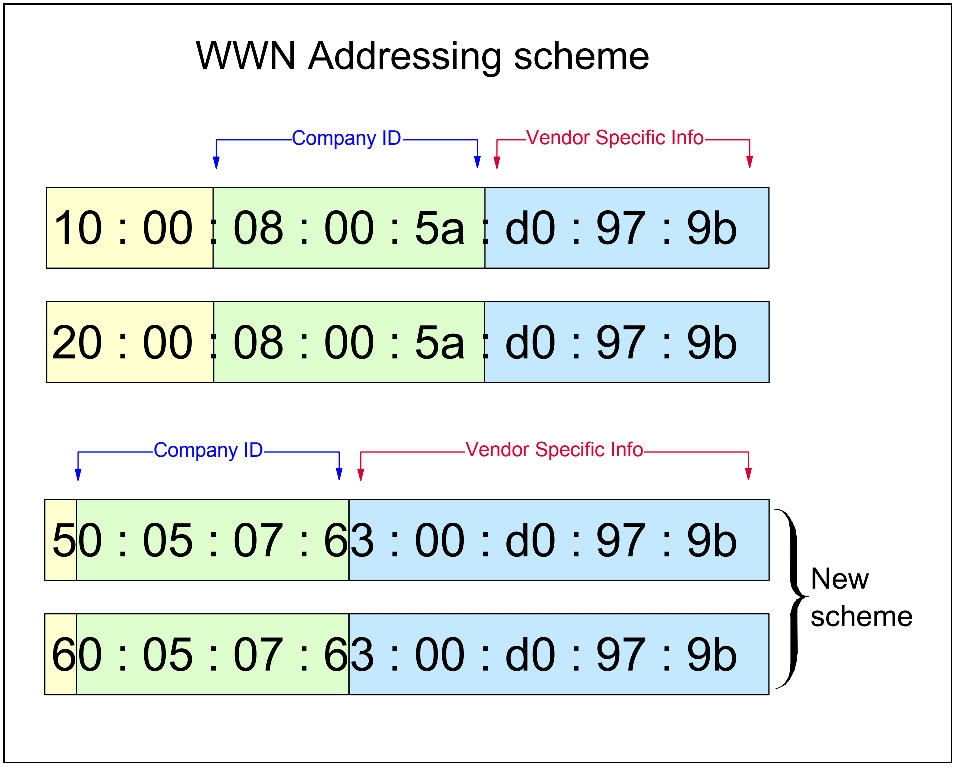

WWN Addressing scheme
Company ID
Vendor Specific Info
10 : 00 : 08 : 00 : 5a : d0 : 97 : 9b
20 : 00 : 08 : 00 : 5a : d0 : 97 : 9b
Company ID
Vendor Specific Info
50 : 05 : 07 : 63 : 00 : d0 : 97 : 9b
60 : 05 : 07 : 63 : 00 : d0 : 97 : 9b
New scheme

Figure 3-9 World Wide Name addressing scheme

The new addressing scheme starts with a hex 5 or 6 in the first half-byte followed by the vendor identifier in the next 3 bytes. The vendor-specific information is then contained in the following fields.

A worldwide node name (WWNN) is a globally unique 64-bit identifier assigned to each Fibre Channel *node* process

Some nodes (devices) may have multiple Fibre Channel adapters, like an ESS, for example. In this case the device also has an identifier for each of its Fibre Channel adapters. This identifier is called the world wide port name (WWPN). It is possible to uniquely identify all Fibre Channel adapters and paths within a device.

3.8.2 Port address

Because of the potential impact on routing performance by using 64-bit addressing, there is another addressing scheme used in Fibre Channel networks. This scheme is used to address ports in the switched fabric. Each port in the switched fabric has its own unique 24-bit address. With this 24-bit address scheme, we get a smaller frame header, and this can speed up the routing process. With this frame header and routing logic, the Fibre Channel is optimized for high-speed switching of frames.

With a 24-bit addressing scheme, this allows for up to 16 million addresses, which is an address space larger than any practical SAN design in existence in today's world. There needs to be some relationship between this 24-bit address and the 64-bit address associated with World Wide Names. We will explain this in the section that follows.

3.8.3 24-bit port address

The 24-bit address scheme removes the overhead of manual administration of addresses by allowing the topology itself to assign addresses. This is not like WWN addressing, in which the addresses are assigned to the manufacturers by the IEEE standards committee, and are built in to the device at the time of manufacture. If the topology itself assigns the 24-bit addresses, then somebody has to be responsible for the addressing scheme from WWN addressing to port addressing.

In the switched fabric environment, the switch itself is responsible for assigning and maintaining the port addresses. When the device with its WWN logs into the switch on a specific port, the switch will assign the port address to that port and the switch will also maintain the correlation between the port address and the WWN address of the device of that port. This function of the switch is implemented by using the Name Server.

The Name Server is a component of the fabric operating system, which runs inside the switch. It is essentially a database of objects in which fabric-attached device registers its values.

Dynamic addressing also removes the partial element of human error in addressing maintenance, and provides more flexibility in additions, moves, and changes in the SAN.

A 24-bit port address consists of three parts:

- Domain (from bits 23 to 16)
- Area (from bits 15 to 08)
- Port or Arbitrated Loop physical address: AL_PA (from bits 07 to 00)

We show how the address is built up in Figure 3-10 on page 61.

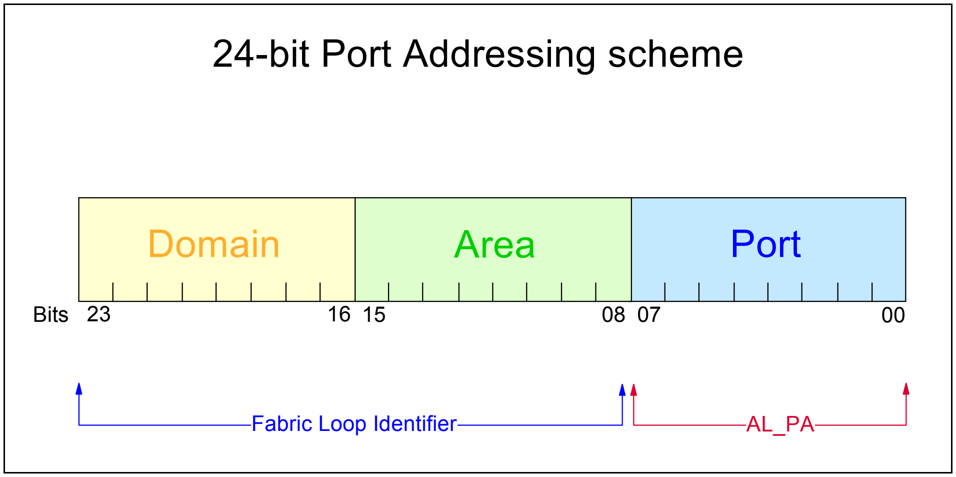


Figure 3-10 Fabric port address

The significance of some of the bits that make up the port address in the are:

- Domain

  The most significant byte of the port address is the domain. This is the address of the switch itself. One byte allows up to 256 possible addresses. Because some of these are reserved, as for the one for broadcast, there are only 239 addresses available. This means that you can theoretically have as many as 239 switches in your SAN environment. The domain number allows each switch to have a unique identifier if you have multiple interconnected switches in your environment.

- Area

  The area field provides 256 addresses. This part of the address is used to identify the individual FL_Ports supporting loops or it can be used as the identifier for a group of F_Ports, for example, a card with more ports on it. This means that each group of ports has a different area number, even if there is only one port in the group.

- Port

  The final part of the address provides 256 addresses for identifying attached N_Ports and NL_Ports.

To arrive at the number of available addresses is a simple calculation based on:

```
Domain x Area x Ports
```

This means that there are `239 x 256 x 256 = 15,663,104` addresses available.

3.8.4 Loop address

An NL_Port, like an N_Port, has a 24-bit port address. If no switch connection exists, the two upper bytes of this port address are zeroes (x’00 00’) and referred to as a private loop. The devices on the loop have no connection with the outside world. If the loop is attached to a fabric and an NL_Port supports a fabric login, the upper two bytes are assigned a positive value by the switch. We call this mode a public loop.

As fabric-capable NL_Ports are members of both a local loop and the greater fabric community, a 24-bit address is needed as an identifier in the network. In this case of public loop assignment, the value of the upper two bytes represents the loop identifier, and this will be common to all NL_Ports on the same loop that performed login to the fabric.

In both public and private arbitrated loops, the last byte of the 24-bit port address refers to the arbitrated loop physical address (AL_PA). The AL_PA is acquired during initialization of the loop and may, in the case of a fabric-capable loop device, be modified by the switch during login.

The total number of the AL_PAs available for arbitrated loop addressing is 127. This number is based on the requirements of 8b/10b running disparity between frames.

3.8.5 FICON address

FICON generates the 24-bit FC port address field in yet another way. When communication is required from the FICON channel port to the FICON CU port, the FICON channel (using FC-SB-2 and FC-FS protocol information) will provide both the address of its port, the source port address identifier (S_ID), and the address of the CU port, the destination port address identifier (D_ID) when the communication is from the channel N_Port to the CU N_Port.

The Fibre Channel architecture does not specify how a server N_Port determines the destination port address of the storage device N_Port with which it requires communication. This is node and N_Port implementation dependent. Basically, there are two ways that a server can determine the address of the N_Port with which it wishes to communicate:

- The *discovery* method, by knowing the World Wide Name (WWN) of the target Node N_Port and then requesting a WWN for the N_Port port address from a Fibre Channel Fabric Service called the fabric Name Server.
- The *defined* method, by the server (processor channel) N_Port having a known predefined port address of the storage device (CU) N_Port with which it requires communication. This later approach is referred to as the *port address definition* approach, and is the approach that is implemented for the

FICON channel in FICON native (FC) mode by the IBM @server® zSeries and the 9672 G5/G6, using either the z/OS HCD function or an IOCP program to define a one-byte switch port, a one-byte FC area field of the 3-byte fiber channel N_Port port address.

The Fibre Channel architecture (FC-FS) uses a 24-bit FC port address, three bytes, for each port in an FC switch. The switch port addresses in a FICON native (FC) mode are always assigned by the switch fabric.

For the FICON channel in FICON native (FC) mode, the Accept (ACC ELS) response to the Fabric Login (FLOGI), in a switched point-to-point topology, provides the channel with the 24-bit N_Port address to which the channel is connected. This N_Port address is in the ACC destination address field (D_ID) of the FC-2 header.

The FICON CU port will also perform a fabric login to obtain its 24-bit FC port address. Figure 3-11 shows the FC-FS 24-bit FC port address identifier is divided into three fields.

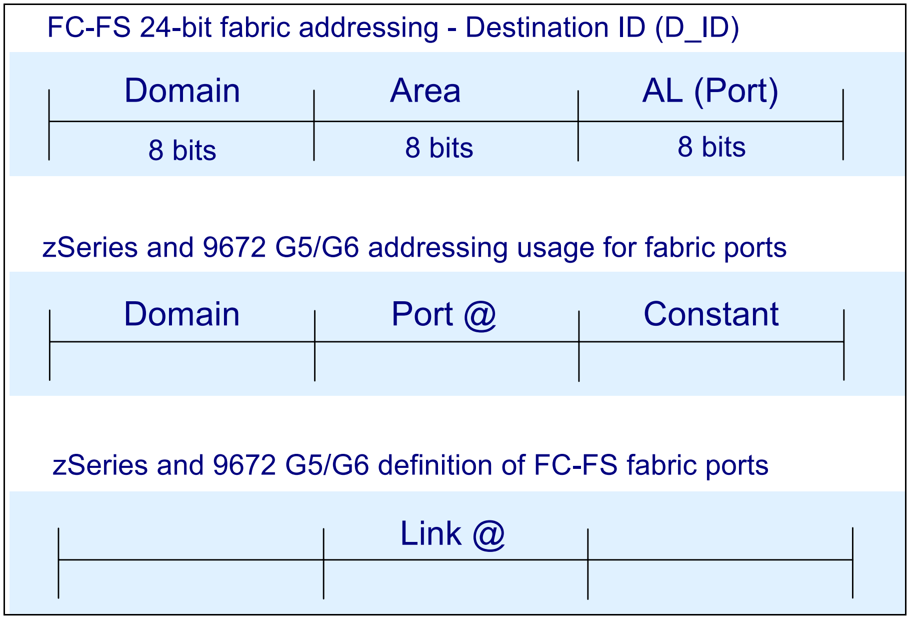

FC-FS 24-bit fabric addressing - Destination ID (D_ID)
Domain
Area
AL (Port)
8 bits
8 bits
8 bits
zSeries and 9672 G5/G6 addressing usage for fabric ports
Domain
Port @
Constant
zSeries and 9672 G5/G6 definition of FC-FS fabric ports
Link @

Figure 3-11 FICON port addressing

It shows the FC-FS 24-bit port address and the definition of usage of that 24-bit address in a zSeries and 9672 G5/G6 environment. Only the eight bits making up the FC port address are defined for the zSeries and 9672 G5/G6 to access a FICON CU. The FICON channel in FICON native (FC) mode working with a switched point-to-point FC topology, single switch, provides the other two bytes that make up the three-byte FC port address of the CU to be accessed.

The zSeries and 9672 G5/G6 processors, when working with a switched point-to-point topology, require that the Domain and the AL_Port (Arbitrated Loop) field values be the same for all the FC F_Ports in the switch. Only the area field value will be different for each switch F_Port.

For the zSeries and 9672 G5/G6 the *area* field is referred to as the F_Port's *port address field*. It is just a one-byte value, and when defining access to a CU that is attached to this port, using the zSeries HCD or IOCP, the port address is referred to as the Link address.

As shown in Figure 3-12, the eight bits for the domain address and the eight-bit constant field are provided from the Fabric Login initialization result, while the eight bits, one byte for the port address (1-byte Link address), are provided from the zSeries or 9672 G5/G6 CU link definition (using HCD and IOCP).

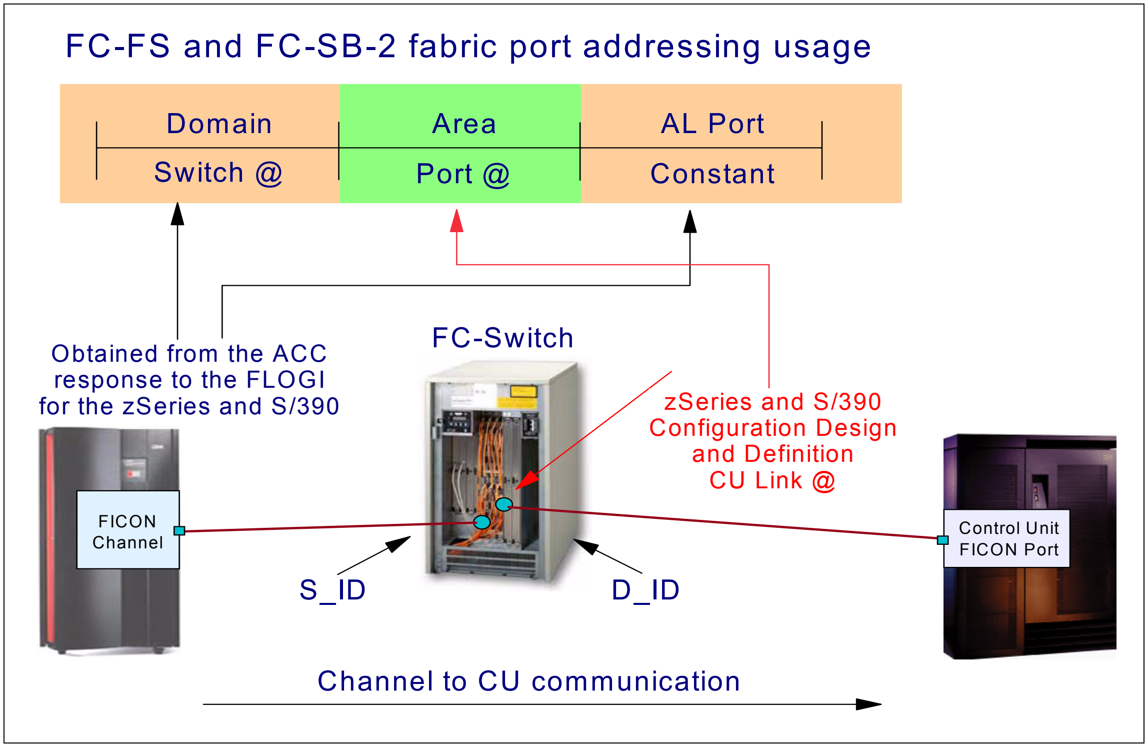

FC-FS and FC-SB-2 fabric port addressing usage
Domain
Switch @
Area
Port @
AL Port
Constant
Obtained from the ACC response to the FLOGI for the zSeries and S/390
FC-Switch
zSeries and S/390 Configuration Design and Definition CU Link @
FICON Channel
Control Unit FICON Port
S_ID
D_ID
Channel to CU communication

Figure 3-12 FICON single switch: Switched point-to-point link address

FICON address support for cascaded switches

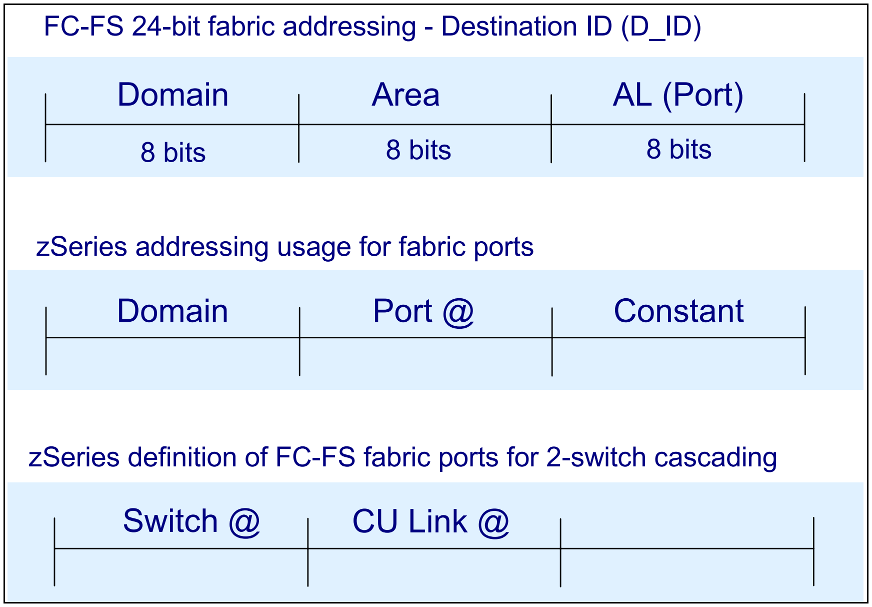
The Fibre Channel architecture (FC-FS) uses a 24-bit FC port address of three bytes for each port in an FC switch. The switch port addresses in a FICON native (FC) mode are always assigned by the switch fabric.

For the FICON channel in FICON native (FC) mode, the Accept (ACC ELS) response to the Fabric Login (FLOGI) in a two-switch cascaded topology, provides the channel with the 24-bit N_Port address to which the channel is connected. This N_Port address is in the ACC destination address field (D_ID) of the FC-2 header.

The FICON CU port will also perform a fabric login to obtain its 24-bit FC port address.

Figure 3-13 on page 66 shows that the FC-FS 24-bit FC port address identifier is divided into three fields:

- Domain
- Area
- AL Port



FC-FS 24-bit fabric addressing - Destination ID (D_ID)
Domain
Area
AL (Port)
8 bits
8 bits
8 bits
zSeries addressing usage for fabric ports
Domain
Port @
Constant
zSeries definition of FC-FS fabric ports for 2-switch cascading
Switch @
CU Link @

Figure 3-13 FICON addressing for cascaded directors

It shows the FC-FS 24-bit port address and the definition usage of that 24-bit address in a zSeries environment. Here, 16 bits making up the FC port address must be defined for the zSeries to access a FICON CU in a cascaded environment. The FICON channel in FICON native (FC) mode working with a cascaded FC topology, two-switch, provides the remaining byte making up the full three-byte FC port address of the CU to be accessed.

It is required that the Domain, switch @, and the AL_Port, Arbitrated Loop, field value be the same for all the FC F_Ports in the switch. Only the area field value will be different for each switch F_Port.

The zSeries domain and area fields are referred to as the F_Port's port address field. It is a two-byte value, and when defining access to a CU that is attached to this port, using the zSeries HCD or IOCP, the port address is referred to as the Link address.

As shown in Figure 3-14 on page 67, the eight bits for the constant field are provided from the Fabric Login initialization result, while the 16 bits for the port address, two-byte Link address, are provided from the zSeries CU link definition using HCD and IOCP.

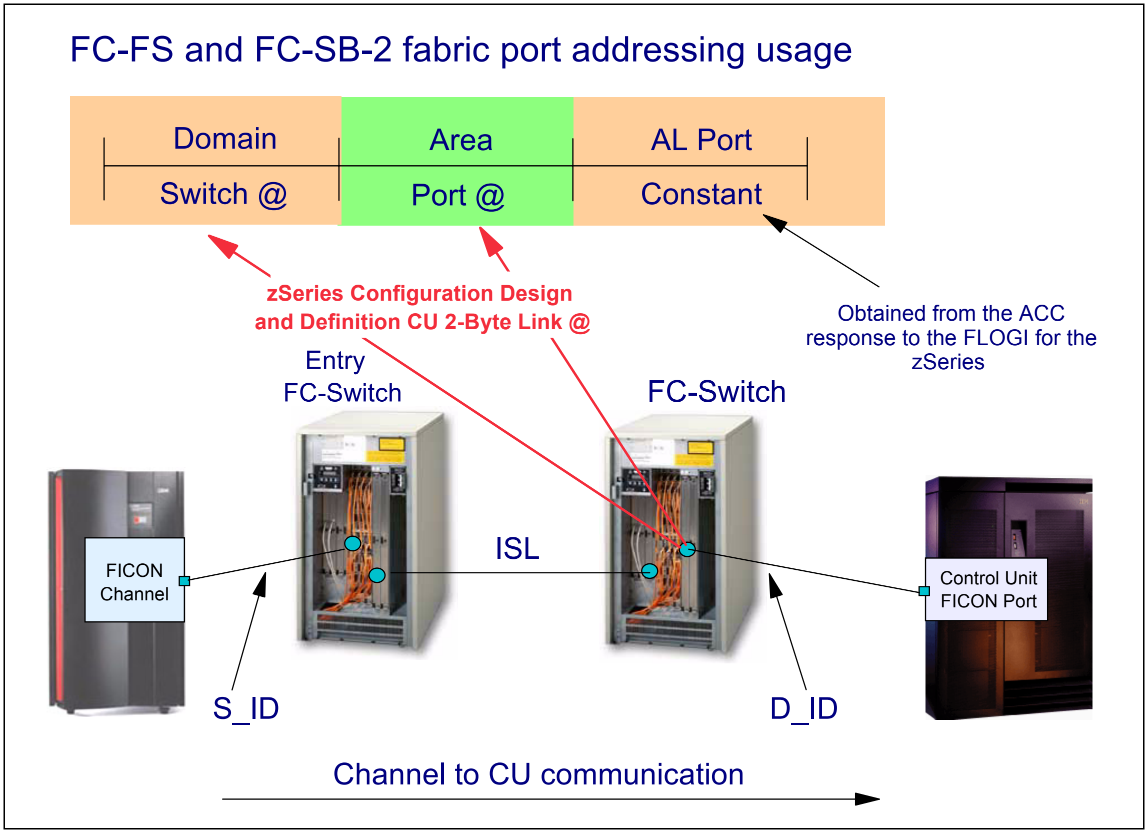

FC-FS and FC-SB-2 fabric port addressing usage
Domain
Switch @
Area
Port @
AL Port
Constant
zSeries Configuration Design and Definition CU 2-Byte Link @
Obtained from the ACC response to the FLOGI for the zSeries
Entry FC-Switch
FC-Switch
ISL
FICON Channel
Control Unit FICON Port
S_ID
D_ID
Channel to CU communication

Figure 3-14 Two cascaded director FICON addressing

As a footnote, FCP connectivity is device-centric and is defined in the fabric using the WWPN of the devices that are allowed to communicate. When an FCP device attaches to the fabric, it queries the Name Server for the list of devices that it is allowed to form connections with (i.e. the zoning information). FICON devices do not query the Name Server for accessible devices because the allowable port/device relationships have been defined in the host, thus the zoning and Name Server information does not need to be retrieved.

4

Topologies and other fabric services

Historically, interfaces to storage consisted of parallel bus architectures (such as SCSI and IBM bus and tag) that supported a small number of devices. Fibre Channel technology provides a means to implement robust storage networks that may consist of hundreds or thousands of devices. Fibre Channel SANs support high-bandwidth storage traffic, at the time of writing up to 10 Gbps.

Storage subsystems, storage devices, and server systems can be attached to a Fibre Channel SAN. Depending on the implementation, several different components can be used to build a SAN. It is, as the name suggests, a network so any combination of devices that are able to interoperate are likely to be utilized.

Given this, a Fibre Channel network may be composed of many different types of interconnect entities, including directors, switches, hubs, routers, gateways, and bridges.

It is the deployment of these different types of interconnect entities that allow Fibre Channel networks of varying scale to be built. In smaller SAN environments you can employ hubs for Fibre Channel arbitrated loop topologies, or switches and directors for Fibre Channel switched fabric topologies. As SANs increase in size and complexity, Fibre Channel directors can be introduced to facilitate a more flexible and fault tolerant configuration. Each of the components that compose a Fibre Channel SAN should provide an individual management capability, as well as participate in an often complex end-to-end management environment.

As we have stated previously, a SAN is a dedicated high-performance network to move data between heterogeneous servers and storage resources. It is a separate dedicated network that avoids any traffic conflicts between clients and servers, which are typically encountered on the traditional messaging network. We show this distinction in Figure 4-1.

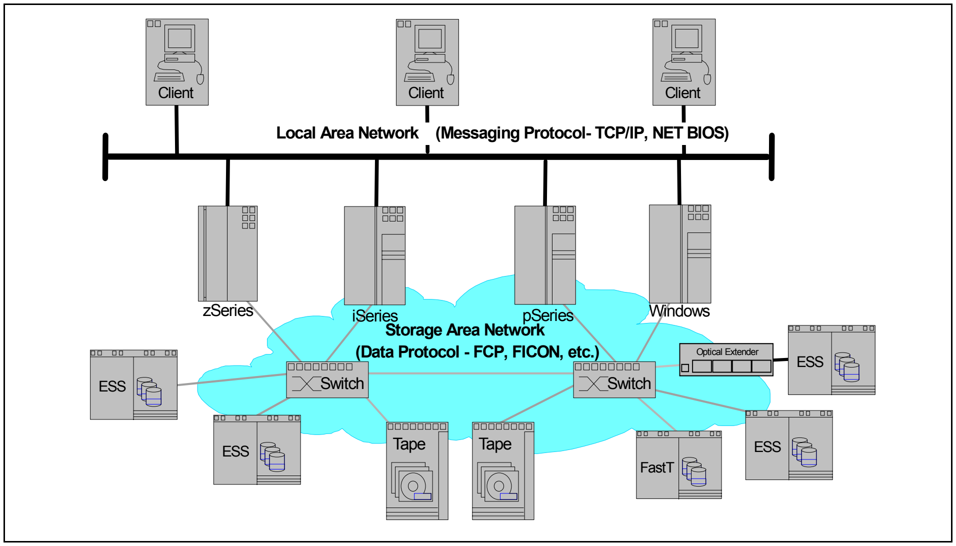


Figure 4-1 The SAN

A Fibre Channel SAN is high performing because of its inherent bandwidth speed, and this is partly aided because of the unique packaging of data, which only consumes about 5 percent of overhead.

4.1 Fibre Channel topologies

Before we describe some of the other physical components of the SAN, it is necessary to introduce Fibre Channel topologies. Fibre Channel based networks share many similarities with other networks, but differ considerably by the absence of topology dependencies. Networks based on Token Ring, Ethernet, and FDDI are topology dependent and cannot share the same media because they have different rules for communication. The only way they can interoperate is through bridges and routers. Each uses its own media-dependent data encoding methods and clock speeds, header format, and frame length restrictions.

Fibre Channel based networks support three types of topologies, which include point-to-point, arbitrated loop, and switched. These can be stand-alone or interconnected to form a fabric.

The three Fibre Channel topologies are:

- Point-to-point
- Arbitrated loop
- Switched fabric

A switched fabric is the most commonly encountered topology today. and will be the focus for the remainder of this redbook.

We will describe these topologies.

4.1.1 Point-to-point

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

Fibre Channel is a full duplex protocol, which means both paths transmit data simultaneously. As an example, Fibre Channel connections based on the 1 Gbps standard are able to transmit at 100 MBps and receive at 100 MBps simultaneously. Again, as an example, for Fibre Channel connections based on the 2 Gbps standard, they can transmit at 200 MBps and receive at 200 MBps simultaneously. This will extend to 4 Gbps and 10 Gbps technologies as well.

Illustrated in Figure 4-2 on page 72 is a simple point-to-point connection.

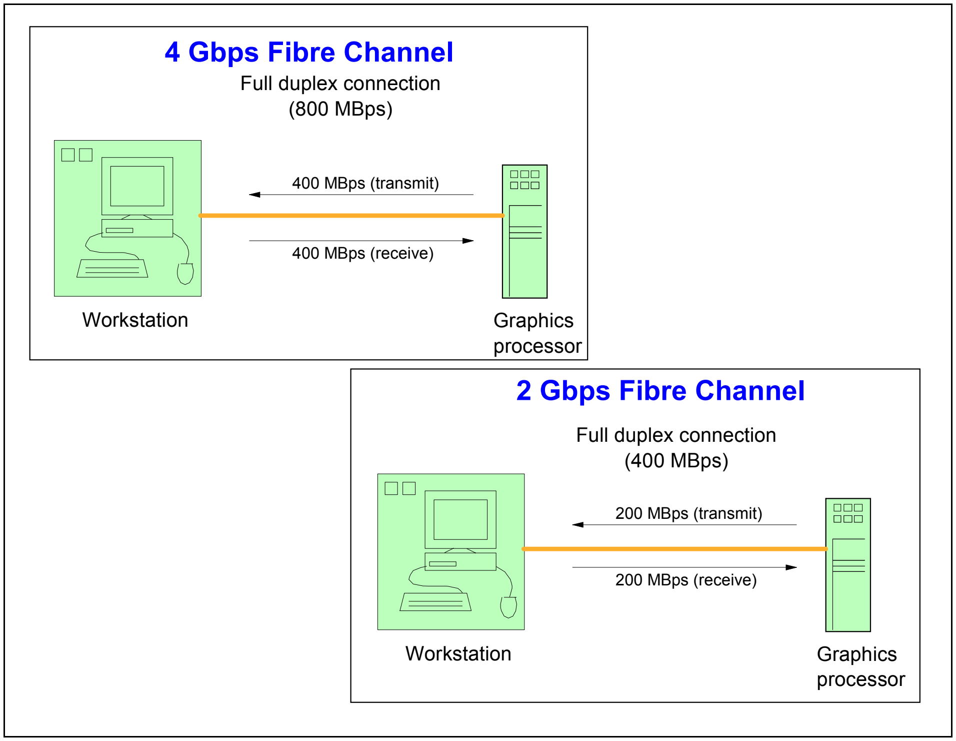


Figure 4-2 Point-to-point

4.1.2 Arbitrated loop

Our second topology is Fibre Channel Arbitrated Loop (FC-AL). FC-AL is more useful for storage applications. It is a loop of up to 126 nodes (NL_Ports) that is managed as a shared bus. Traffic flows in one direction, carrying data frames and primitives around the loop with a total bandwidth of 400 MBps (or 200 MBps for a loop based on 2 Gbps technology).

Using arbitration protocol, a single connection is established between a sender and a receiver, and a data frame is transferred around the loop. When the communication comes to an end between the two connected ports, the loop becomes available for arbitration and a new connection may be established. Loops can be configured with hubs to make connection management easier. A distance of up to 10 km is supported by the Fibre Channel standard for both of these configurations. However, latency on the arbitrated loop configuration is affected by the loop size.

A simple loop, configured using a hub, is shown in Figure 4-3 on page 73.

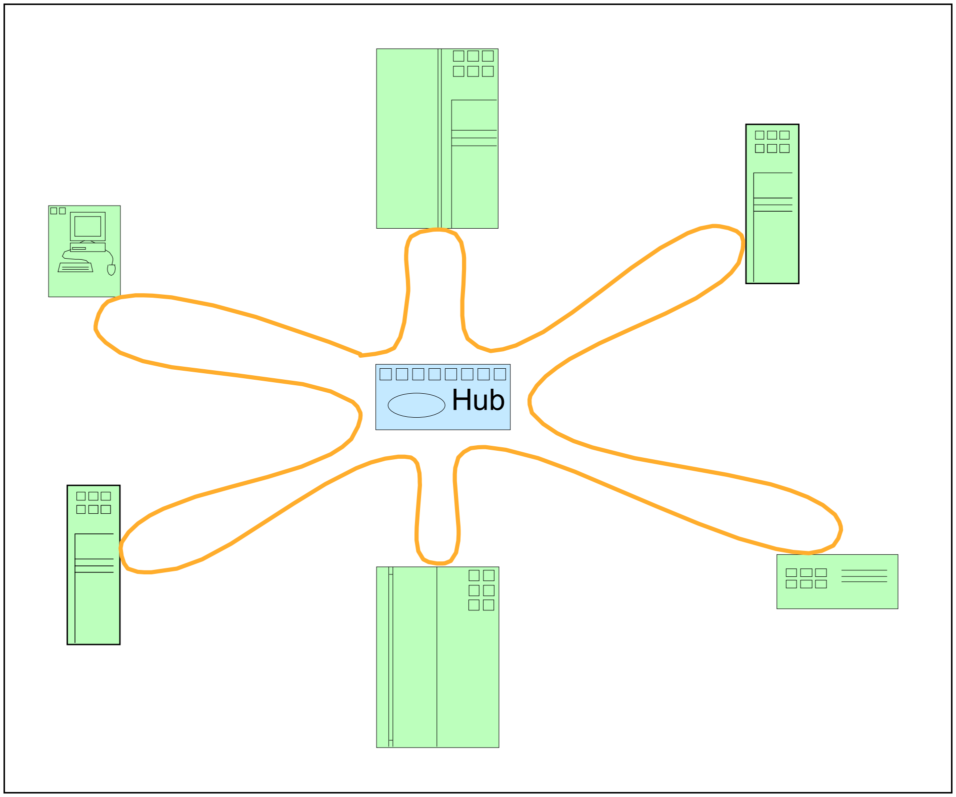


Figure 4-3 Arbitrated loop

We discuss FC-AL in more depth in 4.3, “Fibre Channel Arbitrated Loop protocols” on page 76.

4.1.3 Switched fabric

Our third, and the most useful topology used in SAN implementations, is Fibre Channel Switched Fabric (FC-SW). It applies to switches and directors that support the FC-SW standard, that is, it is not limited to switches as its name suggests. A Fibre Channel fabric is one or more fabric switches in a single, sometimes extended, configuration. Switched fabrics provide full bandwidth per port compared to the shared bandwidth per port in arbitrated loop implementations.

One of the key differentiators is that if you add a new device into the arbitrated loop, you further divide the shared bandwidth. However, in a switched fabric, adding a new device or a new connection between existing ones actually increases the bandwidth. For example, an 8-port switch (for example let’s assume it is based on 2 Gbps technology) with three initiators and three targets can support three concurrent 200 MBps conversations or a total of 600 MBps throughput (1,200 MBps if full-duplex applications were available).

A switched fabric configuration is shown in Figure 4-4.

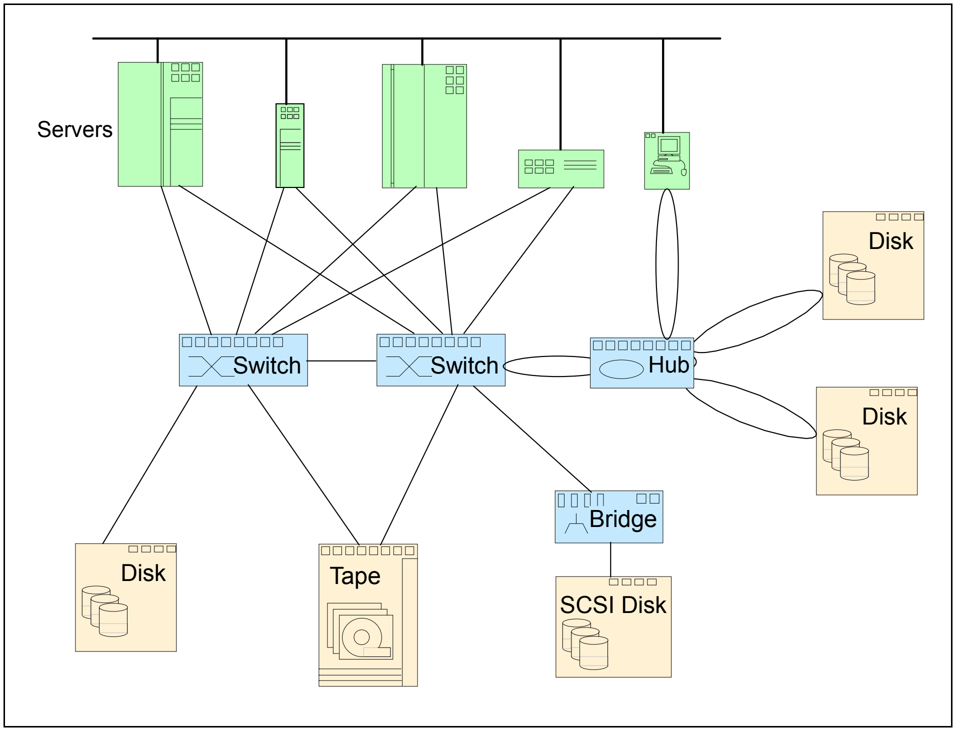


Figure 4-4 Sample switched fabric configuration

This is one of the major reasons why arbitrated loop is fast becoming a legacy SAN topology. A switched fabric is usually referred to as a *fabric.* One way of identifying yourself as a newcomer to the SAN world is to refer to your “switched fabric”, just use the term fabric and you will be fine.

4.2 Port types

The basic building block of the Fibre Channel is the port. The following lists the various types of Fibre Channel port types and their purposes in switches, servers, and storage. These are the types of Fibre Channel ports that are likely to be encountered:

- E_Port: This is an expansion port. A port is designated an E_Port when it is used as an inter-switch expansion port (ISL) to connect to the E_Port of another switch, to enlarge the switch fabric.
- F_Port: This is a fabric port that is not loop capable. It is used to connect an N_Port point-point to a switch.
- FL_Port: This is a fabric port that is loop capable. It is used to connect an NL_Port to the switch in a public loop configuration.
- G_Port: This is a generic port that can operate as either an E_Port or an F_Port. A port is defined as a G_Port after it is connected but has not received a response to *loop* initialization or has not yet completed the link initialization procedure with the adjacent Fibre Channel device.
- L_Port: This is a loop-capable node or switch port.
- U_Port: This is a universal port—a more generic switch port than a G_Port. It can operate as either an E_Port, F_Port, or FL_Port. A port is defined as a U_Port when it is not connected or has not yet assumed a specific function in the fabric.
- N_Port: This is a node port that is not loop capable. It is used to connect an equipment port to the fabric.
- NL_Port: This is a node port that is loop capable. It is used to connect an equipment port to the fabric in a loop configuration through an L_Port or FL_Port.
- MTx_Port: CNT port used as a mirror for viewing the transmit stream of the port to be diagnosed.
- MRx_Port: CNT port used as a mirror for viewing the receive stream of the port to be diagnosed.
- SD_Port: Cisco SPAN port used for mirroring another port for diagnostic purposes.

Figure 4-5 on page 76 represents the most commonly encountered Fibre Channel port types, and where they are likely to be encountered.

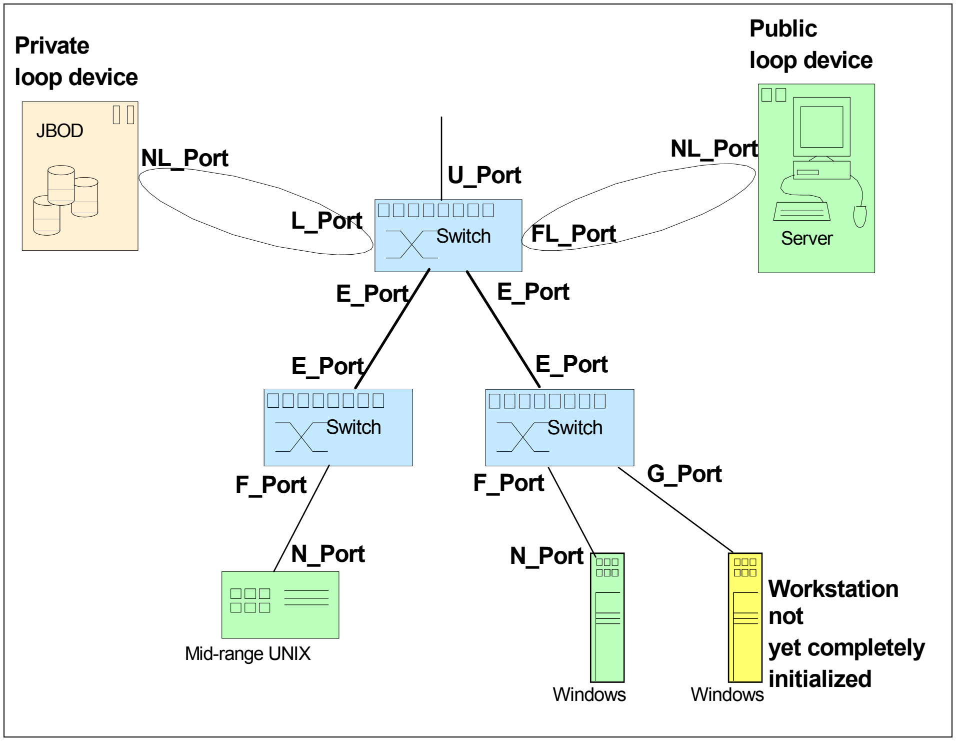


Figure 4-5 Fibre Channel port types

4.2.1 Domain ID

A domain ID is a unique number that identifies the switch or director to a fabric. It can be either static or dynamic. Static (insistent) domain IDs are a requirement for FICON. Each manufacturer will have a range of numbers, and a maximum number of domain IDs that can be used in a fabric.

4.3 Fibre Channel Arbitrated Loop protocols

To support the shared behavior of Fibre Channel Arbitrated Loop (FC-AL), a number of loop-specific protocols are used. These protocols are used to:

- Initialize the loop and assign addresses.
- Arbitrate for access to the loop.
- Open a loop circuit with another port in the loop.
- Close a loop circuit when two ports have completed their current use of the loop.

- Implement the access fairness mechanism to ensure that each port has an opportunity to access the loop.

We discuss some of these topics in the sections that follow.

4.3.1 Fairness algorithm

The way that the fairness algorithm works is based around the IDLE ordered set, and the way that arbitration is carried out. In order to determine that the loop is not in use, an NL_Port waits until it sees an IDLE go by and it can arbitrate for the loop by sending an RB Primitive Signal ordered set. If a higher priority device arbitrates before the first NL_Port sees its own ARB come by, then it loses the arbitration; but if it sees that its own ARB has gone all the way around the loop, then it has won arbitration. It can then open a communication to another NL_Port. When it has finished, it can close the connection and either arbitrate for the loop or send one or more IDLEs. If it complies with the fairness algorithm then it will take the option of sending IDLEs. That will force lower priority NL_Ports to successfully arbitrate for sending IDLEs, and that will allow lower priority NL_Ports to successfully arbitrate for the loop. However, there is no rule that forces any device to operate the fairness algorithm.

4.3.2 Loop addressing

An NL_Port, like an N_Port, has a 24-bit port address. If no switch connection exists, the two upper bytes of this port address are zeroes (x’00 00’) and referred to as a private loop. The devices on the loop have no connection with the outside world. If the loop is attached to a fabric and NL_Port supports a fabric login, the upper two bytes are assigned a positive value by the switch. We call this mode a public loop.

As fabric-capable NL_Ports are members of both a local loop and a greater fabric community, a 24-bit address is needed as an identifier in the network. In the case of public loop assignment, the value of the upper two bytes represents the loop identifier, and this will be common to all NL_Ports on the same loop that performed login to the fabric.

In both public and private Arbitrated Loops, the last byte of the 24-bit port address refers to the Arbitrated Loop physical address (AL_PA). The AL_PA is acquired during initialization of the loop and may, in the case of fabric-capable loop devices, be modified by the switch during login.

The total number of the AL_PAs available for Arbitrated Loop addressing is 127, which is based on the requirements of 8b/10b running disparity between frames.

As a frame terminates with an end-of-frame character (EOF), this will force the current running disparity negative. In the Fibre Channel standard, each transmission word between the end of one frame and the beginning of another frame should also leave the running disparity negative. If all 256 possible 8-bit bytes are sent to the 8b/10b encoder, 134 emerge with neutral disparity characters. Of these 134, seven are reserved for use by Fibre Channel. The 127 neutral disparity characters left have been assigned as AL_PAs. Put another way, the 127 AL_PA limit is simply the maximum number, minus reserved values, of neutral disparity addresses that can be assigned for use by the loop. This does not imply that we recommend this amount, or load, but only that it is possible.

Arbitrated Loop will assign priority to AL_PAs, based on numeric value. The lower the numeric value, the higher the priority is.

It is the Arbitrated Loop initialization that ensures each attached device is assigned a unique AL_PA. The possibility for address conflicts only arises when two separated loops are joined together without initialization.

Note: System z9™ and zSeries servers do not support the arbitrated loop topology.

4.4 Fibre Channel login

There are three different types of login for Fibre Channel. These are:

- Port login
- Process login
- Fabric login

We explain their roles in the SAN in the topics that follow.

4.4.1 Port login

Port login, also known as PLOGI, is used to establish a session between two N_Ports and is necessary before any upper level commands or operations can be performed. During port login, two N_Ports (devices) swap service parameters and make themselves known to each other.

4.4.2 Process login

Process login is also known as PRLI. Process login is used to set up the environment between related processes on an originating N_Port and a responding N_Port. A group of related processes is collectively known as an image pair. The processes involved can be system processes and system images, such as mainframe logical partitions, control unit images, and FC-4 processes. Use of process login is optional from the perspective of the Fibre Channel FC-2 layer, but may be required by a specific upper-level protocol, as in the case of SCSI-FCP mapping.

4.4.3 Fabric login

After the fabric-capable Fibre Channel device is attached to a fabric switch, it will carry out a fabric login (FLOGI).

Similar to port login, FLOGI is an extended link service command that sets up a session between two participants. With FLOGU a session is created between an N_Port or NL_Port and the switch. An N_Port will send a FLOGI frame that contains its Node Name, its N_Port Name, and service parameters to a well-known address of 0xFFFFFE.

The switch accepts the login and returns an accept (ACC) frame to the sender. If some of the service parameters requested by the N_Port or NL_Port are not supported the switch will set the appropriate bits in the ACC frame to indicate this.

NL_Ports derives its AL_PA during the loop initialization process (LIP). The switch then decides if it will accept this AL_PA, if it does not conflict with any previously assigned AL_PA on the loop. If not, a new AL_PA is assigned to the NL_Port, which then causes the start of another LIP.

4.5 Fibre Channel fabric services

There is a set of services available to all devices participating in a Fibre Channel fabric. They are known as fabric services, and include:

- Management services
- Time services
- Simple name server
- Login services
- Registered State Change Notification (RSCN)

These services are implemented by switches and directors participating in the SAN. Generally speaking, the services are distributed across all the devices, and a node can make use of whichever switching device it is connected to.

4.5.1 Management services

This is an in-band fabric service that allows data to be passed from device to management platforms. This will include such information as the topology of the SAN. A critical feature of this service is that it allows management software access to the SNS, bypassing any potential block caused by zoning. This means that a management suite can have a view of the entire SAN. The well-known port used for the Management Server is 0xFFFFFA.

4.5.2 Time services

At the time of writing, this has not been defined. However, the assigned port is OxFFFFFB. It is intended for the management of fabric-wide expiration timers or elapsed time values, and not intended for precise time synchronization.

4.5.3 Simple name server

Fabric switches implement a concept known as the simple name server (SNS). All switches in the fabric keep the Simple Name Server (SNS) updated, and are therefore aware of all devices in the SNS. After a node has successfully logged into the fabric, it performs a PLOGI into a well-known port, 0xFFFFFC. This allows it to register itself and pass on critical information such as class of service parameters, its WWN/address, and the upper layer protocols that it can support.

4.5.4 Login services

In order to do a fabric login, a node communicates with the login server at address 0xFFFFFE.

4.5.5 Registered State Change Notification

This service, Registered State Change Notification (RSCN), is critical, as it propagates information about a change in the state of one node to all other nodes in the fabric. This means that in the event of, for example, a node being shut down, that the other nodes on the SAN will be informed and can take necessary steps to stop communicating with it. This prevents the other nodes from trying to communicate with the node that has been shut down, timing out, and retrying.

4.6 Routing mechanisms

A complex fabric can be made of interconnected switches and directors, perhaps even spanning a LAN/WAN connection. The challenge is to route the traffic with a minimum of overhead, latency, and reliability, and to prevent out-of-order delivery of frames. Here are some of the mechanisms.

4.6.1 Spanning tree

In case of failure, it is important to consider having an alternative path between source and destination available. This will allow data to still reach its destination. However, having different paths available could lead to the delivery of frames being out of the order, due to frame taking a different path and arriving earlier than one of its predecessors.

A solution, which can be incorporated into the meshed fabric, is called a spanning tree and is an IEEE 802.1 standard. This means that switches keep to certain paths, as the spanning tree protocol will block certain paths to produce a simply connected active topology. Then the shortest path in terms of hops is used to deliver the frames, and only one path is active at a time. This means that all associated frames go over the same path to the destination. The paths that are blocked can be held in reserve and used only if, for example, a primary path fails.

The most commonly used path selection protocol is fabric shortest path first (FSPF). This type of path selection is usually performed at boot time, and no configuration is needed. All paths are established at start time, and only if no inter-switch link (ISL) is broken or added will reconfiguration take place.

4.6.2 Fabric shortest path first

According to the FC-SW-2 standard, fabric shortest path first (FSPF) is a link state path selection protocol. The concepts used in FSPF were first proposed by Brocade, and have since been incorporated into the FC-SW-2 standard. Since then it has been adopted by most, if not all, manufacturers.

What FSPF is

FSPF keeps track of the links on all switches in the fabric and associates a cost with each link. The cost is always calculated as being directly proportional to the number of hops. The protocol computes paths from a switch to all other switches in the fabric by adding the cost of all links traversed by the path, and choosing the path that minimizes the cost.

How FSPF works

The collection of link states (including cost) of all switches in a fabric constitutes the topology database (or link state database). The topology database is kept in all switches in the fabric, and they are maintained and synchronized to each other. There is an initial database synchronization, and an update mechanism.

The initial database synchronization is used when a switch is initialized, or when an ISL comes up. The update mechanism is used when there is a link state change. This ensures consistency among all switches in the fabric.

How FSPF helps

In the situation where there are multiple routes, FSPF will ensure that the route that is used is the one with the lowest number of hops. If all the hops:

- Have the same latency
- Operate at the same speed
- Have no congestion

then FSPF will ensure that the frames get to their destinations by the fastest route.

We discuss FSPF in greater depth in 6.5, “Inter-switch links” on page 121.

4.7 Zoning

Zoning allows for finer segmentation of the switched fabric. Zoning can be used to instigate a barrier between different environments. Only the members of the same zone can communicate within that zone and all other attempts from outside are rejected.

For example, it might be desirable to separate a Microsoft Windows NT environment from a UNIX environment. This is very useful because of the manner in which Windows attempts to claim all available storage for itself. Because not all storage devices are capable of protecting their resources from any host seeking available resources, it makes sound business sense to protect the environment in another manner. We show an example of zoning in Figure 4-6 on page 83 where we have separated AIX from NT and created Zone 1 and Zone 2. This diagram also shows how a device can be in more than one zone.

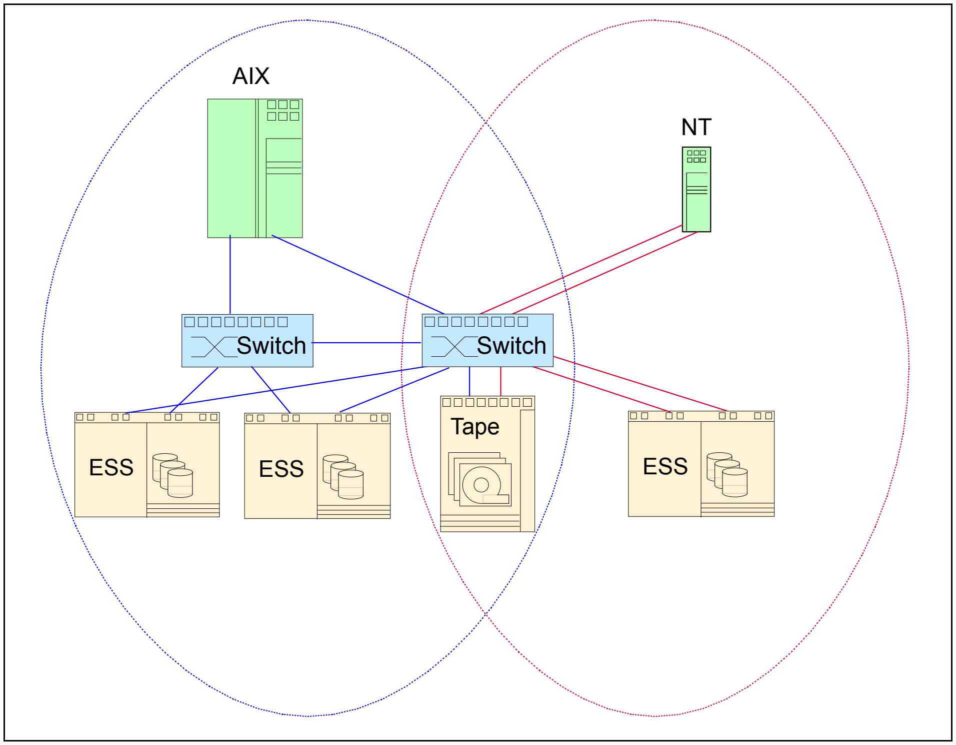


Figure 4-6 Zoning

Looking at zoning in this way, it could also be considered as a security feature, and not just for separating environments. Zoning could also be used for test and maintenance purposes. For example, not many enterprises will mix their test and maintenance environments with their production environment. Within a fabric, you could easily separate your test environment from your production bandwidth allocation on the same fabric using zoning.

An example of zoning is shown in Figure 4-7 on page 84. In this case:

- Server A and Storage A can communicate with each other.
- Server B and Storage B can communicate with each other.
- Server A cannot communicate with Storage B.
- Server B cannot communicate with Storage A.
- Both servers and both storage devices can communicate with the tape.

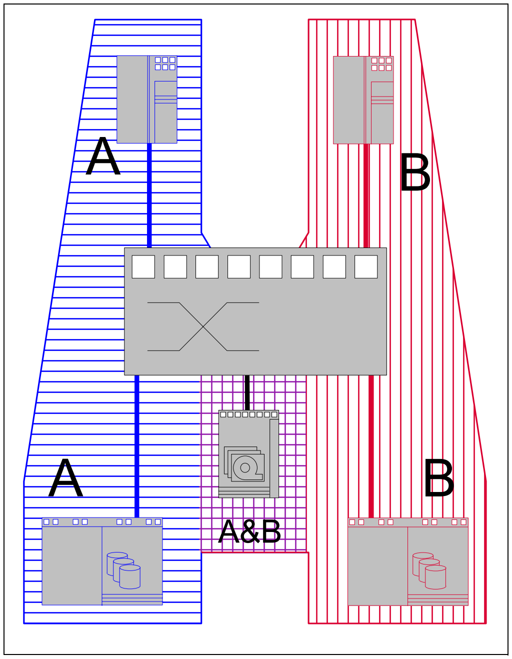


Figure 4-7 An example of zoning

Zoning also introduces the flexibility to manage a switched fabric to meet different user groups objectives.

Zoning can be implemented in two ways:

- Hardware zoning
- Software zoning

These forms of zoning are different, but are not necessarily mutually exclusive. Depending upon the particular manufacturer of the SAN hardware, it is possible for hardware zones and software zones to overlap. While this adds to the flexibility, it can make the solution complicated, increasing the need for good management software and documentation of the SAN.

4.7.1 Hardware zoning

Hardware zoning is based on the physical fabric port number. The members of a zone are physical ports on the fabric switch. It can be implemented in the following configurations:

- One-to-one
- One-to-many
- Many-to-many

Figure 4-8 shows an example of zoning based on the switch port numbers.

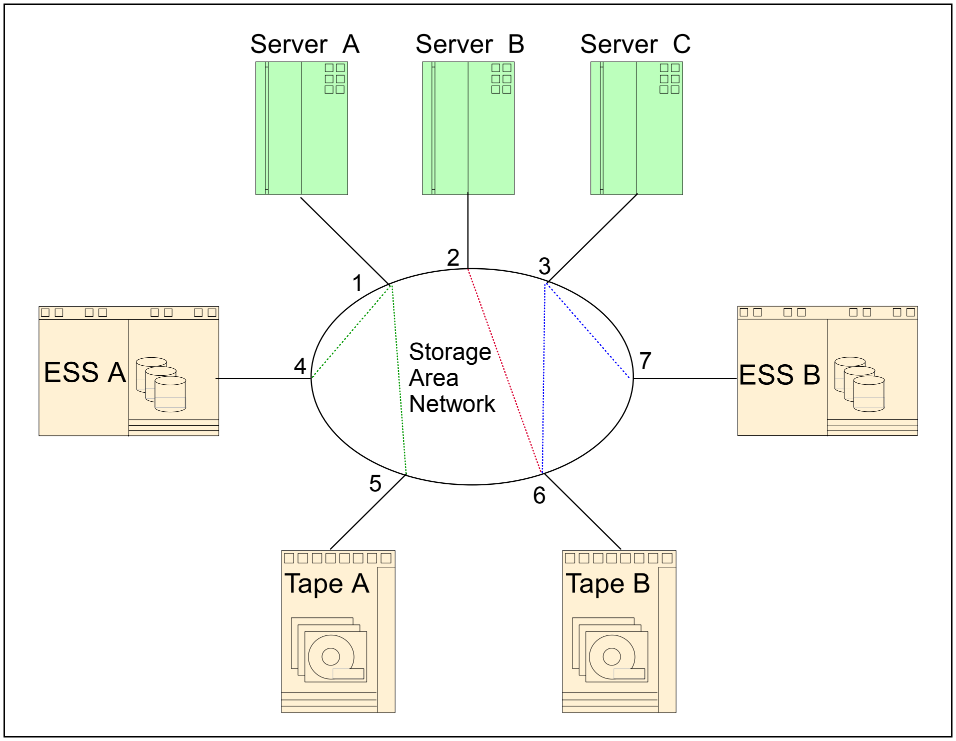


Figure 4-8 Zoning based on the switch port number

In this example, port-based zoning is used to restrict Server A to only see storage devices that are zoned to port 1: ports 4 and 5.

Server B is also zoned so that it can only see from port 2 to port 6.

Server C is zoned so that it can see both ports 6 and 7, even though port 6 is also a member of another zone.

A single port can also belong to multiple zones.

We show an example of hardware zoning in Figure 4-9 on page 86. This example illustrates another way of considering the hardware zoning as an array of connections.

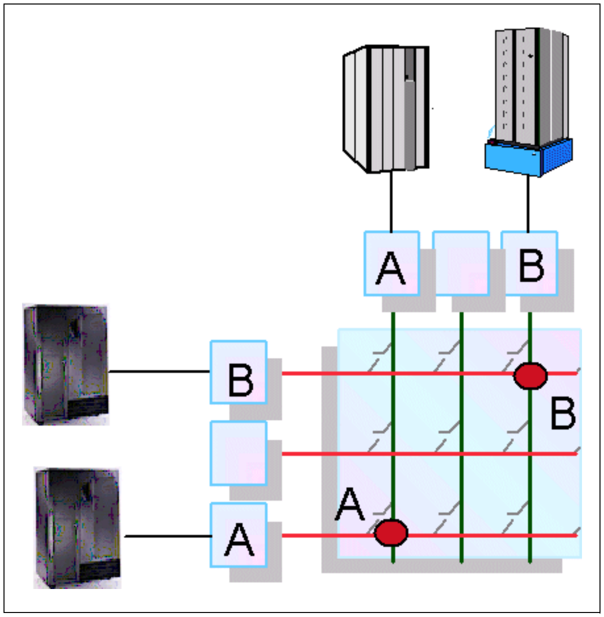


Figure 4-9 Hardware zoning

In this example, device A can only access storage device A through connection A. Device B can only access storage device B through connection B.

In a hardware-enforced zone, switch hardware, usually at the ASIC level, ensures that there is no data transferred between unauthorized zone members. However, devices can transfer data between ports within the same zone. Consequently, hard zoning provides the highest level of security. The availability of hardware-enforced zoning and the methods to create hardware-enforced zones depends on the switch hardware.

One of the disadvantages of hardware zoning is that devices have to be connected to a specific port, and the whole zoning configuration could become unusable when the device is connected to a different port. In cases where the device connections are not permanent, the use of software zoning is likely to make life easier.

The advantage of hardware zoning is that it can be implemented into a routing engine by filtering. As a result, this kind of zoning has a very low impact on the performance of the routing process.

If possible, the designer can include some unused ports in a hardware zone. So, in the event of a particular port failing, maybe caused by a GBIC or transceiver problem, the cable could be moved to a different port in the same zone. This would mean that the zone would not need to be reconfigured.

4.7.2 Software zoning

Software zoning is implemented by the fabric operating systems within the fabric switches. They are almost always implemented by a combination of the name server and the Fibre Channel Protocol. when a port contacts the name server, the name server will only reply with information about ports in the same zone as the requesting port. A soft zone, or software zone, is not enforced by hardware. What this means is that if a frame is incorrectly delivered (addressed) to a port that it was not intended to, then it will be delivered to that port. This is in contrast to hard zones.

When using software zoning the members of the zone can be defined using their World Wide Names:

- Node WWN
- Port WWN

Usually, zoning software also allows you to create symbolic names for the zone members and for the zones themselves. Dealing with the symbolic name or aliases for a device is often easier than trying to use the WWN address.

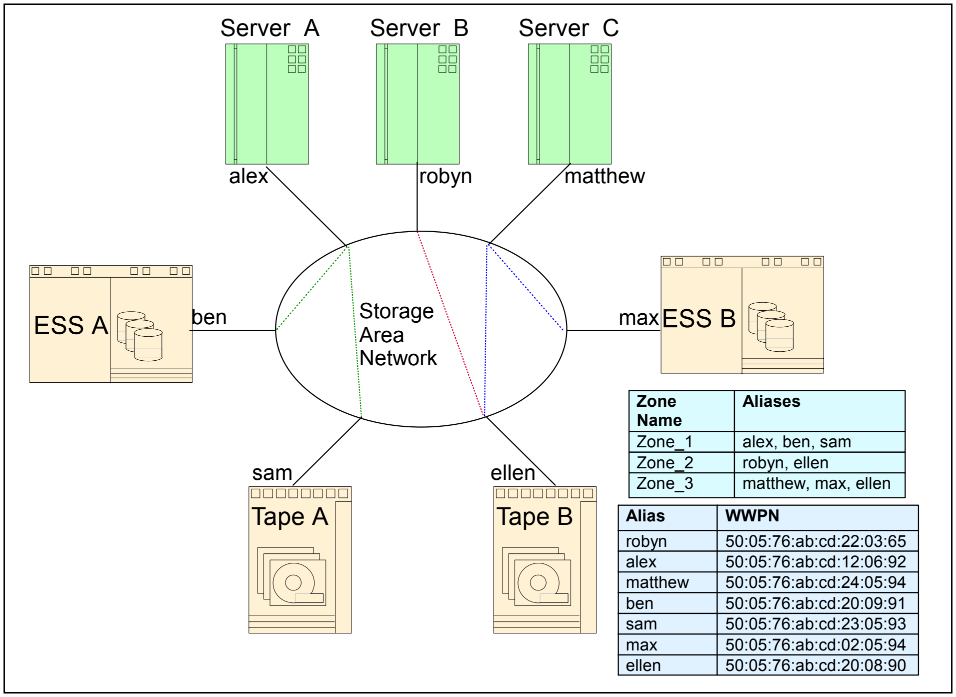
The number of members possible in a zone is limited only by the amount of memory in the fabric switch. A member can belong to multiple zones. You can define multiple sets of zones for the fabric, but only one set can be active at any time. You can activate another zone set any time you want, without the need to power down the switch.

With software zoning there is no need to worry about the physical connections to the switch. If you use WWNs for the zone members, even when a device is connected to another physical port, it will still remain in the same zoning definition, because the device’s WWN remains the same. The zone follows the WWN.

Important: However, by stating this, it does not automatically mean that if you unplug a device, such as a disk subsystem, and plug it into another switch port, that your host will still be able to communicate with your disks (until you either reboot or unload and load your operating system device definitions), even if the device remains member of that particular zone. This depends on components you use in your environment, like operating system and multipath software.

Shown in Figure 4-10 is an example of WWN-based zoning. In this example, symbolic names are defined for each WWN in the SAN to implement the same zoning requirements, as shown in the previous Figure 4-8 on page 85 for port zoning:

- Zone_1 contains the aliases alex, ben, and sam, and is restricted to only these devices.
- Zone_2 contains the aliases robyn and ellen, and is restricted to only these devices.
- Zone_3 contains the aliases matthew, max, and ellen, and is restricted to only these devices.



Zone Name	Aliases
Zone_1	alex, ben, sam
Zone_2	robyn, ellen
Zone_3	matthew, max, ellen

Alias	WWPN
robyn	50:05:76:ab:cd:22:03:65
alex	50:05:76:ab:cd:12:06:92
matthew	50:05:76:ab:cd:24:05:94
ben	50:05:76:ab:cd:20:09:91
sam	50:05:76:ab:cd:23:05:93
max	50:05:76:ab:cd:02:05:94
ellen	50:05:76:ab:cd:20:08:90

Figure 4-10 Zoning based on the devices' WWNs

There are some potential security leaks with software zoning:

- When a specific host logs into the fabric and asks for available storage devices, the simple name server (SNS) looks in the software zoning table to see which devices are allowable. The host only sees the storage devices defined in the software zoning table. But, the host can also make a direct connection to the storage device, using device discovery, without asking SNS for the information.

- It is possible for a device to define the WWN that it will use, rather than using the one designated by the manufacturer of the HBA. This is known as *WWN spoofing*. An unknown server could masquerade as a trusted server and thus gain access to data on a particular storage device. Some fabric operating systems allow the fabric administrator to prevent this risk by allowing the WWN to be tied to a particular port.
- Any device that does any form of probing for WWNs is able to discover devices and talk to them. A simple analogy is that of an unlisted telephone number. Although the telephone number is not publicly available, there is nothing to stop a person from dialing that number, whether by design or accident. The same holds true for WWN. There are devices that randomly probe for WWNs to see if they can start a conversation with them.

A number of switch vendors offer hardware-enforced WWN zoning, which can prevent this security exposure. Hardware-enforced zoning uses hardware mechanisms to restrict access rather than relying on the servers to follow the fibre channel protocols.

> **Note:** When a device logs in to a software-enforced zone, it queries the name server for devices within the fabric. If zoning is in effect, only the devices in the same zone or zones are returned. Other devices are hidden from the name server query reply. When using software-enforced zones, the switch does not control data transfer and there is no guarantee of data being transferred from unauthorized zone members. Use software zoning where flexibility and security are ensured by the cooperating hosts.

Frame filtering

Zoning is a fabric management service that can be used to create logical subsets of devices within a SAN and enable partitioning of resources for management and access control purposes. *Frame filtering* is another feature that enables devices to provide zoning functions with finer granularity. Frame filtering can be used to set up port-level zoning, world wide name zoning, device-level zoning, protocol-level zoning, and LUN-level zoning. Frame filtering is commonly performed by an ASIC. This has the result that, after the filter is set up, the complicated function of zoning and filtering can be achieved at wire speed.

4.7.3 LUN masking

The term *logical unit number* (LUN) was originally used to represent the entity within a SCSI target which executes I/Os. A single SCSI device usually only has a single LUN, but some devices, such as tape libraries, might have more than one LUN.

In the case of a storage array, the array makes virtual disks available to servers. These virtual disks are identified by LUNs.

It is absolutely possible for more than one host to see the same storage device or LUN. This is potentially a problem, both from a practical and a security perspective. Another approach to securing storage devices from hosts wanting to take over already assigned resources is logical unit number (LUN) masking. Every storage device offers its resources to the hosts by means of LUNs.

For example, each partition in the storage server has its own LUN. If the host server wants to access the storage, it needs to request access to the LUN in the storage device. The purpose of LUN masking is to control access to the LUNs. The storage device itself accepts or rejects access requests from different hosts.

The user defines which hosts can access which LUN by means of the storage device control program. Whenever the host accesses a particular LUN, the storage device checks its access list for that LUN, and it allows or disallows access to the LUN.

5

IP storage networking

Current SAN implementations are highly dependent on Fibre Channel technology. Anybody that remembers the “SAN versus NAS debate” may be feeling a sense of déjà vu at the moment though. (If you do not remember the SAN versus NAS debate, it is similar to the VHS versus Betamax debate in the video arena.) iSCSI, iFCP, and FCIP are prompting the same sort of debate in the marketplace. As with the SAN/NAS debate, each has an argument based on market direction, technical, architectural, and ease of adoption merits.

When any new technology is mooted or introduced into the marketplace there will be opinions formed, opinions divided, and sometimes even opinions in violent agreement.

So what are some of the divisions that have formed around these IP storage transport protocols? Firstly, let’s look at these technologies at a high level.

5.1 Fibre Channel over IP

Fibre Channel over IP (FCIP) is also known as Fibre Channel tunneling or storage tunneling. It is a method for allowing the transmission of Fibre Channel information to be tunneled through the IP network. Because most organizations already have an existing IP infrastructure, the attraction of being able to link geographically dispersed SANs, at relatively low cost, is enormous. FCIP encapsulates Fibre Channel block data and subsequently transports it over a

TCP socket, or tunnel. TCP/IP services are utilized to establish connectivity between remote SANs. Any congestion control and management, as well as data error and data loss recovery, is handled by TCP/IP services and does not affect FC fabric services. The major point with FCIP is that is does not replace FC with IP, it simply allows deployments of FC fabrics using IP tunneling. The assumption that this might lead to is that the “industry” has decided that FC based SANs are more than appropriate, and that the only need for the IP connection is to facilitate any distance requirement that is beyond the current scope of an FCP SAN.

The main advantages are that FCIP overcomes the distance limitations of native Fibre Channel and enables geographically distributed SANs to be linked using the existing IP infrastructure, while keeping the fabric services intact.

5.2 iFCP

Internet Fibre Channel Protocol (iFCP) is a mechanism for transmitting data to and from Fibre Channel storage devices in a SAN, or on the Internet using TCP/IP. iFCP gives the ability to incorporate already existing SCSI and Fibre Channel networks into the Internet. iFCP is able to be used in tandem with existing Fibre Channel protocols, such as FCIP, or it can replace them. Whereas FCIP is a tunneled solution, iFCP is an FCP routed solution. A significant difference to FCIP is that iFCP actually replaces the lower-layer Fibre Channel transport with TCP/IP and Gigabit Ethernet. With iFCP, Fibre Channel devices connect to an iFCP gateway or switch and each Fibre Channel session is terminated at the local gateway and converted to a TCP/IP session via iFCP. A second gateway or switch receives the iFCP session and initiates a Fibre Channel session.

The appeal of iFCP is that for customers that have a wide range of FC devices, and who want to be able to connect these with the IP network, iFCP gives the ability to permit this. iFCP can interconnect FC SANs with IP networks, and also allows customers to use the TCP/IP network in place of the SAN. iFCP is a gateway-to-gateway protocol, and does not simply encapsulate FC block data. Gateway devices are used as the medium between the FC initiators and targets. As these gateways can either replace or be used in tandem with existing FC fabrics, iFCP could be used to help migration from a Fibre Channel SAN to an IP SAN, or allow a combination of both.

The iFCP protocol uses TCP/IP switching and routing elements to complement and enhance, or replace, Fibre Channel SAN components. This enables existing Fibre Channel storage devices or SANs to attach to an IP network as each is assigned an IP address. iFCP can replace, or be used in conjunction with, existing technologies such as FCIP, as it addresses some issues that FCIP cannot handle. For example, FCIP only works within a Fibre Channel environment, whereas iFCP can be implemented in Fibre Channel or hybrid networks.

The main advantages of iFCP are that it overcomes distance limitations, allows distributed SANs to be connected, and integrates FC capable devices into a familiar IP infrastructure.

5.3 iSCSI

Internet SCSI (iSCSI) is a transport protocol that carries SCSI commands from an initiator to a target. It is a data storage networking protocol that transports standard Small Computer System Interface (SCSI) requests over the standard Transmission Control Protocol/Internet Protocol (TCP/IP) networking technology. SCSI data and commands are encapsulated in TCP/IP packets. iSCSI enables the implementation of IP-based storage area networks (SANs), enabling customers to use the same networking technologies—from the box level to the Internet—for both storage and data networks. Taking this approach enables users to get full access to IP management and routing, and, as it uses TCP/IP, iSCSI is also well suited to run over almost any physical network. By eliminating the need for a second network technology just for storage, iSCSI can lower the costs of deploying networked storage and increase its potential market. ISCSI is a native IP interconnect that wraps SCSI data and commands in TCP/IP packets. Latency, introduced when putting storage on the same route as other network traffic, and security have been big concerns for iSCSI. The receiving device takes the command out of the IP packet and passes it to the SCSI controller, which forwards the request to the storage device. Once the data is retrieved, they are again wrapped in an IP packet and returned to the requesting device.

The main advantages of iSCSI are, when compared to a Fibre Channel installation, it is a low cost implementation, and there are no distance limitations.

5.4 The FCIP, iFCP or iSCSI conundrum

Now that we have introduced the protocols at a high level, what are the strategic differences between them all? Do I need them all, any, or none? What are some of the benefits that these technologies will bring me? Some of the benefits that can be realized are:

- Departmental isolation and resource sharing alleviation
- Technology migration and integration
- Remote replication of disk systems
- Remote access to disk and tape systems

- Low-cost connection to SANs
- Inter fabric routing
- Overcoming distance limitations

In these early days of protocol introduction, and before market adoption decides whether there will be a dominant protocol, it is likely that take-up will be somewhat slow and hesitant. The reason for this is that people do not want to make any large financial investment without any guarantees that there will be some form of return. However, the beauty of these protocols is that they immediately bring benefits. As these are standards based protocols they allow the leveraging of both the existing TCP/IP and FCP infrastructure, they support existing FC devices, and enable simplification of the infrastructure by removing any SAN islands.

There is a school of thought that also sees iFCP and iSCSI as a replacement for Fibre Channel fabrics, but this remains to be seen. For the discussion and introduction of these protocols within this redbook, it is immaterial to us one way or the other as to if/when Fibre Channel fabrics will or won't be replaced.

It would be unwise to start to write off Fibre Channel just yet. It is certain that a number of protocols will each exist in the storage architecture marketplace, each complementing, and sometimes competing.

5.5 The multiprotocol environment

As with any technology it comes with its unique jargon and terminology. Typically it is borrowed from the networking world, but may have a different meaning. It is not our intent to cover each and every unique description, but we will make some distinctions that we feel are important for a basic introduction to routing in an IP SAN.

5.5.1 Fibre Channel switching

A Fibre Channel switch filters and forwards packets between Fibre Channel connections on the *same* fabric, but it cannot transmit packets between fabrics. As soon as you join two switches together, you merge the two fabrics into a single fabric with one set of fabric services.

5.5.2 Fibre Channel routing

A router forwards data packets *between* two or more fabrics. Routers use headers and forwarding tables to determine the best path for forwarding the packets.

Separate fabrics each have their own addressing schemes. When they are joined by a router, there must be a way to translate the addresses between the two fabrics. This mechanism is called *network address translation* (NAT) and is inherent in all the IBM System Storage™ multiprotocol switch/router products. It is sometimes referred to as FC-NAT to differentiate it from a similar mechanism which exists in IP routers.

5.5.3 Tunneling

Tunneling is a technique that allows one network to send its data via another network's connections. Tunneling works by encapsulating a network protocol within packets carried by the second network. For example, in a Fibre Channel over Internet Protocol (FCIP) solution, Fibre Channel packets can be encapsulated inside IP packets. Tunneling raises issues of packet size, compression, out-of-order packet delivery, and congestion control.

5.5.4 Routers and gateways

When a Fibre Channel router needs to provide protocol conversion or tunneling services, it is a *gateway* rather than a router. However, it has become common usage to broaden the term *router* to include these functions. FCIP is an example of tunneling, while Small Computer System Interface over IP (iSCSI) and Internet Fibre Channel Protocol (iFCP) are examples of protocol conversion.

5.5.5 Internet Storage Name Service

The Internet Storage Name Service (iSNS) protocol facilitates automated discovery, management, and configuration of iSCSI and Fibre Channel devices that exist on a TCP/IP network. iSNS provides storage discovery and management services comparable to those that are found in Fibre Channel networks. What this means is that the IP network appears to operate in a similar capacity as a SAN. Coupling this with its ability to emulate Fibre Channel fabric services, iSNS allows for a transparent integration of IP and Fibre Channel networks as it can manage both iSCSI and Fibre Channel devices.

5.6 Deeper into the protocols

We have introduced all the protocols at a high level. Now we will show how they do what they do with the Fibre Channel traffic in greater depth.

5.6.1 FCIP

FCIP is a method for tunneling Fibre Channel packets through an IP network. FCIP encapsulates Fibre Channel block data and transports it over a TCP socket, or tunnel. TCP/IP services are used to establish connectivity between remote devices. The Fibre Channel packets are not altered in any way. They are simply encapsulated in IP and transmitted.

Figure 5-1 shows FCIP tunneling, assuming that the Fibre Channel packet is small enough to fit inside a single IP packet.

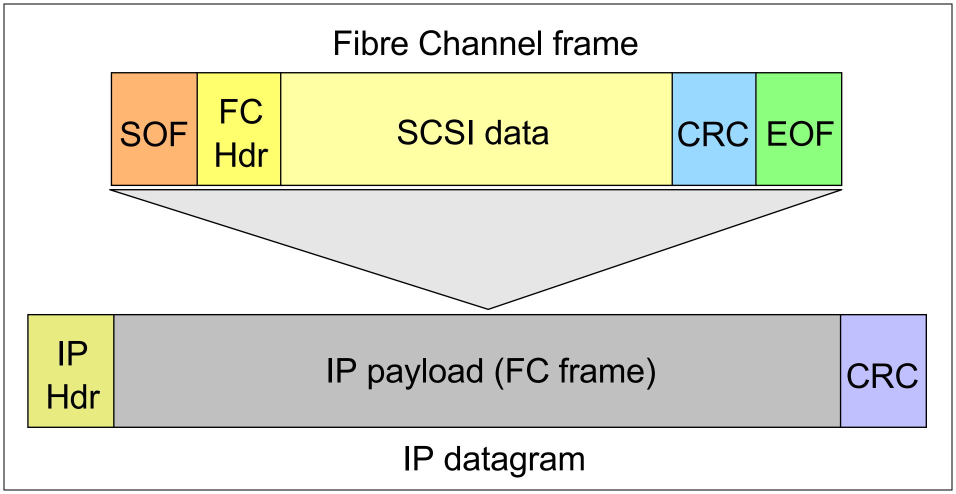


Figure 5-1 FCIP encapsulates the Fibre Channel frame into IP packets

The main advantage is that FCIP overcomes the distance limitations of native Fibre Channel. It also enables geographically distributed devices to be linked using the existing IP infrastructure, while keeping fabric services intact.

The architecture of FCIP is outlined in the Internet Engineering Task Force (IETF) Request for Comment (RFC) 3821, “Fibre Channel over TCP/IP (FCIP)”, available on the Web at:

http://ietf.org/rfc/rfc3821.txt

Because FCIP simply tunnels Fibre Channel, creating an FCIP link is like creating an inter-switch link (ISL), and the two fabrics at either end are merged into a single fabric. This creates issues in situations where you do not want to merge the two fabrics for business reasons, or where the link connection is prone to occasional fluctuations.

Many corporate IP links are robust, but it can be difficult to be sure because traditional IP-based applications tend to be retry-tolerant. Fibre Channel fabric services are not as retry-tolerant. Each time the link disappears or reappears, the switches re-negotiate and the fabric is reconfigured.

By combining FCIP with FC-FC routing, the two fabrics can be left “un-merged”, each with its own separate Fibre Channel services.

5.6.2 iFCP

iFCP is a gateway-to-gateway protocol. It provides Fibre Channel fabric services to Fibre Channel devices over a TCP/IP network. iFCP uses TCP to provide congestion control, error detection, and recovery. iFCP’s primary purpose allows interconnection and networking of existing Fibre Channel devices at wire speeds over a IP network.

Under iFCP, IP components and technology replace the Fibre Channel switching and routing infrastructure. iFCP was originally developed by Nishan Systems who were acquired by McDATA in September 2003.

To learn more about the architecture and specification of iFCP, refer to the document at the following IETF Web site:

http://www.ietf.org/internet-drafts/draft-ietf-ips-ifcp-14.txt

There is a popular myth that iFCP does not use encapsulation. In fact, iFCP encapsulates the Fibre Channel packet in much the same way that FCIP does. In addition, it maps the Fibre Channel header to the IP header and a TCP session, as shown in Figure 5-2.

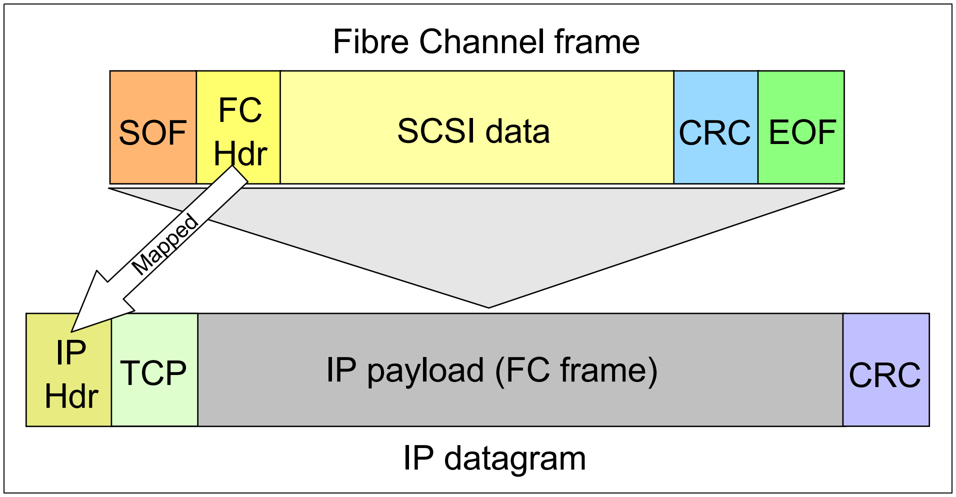


Figure 5-2 iFCP encapsulation and header mapping iFCP uses the same Internet Storage Name Server (iSNS) mechanism that is used by iSCSI.

iFCP also allows data to fall across IP packets and share IP packets. Some FCIP implementations can achieve a similar result when running software compression, but not otherwise. FCIP typically break each large Fibre Channel packet into two dedicated IP packets. iFCP compression is payload compression only. Headers are not compressed to simplify diagnostics.

iFCP uses one TCP connection per fabric login (FLOGI), while FCIP typically uses one connection per router link (although more are possible). A FLOGI is the process by which an N_PORT registers its presence on the fabric, obtains fabric parameters such as classes of service supported, and receives its N_PORT address. Because under iFCP there is a separate TCP connection per N_PORT to N_PORT couple, each connection can be managed to have its own Quality of Service (QoS) identity. A single incidence of congestion does not need to drop the sending rate for all connections on the link.

While all iFCP traffic between a given remote and local N_PORT pair must use the same iFCP session, that iFCP session can be shared across multiple gateways or routers.

5.6.3 iSCSI

The Small Computer Systems Interface (SCSI) protocol has a client/server architecture. Clients (called *initiators*) issue SCSI commands to request services from logical units on a server known as a *target*. A SCSI *transport* maps the protocol to a specific interconnect.

The SCSI protocol has been mapped over various transports, including Parallel SCSI, Intelligent Peripheral Interface (IPI), IEEE-1394 (firewire), and Fibre Channel. All of these transports are ways to pass SCSI commands. Each transport is I/O specific and has limited distance capabilities.

The iSCSI protocol is a means of transporting SCSI packets over TCP/IP to take advantage of the existing Internet infrastructure.

A session between a iSCSI initiator and an iSCSI target is defined by a session ID that is a combination of an initiator part (ISID) and a target part (Target Portal Group Tag).

The iSCSI transfer direction is defined with respect to the initiator. Outbound or outgoing transfers are transfers from an initiator to a target. Inbound or incoming transfers are transfers from a target to an initiator.

For performance reasons, iSCSI allows a “phase-collapse”. A command and its associated data may be shipped together from initiator to target, and data and responses may be shipped together from targets.

An iSCSI name specifies a logical initiator or target. It is not tied to a port or hardware adapter. When multiple network interface cards (NICs) are used, they should generally all present the same iSCSI initiator name to the targets, because they are simply paths to the same SCSI layer. In most operating systems, the named entity is the operating system image.

The architecture of iSCSI is outlined in IETF RFC 3720, “Internet Small Computer Systems Interface (iSCSI)”, which you can find on the Web at:

http://www.ietf.org/rfc/rfc3720.txt

Figure 5-3 shows the format of the iSCSI packet.

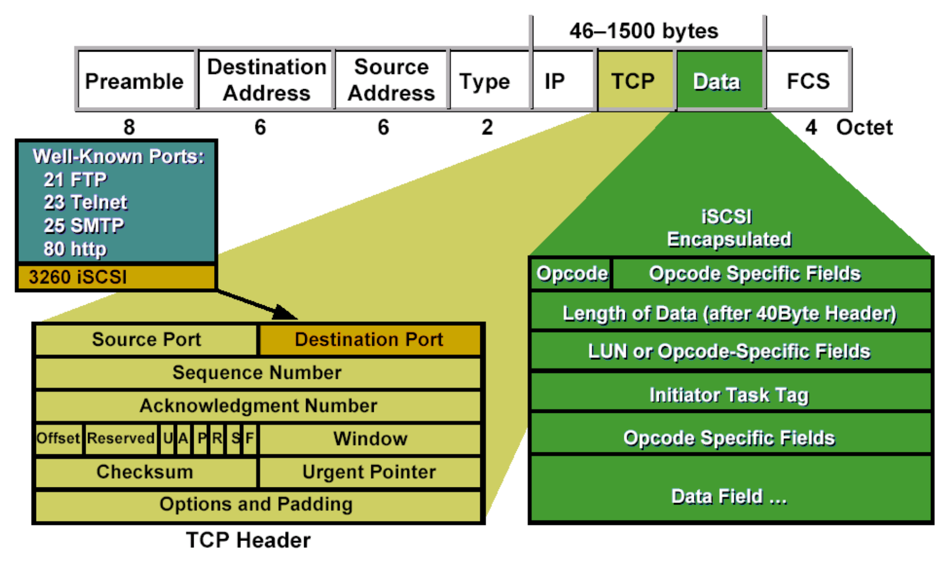


Figure 5-3 iSCSI packet format

Testing on iSCSI latency has shown a difference of up to 1 ms of additional latency for each disk I/O as compared to Fibre Channel. This does not include such factors as trying to do iSCSI I/O over a shared, congested or long-distance IP network, all of which may be tempting for some customers. iSCSI generally uses a shared 1 Gbps network.

iSCSI naming and discovery

Although we do not propose to go for an iSCSI deep dive in this redbook, there are three ways for an iSCSI initiator to understand which devices are in the network:

- In small networks, you can use the **`sendtargets`** command.
- In larger networks, you can use the Service Location Protocol (SLP, multicast discovery).
- In large networks, we recommend that you use Internet Storage Name Service (iSNS).

> **Note:** At time of writing, not all vendors' have delivered iSNS.

You can find a range of drafts that cover iSCSI naming, discovery, and booting on the following Web site:

http://www.ietf.org/proceedings/02mar/220.htm

5.7 Routing considerations

As you would expect with any technology there are going to be a unique set of characteristics that need to be given consideration. The topics that follow briefly describe some of the issues, or items, that are considerations in a multiprotocol Fibre Channel environment.

5.7.1 Packet size

The standard size of a Fibre Channel packet is 2148 bytes, and the standard IP packet size is 1500 bytes (with a 1460 byte payload). It does not take an Einstein to work out that one is larger than the other and will need to be accommodated somehow.

When transporting Fibre Channel over IP, you can use jumbo IP packets to accommodate larger Fibre Channel packets. Keep in mind that jumbo IP packets must be turned on for the whole data path. In addition, a jumbo IP packet is not compatible with any devices in the network that do not have jumbo IP packets enabled.

Alternatively, you can introduce a variety of schemes to split Fibre Channel packets across two IP packets. Some compression algorithms can allow multiple small Fibre Channel packets or packet segments to share a single IP packet.

Each technology and each vendor may implement this differently. But the key point is this: they all try to avoid sending small inefficient packets.

5.7.2 TCP congestion control

Sometimes standard TCP congestion mechanisms may not be suitable for tunneling storage. Standard TCP congestion control is designed to react quickly and severely to network congestion, but recover slowly. This is well suited to traditional IP networks being somewhat variable and unreliable. But for storage applications, this approach is not always appropriate and may cause disruption to latency-sensitive applications.

When three duplicate unanswered packets are sent on a traditional TCP network, the sending rate backs-off by 50%. When packets are successfully sent, it does a slow-start linear ramp-up again.

Some vendors tweak the back-off and recovery algorithms. For example, the tweak causes the send rate to drop by 12.5% each time congestion is encountered, and then to recover rapidly to the full sending rate by doubling each time until full rate is regained.

Other vendors take a simpler approach to achieve much the same end.

If you are sharing your IP link between storage and other IP applications, then using either of these storage friendly congestion controls may impact your other applications.

You can find the specification for TCP congestion control on the Web at:

`http://www.ietf.org/rfc/rfc2581.txt`

5.7.3 Round-trip delay

Round-trip link latency is the time it takes for a packet to make a round-trip across the link. The term *propagation delay* is also sometime used. Round-trip delay generally includes both inherent latency and delays due to congestion.

Fibre Channel cable has an inherent latency of approximately five microseconds per kilometer each way. Typical Fibre Channel devices, like switches and routers, have inherent latencies of around five microseconds each way. IP routers might vary between five and one hundred microseconds in theory, but when tested with filters applied, the results are more likely to be measured in milliseconds.

This is the essential problem with tunneling Fibre Channel over IP. Fibre Channel applications are generally designed for networks that have round-trip delays measured in microseconds. IP networks generally deliver round-trip delays measured in milliseconds or tens of milliseconds. Internet connections often have round-trip delays measured in hundreds of milliseconds.

Any round-trip delay caused by additional routers and firewalls along the network connection also needs to be added to the total delay. The total round-trip delay varies considerably depending on the models of routers or firewalls used, and the traffic congestion on the link.

So how does this affect you? If you are purchasing the routers or firewalls yourself, we recommend that you include the latency of any particular product in the criteria that you use to choose the products. If you are provisioning the link from a service provider, we recommend that you include at least the maximum total round-trip latency of the link in the service-level agreement (SLA).

Time of frame in transit

The time of frame in transit is the actual time that it takes for a given frame to pass through the slowest point of the link. Therefore it depends on both the frame size and link speed.

The maximum size of the payload in a Fibre Channel frame is 2112 bytes. The Fibre Channel headers add 36 bytes to this, for a total Fibre Channel frame size of 2148 bytes. When transferring data, Fibre Channel frames at or near the full size are usually used.

If we assume that we are using jumbo frames in the Ethernet, the complete Fibre Channel frame can be sent within one Ethernet packet. The TCP and IP headers and the Ethernet medium access control (MAC) add a minimum of 54 bytes to the size of the frame, giving a total Ethernet packet size of 2202 bytes, or 17616 bits.

For smaller frames, such as the Fibre Channel acknowledgement frames, the time in transit is much shorter. The minimum possible Fibre Channel frame is one with no payload. With FCIP encapsulation, the minimum size of a packet with only the headers is 90 bytes, or 720 bits.

Table 5-1 details the transmission times of this FCIP packet over some common wide area network (WAN) link speeds.

Table 5-1 FCIP packet transmission times over different WAN links

Link type	Link speed	Large packet	Small packet
Gigabit Ethernet	1250 Mbps	14 µs	0.6 µs
OC-12	622.08 Mbps	28 µs	1.2 µs
OC-3	155.52 Mbps	113 µs	4.7 µs
T3	44.736 Mbps	394 µs	16.5 µs
E1	2.048 Mbps	8600 µs	359 µs

Link type	Link speed	Large packet	Small packet
T1	1.544 Mbps	11 400 µs	477 µs

If we cannot use jumbo frames, each large Fibre Channel frame needs to be divided into two Ethernet packets. This doubles the amount of TCP, IP, and Ethernet MAC overhead for the data transfer.

Normally each Fibre Channel operation transfers data in only one direction. The frames going in the other direction are close to the minimum size.

5.7.4 Write acceleration

Write acceleration, or Fast Write as it is sometimes called, is designed to mitigate the problem of the high latency of long distance networks. Write acceleration eliminates the time spent waiting for a target to tell the sender that it is ready to receive data. The idea is to send the data before receiving the ready signal, knowing that the ready signal will almost certainly arrive as planned. Data integrity is not jeopardized because the write is not assumed to have been successful until the final acknowledgement has been received.

Figure 5-4 shows a standard write request.

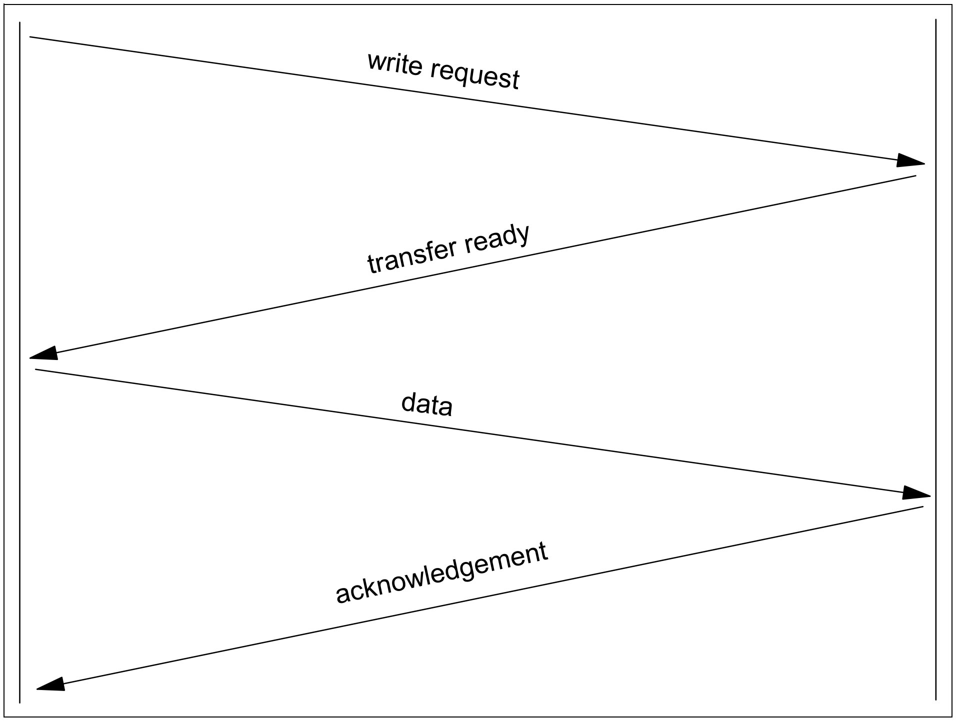


Figure 5-4 A standard write request

Figure 5-5 shows an accelerated write request.

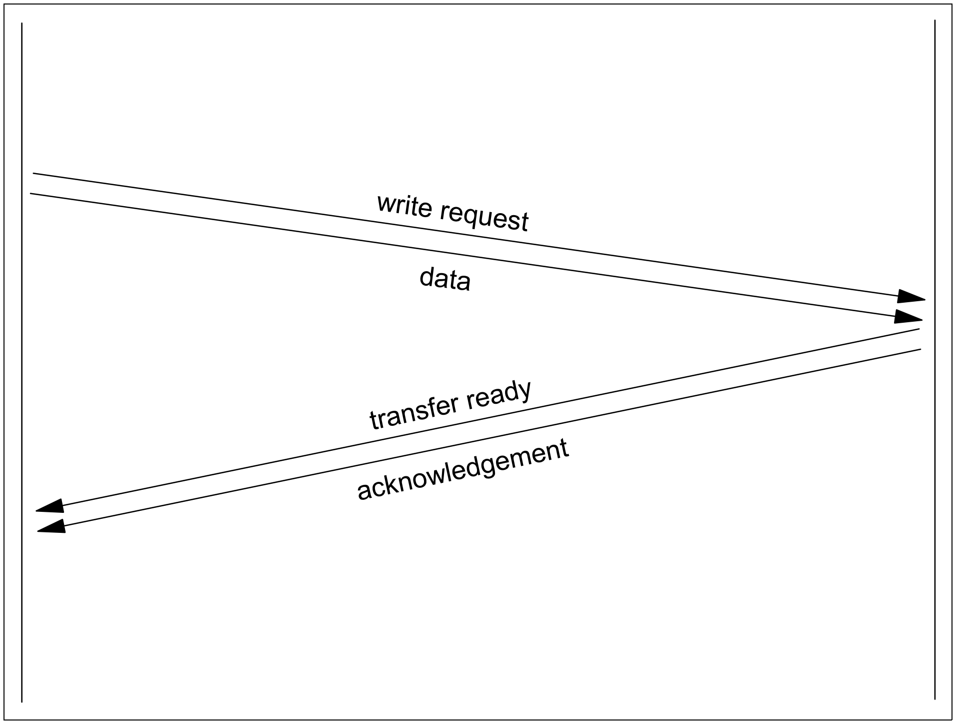


Figure 5-5 Write acceleration or Fast Write request

5.7.5 Tape acceleration

Tape acceleration (TA) takes write acceleration one step further by “spoofing” the transfer ready and the write acknowledgement. This gives the tape transfer a better chance of streaming rather than running stop and start. The risk here is that writes have been acknowledged but may not have completed successfully.

Without tape acceleration a sophisticated backup/restore application, such as Tivoli® Storage Manager, can recover and restart from a broken link. However, with TA, Tivoli Storage Manager believes that any write for which it has received an acknowledgement must have completed successfully. The restart point is therefore set after that last acknowledgement. With TA, that acknowledgement was spoofed so it might not reflect the real status of that write.

Tape acceleration provides faster tape writes at the cost of recoverability. While the write acknowledgments are spoofed, the writing of the final tape mark is never spoofed. This provides some degree of integrity control when using TA.

Figure 5-6 shows how you can use tape acceleration to improve data streaming.

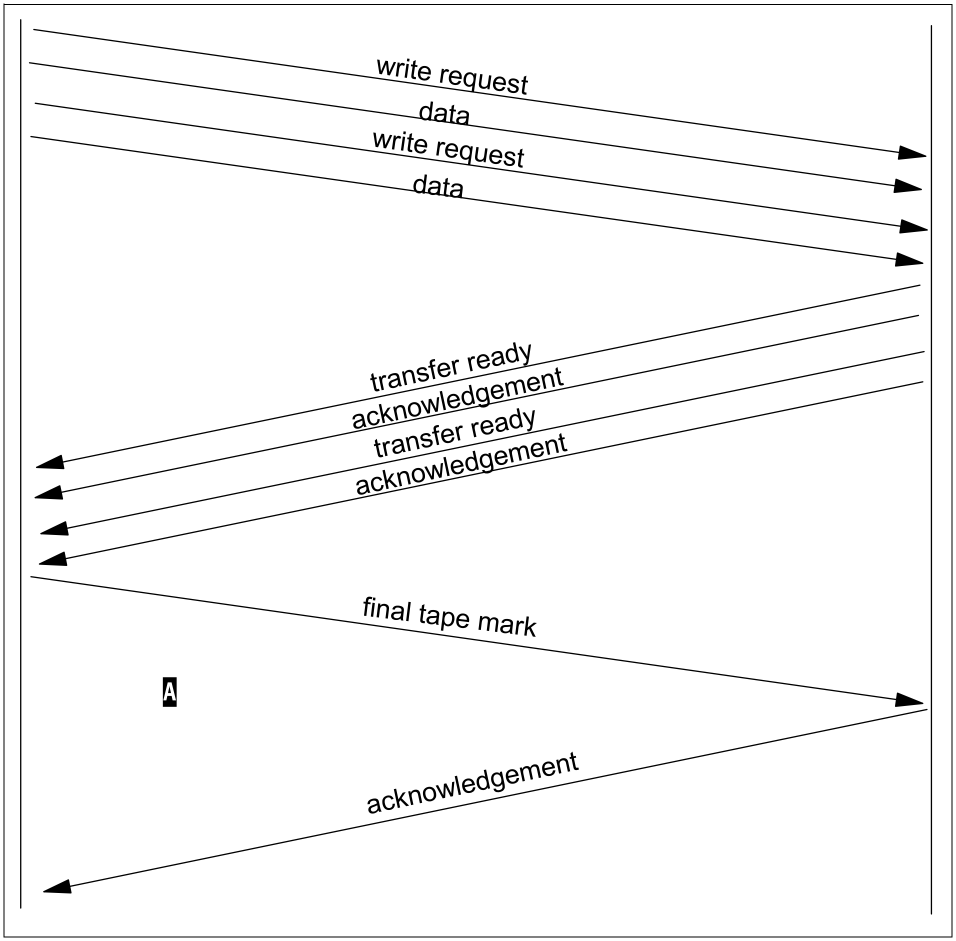


Figure 5-6 Tape acceleration example

5.8 Multiprotocol solution briefs

The solution briefs in the following sections show how you can use multiprotocol routers.

5.8.1 Dividing a fabric into sub-fabrics

Let us suppose that you have eight switches in your data center, and they are grouped into two fabrics of four switches each. Two of the switches are used to connect the development/test environment, two are used to connect a joint-venture subsidiary company, and four are used to connect the main production environment.

The development/test environment does not follow the same change control disciplines as the production environment. Also systems and switches can be upgraded, downgraded, or rebooted on occasions, usually unscheduled and without any form of warning.

The joint-venture subsidiary company is up for sale. The mandate is to provide as much separation and security as possible between it and the main company, and the subsidiary. The backup/restore environment is shared between the three environments.

In summary, we have a requirement to provide a degree of isolation, and a degree of sharing. In the past this would have been accommodated through zoning. Some fabric vendors may still recommend that approach as the simplest and most cost-effective. However as the complexity of the environment grows, zoning can become complex. Any mistakes in setup can disrupt the entire fabric. Adding FC-FC routing to the network allows each of the three environments to run separate fabric services and provides the capability to share the tape backup environment.

In larger fabrics with many switches and separate business units, for example in a shared services hosting environment, separation and routing are valuable in creating a larger number of simple fabrics, rather than fewer more complex fabrics.

5.8.2 Connecting a remote site over IP

Suppose you want to replicate your disk system to a remote site, perhaps 50 km away synchronously, or 500 km away asynchronously. Using FCIP tunneling or iFCP conversion, you can transmit your data to the remote disk system over a standard IP network. The router includes Fibre Channel ports to connect back-end devices or switches and IP ports to connect to a standard IP wide area network router. Standard IP networks are generally much lower in cost to provision than traditional high quality dedicated dense wavelength division multiplexing (DWDM) networks. They also often have the advantage of being well understood by internal operational staff.

Similarly you might want to provision storage volumes from your disk system to a remote site by using FCIP or iFCP.

Note: FCIP and iFCP can provide a low cost way to connect remote sites using familiar IP network disciplines.

5.8.3 Connecting hosts using iSCSI

Many hosts do not require high bandwidth low latency access to storage. For such hosts, iSCSI may be a more cost-effective connection method. iSCSI can be thought of as an IP SAN. There is no requirement to provide a Fibre Channel switch port for every server, nor to purchase Fibre Channel host bus adapters (HBAs), nor to lay Fibre Channel cable between storage and servers.

The iSCSI router has both Fibre Channel ports and Ethernet ports to connect to servers located either locally on the Ethernet or remotely over a standard IP wide area network connection.

The iSCSI connection delivers block I/O access to the server so it is application independent. That is, an application cannot really tell the difference between direct SCSI, iSCSI, or Fibre Channel, since all three are delivery SCSI block I/Os.

Different router vendors quote different limits on the number of iSCSI connections that are supported on a single IP port.

iSCSI places a significant packetizing and depacketizing workload on the server CPU. This can be mitigated by using TCP/IP offload engine (TOE) Ethernet cards. However since these cards can be expensive, they somewhat undermine the low-cost advantage of iSCSI.

Note: iSCSI can be used to provide low-cost connections to the SAN for servers that are not performance critical.

6

Fibre Channel products and technology

In this chapter we describe the Fibre Channel SAN products and technology that are encountered. For a description of the IBM products that have been inducted into the IBM System Storage and TotalStorage portfolio, refer to Chapter 9, “The IBM product portfolio” on page 175.

6.1 The environment

Historically, interfaces to storage consisted of parallel bus architectures (such as SCSI and IBM bus and tag) that supported a small number of devices. Fibre Channel technology provides a means to implement robust storage networks that may consist of hundreds or thousands of devices. Fibre Channel SANs support high-bandwidth storage traffic, at the time of writing up to 10 Gbps.

Storage subsystems, storage devices, and server systems can be attached to a Fibre Channel SAN. Depending on the implementation, several different components can be used to build a SAN. It is, as the name suggests, a network so any combination of devices that are able to interoperate are likely to be utilized.

Given this, a Fibre Channel network may be composed of many different types of interconnect entities, including directors, switches, hubs, routers, gateways, and bridges.

It is the deployment of these different types of interconnect entities that allow Fibre Channel networks of varying scale to be built. In smaller SAN environments you can employ hubs for Fibre Channel arbitrated loop topologies, or switches and directors for Fibre Channel switched fabric topologies. As SANs increase in size and complexity, Fibre Channel directors can be introduced to facilitate a more flexible and fault tolerant configuration. Each of the components that compose a Fibre Channel SAN should provide an individual management capability, as well as participate in an often complex end-to-end management environment.

6.2 SAN devices

A Fibre Channel SAN employs a fabric to connect devices, or end points. A fabric can be as simple as a single cable connecting two devices, akin to server attached storage. However, the term is most often used to describe a more complex network to connect servers and storage utilizing switches, directors, and gateways.

Independent from the size of the fabric, a good SAN environment starts with good planning, and always includes an up-to-date map of the SAN.

Some of the items to consider are:

- How many ports do I need now?
- How fast will I grow in two years?
- Are my servers and storage in the same building?

- Do I need long distance solutions?
- Do I need redundancy for every server or storage?
- How high are my availability needs and expectations?
- Will I connect multiple platforms to the same fabric?

We show a high-level view of a fabric in Figure 6-1.

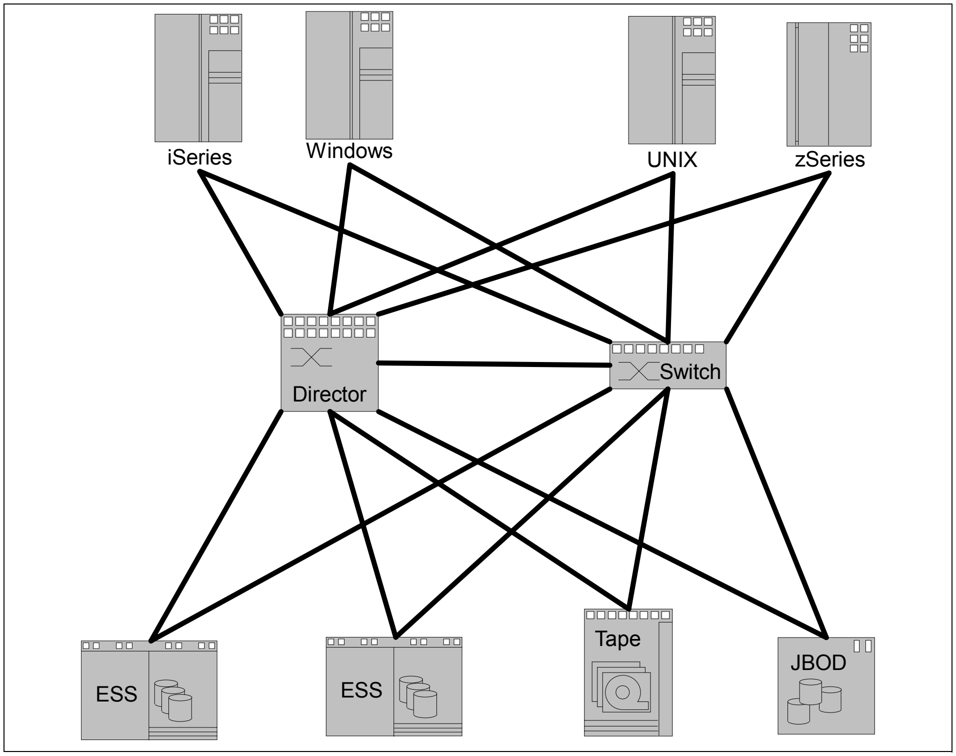


Figure 6-1 High-level view of a fabric

6.2.1 Bridges and gateways

A bridge is a device that converts signals and data from one form to another. In the specific case of SAN, a bridge is a unit converting between Fibre Channel and legacy (or existing) storage protocols such as:

- SCSI
- SSA

The offerings from IBM are called SAN Data Gateways or SAN Data Gateway Routers. Depending on the particular bridge, it may be possible to have only a single Fibre Channel port, whereas some will support multiple ports.

6.2.2 Arbitrated loop hubs

FC-AL topology allows devices be connected using discreet cabling, or an Arbitrated Loop hub.

In FC-AL all devices on the loop share the bandwidth. The total number of devices that may participate in the loop is 126, without using any hubs or fabric. For practical reasons, however, the number tends to be limited to no more than 10 and 15.

Hubs are typically used in a SAN to attach devices or servers that do not support switched fabric only FC-AL. They may be unmanaged hubs, managed hubs, or switched hubs.

Unmanaged hubs serve as cable concentrators and as a means to configure the Arbitrated Loop based on the connections it detects. When any of the hub's interfaces, usually GBIC, senses no cable connected, that interface shuts down and the hub port is bypassed as part of the Arbitrated Loop configuration.

Managed hubs offer all the benefits of unmanaged hubs, but in addition offer the ability to manage them remotely, using SNMP.

6.2.3 Switched hubs

Switched hubs allow devices to be connected in its own Arbitrated Loop. These loops are then internally connected by a switched fabric.

A switched hub is useful to connect several FC-AL devices together, but to allow them to communicate at full Fibre Channel bandwidth rather than them all sharing the bandwidth.

Switched hubs are usually managed hubs.

Note: In its early days, FC-AL was described as "SCSI on steroids". Although FC-AL has the bandwidth advantage over SCSI, it does not come anywhere close to the speeds that can be achieved and sustained on a per port basis in a switched fabric. For this reason, FC-AL implementations are, by some observers, considered as legacy SANs.

6.2.4 Switches and directors

Switches and directors allow Fibre Channel devices to be connected (cascaded) together, implementing a switched fabric topology between them. The switch intelligently routes frames from the initiator to responder and operates at full Fibre Channel bandwidth.

It is possible to connect switches together in cascades and meshes using inter-switch links (E_Ports). It should be noted that devices from different manufacturers may not interoperate fully.

The switch also provides a variety of fabric services and features such as:

- Name service
- Fabric control
- Time service
- Automatic discovery and registration of host and storage devices
- Rerouting of frames, if possible, in the event of a port problem
- Storage services (virtualization, replication, extended distances)

It is common to refer to switches as either *core* switches or *edge* switches depending on where they are located in the SAN. If the switch forms, or is part of the SAN backbone, then it is the core switch. If it is mainly used to connect to hosts or storage then it is called an edge switch. Like it or not, directors are also sometimes referred to as switches. Whether this is appropriate or not is a matter for debate outside of this book.

6.2.5 Multiprotocol routing

Beginning to make an impact on the market are devices that are multiprotocol routers and devices. These provide improved scalability, security, and manageability by enabling devices in separate SAN fabrics to communicate *without* merging fabrics into a single, large SAN fabric. Depending on the manufacturer, they support a number of protocols and have their own features, such as zoning. As their name suggests, the protocols supported include:

- FCP
- FCIP
- iFCP
- iSCSI
- IP

6.2.6 Service modules

Increasingly, with the demand for the intermix of protocols and the introduction to the marketplace of new technologies, SAN vendors are starting to adopt a modular system approach to their components. What this means is that service modules can be plugged into a slot on the switch or director to provide functions and features such as virtualization, the combining of protocols, storage services, and so on.

6.2.7 Multiplexers

Multiplexing is the process of simultaneously transmitting multiple signals over the same physical connection. There are two common types of multiplexing used for fiber optic connections based on either time or wavelength:

- Time Division Multiplexing (TDM)
- Wavelength Division Multiplexing (WDM)
- Dense Wavelength Division Multiplexing (DWDM)

6.2.8 Storage considered as legacy

In the context of a SAN, legacy (or existing) equipment consists of devices that do not inherently support Fibre Channel. As an example, SCSI disk arrays, tape drives, and SSA devices may be considered as legacy (or existing) equipment.

In order to protect your investment, it is often a requirement that legacy (or existing) equipment gets reused after the implementation of SAN.

A bridge, router, or gateway device is used to convert between these protocols. These have Fibre Channel connectivity offering connections to the legacy (or existing) equipment at the same time.

6.3 Componentry

There are, of course, a number of components that have to come together to make a SAN, well, a SAN. We will identify some of the components that are likely to be encountered.

6.3.1 ASIC

The fabric electronics utilize personalized application-specific integrated circuits (ASIC or ASICs) and its predefined set of elements, such as logic functions, I/O circuits, memory arrays, and backplane to create specialized fabric interface components.

An ASIC provides services to Fibre Channel ports that may be used connect to external N_Ports (such as an F_Port or FL_Port), external loop devices (such as an FL_Port), or to other switches such as an E_Port). The ASIC contains the Fibre channel interface logic, message/buffer queuing logic, and receive buffer memory for the on-chip ports, as well other support logic.

Frame filtering

Frame filtering is a feature that enables devices to provide zoning functions with finer granularity. Frame filtering can used to set up port level zoning, world wide name zoning, device level zoning, protocol level zoning, and LUN level zoning. Frame filtering is commonly carried out by an ASIC. This has the result that, after the filter is set up, the complicated function of zoning and filtering can be achieved at wire speed.

6.3.2 Fibre Channel transmission rates

Sometimes referred to as feeds and speeds. The current set of vendor offerings for switches, host bus adapters, and storage devices is constantly increasing. Currently the supported maximum for an IBM SAN is 10 Gbps. Most hardware can auto-negotiate to accommodate a range of speeds up to and including 10 Gbps, if necessary, to support older hardware.

6.3.3 SerDes

The communication over a fiber, whether optical or copper, is serial. Computer busses, on the other hand, use parallel busses. This means that Fibre Channel devices need to be able to convert between these two. For this, they use a serializer/deserializer, which is commonly referred as a SerDes.

6.3.4 Backplane and blades

Rather than having a single printed circuit assembly containing all the components in a device, sometimes the design used is that of a *backplane* and *blades*. For example, directors and large core switches usually implement this technology.

The backplane is a circuit board with multiple connectors into which other cards may be plugged. These other cards are usually referred as blades, but other terms could be used.

If the backplane is in the center of the unit with blades being plugged in at the back and front, then it would usually be referred to as a midplane.

6.4 Gigabit transport technology

In Fibre Channel technology, frames are moved from source to destination using gigabit transport, which is a requirement to achieve fast transfer rates. To communicate with gigabit transport, both sides have to support this type of communication. This can be accomplished by installing this feature into the device or by using specially designed interfaces that can convert other communication transport into gigabit transport. Gigabit transport can be used in copper or fibre optic infrastructure.

The interfaces that are used to convert the internal communication transport to gigabit transport are as follows:

- XFP
- Small form-factor pluggable media (SFP)
- Gigabit interface converters (GBIC)
- Gigabit Link Modules (GLM)
- Media Interface Adapters (MIA)
- 1x9 transceivers

6.4.1 Ten Gigabit small form-factor pluggable

The Ten (X) Gigabit small form-factor pluggable (XFP) specification defines the electrical, management, and mechanical interfaces of the XFP module. The module is a hot pluggable small footprint serial-to-serial data-agnostic multi rate optical transceiver, intended to support Telecom (SONET OC-192 and G.709 “OTU-2”) and Datacom applications (10 Gbps Ethernet and 10 Gbps Fibre Channel). Nominal data rates range from 9.95 Gbps, 10.31 Gbps, 10.52 Gbps, 10.70 Gbps, and the emerging 11.09 Gbps. The modules support all data encoding’s for these technologies. The modules may be used to implement single mode or multi-mode serial optical interfaces at 850 nm, 1310 nm, or 1550 nm. The XFP module design may use one of several different optical connectors. An adaptable heatsink option allows a single module design to be compatible with a variety of hosts.

6.4.2 Small form-factor pluggable media

The most common fibre channel interconnect in use today is small form-factor pluggable media, as shown in Figure 6-2 on page 117. This component is hot pluggable on the I/O module or the HBA, and the cable is also hot-pluggable. The SFP optical transceivers can use short or long-wavelength lasers.

Another version of the transceivers are called Small Form Fixed optical transceivers, and are mounted on the I/O module of the HBA via pin-through-hole technology, as shown in Figure 6-2 on page 117. The transceivers, which are designed for increased densities, performance, and reduced power, are well-suited for Gigabit Ethernet, Fibre Channel, and 1394b applications.

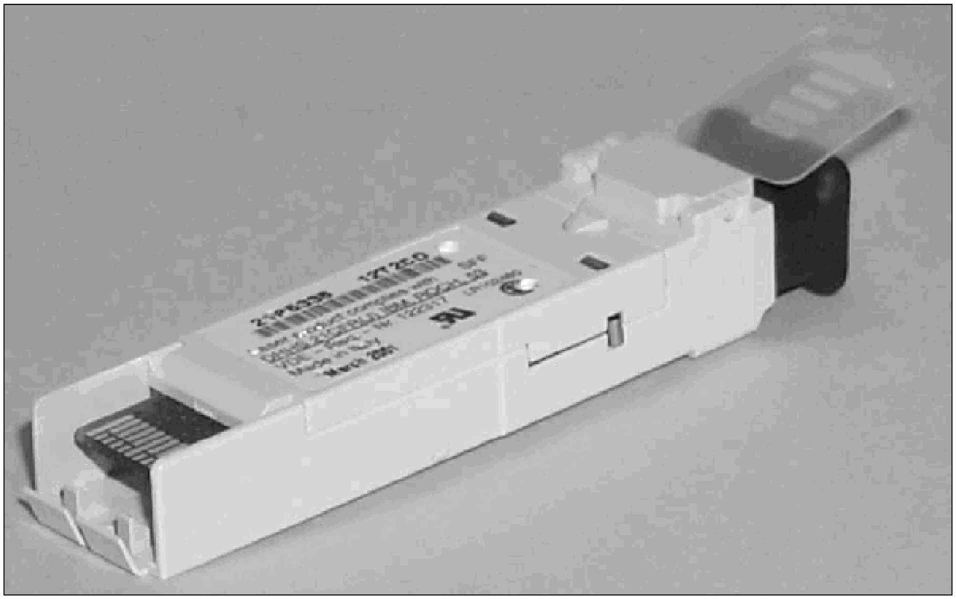


Figure 6-2 SFP

The small dimension of the SFP optical transceivers are ideal in switches and other products where many transceivers have to be configured in a small space.

6.4.3 Gigabit interface converters

Gigabit interface converters (GBICs) are integrated fibre-optic transceivers providing a high-speed serial electrical interface for connecting processors, switches, and peripherals through an optic fibre cable. In SANs they can be used for transmitting data between peripheral devices and processors.

IBM offers laser-based, hot-pluggable, data communications transceivers for a wide range of networking applications requiring high data rates. The transceivers, which are designed for ease of configuration and replacement, are well-suited for Gigabit Ethernet, Fibre Channel, and 1394b applications. Current GBICs are capable of 2125 Mbps transmission rates. GBICs are available in both short wavelength and long wavelength versions, providing configuration flexibility.

GBICs are usually hot-pluggable, and easy to configure and replace. On the optical side they use low-loss, SC type, push-pull connectors. They are mainly used in hubs, switch directors, and gateways. The transfer rates are typically in the range of 1063 Mbps, 1250 Mbps, 2125 Mbps, or 2500 Mbps. A GBIC is shown in Figure 6-3 on page 118.

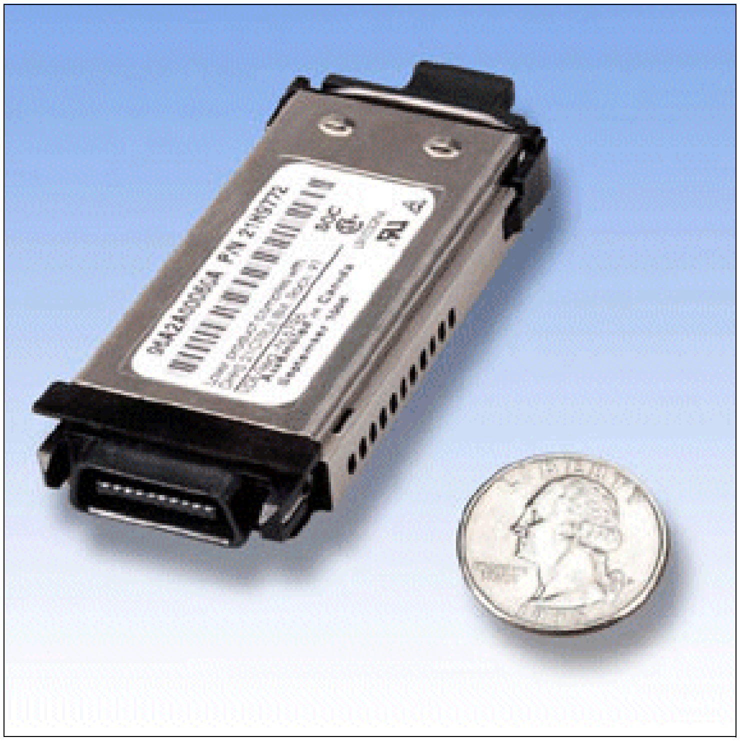


Figure 6-3 GBIC

6.4.4 Gigabit Link Modules

Gigabit Link Modules (GLMs) were used in early Fibre Channel applications. They are low-cost alternative to GBICs, but they are not hot-pluggable. They use the same fibre optic for the transport of optical signals as GBICs. The GLM converts encoded data that has been serialized into pulses of laser for transmission into optical fibre. A GLM at a second optical link, running at the same speed as the sending GLM, receives these pulses, along with the requisite synchronous clocking signals.

A GLM is shown in Figure 6-4 on page 119.

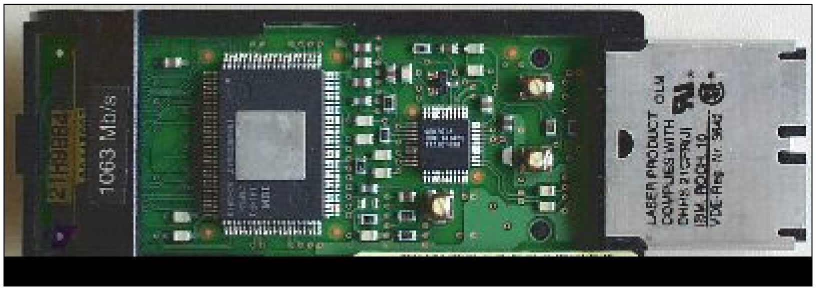


Figure 6-4 GLM

6.4.5 Media Interface Adapters

Media Interface Adapters (MIAs) can be used to facilitate conversion between optical and copper interface connections. Typically MIAs are attached to host bus adapters, but they can also be used with switches and hubs. If a hub or switch only supports copper or optical connections, MIAs can be used to convert the signal to the appropriate media type, copper or optical.

A MIA is shown in Figure 6-5.

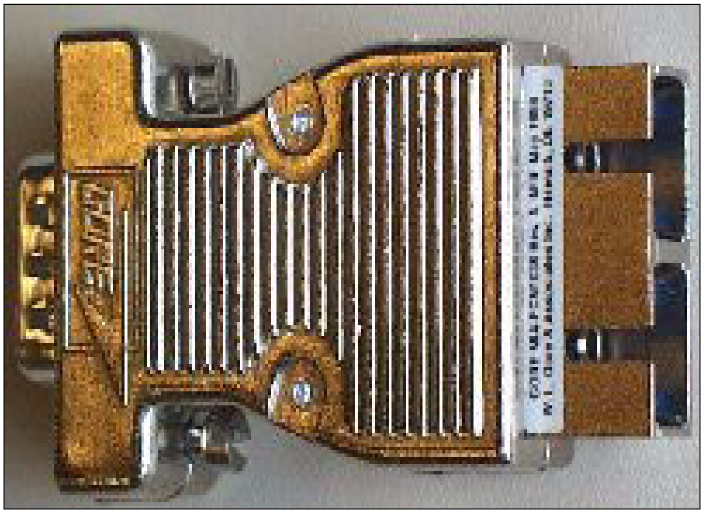


Figure 6-5 MIA

6.4.6 1x9 Transceivers

Early FC implementations sometimes relied on 1x9 transceivers for providing SC connection to their devices. These are typically no longer used but are shown in Figure 6-6.

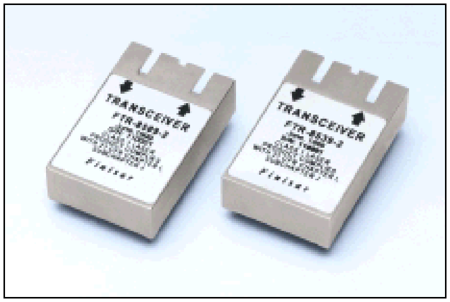


Figure 6-6 1x9 transceivers

6.4.7 Host bus adapters

The device that acts as an interface between the fabric of a SAN and either a host or a storage device is a host bus adapter (HBA).

The HBA connects to the bus of the host or storage system. It has some means of connecting to the cable or fibre leading to the SAN. Some devices offer more than one Fibre Channel connection. The function of the HBA is to convert the parallel electrical signals from the bus into a serial signal to pass to the SAN.

A host bus adapter is shown in Figure 6-7 on page 121.

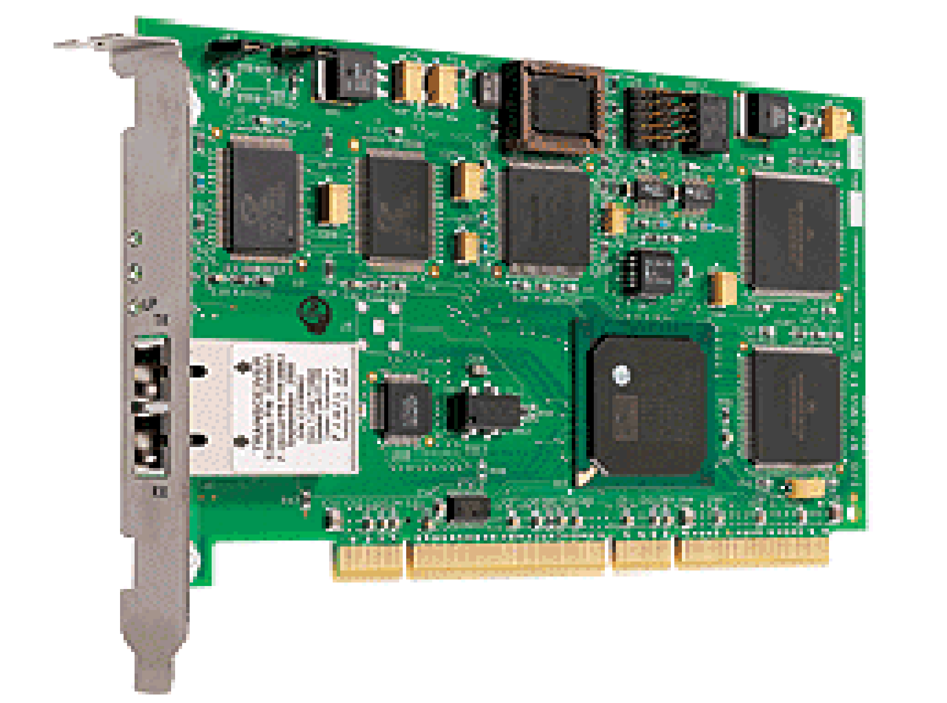


Figure 6-7 HBA

There are several manufacturers of HBAs, and an important consideration when planning a SAN is the choice of HBAs. HBAs may have more than one port, may be supported by some equipment and not others, may have parameters that can be used to tune the system, and many other features. The choice of an HBA is a critical one.

6.5 Inter-switch links

A link that joins a port on one switch to a port on another switch (referred to as E_Ports) is called an inter-switch link (ISL).

ISLs carry frames originating from the node ports, and those generated within the fabric. The frames generated within the fabric serve as control, management, and support for the fabric.

Before an ISL can carry frames originating from the node ports, the joining switches have to go through a synchronization process on which operating parameters are interchanged. If the operating parameters are not compatible, the switches may not join, and the ISL becomes *segmented*. Segmented ISLs cannot carry traffic originating on node ports, but they can still carry management and control frames.

6.5.1 Cascading

Expanding the fabric is called switch cascading, or just cascading. Cascading is basically interconnecting Fibre Channel switches and/or directors using ISLs. The cascading of switches provides the following benefits to a SAN environment:

- The fabric can be seamlessly extended. Additional switches can be added to the fabric, without powering down existing fabric.
- You can easily increase the distance between various SAN participants.
- By adding more switches to the fabric, you increase connectivity by providing more available ports.
- Cascading provides high resilience in the fabric.
- With inter-switch links (ISLs), you can increase the bandwidth. The frames between the switches are delivered over all available data paths. So the more ISLs you create, the faster the frame delivery will be, but careful consideration must be employed to ensure that a bottleneck is not introduced.
- When the fabric grows, the name server is fully distributed across all the switches in fabric.
- With cascading, you also provide greater fault tolerance within the fabric.

6.5.2 Hops

When FC traffic traverses an ISL, this is known as a *hop*. Or, to put it another way, traffic going from one switch over an ISL to another switch is one hop.

We show a sample configuration that illustrates this in Figure 6-8 on page 123, with a kangaroo to illustrate the hop count.

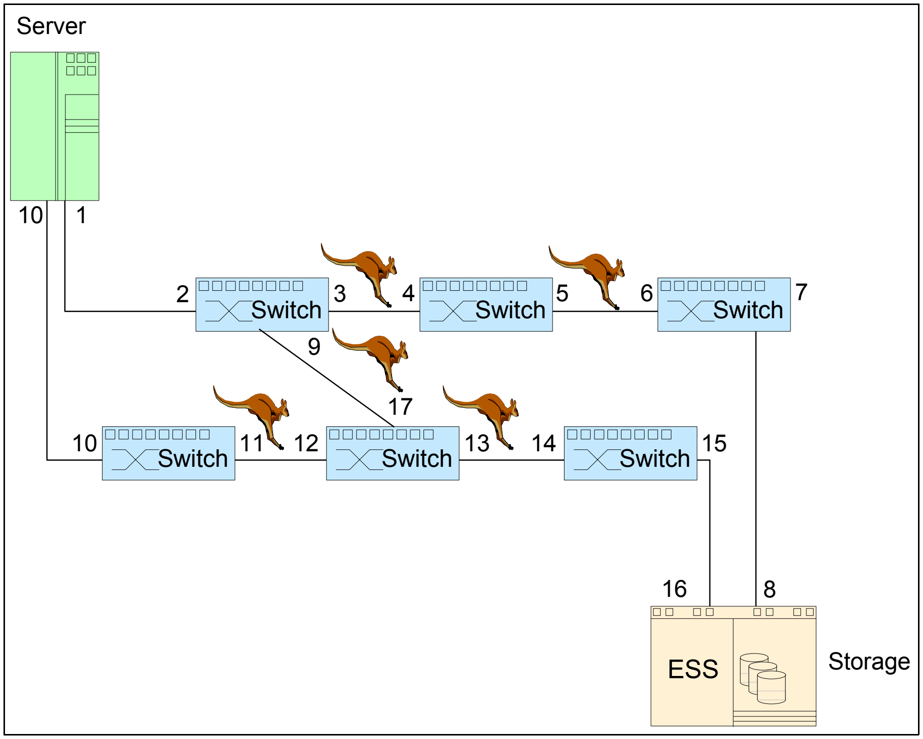


Figure 6-8 Kangaroo illustrating hops in a cascaded fabric

As with a lot of things in life, there is a hop count limit. This is set by the fabric operating system and is used to derive a frame holdtime value for each switch. This holdtime value is the maximum amount of time that a frame can be held in a switch before it is dropped, or if the fabric indicates that it is too busy. The hop count limits needs to be investigated and considered in any SAN design work as it will have a major effect on the proposal.

6.5.3 Fabric shortest path first

Although not strictly speaking a physical component, it makes sense to introduce fabric shortest path first now. According to the FC-SW-2 standard, fabric shortest path first (FSPF) is a link state path selection protocol. FSPF keeps track of the links on all switches in the fabric (in routing tables) and associates a cost with each link. The protocol computes paths from a switch to all the other switches in the fabric by adding the cost of all links traversed by the path, and choosing the path that minimizes the cost.

For example, as shown in Figure 6-9, if we need to connect a port in switch A to a port in switch D, it will take the ISL from A to D (as indicated by the dotted line). It will not go from A to B to D, nor from A to C to D.

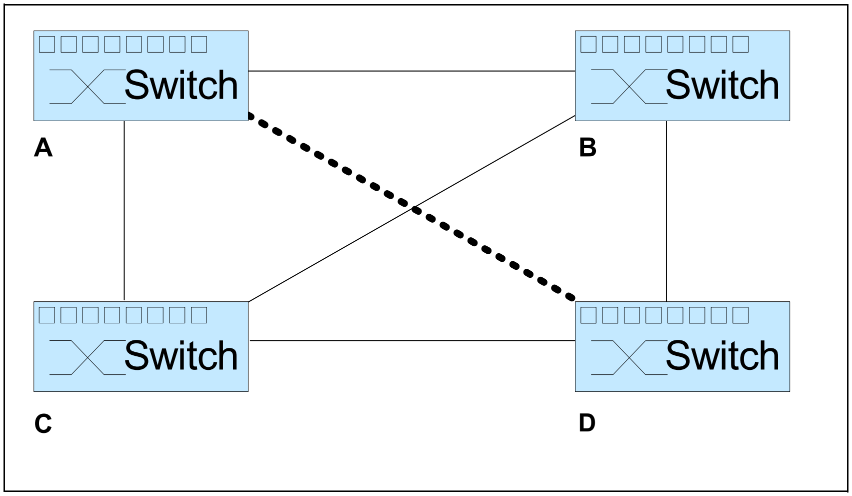


Figure 6-9 Four-switch fabric

FSPF is currently based on the hop count cost.

The collection of link states, including cost, of all switches in a fabric constitutes the topology database, or *link state* database. The topology database is kept in all switches in the fabric, and they are maintained and synchronized to each other. There is an initial database synchronization, and an update mechanism. The initial database synchronization is used when a switch is initialized, or when an ISL comes up. The update mechanism is used when there is a link state change, for example, an ISL going down or coming up, and on a periodic basis. This ensures consistency among all switches in the fabric.

If we look again at the example in Figure 6-9, and we imagine that the link from A to D goes down, switch A will now have four routes to reach D:

- A-B-D
- A-C-D
- A-B-C-D
- A-C-B-D

A-B-D and A-C-D will be selected because they are the shortest paths based on the hop count cost. The update mechanism ensures that switches B and C will also have their databases updated with the new routing information.

6.5.4 Blocking

To support highly performing fabrics, the fabric components, switches or directors must be able to move data around without any impact to other ports, targets, or initiators that are on the same fabric. If the internal structure of a switch or director cannot do so without impact, we end up with blocking.

Because the fabric components do not typically read the data that they are transmitting or transferring. This means that as data is being received, data is being transmitted. Because the potential can be as much as 10 Gbps bandwidth for each direction of the communication, a fabric component will need to be able to support this. So that data does not get delayed within the SAN fabric component itself, switches, directors, and hubs may employ a non-blocking switching architecture. Non-blocking switches provide for multiple connections travelling through the internal components of the switch concurrently.

Blocking means that the data does not get to the destination. This is opposed to congestion, where data will still be delivered, albeit with a delay. Switches and directors may employ a non-blocking switching architecture. Non-blocking switches and directors are the Ferraris on the SAN racetrack.

We illustrate this concept in Figure 6-10.

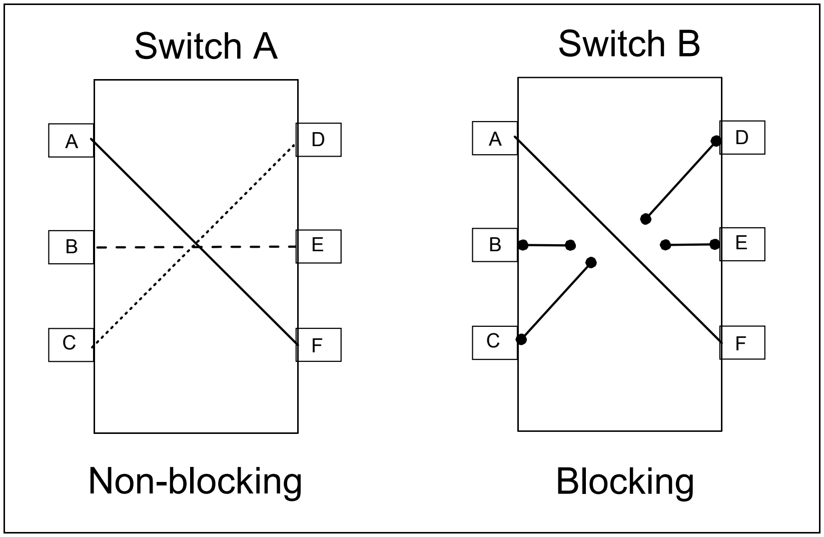


Figure 6-10 Non-blocking and blocking switching

In this example, nonblocking Switch A, port A speaks to port F. Switch B speaks to E, and C speaks to D without any form of suspension of communication or delay. That is to say, the communication is not blocked. In the blocking Switch B, while port A is speaking to F, all other communication has been stopped or blocked.

6.5.5 Latency

Typically, in the SAN world, latency is the time that it takes for a FC frame to traverse the fabric. The more ISLs, the more the latency if the FC frame has to traverse the fabric using ISLs. By fabric, we mean the FC components, and in any latency discussion related to the SAN, it is unusual if the host or storage is included in the equation. Usually the time taken is expressed in microseconds, which gives an indication as to the performance characteristics of the SAN fabric. It will often be given at a switch level, and sometimes a fabric level.

6.5.6 Oversubscription

We use the term oversubscription to describe the occasion when we have several ports trying to communicate with each other, and when the total throughput is higher than what that port can provide.

This can happen on storage ports and ISLs. When designing a SAN it is important to consider the possible traffic patterns to determine the possibility of oversubscription, which may result in degraded performance. Oversubscription of an ISL may be overcome by adding a parallel ISL. Oversubscription to a storage device may be overcome by adding another adapter to the storage array and connecting into the fabric.

6.5.7 Congestion

When oversubscription occurs, it leads to a condition called congestion. When a node is unable to utilize as much bandwidth as it would like to, due to contention with another node, then there is a congestion. A port, link, or fabric can be congested.

6.5.8 Trunking

Trunking is a feature of switches that enables traffic to be distributed across available inter-switch links (ISLs) while still preserving in-order delivery. On some Fibre Channel protocol devices, frame traffic between a source device and destination device must be delivered in order within an exchange.

This restriction forces current devices to fix a routing path within a fabric. Consequently, certain traffic patterns in a fabric can cause all active routes to be allocated to a single available path and leave other paths unused. Trunking implementation usually creates a trunking group (a set of available paths linking two adjacent switches). Ports in the trunking group are called trunking ports.

We illustrate the concepts of trunking in Figure 6-11.

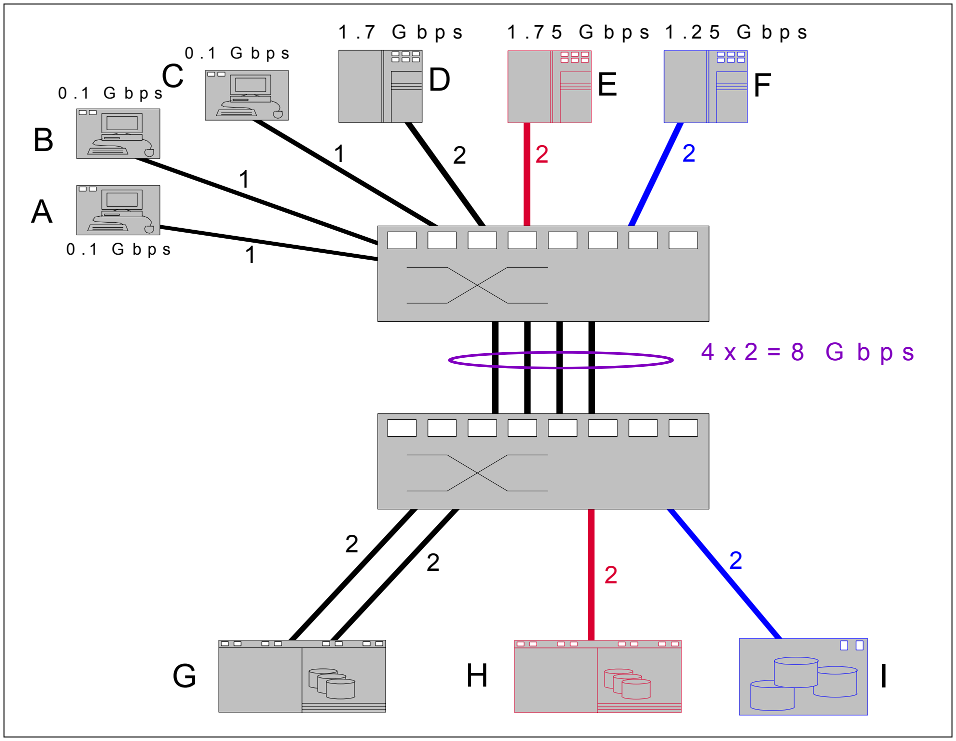


Figure 6-11 Trunking

In this example we have six computers that are accessing three storage devices. Computers A, B, C, and D are communicating with Storage G. Server E is communicating with storage H, and server F uses disks in storage device I.

The speeds of the links are shown in Gbps, and the target throughput for each computer is shown. If we let FSPF alone decide the routing for us, we could have a situation where servers D and E were both utilizing the same ISL. This would lead to oversubscription and hence congestion, as 1.7 added to 1.75 is greater than 2.

If all of the ISLs are gathered together into a trunk, then effectively they can be seen as a single, big ISL. In effect, they appear to be an 8 Gbps ISL. This bandwidth is greater than the total requirement of all of the servers. In fact, the nodes require an aggregate bandwidth of 5 Gbps, so we could even suffer a failure of one of the ISLs and still have enough bandwidth to satisfy their needs.

When the nodes come up, FSPF will simply see one route, and they will all be assigned a route over the same trunk. The fabric operating systems in the switches will share the load over the actual ISLs, which combine to make up the trunk. This is done by distributing frames over the physical links, and then re-assembling them at the destination switch so that in-order delivery can be assured, if necessary. To FSPF, a trunk will appear as a single, low-cost ISL.

7

Management

Management is one of the key issues behind the concept of infrastructure simplification. The ability to manage heterogeneous systems at different levels as though they were a fully-integrated infrastructure offering the system administrator a unified view of the whole SAN, is a goal that many vendors and developers have been striving to achieve.

In this chapter, we look at some of the initiatives that have been, and are being, developed in the field of SAN management, and these will incrementally smooth the way towards infrastructure simplification.

7.1 Management principles

SAN management systems typically comprise a set of multiple-level software components that provide tools for monitoring, configuring, controlling (performing actions), diagnosing, and troubleshooting a SAN. In this section, we briefly describe the different types and levels of management that can be found in a typical SAN implementation, as well as the efforts that are being made towards the establishment of open and general-purpose standards for building interoperable, manageable components.

In this section it is also shown that, despite these efforts, the reality of a “one pill cures all” solution is a long way off. Typically, each vendor and each device has its own form of software and hardware management techniques. These are usually independent of each other, and to pretend that there is one SAN management solution that will provide a single point of control, capable of performing every possible action, would be premature at this stage. This redbook does not aim at fully describing each vendor’s own standard(s), but at presenting the reader with an overview of the myriad of possibilities that they might find in the IT environment. That stated, the high-level features of any SAN management solution are likely to include most, if not all, of the following:

- Cost effectiveness
- Open approach
- Device management
- Fabric management
- Pro-active monitoring
- Fault isolation and troubleshooting
- Centralized management
- Remote management
- Adherence to standards
- Resource management
- Secure access
- Standards compliant

7.1.1 Management types

There are essentially two philosophies used for building management mechanisms: *in-band management* and *out-of-band management*. They can be defined as:

In-band management

This means that the management data, such as status information, action requests, events, and so on, flows through the same path as the storage data itself.

Out-of-band management

This means that the management data flows through a dedicated path, therefore not sharing the same physical path used by the storage data.

In-band and out-of-band models can be illustrated as shown in Figure 7-1. These models are not mutually exclusive. In many environments a combination of both may be desired.

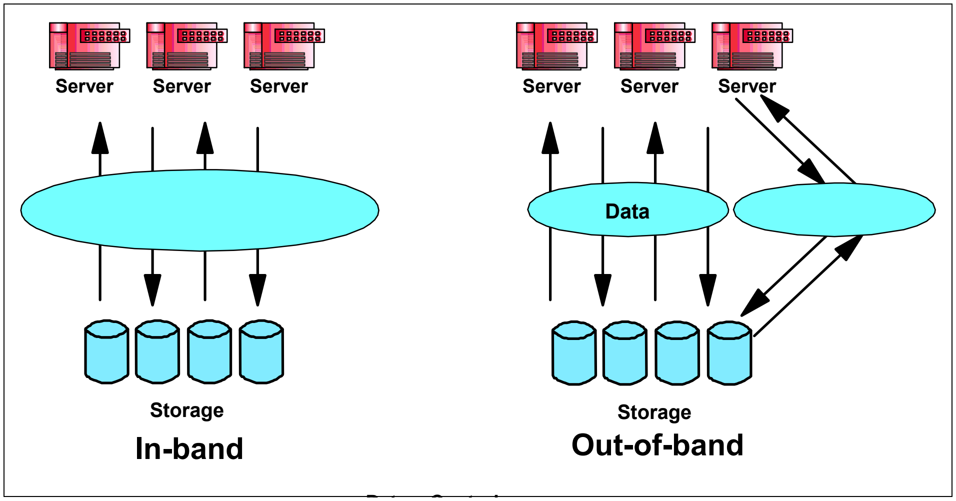


Figure 7-1 In-band and out-of-band models

The in-band approach is simple to implement, requires no dedicated channels (other than LAN connections) and has inherent advantages, such as the ability for a switch to initiate a SAN topology map by means of queries to other fabric components. However, in the event of a failure of the Fibre Channel transport itself, the management information cannot be transmitted. Therefore access to devices is lost, as is the ability to detect, isolate, and recover from network problems. This problem can be minimized by a provision of redundant paths between devices in the fabric.

In-band management is evolving rapidly. Proposals exist for low-level interfaces such as Return Node Identification (RNID) and Return Topology Identification (RTIN) to gather individual device and connection information, and for a management server that derives topology information. In-band management also allows attribute inquiries on storage devices and configuration changes for all elements of the SAN. Since in-band management is performed over the SAN itself, administrators are not required to manage any additional connections.

On the other hand, out-of-band management does not rely on the storage network; its main advantage is that management commands and messages can be sent even if a loop or fabric link fails. Integrated SAN management facilities are more easily implemented. However, unlike in-band management, it cannot automatically provide SAN topology mapping.

In summary, we can say that In-band management has these main advantages:

- Device installation, configuration, and monitoring
- Inventory of resources on the SAN
- Automated component and fabric topology discovery
- Management of the fabric configuration, including zoning configurations
- Health and performance monitoring

Out-of-band management has these main advantages:

- It keeps management traffic out of the FC path, so it does not affect the business-critical data flow on the storage network.
- It makes management possible, even if a device is down.
- It is accessible from anywhere in the routed network.

7.1.2 SAN management levels

The SAN management architecture can be divided into three distinct levels:

- Storage level
- Network level
- Enterprise level

In Figure 7-2 on page 133 we illustrate the three management levels in a SAN solution.

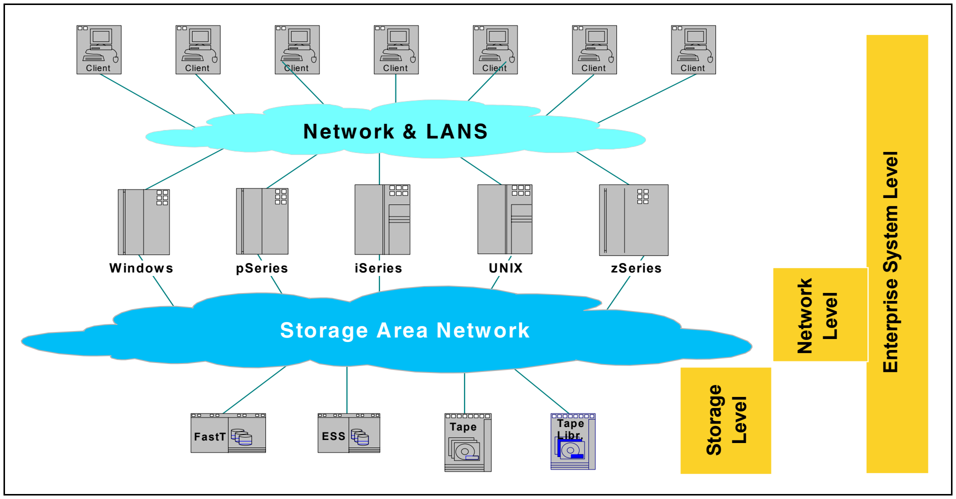


Figure 7-2 SAN management levels

Storage level

The SAN storage level is comprised of the storage devices that integrate the SAN, such as disks, disk arrays, tapes, and tape libraries. As the configuration of a storage resource must be integrated with the configuration of the server's logical view of them, the SAN storage level management may also span both storage resources and servers.

Network level

The SAN network level is comprised by all the components that provide connectivity, such as cables, switches, inter-switch links, gateways, routers and HBAs.

Enterprise level

The enterprise level comprises all devices and components present in a SAN environment, as well as the workstations indirectly connected to it. As such, it implies the ability to manage the whole system from a single perspective or, in other words, it comprises the integration of all-level components in a single manageable structure that can be operated from a single console.

A number of initiatives and standards, such as Web-Based Enterprise Management (WBEM), Common Information Model (CIM), Desktop Management Interface (DMI), and Java™ Management Application Programming Interface (JMAPI) are being defined, and deployed today in order to create enterprise-level of standardization on management.

7.1.3 SAN fault isolation and troubleshooting

In addition to providing tools for monitoring and configuring a SAN, one of the key benefits that a well-designed management mechanism can bring is the ability to efficiently detect, diagnose and solve problems in a SAN.

There are many tools to collect the necessary data in order to perform problem determination and problem source identification (PD/PSI) in a SAN. Generally speaking, these tools offer the ability to:

- Interrogate the fabric.
- Find devices or LUNs that have gone missing.
- Identify storage device or server failures.
- Identify fabric failures (usually by means of **`ping`** and **`traceroute`**).
- Interpret message and error logs.
- Fire events when failures are detected.
- Check for zoning conflicts.

Although a well-designed management system can provide such invaluable facilities, an easy-to-troubleshoot SAN still relies heavily on a good design, and on good documentation; in terms of PD/PSI, this means that configuration design information is understandable, available at any support level, and is always updated with respect to the latest configuration. There must also be a database where all the information about connections, naming conventions, device serial numbers, WWN, zoning, system applications, and so on, is safely stored. Last, but not least, there should be a responsible person in charge of maintaining this infrastructure, and monitoring the SAN health status.

7.2 Management interfaces and protocols

In this section we present the main protocols and interfaces that have been developed to support management mechanisms.

7.2.1 SNIA initiative

The Storage Networking Industry Association (SNIA) is using its Storage Management Initiative (SMI) to create and promote adoption of a highly functional interoperable management interface for multivendor storage networking products. The SNIA strategic imperative is to have all storage managed by the SMI interface. The adoption of this interface will allow the focus to switch to the development of value-add functionality. IBM is one of the industry vendors promoting the drive towards this vendor-neutral approach to SAN management.

In 1999, the SNIA and Distributed Management Task Force (DMTF) introduced open standards for managing storage devices. These standards use a common protocol called the Common Information Model (CIM) to enable interoperability. The Web-based version of CIM (WBEM) uses XML to define CIM objects and process transactions within sessions. This standard proposes a CIM Object Manager (CIMOM) to manage CIM objects and interactions. CIM is used to define objects and their interactions. Management applications then use the CIM object model and XML over HTTP to provide for the management of storage devices. This enables central management through the use of open standards.

SNIA uses the xmlCIM protocol to describe storage management objects and their behavior. CIM allows management applications to communicate with devices using object messaging encoded in xmlCIM.

The Storage Management Interface Specification (SMI-S) for SAN-based storage management provides basic device management, support for copy services, and virtualization. As defined by the standard, the CIM services are registered in a directory to make them available to device management applications and subsystems.

For more information about SMI-S go to:

http://www.snia.org

Additionally, SNIA and the International Committee for Information Technology Standards (INCITS) announced in October 2004 that the Storage Management Initiative Specification (SMI-S) has been approved as a new INCITS standard. The standard was approved by the INCITS executive board and has been designated as ANSI INCITS 388-2004, *American National Standard for Information Technology Storage Management*.

ANSI INCITS 388-2004 was developed through a collaborative effort by members of SNIA representing a cross section of the industry. Today, the standard focuses on storage management of SANs and will be extended to include Network Attached Storage (NAS), Internet Small Computer System Interface (iSCSI), and other storage networking technologies.

The ANSI INCITS 388-2004 standard can be purchased through the INCITS Web site at:

http://www.incits.org

Open storage management with CIM

SAN management involves configuration, provisioning, logical volume assignment, zoning, and Logical Unit Number (LUN) masking, as well as monitoring and optimizing performance, capacity, and availability. In addition, support for continuous availability and disaster recovery requires that device copy services are available as a viable failover and disaster recovery environment. Traditionally, each device provides a command line interface (CLI) as well as a graphical user interface (GUI) to support these kinds of administrative tasks. Many devices also provide proprietary APIs that allow other programs to access their internal capabilities.

For complex SAN environments, management applications are now available that make it easier to perform these kinds of administrative tasks over a variety of devices.

The CIM interface and SMI-S object model adopted by SNIA provide a standard model for accessing devices, which allows management applications and devices from a variety of vendors to work with each other's products. This means that customers have more choice as to which devices will work with their chosen management application, and which management applications they can use with their devices.

IBM has embraced the concept of building open standards-based storage management solutions. IBM management applications are designed to work across multiple vendors' devices, and devices are being CIM-enabled to allow them to be controlled by other vendors' management applications.

CIM Object Manager

The SMI-S standard designates that either a proxy or an embedded agent may be used to implement CIM. In each case, the CIM objects are supported by a CIM Object Manager (CIMOM). External applications communicate with CIM via HTTP to exchange XML messages, which are used to configure and manage the device.

In a proxy configuration, the CIMOM runs outside of the device and can manage multiple devices. In this case, a *provider* component is installed into the CIMOM to enable the CIMOM to manage specific devices.

The providers adapt the CIMOM to work with different devices and subsystems. In this way, a single CIMOM installation can be used to access more than one device type, and more than one device of each type on a subsystem.

The CIMOM acts as a catcher for requests that are sent from storage management applications. The interactions between catcher and sender use the language and models defined by the SMI-S standard. This allows storage management applications, regardless of vendor, to query status and perform command and control using XML-based CIM interactions.

IBM has developed its storage management solutions based on the CIMOM architecture, as shown in Figure 7-3.

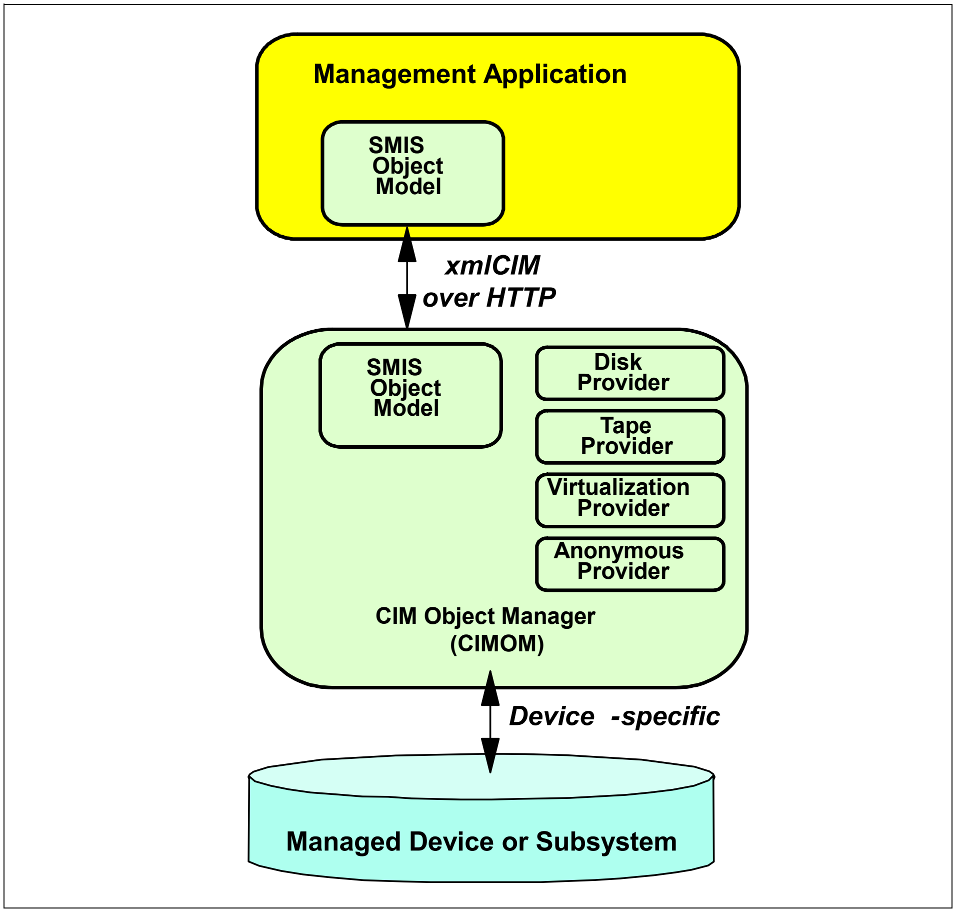


Figure 7-3 CIMOM component structure

7.2.2 Simple Network Management Protocol

SNMP, which is an IP-based protocol, has a set of commands for obtaining the status and setting the operational parameters of target devices. The SNMP management platform is called the SNMP manager, and the managed devices have the SNMP agent loaded. Management data is organized in a hierarchical data structure called the Management Information Base (MIB). These MIBs are defined and sanctioned by various industry associations. The objective is for all vendors to create products in compliance with these MIBs, so that inter-vendor interoperability at all levels can be achieved. If a vendor wants to include additional device information that is not specified in a standard MIB, then that is usually done through MIB extensions.

This protocol is widely supported by LAN/WAN routers, gateways, hubs, and switches, and is the predominant protocol used for multivendor networks. Device status information (vendor, machine serial number, port type and status, traffic, errors, and so on) can be provided to an enterprise SNMP manager. A device can generate an alert by SNMP, in the event of an error condition. The device symbol, or icon, displayed on the SNMP manager console, can be made to turn red or yellow, or any warning color, and messages can be sent to the network operator.

Out-of-band developments

Two primary SNMP MIBs are being implemented for SAN fabric elements that allow out-of-band monitoring. The ANSI Fibre Channel Fabric Element MIB provides significant operational and configuration information on individual devices. The emerging Fibre Channel Management MIB provides additional link table and switch zoning information that can be used to derive information about the physical and logical connections between individual devices. Even with these two MIBs, out-of-band monitoring is incomplete. Most storage devices and some fabric devices do not support out-of-band monitoring. In addition, many administrators simply do not attach their SAN elements to the TCP/IP network.

7.2.3 Service Location Protocol

The Service Location Protocol (SLP) provides a flexible and scalable framework for providing hosts with access to information about the existence, location, and configuration of networked services. Traditionally, users have had to find devices by knowing the name of a network host that is an alias for a network address. SLP eliminates the need for a user to know the name of a network host supporting a service. Rather, the user supplies the desired type of service and a set of attributes that describe the service. Based on that description, the Service Location Protocol resolves the network address of the service for the user.

SLP provides a dynamic configuration mechanism for applications in local area networks. Applications are modeled as clients that need to find servers attached to any of the available networks within an enterprise. For cases where there are many different clients and/or services available, the protocol is adapted to make use of nearby Directory Agents that offer a centralized repository for advertised services.

The IETF's Service Location (srvloc) Working Group is developing SLP. SLP is defined in RFC 2165 (Service Location Protocol, June 1997) and updated in RFC 2608 (Service Location Protocol, Version 2, June 1999). More information can be found in this text document:

http://www.ietf.org/rfc/rfc2608.txt

SCSI Enclosure Services

In SCSI legacy (or existing) systems, a SCSI protocol runs over a limited length parallel cable, with up to 15 devices in a chain. The latest version of SCSI-3 serial protocol offers this same disk read/write command set in a serial format, which allows for the use of Fibre Channel, as a more flexible replacement of parallel SCSI. The ANSI SCSI Enclosure Services (SES) proposal defines basic device status from storage enclosures. For example, DIAGNOSTICS and RECEIVE DIAGNOSTIC RESULTS commands can be used to retrieve power supply status, temperature, fan speed, and other parameters from the SCSI devices.

The ANSI SCSI-3 serial protocol previously used in SCSI connections is now used over FCP by many SAN vendors in order to offer higher speeds, longer distances, and greater device population for SANs, with few changes in the upper level protocols. The ANSI SCSI-3 serial protocol has also defined a new set of commands called SCSI Enclosure Services (SES) for basic device status from storage enclosures.

SES has a minimal impact on Fibre Channel data transfer throughput. Most SAN vendors deploy SAN management strategies using Simple Network Management Protocol (SNMP) based, and non-SNMP based (SES), protocols.

7.2.4 Vendor-specific mechanisms

These are some of the vendor-specific mechanisms that have been deployed by major SAN device providers.

Application Program Interface

As we know, there are many SAN devices from many different vendors and everyone has their own management and/or configuration software. In addition to this, most of them can also be managed via a command line interface (CLI) over a standard telnet connection, where an IP address is associated with the SAN device, or they can be managed via a RS-232 serial connection.

With different vendors and the many management and/or configuration software tools, we have a number of different products to evaluate, implement, and learn. In an ideal world there would be one product to manage and configure all the actors on the SAN stage.

Application Program Interfaces (APIs) are one way to help this become a reality. Some vendors make the API of their product available for other vendors in order to make it possible for common management in the SAN. This allows for the development of upper level management applications capable of interacting with multiple-vendor devices and offering the system administrator a single view of the SAN infrastructure.

Tivoli Common Agent Services

The Tivoli Common Agent Services are a new component designed to provide a way to deploy agent code across multiple end-user machines or application servers throughout an enterprise. The agents collect data from and perform operations on managed resources for Fabric Manager.

The Tivoli Common Agent Services agent manager provides authentication and authorization and maintains a registry of configuration information about the agents and resource managers in the SAN environment. The resource managers are the server components of products that manage agents deployed on the common agent. Management applications use the services of the agent manager to communicate securely with and to obtain information about the computer systems running the Tivoli common agent software, referred to in this document as the agent.

Tivoli Common Agent Services also provide common agents to act as containers to host product agents and common services. The common agent provides remote deployment capability, shared machine resources, and secure connectivity.

Tivoli Common Agent Services is comprised of two subcomponents:

- Agent manager

  The agent manager handles the registration of managers and agents, and security (such as the issuing of certificates and keys and the performing of authentication). It also provides query APIs for use by other products. One agent manager instance can manage multiple resource managers and agents. The agent manager can be on the same machine as Fabric Manager or on a separate machine.

- Common agent

  The common agent resides on the agent machines of other Tivoli products. One common agent can manage multiple product agents on the same machine. It provides monitoring capabilities and can be used to install and update product agents.

7.3 Management features

SAN management requirements are typified by having a common purpose, but are implemented in different fashions by the differing vendors. Some prefer to use Web browser interfaces, some prefer to use embedded agents, and some prefer to use the CLI, some use a combination of all. There is no right or wrong way. Usually the selection of SAN components is based on a combination of what the hardware and software will provide, not on the ease of use of the management solution. As we have stated previously, the high-level features of any SAN management solution are likely to include most of the following:

- Cost effectiveness
- Open approach
- Device management
- Fabric management
- Pro-active monitoring
- Fault isolation and troubleshooting
- Centralized management
- Remote management
- Adherence to standards
- Resource management
- Secure access
- Standards compliant

7.3.1 IBM TotalStorage Productivity Center

The IBM TotalStorage Productivity Center is an open storage infrastructure management solution designed to help:

- Reduce the effort of managing complex, heterogeneous storage infrastructures
- Improve storage capacity utilization
- Improve administrative efficiency

TPC provides reporting capabilities, identifying data usage and its location, and provisioning. It also provides a central point of control to move the data based on business needs to more appropriate online or offline storage, and centralizes the management of storage infrastructure capacity, performance and availability.

The IBM TotalStorage Productivity Center is comprised of:

IBM TotalStorage Productivity Center for Data

The IBM TotalStorage Productivity Center for Data is a Java and Web-based solution designed to help identify, evaluate, control, and predict enterprise storage management needs. It supports today's complex heterogeneous environment, including Direct Access Storage (DAS), Network Attached Storage (NAS), and Storage Area Network (SAN) storage including intelligent disk subsystems and the IBM TotalStorage 3584 Tape Libraries. Productivity Center for Data supports leading databases and provides chargeback capabilities based on storage usage.

IBM TotalStorage Productivity Center for Fabric

The IBM TotalStorage Productivity Center for Fabric is a comprehensive management solution for multivendor SANs. It includes automatic resource and topology discovery, monitoring performance and alerts, and zone control. Productivity Center for Fabric is an enterprise scalable solution architected to ANSI SAN standards, allowing you to choose best of breed products for your storage infrastructure.

TPC includes advanced performance management of the SAN environment to help customers enforce storage service levels. It allows you to specify throughput thresholds and alert storage administrators to potential bottlenecks.

IBM TotalStorage Productivity Center for Disk

The TotalStorage Productivity Center for Disk is designed to help reduce the complexity of managing SAN storage devices by allowing administrators to configure, manage and performance monitor storage from a single console. Productivity Center for Disk has been specifically designed to configure multiple storage devices from a single console. Commitment to open standards means that TPC allows the ability to monitor and track the performance of SAN attached Storage Management Interface Specification (SMI-S) compliant storage devices. It enables proactive performance management by setting performance thresholds based on performance metrics and the generation of alerts.

This solution is architected to SNIA's (Storage Networking Industry Association) SMI-S and supports compliant SAN components.

IBM TotalStorage Productivity Center for Replication

The IBM TotalStorage Productivity Center for Replication is designed to help simplify copy services management for the ESS. It is designed to provide configuration and management of the FlashCopy® and Synchronous Metro Mirror capabilities of the ESS.

Data replication is a core function required for data protection and disaster recovery. The Productivity Center for Replication is designed to control and monitor copy services operations in storage environments. It also provides advanced copy services functions for supported storage subsystems on the SAN.

IBM TotalStorage Productivity Center Standard Edition

As a single integrated solution of the IBM TotalStorage Productivity Center, IBM TotalStorage Productivity Center Standard Edition is designed to help you improve your storage TCO and ROI by:

- Combining the assets, capacity, performance and operational management, traditionally offered by separate Storage Resource Management, SAN Management and Device Management applications into a single platform
- Monitoring and tracking the performance of SAN attached SMI-S compliant storage devices
- Managing the capacity utilization and availability of file systems, databases, storage, and IBM TotalStorage 3584 Family of tape libraries
- Monitoring, managing and controlling (zoning) SAN fabric components
- Assisting in the support of data classification

The IBM TotalStorage Productivity Center Standard Edition consists of:

- IBM TotalStorage Productivity Center for Data
- IBM TotalStorage Productivity Center for Fabric
- IBM TotalStorage Productivity Center for Disk

IBM TotalStorage Productivity Center Limited Edition

As a component of the IBM TotalStorage Productivity Center, IBM TotalStorage Productivity Center Limited Edition is packaged with the IBM TotalStorage Family of DS Platforms, including the DS8000™, DS6000™ and DS4000™. It is also available with the SAN Volume Controller and select IBM tape libraries.

The TotalStorage Productivity Center Limited Edition has been designed to:

- Discover and configure IBM and heterogeneous SMI-S supported devices
- Perform event gather, error logging and launch device element managers
- Provide basic asset and capacity reporting
- Display an end-to-end topology view of your storage infrastructure and health console
- Enable a simple upgrade path to IBM TotalStorage Productivity Center Standard Edition (or single priced modules)

It is a tool that provides storage administrators with a simple way to conduct device management for multiple storage arrays and SAN fabric components. All this is achieved from a single integrated console that is the base of operations for the IBM TotalStorage Productivity Center suite. This offering is available with:

- IBM TotalStorage Productivity Center for Data
- IBM TotalStorage Productivity Center for Fabric
- IBM TotalStorage Productivity Center for Disk

7.4 Vendor management applications

Each vendor in the IBM TotalStorage SAN portfolio brings their own bespoke applications to manage and monitor the SAN. In the topics that follow we give a high-level overview of each of them.

7.4.1 IBM TotalStorage b-type family

The b-type family switch management framework is designed to support the widest range of solutions, from the very small workgroup SANs up to very large enterprise SANs. The switch management options include browser-based WEBTOOLS, Fabric Manager, and open standards-based interfaces to enterprise SAN managers.

The following management interfaces allow you to monitor fabric topology, port status, physical status, and other information to aid in system debugging and performance analysis:

- Command-line interface (CLI) through a Telnet connection
- Advanced Web Tools
- SCSI Enclosure Services (SES)
- SNMP applications
- Management server

You can use all these management methods either in-band (Fibre Channel) or out-of-band (Ethernet), except for SES, which can be used for in-band only.

7.4.2 Cisco

Fabric Manager and Device Manager are the centralized tools used to manage the Cisco SAN fabric and the devices connected to it. Fabric Manager can be used to manage fabric-wide settings such as zoning and also manage settings at an individual switch level.

Fabric Manager provides high-level summary information about all switches in a fabric, automatically launching the Web Tools interface when more detailed information is required. In addition, Fabric Manager provides improved performance over Web Tools alone.

Some of the capabilities of fabric manager are:

- Configures and manages the fabric on multiple efficient levels
- Intelligently groups multiple SAN objects and SAN management functions to provide ease and time efficiency in administering tasks

- Identifies, isolates, and manages SAN events across multiple switches and fabrics
- Provides drill-down capability to individual SAN components through tightly coupled Web Tools and Fabric Watch integration
- Discovers all SAN components and views the real-time state of all fabrics
- Provides multi-fabric administration of secure Fabric OS SANs through a single encrypted console
- Monitors ISLs
- Manages switch licenses
- Performs fabric stamping

7.4.3 IBM TotalStorage e-type family

The Emulex InSpeed-based family of SAN storage switches combine extremely high performance levels with the simplicity needed for entry-level SANs. There are different levels of management possible with InSpeed SAN storage switches.

The lowest level of management includes an RS-232C or similar standard interface. Included is some active monitoring of ports and the ability to generate an interrupt via the external communications ports in order to allow logging of events.

The next level of management has complete monitoring and control capabilities and can generate events. In addition, advanced policy management features such as Port-Test-Before-Insert, health monitoring, and zone recovery are available.

The highest form of management includes all the previous management capabilities, along with SNMP, XML, and a Web server interface.

7.4.4 IBM TotalStorage m-type family

McDATA’s Enterprise Fabric Connectivity Manager (EFCM) application provides a common java-based GUI to configure and manage McDATA storage area networks. It is intended to give a fabric-wide view of the SAN from a single console, and can discover non-McDATA switches, provided the principal switch in a fabric is a McDATA switch. The application is accessed on the EFC server through a network connection from a remote user workstation. The application operates independently from the director, switch, or other product managed by the EFC Server.

One of the major factors for enterprise business to use the EFC server to manage their SAN is the capability to back up and restore device and fabric configuration for all the products managed by the local or remote EFC server. It enables the enterprise SAN to become disaster proof.

7.5 SAN multipathing software

In a well-designed SAN, it is likely that you will want a device to be accessed by the host application over more than one path in order to potentially obtain better performance, and to facilitate recovery in the case of adapter, cable, switch, or GBIC failure.

In Figure 7-4 on page 147 we show a high-level view of typical configuration in a core-edge SAN environment.

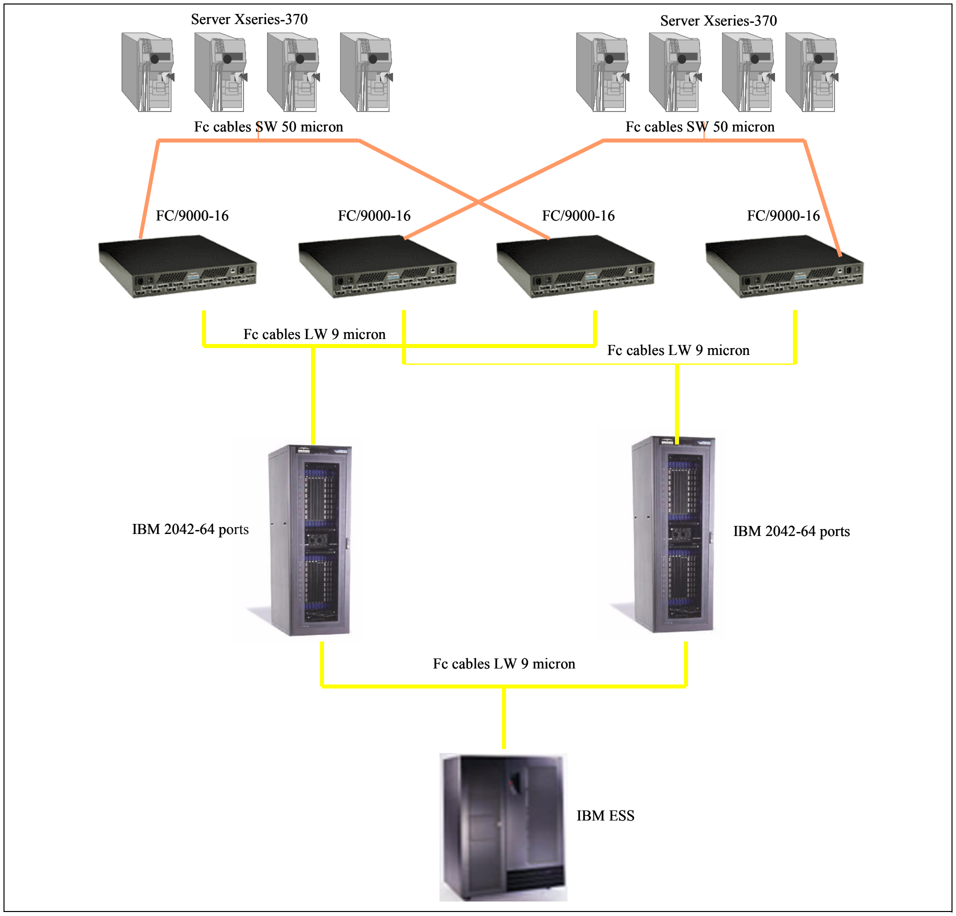


Figure 7-4 Core-edge SAN environment

In Figure 7-5 on page 148 we show this in more detail.

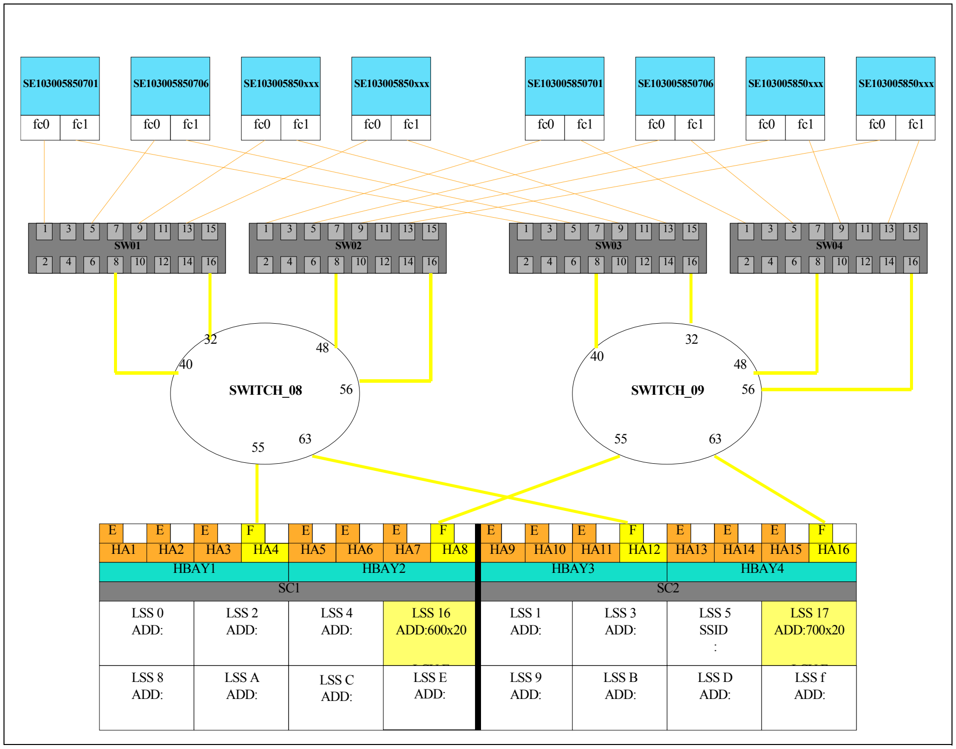


Figure 7-5 Core-edge SAN environment details

In this case, the same logical volume within a storage device (LUN) may be presented many times to the host through each of the possible paths to the LUN. In order to avoid this and make the device easier to administrate and to eliminate confusion, multipathing software will be needed. This will be responsible for making each LUN visible only once from the application and operating system point of view (similar to the concepts introduced in the XA architecture in the MVS™ operating system). In addition to this, the multipathing software is also responsible for fail-over recovery, and load balancing:

- Fail-over recovery: In a case of the malfunction of a component involved in making the LUN connection, the multipathing software must redirect all the data traffic onto other available paths.
- Load balancing: The multipathing software must be able to balance the data traffic equitably over the available paths from the hosts to the LUNs.

There are different kinds of multipathing software available from different vendors.

7.6 Storage virtualization in the SAN

In this section, we present the basics about one of the key features that a comprehensive management application should provide: SAN storage virtualization.

7.6.1 SANs and storage virtualization

Much attention is being focused on storage virtualization. SNIA defines storage virtualization as:

> The act of integrating one or more (back-end) services or functions with additional (front-end) functionality for the purpose of providing useful abstractions. Typically, virtualization hides some of the back-end complexity, or adds or integrates new functionality with existing back-end services. Examples of virtualization are the aggregation of multiple instances of a service into one virtualized service, or to add security to an otherwise insecure service. Virtualization can be nested or applied to multiple layers of a system.

Putting it in more practical terms, storage virtualization is the pooling of physical storage from multiple network storage devices into what appears to be a single storage device that is managed from a central console.

Storage virtualization techniques are becoming increasingly more prevalent in the IT industry today. Storage virtualization forms one of several layers of virtualization in a storage network, and can be described as “the abstraction from physical volumes of data storage to a logical view of data storage.”

This abstraction can be made on several levels of the components of storage networks and is not limited to the disk subsystem. Storage virtualization separates the representation of storage to the operating system (and its users) from the actual physical components. Storage virtualization has been represented and taken for granted in the mainframe environment for many years.

The SAN is making it easier for customers to spread their IT systems out geographically, but even in local networks, different types of servers that use different operating systems do not get the full benefit of sharing storage. Instead, the storage is partitioned to each different type of server, which creates complex management and inefficient use of storage. When storage must be added, applications are often disrupted. At the same time, the reduced cost of storage and the technology of storage networks with faster data transfer rates have enabled customers to use increasingly sophisticated applications, such as digital media. This has caused even greater complexity and difficulty of management as the amount of storage required grows at unprecedented rates.

7.6.2 Virtualization levels

We will define the different levels that virtualization can be achieved at in a storage network, as illustrated in Figure 7-6.

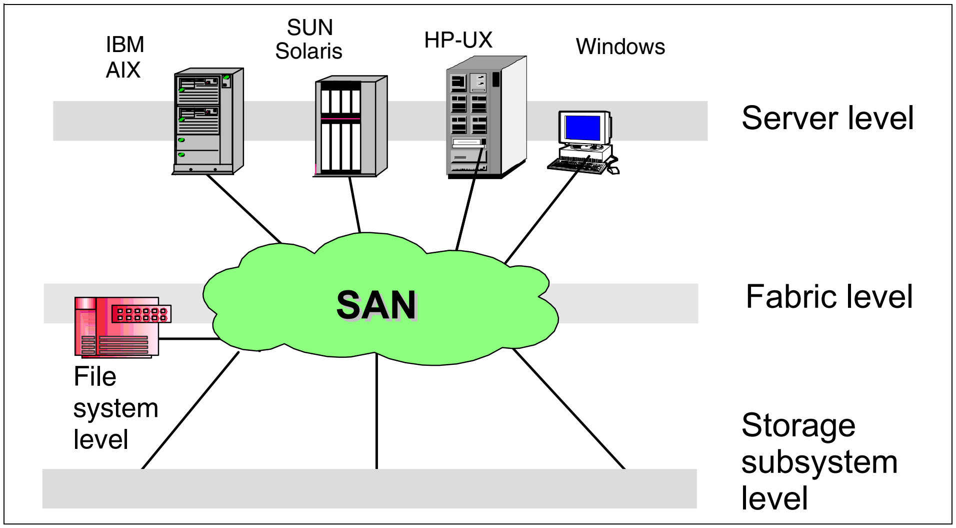


Figure 7-6 Levels that storage virtualization can be applied at

Fabric level

At the fabric level, virtualization can enable the independence of storage pools from heterogeneous servers. The SAN fabric would be zoned to allow the virtualization appliances to see the storage subsystems, and for the servers to see the virtualization appliances. Servers would not be able to directly see or operate on the storage subsystems.

Storage subsystem level

As seen before in this chapter, a LUN is a logical volume within a storage device. Disk storage systems can provide some level of virtualization already by subdividing disks into LUNs. Conversely, multiple storage devices could be consolidated together to form one large virtual drive. RAID subsystems are an example of virtualization at the storage level. Storage virtualization can take this to the next level by enabling the presentation and the management of disparate storage systems.

Server level

Abstraction at the server level is by means of the logical volume management of the operating systems on the servers. At first sight, increasing the level of abstraction on the server seems well suited for environments without storage networks, but this can be vitally important in storage networks too.

Since the storage is no longer controlled by individual servers, it can be used by any server as needed. Capacity can be added or removed on demand without affecting the application servers. Storage virtualization simplifies storage management and reduces the cost of managing the SAN environment.

File system level

Virtualization at the file system level provides the highest level of virtual storage. It can also provide the highest benefit, because it is data (not volumes) that is to be shared, allocated, and protected.

7.6.3 Virtualization models

Storage virtualization, just like general-purpose management mechanisms, can be implemented in an *in-band* or *out-of-band* fashion.

When we implement an in-band virtual storage network, both data and control flow over the same path. Levels of abstraction exist in the data path, and storage can be pooled under the control of a domain manager. In general, in-band solutions are perceived to be simpler to implement, especially because they do not require special software to be installed in servers (other than conventional multi-pathing software). In-band solutions can also provide caching and advanced functions within the storage network. This can help to improve the performance of existing disk systems and can extend their useful life, and reduce the cost of new storage capacity by enabling the use of lower function, lower cost disk systems, without the loss of performance.

Other advantages include:

- Ability to off load function from the host
- Providing storage management for the SAN
- Performing performance optimizations in the data path
- Supporting host systems not in a cluster
- Supporting multiple heterogeneous hosts
- Integrating well with storage management software
- Releasing the customer from a particular vendor's storage
- Integrating with storage to create a better management picture
- Offering excellent scalability

In an out-of-band implementation, the data flow is separated from the control flow. This is achieved by separating the data and meta-data (data about the data) into different places. Out-of-band virtualization involves moving all mapping and locking tables to a separate server (the meta-data controller) that contains the meta-data of the files.

In an out-of-band solution the servers request authorization to data from the meta-data controller, which grants it, handles locking, and so on. Once they are authorized, servers access the data directly without any meta-data controller intervention. Once a client has obtained access to a file, all I/O will go directly over the SAN to the storage devices. For many operations, the meta-data controller does not even intervene.

Separating the flow of control and data in this manner allows the I/O to use the full bandwidth that a SAN provides, while control could go over a separate network or routes in the SAN that are isolated for this purpose. This results in performance that is nearly equal to local file system performance with all of the benefits and added functionality that come with an out-of-band implementation.

Other advantages include:

- Releasing the customer from a particular vendor’s storage
- Providing storage management for the SAN
- Offering excellent scalability
- Off loading host processing
- Supporting storage management from multiple vendors
- Integrating well with storage management software
- Supporting multiple heterogeneous hosts
- Relatively low overhead in the data path

7.6.4 Virtualization strategies

The IBM strategy is to move the storage device management intelligence out of the server, reducing the dependency of having to implement specialized software, like Logical Volume Managers (LVM), at the server level. We also intend to reduce the requirement for intelligence at the storage subsystem level, which will decrease the dependency on having to implement intelligent storage subsystems.

By implementing at a fabric level, storage control is moved into the network, which gives the opportunity to all for virtualization, and at the same time reduces complexity by providing a single view of storage. The storage network can be used to leverage all kinds of services across multiple storage devices, including virtualization.

By implementing at a file system level, file details are effectively stored on the storage network instead of in individual servers. This design means the file system intelligence is available to all application servers. Doing so provides immediate benefits: a single namespace and a single point of management. This eliminates the need to manage files on a server-by-server basis.

8

Security

Security has always been a major concern for networked systems administrators and users. Even for specialized networked infrastructures, such as SANs, special care has to be taken so that information does not get corrupted, either accidentally or deliberately, or fall into the wrong hands. And, we also need to make sure that at a fabric level the correct security is in place, for example, to make sure that a user does not inadvertently change the configuration incorrectly.

Now that SANs have “broken” the traditional direct-attached storage paradigm of servers being cabled directly to servers, the inherent security that this provided has been lost. The SAN and its resources may be shared by many users and many departments. The SAN may be shared by different operating systems that have differing ideas as to who owns what storage. To protect the privacy and safeguard the storage, SAN vendors came up with a segmentation tool, zoning, to overcome this. The fabric itself would enforce the separation of data so that only those users intended to have access could communicate with the data they were supposed to.

Zoning, however, does not provide security. For example, if data is being transmitted over a link it would be possible to “sniff” the link with an analyzer and steal the data. This is a vulnerability that becomes even more evident when the data itself has to travel outside of the data center, and over long distances. This will often involve transmission over networks that are owned by different carriers.

More often than not, not all data is not encrypted before being sent and this itself means that stealing data could be extremely fruitful. The introduction of multiprotocol devices that allow Fibre Channel hosts to connect to Gigabit Ethernet switches have, in a touch of irony to those that see these as competing technologies, allowed the introduction of IP Security (IPSec) into the Fibre Channel world. There are also third-party devices that will provide encryption using tried and trusted algorithms.

It would be naive to expect that the required level of security can be achieved from any one of the methodologies and technologies, independent of all others, that we will discuss in this chapter. The storage architect needs to understand, and administrators accept, that in a SAN environment, often with a combination of diverse operating systems and vendor storage devices, that some combination of technologies could be required to ensure that the SAN is secure from unauthorized systems and users.

In terms of making the fabric and the data as secure as possible, these are some of the questions that need to be answered:

- How do we stop hosts in the SAN from taking over all the storage (LUNs) that they see?
- How can we segregate operating systems and at what level?
- How can we segregate different applications on the fabric?
- How do we allow a host to access some LUNs and not others?
- How do we provide switch-to-switch security?
- How do we protect the fabric from unauthorized hosts logging in?
- How do we provide users with different authorization levels?
- How do we track, or audit, and changes?

SANs and their ability to make data highly available, need to be tempered by well thought out, and more importantly implementing, security policies that manage how devices interact within the SAN. It is essential that the SAN environment implements a number of safeguards to ensure data integrity, and to prevent unwanted access from unauthorized systems and users.

In the discussions that follow we briefly explore some of the technologies and their associated methodologies that can be used to ensure data integrity, and to protect and manage the fabric. Each technology has advantages and disadvantages; and each must be considered based on a well thought out SAN security strategy, developed during the SAN design phase.

8.1 Security principles

It is a well-known fact that “a chain is only as strong as its weakest link” and when talking about computer security, the same concept applies: there is no point in locking all the doors and then leaving a window open. A secure, networked infrastructure must protect information at many levels or layers, and have no single point of failure.

The levels of defense need to be complementary, and work in conjunction with each other. If you have a SAN, or any other network for that matter, that crumbles after a single penetration, then this is not a recipe for success.

There are a number of unique entities that need to be given consideration in any environment. We discuss some of the most important ones in the topics that follow.

8.1.1 Access control

Access control can be performed both by means of *authentication* and *authorization* techniques:

Authentication Means that the secure system has to challenge the user (usually by means of a password) so that he or she identifies himself.

Authorization Having identified a user, the system will be able to “know” what this user is allowed to do and what they are not.

As true as it is in any IT environment, it is also true in a SAN environment that access to information, and to the configuration or management tools, must be restricted to only those people that are need to have access, and authorized to make changes. Any configuration or management software is typically protected with several levels of security, usually starting with a user ID and password that must be assigned appropriately to personnel based on their skill level and responsibility.

8.1.2 Auditing and accounting

It is essential that an audit trail is maintained for auditing and troubleshooting purposes. Logs should be inspected on a regular basis and archived.

8.1.3 Data security

Whether at rest or in-flight, data security comprises of both data *confidentiality* and *integrity*:

Data confidentiality the system has to guarantee that the information cannot be accessed by unauthorized people, remaining confidential for them but available for only authorized personnel. As shown in the next section, this is usually accomplished by data encryption.

Data integrity the system has to guarantee that the data stored or processed within its boundaries is not altered or tampered with in any way.

This is a security and integrity requirement aiming to guarantee that data from one application or system does not become overlaid, corrupted, or otherwise destroyed, whether intentionally or by accident, by other applications or systems. This may involve some form of authorization, and/or the ability to fence off one system's data from another systems.

This has to be balanced with the requirement for the expansion of SANs to enterprise-wide environments, with a particular emphasis on multi-platform connectivity. The last thing that we want to do with security is to create SAN islands, as that would destroy the essence of the SAN. True cross-platform data sharing solutions, as opposed to data partitioning solutions, are also a requirement. Security and access control also need to be improved to guarantee data integrity.

In the topics that follow, we overview some of the common approaches to securing data that are encountered in the SAN environment. The list is not meant to be an in-depth discussion, but merely an attempt to acquaint the reader with the technology and terminology likely to be encountered when a discussion on SAN security takes place.

8.1.4 Encryption

Encryption is the translation of data into a secret code and is the most effective way to achieve data security. To read an encrypted file you must have access to a secret key, or password or passphrase, that enables you to decrypt it. Unencrypted data is called plain text; encrypted data is referred to as cipher text.

There are two main types of encryption: symmetric encryption and asymmetric encryption (also called public-key encryption).

Symmetric When the same secret password, or key, is used to encrypt a message and decrypt the corresponding cipher text

Asymmetric When one key is used to encrypt a message and another to decrypt the corresponding cipher text.

Figure 8-1 illustrates these two cryptographic schemes.

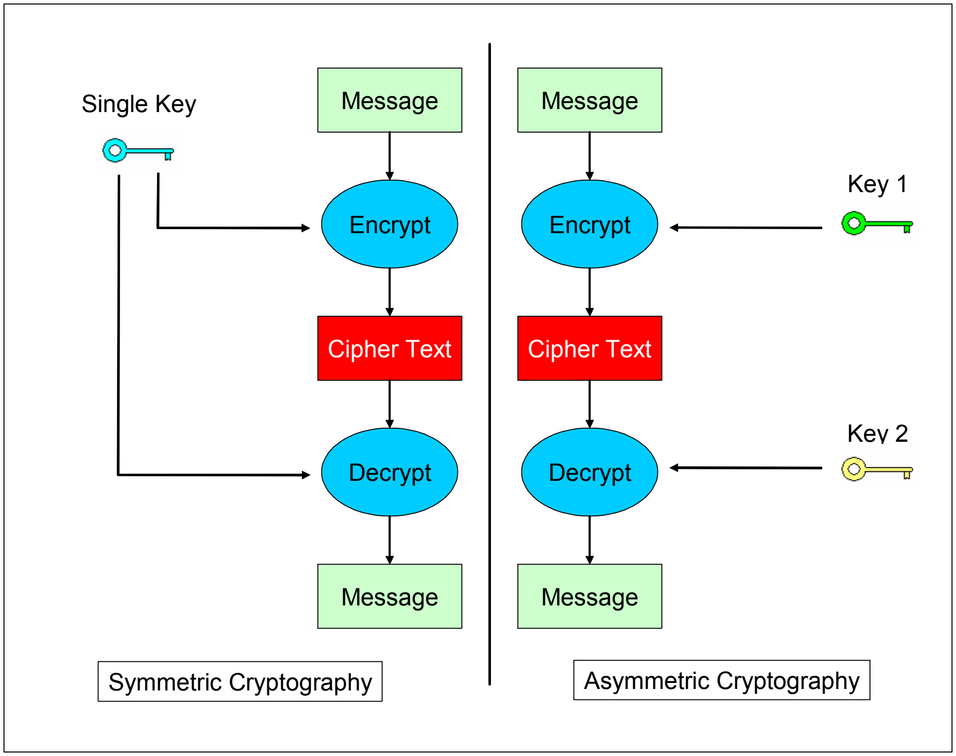


Figure 8-1 Symmetric and asymmetric cryptography

A symmetric crypto-system follows a fairly straightforward philosophy: two parties can securely communicate as long as both use the same *cryptographic algorithm* and possess the same secret key to encrypt and decrypt messages. This is the simplest and most efficient way of implementing secure communication, as long as the participating parties are able to securely exchange secret keys (or passwords).

An asymmetric (or public-key) crypto-system is a cryptographic system that uses a pair of unique keys, usually referred to as public and private keys. Each individual is assigned a pair of these keys to encrypt and decrypt information. A message encrypted by one of these keys can only be decrypted by the other key and vice-versa:

- One of these keys is called a “public key” because it is made available to others for use when encrypting information that will be sent to an individual. For example, people can use a person's public key to encrypt information they want to send to that person. Similarly, people can use the user's public key to decrypt information sent by that person.

- The other key is called “private key” because it is accessible only to its owner. The individual can use the private key to decrypt any messages encrypted with the public key. Similarly, the individual can use the private key to encrypt messages, so that the messages can only be decrypted with the corresponding public key.

This means that exchanging keys is not a security concern: if *A* has a public key and a private key. *A* can send the public key to anyone else. With that public key, *B* can encrypt data to be sent to *A*. Since the data was encrypted with *A’s* public key, *only A* can decrypt that data with his private key. If *A* wants to encrypt data to be sent to *B*, *A* needs *B’s* public key. Figure 8-2 illustrates this process.

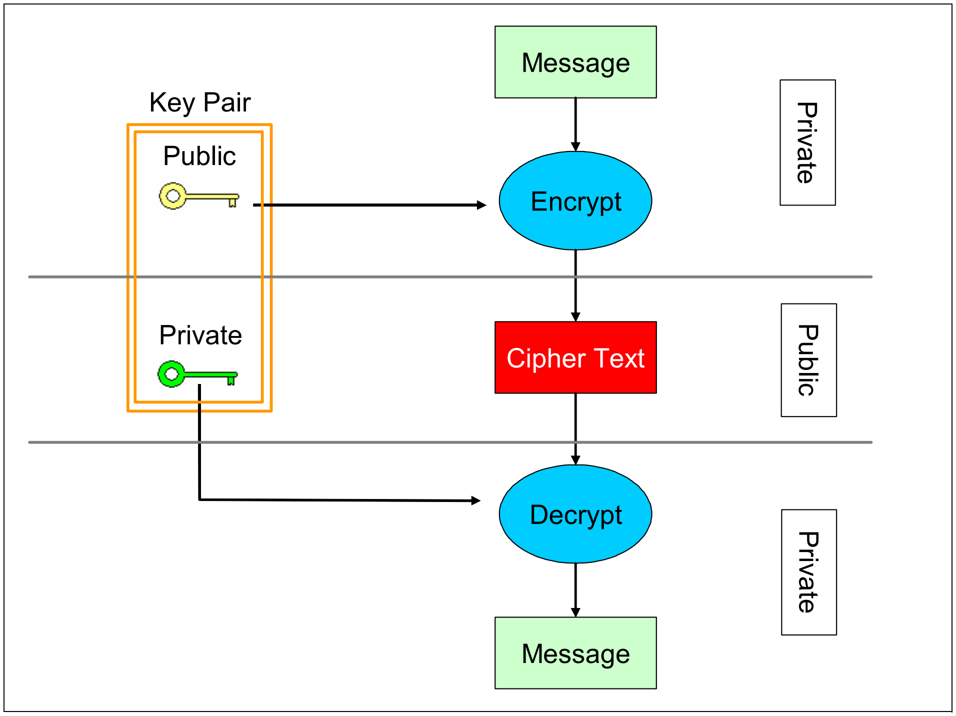


Figure 8-2 Securing a message using asymmetric cryptography

If *A* wants to testify that they were the person that actually sent a document, *A* will encrypt and protect the document with his private key, while others can decrypt it using *A’s* public key; they will know that in this case only *A* could have encrypted this document. Figure 8-3 on page 161 illustrates this process.

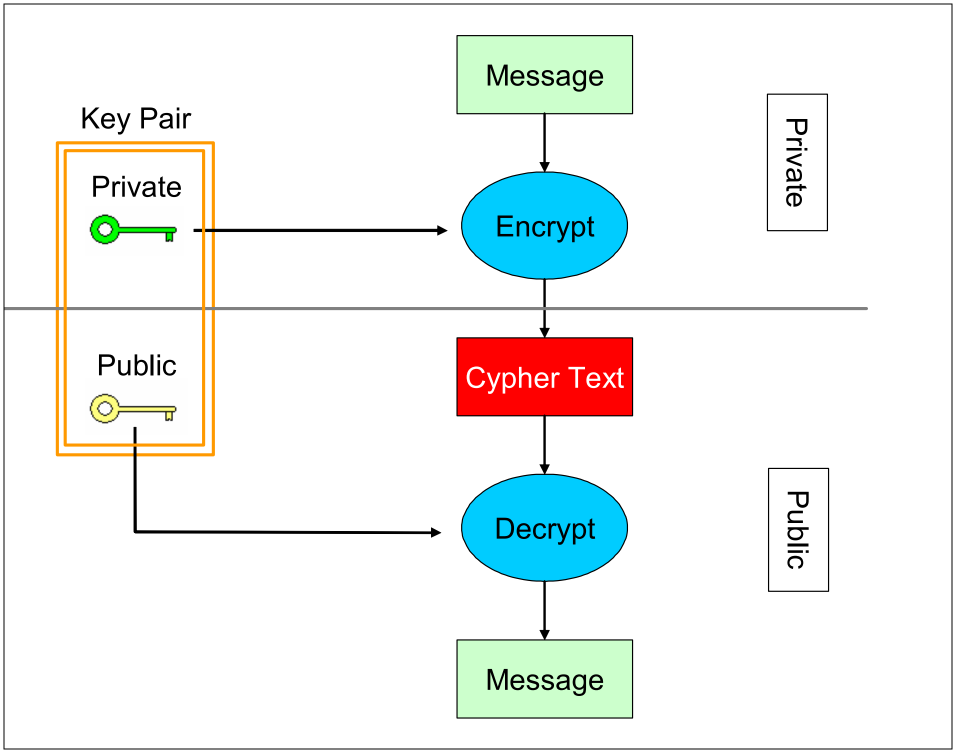


Figure 8-3 Digital signing

The main disadvantage of public-key encryption when compared to symmetric encryption is that it demands much higher computing power to be performed as efficiently. For this reason, most of the current security systems use public-key mechanisms as a way to securely exchange symmetric encryption keys between parties, that then use symmetric encryption for their communication. In this case, the exchanged symmetric secret key (or password) is called *session key*.

However, there is still an issue when talking about public-key crypto-systems: when you initially receive someone's public key for the first time, how do you know that this individual is really who he or she claims to be? If “spoofing” someone's identity is so easy, how do you knowingly exchange public keys? The answer is to use a *digital certificate*. A digital certificate is a digital document issue by a trusted institution that vouches for the identity and key ownership of an individual—it guarantees authenticity and integrity.

In the next sections, some of the most common encryption algorithms and tools are presented.

8.1.5 Encryption schemes

Some of the most popular encryption algorithms in use today are:

DES Data Encryption Standard (DES) is a widely used private-key cryptographic method of data encryption. It uses a 64 bits private key (or password) and, due to its strength, it was restricted from exportation to some countries by the U.S. government. DES originated at IBM in 1977 and was adopted by the U.S. Department of Defense. It is specified in the ANSI X3.92 and X3.106 standards and in the Federal FIPS 46 and 81 standards.

3DES Triple DES or 3DES is based on the DES algorithm developed by an IBM team in 1974 and adopted as a national standard in 1977. 3DES uses three 64-bit long keys (overall key length is 192 bits). The data is encrypted with the first key, decrypted with the second key, and finally encrypted again with the third key. This makes 3DES three times slower than standard DES but offers much greater security. It is the most secure of the DES combinations.

AES Advanced Encryption Standard (AES) is a symmetric 128-bit block data encryption technique developed by Belgian cryptographers Joan Daemen and Vincent Rijmen. The U.S. government adopted the algorithm as its encryption technique in October 2000, replacing the DES encryption it used. AES works at multiple network layers simultaneously. The National Institute of Standards and Technology (NIST) of the U.S. Department of Commerce selected the algorithm, called Rijndael (pronounced Rhine Dahl or Rain Doll), out of a group of five algorithms under consideration.

RSA The RSA is an algorithm for public-key encryption described in 1977 by Ron Rivest, Adi Shamir and Len Adleman at MIT; the letters RSA are the initials of their surnames. It was the first algorithm known to be suitable for digital signing as well as data encryption, and one of the first great advances in public key cryptography. RSA is still widely used in electronic commerce protocols, and is believed to be secure given sufficiently long keys.

ECC Elliptic curve cryptography is an approach to public-key cryptography based on the mathematics of elliptic curves over finite fields. The use of elliptic curves in cryptography was suggested independently by Neal Koblitz and Victor

S. Miller in 1985. Elliptic curves are also used in several integer factorization algorithms that have applications in cryptography, such as, for instance, Lenstra elliptic curve factorization, but this use of elliptic curves is not usually referred to as "elliptic curve cryptography."

Diffie-Hellman

Diffie-Hellman (D-H) key exchange is a cryptographic protocol which allows two parties that have no prior knowledge of each other to jointly establish a shared secret key over an insecure communications channel. This key can then be used to encrypt subsequent communications using a symmetric key cipher.

DSA

The Digital Signature Algorithm (DSA) is a United States Federal Government standard for digital signatures. It was proposed by the National Institute of Standards and Technology (NIST) in August 1991 for use in their Digital Signature Standard (DSS), specified in FIPS 186, adopted in 1993. A minor revision was issued in 1996 as FIPS 186-1, and the standard was expanded further in 2000 as FIPS 186-2. DSA is covered by U.S. Patent 5,231,668, filed July 26, 1991, and attributed to David W. Kravitz, a former NSA employee.

SHA

The Secure Hash Algorithm (SHA) family is a set of related cryptographic hash functions. The most commonly used function in the family, SHA-1, is employed in a large variety of popular security applications and protocols, including TLS, SSL, PGP, SSH, S/MIME, and IPSec. SHA algorithms were designed by the National Security Agency (NSA) and published as a U.S. government standard.

8.1.6 Encryption tools and systems

These are some of the most popular tools and systems that apply the concepts seen in the previous sections to deploy secure functionality to networked environments.

PKI

In cryptography, a public key infrastructure (PKI) is an arrangement which provides for third-party vetting of, and vouching for, user identities. It also allows binding of public keys to users. This is usually carried out by software at a central location together with other coordinated software at distributed locations. The public keys are typically in certificates. The term is used to mean both the certificate authority and related arrangements as well as, more broadly and somewhat confusingly, the use of public key algorithms in electronic communications. The latter sense is erroneous since PKI methods are not required to use public key algorithms.

PGP

Pretty Good Privacy (PGP) is a computer program which provides cryptographic privacy and authentication. The first released version of PGP, by designer and developer Phil Zimmerman, became available in 1991. Subsequent versions have been developed by Zimmerman and others.

SSL

Secure Sockets Layer (SSL) is a protocol developed by Netscape for transmitting private documents over the Internet. SSL works by using a private key to encrypt data that is transferred over a socket connection. Both Netscape Navigator and Internet Explorer® support SSL, and many Web sites use the protocol to obtain confidential user information. URLs that require an SSL connection start with https: instead of http.

TLS

Transport Layer Security is a protocol that guarantees privacy and data integrity between client/server applications communicating over the Internet. The TLS protocol is made up of two layers: the TLS Record Protocol which is layered on top of a reliable transport protocol, such as TCP, it ensures that the connection is private by using symmetric data encryption and it ensures that the connection is reliable. The TLS Record Protocol is also used for encapsulation of higher-level protocols, such as the TLS Handshake Protocol. The TLS Handshake Protocol allows authentication between the server and client and the negotiation of an encryption algorithm and cryptographic keys before the application protocol transmits or receives any data.

SFTP

Secure version of the widely known FTP protocol, also written as S/FTP. SFTP uses SSL to encrypt the entire user session, thereby protecting the contents of files and the user's login name and password from network sniffers. Through normal FTP, usernames, passwords, and file contents are all transferred in plain text.

SSH

Secure Shell (SSH) was developed by SSH Communications Security Ltd., and is a program to log into another computer over a network, to execute commands in a remote machine, and to move files from one machine to another. It provides strong authentication and secure communications over insecure channels. It is a replacement for **rlogin, rsh, rcp,** and **rdist**. SSH protects a network from attacks such as IP spoofing, IP source routing, and DNS spoofing. An attacker who has managed to take over a network can only force ssh to disconnect. He or she cannot play back the traffic or hijack the connection when encryption is enabled. When using ssh's slogin (instead of rlogin), the entire login session, including transmission of password, is encrypted; therefore it is virtually impossible for an outsider to collect passwords.

8.2 Security mechanisms

In this section we briefly introduce some concepts on IP security, which has been around since the first implementations of networked secure systems.

8.2.1 IP security

There a number of standards and products originally developed for local IP networks that have gained a major importance as such networks became world-widely prevalent. In this section, we discuss some of the standards associated with secure network management, which is the ultimate use for IP networks on SANs.

The Simple Network Management Protocol (SNMP) was extended for security functions to SNMPv3. The SNMPv3 specifications were approved by the Internet Engineering Steering Group (IESG) as a full Internet standard in March 2002.

IP security (IPSec) uses cryptographic techniques obtaining management data that can flow through an encrypted tunnel. Encryption makes sure that only the intended recipient can make use of it (RFC 2401). IPSec is widely used to implement Virtual Private Networks (VPN).

Other cryptographic protocols for network management are Secure Shell (SSH) and Transport Layer Security (TLS, RFC 2246). TLS was formerly known as Secure Sockets Layer (SSL). They help ensure secure remote login and other network services over insecure networks.

A common method to build trusted areas in IP networks is the use of firewalls. A firewall is an agent that screens network traffic and blocks traffic it believes to be inappropriate or dangerous. You will use a firewall to filter out addresses and protocols you do not want to pass into your LAN. A firewall will protect the switches connected to the management LAN, and allows only traffic from the management stations and certain protocols that you define.

Finally, another important IP-based security mechanism is the Remote Authentication Dial-In User Service (RADIUS). RADIUS is a distributed security system developed by Lucent Technologies InterNetworking Systems and is used nowadays as a common industry standard for user authentication, authorization, and accounting (RFC 2865). A RADIUS Network Access Server (NAS), which acts as an IP-router or switch in LANs and a SAN switch in SANs, is responsible for performing such functions.

More information about the IPSec working group can be found at:

http://www.ietf.org/html.charters/ipsec-charter.html

8.3 Fibre Channel security

In this topic we discuss Fibre Channel security, that has borrowed most of its security mechanisms from its predecessor technologies, and leverages access and data security on today's SANs.

Since April 2002, the ANSI T11 group has been working on FC-SP, a proposal for the development of a set of methods that allow security techniques to be implemented in a SAN.

Up until now, fabric access of Fibre Channel components was attended to by identification (who are you?). This information could be used later to decide if a device was allowed to attach to storage (by zoning), or if it was just for the propagation of information (for example, attaching a switch to a switch), but it was not a criteria to refuse an inter-switch connection.

As the fabric complexity increases, more stringent controls are required for guarding against malicious attacks and accidental configuration changes. Additionally, increasingly more in-fabric functionality is being proposed and implemented that requires a closer focus on security.

8.3.1 Securing a fabric

In this section some of the current methods for securing a SAN fabric are presented.

Fibre Channel Authentication Protocol

The Switch Link Authentication Protocol (SLAP/FC-SW-3) establishes a region of trust between switches. For an end-to-end solution to be effective, this region of trust must extend throughout the SAN, which requires the participation of fabric-connected devices, such as HBAs. The joint initiative between Brocade and Emulex establishes Fibre Channel Authentication Protocol (FCAP) as the next-generation implementation of SLAP. Customers gain the assurance that a region of trust extends over the entire domain.

FCAP has been incorporated into its fabric switch architecture and has proposed the specification as a standard to ANSI T11 (as part of FC-SP). FCAP is a Public Key Infrastructure (PKI)-based cryptographic authentication mechanism for establishing a common region of trust among the various entities (such as switches and HBAs) in a SAN. A central, trusted third party serves as a guarantor to establish this trust. With FCAP, certificate exchange takes place among the switches and edge devices in the fabric to create a region of trust consisting of switches and HBAs.

The fabric authorization database is a list of the WWNs and associated information like domain IDs of the switches that are authorized to join the fabric.

The fabric authentication database is a list of the set of parameters that allows the authentication of a switch within a fabric. An entry of the authentication database holds at least the switch WWN, authentication mechanism Identifier, and a list of appropriate authentication parameters.

Virtual SANs

In 2004 the T11 committee of the International Committee for Information Technology Standards selected Cisco's Virtual SAN (VSAN) technology for approval by the American National Standard Institute (ANSI) as the industry standard for implementing virtual fabrics. In simple terms, this gives the ability to segment a single physical SAN fabric into many logical, independent SANs.

VSANs offer the capability to overlay multiple hardware-enforced virtual fabric environments within a single physical fabric infrastructure. Each VSAN contains separate (dedicated) fabric services designed for enhanced scalability, resilience, and independence among storage resource domains. This is especially useful in segregating service operations and failover events between high-availability resource domains allocated to different VSANs. Each VSAN contains its own complement of hardware-enforced zones, dedicated fabric services, and management capabilities, just as though the VSAN were configured as a separate physical fabric. Therefore, VSANs are designed to allow more efficient SAN utilization and flexibility, because SAN resources may be allocated and shared among more users, while supporting secure segregation of traffic and retaining independent control of resource domains on a VSAN-by-VSAN basis. Within each VSAN it has its own separate zoning configurations.

Zoning

Zoning allows for finer segmentation of the switched fabric. Zoning can be used to instigate a barrier between different environments. Only the members of the same zone can communicate within that zone, and all other attempts from outside are rejected. One important point to be aware of is that it is not providing a security feature, it provides separation. We include it in the security chapter to highlight it as an example of something that is commonly used to provide "security", but in actual fact does not.

For example, it may be desirable to separate a Windows NT environment from a UNIX environment. This is very useful because of the manner in which Windows attempts to claim all storage for itself.

Zoning also introduces the flexibility to manage a switched fabric to meet different user group objectives.

Zoning can be implemented in two ways:

- Hardware zoning
- Software zoning

These forms of zoning are different, but are not necessarily mutually exclusive. Depending upon the particular manufacturer of the SAN hardware, it is possible for hardware zones and software zones to overlap. This adds to flexibility, but can make the solution complicated, increasing the need for good management software and documentation of the SAN.

Hardware zoning is based on the physical fabric port number. The members of a zone are physical ports on the fabric switch.

The availability of hardware enforced zoning and the methods to create hardware enforced zones depend upon the switch hardware used.

Figure 8-4 on page 169 shows an example of zoning based on the switch port numbers.

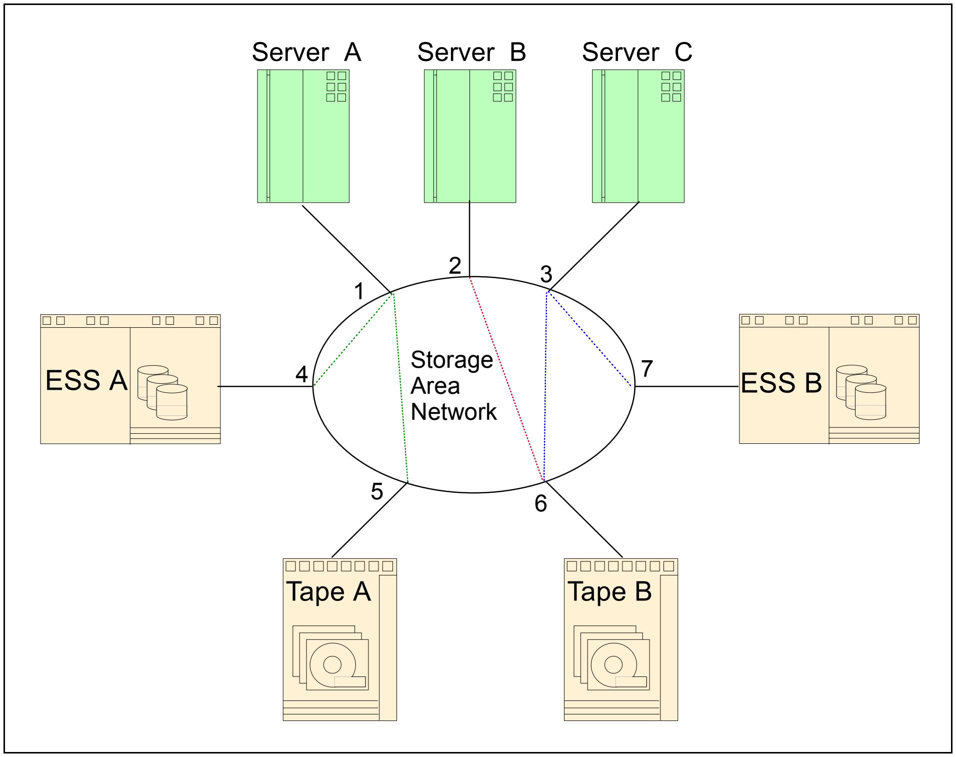


Figure 8-4 Zoning based on the switch port number

In this example, port-based zoning is used to restrict server A to only see storage devices that are zoned to port 1, that is, ports 4 and 5.

Server B is also zoned so that it can only see from port 2 through to port 6.

Server C is zoned so that it can see both ports 6 and 7, even though port 6 is also a member of another zone.

A single port can also belong to multiple zones.

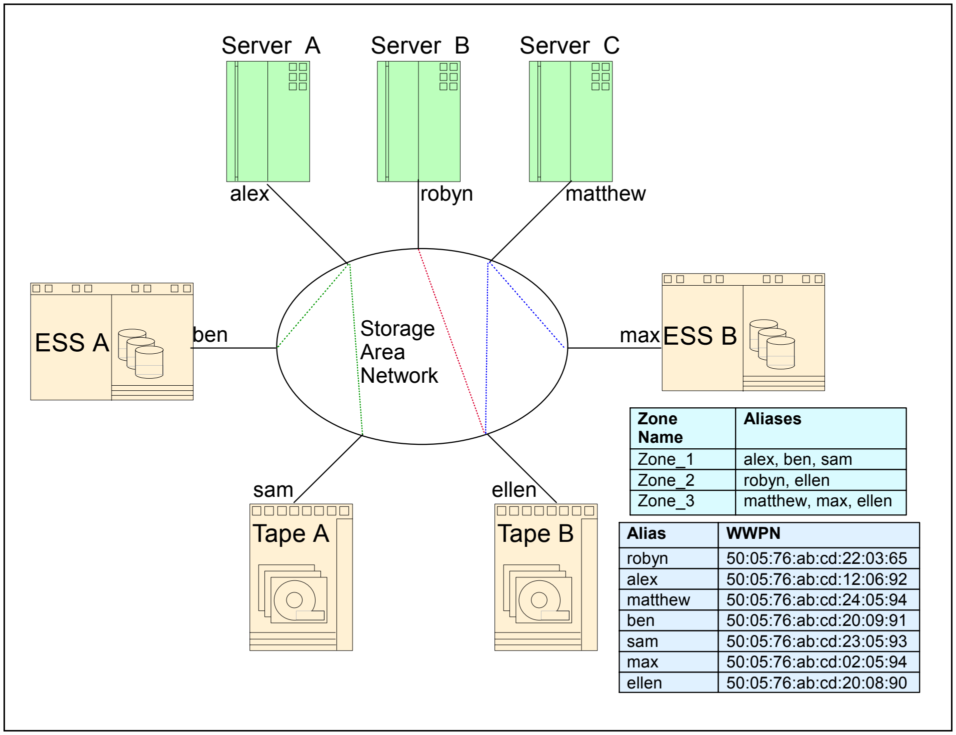
One of the disadvantages of hardware zoning is that devices have to be connected to a specific port, and the whole configuration could become unusable when the device is connected to a different port. In cases where the device connections are not permanent, the use of software zoning is recommended.

Software zoning is implemented by the fabric operating systems within the fabric switches. When using software zoning, the members of the zone can be defined using their WWN and WWPN.

With software zoning there is no need to worry about the device’s physical connections to the switch. If you use WWNs for the zone members, even when a device is connected to another physical port, it will still remain in the same zoning definition, because the device’s WWN remains the same. The zone follows the WWN.

Shown in Figure 8-5 is an example of WWN-based zoning. In this example symbolic names (aliases) are defined for each WWN in the SAN to implement the same zoning requirements, as shown in the previous Figure 8-4 on page 169 for port zoning:

- Zone_1 contains the aliases *alex*, *ben*, and *sam*, and is restricted to only these devices.
- Zone_2 contains the aliases *robyn* and *ellen*, and is restricted to only these devices.
- Zone_3 contains the aliases *matthew*, *max*, and *ellen*, and is restricted to only these devices.



Zone Name	Aliases
Zone_1	alex, ben, sam
Zone_2	robyn, ellen
Zone_3	matthew, max, ellen

Alias	WWPN
robyn	50:05:76:ab:cd:22:03:65
alex	50:05:76:ab:cd:12:06:92
matthew	50:05:76:ab:cd:24:05:94
ben	50:05:76:ab:cd:20:09:91
sam	50:05:76:ab:cd:23:05:93
max	50:05:76:ab:cd:02:05:94
ellen	50:05:76:ab:cd:20:08:90

Figure 8-5 Zoning based on the device’s WWN

There are some potential security issues with software zoning. For example, a specific host can make a direct connection to the storage device, while doing device discovery, without asking SNS for the information it has in the software zoning table to see which storage devices are allowed for that host.

Hardware zoning provides the highest level of security.

Persistent binding

Server-level access control is called persistent binding. Persistent binding uses configuration information stored on the server, and is implemented through the server's HBA driver. The process binds a server device name to a specific Fibre Channel storage volume or logical unit number (LUN), through a specific HBA and storage port WWN. Or, put in more technical terms, it is a host-centric way to direct an operating system to assign certain SCSI target IDs and LUNs.

LUN masking

One approach to securing storage devices from hosts wishing to take over already assigned resources is logical unit number (LUN) masking. Every storage device offers its resources to the hosts by means of LUNs. For example, each partition in the storage server has its own LUN. If the host (server) wants to access the storage, it needs to request access to the LUN in the storage device. The purpose of LUN masking is to control access to the LUNs. The storage device itself accepts or rejects access requests from different hosts. The user defines which hosts can access which LUN by means of the storage device control program. Whenever the host accesses a particular LUN, the storage device will check its access list for that LUN, and it will allow or disallow access to the LUN.

Port binding

To provide a higher level of security, you can also use port binding to bind a particular device (as represented by a WWN) to a given port that will not allow any other device to plug into the port, and subsequently assume the role of the device that was there. The reason for this is that the “rogue” device that was inserted will have a different WWN that the port was bound to.

8.3.2 Zoning, masking and binding

Although neither of these can be classed as security products or mechanisms, combining all their functionality together can make the SAN more secure than it would be without them.

8.3.3 Data security

In order to provide the equivalent security functions that are implemented in the LAN, the ANSI T11-group is considering a range of proposals for connection authentication and integrity, which can be recognized as the FC adoption of the IP security standards. These standards propose to secure FC traffic between all FC ports and the domain controller. These are some of the methods that will be used:

- FCPAP refers to Secure Remote Password Protocol (SRP), RFC 2945.
- DH-CHAP refers to Challenge Handshake Authentication Protocol (CHAP), RFC 1994.
- FCSec refers to IP Security (IPsec), RFC 2406.

  The FCSec aim is to provide authentication of these entities:

  – Node-to-node
  – Node-to-switch
  – Switch-to-switch

An additional function that may be possible to implement is frame level encryption.

The ability to perform switch-to-switch authentication in FC-SP enables a new concept in Fibre Channel: The secure fabric. Only switches that are authorized and properly authenticated are allowed to join the fabric.

Whereas, authentication in the secure fabric is twofold: The fabric wants to verify the identity of each new switch before joining the fabric, and the switch that is wanting to join the fabric wants to verify that it is connected to the right fabric. Each switch needs a list of the WWNs of the switches authorized to join the fabric, and a set of parameters that will be used to verify the identity of the other switches belonging to the fabric.

Manual configuration of such information within all the switches of the fabric is certainly possible, but not advisable in larger fabrics. And there is the need of a mechanism to manage and distribute information about authorization and authentication across the fabric.

Accountability

Although not a method for protecting data, it is a method by which an administrator is able to track any form of change within the network.

8.4 Best practices

As we said before, you may have the most sophisticated security system installed in your house — but it is not worth anything if you leave the window open. Some of the security best practices at a high level, that you would expect to see at the absolute minimum, are:

- Default configurations and passwords should be changed.
- Configuration changes should be checked and double checked to ensure that only the data that is supposed to be accessed can be accessed.
- Management of devices usually takes a “telnet” form—with encrypted management protocols being used.
- Remote access often relies on unsecured networks. Make sure that the network is secure and that some form or protection is in place to guarantee only those with the correct authority are allowed to connect.
- Make sure that the operating systems that are connected are as secure as they ought to be, and if the operating systems are connected to an internal and external LAN, that this cannot be exploited. Access may be gotten by exploiting loose configurations.
- Assign the correct roles to administrators.
- Ensure the devices are in physically secure locations.
- Make sure the passwords are changed if the administrator leaves. Also ensure they are changed on a regular basis.

Finally, the SAN security strategy in its entirety must be periodically addressed as the SAN infrastructure develops, and as new technologies emerge and are introduced into the environment.

These will not absolutely guarantee that your information is 100 percent secure, but they will go some way to ensuring that all but the most ardent “thieves” are kept out.

9

The IBM product portfolio

This chapter provides an overview of the IBM TotalStorage SAN components, that either IBM OEMs, or a reseller, has an agreement for. We include some products that have been withdrawn from marketing, as it is likely that they will still be encountered. We also provide descriptions of the storage and virtualization devices that are commonly encountered in the SAN environment.

9.1 Why an IBM TotalStorage SAN?

IBM TotalStorage SAN solutions provide integrated small and medium business (SMB) and enterprise solutions with multiprotocol local, campus, metropolitan, and global storage networking. IBM provides the choice of Brocade, Cisco, CNT, Emulex, and McDATA switches and directors. IBM SAN solutions are offered with worldwide service and end-to-end support by IBM Business Partners and IBM Global Services.

9.2 Entry SAN switches

These switches provide solutions for the SMB customer:

- Very cost conscious SMB customers with limited technical skills
- Integrated, simple storage consolidation and data protection solutions
  - Homogenous Windows/Linux servers
  - xSeries server sales channels
  - IBM DS Series and LTO Storage
  - High availability with dual fabric deployment
- Support of IBM TotalStorage devices and IBM Tivoli Storage Manager
- Integrated solutions with worldwide IBM support

9.2.1 IBM System Storage SAN10Q

The IBM System Storage SAN10Q (6918-10X) is an affordable, capable, and extremely easy to use switch that allows even the smallest businesses to enjoy the benefits of networked storage without having to hire a SAN expert.

With a small form factor of only 0.9kg (2lbs) and 1U high, the IBM System Storage SAN10Q is a true space-saver—requiring only one-half a rack slot. This means two Fibre Channel switches can be racked in a single slot for a total of 20 ports. All ports are auto-discovering and self-configuring, which helps allow maximum port density and power with a minimum investment.

Simple software wizards lead you through the installation and configuration of a SAN. SANSurfer Switch Manager or SANSurfer Express management application software is included and provides convenient access to your IBM switches and HBAs. It is designed to allow you to manage and provision your compliant storage from one tool.

The common features are:

- Fabric port speed: 4Gbps, full-duplex, auto-negotiating for compatibility with existing 2-Gbps and 1-Gbps devices.
- Fabric latency: Fabric Point-to-Point Bandwidth: up to 848MBps full duplex per port.
- Fabric Aggregate Bandwidth: Single chassis—over 40Gbps (full duplex) end-to-end.
- Maximum frame sizes: 2148 bytes (2112-byte payload).
- Per-port buffering: ASIC-embedded memory (non-shared) and 8-credit zero wait for each port.
- Small and capable — 1U, 4Gbps, 10-port, half-width rack.
- Flexible — rack or standalone form factor.
- Designed to improve manageability — no-wait routing helps maximize performance independent of data traffic.
- Simple to use — auto-sensing, self-configuring ports.
- Logical choice — intuitive and affordable migration from direct attached storage to SAN.
- Complete package — SANSurfer Express software helps simplify switch installation, managing and fabric scaling.

Figure 9-1 shows the SAN10Q.

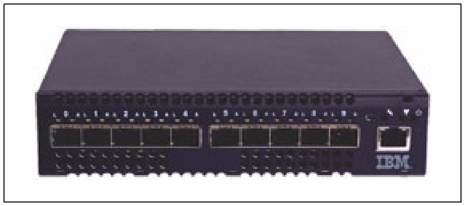


Figure 9-1 SAN10Q

9.2.2 IBM TotalStorage SAN16B-2

The IBM TotalStorage SAN16B02 (2005-B16), shown in Figure 9-2, is a high performance, scalable and simple-to-use fabric switch designed to be the foundation for small to medium-size SANs. It provides an 8, 12 or 16 port 4 Gigabits per second (Gbps) fabric for servers running Microsoft Windows, UNIX, Linux, NetWare and OS/400 operating systems, server clustering, infrastructure simplification and business continuity solutions. The SAN16B-2 includes Advanced Web Tools, an easy-to-use configuration wizard designed to simplify setup and ongoing maintenance for novice users. Optional advanced functions are available for intelligent SAN management and monitoring plus full participation in an IBM TotalStorage SAN b-type extended fabric.

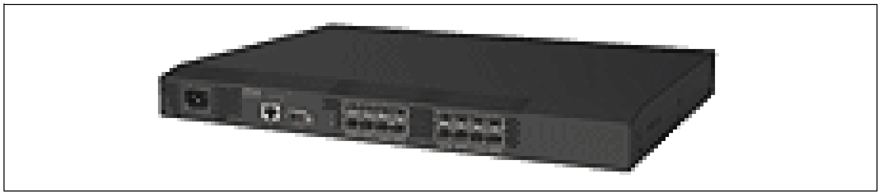


Figure 9-2 SAN16B-2

The IBM TotalStorage SAN16B-2 fabric switch is designed specifically to address the needs of small to medium-size SAN environments. It can be used to create a wide range of high performance SAN solutions, from simple single-switch configurations to larger multi-switch configurations which support fabric connectivity and advanced business continuity capabilities. Some of the main features are:

- 4 Gb Long Wave Transceivers supporting either 4 or 10 km distances.
- Extended Fabric Activation.
- Web Tools EZQuickly guides novice users through setup Help simplify administration and configuration.
- Auto-sensing 4, 2 and 1 Gbps ports designed to:
  - Enable high performance and improved utilization.
  - Enable rapid switch deployment with minimal configuration.
  - Provide easy installation and manageability.
  - Offer future readiness for 4 Gbps capable server and storage devices.
- Hot swappable SFPs that provide fast and easy replacement of optical transceivers with no tools required and support for both short wave and long wave optical transceivers.
- Provides Inter-Switch Link (ISL) Trunking support for up to 4-port ISLs provides total bandwidth of up to 16 Gbps, which helps eliminate packet rerouting and possible application failure due to link failure.
- Efficient 1U design providing high-density configurations in both 8 and 16-port configurations.
- N+1 cooling to help maximize fabric uptime when a fan fails.
- Provides Hot code activation functions that are designed to:
  - Significantly reduces software installation time.
  - Enable fast software upgrades.
  - Help eliminate disruption to existing fabric.

  - Help reduce administrative overhead.

- Provides Advanced Hardware-enforced zoning to help protect against non-secure, unauthorized and unauthenticated network and management access and World Wide Name spoofing.
- Offers upgradeable architecture for scalable SAN fabrics. Optional features enable upgrades to support more complex fabrics and modular upgrades enable a "pay as you grow" approach to adding features when needed.
- Investment protection for existing fabrics, providing backward compatibility with all IBM SAN b-type products, and allowing interoperation with other SAN products manufactured by Brocade.

IBM TotalStorage SAN16B-2 Express Model

The IBM TotalStorage SAN16B-2 Express Model fabric switch is designed specifically to meet the needs of small to medium-size SAN environments. It can be used to create a wide range of high performance SAN solutions, from simple single-switch configurations to larger multi-switch configurations which support fabric connectivity and advanced business continuity capabilities.

A single SAN16B-2 Express Model switch can serve as the cornerstone of a Storage Area Network for those who want to obtain the benefits of storage consolidation and are just beginning to implement Fibre Channel storage systems. Such an entry-level configuration can consist of one or two Fibre Channel links to a disk storage array or to an LTO tape drive. An entry-level eight-port storage consolidation solution can support up to seven servers with a single path to either disk or tape. The Ports on Demand feature is designed to enable a base switch to grow to sixteen ports to support more servers and more storage devices without taking the switch offline.

A high-availability solution can be created with redundant switches. This capability is ideal for server clustering environments. Such a configuration can support from six to fourteen servers, each with dual Fibre Channel adapters cross-connected to redundant SAN16B-2 switches which are cross-connected to a dual-controller storage system.

While the focus of the SAN16B-2 is as the foundation of small to medium-sized SANs, it can be configured to participate as a full member in an extended fabric configuration with other members of the IBM TotalStorage SAN b-type family. This capability provides investment protection as SAN requirements evolve and grow over time. Common features include:

- Extended Fabric Activation.
- Standard configuration includes 8-port activation, eight shortwave SFPs and capability to attach to hosts and storage devices

- Four-port activation option enables "pay-as-you-grow" scalability to 12 or 16 ports.
- No previous SAN experience needed to use EZSwitchSetup wizard.
- Little or no management required in small fabric configuration.
- Advanced Web Tools provides intuitive, graphic switch management.
- Supports IBM e-server xSeries, and pSeries servers plus selected non-IBM servers.
- Full compatibility with IBM TotalStorage SAN b-type switches and directors (IBM 2005, 2109 and 3534) helps protect switch investment.
- Ideal for use as an edge-switch in larger SANs.

9.2.3 IBM TotalStorage SAN16M-2

The IBM TotalStorage SAN16M-2 is a high-performance, scalable and simple-to-use fabric switch designed to be the foundation for small to medium-sized SANs. The SAM16M-2 provides an 8, 12 or 16 port 4 Gigabits per second (Gbps) fabric for servers running almost any operating system. The base switch includes dual power, zoning, EFCM basic management and an easy-to-use installation wizard designed to simplify setup and ongoing maintenance for novice users.

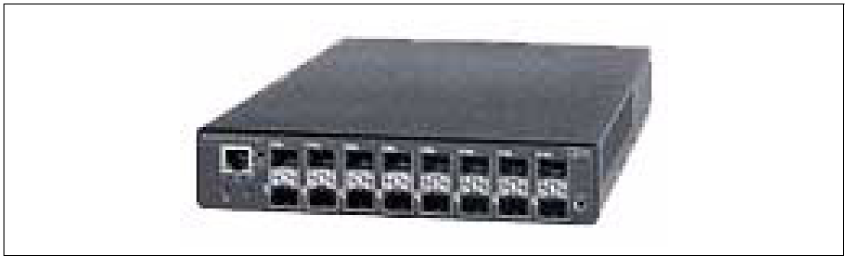


Figure 9-3 SAN16M-2 fabric switch

The IBM TotalStorage SAN16M-2 fabric switch, as shown in Figure 9-3 on page 180, can be used to create a wide range of high performance SAN solutions, from simple single-switch configurations to larger multi-switch configurations which support fabric connectivity and advanced business continuity capabilities. Business continuity solutions include data protection with IBM TotalStorage tape libraries and devices and IBM Tivoli Storage Manager data protection software. Infrastructure simplification solutions for IBM e-server xSeries servers include storage consolidation and high-availability server clustering with IBM TotalStorage disk storage arrays.

A single SAN16M-2 switch can serve as the cornerstone of a Storage Area Network for those who want to obtain the benefits of storage consolidation and are just beginning to implement Fibre Channel storage systems. Such an entry-level configuration can consist of one or two Fibre Channel links to a disk storage array or to an LTO tape drive. An entry-level eight-port storage consolidation solution can support up to seven servers with a single path to either disk or tape. The FlexPort feature is designed to enable a base switch to grow to sixteen ports, in four port increments, to support more servers and more storage devices without taking the switch offline.

A high-availability solution can be created with redundant switches. This capability is ideal for server clustering environments. Such a configuration can support from six to fourteen servers, each with dual Fibre Channel adapters cross-connected to redundant SAN16M-2 switches which are cross-connected to a dual-controller storage system.

While the focus of the SAN16M-2 is as the foundation of small to medium-sized SANs, it can be configured to participate as a full member in a tier enterprise SAN with other members of the IBM TotalStorage SAN m-type family. This capability provides investment protection as SAN requirements evolve and grow over time. Some of the main features are:

- Setup and backup/restore Wizards to quickly guide novice users through setup, simplifying administration and configuration tasks.
- Auto-sensing 4, 2 and 1 Gbps ports designed to:
  - Enable high performance and improved utilization
  - Enable rapid switch deployment with minimal configuration.
  - Provide easy installation and manageability
  - Offer future readiness for new 4 Gbps server and storage devices
- 4 Gbps longwave 4/10 km SFP transceivers for the IBM TotalStorage SAN16M-2
- Storage Network Services (SNS) Open Systems firmware package
- Provides hot swappable SFPs for fast and easy replacement of optical transceivers with no tools required.
- Designed for high density packaging to enable redundant switch installation on single 1RU shelf.
- Designed with redundant, hot pluggable power that allows switch to remain online if one power supply fails.
- Avoids switch replacement with risk of re-cabling errors.
- Provides HotCAT online code activation that will help significantly to reduce software installation time, enable fast software upgrades, help eliminate disruption to existing fabric, and help reduce administrative overhead.

- The FlexPort Expansion feature provides modular upgrades enabling a “pay as you grow” strategy as workload increases.
- The Open Trunking feature is designed to help simplify deployment of tiered enterprise SAN solutions and improve throughput with automatic traffic management.
- The Element Manager feature enables deployment in tiered SAN infrastructures with EFCM management as requirement grow.
- Wide range of IBM supported servers and storage products
- Flexible choice of servers, storage, software and switches for creation of infrastructure simplification and business continuity solutions with worldwide IBM service and support.
- Provides compatibility with IBM SAN m-type products giving switch investment protection for existing fabrics.
- Allows interoperation with other IBM SAN products manufactured by McDATA.

IBM TotalStorage SAN16M-2 Express Model

The IBM TotalStorage SAN16M-2 Express Model fabric switch is designed specifically to meet the needs of small to medium-size SAN environments. It can be used to create a wide range of high performance SAN solutions, from simple single-switch configurations to larger multi-switch configurations which support fabric connectivity and advanced business continuity capabilities.

A single SAN16M-2 switch can serve as the cornerstone of a Storage Area Network for those who want to obtain the benefits of storage consolidation and are just beginning to implement Fibre Channel storage systems. Such an entry-level configuration can consist of one or two Fibre Channel links to a disk storage array or to an LTO tape drive. An entry-level eight-port storage consolidation solution can support up to seven servers with a single path to either disk or tape. The FlexPort feature is designed to enable a base switch to grow to sixteen ports, in four port increments, to support more servers and more storage devices without taking the switch offline.

A high-availability solution can be created with redundant switches. This capability is ideal for server clustering environments. Such a configuration can support from six to fourteen servers, each with dual Fibre Channel adapters cross-connected to redundant SAN16M-2 switches which are cross-connected to a dual-controller storage system. Features include:

- 4 Gb Long Wave Transceivers supporting either 4 or 10 km distances.

9.3 Midrange SAN switches

The IBM Midrange SAN switches provide scalable and affordable SMB and enterprise solutions.

- Cost conscious SMB customers with limited technical skills
- Integrated, scalable, high-availability IBM Virtualization family solutions
- Heterogeneous Windows, Linux, iSeries, UNIX, and mainframe servers
- xSeries, iSeries, pSeries, and zSeries Server sales channels
- IBM FAStT, ESS, LTO, and ETS storage
- Support the IBM TotalStorage Virtualization family, TotalStorage devices, IBM Tivoli Storage Manager, SAN Manager, SRM and Multiple Device Manager
- Integrated solutions at affordable prices with worldwide IBM support and IBM TotalStorage Solution Center (TSSC) services

9.3.1 IBM TotalStorage SAN32B-2

The IBM TotalStorage SAN32B-2 (2005-B32) high-performance fabric switch provides 16, 24, and 32 port, 4 Gbps switching for open server midrange and enterprise infrastructure simplification and business continuity SAN solutions. It provides full interoperability with IBM TotalStorage SAN b-type switches and directors that help protect switch investments. This is shown in Figure 9-4

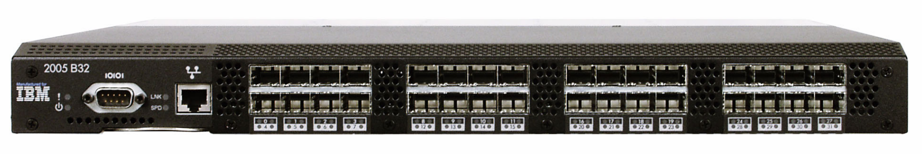


Figure 9-4 SAN32B-2

IBM TotalStorage SAN b-type switches and directors provide the performance, scalability, high availability and manageability required to evolve entry SAN solutions into large enterprise SAN solutions. It gives the opportunity to initially deploy separate SAN solutions at the departmental and data center levels and then to consolidate them into an integrated enterprise SAN as experience and requirements grow and change. The main features are:

- Simple-to-use midrange and enterprise infrastructure simplification and business continuity SAN solutions.
- 4 Gb Long Wave Transceivers supporting either 4 or 10 km distances

- Designed for high-performance with 4 Gigabit per second (Gbps) ports and enhanced inter-switch link (ISL) trunking with up to 32 Gbps per data path.
- Pay-as-you-grow scalability with Ports on Demand features.
- Designed to support high availability with redundant, hot-swappable fans and power supplies and nondisruptive software upgrades.
- Multiple-management options for first-time storage area network (SAN) users and complex enterprise SAN consolidation solutions.
- Interoperability with IBM TotalStorage SAN b-type switch family helps protect switch investment.

IBM TotalStorage SAN32B-2 Express Model

The SAN32B-2 Express Model is a high performance midrange fabric switch provides 16, 24 and 32-port, 4 Gigabit per second (Gbps) fabric switching for Windows NT/2000, UNIX and OS/400 servers. The base switch offers Advanced Zoning, Fabric Watch, WEBTOOLS, dual replaceable power supplies and 16-ports activated. Ports on Demand features support "pay-as-you-grow" scalability. This Model is particularly suitable for SMB customers.

A wide range of IBM TotalStorage midrange storage area network (SAN) infrastructure simplification and business continuity solutions can be created with the IBM TotalStorage SAN32B-2 Express Model fabric switch. Infrastructure simplification solutions for IBM @server xSeries, iSeries and pSeries families of servers include storage consolidation and high-availability server clustering with IBM TotalStorage disk storage arrays. Business continuity solutions include data protection with IBM TotalStorage tape libraries and devices, and IBM Tivoli Storage Manager data protection software. Some of the main features are:

- Multiple management options for first-time SAN users.
- 4 Gb longwave SFP transceivers supporting 4 and 10 Km.
- Designed for high performance with 4 Gigabit per second (Gbps) throughput on all ports and enhanced ISL-Trunking with up to 32 Gbps per data path.
- Designed to support high availability with redundant, hot-swappable fans and power supplies and nondisruptive software upgrades.
- Standard configuration enables 16 ports and includes sixteen 4 Gbps shortwave small form-factor pluggable (SFP) optical transceivers, Advanced Zoning, Fabric Watch, WEBTOOLS and dual replaceable power supplies.
- Each port can auto-negotiate to 1 Gbps, 2 Gbps or 4 Gbps speed depending on the speed of the device at the other end of the link.
- 8-port Upgrade Express Option enables "pay-as-you-grow" upgrades to 24 and 32 ports and includes eight 4 Gbps shortwave SFPs.

- Up to 256 Gbps throughput is possible with a 32-port configuration.
- Optional longwave SFPs are available for link distances of up to 10 km, 35 km or 80 km.
- Optional Advanced Security Activation, Extended Fabric Activation, Fabric Manager V4 Maximum Domains, and Remote Switch Activation features provide enhanced capabilities needed in larger SAN and extended distance configurations.
- Interoperability with IBM TotalStorage SAN b-type switch family helps protect switch investment.

9.3.2 IBM System Storage SAN64B-2

The IBM System Storage SAN64B-2 fabric switch, as shown in Figure 9-5, joins the SAN16B-2, SAN32B-2 and SAN256B as a new member of the System Storage SAN b-type 4 Gbps switch family. The SAN64B-2 addresses the needs of storage networking demand by providing improved performance with higher throughput capability, and enhanced port density.

Figure 9-5 The IBM System Storage SAN64B-2

The main features of the SAN64B-2 are:

- 4 Gbps port-to-port throughput with auto-sensing capability for connecting to existing 1, 2, and future 4 gigabit host servers, storage, and switches.
- Up to 64 non-blocking ports with full duplex throughput at 1, 2, or 4 Gbps link speeds.
- High availability features - automatic path routing, and nondisruptive firmware upgrades.
- Scalable ports on demand 32, 48 or 64 ports to accommodate a broad range of connectivity solutions, in 2U form factor for enhanced port density and space utilization.
- Scalability mid-range to large SAN fabric environments
- Open FCP support.

- Full compatibility with existing TotalStorage SAN b-type switches (IBM 2005, 2109, and 3534)
- Base B64 model firmware features - Fabric Watch, Full Fabric, Advanced Zoning and Web Tools.
- Optional features - Additional Port Activation, Advanced Security, Enhanced ISL Trunking, and Performance Monitoring.
- Support for 4 Gbps long wave and 4 Gbps short wave small form-factor pluggable (SFP) optic transceivers

9.3.3 IBM TotalStorage SAN32M-2

The IBM TotalStorage SAN32M-2 is a simple-to-use SAN switch with ease-of-installation and ease-of-use features designed specifically for the needs of medium-size and enterprise environments. The SAN32M-2 provides a foundation for new infrastructure simplification and business continuity solutions for servers running Microsoft Windows, Linux, NetWare and OS/400,   AIX, z/OS operating systems.

Figure 9-6 The IBM TotalStorage SAN32M-2 Fabric switch

The IBM TotalStorage SAN32M-2 fabric switch is designed specifically to address the needs of medium-size and enterprise SAN environments. It can be used to create a wide range of high performance SAN solutions, from simple single-switch configurations to larger multi-switch configurations which support fabric connectivity and advanced business continuity capabilities. Infrastructure simplification solutions for IBM xSeries, iSeries, pSeries and zSeries servers include storage consolidation and high-availability server clustering with IBM TotalStorage disk storage arrays. Business continuity solutions include data protection with IBM TotalStorage tape libraries and devices and IBM Tivoli Storage Manager data protection software.

A single SAN32M-2 switch can serve as the cornerstone of a Storage Area Network for those who want to obtain the benefits of storage consolidation and are just beginning to implement Fibre Channel storage systems. Such an entry-level configuration can consist of one or two Fibre Channel links to a disk storage array or to an LTO tape drive. An entry-level sixteen-port storage consolidation solution can support up to fifteen servers with a single path to either disk or tape. The FlexPort feature is designed to enable a base switch to grow to sixteen ports, in eight port increments, to support more servers and more storage devices without taking the switch offline. A high-availability solution can be created with redundant switches. This capability is ideal for server clustering environments. Such a configuration can support from fourteen to thirty servers, each with dual Fibre Channel adapters cross-connected to redundant SAN32M-2 switches which are cross-connected to a dual-controller storage system.

While the SAN32M-2 can be the foundation of medium-sized SANs, it can be configured to participate as a full member in a tier enterprise SAN with other members of the IBM TotalStorage SAN m-type family. This capability helps provide investment protection as SAN requirements evolve and grow over time. The main features are:

- Setup and backup/restore Wizards quickly guide novice users through setup to help simplify administration and configuration
- Auto-sensing 4, 2 and 1 Gbps ports are designed to enable: high performance and improved utilization, rapid switch deployment with minimal
- configuration needed
- Offer future readiness for new 4 Gbps server and storage devices
- Supports 4 Gbps longwave 4/10 km SFP transceivers for the IBM TotalStorage SAN32M-2 switch
- Storage Network Services (SNS) Open Systems, Mainframe and Mainframe Cascading firmware packages
- FICON CUP Zoning
- Hot swappable SFPs provide fast and easy replacement of optical transceivers with no tools required
- The SAN32M02 is designed with high density packaging that enables installation of the 32-port switch in a single 1RU height rack
- Redundant, hot pluggable power designed to allow switch to remain online if one power supply fails, and to avoid switch replacement with risk of re-cabling errors
- HotCAT online code activation that is designed to help significantly reduce software installation time, enables fast software upgrades and helps to eliminate disruption to existing fabric when installing updates
- The FlexPort Expansion feature offers modular upgrades enabling a "pay as you grow" strategy as workload increases
  - 8 Port FlexPort Expansion kit

- The Open Trunking feature is designed to help simplify deployment of tiered enterprise SAN solutions and improve throughput with automatic traffic management
- The Element Manager feature enables deployment in tiered SAN infrastructures with EFCM management as requirements grow.
- The FICON Management Server enables cost-effective, switch-based mainframe FICON solutions with IBM TotalStorage DS6000 series.
- Compatibility with IBM SAN m-type products, helping to provide switch investment protection for existing fabrics.
- Allows interoperation with other IBM SAN products manufactured by McDATA.

IBM TotalStorage SAN32M-2 Express Model

The IBM TotalStorage SAN32M-2 Express Model is a simple-to-use SAN switch with ease-of-installation and ease-of-use features designed specifically for the needs of medium-size and enterprise environments. The SAN32M-2 provides a foundation for new infrastructure simplification and business continuity solutions for servers running Microsoft, Windows, Linux, NetWare operating systems. Its high-performance 4 Gigabit per second links with pay-as-you-grow FlexPort scalability enables growth from 16 to 24 to 32 ports.

The IBM TotalStorage SAN32M-2 Express fabric switch can be used to create a wide range of high performance SAN solutions, from simple single-switch configurations to larger multi-switch configurations which support fabric connectivity and advanced business continuity capabilities. Infrastructure simplification solutions for IBM xSeries servers include storage consolidation and high-availability server clustering with IBM TotalStorage disk storage arrays. Business continuity solutions include data protection with IBM TotalStorage tape libraries and devices and IBM Tivoli Storage Manager data protection software.

A single SAN32M-2 switch can serve as the cornerstone of a Storage Area Network for those who want to obtain the benefits of storage consolidation and are just beginning to implement Fibre Channel storage systems. Such an entry-level configuration can consist of one or two Fibre Channel links to a disk storage array or to an LTO tape drive. An entry-level sixteen-port storage consolidation solution can support up to fifteen servers with a single path to either disk or tape. The FlexPort feature is designed to enable a base switch to grow to sixteen ports, in eight port increments, to support more servers and more storage devices without taking the switch offline.

A high-availability solution can be created with redundant switches. This capability is ideal for server clustering environments. Such a configuration can support from fourteen to thirty servers, each with dual Fibre Channel adapters cross-connected to redundant SAN32M-2 switches which are cross-connected to a dual-controller storage system. While the SAN32M-2 can be the foundation of medium-sized SANs, it can be configured to participate as a full member in a tier enterprise SAN with other members of the IBM TotalStorage SAN m-type family. This capability provides investment protection as SAN requirements evolve and grow over time. Features include:

- 4 Gb Long Wave Transceivers supporting either 4 or 10 km distances.

9.3.4 Cisco MDS 9120 and 9140 Multilayer Fabric Switch

The Cisco MDS 9120 and 9140 Multilayer Fabric Switches (2061-020 and 2061-040) provide 20 and 40 ports, with 2 Gbps performance, space saving design, and improved reliability capabilities. The host-optimized and target-optimized Fibre Channel ports help to reduce TCO. Intelligent networking services such as Virtual SAN (VSAN) and comprehensive security help simplify entry and midrange SAN management and integration into large core-to-edge SANs.

Shown in Figure 9-7 is the MDS 9120 switch.

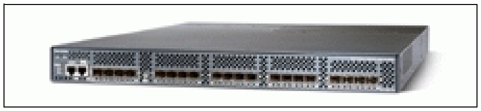


Figure 9-7 9120

Shown in Figure 9-8 is the MDS 9140 switch.

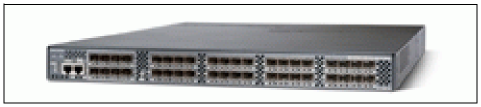


Figure 9-8 9140

The main features are:

- Entry and midrange Fibre Channel SAN solutions.
- Up to 2 Gbps per port throughput and Port Channel support high-performance core-edge SAN deployments.
- Simplified SAN management with host-optimized and target-optimized ports can help to reduce total cost of ownership.
- MDS 9000 inter-family compatibility supports scalability and consistent service as the SAN grows.
- Compact 20 and 40 port design with high-availability capabilities.

- Built-in intelligent network services can help simplify SAN management and reduce total cost of ownership.
- Comprehensive security features support SAN consolidation.
- Virtual SAN (VSAN) capability is designed to create virtual SAN islands on a single physical fabric.
- Offers interoperability with a broad range of IBM servers as well as disk and tape storage devices.

9.3.5 Cisco MDS 9216i and 9216A Multilayer Fabric Switch

The Cisco MDS 9216i Multilayer Fabric Switch is designed to support business continuance solutions in a cost-effective manner. It offers a multiprotocol capable integrated Fibre Channel and IP Storage Services architecture including fourteen 2-Gbps Fibre Channel interfaces for high-performance SAN connectivity, and two 1-Gbps Ethernet ports enabling Fibre Channel over IP (FCIP) and iSCSI storage services. The 9216i base model includes iSCSI capability to extend the benefits of Fibre Channel SAN-based storage to Ethernet-attached servers at a lower cost than Fibre Channel interconnect alone. Additionally, the MDS 9216i offers optional FCIP activation for remote SAN extension. This capability can help simplify data protection and business continuance strategies by enabling backup, remote replication, and other disaster recovery services over wide area network (WAN) distances using open-standard FCIP tunneling. The 9216i is shown in Figure 9-9 on page 190.

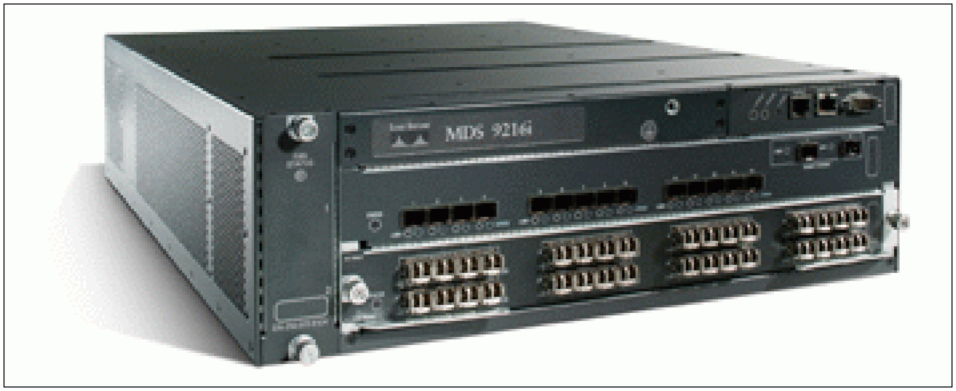


Figure 9-9 9216i

The Cisco MDS 9216A Multilayer Fabric Switch is designed for building mission-critical enterprise SANs where scalability, multilayer capability, resiliency, robust security, and ease of management are required. The 9216A base model includes an updated internal backplane, designed with built-in scalability and growth options to accommodate next-generation Cisco line cards.

The internal backplane within the 9216A base model provides support for next-generation advanced functionality.

Shown in Figure 9-10 is the MDS 9216A switch.

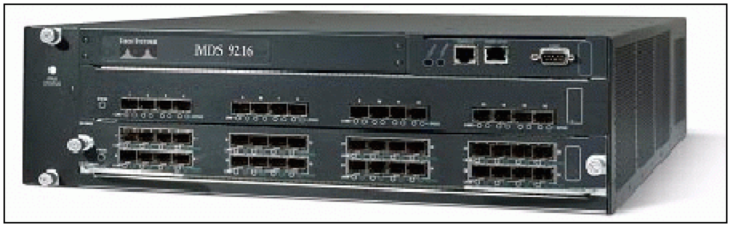


Figure 9-10 9216A

Both models highlight an advanced modular design, to provide built-in scalability and support future growth by featuring an enhanced internal backplane design, and accommodate expansion with the full line of optional switching modules and IP multiprotocol switching modules. The 9216i and 9216A models are designed to provide 1/2 Gbps Fibre Channel compatibility and performance with advanced intelligence to address security, performance, and manageability requirements to consolidate geographically dispersed SAN islands into a large SAN enterprise.

The main features are:

- Integrated IP and Fibre Channel SAN solutions
- Simplified large storage network management and improved SAN fabric utilization can help reduce total cost of ownership
- Provides throughput of up to 2 Gbps per port and up to 32 Gbps with each PortChannel ISL connection
- Offers 4-port 10 Gbps Fibre Channel Switching Module
- Offers 12, 24 and 48-port 1/2/4 Gbps Fibre Channel Switching Modules
- Offers scalability
- Offers Gigabit Ethernet ports for iSCSI or FCIP connectivity
- Features modular design with excellent availability capabilities
- Uses intelligent network services to help simplify storage area network (SAN) management and reduce total cost
- Helps provide security for large enterprise SANs
- Includes Virtual SAN (VSAN) capability for SAN consolidation into virtual SAN islands on a single physical fabric
- Offers compatibility with a broad range of IBM servers as well as disk and tape storage devices

9.3.6 Cisco MDS 9020 Fabric Switch

The Cisco MDS 9020 Multilayer SAN Switch is the next generation midrange SMB switch designed to address the needs of small and medium-sized businesses with a wide range of SAN capabilities. The Cisco MDS 9020 Multilayer SAN Switch offers 1, 2 and 4 Gbps Fibre Channel switch connectivity and intelligent network services to help improve the security, performance, and manageability required to consolidate geographically dispersed storage devices.

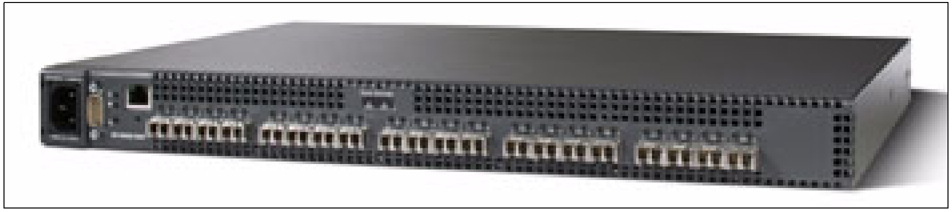


Figure 9-11 Cisco MDS 9020 Fabric Switch

The Cisco MDS 9020 Fabric Switch, as shown in Figure 9-11, can be used as part of SAN solutions from simple single-switch configurations to larger multi-switch configurations in support of fabric connectivity and advanced business continuity capabilities. Fabric connectivity capabilities can be the basis for infrastructure simplification solutions for IBM e-server, iSeries and pSeries servers and storage consolidation and high-availability server clustering with IBM System Storage disk storage arrays. Business continuity capabilities can help businesses protect valuable data with IBM System Storage tape libraries and devices and IBM Tivoli Storage Manager data protection software.

A single MDS 9020 switch can serve as an initial building block for a Storage Area Network for those who want to obtain the benefits of storage consolidation and are just beginning to implement Fibre Channel storage systems. An entry-level configuration, for example, could consist of one or two Fibre Channel links to a disk storage array, or to an LTO tape drive. An entry-level, eight-port storage consolidation solution could support up to seven servers with a single path to either disk or tape. The shortwave optical transceiver feature is designed to enable a base switch to grow to 20 ports, in single port increments, to support more servers and more storage devices without taking the switch offline.

Higher availability solutions can be created using multiple MDS 9020 switches. Such implementations would be well-suited to server clustering environments. Such a configuration could support from six to 18 servers, each with dual Fibre Channel adapters cross-connected to redundant 9020 switches which are cross-connected to a dual-controller storage system.

While the focus of the MDS 9020 switch is as the foundation of medium-sized SMB SANs, it can also be configured to participate as a component of a tiered enterprise SAN with other members of the Cisco MDS 9000 family. This capability helps provide investment protection as SAN requirements evolve and grow over time.

Some of the main features are:

- The Cisco Fabric Manager and Cisco CLI designed to help simplify installation and fabric management for Cisco MDS 9000 customers and to simplify SAN consolidation with Cisco MDS 9000 family SAN fabrics.
- Provides Auto-sensing 4, 2 and 1 Gbps ports designed to help:
  - Enable high performance and improved utilization.
  - Enable rapid switch deployment with minimal configuration.
  - Provide easy installation and manageability.
  - Support new 4 Gbps server and storage devices.
- Provides Hot-swappable SFP transceivers that supports fast and easy replacement of optical transceivers with no tools required.
- Fibre Channel 4 Gb LW 4 km and 10 km SFP transceivers for FC connections at distances up to 4 km (2.5 miles) and 10 km (6.2 miles), respectively.
- High density package Provides 25% more ports than traditional 16 port switches.
- Nondisruptive firmware upgrades designed to help significantly reduce software installation time, enable fast software upgrades, avoid disruption to existing fabric when installing updates, and help reduce administrative overhead.
- Compatibility with Cisco MDS 9000 family, supporting interoperation with Cisco MDS 9000 products.

9.4 Enterprise SAN directors

The IBM Enterprise SAN director class provides:

- Highest availability and scalability, and intelligent software to simplify management of complex, integrated enterprise SANs
- Heterogeneous Windows, Linux, iSeries, UNIX, and mainframe servers
  - xSeries, iSeries, pSeries, and zSeries Server sales channels
  - IBM FAStT, ESS, LTO, and ETS storage
- Supports the IBM TotalStorage Virtualization family and storage systems, IBM Tivoli Storage Manager, SAN Manager Storage Resource Manager, and Multiple Device Manager
- Offers customized solutions with competitive prices, worldwide IBM support, and IBM Global Services and IBM TSSC services

9.4.1 IBM TotalStorage SAN Director M14

The IBM TotalStorage SAN Director M14 (2109-M14) with its next-generation director technology is designed to provide improved performance, enhanced scalability, and has a design ready for future higher performance and expanded capability features. The director is well-suited to address enterprise SAN customer requirements for infrastructure simplification and improved business continuity. It is also designed to be interoperable with other members of the IBM TotalStorage SAN b-type switch family. It can be used to configure a wide range of highly scalable solutions that address today's demands for integrated, heterogeneous mainframe and open server enterprise SANs. This is shown in Figure 9-12 on page 194.

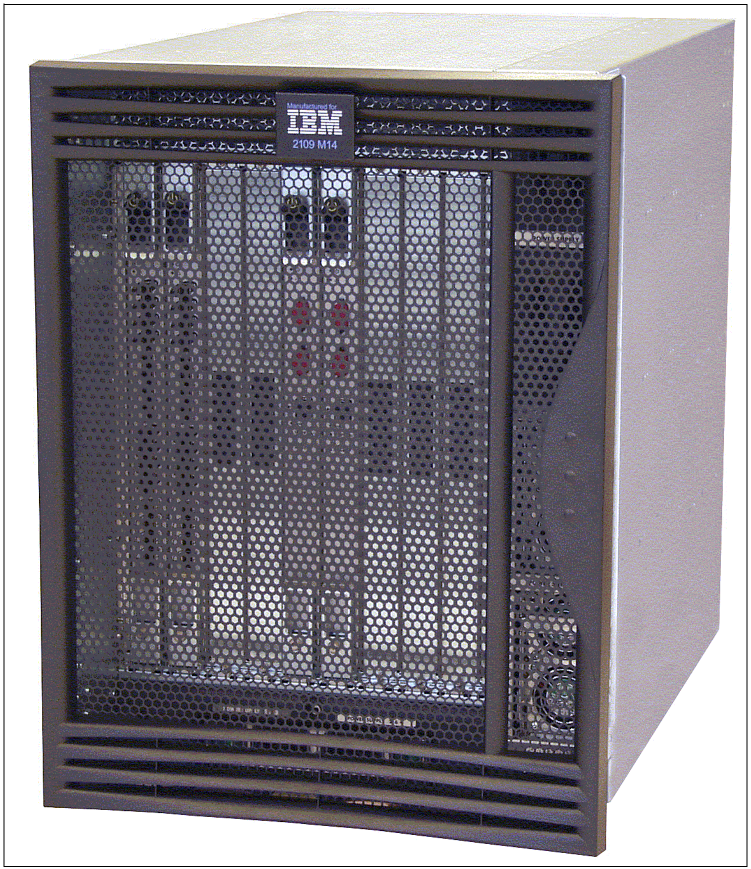


Figure 9-12 M14

The director provides 32–128 ports in a single fabric; 2 Gbps fabric switching for Windows NT/2000 and UNIX; FICON switching for mainframe server clustering;

and provides infrastructure simplification alongside business continuity solutions. The base director includes zoning, WEBTOOLS, Fabric Watch, ISL-Trunking, and Performance Monitoring. Its Fabric Manager feature simplifies complex fabric management. The main features the M14 provides are:

- New 4 Gbps Long Wave SFP Transceivers supporting 4 and 10 km distances.
- High-availability director with built-in redundancy designed to avoid single points of failure.
- Highly scalable director with 32 to 128 ports in a single domain.
- FICON Director switching with Fibre Channel/FICON intermix, FICON CUP (Control Unit Port), and FICON cascading.
- Interoperable with IBM TotalStorage SAN b-type switches.
- Offers advanced fabric services such as end-to-end performance monitoring and fabric-wide health monitoring.
- Remote Switch Activation extends the distance of SAN fabrics by enabling two Fibre Channel switches to interconnect over a Wide Area Network (WAN).
- Fabric Manager helps simplify management, reduce cost of administration, and accelerate deployment and provisioning.

The M14 is designed to provide Fibre Channel connectivity to:

- IBM @server iSeries, zSeries, pSeries, xSeries
- Other Intel processor-based servers with Windows NT, Windows 2000, Windows 2003, NetWare, and Linux
- Selected Sun™ and HP Servers
- IBM TotalStorage Enterprise Storage Server (ESS), IBM TotalStorage DS8000 Series, TotalStorage DS6000 Series, TotalStorage DS4000 Series
- IBM TotalStorage 3590 and 3592 Tape Drives and IBM TotalStorage 3494 Tape Library
- IBM TotalStorage 3582 and 3583 Ultrium Tape Libraries and IBM TotalStorage 3584 UltraScalable Tape Library
- IBM TotalStorage SAN Switches

Inter-switch link trunking

As a standard feature this enables as many as four Fibre Channel links between the M14 and M12 directors, b-type switches, to be combined to form a single logical ISL with an aggregate speed of up to 8 Gbps. ISL trunking provides additional scalability by enabling M14 and M12 directors to be networked in an expandable core, in a core-to-edge SAN fabric.

Performance monitoring

This is another standard feature that provides support for frame filtering-based performance monitoring tools for enhanced end-to-end performance monitoring.

As core-to-edge SAN fabrics scale up to thousands of devices, ISL trunking and frame filtering can help simplify storage management and reduce the overall cost of the storage infrastructure.

FICON Director operation

FICON Director switching includes FICON servers, intermixed FICON and Open servers and FICON cascading between two directors. FICON CUP Activation provides a Control Unit port (CUP) in-band management function designed to allow mainframe applications to perform configurations, monitoring, management, and statistics collection. These applications include System Automation for OS/390 (SA/390), Dynamic Channel Management Facility (DCM) and Resource Measurement Facility (RMF™), Enhanced Call Home, and RAS capabilities, all of which can help simplify management. Hardware enforced FICON and FCP port zoning enhances separation with intermix operations. ISL trunking, with self-optimizing traffic management, can enhance the performance and availability of FICON cascading. Advanced Security Activation is required for FICON cascading. The FICON CUP and Security Activation Bundle provides an affordable bundle of these features.

Advanced Security

This feature can help create a secure storage networking infrastructure required to control and manage fabric access. External threats and internal operational events can compromise valuable enterprise data assets and create data integrity exposures.

Advanced Security Activation

This feature can help create a secure storage networking infrastructure required for multiple protocol operation and SAN island consolidation. Advanced Security extends basic fabric security provided by Advanced Zoning hardware-enforced WWN zoning. It provides a policy-based security system for IBM SAN Switch fabrics with Fabric OS Versions 3 and 4.

WEBTOOLS

This is designed to provide a comprehensive set of management tools that support a Web browser interface for flexible, easy-to-use integration into existing enterprise storage management structures. WEBTOOLS supports security and data integrity by limiting (zoning) host system attachment to specific storage systems and devices.

9.4.2 IBM TotalStorage SAN140M

The IBM TotalStorage SAN140M (2027-140) is a 2 Gbps FICON and Fibre Channel director with scalability from 16 to 140 ports for the highest availability, enterprise infrastructure simplification, and business continuity solutions. The SAN24M-1 and SAN32M-1 can be used as edge switches for highly scalable m-type enterprise-to-edge SAN solutions.

The IBM TotalStorage SAN140M is shown in Figure 9-13 on page 197.

.

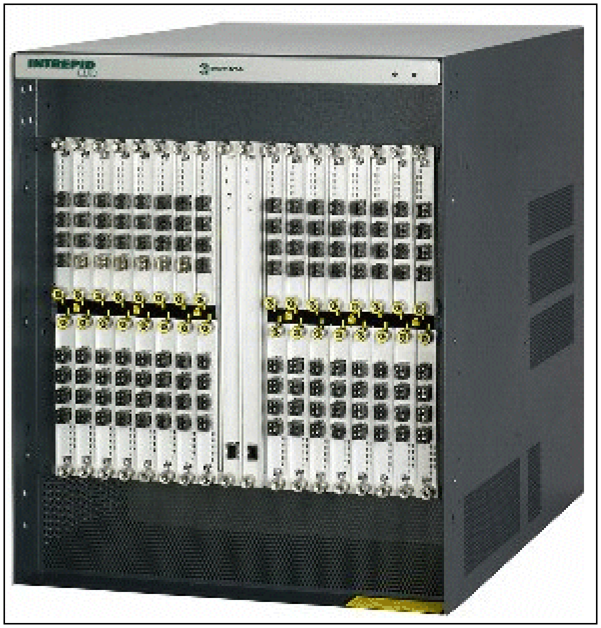


Figure 9-13 SAN140M

The main features are:

- 4 Gbps QPM Port Modules, a cost-effective bandwidth per port that can be used between McDATA Intrepid 6140 Directors, IBM TotalStorage SAN140M directors, IBM TotalStorage SAN32M-2 and SAN16M-2 switches, or any other supported 4 Gbps standards-compliant device
- Storage Network Services (SNS) firmware packages that are value-based bundles for simple ordering and implementation of McDATA's advanced firmware functions that manage Open Systems, Mainframe and Mainframe Cascading
- FICON CUP zoning
- Simple-to-use enterprise infrastructure simplification and business continuity solutions for IBM @server xSeries, iSeries, and pSeries
- Provides highly scalable 16–140 port switching backbone for advanced enterprise infrastructure simplification and business continuity solutions, including mainframe FICON disk and tape storage

- Designed to support highest-availability with redundancy of all active components including hot-swappable processors, fans, and power supplies, HotCAT online code activation and call-home with EFCM software
- Designed for high performance with 1 and 2 Gbps throughput on all ports
- Enterprise Fabric Connectivity Manager, FICON Management Server (CUP) and Open Systems Management Server software help simplify management of complex SAN infrastructures
- Easy-to-manage enterprise infrastructure simplification and business continuity solutions for IBM @server xSeries, iSeries, pSeries and zSeries servers
- Supports 4 Gbps longwave 4/10 km SFP transceivers for the IBM TotalStorage SAN140M switch
- 8 Port FlexPort Expansion kit

9.4.3 IBM TotalStorage SANC40M

The IBM TotalStorage SANC40M (2027-C40) replaces the McDATA Fabricenter FC-512) and provides a space-saving cabinet for IBM TotalStorage SAN m-type directors and switches. It features redundant power distribution for high-availability directors, space for 1U rack mount server and 39U for directors and switches, and its cabling flexibility supports up to 512 Fibre Channel ports.

Its design provides the required airflow and power for high-availability operation. The cabinet comes complete with 28 individual power connections or 14 power connections with dual independent power distribution and dual line cords.

The cabinet supports up to two IBM TotalStorage SAN256M directors; three IBM TotalStorage SAN140M directors; or a combination of up to 14 high-availability, dual power-connected IBM TotalStorage SAN m-type switches and directors. Its main features are:

- Space-saving cabinet for IBM TotalStorage SAN m-type directors and switches
- Dual power distribution system designed for high availability
- Space for 1U rack mount management server and 39U for directors and switches
- 1U Management Server with Microsoft Windows and McDATA Enterprise Fabric Connectivity Manager (EFCM) software features. The 1U Management Server offers as the latest recommended hardware platform for McDATA's storage network management software, EFCM. Microsoft Windows Server® 2003 Standard Edition, as well as the latest version of EFCM, are shipped pre-installed on the 1U Server

- Future-ready design
- Flexible configuration options

Flexible configuration options

The Enterprise Fabric Connectivity Manager (EFCM) software is designed to provide an enterprise-to-edge view of the entire SAN, allowing IT administrators to monitor and control all switched enterprise components from a single console.

A 1U rack mount management server with the EFCM software helps centralize management of multiple directors in the fabric and monitors their operations. The server provides two Ethernet LAN connections—one for a private LAN that supports communication with the directors and switches, and the second for an optional connection to a corporate intranet for remote workstation access. The server supports continuous director monitoring, logging, and alerting; centralizes log files with the EFCM software, configuration databases and firmware distribution. It supports centralized “call-home”, e-mail, service, and support operations. As many as 48 directors and switches can be managed from a single server, and up to eight concurrent users can access the server.

A 24-port Ethernet hub (included and located in the IBM TotalStorage SANC40M cabinet) supports the two connections required by the high-availability function in the IBM TotalStorage SAN256M and IBM TotalStorage SAN140M directors. This hub also supports multiple directors and switches connected to a private LAN. One LAN connection is required for each control processor card in the directors.

9.4.4 IBM TotalStorage SAN256M

The IBM TotalStorage SAN256M high-availability enterprise director (2027-256) provides 10 Gbps backbone connections, 64-256 ports 2 Gbps fabric switching for Windows NT/2000 and UNIX; and FICON switching for mainframe server clustering, infrastructure simplification and business continuity solutions. The Enterprise Fabric Connectivity Manager provides integrated management of complex IBM SAN m-type tiered enterprise fabrics. The SAN256M is shown in Figure 9-14 on page 200.

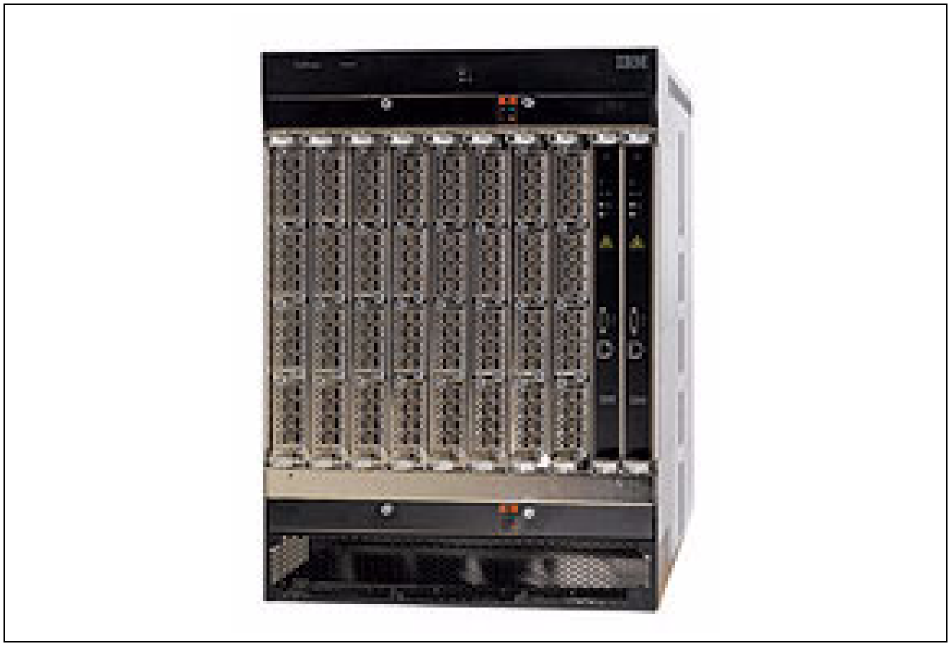


Figure 9-14 SAN256M

A wide range of IBM TotalStorage enterprise storage area network (SAN) infrastructure simplification and business continuity solutions can be created with the IBM TotalStorage SAN 256M enterprise director.

These business continuity solutions include data protection with IBM TotalStorage Ultrium 2 Linear Tape-Open (LTO) and IBM TotalStorage 3592 Tape Drives and tape libraries with IBM Tivoli Storage manager data protection software. The standard director features and capabilities may be combined with IBM TotalStorage disk, tape, and software products to create metro mirroring and global mirroring solutions designed to be disaster-tolerant. Enterprise-to-edge SAN management features help simplify management of large enterprise solutions. The main features are:

- Easy-to-manage tiered enterprise infrastructure simplification and business continuity solutions for IBM @server xSeries, iSeries, pSeries and zSeries.
- Highly scalable 64–256 port switching backbone for tiered global enterprise storage area networks (SANs).
- Designed to provide high availability with concurrent hardware and firmware upgrades and call-home with McDATA Enterprise Fabric Connectivity Manager, EFCM.

- Director FlexPar, designed to provide dynamic application network provisioning, can help simplify Fibre Channel and mainframe FICON SAN consolidation.
- Helps to provide global business continuity solutions with 10 Gbps links up to 190 km.
- EFCM and FICON Management Server (CUP) software can help simplify management of complex SAN infrastructures.
- Infrastructure simplifications solutions.
- Business continuity solutions.
- High-availability design for tiered SAN director backbone solutions.
- EFCM designed to help simplify management of a tiered enterprise SAN infrastructure.
- FICON Management Server and FICON Cascading.
- SANtegrity Binding (standard feature) is required to enable cascading between two directors for mainframe FICON protocol applications.

9.4.5 IBM TotalStorage SAN256B

The IBM TotalStorage SAN256B with next-generation director technology is designed to provide outstanding performance, enhanced scalability and a design ready for future high performance 4 Gbps capable hardware and expanded capability features. The SAN256B is well suited to address enterprise SAN customer requirements for infrastructure simplification and improved business continuity.

The SAN256B director can interoperate with other members of the IBM TotalStorage SAN b-type family. It can be configured with a wide range of highly scalable solutions that address demands for integrated zSeries and open system server enterprise SANs. The SAN256B is shown in Figure 9-15 on page 202.

Figure 9-15 IBM TotalStorage SAN256B

The IBM TotalStorage SAN256B is a high availability director with built-in redundancy designed to avoid single points of failure. It's a highly scalable director with 16 or 32 ports per port blade, and from 16 to 256 ports in a single domain. Other features include:

- New Fibre Channel Routing Blade supporting FCIP and FCR
- New 4 Gb Long Wave SFP Transceivers supporting 4 and 10 km distances.
- New Quad Rate SFP Transceivers supporting 1 Gb Ethernet and Fibre Channel speeds up to 4 Gb.
- Additional Power Supplies to support Fibre Channel Routing and improve availability.
- FICON Director switching with Fibre Channel/FICON intermix, FICON CUP (Control Unit Port) and FICON cascading.
- Interoperable with other IBM TotalStorage SAN b-type switches and directors.
- Offers advanced security with comprehensive policy-based security capabilities.
- Offers advanced fabric services such as end-to-end performance monitoring and fabric-wide health monitoring.

- Designed with redundant control processors, power supplies and fans to avoid single points of failure. This can help improve system availability.
- Supports nondisruptive control processor failover and concurrent firmware activation, helping to improve availability through planned and unplanned maintenance.
- Auto-sensing 4, 2 and 1 Gbps ports with high link speeds offering extraordinary performance.
- Auto-negotiation offers easy interoperability with a wide range of servers, storage systems, switches and directors.
- Scalable from 16 to 256 ports in a single domain that allows you to manage fewer elements in large scale SANs.
  - Scalable in 16 or 32-port increments.
  - Install Longwave or shortwave transceivers on a port-by-port basis
- Provides advanced Inter-Switch Link (ISL) Trunking that allows you to combine as many as eight links to from a single logical ISL, capable of forming aggregate bandwidth of 32 Gbps.
- Frame Filtering enables end-to-end performance analysis, allowing specific devices or ports to be monitored.
- Supports the interconnection of multiple IBM SAN switches and directors.
- Implement fault tolerant storage solutions.

9.4.6 Cisco MDS 9506 Multilayer Director

The Cisco MDS 9506 Multilayer Director (2062-D04) supports 1, 2 and 4 Gbps Fibre Channel switch connectivity and intelligent network services to help improve the security, performance and manageability required to consolidate geographically dispersed storage devices into a large enterprise SAN. Administrators can use the Cisco MDS 9506 to help address the needs for high performance and reliability in SAN environments ranging from small workgroups to very large, integrated global enterprise SANs.

The Cisco MDS 9506 Multilayer Director utilizes two Supervisor-2 Modules designed for high availability and performance. The Supervisor-2 Module combines an intelligent control module and a high-performance crossbar switch fabric in a single unit. It uses Fabric Shortest Path First (FSPF) multipath routing, which provides intelligence to load balance across a maximum of 16 equal-cost paths and to dynamically reroute traffic if a switch fails.

Each Supervisor-2 Module provides the necessary crossbar bandwidth to deliver full system performance in the MDS 9506 director with up to four Fibre Channel switching modules. It is designed to provide that loss or removal of a single crossbar module has no impact on system performance.

Fibre Channel switching modules are designed to optimize performance, flexibility and density. The Cisco MDS 9506 Multilayer Director requires a minimum of one and allows a maximum of four switching modules. These modules are available in either a 12-, 24- and 48-port 4 Gbps configurations, allowing the Cisco MDS 9506 to support 12 to 192 Fibre Channel ports per chassis. Optionally, a 4-port 10 Gbps Fibre Channel module is available for high performance inter-switch link (ISL) connections over metro optical networks.

The Fibre Channel switching modules provide auto-sensing 1 Gbps and 2 Gbps interfaces for high-performance connectivity and compatibility with legacy devices. Switching modules are hot-swappable with small form-factor pluggable (SFP) optic transceivers and support LC interfaces. Individual ports can be configured with either shortwave SFPs for connectivity up to 300 meters at 2 Gbps (500 meters at 1 Gbps) or longwave SFPs for connectivity up to 10 km (at either 1 Gbps or 2 Gbps). Ports can be configured to operate in standard expansion port (E_Port), fabric port (F_Port) and fabric loop port (FL_Port) modes as well as in unique Cisco port modes.

Advanced traffic management capabilities are integrated into the switching modules to help simplify deployment and to optimize performance across a large fabric. The PortChannel capability allows users to aggregate up to 16 physical 2 Gbps Inter-Switch Links into a single logical bundle, providing optimized bandwidth utilization across all links. The bundle may span any port from any 16-port switching module within the chassis, providing up to 32 Gbps throughput.

Highlights include:

- Provides Fibre Channel throughput of up to 4 gigabits per second, Gbps per port and up to 64 Gbps with each PortChannel Inter-Switch Link connection
- Offers scalability from 12 to 192 Fibre Channel ports
- Offers 10 Gbps ISL ports for inter-Data Center links over metro optical networks
- Offers Gigabit Ethernet IP, GbE ports for iSCSI or FCIP connectivity over global networks
- Includes Virtual SAN (VSAN) capability for SAN consolidation into virtual SAN islands on a single physical fabric
- Includes high-availability design with nondisruptive firmware upgrades
- Enterprise, SAN Extension over IP, Mainframe, Storage Services Enabler and Fabric Manager Server Packages provide added intelligence and value

The MDS 9506 Multilayer Director is shown in Figure 9-16.

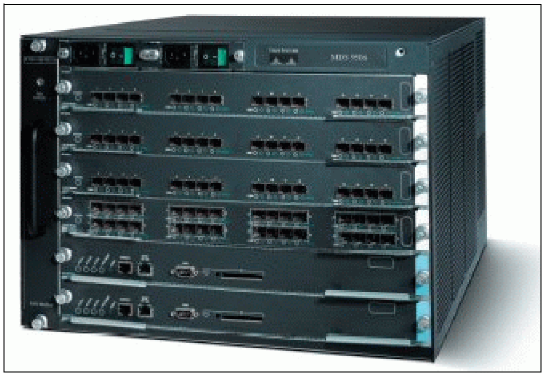


Figure 9-16 9506

9.4.7 Cisco MDS 9509 Multilayer Director

The Cisco MDS 9509 Multilayer Director (2062-D07) provides 1, 2 and 4 Gbps Fibre Channel switch connectivity and intelligent network services to help improve the security, performance and manageability required to consolidate geographically dispersed storage devices into a large enterprise SAN. Administrators can use the Cisco MDS 9509 to help address the needs for high performance and reliability in SAN environments ranging from small workgroups to very large, integrated global enterprise SANs.

The Cisco MDS 9509 Multilayer Director utilizes two Supervisor-2 Modules, designed to support high availability and performance. The Supervisor-2 Module combines an intelligent control module and a high-performance crossbar switch fabric in a single unit. It uses Fabric Shortest Path First (FSPF) multipath routing, which supports load balancing across a maximum of 16 equal-cost paths designed to dynamically reroute traffic if a switch fails.

Each Supervisor-2 Module provides the necessary crossbar bandwidth to deliver full system performance in the MDS 9509 director with up to seven Fibre Channel switching modules. It is designed to provide that loss or removal of a single crossbar module has no impact on system performance.

Fibre Channel switching modules are designed to optimize performance, flexibility and density. The Cisco MDS 9509 Multilayer Director requires a minimum of one and allows a maximum of seven switching modules. These modules are available in 12-, 24- and 48-port 4 Gbps configurations, allowing the Cisco MDS 9509 to support 12 to 336 Fibre Channel ports per chassis. Optionally, a 4-port 10 Gbps Fibre Channel module is available for high performance inter-switch link (ISL) connections over metro optical networks.

Highlights include:

- Supports throughput of up to 4 Gbps per port and up to 64 Gbps with each PortChannel Inter-Switch Link (ISL) connection
- Offers scalability from 12 to 336 4-Gbps Fibre Channel ports
- Offers Gigabit Ethernet IP (GbE) ports for iSCSI or FCIP connectivity over global networks
- High-availability design with support for nondisruptive firmware upgrades
- Includes Virtual SAN (VSAN) capability for SAN consolidation into virtual SAN islands on a single physical fabric
- Enterprise, SAN Extension over IP, Mainframe and Storage Services Enabler and Fabric Manager Server Packages provide added intelligence and value

The MDS 9509 Multilayer Director is shown in Figure 9-17.

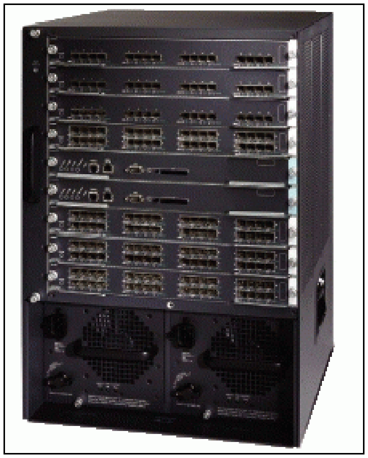


Figure 9-17 9509

9.4.8 Cisco MDS 9513 Multilayer Director

The Cisco MDS 9513 Multilayer Director (IBM 2062-E11) supports 1, 2 and 4 Gbps Fibre Channel switch connectivity and intelligent network services to help improve the security, performance and manageability required to consolidate dispersed SAN islands into a large enterprise SAN. Administrators can use the Cisco MDS 9513 to help address the needs for high performance, scalability and availability in SAN environments ranging from single site environments to very large multiple site metropolitan environments.

The Cisco MDS 9513 Multilayer Director utilizes two Supervisor-2 Modules, designed to support high availability. The Supervisor-2 Module is designed to provide industry leading scalability, intelligent SAN services, non-disruptive software upgrades, stateful process restart and failover, and redundant operation. Dual crossbar switching fabric modules provide a total internal switching bandwidth of 2.4 Tbps for inter-connection of up to eleven Fibre Channel switching modules.

Fibre Channel switching modules are designed to improve performance, flexibility and density. The Cisco MDS 9513 Multilayer Director requires a minimum of one, and allows a maximum of eleven, Fibre Channel switching modules. These modules are available in 12-, 24- or 48-port 4 Gbps configurations, allowing the Cisco MDS 9513 to support 12 to 528 Fibre Channel ports per chassis. Optionally, a 4-port 10 Gbps Fibre Channel module is available for high performance inter-switch link (ISL) connections over metro optical networks.

Highlights include:

- Supports Fibre Channel throughput of up to 4 gigabit per second (Gbps), per port and up to 64 Gbps with each PortChannel Inter-Switch Link (ISL) connection
- Offers Gigabit Ethernet (GbE) IP ports for iSCSI or FCIP connectivity over global networks
- Offers scalability from 12 to 528 4 Gbps Fibre Channel ports
- High-availability design with support for non-disruptive firmware upgrades Includes Virtual SAN (VSAN) capability for SAN consolidation into virtual SAN ‘islands’ on a single physical fabric
- Enterprise, SAN Extension over IP, Mainframe, Storage Services Enabler and Fabric Manager Server Packages provide added intelligence and value

The Cisco MDS 9513 Multilayer Director is shown in Figure 9-18.

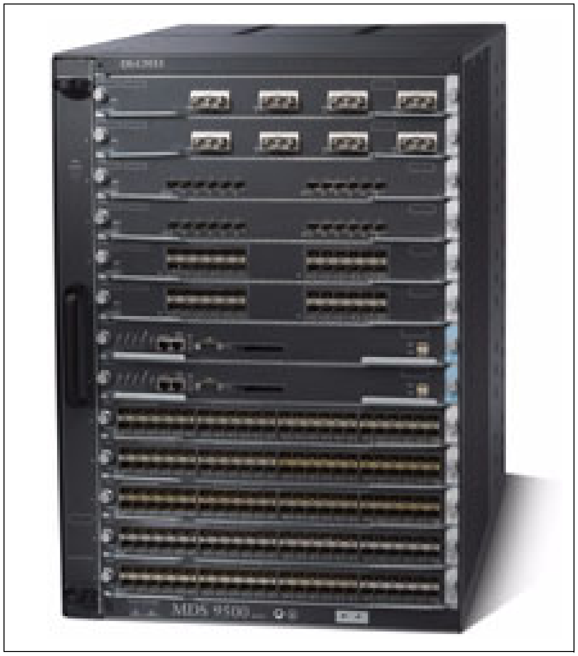


Figure 9-18 Cisco MDS 9513 Multilayer Director

9.5 Multiprotocol routers

IBM currently has three multiprotocol routers in its portfolio.

9.5.1 IBM TotalStorage SAN04M-R multiprotocol SAN router

IBM TotalStorage SAN04M-R (2027-R04) provides two Fibre Channel ports for attachment to existing SANs and two IP ports with Internet Fibre Channel Protocol (iFCP) for high performance metro and global business continuity solutions and Internet SCSI (iSCSI) for cost-effective infrastructure simplification solutions. The router includes four tri-mode shortwave transceivers, iSCSI services and SANvergence Manager. The SAN04M-R is shown on Figure 9-19.

Figure 9-19 IBM TotalStorage SAN04M-R multiprotocol SAN router

Simplify storage infrastructure and protect business continuity with the IBM TotalStorage SAN04M-R multiprotocol SAN router. Cost-effective business continuity solutions over metropolitan and global IP networks include remote IBM TotalStorage tape libraries with IBM Tivoli Storage Manager data protection software and remote mirroring with IBM TotalStorage Resiliency Family. Infrastructure simplification solutions for IBM e-server, xSeries, iSeries and pSeries include iSCSI server integration with IBM TotalStorage.

Midrange and large enterprises with remote site business continuity requirements require cost-effective, simple and secure SAN extension capability. SAN routing provides connectivity between local sites and remote sites over existing IP infrastructures to help enhance data protection and disaster tolerance.

Some of the main features of the TotalStorage SAN04M-R multiprotocol SAN router are:

- Extends the SAN infrastructure over IP networks for cost-effective metro and global business continuity solutions.
- Offers iSCSI server SAN connectivity for cost-effective infrastructure simplification solutions
- Designed for high throughput with 1 Gigabit per second (Gbps) Fibre Channel and Gigabit Ethernet (GbE) with Fast Write and compression to help improve performance over long distances
- Interoperability with IBM TotalStorage SAN m-type (McDATA) family helps provide switch investment protection.
- SAN extension over IP networks helps create cost-effective business continuity solutions.
- Metro and global mirror can utilize existing IP infrastructures.
- Isolation of local and remote SANs improves availability since events in one location do not effect the other location.
- Designed to provide wire rate performance over Gigabit Ethernet connections.

- Fast Write and data compression improves IP network bandwidth utilization help reduce communication cost.
- Intelligent management helps improve performance over extended distances.
- IBM SAN16M-R and IBM SAN04M-R interoperability, extending business continuity solutions to remote sites at lower TCO.
- Flexible configuration options with a common management system.
- iSCSI server SAN connectivity helps create cost-effective infrastructure simplification solutions.
- Utilize existing IP infrastructure to concentration multiple iSCSI servers onto on IP router port.
- Extend benefits of centralized management to edge of the enterprise and beyond.
- SANvergence Manager provides centralized management of multiple sites helps improve TCO and reduce resolution of issues.
- Enhances IBM TotalStorage SAN m-type (McDATA) family which helps provide switch investment protection.
- Offers Pay-as-you grow features like, Entry iSCSI services expandable with multiple SAN extension and management features.
- Redundant power and cooling. Component failures do not impact availability.

9.5.2 IBM TotalStorage SAN 16B-R multiprotocol SAN router

The IBM TotalStorage SAN16B-R (2109-A16) multiprotocol router provides 8 or 16 Fibre Channel and IP ports with Fibre Channel to Fibre Channel, FC-FC Routing Service for SAN island consolidation, Fibre Channel over IP, FCIP Tunneling Service for metro and global business continuity solutions, and an iSCSI Gateway Service for low-cost infrastructure simplification solutions. It provides full interoperability, and integrated management with IBM TotalStorage SAN b-type switches and directors helps protect investments. This is shown in Figure 9-20 on page 210.

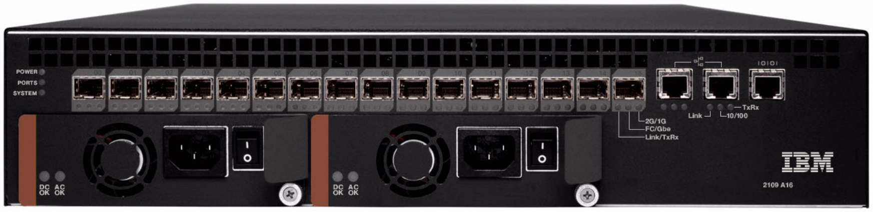


Figure 9-20 SAN16B-R

The IBM TotalStorage SAN16B-R multiprotocol router provides improved scalability, security, and manageability by enabling devices in separate SAN fabrics to communicate without merging fabrics into a single, large SAN fabric. This capability enables customers to initially deploy separate SAN solutions at the departmental and data center levels and then to consolidate them into large enterprise SAN solutions as their experience and requirements grow and change.

- Integrated switch and router management helps simplify deployment and operation of large enterprise SAN infrastructure simplification, business continuity, and information lifecycle management solutions.
- Designed for high performance with 1 and 2 Gbps Fibre Channel and Gigabit Ethernet IP ports.
- Pay-as-you-grow scalability with Ports on Demand feature.
- Designed to support high availability with redundant, hot-swappable fans and power supplies and hot-pluggable optical transceivers.
- Enables SAN island consolidation for infrastructure simplification without compromising security.
- Utilizes existing IP WAN infrastructures for metro and global business continuity solutions.
- Offers iSCSI server SAN connectivity for low-cost infrastructure simplification solutions.
- Integrated router and IBM TotalStorage SAN b-type (Brocade) switch management helps simplify installation and operation and helps provide switch investment protection.

9.5.3 IBM TotalStorage SAN16M-R multiprotocol SAN router

IBM TotalStorage SAN16M-R multiprotocol SAN router (2027-R16) provides 12 Fibre Channel and 4 IP ports with SAN Routing for SAN island consolidation, and Internet Fibre Channel Protocol (iFCP) for high-performance metro and global business continuity solutions and Internet SCSI, iSCSI for low-cost infrastructure simplification solutions. This is shown in Figure 9-21.

Figure 9-21 SAN16M-R

The router includes zoning, two SAN routing ports, iSCSI, and SANvergence Manager. Optional features include SAN routing on 12 Fibre Channel SAN ports, iFCP with Fast Write and compression on four IP ports, and SANvergence Manager Enterprise. The main features are:

- Enables SAN island consolidation for secure data center infrastructure simplification solutions
- Offers iSCSI server SAN connectivity for low-cost infrastructure simplification solutions
- Provides SAN routing over distance for metro and global business continuity solutions
- Designed for high throughput with 1 and 2 Gbps Fibre Channel and Gigabit Ethernet (GbE) with Fast Write and compression
- Interoperability with IBM TotalStorage SAN m-type (McDATA) family provides switch investment protection
- Includes SANvergence Manager for router and network management.

9.5.4 IBM System Storage SAN18B-R multiprotocol router

The IBM System Storage SAN18B-R Router (2005-R18) is designed to deliver up to 4 Gbps Fibre Channel Routing and 1 Gbps Ethernet in an easy to manage compact design. The SAN18B-R shown in Figure 9-22, is based on next-generation 4 Gbps Fibre Channel ASIC technology coupled with hardware-assisted traffic processing to provide industry-leading functionality on both FC routing and FCIP services.

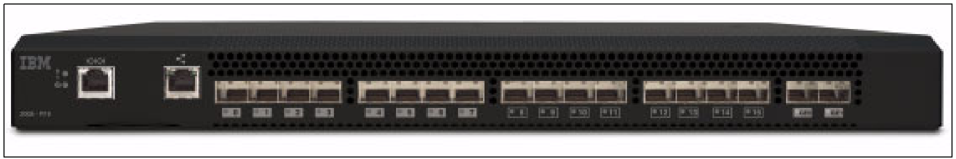


Figure 9-22 IBM System Storage SAN18B-R router

The SAN18B-R Router is designed to seamlessly integrate into existing IBM SAN b-type infrastructures, whether Router or switch-based, the FC routing and FCIP capabilities provide the flexibility for a variety of architectures while extending the reach of the IBM SAN b-type product family.

The SAN18B-R Router provides the industry's first 4 Gbps FC routing capability coupled with hardware-assisted traffic forwarding for FCIP. Each SAN18B-R Router provides 16 4 Gb FC ports and two 1 Gigabit Ethernet ports-offering the high-performance feature set required to run storage applications at line-rate speed, whether the medium is Fibre Channel or Gigabit Ethernet. In the case of running FCIP over high-latency low-speed links, the router offers additional features such as hardware-based compression, up to 8 FC-IP tunnels per GbE port and extensive buffering.

Other unique features include hierarchical routing services, SAN isolation from WAN failures, scalable remote site fan-in, hardware-based encryption, and write acceleration for faster replication performance. The main features include:

- 16 Fibre Channel ports supporting 1, 2 or 4 Gb per second
- 2 Ethernet Ports supporting 1 Gb per second
- FC routing combined with SAN extension
- Optional Fibre Channel over IP Activation
- New 4 Gb Long Wave SFP Transceivers supporting 4 and 10 km
- New Quad Rate SFP Transceivers supporting 1 Gb Ethernet and Fibre Channel speeds up to 4 Gb on the SAN18B-R

Optional features available with the SAN18B-R include:

- A wide range of SFP Transceivers capable of up to 4 Gb Fibre Channel and 1 Gb Ethernet.
- R18 Advanced Security Activation - enables policy-based security mechanisms integrated within Fabric Operating System.
- R18 Performance Bundle Activation is plant order only, and provides both Enhanced Inter-Switch Link (ISL) Trunking and Performance Monitoring capabilities.
- Provides Performance Monitoring capability to help identify end-to-end bandwidth usage.
- Trunking Activation enables Fibre Channel packets to be efficiently distributed across multiple Inter-Switch connections (links) between two SAN b-type fabric switches, while preserving in-order delivery.
- R18 FCIP Activation provides Fibre Channel over Internet Protocol activation.

9.6 IBM TotalStorage DS family

IBM has brought together into one family a broad range of disk systems to help small to large-size enterprises select the right solutions for their needs. The new IBM TotalStorage DS family combines the high-performance heritage of the IBM TotalStorage DS6000 and DS8000 series enterprise servers with the newly enhanced IBM TotalStorage DS4000 series of mid-range systems (formerly called the FAStT family) with newly introduced low-priced entry systems

This family is complemented by a full range of IBM TotalStorage capabilities such as advanced copy services, management tools, and virtualization services to help protect your data.

9.6.1 Entry-level disk systems

The new IBM TotalStorage DS300 and DS400 disk systems are designed to deliver advanced functionality at a breakthrough price. An exceptional solution for workgroup storage applications such as e-mail, file, print, and Web servers, as well as collaborative databases and remote boot for diskless servers.

IBM TotalStorage DS300

The IBM TotalStorage DS300 featuring the iSCSI host connection is an entry-level, cost-effective workgroup SAN storage for IBM @server xSeries and BladeCenter servers.

Designed to deliver advanced functionality at a breakthrough price, the DS300 provides an exceptional solution for workgroup storage applications, such as file, print, and Web serving, as well as remote boot storage for diskless servers. Select configurations of the DS300 are part of the IBM Express portfolio, designed, developed, and priced to meet the specific needs of mid-sized businesses.

Table 9-1 IBM TotalStorage DS300 description

Feature	DS300
Product	DS300
Machine/model	1701-2RD
Platform support	Windows 2000, Windows 2003, Linux, NetWare
Host connectivity	iSCSI
SAN support	Direct, Switched Ethernet
Copy services	FlashCopy, Metro Mirror
Availability features	Fault Tolerant, RAID, Redundant Hotswap Power, Hotswap drives, Dual controller, dual pathing drivers
Controller	Dual Active 1-GB iSCSI RAID Controllers
Cache (min, max)	256 MB, 1 GB (Single), 512 MB, 2 GB (Dual) - Battery Back-up
RAID support	0,1, 5, 10, 50
Capacity (min, max)	36 GB, 2 TB
Drive interface	Ultra320 iSCSI

Feature	DS300
Drive support	36 GB, 73 GB, 146 GB 10 K RPM Disk Drives; 36 GB, 73 GB, 15 K RPM Disk Drives
Certifications	Microsoft Windows MSCS

IBM TotalStorage DS400

The IBM TotalStorage DS400 featuring the 2 Gb Fibre Channel host connection is an entry-level, cost-effective workgroup SAN storage for IBM @server xSeries and BladeCenter servers. Designed to deliver advanced functionality at a breakthrough price, the DS400 provides an exceptional solution for workgroup storage applications, such as e-mail, file, print, and Web servers, as well as collaborative databases. Select configurations of the DS400 are part of the IBM Express portfolio—designed, developed, and priced to meet the specific needs of mid-sized businesses.

Table 9-2 IBM TotalStorage DS400 description

Feature	DS400
Product	DS400
Machine/model	1700-2RD
Platform support	Windows 2000, Windows 2003, Linux, NetWare, VMware
Host connectivity	Fibre Channel
SAN support	Direct, FC-AL, Switched Fabric
Copy services	FlashCopy, Metro Mirror
Availability features	Fault Tolerant, RAID, Redundant Hotswap Power, Hotswap drives, Dual controller, dual pathing drivers
Controller	Dual Active 2 GB FC RAID Controllers
Cache (min, max)	256 MB, 1 GB (Single), 512 MB, 2 GB (Dual) - Battery Back-up
RAID support	0,1, 5, 10, 50
Capacity (min, max)	36 GB, 5.8 TB with 2 EXP400 Expansion Units
Drive interface	Ultra320 SCSI

Feature	DS400
Drive support	36 GB, 73 GB, 146 GB 10 K RPM Disk Drives; 36 GB, 73 GB, 15 K RPM Disk Drives
Certifications	Microsoft Windows MSCS

IBM TotalStorage DS4100

Most businesses today must retain a growing volume of valuable data at an affordable price. With the DS4100 single controller you get entry-level storage shipping with 750 GB for an attractive price. A single controller supports up to 3.5 TB, and a dual controller supports up to 28 TB with the DS4000 EXP100. This general purpose or near-line application storage must keep up with demand. For small and medium-sized enterprises, predicting storage needs and controlling costs can be especially challenging as business grows.

The IBM TotalStorage DS4100 (formerly FAStT100) is designed to give cost-conscious enterprises an entry-level server that can help address storage consolidation and near-line application storage needs without undue expense, while leaving them room to grow. The single controller model supports up to 3.5 TB, while the dual controller model supports up to 28 TB of Serial ATA (SATA) physical disk storage with DS4000 EXP100—provided by up to 14 internal 250 GB disk drives inside the controller. The DS4100 can provide ample yet scalable storage without the cost of extra expansion units. This disk system is designed to help consolidate direct-attached storage into a centrally managed, shared, or storage area network (SAN) environment. With four Fibre Channel ports to attach to servers on a dual controller, the need for additional switches is reduced or eliminated for potential cost savings.

The DS4100 is designed to interoperate with IBM @server pSeries and xSeries servers as well as with Intel processor-based and UNIX-based servers. To help make integration easy, IBM tests the DS4100 for interoperability with IBM servers, as well as with many other brands of servers and switches.

The DS4100 also includes management tools and automation features that can help lower administration costs. At the same time, the SATA-based DS4100 supports increased reliability and performance compared to older, non-redundant parallel Advanced Technology Attachment (ATA) products.

9.6.2 IBM TotalStorage DR550 Express

The IBM TotalStorage DR550 Express is an entry level integrated offering for clients that need to retain and preserve electronic business records. It has many of the features and benefits of the DR550, but at a much lower cost making it ideal for small to medium businesses. The most significant difference between the DR550 and the DR550 Express is the lack of external disk storage. The DR550 Express uses only internal disks of the p5 520 (eigth146 GB UltraSCSI 3 disk drives) and comes in two versions: disk only and tape ready.

Integrating IBM @server pSeries POWER5™ processor-based servers and IBM Tivoli Storage Manager for Data Retention software, this offering is designed to provide a central point of control to help manage growing compliance and data retention needs. The powerful system, which is designed to be mounted into a standard 19 inch rack (you may want to consider a lockable rack for added security - the cabinet is not included with DR550 Express, but can be purchased separately if needed), supports the ability to retain data and prevents tampering or alteration. The system's compact design can help with fast and easy deployment, and incorporates an open and flexible architecture.

To help clients better respond to changing business environments as they transform their infrastructure, the DR550 Express is shipped with approximately 1 terabyte of physical capacity and can be expanded (add an additional 1 TB), if required.

9.6.3 IBM TotalStorage EXP24

The IBM TotalStorage EXP24 is a scalable near-line storage for small to midrange computing environments.

The IBM TotalStorage EXP24 is designed to give cost-conscious businesses an entry-level server that can help meet storage and near-line application storage needs without undo expense, while leaving them room to grow. The EXP24 is scalable up to 7.2TB of Ultra™ 320 SCSI physical disk storage with 24 internal 300 GB disks. The EXP24 is designed to provide ample yet scalable storage without the cost of extra expansion units.

This storage server also is designed to help consolidate storage into a centrally managed environment. The EXP24 is designed to interoperate with IBM @server pSeries servers. Highlights include:

- Well-suited for small to medium-sized enterprises that need affordable, entry-level storage
- Supports storage expansion of up to 24 Ultra 320 SCSI disk drives
- Expandable to over 7 Terabytes of storage capacity
- Offers high-availability features, including redundant power supplies, cooling and hot swap DASD
- Provides up to four SCSI initiator host ports designed to economically attachment of up to 8 servers to a single EXP24 or for failover capabilities
- Configurable as either 4 groups of 6 drives or 2 groups of 12 drives with either single or dual connection any group of drives.

9.6.4 Mid-range disk systems

IBM offers a full range of disk products within a single TotalStorage family to help small to large-size enterprises select the right solutions for their needs. This family includes the IBM TotalStorage Enterprise Storage Server (ESS), the FAStT products (now called the IBM TotalStorage DS4000 series), and new low-priced entry-level products called the IBM TotalStorage DS300 and DS400.

The DS4000 series (formerly FAStT) has been enhanced to complement the entry and enterprise disk system offerings with the DS4000 Storage Manager V9.10, enhanced remote mirror option, and the DS4100 option for larger capacity configurations, along with support for EXP100 serial ATA expansion units attached to DS4400s.

IBM TotalStorage DS4000

The IBM TotalStorage DS4000 EXP710 Fibre Channel Storage Expansion Unit, available for selected DS4000 Midrange Disk Systems, expands the IBM TotalStorage DS4000 family by offering a new 14-bay, 2 Gbps, rack-mountable Fibre Channel (FC) drive enclosure. Enhancements include support for improved reliability and efficiency utilizing internal switch technology to attach to each disk drive within the EXP710 enclosure.

IBM TotalStorage DS4100

Most businesses today must retain a growing volume of valuable data at an affordable price. With the DS4100 single controller you get entry-level storage shipping with 750 GB for an attractive price. The single controller supports up to 3.5 TB, and the dual controller supports up to 28 TB with the DS4000 EXP100. This general purpose or near-line application storage must keep up with demand. For small and medium-sized enterprises, predicting storage needs and controlling costs can be especially challenging as business grows.

The IBM TotalStorage DS4100 (formerly FAStT100) is designed to give cost-conscious enterprises an entry-level server that can help address storage consolidation and near-line application storage needs without undue expense, while leaving them room to grow. The single controller model supports up to 3.5 TB, while the dual controller model supports up to 28 TB of Serial ATA (SATA) physical disk storage with DS4000 EXP100—provided by up to 14 internal 250 GB disk drives inside the controller. The DS4100 can provide ample yet scalable storage without the cost of extra expansion units. This disk system is designed to help consolidate direct-attached storage into a centrally managed, shared, or storage area network (SAN) environment. With four Fibre Channel ports to attach to servers on the dual controller, the need for additional switches is reduced or eliminated for additional potential cost savings.

The DS4100 is designed to interoperate with IBM @server pSeries and xSeries servers as well as with Intel processor-based and UNIX-based servers. To help make integration easy, IBM tests the DS4100 for interoperability with IBM servers, as well as with many other brands of servers and switches.

The DS4100 also includes management tools and automation features that can help lower administration costs. At the same time, the SATA-based DS4100 supports increased reliability and performance compared to older, non-redundant parallel Advanced Technology Attachment (ATA) products.

Table 9-3 IBM TotalStorage DS4100 description

Feature	DS4100
Product	DS4100 (formerly FAStT100 Storage Servers
Machine/model	1724-100/1SC
Platform support	Windows Server 2003, Windows 2000 Server and Adv.Server, Novell netWare 5.1 w/SP6, Red hat Enterprise Linux 3.0, SUSE LINUX Enterprise Server 8, VMWare ESX 2.1, AIX 5.1/5.2, HP-UX 11/11i, Solaris™ 8/9
Host connectivity	Fibre Channel
SAN support	Direct, FC-AL, Switched Fabric
Copy services	FlashCopy option
Availability features	Fault-tolerant, RAID, redundant power/cooling, hot-swap drives, dual controllers, concurrent microcode update capability, dual-pathing driver
Controller	Single/dual active 2 GB RAID controllers
Cache (min, max)	256 MB, 512 MB
RAID support	0, 1, 3, 5, 10
Capacity (min, max)	250 GB, dual controller supports 28 TB with seven expansion units, single controller supports 3.5 TB
Drive interface	2 Gb FC-AL
Drive support	250 GB 7,200 rpm SATA disk drives

Feature	DS4100
Certifications	Microsoft Clustering Services, IBM SAN Volume Controller 1.1.1

IBM TotalStorage DS4300

The IBM TotalStorage DS4300 (formerly FAStT600) is a mid-level disk system that can scale to over eight terabytes of Fibre Channel disk using three EXP700s, over sixteen terabytes of Fibre Channel disk with the Turbo feature using seven EXP700s. It uses the latest in storage networking technology to provide an end-to-end 2 Gbps Fibre Channel solution. As part of the IBM DS4000 mid-range disk systems family, the Model 4300 with Turbo uses the same common storage management software and high-performance hardware design, providing clients with enterprise-like capabilities found in high-end models, but at a much lower cost.

The new DS4000 Storage Manager enables up to 256 logical volumes (LUNs) per storage partition, definition of array groups greater than 2 TB, and SCSI-3 Persistent Reservation. FlashCopy with VolumeCopy is a new function for a complete logical volume copy within the DS4000 is available on the DS4300 Turbo.

Coupled with the EXP100, the DS4300 and DS4300 with Turbo feature allows you to configure RAID-protected storage solutions of up to 28 TB to help provide economical and scalable storage for your rapidly growing application needs for limited access, data reference, and near-line storage.

The IBM TotalStorage DS4300 and the IBM TotalStorage DS4000 EXP700 have been issued a Certificate of Conformance by National Technical Systems, Inc., for conformance to Network Equipment-Building System requirements for Level 3 type 2 and 4 equipment (NEBS 3).

Table 9-4 IBM TotalStorage DS4300 description

Feature	DS4300
Product	DS4300 (formerly FAStT600 Storage Servers
Machine/model	1722-60U/6LU
Platform support	pSeries, xSeries, Windows 2000; optional support for AIX, Solaris, HP-UX, NetWare, Linux, VMWare
Host connectivity	Fibre Channel
SAN support	Direct, FC-AL, Switched Fabric

Feature	DS4300
Copy services	Enhanced Remote Mirroring, FlashCopy, VolumeCopy (turbo option only)
Availability features	Fault-tolerant, RAID, redundant power/cooling, hot-swap drives, single/dual controllers, concurrent microcode update capability, dual-pathing driver
Controller	Single/dual active 2 GB RAID controllers; optional turbo feature
Cache (min, max)	256 MB, 256 MB (single), 512 MB, 512 MB (base) 2 GB, 2 GB (turbo option)
RAID support	0, 1, 3, 5, 10
Capacity (min, max)	Base Single: 2 TB Base: 32 GB, 8.2 TB via EXP700 (FC), 250 GB, 28TB via EXP 100 (Serial ATA) Turbo option: 32 GB, 16.4 TB via EXP700/EXP710 (FC), 250 GB, 28 TB via EXP100 (Serial ATA)
Drive interface	2 Gb FC-AL
Drive support	36.4 GB, 73.4 GB and 146.8 GB 10 K disk drives; 18.2 GB, 34.4 GB and 73.4 GB 15 K disk drives
Certifications	Microsoft RAID, Cluster and Data Center, HACMP™, VERITAS Clustering

IBM TotalStorage DS4400

The IBM TotalStorage DS4400 (formerly FAStT700) delivers superior performance with 2 Gbps Fibre Channel technology. The DS4400 is designed to offer investment protection with advanced functions and flexible features, and scales from 36 GB to over 32 TB to support the growing storage requirements created by e-business applications. It also offers advanced replication services to support business continuance, and is an effective disk system for any enterprise seeking performance without borders.

Table 9-5 IBM TotalStorage DS4400 description

Feature	DS4400
Product	DS4400 (formerly FAStT700 Storage Servers

Feature	DS4400
Machine/model	1742-1RU
Platform support	pSeries, select RS/6000® servers, xSeries, select Netfinity® servers, Windows NT, Windows 2000, netWare, Linux, AIX, HP-UX, Solaris, VMWare
Host connectivity	Fibre Channel
SAN support	Direct, FC-AL, Switched Fabric
Copy services	Enhanced Remote Mirroring, FlashCopy, VolumeCopy
Availability features	Fault-tolerant, RAID, redundant power/cooling, hot-swap drives, dual controllers, concurrent microcode update capability, dual-pathing driver
Controller	Dual active 2 GB RAID controllers
Cache (min, max)	2 GB, 2 GB
RAID support	0, 1, 3, 5, 10
Capacity (min, max)	32 GB, 32 TB via EXP700/EXP710 (FC)
Drive interface	2 Gb FC-AL
Drive support	36.4 GB, 73.4 GB and 146.8 GB 10 K disk drives; 18.2 GB, 34.4 GB and 73.4 GB 15 K disk drives
Certifications	Microsoft RAID, Cluster and Data Center, NetWare Cluster, HACMP, VERITAS Clustering

IBM TotalStorage DS4500

The IBM TotalStorage DS4500 (formerly FAStT900) delivers breakthrough disk performance and outstanding reliability for demanding applications in data-intensive computing environments. The DS4500 is designed to offer investment protection with advanced functions and flexible features. Designed for today's on demand business needs, it offers up to 32 TB of Fibre Channel disk storage capacity with the EXP700. It also offers advanced replication services to support business continuance and disaster recovery.

Coupled with the EXP100, it allows you to configure RAID-protected storage solutions of up to 56 TB to help provide economical and scalable storage for your rapidly growing application needs for limited access, data reference, and near-line storage.

Table 9-6 IBM TotalStorage DS4500 description

Feature	DS4500
Product	DS4500 (formerly FAStT900 Storage Servers
Machine/model	1742-90U
Platform support	pSeries, select RS/6000 servers, xSeries, select netfinity servers, select Sun and HP UNIX servers and other Intel processor-based servers, Windows NT, Windows 2000, netWare, VMWare, Linux, AIX, Solaris, HP-UX
Host connectivity	Fibre Channel
SAN support	Direct, FC-AL, Switched Fabric
Copy services	Enhanced Remote Mirroring, FlashCopy, VolumeCopy
Availability features	Fault-tolerant, RAID, redundant power/cooling, hot-swap drives, dual controllers, concurrent microcode update capability, dual-pathing driver
Controller	Dual active 2 GB RAID controllers
Cache (min, max)	2 GB, 2 GB
RAID support	0, 1, 3, 5, 10
Capacity (min, max)	32 GB, 32 TB via EXP700/EXP710 (FC) 250 GB, 56 TB via EXP100 (Serial ATA)
Drive interface	2 Gb FC-AL
Drive support	36.4 GB, 73.4 GB and 146.8 GB 10 K disk drives; 18.2 GB, 34.4 GB and 73.4 GB 15 K disk drives
Certifications	Microsoft RAID, Cluster, NetWare Cluster, HACMP, VERITAS Clustering

9.6.5 IBM System Storage DR550

The IBM System Storage DR550 is designed to help businesses meet these growing challenges of managing and protecting retention managed data and other critical information assets with operational efficiency. Well suited for archiving a broad range of electronic based records, including e-mail, digital images, database applications, instant messages, account records, contracts or insurance claim documents, and other types of storage records, DR550 is designed to provide advanced storage management technology to enable the management and enforcement of data retention policies.

The DR550’s policy-based, archive data retention capabilities are designed to support nonerasable, non-rewritable data storage, and help address the needs of regulated industries and other businesses with long-term data retention and protection requirements.

The DR550 brings together off-the-shelf IBM hardware and software products. The hardware comes pre-mounted in a secure rack; the software is preinstalled and to a large extent preconfigured. This offering is designed to be easy to deploy.

The DR550 is available in seven configurations and is now available with Enhanced Remote Mirroring. There are two single-node and five dual-node configurations that vary depending on installed disk storage capacity. The customer also has the choice between a so-called disk-only version and a disk-and-tape-ready version that provides additional hardware required to attach tape devices. IBM recommends the IBM TotalStorage 3592 Enterprise Tape Drive or the LTO3 generation tape drive in combination with WORM media to extend the DR550 characteristics for non-erasable and non-rewritable data also to the tape storage pool.

The base single-node configuration consists of one IBM @server p5 520 (1-way), one IBM TotalStorage DS4100 Storage Server with 3.5 TB of raw disk capacity, one IBM TotalStorage SAN Switch 2005-H08 that connects the p5 520 servers to the storage, and a convenient Hardware Management Console, with everything preinstalled in an IBM 7014 rack. The second single-node configuration comes with one additional EXP100 expansion enclosure, increasing the total disk storage capacity to 7 TB.

Designed for high availability, the dual-node configurations consist of two IBM @server p5 520s (1-way) configured in an HACMP active/passive cluster, two IBM TotalStorage SAN Switches (2005-H08s), and one or two DS4100 Storage Servers. The dual-node configurations scale from 3.5 TB to 7 TB, 14 TB, 28 TB, and a maximum of 56 TB total disk storage capacity. The 56 TB configuration features a second 7014 rack to accommodate the second DS4100 Storage Server and the additional seven EXP100 expansion units.

All of these configurations, with the exception of the 56 TB dual-node setup, are now available with an Enhanced Remote Mirroring (ERM) option. With this option, the 2005 H08s are replaced by 2005 H16s to allow for the additional ports needed. Also included are the ERM activation keys for the DS4100s. Please note that this option requires two DR550s, preferably at different sites.

The software bundle includes the IBM AIX 5L™ Version 5.2 operating system, HACMP cluster software (dual node only), IBM Tivoli Storage Manager 5.3 for Data Retention, and IBM TotalStorage DS4000 Storage Manager 9.12, customized for additional protection, all running on the p520 server or servers. The only way to store and retrieve data is through the IBM Tivoli Storage Manager Client API. As such, document management applications need to be capable of communicating with this API in order to archive data on the DR550.

9.6.6 Enterprise disk systems

The IBM TotalStorage DS6000 series, along with the DS8000 series, delivers an enterprise storage continuum of systems with shared replication services and common management interfaces.

The DS6000 and DS8000 series systems are designed to help your company simplify its storage infrastructure, support business continuity, and optimize information lifecycle management.

IBM TotalStorage DS6000

The IBM TotalStorage DS6000 series is a very affordable customer setup storage system designed to help reduce the costs and complexities associated with implementing a pooled storage solution.

The modular design of the DS6000 and intuitive GUI configuration and management software can help companies quickly deploy and realize the system's benefits.

Table 9-7 IBM TotalStorage DS6000 description

Feature	DS6800
Product	IBM TotalStorage DS6800
Machine/model	1750/511

Feature	DS6800
Platform support	xSeries, iSeries, AS/400®, pSeries, RS/6000, zSeries, S/390, i5/OS®, OS/400, AIX, Solaris, HP-UX, Windows 2000, Windows Server 2003, Linux for S/390, z/OS, z/VM®, VSE/ESA™, TPF, Linux for iSeries, Linux for pSeries, Linux for Intel systems, OpenVMS, TRU64, NetWare, VMWare, Apple Macintosh OS X, Fujitsu Primepower
Host connectivity	1 Gb and 2 Gb Fibre Channel/FICON
SAN support	Direct, FC-AL, Switched Fabric
Copy services	FlashCopy, Metro Mirror, Global Mirror, Global Copy, as target for z/OS Global Mirror -Interoperable with ESS 800, ESS 750 and DS8000 Series
Availability features	Fault Tolerant, dual redundant and hot-swap RAID Controller CArds, Battery Backup Units, Fibre Channel switch controllers, power supplies, nondisruptive hardware and software code load updates, multi-pathing device driver‘
Controller	Dual active/active
Cache (min, max)	4 GB
RAID support	5, 10
Capacity (min, max)	292 GB, 67.2 TB
Drive interface	2 Gb Fibre Channel
Drive support	73 GB 15 K, 146 GB 10 K, 300 GB 10 K

IBM TotalStorage DS8000

The IBM TotalStorage DS8000 series are high- performance, high-capacity storage systems designed to break through to an entirely new dimension in scalability, resiliency, and overall total value. Incorporating dual-clustered POWER5 servers, new four-port 2 Gb Fibre Channel/ FICON host adapters, up to 256 GB cache, and new Fibre Channel disk drives, the DS8000 series is designed for outstanding performance.

Table 9-8 IBM TotalStorage DS8000 description

	DS8300	**DS8100**
Product	IBM TotalStorage DS8000 Series	IBM TotalStorage DS8000 Series
Machine/model	2107/922/9A2	2107/921
Platform support	xSeries, iSeries, AS/400, pSeries, RS/6000, zSeries, S/390, i5/OS, OS/400, AIX, Solaris, HP-UX, Windows 2000, Windows Server 2003, Linux for S/390, z/OS, z/VM, VSE/ESA, TPF, Linux for iSeries, Linux for pSeries, Linux for Intel systems, OpenVMS, TRU64, NetWare, VMWare, Apple Macintosh OS X, Fujitsu Primepower, SGI IRIX	xSeries, iSeries, AS/400, pSeries, RS/6000, zSeries, S/390, i5/OS, OS/400, AIX, Solaris, HP-UX, Windows 2000, Windows Server 2003, Linux for S/390, z/OS, z/VM, VSE/ESA, TPF, Linux for iSeries, Linux for pSeries, Linux for Intel systems, OpenVMS, TRU64, NetWare, VMWare, Apple Macintosh OS X, Fujitsu Primepower, SGI IRIX
Host connectivity	1 Gb and 2 Gb Fibre Channel/FICON, ESCON, SCSI	1 Gb and 2 Gb Fibre Channel/FICON, ESCON
SAN support	Direct, FC-AL, Switched Fabric	Direct, FC-AL, Switched Fabric
Copy services	FlashCopy, Metro Mirror, Global Mirror, Global Copy, z/OS Global Mirror. Interoperable with ESS 750 and DS6000 Series	FlashCopy, Metro Mirror, Global Mirror, Global Copy, z/OS Global Mirror. Interoperable with ESS 750 and DS6000 Series
Availability features	Fault Tolerant, dual redundant and hot-swap RAID Controller Cards, Battery Backup Units, Fibre Channel switch controllers, power supplies, nondisruptive hardware and software code load updates, multi-pathing device driver	Fault Tolerant, dual redundant and hot-swap RAID Controller Cards, Battery Backup Units, Fibre Channel switch controllers, power supplies, nondisruptive hardware and software code load updates, multi-pathing device driver‘
Controller	Dual active/active	Dual active/active

	DS8300	DS8100
Cache (min, max)	32/356 GB	16/128 GB
RAID support	5, 10	5, 10
Capacity (min, max)	1.1 TB, 192 TB	1.1 TB, 115 TB
Drive interface	2 Gb Fibre Channel	2 Gb Fibre Channel
Drive support	73 GB 15 K, 146 GB 10 K, 300 GB 10 K	73 GB 15 K, 146 GB 10 K, 300 GB 10 K

IBM TotalStorage Enterprise Storage Server

To address the unique requirements of the on demand world, the IBM TotalStorage Enterprise Storage Server (ESS) helps set new standards for performance, automation, and integration, as well as for capabilities that support continuous data availability. This storage system also supports many advanced copy functions, which can be critical for increasing data availability by providing important disaster recovery and backup protection.

Highly scalable enterprise storage

The ESS Model 800, with an optional turbo feature, is designed to offer the performance, accessibility, security, and reliability needed to support 24x7 operations of the on demand world. Add the flexibility, ease of management, and price/performance that comes standard with the ESS Model 800, and you have both a world-class product, and a low total cost of ownership (TCO) as well.

Mid-range disk systems

The ESS Model 750, as shown includes many of the functions and all the reliability of the ESS Model 800. The ESS Model 750 is designed to provide outstanding price/performance, scaling from 1.1 TB up to 4.6 TB of physical capacity.

Table 9-9 IBM TotalStorage ESS800 and ESS750 descriptions

Feature	ESS800	ESS 750
Product	Enterprise Storage Server	Enterprise Storage Server
Machine/model	2105/750	2105/750

Feature	ESS800	ESS 750
Platform support	xSeries, iSeries, AS/400, pSeries, RS/6000, zSeries, S/390, i5/OS, OS/400, AIX, Solaris, HP-UX, Dynix, OpenVMS, Tru64, Windows NT, Windows 2000, Windows Server 2003, NetWare, VMWare, Linux for S/390, z/OS, z/VM, OS/390, VM/ESA®, VSE/ESA, TPF, Linux for Intel systems, Fujitsu Primepower, SGI Origin IRIX	xSeries, iSeries, AS/400, pSeries, RS/6000, zSeries, S/390, i5/OS, OS/400, AIX, Solaris, HP-UX, Dynix, OpenVMS, Tru64, Windows NT, Windows 2000, Windows Server 2003, NetWare, VMWare, Linux for S/390, z/OS, z/VM, OS/390, VM/ESA, VSE/ESA, TPF, Linux for Intel systems, Fujitsu Primepower, SGI Origin IRIX
Host connectivity	1 Gb and 2 Gb Fibre Channel/FICON, ESCON, SCSI	1 Gb and 2 Gb Fibre Channel/FICON, ESCON
SAN support	Direct, FC-AL, Switched Fabric	Direct, FC-AL, Switched Fabric
Copy services	FlashCopy, Metro Mirror, Global Mirror, Global Copy, z/OS Global Mirror. Interoperable with DS8000 and DS6000	FlashCopy, Metro Mirror, Global Mirror, Global Copy, interoperable with DS8000 and DS6000
Availability features	Fault-tolerant, RAID redundant power/cooling, hot-sw2ap drives, dual controllers, concurrent microcode update capability, dual-pathing driver	Fault-tolerant, RAID redundant power/cooling, hot-sw2ap drives, dual controllers, concurrent microcode update capability, dual-pathing driver
Controller	SMB dual active; optional turbo feature	SBM dual active
Cache (min, max)	8 GB, 64 GB	8 GB, 16 GB
RAID support	5, 10	5, 10
Capacity (min, max)	582 GB, 55.9 TB (physical capacity)	1.1 TB, 4, 6 TB
Drive interface	SSA	SSA

Feature	ESS800	ESS 750
Drive support	18.2 GB, 36.4 GB, 72.8 GB and 145.6 GB 10,000 rpm disk drives 18.2 GB, 36.4 GB and 72.8 GB 15,000 rpm disk drives	72.8 GB, 145.6 GB (10,000 rpm)
Certifications	Microsoft RAID, Cluster and Data Center, GDPS®, HACMP, NetWare, Linux	Microsoft RAID, Cluster and Data Center, GDPS, HACMP, NetWare, Linux

Note: For detailed information relating to the IBM TotalStorage DS Series portfolio, visit the following Web site:

`http://www-1.ibm.com/servers/storage/disk/index.html`

9.7 IBM Tape Storage Systems

As information technology budgets shrink, low-cost tape storage has become more attractive as a way to manage ever-increasing corporate data growth. From a single tape drive to libraries capturing petabytes of data, IBM TotalStorage Tape Drives, Tape Libraries, and Virtual Tape Servers offer a range of solutions to meet data management needs and to address growing business continuance concerns and regulatory requirements. IBM Tape autoloaders.

We give an overview tape autoloaders in the topics that follow.

3581 Tape Autoloader

The IBM TotalStorage 3581 Tape Autoloader can provide an excellent solution for businesses looking for high-density and performance tape autoloader storage in constrained rack or desktop environments. The 3581 offers a single Ultrium 2 or 3 tape drive and storage for up to eight tape cartridges in a 2U form factor. Optional features are available and designed to help enable the operation of the autoloader as a small library.

Table 9-10 3581 Tape Autoloader description

Feature	3581
Product	Tape Autoloader

Feature	3581
Machine model	3581 L28 Ultrium 2 F28 Ultrium 2 L23 Ultrium 2 H23 Ultrium 2
Product strengths	Open systems attach, multiplatform, high capacity, optional barcode reader
Technology	Longitudinal Serpentine
Number of heads/tracks	Ultrium 2:8/512
Number of drives	1
Max number of carthidges	8 - L28, F28 7 - L23, H23
Carthidge capacity native/compressed	Ultrium 2: 200/400 GB
Max system capacity compressed	L23, H23:2.8 TB L28, F28:3.2 TB
Time to data	Ultrium 2: 49 seconds
Interface	L28, F28-FC fabric L23, LVD H23, HVD
SAN-ready	pSeries, RS/6000, Windows NT, Sun
Supported platforms	xSeries, netfinity, iSeries, AS/400, pSeries, RS/6000, Windows 2000, Windows Server 2003, Sun HP, Linux
Media part number	Ultrium 2: 08L9870

3582 Tape Library

The IBM TotalStorage 3582 Tape Library can provide an outstanding automation solution for addressing the storage needs of small to medium-sized environments. With up to two tape drives and 24 tape cartridges, the 3582 Tape Library is designed to leverage the LTO technology to cost-effectively handle growing storage requirements.

Table 9-11 3582 Tape Library description

Feature	3582
Product	Tape Library
Machine model	3582 L23 Ultrium 2 F23 Ultrium 2
Product strengths	Open systems attach, multiplatform, scalable, high capacity, direct-to-SAN attach
Technology	Longitudinal Serpentine
Number of heads/tracks	Ultrium 2:8/512
Number of drives	1–2
Max Number of Carthidges	24
Carthidge capacity native/compressed	Ultrium 2: 200/400 GB
Max system capacity compressed	Ultrium 2:9.6 TB
Time to data	Ultrium 2: 49 seconds
Interface	LVD, HVD, FC fabric
SAN-ready	pSeries, RS/6000, Windows NT, Sun
Supported platforms	xSeries, netfinity, iSeries, AS/400, pSeries, RS/6000, Windows 2000, Windows Server 2003, Sun HP, Linux
Media Part number	Ultrium 2: 08L9870

3583 Tape Library

By fully leveraging advanced Linear Tape-Open (LTO) technology, the IBM TotalStorage 3583 Tape Library can provide an outstanding solution for cost-effectively handling a wide range of backup, archive, and disaster recovery data storage needs. The breakthrough reliability, capacity, and performance of LTO offers an excellent alternative to DLT, 8 mm, 4 mm, or 1/4-inch tape drives for streaming data applications such as backup.

Table 9-12 3583 Tape Library description

Feature	**3583**
Product	Tape Library
Machine model	3583 L18-18 Carts L36-36 Cards L72-72 cARDS
Product strengths	Open systems attach, multiplatform, scalable, high capacity, direct-to-SAN attach
Technology	Longitudinal Serpentine
Number of heads/tracks	Ultrium 2:8/512
Number of drives	1–6
Max number of carthidges	72
Carthidge capacity native/compressed	Ultrium 2: 200/400 GB
Max system capacity compressed	Ultrium 2:28.8 TB
Time to data	Ultrium 2: 49 seconds
Interface	LVD, HVD, FC fabric
SAN-ready	pSeries, RS/6000, Windows NT, Sun
Supported platforms	xSeries, netfinity, iSeries, AS/400, pSeries, RS/6000, Windows 2000, Windows Server 2003, Sun HP, Linux
Media Part number	Ultrium 2: 08L9870

7212 Storage Device Enclosure

The IBM TotalStorage 7212 Storage Enclosure is a versatile product that provides efficient and convenient storage expansion for IBM @server iSeries and pSeries servers. The 7212 offers two models designed for either manual operation or automation. The 7212 can mount in 1 EIA unit of a standard 19-inch rack using an optional rack hardware kit, or it can be configured for desktop mounting.

Table 9-13 7212 Storage Device Enclosure description

Feature	7212-102
Product	Storage Device Enclosure
Machine model	7212 102
Product strengths	Rack-mountable 2-drive enclosure utilizes only 1U (1.75”) of space
Technology	DDS, DVD, VXA
Number of heads/tracks	DDS/VXA: Rotating Drum
Number of drives	1–2
Max number of carthidges	2
Carthidge capacity native/compressed	DDS-4: 20/40 GB DAT72: 36/72 GB VXA-2: 80/160 GB
Max system capacity compressed	160 GB with two VXA-2 drives; 18.8 GB with two DVD-RAM drives
Time to data	VXA-2: 40 seconds VXA: 40 SECONDS DAT72: 50 seconds
Interface	SCSI-3 ULTRA LVS/SE, 160/320
Supported platforms	pSeries, RS/6000, iSeries, AS/400
Media Part number	DDS-4: 59H4456 DAT72: 18P7912 VXA-2: 19P4876 SLR60: 19P4209 SLR100: 35L0980

7332 4 mm DDS Tape Cartridge Autoloader

The 7332 family of 4 mm tape cartridge autoloaders provide an externally attached cost-effective tape storage solution consisting of proven 4 mm tape drive technology and a tape autoloader. The table-top unit, which is designed to attach to IBM RS/6000 servers and workstations, includes the same DDS 4 mm tape drive used in 7206, robotics, and a cartridge magazine with slots for 4 mm cartridges. It operates in streaming mode by using the autoloader to sequentially feed cartridges.

Table 9-14 3583 Tape Library description

Feature	3583
Product	Tape Library
Machine model	3583 L18-18 Carts L36-36 Cards L72-72 cARDS
Product strengths	Open systems attach, multiplatform, scalable, high capacity, direct-to-SAN attach
Technology	Longitudinal Serpentine
Number of heads/tracks	Ultrium 2:8/512
Number of drives	1–6
Max number of carthidges	72
Cartridge capacity native/compressed	Ultrium 2: 200/400 GB
Max system capacity compressed	Ultrium 2:28.8 TB
Time to data	Ultrium 2: 49 seconds
Interface	LVD, HVD, FC fabric
SAN-ready	pSeries, RS/6000, Windows NT, Sun
Supported platforms	xSeries, netfinity, iSeries, AS/400, pSeries, RS/6000, Windows 2000, Windows Server 2003, Sun HP, Linux
Media Part number	Ultrium 2: 08L9870

9.7.1 Tape drives

In this section we give an overview of tape drives.

IBM 3592 Tape Drive

The IBM TotalStorage 3592 Tape Drive Model J1A is designed to provide high capacity and performance for storing mission-critical data. By offering significant advancements in capacity and data transfer rates, the 3592 Tape Drive helps address storage requirements that are often filled by two types of drives—those that provide fast access for data access and those that provide high capacity for backups. The 3592 Tape Drive handles both types of use, helping simplify your tape infrastructure. Additionally, the 3592 Tape Drive Model J1A offers Write Once, Read Many (WORM) functionality, which is designed to help support data retention needs and applications requiring an audit trail.

Table 9-15 3592 Tape Drive description

Feature	**3592**
Product	1/2” Tape Drive
Machine model	3592 J1A
Product strengths	Multipurpose drive, capacity, performance, Write Once Read Many (WORM) support
Technology	longitudinal Serpentine
Number of heads/tracks	8/512
Number of drives	1
Cartridge capacity native/compressed	300/900 GB 300/900 GB WORM 60/180 GB 60/180 GB WORM
Max drive data rate native/compressed	40/120 MBps
Time to Data	Cartridge dependent
Interface	FC, ESCON, FICON, 2 Gb FICON
SAN-ready	pSeries, iSeries, RS/6000, Sun, HP, Windows NT, Linux, Windows 2000
Supported platforms	iSeries, AS/400, pSeries, RS/6000, zSeries, S/390, Linux, Windows 2000, Sun, HP
Media part number	18P7534

IBM 3590 Tape Drive

The IBM TotalStorage 3590 Tape Drive provides high levels of performance and reliability and exemplifies IBM's continued leadership in storage products. Since its first shipment in September 1995, it has met with wide marketplace acceptance. Over 100,000 3590 Tape Drives are installed in both IBM and non-IBM systems across industry sectors.

Table 9-16 3590 1/2" Tape Drive description

Feature	**3590**
Product	1/2" Tape Drive
Machine model	3590 J1A
Product strengths	Multipurpose drive, capacity, performance, Write Once Read Many (WORM) support
Technology	longitudinal Serpentine
Number of heads/tracks	8/512
Number of drives	1
Cartridge capacity native/compressed	300/900 GB 300/900 GB WORM 60/180 GB 60/180 GB WORM
Max drive data rate native/compressed	40/120 MBps
Time to data	Cartridge dependent
Interface	FC, ESCON, FICON, 2 Gb FICON
SAN-ready	pSeries, iSeries, RS/6000, Sun, HP, Windows NT, Linux, Windows 2000
Supported platforms	iSeries, AS/400, pSeries, RS/6000, zSeries, S/390, Linux, Windows 2000, Sun, HP
Media part number	18P7534

IBM 3580 Tape Drive

The IBM TotalStorage 3580 model L33 Tape Drive is an external drive incorporating the third and latest generation of IBM LTO technology. This is an external stand-alone or rack-mountable unit, similar to previous models of the 3580, and is the entry point for the family of IBM Ultrium tape products. The 3580 Tape Drive provides an excellent migration path from digital linear tape (DLT or

SDLT), 1/4-in., 4 mm, or 8 mm tape drives. The 3580 model L33 can read and write LTO Ultrium 2 Data Cartridges and read LTO Ultrium 1 Data Cartridges.

Table 9-17 3580 Tape Drive description

Feature	**3580**
Product	Tape Drive
Machine model	3580 L33 Ultrium 3 L23 Ultrium 2 H23 Ultrium 2
Product strengths	Open systems attach, high capacity, fast data transfer rate, desktop footprint
Technology	Longitudinal Serpentine
Number of heads/tracks	Ultrium 3: 16/704 Ultrium 2: 8/512
Number of drives	1
Cartridge capacity native/compressed	Ultrium 3: 400 GB Ultrium 2: 200/400 GB
Max drive data rate native/compressed	Ultrium 3: 80 MBps Ultrium 2: 35/70 MBps
Time to data	Ultrium 3: 49 seconds Ultrium 2: 49 seconds
Interface	L33, L23, LVD, H23, HVD
SAN-ready	pSeries, RS/6000, Windows NT, Sun
Supported platforms	xSeries, Netfinity, iSeries, AS/400, pSeries, RS/6000, Windows 2000, Windows Server 2003, Sun, HP, Linux
Media part number	Ultrium 3: 24R1922 Ultrium 2: 08L9870

IBM 7205 External SDLT Tape Drive Model 550

The IBM 7205 Model 550 delivers fast and dependable tape backup, restore, and archive functions in a pSeries and RS/6000 environment. This stand-alone digital linear tape drive is an excellent alternative to other tape technologies in the industry.

Table 9-18 7205 External SDLT Tape Drive Model 550 description

Feature	7205
Product	SDLT Drive Tape Drive
Machine model	7205 550
Product strengths	Cost-effective save/restore/archive solution
Technology	Longitudinal Serpentine
Number of heads/tracks	8/536
Number of drives	1
Max number of cartridges	1
Cartridge capacity native/compressed	160/320 GB
Max drive data rate native/compressed	16/32 MBps
Time to data	70 seconds
Interface	SCSI-2 F/W, Diff PCI, Ultra2 SCSI LVD
SAN-ready	pSeries, RS/6000
Supported platforms	pSeries, RS/6000
Media part number	35L1119

IBM 7206 External Tape Drive

The 7206 family of tape drives are externally attached streaming tape drives designed for use with IBM RS/6000 workstations and servers.

Table 9-19 7206 External Tape Drive description

Feature	7206	7206-VX2
Product	4 mm DDS-4 4 mm Gen 5 (DAT72)	8 mm VXA-2 Tape Drive
Machine model	7206 220-DDS4 336-DDS Gen 5	7206 VX2

Feature	7206	7206-VX2
Product strengths	Cost-effective streaming tape drive	Low-cost, high-capacity VXA-2 technology
Technology	Helical Scan	Helical Scan
Number of heads/tracks	Rotating Drum	Rotating Drum
Number of drives	1	1
Max number of carthidges	1	1
Cartridge capacity native/compressed	220: 20/40 GB 336: 36/72 GB	20/40 GB 59/118 GB 80/160 GB
Max system capacity compressed	220: 40 GB 336: 72 GB	160 GB
Max drive data rate native/compressed	220: 2/6 MBps 336: 3/6 MBps	6/12 GBps
Time to data	50 seconds	50 seconds
Interface	SCSI-2 F/W SE, LVD/SE	SCSI-3 ULTRA, LVD/SE, 160/320
SAN-ready	pSeries, RS/6000	pSeries, RS/6000, iSeries, AS/400
Supported platforms	pSeries, RS/6000	pSeries, RS/6000, iSeries, AS/400
Media part number	DDS-4: 59H4456 DAT72: 18P7912	19P4876 (230M), 19P4877 (170M) 19P4878 (62M)

IBM 7207 External Tape Drive

The 7207 provides affordable backup, archival storage, and data interchange for iSeries/AS400 and pSeries/RS6000 systems.

Table 9-20 7207 External Tape Drive description

Feature	7207
Product	SLR (QIC) Compatible External Tape Drive

Feature	7207
Machine model	7207 122 330
Product strengths	Backward read/write compatible with iSeries Internal
Technology	SLR (QIC format)
Number of heads/tracks	1/1
Number of drives	1
Max number of carthidges	1
Cartridge capacity native/compressed	122: 4/8 GB 330: 30/60 GB
Max system capacity compressed	122: 4/8 GB 330: 30/60 GB
Max drive data rate native/compressed	122: .38/.76 MBps 330: 4/8 MBps
Time to data	122: 85 seconds 330: 50 seconds
Interface	SCSI-2 SE, ULTRA, LVD/SE, 160/320
SAN-ready	iSeries, AS/400, pSeries, RS/6000
Supported platforms	iSeries, AS/400, pSeries, RS6000
Media Part number	122: 59H3660 330: 19P4209
Interface	SCSI-2 SE, ULTRA, LVD/SE, 160/320

IBM 7208 Model 345 External 8 mm Tape Drive

The IBM 7208 8 mm tape drive provides an excellent solution to users of 8 mm tape who require a larger capacity or higher performance tape backup. The IBM 7208 tape drive is based on enhanced 8 mm Mammoth tape drive technology, and delivers fast and dependable save and restore functions on both pSeries and iSeries servers.

Table 9-21 7208 Model 345 External 8 mm Tape Drive description

Feature	7208
Product	8 mm Mammoth Tape Drive
Machine model	7208 345-Mammoth-2
Product strengths	High-performance 8 mm technology
Technology	Helical Scan
Number of heads/tracks	Rotating Drum
Number of drives	1
Max number of carthidges	1
Cartridge capacity native/compressed	60/150 GB
Max system capacity compressed	150 GB
Max drive data rate native/compressed	12/30 MBps
Time to data	93 seconds
Interface	SCSI-2 Ultra2 LVD
SAN-ready	pSeries, RS/6000, iSeries, AS/400, Windows NT
Supported platforms	pSeries, RS/6000, iSeries, AS/400
Media part number	18P6485-60 GB

IBM Virtualization Engine TS7510

The IBM Virtualization Engine™ TS7510 is the first member of the IBM Virtualization Engine TS7000 Series of virtual tape libraries. The TS7510 combines hardware and software into an integrated solution designed to provide tape virtualization for open systems servers connecting over Fibre Channel physical connections.

The TS7510 combines IBM server technology, disk technology and tape technology, and is designed to virtualize, or emulate tape libraries, tape drives, and tape media. Real tape resources can then be attached to the TS7510 to help address information lifecycle management and business continuance. The TS7510 is designed to help customers achieve the following throughput efficiencies:

- Reduce backup window
- Improve restore process
- Facilitate data sharing
- Low total cost of ownership (TCO)

Table 9-22 TS7510 benefits

TS 7510 Feature	Benefit
Up to 128 virtual libraries per system	Designed to allow each backup server to allocate its own virtual library
Up to 1024 virtual drives per system	Designed to allow for a substantial number of mount points for very high performance
Up to 8,192 virtual cartridges per system	Designed to allow for substantial number of virtual cartridges, allowing for capacity on demand growth.
Up to 32 concurrent backups per system	Designed to allow multiple backup/restore jobs to run simultaneously on a single TS7510 for high performance and infrastructure simplification,
2:1 Compression capability	Provides up to 92TB Cache Physical Capacity at 2:1 Compression,
Support for major operating systems and hardware platforms	Supports your diverse IT infrastructure.
8 FC ports for host/tape attachment per system	High FC port count is designed to offer low ISV costs.
Optional IP remote replication function, with compression and encryption	Designed to allow for remote disaster recovery site copies. Designed to allow for electronic vaulting, with high security features using encryption.
Dual Node server with failover/failback capacity	Designed to provide for automatic server failover and failback for high availability requirements.
Dual core 3.0 GHz processors	High CPU speeds and dual processors address processor intensive functions like compression/encryption with minimal performance impact.

IBM TotalStorage Virtual Tape Server

As an option to the 3494 library, IBM offers a product called the IBM TotalStorage Virtual Tape Server (VTS). Because of the way z/OS and its predecessors organize tape storage, ordinarily only a single volume may be stored on a single tape. In some environments, this can result in the extremely inefficient usage of tape capacity. The VTS option for the 3594 library replaces the control units that would usually be used to communicate with the tape drives. It stores the data received from the host in internal disk storage, and then stores that data to tape in a manner that results in the tape being full, or nearly so. This greatly reduces the number or tape cartridges required for data backups.

The IBM TotalStorage 3494 Virtual Tape Server (VTS) is an enterprise tape solution designed to enhance performance and provide the capacity required for today's backup requirements. The adoption of this solution can help reduce batch processing time, total cost of ownership, and management cost. A VTS Peer-to-Peer (PtP) configuration can provide redundancy for greater disaster tolerance and help fulfill business continuity requirements.

Table 9-23 3494-VTS description

Feature	**3494-VTS**
Product	Virtual Tape Server
Machine model	3494 B10, B20
Product strengths	Virtualtape, fastaccess peer-to-peer copy
Technology	Disk: Emulated 3490E Tape: 3592, 3590 1/2”
Number of drives	62-256 (virtual)
Max number of carthidges	500,000 (virtual volumes)
Cartridge capacity native/compressed	Drive dependent
Max system capacity compressed	Disk: Up to 5.2 TB Tape: 5616 TB
Max drive data rate native/compressed	VTS: Drive dependent
Time to data	VTS: 1–3 seconds if data in VTS Cache
Interface	Ultra SCSI, ESCON, FICON
SAN-ready	pSeries, RS/6000, Windows NT, Windows 2000, Sun

Feature	3494-VTS
Supported platforms	pSeries, RS/6000, zSeries, S/390, Windows 2000, Sun, HP
Media part number	18P7534 05H4434 X: 05H3188

Note: For detailed information relating to the IBM Tape Storage Systems portfolio, go to the following Web site:

`http://www-1.ibm.com/servers/storage/tape/index.html`

9.8 IBM TotalStorage virtualization

The IBM TotalStorage Open Software family combines the power of our storage virtualization software and Tivoli storage management software to help to simplify storage infrastructure, optimize storage utilization, and enables businesses to adapt quickly and dynamically to variable environments.

9.8.1 IBM TotalStorage SAN Volume Controller

SANs enable companies to share homogeneous storage resources across the enterprise. But for many companies, information resources are spread over a variety of locations and storage environments, often with products from different vendors, who supply everything from mainframes to laptops. To achieve higher utilization of resources, companies now need to share their storage resources from all their environments, regardless of the vendor. While storage needs rise rapidly, and companies operate on lean budgets and staffing, the best solution is one that leverages the investment already made and that provides growth when needed. SVC contributes towards this goal of a solution that can help strengthen existing SANs by increasing storage capacity, efficiency, uptime, administrator productivity, and functionality.

The IBM TotalStorage SAN Volume Controller (SVC) is designed to:

- Provide a centralized control point for managing an entire heterogeneous SAN, including storage volumes from disparate vendor devices.
- Help optimize existing IT investments by virtualizing storage and centralizing management.
- Reduce downtime for planned and unplanned outages, maintenance, and backups.

- Increase storage capacity utilization, uptime, administrator productivity, and efficiency.
- Provide a single set of advanced copy and backup services for multiple storage devices.

Centralized point of control

SVC is designed to help IT administrators manage storage volumes from their storage area networks (SANs). It helps combine the capacity of multiple storage controllers, including storage controllers from other vendors, into a single resource, with a single view of the volumes.

Reduction of downtime

SVC is designed to provide IT administrators with the ability to migrate storage from one device to another without taking the storage offline, and allow them to better reallocate, scale, upgrade and back up storage capacity without disrupting applications.

Improved resource utilization

SVC is designed to help increase storage capacity and uptime, as well as administrator productivity and efficiency, while leveraging existing storage investments through virtualization and centralization of management.

A single, cost-effective set of advanced copy services

SVC is designed to support advanced copy services across all attached storage, regardless of the intelligence of the underlying controllers.

SVC architecture

The IBM TotalStorage SAN Volume Controller (SVC) is based on the COMmodity PArts Storage System (COMPASS), Compass architecture developed at the IBM Almaden Research Center. The overall goal of the Compass architecture is to create storage subsystem software applications that require minimal porting effort to leverage a new hardware platform. The approach is to minimize the dependency on unique hardware, and to allow exploitation of, or migration to, new SAN interfaces simply by plugging in new commodity adapters.

9.8.2 IBM TotalStorage SAN File System

IBM TotalStorage SAN File System is a common SAN-wide file system that permits centralization of management and improved storage utilization at the file level. The IBM TotalStorage SAN File System is configured in a high-availability configuration based on xSeries with clustering for the metadata controllers, providing redundancy and fault tolerance.

The IBM TotalStorage SAN File System is designed to provide policy-based storage automation capabilities for provisioning and data placement, nondisruptive data migration, and a single point of management for files on a storage network.

SANs have gained wide acceptance. Interoperability issues between components from different vendors connected by a SAN fabric have received attention and have mostly been resolved, but the problem of managing the data stored on a variety of devices from different vendors is still a major challenge to the industry.

Data sharing in a SAN

The term *data sharing* is used somewhat loosely by users and some vendors. It is sometimes interpreted to mean the replication of files or databases to enable two or more users, or applications, to concurrently use separate copies of the data. The applications concerned may operate on different host platforms.

Data sharing may also be used to describe multiple users accessing a single copy of a file.This could be called “true data sharing”. In a homogeneous server environment, with appropriate application software controls, multiple servers may access a single copy of data stored on a consolidated storage subsystem. If attached servers are heterogeneous platforms (for example, a mix of UNIX and Windows), sharing of data between such unlike operating system environments is complex. This is due to differences in file systems, access controls, data formats, and encoding structures.

SAN File System architecture

The IBM TotalStorage SAN File System architecture makes it possible to bring the benefits of the existing mainframe system-managed storage (SMS) to the SAN environment. Features such as policy-based allocation, volume management, and file management have long been available on IBM mainframe systems. However, the infrastructure for such centralized, automated management has been lacking in the open systems world of Linux, Windows, and UNIX. On conventional systems, storage management is platform dependent. IBM TotalStorage SAN File System provides a single, centralized point of control to better manage files and data, and is platform independent. Centralized file and data management dramatically simplifies storage administration and lowers TCO.

The SAN File System is a common file system specifically designed for storage networks. By managing file details (via the metadata server) on the storage network instead of in individual servers, the SAN File System design moves the file system intelligence into the storage network where it can be available to all application servers. The file level virtualization aggregation provides immediate benefits: A single global namespace and a single point of management. This eliminates the need to manage files on a server-by-server basis. A global namespace is the ability to access any file from any client system using the same name.

The SAN File System automates routine and error-prone tasks such as file placement, and handles out of space conditions. The IBM TotalStorage SAN File System will allow true heterogeneous file sharing—where read and write on the exact same data can be done by different operating systems.

10

Solutions

The added value of a SAN lies in the exploitation of its technology to provide tangible and desirable benefits to the business. Benefits range from increased availability and flexibility to additional functionality that can reduce application downtime. This chapter contains only a description of general SAN applications, and the kinds of components required to implement them. There is far more complexity than is presented here. For instance, this text will not cover how to choose one switch over another, or how many ISLs are necessary for a given SAN design. For detailed case studies refer to *IBM TotalStorage: SAN Product, Design, and Optimization Guide*, SG24-6384.

10.1 Introduction

There are many categories in which SAN solutions can be classified. We have chosen to classify ours as: infrastructure simplification, business continuity and information lifecycle management. In this chapter we discuss the use of basic SAN design patterns to build solutions for different requirements, ranging from simple data movement techniques, frequently employed as a way to improve business continuity, to sophisticated storage pooling techniques, used to simplify complex infrastructures.

Before we do that we will present some basic principles to be considered when planning a SAN implementation, or upgrade.

10.2 Basic solution principles

A number of important decisions need to be made by the system architect either when a new SAN is being designed or when an existing SAN is being expanded; such decisions usually refer to the choice of the connectivity technology, the best practices for adding capacity to a SAN or the more suitable technology for achieving data integration. This section discusses some of these aspects.

10.2.1 Connectivity

Connecting servers to storage devices through a SAN fabric is often the first step taken in a phased SAN implementation. Fibre Channel attachments have the following benefits:

- Running SCSI over Fibre Channel for improved performance
- Extended connection distances (sometimes called remote storage)
- Enhanced addressability

Many implementations of Fibre Channel technology are simple configurations that remove some of the restrictions of existing storage environments, and allow you to build one common physical infrastructure. The SAN uses common cabling to the storage and the other peripheral devices. The handling of separate sets of cables, such as OEMI, ESCON, SCSI single-ended, SCSI differential, SCSI LVD, and others have caused the IT organization management much trauma as it attempted to treat each of these differently. One of the biggest problems is the special handling that is needed to circumvent the various distance limitations.

Installations without SANs commonly use SCSI cables to attach to their storage. SCSI has many restrictions, such as limited speed, a very small number of devices that can be attached, and severe distance limitations. Running SCSI over Fibre Channel helps to alleviate these restrictions. SCSI over Fibre Channel helps improve performance and enables more flexible addressability and much greater attachment distances compared to normal SCSI attachment.

A key requirement of this type of increased connectivity is providing consistent management interfaces for configuration, monitoring, and management of these SAN components. This type of connectivity allows companies to begin to reap the benefits of Fibre Channel technology, while also protecting their current storage investments.

10.2.2 Adding capacity

The addition of storage capacity to one or more servers may be facilitated while the device is connected to a SAN. Depending on the SAN configuration and the server operating system, it may be possible to add or remove devices without stopping and restarting the server.

If new storage devices are attached to a section of a SAN with loop topology (mainly tape drives), the LIP may impact the operation of other devices on the loop. This may be overcome by quiescing operating system activity to all the devices on that particular loop before attaching the new device. This is far less of a problem with the latest generation of loop-capable switches. If storage devices are attached to a SAN by a switch, using the switch and management software it is possible to make the devices available to any system connected to the SAN.

10.2.3 Data movement and copy

Data movement solutions require that data be moved between similar or dissimilar storage devices. Today, data movement or replication is performed by the server or multiple servers. The server reads data from the source device, perhaps transmitting the data across a LAN or WAN to another server, and then the data is written to the destination device. This task ties up server processor cycles and causes the data to travel twice over the SAN—once from source device to a server, and then a second time from a server to a destination device.

The objective of SAN data movement solutions is to be able to avoid copying data through the server (server-free), and across a LAN or WAN (LAN-free), thus freeing up server processor cycles and LAN or WAN bandwidth. Today, this data replication can be accomplished in a SAN through the use of an intelligent gateway that supports the third-party copy SCSI-3 command. Third-party copy implementations are also referred to as outboard data movement or copy implementations.

The following sections list some of the copy services available.

Traditional copy

One of the most frequent tasks for a space manager is moving files through the use of various tools. Another frequent user of traditional copy is space management software, such as Tivoli's TSM, during the reclamation or recycle process. With SAN outboard data movement, traditional copy can be performed server-free, therefore making it easier to plan and faster to execute.

T-0 copy

Another outboard copy service enabled by Fibre Channel technology is T-0 (time=zero) copy. This is the process of taking a snapshot, or freezing the data (databases, files, or volumes) at a certain time, and then allowing applications to update the original data while the frozen copy is duplicated. With the flexibility and extensibility that Fibre Channel brings, these snapshot copies can be made to local or remote devices. The requirement for this type of function is driven by the need for 24x7 availability of key database systems.

Remote copy

Remote copy is a business requirement used in order to protect data from disasters, or to migrate data from one location to avoid application downtime for planned outages such as hardware or software maintenance.

Today, remote copy solutions are either synchronous or asynchronous, and they require different levels of automation in order to guarantee data consistency across disks and disk subsystems. Today's remote copy solutions are implemented only for disks at a physical or logical volume level.

In the future, with more advanced storage management techniques such as outboard hierarchical storage management and file pooling, remote copy solutions need to be implemented at the file level. This implies more data to be copied, and requires more advanced technologies to guarantee data consistency across files, disks, and tape in multi-server heterogeneous environments. A SAN is required to support bandwidth and management of such environments

10.2.4 Upgrading to faster speeds

One of the other considerations of any SAN environment is how newer, faster technology is to be introduced. Both 4 Gbps and 10 Gbps products are beginning to reach the market. For most applications this will not mean that they will immediately benefit. Applications that have random or "bursty" I/O will not necessarily see a benefit. Applications that stream large amounts of data are likely to see the most immediate benefits. One place that makes sense for 4 Gbps to be utilized is the inter-switch link. This has two advantages: firstly, the increased speed between switches is the obvious one, and secondly the advantage is that it may be possible to have fewer ISls with the increased bandwidth. This means that it may be possible to “rescue” ISLs and use them to attach hosts or storage.

Most businesses should have a plan in place to transition to 4 Gbps, or at the very least they should be attempting to identify the areas that should become 4 Gbps first. In terms of where the migration should take place first, it makes sense to transition to 4 Gbps switches and directors first (these were the first to market), then to 4 Gbps capable HBAs (which are beginning to come to the market), and then finally 4 Gbps capable devices.

10.3 Infrastructure simplification

Few would question the statement that IT infrastructures have grown more complex in recent years. The dramatic growth in the use of IT, combined with distributed computing architectures, is part of the reason. But business processes have also become more complex and integrated too, driving a greater need for complex interconnections among systems.

The added complexity that accompanies growth can stand in the way of fully realizing the benefits of IT. Infrastructure simplification is a way to look at the entire IT operation and help eliminate the complexity that can raise costs, reduce reliability, and create dependencies on specialized skills—factors making it harder to operate as an on demand business.

In this section we present some of the main solution designs for achieving infrastructure simplification.

10.3.1 Storage pooling

Before SANs, the concept of the physical pooling of devices in a common area of the computing center was often just not possible, and when it was possible, it required expensive and unique extension technology. By introducing a network between the servers and the storage resources, this problem is minimized. Hardware interconnections become common across all servers and devices. For example, common trunk cables can be used for all servers, storage, and switches.

This section describes the two main types of storage device pooling: *disk pooling* and *tape pooling*.

Disk pooling

Disk pooling allows multiple servers to utilize a common pool of SAN-attached disk storage devices. Disk storage resources are pooled within a disk subsystem or across multiple IBM and non-IBM disk subsystems, and capacity is assigned to independent file systems supported by the operating systems on servers. The servers are potentially a heterogeneous mix of UNIX, Windows, and even OS/390.

Storage can be dynamically added to the disk pool and assigned to any SAN-attached server when and where it is needed. This provides efficient access to shared disk resources without a level of indirection associated with a separate file server, since storage is effectively *directly attached* to all the servers, and efficiencies of scalability result from consolidation of storage capacity.

When storage is added, zoning can be used to restrict access to the added capacity. As many devices (or LUNs) can be attached to a single port, access can be further restricted using LUN-masking, that is, specifying who can access a specific device or LUN.

Attaching and detaching storage devices can be done under the control of a common administrative interface. Storage capacity can be added without stopping the server, and can be immediately made available to applications.

Figure 10-1 shows an example of disk storage pooling across two servers.

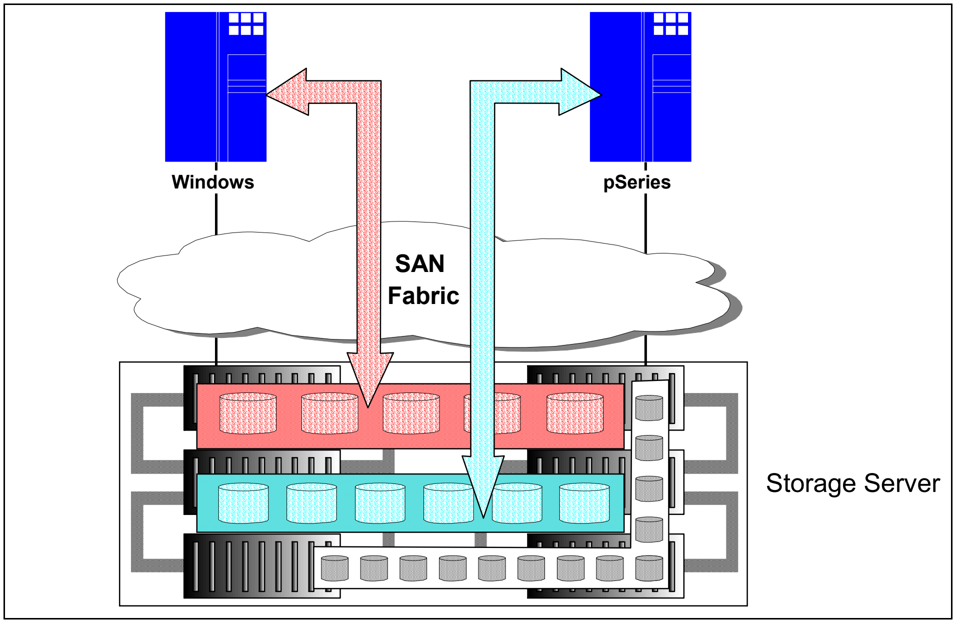


Figure 10-1   Disk pooling

One server is assigned a pool of disks formatted to the requirements of the file system, and the second server is assigned another pool of disks, possibly in another format. The third pool shown may be space not yet allocated or pre-formatted disk for future use.

Tape pooling

Tape pooling addresses the problem faced today in an open systems environment in which multiple servers are unable to share tape resources across multiple hosts. Older methods of sharing a device between hosts consist of either manually switching the tape device from one host to the other, or writing applications that communicate with connected servers through distributed programming.

Tape pooling allows applications on one or more servers to share tape drives, libraries, and cartridges in a SAN environment in an automated, secure manner. With a SAN infrastructure, each host can directly address the tape device as though it were connected to all of the hosts.

Tape drives, libraries, and cartridges are owned by either a central manager or a peer-to-peer management implementation, and are dynamically allocated and reallocated to systems as required, based on demand. Tape pooling allows for resource sharing, automation, improved tape management, and added security for tape media.

Software is required to manage the assignment and locking of the tape devices in order to serialize tape access. Tape pooling is a very efficient and cost-effective way of sharing expensive tape resources, such as automated tape libraries. Tape libraries can even be shared between operating systems.

At any particular instant in time, a tape drive can be owned by one system, as shown in Figure 10-2 on page 256.

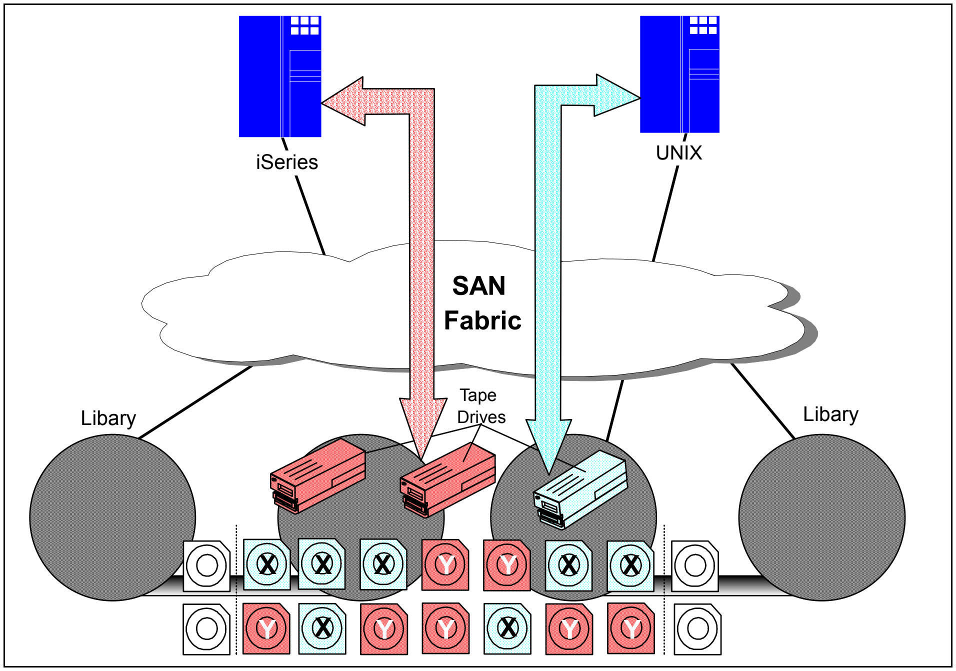


Figure 10-2 Tape pooling

In this example, the iSeries server currently has two tape drives assigned, and the UNIX server has only one drive assigned. The tape cartridges, physical or virtual, in the libraries are assigned to different applications or groups and contain current data, or are assignable to servers (in scratch groups) if they are not yet used, or they no longer contain current data.

10.3.2 Data sharing

The term *data sharing* refers to accessing the same data from multiple systems and servers. It is often used synonymously with storage partitioning and disk pooling. True data sharing goes a step beyond sharing storage capacity with pooling, in that multiple servers are actually sharing the data on the storage devices. The architecture that the zSeries servers are built on have supported data sharing since the early 1970s.

While data sharing is not a solution that is exclusive to SANs, the SAN architecture can take advantage of the connectivity of multiple hosts to the same storage in order to enable data to be shared more efficiently than through the services of a file server or NAS unit, as is often the case today. SAN connectivity has the potential to provide sharing services to heterogeneous hosts, including UNIX, Windows, and z/OS.

Storage partitioning is usually the first stage towards true data sharing, usually implemented in a server and storage consolidation project. There are multiple stages or phases towards true data sharing:

- Logical volume partitioning
- File pooling
- True data sharing

Logical volume partitioning

Storage partitioning does not represent a true data sharing solution. It is essentially just a way of splitting the capacity of a single storage server into to multiple pieces. The storage subsystems are connected to multiple servers, and storage capacity is partitioned among the various subsystems.

Logical disk volumes are defined within the storage subsystem and assigned to servers. The logical disk is addressable from the server. A logical disk may be a subset or superset of disks only addressable by the subsystem itself. A logical disk volume can also be defined as subsets of several physical disks (striping). The capacity of a disk volume is set when defined. For example, two logical disks, with different capacities (for example, 50 GB and 150 GB) may be created from a single 300 GB hardware addressable disk, with each being assigned to a different server, leaving 100 GB of unassigned capacity. A single 2000 GB logical disk may also be created from multiple real disks that exist in different storage subsystems. The underlying storage controller must have the necessary logic to manage the volume grouping, and guarantee access securely to the data.

Figure 10-3 on page 258 shows multiple servers accessing logical volumes created using the different alternatives mentioned above. (The logical volume *Another volume* is not assigned to any server.)

.

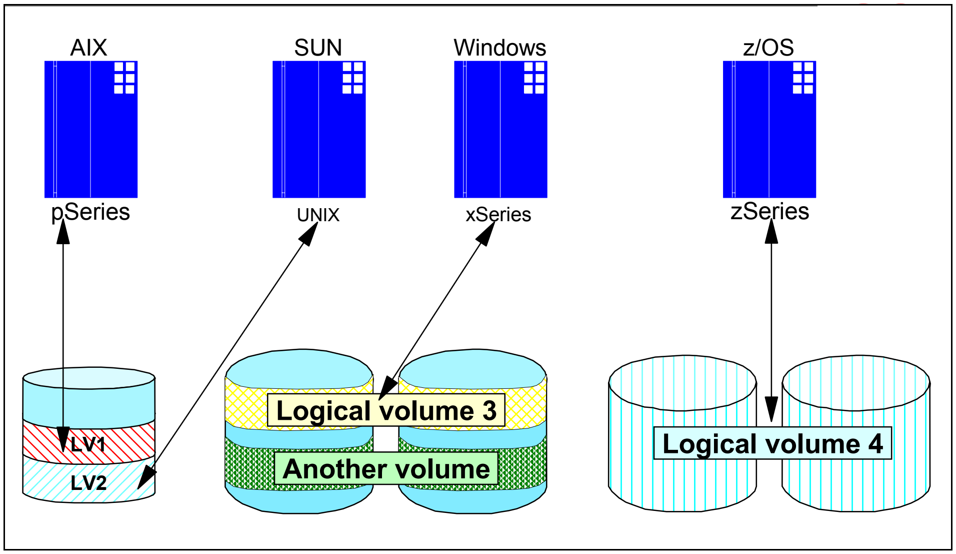


Figure 10-3 Logical volume partitioning

File pooling

File pooling assigns disk space (as needed) to contain the actual file being created. Instead of assigning disk capacity to individual servers on a physical or logical disk basis, or by using the operating system functions (as in z/OS, for example) to manage the capacity, file pooling presents a mountable name space to the application servers. This is similar to the way NFS behaves today. The difference is that there is direct channel access, not network access as with NFS, between the application servers and the disk(s) where the file is stored. Disk capacity is assigned only when the file is created and released when the file is deleted. The files can be shared between servers in the same way (operating system support, locking, security, and so on) as though they were stored on a shared physical or logical disk.

Figure 10-4 on page 259 shows multiple servers accessing files in shared storage space. The unassigned storage space can be reassigned to any server on an as-needed basis when new files are created.

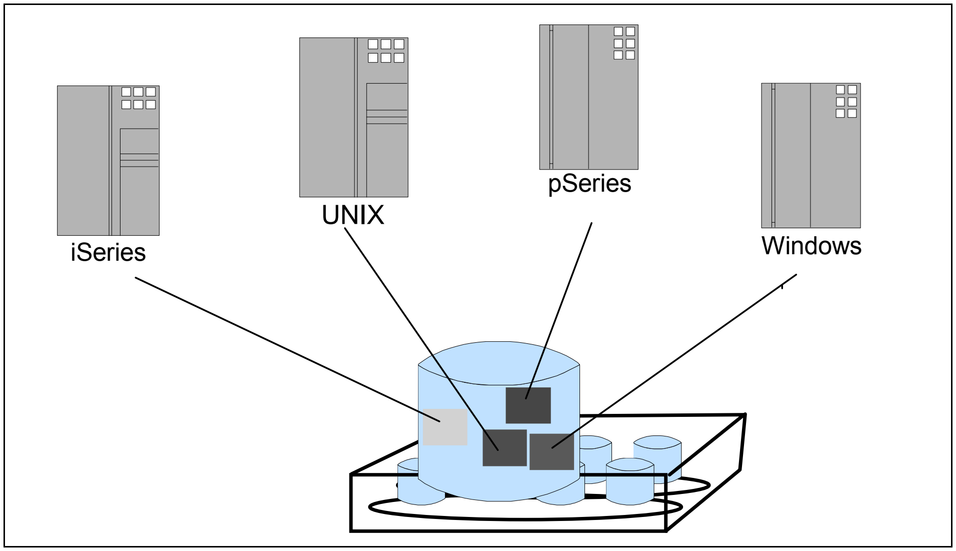


Figure 10-4 File pooling

True data sharing

In true data sharing, the same copy of data is accessed concurrently by multiple servers. This allows for considerable storage savings, and may be the basis upon which storage consolidation can be built. There are various levels of data sharing:

- Sequential, point-in-time, or one-at-a-time access. This is really the serial reuse of data. It is assigned first to one application, then to another application, in the same or a different server, and so on.
- Multi-application simultaneous read access. In this model, multiple applications in one or multiple servers can read data, but only one application can update it, thereby eliminating any integrity problems.
- Multi-application simultaneous read and write access. This is similar to the situation described above, but all hosts can update the data. There are two versions of this—one where all applications are on the same platform (homogeneous), and one where the applications are on different platforms (heterogeneous).

With true data sharing, multiple reads and writes can happen at the same time. Multiple read operations are not an issue, but multiple write operations can potentially access and overwrite the same information. A serialization mechanism is needed to guarantee that the data written by multiple applications is written to the disk in an orderly way. Serialization mechanisms may be defined to serialize from a group of physical or logical disk volumes to an individual physical block of data within a file or database.

Such a form of data sharing requires complicated co-ordination across multiple servers on a level far greater scale than mere file-locking.

Homogeneous data sharing

Figure 10-5 shows multiple hosts accessing and sharing the same data. The data encoding mechanism across these hosts is common and usually platform dependent. The hosts or the storage subsystem must provide a serialization mechanism for accessing the data to ensure write integrity and serialization.

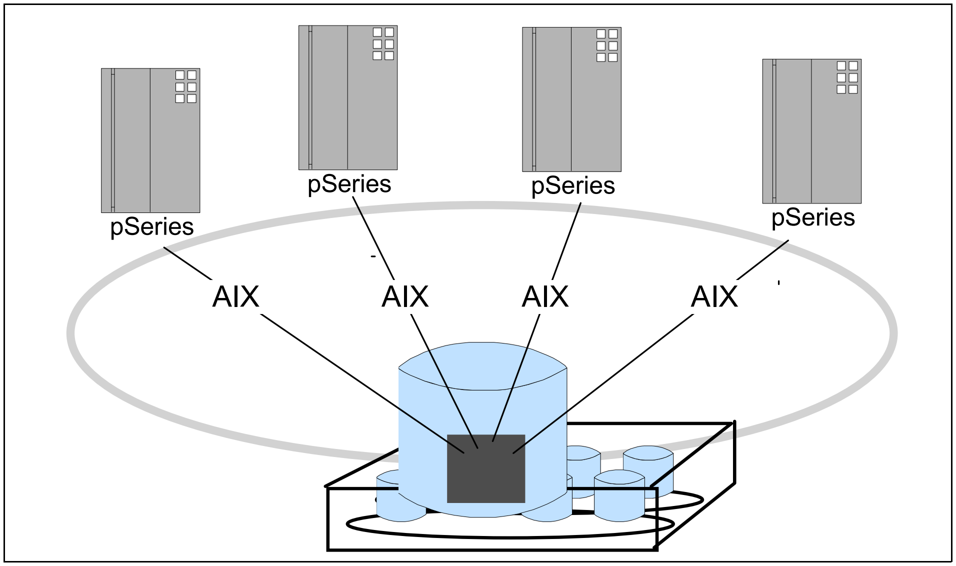


Figure 10-5 Homogeneous data sharing

Heterogeneous data sharing

In heterogeneous data sharing (as illustrated in Figure 10-6 on page 261) different operating systems access the same data. The issues are similar to those in homogeneous data sharing with one major addition: The data must be stored in a common file system, but may be with a common encoding and other conventions; or the file system logic will be needed to perform the necessary conversions of EBCDIC or ASCII, and any other differences. Thus, we have the requirement for a SAN distributed file system. With the appropriate technology it will be possible to provide access to the files.

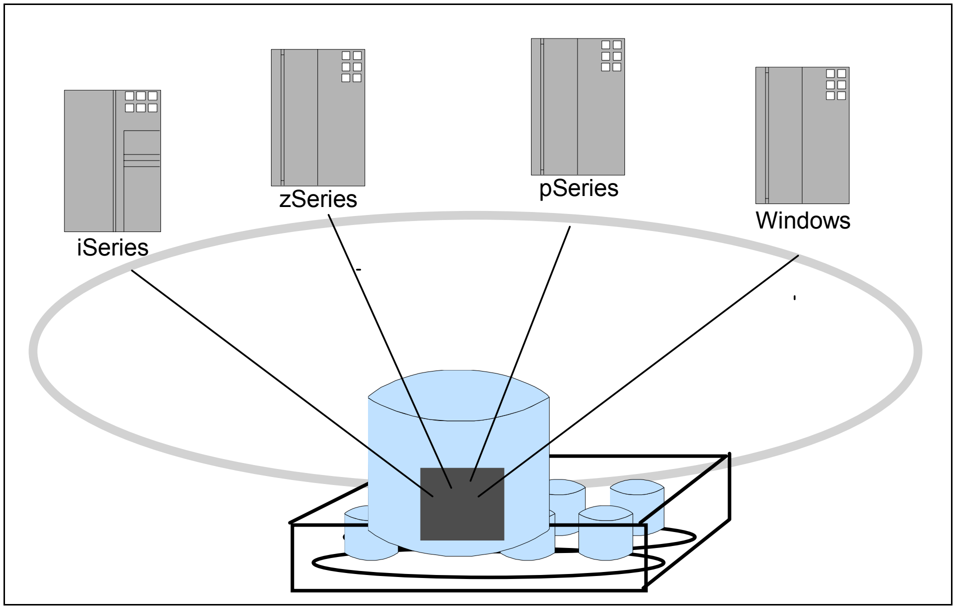


Figure 10-6 Heterogeneous data sharing

10.3.3 Clustering

SAN architecture naturally lends itself to scalable server clustering in a share-all situation, because a cluster of homogenous servers can see a single system image of the data. While this was possible with SCSI using multiple pathing, scalability is an issue because of the distance constraints of SCSI. SCSI allows for distances of up to 25 meters, and the size of SCSI connectors limits the number of connections to servers or subsystems.

A SAN allows for efficient load balancing in distributed processing application environments. Applications that are processor-constrained on one server can be executed on a different server with more processor capacity. In order to do this, both servers must be able to access the same data volumes, and serialization of access to data must be provided by the application or operating system services. Today, the S/390 Parallel Sysplex® provides services and operating system facilities for seamless load balancing across many members of a server complex.

In addition to this advantage, SAN architecture also lends itself to exploitation in a failover situation, whereby the secondary system can take over upon failure of the primary system and have direct addressability to the storage that was used by the primary system. This improves reliability in a clustered system environment, because it eliminates downtime due to processor unavailability.

Figure 10-7 shows an example of clustering. Servers S1 and S2 share IBM Enterprise Storage Servers #1 and 2. If S1 fails, S2 can access the data on ESS #1. The example also shows that ESS #1 is mirrored to ESS #2. Moving the standby server, S2, to the remote WAN connected site would allow for operations to continue in the case of a disaster being declared.

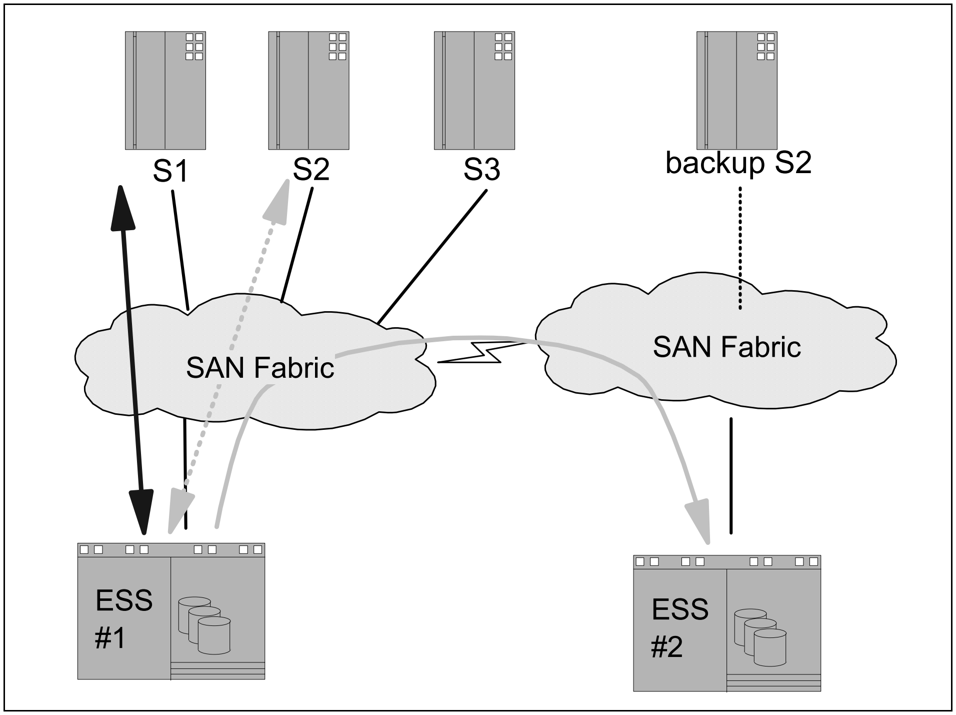


Figure 10-7 Server clustering

10.3.4 Consolidation

We can improve scalability, security, and manageability by enabling devices in separate SAN fabrics to communicate without merging fabrics into a single, large SAN fabric. This capability enables customers to initially deploy separate SAN solutions at the departmental and data center levels and then to consolidate them into large enterprise SAN solutions as their experience and requirements grow and change.

Customers have deployed multiple SAN islands for different applications with different fabric switch solutions. Growing availability of iSCSI server capabilities has created the opportunity for low-cost iSCSI server integration and storage consolidation. Additionally, depending on the choice of router, they will provide FCIP or iFCP capability.

The new SAN routers provide an iSCSI Gateway Service to integrate low-cost Ethernet-connected servers to existing SAN infrastructures. It also provides Fibre Channel, FC-FC Routing Service to interconnect multiple SAN islands without requiring the fabrics to merge into a single large SAN.

In Figure 10-8 we show an example using a multiprotocol router to extend SAN capabilities across the enterprise.

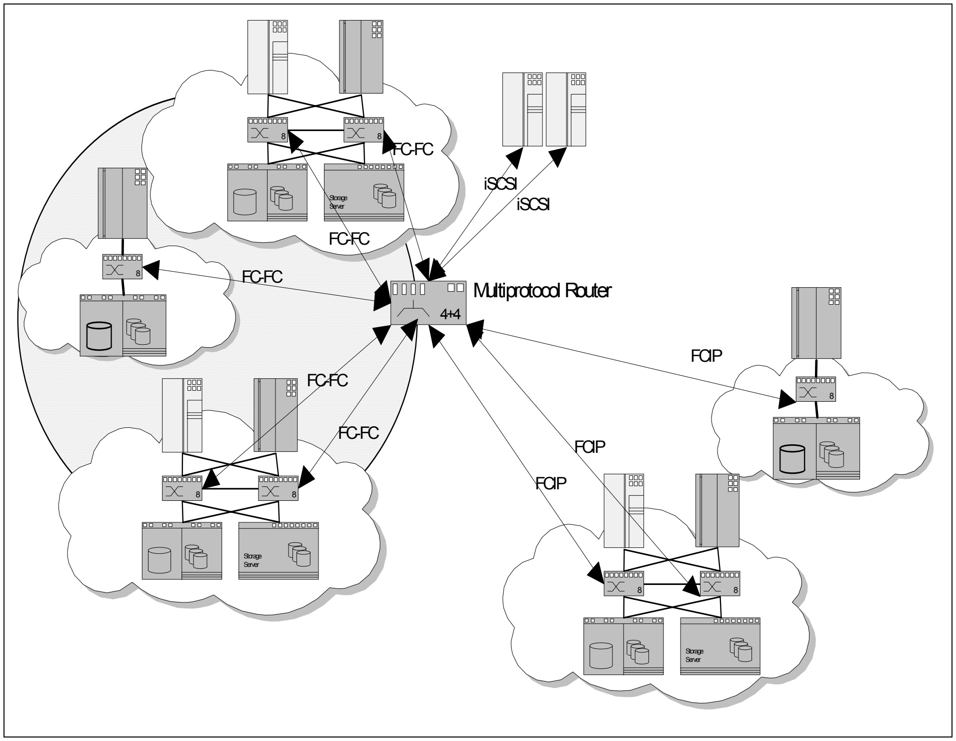


Figure 10-8 SAN consolidation

A multiprotocol capable router solution brings a number of benefits to the marketplace. In the example shown there are a number of discrete SAN islands, and a number of different protocols involved. For a merge of the SAN fabric to take place, it would involve a number of disruptive and potentially expensive actions such as:

- Downtime
- Purchase of additional switches/ports
- Purchase of HBAs
- Migration costs
- Configuration costs
- Purchase of additional licenses

- Ongoing maintenance

However, by installing a multiprotocol router the advantages are:

- Least disruptive method
- No need to purchase extra HBAs
- Minimum number of ports to connect to the router
- No expensive downtime
- No expensive migration costs
- No ongoing maintenance costs other than router
- Support of other protocols
- Increases ROI by consolidating resource
- Can be used to isolate the SAN environment to be more secure

There are many more benefits that the router can provide. In this example, an FC-FC routing service that negates the need for a costly SAN fabric merge exercise, the advantages are apparent and real. The router can also be used to provide:

- Device connectivity across multiple SANs for infrastructure simplification
- Tape backup consolidation for information lifecycle management
- Long-distance SAN extension for business continuance
- Low-cost server connectivity to SAN resources

More information can be found at:

http://www-03.ibm.com/servers/storage/solutions/is/index.html

10.4 Business continuity

On demand businesses rely on their IT systems to conduct business. Everything must be working all the time! In this scenario, a sound and comprehensive business continuity strategy encompasses high availability, near continuous operations, and disaster recovery.

Today, data protection of multiple network-attached servers is performed according to one of two backup and recovery paradigms: Local backup and recovery, or network backup and recovery.

The local backup and recovery paradigm has the advantage of speed, because the data does not travel over the network. However, with a local backup and recovery approach, there are costs for overhead (because local devices must be acquired for each server, and are thus difficult to utilize efficiently), and management overhead (because of the need to support multiple tape drives, libraries, and mount operations).

The network backup and recovery paradigm is more cost-effective, because it allows for the centralization of storage devices using one or more network-attached devices. This centralization allows for a better return on investment, as the installed devices will be utilized more efficiently. One tape library can be shared across many servers. Management of a network backup and recovery environment is often simpler than the local backup and recovery environment, because it eliminates the potential need to perform manual tape mount operations on multiple servers.

SANs combine the best of both approaches. This is accomplished by central management of backup and recovery, assigning one or more tape devices to each server, and using FC protocols to transfer data directly from the disk device to the tape device, or vice versa over the SAN.

In the following sections we discuss these approaches in more detail.

10.4.1 LAN-free data movement

The network backup and recovery paradigm implies that data flows from the backup and recovery client (usually a file or database server) to the backup and recovery server, or between backup and recovery servers, over a network connection. The same is true for archive or hierarchical storage management applications. Often the network connection is the bottleneck for data throughput. This is due to network connection bandwidth limitations. The SAN can be used instead of the LAN as a transport network.

Tape drive and tape library sharing

A basic requirement for LAN-free or server-free backup and recovery is the ability to share tape drives and tape libraries, as described in “Tape pooling” on page 255, between backup and recovery servers, and between a backup and recovery server, and its backup and recovery client (usually a file or database server). Network-attached end-user backup and recovery clients will still use the network for data transportation.

In the tape drive and tape library sharing approach, the backup and recovery server or client that requests a backup copy to be copied to or from tape will read or write the data directly to the tape device using SCSI commands. This approach bypasses the network transport’s latency and network protocol path length; therefore, it can offer improved backup and recovery speeds in cases where the network is the constraining factor. The data is read from the source device and written directly to the destination device.

Figure 10-9 on page 266 shows an example of tape drive or tape library sharing.

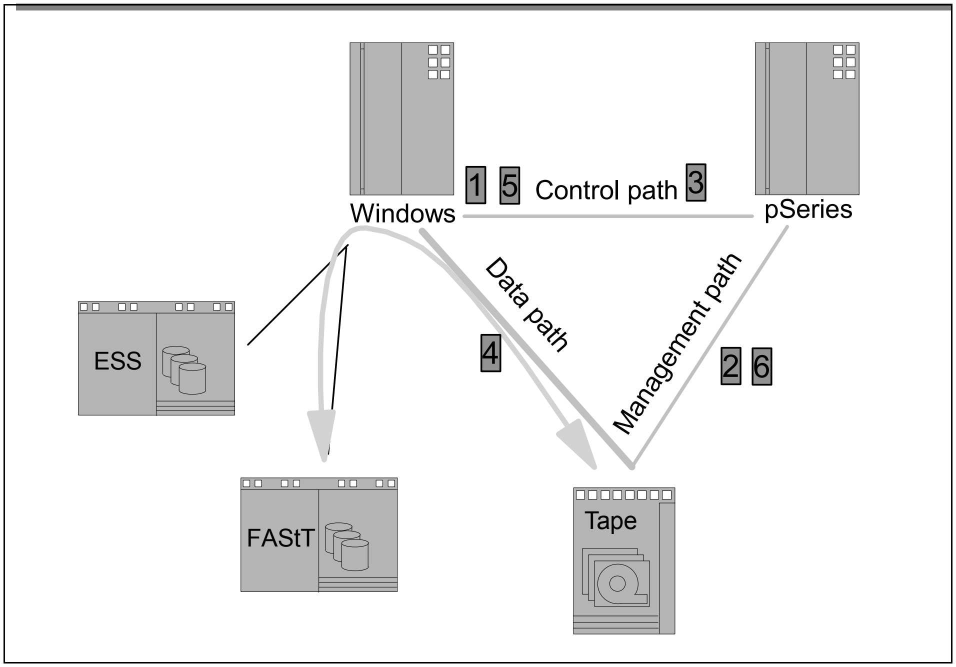


Figure 10-9 LAN-less backup and recovery

Where:

1. A backup and recovery client requests one or more tapes to perform the backup operations. This request is sent over a control path, which could be a standard network connection between client and server. The backup and recovery server then assigns one or more tape drives to the client for the duration of the backup operation.
2. The server then requests the tapes required to be mounted into the drives using the management path.
3. The server then notifies the client that the tapes are ready.
4. The client performs the backup or recovery operations over the data path.
5. When the client completes the operations, it notifies the server that it has completed the backup or recovery, and the tapes can be released.
6. The server requests the tape cartridges to be dismounted, using the management path for control flow.

10.4.2 Server-free data movement

In the preceding approaches, server intervention was always required to copy the data from source device to target device. The data was read from the source device into the server memory, and then written from the server memory to the target device. The server-free data movement approach avoids the use of any server or IP network for data movement, only using the SAN for carrying out the SCSI-3 third-party copy function.

Figure 10-10 illustrates this approach. Management of the tape drives and cartridges is handled as in the preceding example. The client issues a third-party copy SCSI-3 command that will cause the data to be copied from the source device to the target device. No server processor cycle or IP-network bandwidth is used.

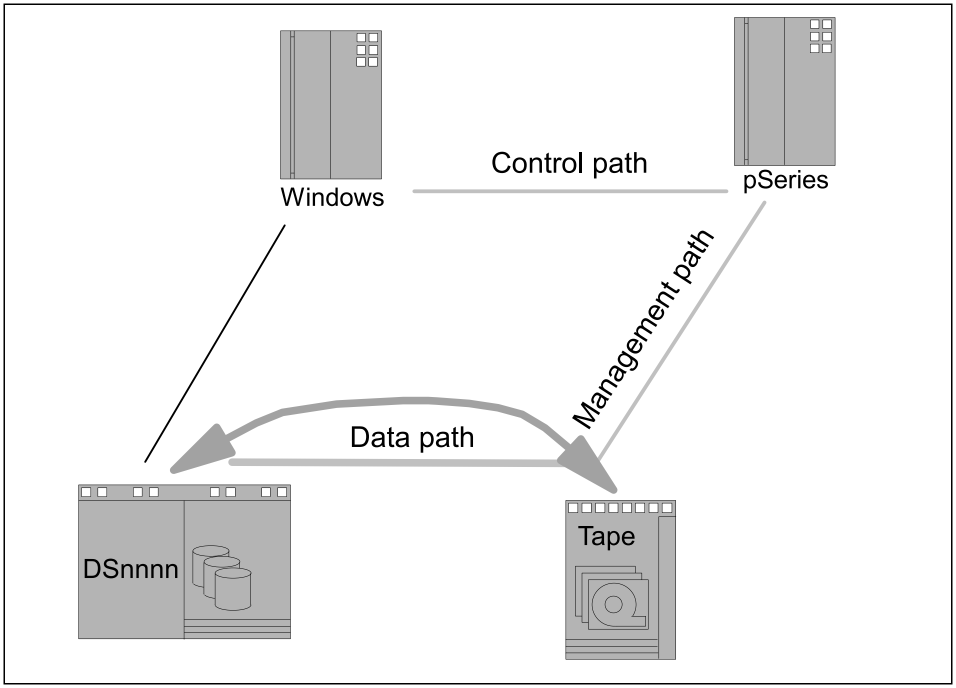


Figure 10-10 Server-free data movement for backup and recovery

In the example shown in Figure 10-10, the backup and recovery client issued the third-party copy command to perform a backup or recovery using tape pooling. Another implementation would be for the backup and recovery server to initiate the third-party copy SCSI-3 command on request from the client, using disk pooling.

The third-party copy SCSI-3 command defines block-level operations, as is the case for all SCSI commands. The SCSI protocol is not aware of the file system or database structures. Using third-party copy for file-level data movement requires the file systems to provide mapping functions between file system files and device block addresses. This mapping is a first step towards sophisticated database backup and recovery, log archiving, and so on.

The server part of a backup and recovery application also performs many other tasks requiring server processor cycles for data movement; for example, data migration and reclamation/recycle. During reclamation, data is read from the tape cartridge to be reclaimed into server memory, and then written from server memory to a new tape cartridge.

The server-free data movement approach avoids the use of extensive server processor cycles for data movement, as shown in Figure 10-11.

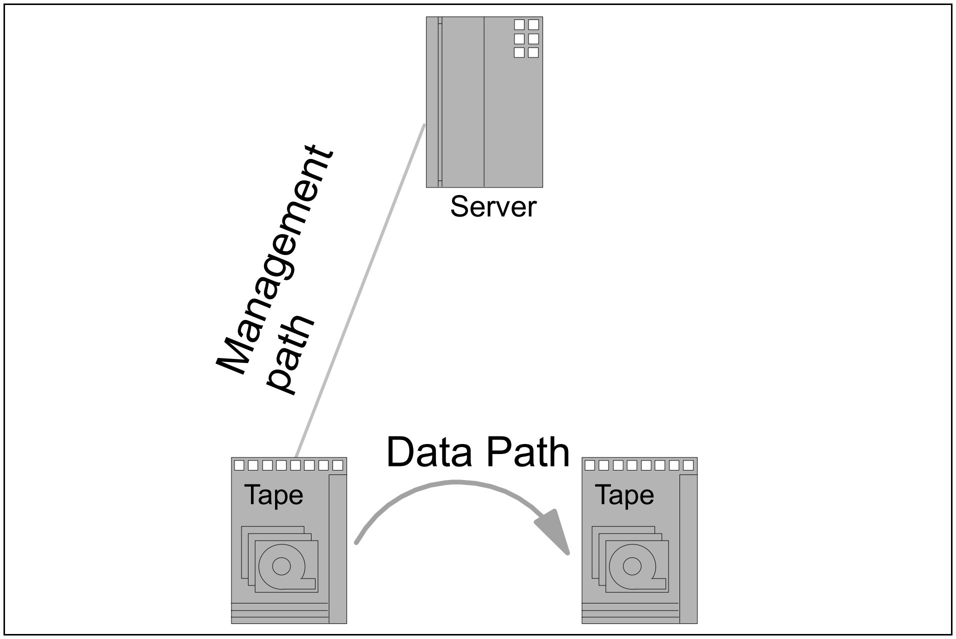


Figure 10-11 Server-free data movement for tape reclamation

10.4.3 Disaster backup and recovery

SANs can facilitate disaster backup solutions because of the greater flexibility allowed in connecting storage devices to servers, and also the greater distances that are supported when compared with SCSI's restrictions. It is possible when using a SAN infrastructure to perform extended distance backups for disaster recovery within a campus or city, as shown in Figure 10-12.

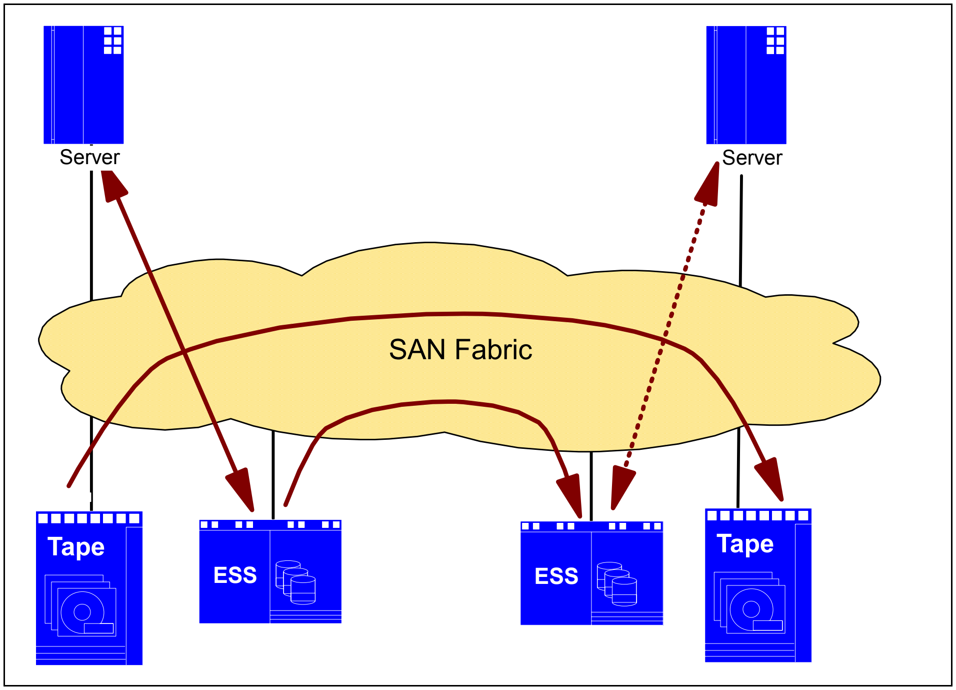


Figure 10-12 Disaster backup and recovery

When longer distances are required, SANs must be connected using gateways and WANs, similar to the situation discussed in 10.3.3, “Clustering” on page 261, and shown in Figure 10-7 on page 262.

Depending on business requirements, disaster protection implementations may make use of copy services implemented in disk subsystems and tape libraries (that might be implemented using SAN services), SAN copy services, and most likely a combination of both.

Additionally, services and solutions—similar to Geographically Dispersed Parallel Sysplex™ (GDPS) for zSeries servers, available today from IBM Global Services—will be required to monitor and manage these environments.

More information can be found at:

`http://www-03.ibm.com/servers/storage/solutions/business_continuity/index.html`

10.5 Information lifecycle management

Information lifecycle management (ILM) is a process for managing information through its lifecycle, from conception until disposal, in a manner that optimizes storage and access at the lowest cost.

ILM is not just hardware or software—it includes processes and policies to manage the information. It is designed upon the recognition that different types of information can have different values at different points in their lifecycle. Predicting storage needs and controlling costs can be especially challenging as the business grows.

The overall objectives of managing information with ILM are to help reduce the total cost of ownership (TCO) and help implement data retention and compliance policies. In order to effectively implement ILM, owners of the data need to determine how information is created, how it ages, how it is modified, and if/when it can safely be deleted. ILM segments data according to value, which can help create an economical balance and sustainable strategy to align storage costs with businesses objectives and information value.

10.5.1 ILM elements

To manage the data lifecycle and make your business ready for on demand, there are four main elements that can address your business in an ILM structured environment. They are:

- Tiered storage management
- Long-term data retention
- Data lifecycle management
- Policy-based archive management

In the next sections we describe each of these elements in greater detail.

10.5.2 Tiered storage management

Most organizations today seek a storage solution that can help them manage data more efficiently. They want to reduce the costs of storing large and growing amounts of data and files and maintain business continuity. Through tiered storage, you can reduce overall disk-storage costs, by providing benefits like:

- Reducing overall disk-storage costs by allocating the most recent and most critical business data to higher performance disk storage, while moving older and less critical business data to lower cost disk storage.
- Speeding business processes by providing high-performance access to most recent and most frequently accessed data.
- Reducing administrative tasks and human errors. Older data can be moved to lower cost disk storage automatically and transparently.

Typical storage environment

Storage environments typically have multiple tiers of *data value*, such as application data that is needed daily, and archive data that is accessed infrequently. However, typical storage configurations offer only a single tier of storage, as shown in Figure 10-13, which limits the ability to optimize cost and performance.

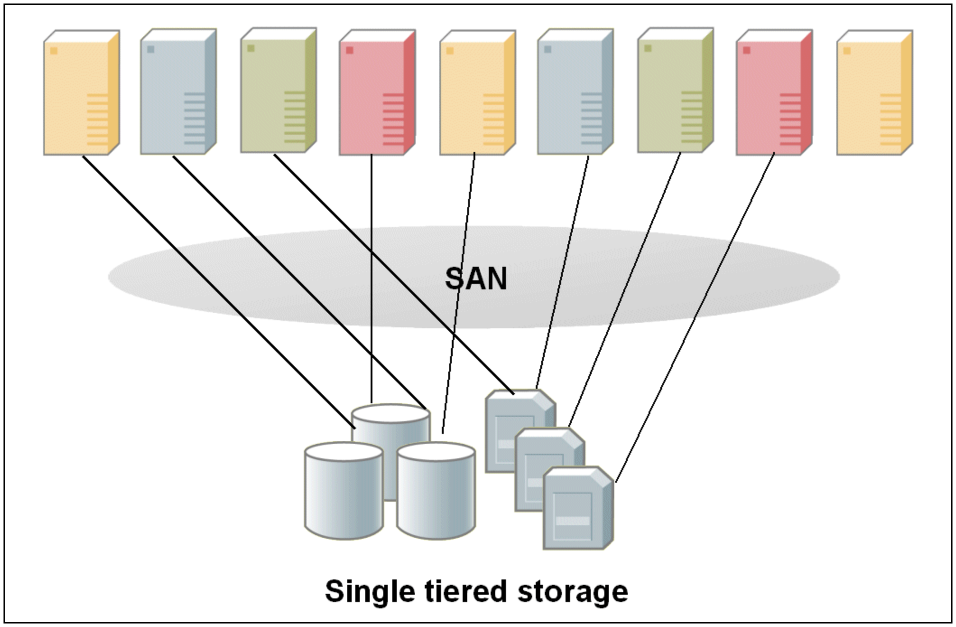


Figure 10-13 Traditional non-tiered storage environment

Multi-tiered storage environment

A tiered storage environment that utilizes the SAN infrastructure affords the flexibility to align storage cost with the changing value of information. The tiers will be related to data value. The most critical data is allocated to higher performance disk storage, while less critical business data is allocated to lower cost disk storage.

Each storage tier will provide different performance metrics and disaster recovery capabilities. Creating classes and storage device groups is an important step to configure a tiered storage ILM environment.

Figure 10-14 shows a multi-tiered storage environment.

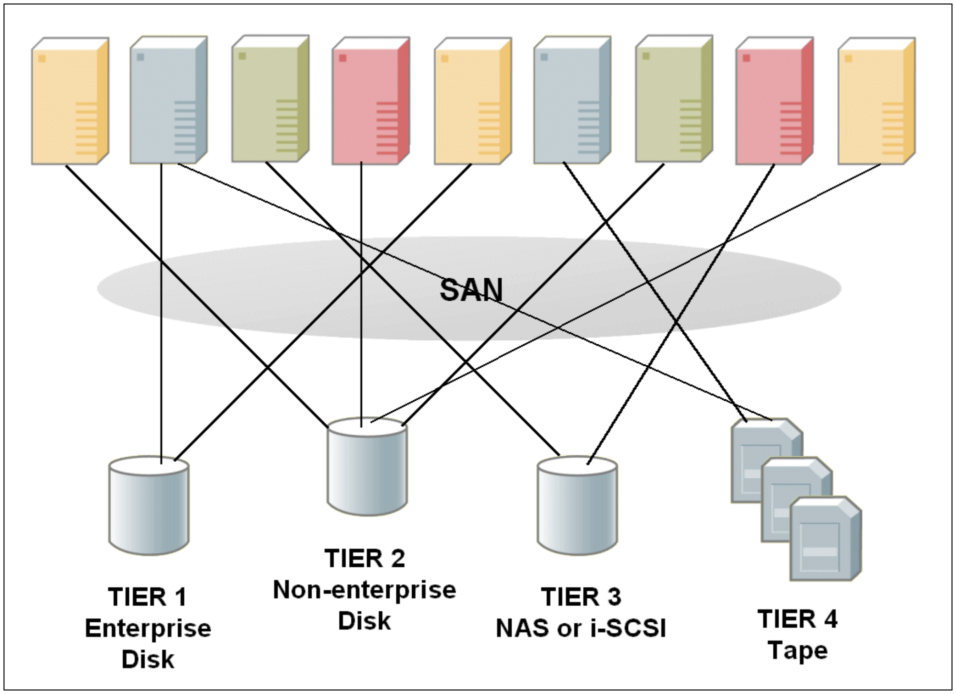


Figure 10-14 ILM tiered storage environment

An IBM ILM solution in a tiered storage environment is designed to:

- Reduce the total cost of ownership (TCO) of managing information. It can help optimize data costs and management, freeing expensive disk storage for the most valuable information.
- Segment data according to value. This can help create an economical balance and sustainable strategy to align storage costs with business objectives and information value.
- Help make decisions about moving, retaining, and deleting data, because ILM solutions are closely tied to applications.
- Manage information and determine how it should be managed based on content, rather than migrating data based on technical specifications. This approach can help result in more responsive management, and offers you the ability to retain or delete information in accordance with business rules.

- Provide the framework for a comprehensive enterprise content management strategy.

Key products of IBM for tiered storage solutions and storage virtualization solutions are:

- IBM TotalStorage SAN Volume Controller (SVC)
- IBM TotalStorage DS family of disk storage - DS4x000, DS6000, and DS8000
- IBM TotalStorage tape drives, tape libraries, and virtual tape solutions

10.5.3 Long-term data retention

There is a rapidly growing class of data that is best described by the way in which it is managed rather than the arrangement of its bits. The most important attribute of this kind of data is its retention period, hence it is called *retention managed data*, and it is typically kept in an archive or a repository. In the past it has been variously known as archive data, fixed content data, reference data, unstructured data, and other terms implying its read-only nature. It is often measured in terabytes and is kept for long periods of time, sometimes forever.

In addition to the sheer growth of data, laws and regulations governing the storage and secure retention of business and client information are increasingly becoming part of the business landscape, making data retention a major challenge to any institution. An example of these is the Sarbanes-Oxley Act in the US, of 2002.

Businesses must comply with these laws and regulations. Regulated information can include e-mail, instant messages, business transactions, accounting records, contracts, or insurance claims processing, all of which can have different retention periods, for example, for 2 years, for 7 years, or retained forever. Data is an asset when it needs to be kept; however, data kept past its mandated retention period could also become a liability. Furthermore, the retention period can change due to factors such as litigation. All these factors mandate tight coordination and the need for ILM.

Not only are there numerous state and governmental regulations that must be met for data storage, but there are also industry-specific and company-specific ones. And of course these regulations are constantly being updated and amended. Organizations need to develop a strategy to ensure that the correct information is kept for the correct period of time, and is readily accessible when it needs to be retrieved at the request of regulators or auditors.

It is easy to envisage the exponential growth in data storage that will result from these regulations and the accompanying requirement for a means of managing this data. Overall, the management and control of retention managed data is a significant challenge for the IT industry when taking into account factors such as cost, latency, bandwidth, integration, security, and privacy.

IBM ILM data retention strategy

Regulations and other business imperatives stress the need for an Information Lifecycle Management process and tools to be in place. Key products of IBM for data retention and compliance solutions are:

- IBM Tivoli Storage Manager, including IBM System Storage Archive Manager
- IBM DB2 Content Manager Family, which includes DB2 Content Manager, Content Manager OnDemand, CommonStore for Exchange Server, CommonStore for Lotus® Domino®, and CommonStore for SAP
- IBM DB2 Records Manager
- IBM TotalStorage DS4000 with S-ATA disks
- IBM System Storage DR550
- IBM TotalStorage Tape (including WORM) products

10.5.4 Data lifecycle management

At its core, the process of ILM moves data up and down a path of tiered storage resources, including high-performance, high-capacity disk arrays, lower-cost disk arrays such as serial ATA (SATA), tape libraries, and permanent archival media where appropriate. Yet ILM involves more than just data movement; it encompasses scheduled deletion and regulatory compliance as well. Because decisions about moving, retaining, and deleting data are closely tied to application use of data, ILM solutions are usually closely tied to applications.

ILM has the potential to provide the framework for a comprehensive information-management strategy, and helps ensure that information is stored on the most cost-effective media. This helps enable administrators to make use of tiered and virtual storage, as well as process automation. By migrating unused data off of more costly, high-performance disks, ILM is designed to help:

- Reduce costs to manage and retain data.
- Improve application performance.
- Reduce backup windows and ease system upgrades.
- Streamline™ data management.
- Allow the enterprise to respond to demand—in real-time.
- Support a sustainable storage management strategy.
- Scale as the business grows.

ILM is designed to recognize that different types of information can have different value at different points in their lifecycle. As shown in Figure 10-15 on page 275, data can be allocated to a specific storage level aligned to its cost, with policies defining when and where data will be moved.

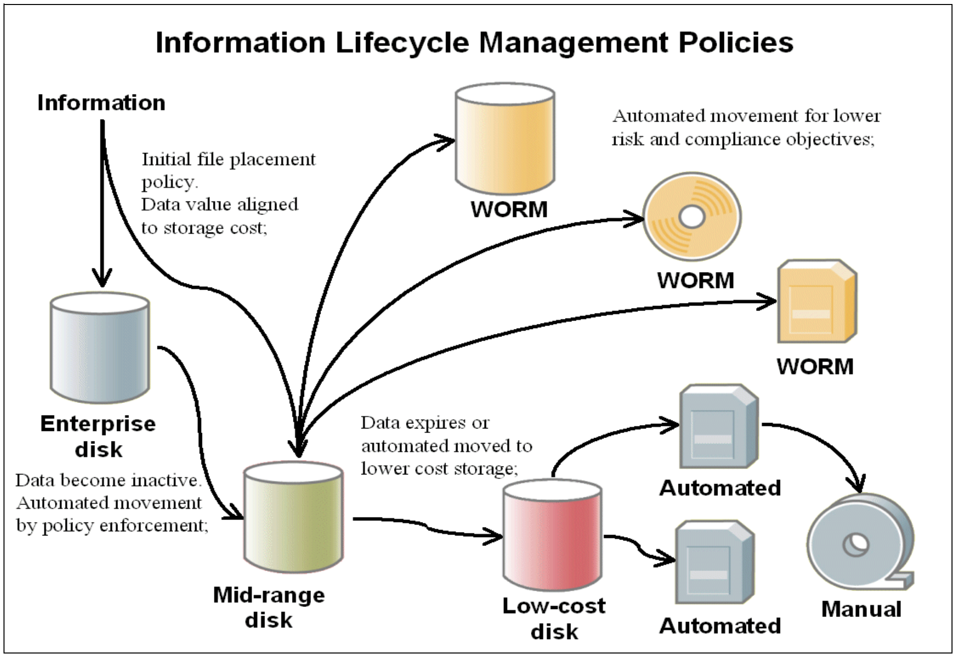

Information Lifecycle Management Policies
Information
Initial file placement policy.
Data value aligned to storage cost;
WORM
Automated movement for lower risk and compliance objectives;
WORM
WORM
Enterprise disk
Data become inactive.
Automated movement by policy enforcement;
Data expires or automated moved to lower cost storage;
Automated
Mid-range disk
Low-cost disk
Automated
Manual

Figure 10-15 ILM policies

Key products of IBM for lifecycle management are:

- IBM TotalStorage Productivity Center
- IBM TotalStorage SAN Volume Controller (SVC)
- IBM Tivoli Storage Manager including IBM System Storage Archive Manager
- IBM Tivoli Storage Manager for Space Management

10.5.5 Policy-based archive management

As businesses of all sizes migrate to e-business solutions and a new way of doing business, they already have mountains of data and content that have been captured, stored, and distributed across the enterprise. This wealth of information provides a unique opportunity. By incorporating these assets into e-business solutions, and at the same time delivering newly generated information media to their employees and clients, a business can reduce costs and information redundancy and leverage the potential profit-making aspects of their information assets.

Growth of information in corporate databases such as Enterprise Resource Planning (ERP) systems and e-mail systems makes organizations think about moving unused data off the high-cost disks. They need to:

- Identify database data that is no longer being regularly accessed and move it to an archive where it remains available.
- Define and manage what to archive, when to archive, and how to archive from the mail system or database system to the back-end archive management system.

Database archive solutions can help improve performance for online databases, reduce backup times, and improve application upgrade times.

E-mail archiving solutions are designed to reduce the size of corporate e-mail systems by moving e-mail attachments and/or messages to an archive from which they can easily be recovered if needed. This action helps reduce the need for end-user management of e-mail, improves the performance of e-mail systems, and supports the retention and deletion of e-mail.

The way to do this is to migrate and store all information assets into an e-business enabled content manager. ERP databases and e-mail solutions generate large volumes of information and data objects that can be stored in content management archives. An archive solution allows you to free system resources, while maintaining access to the stored objects for later reference. Allowing it to manage and migrate data objects gives a solution the ability to have ready access to newly created information that carries a higher value, while at the same time still being able to retrieve data that has been archived on less expensive media, as shown in Figure 10-16 on page 277.

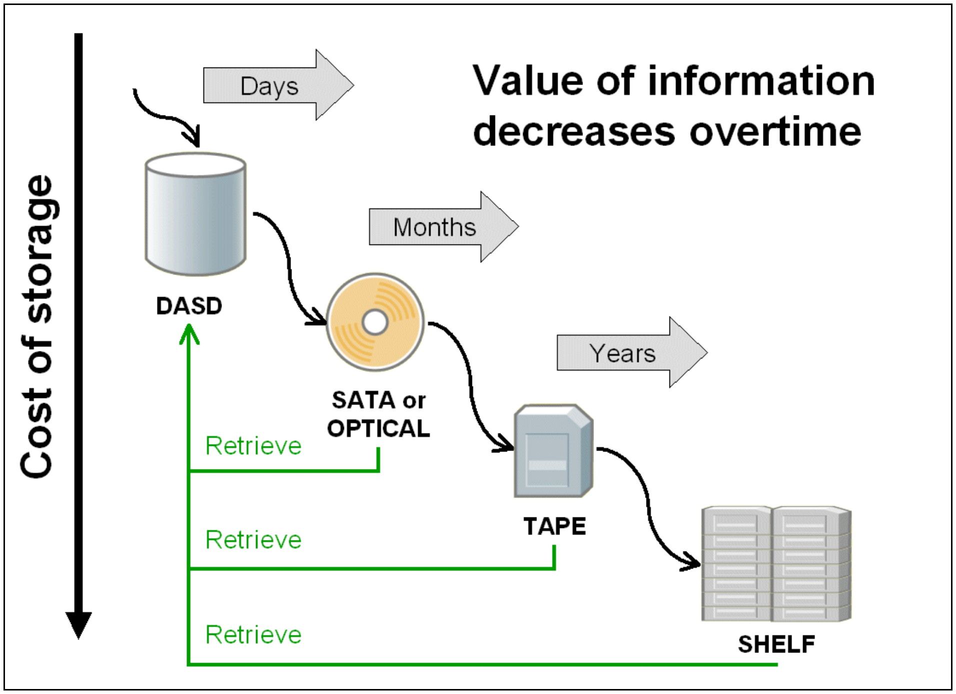


Figure 10-16 Value of information and archive/retrieve management

Key products of IBM for archive management are:

- IBM Tivoli Storage Manager including IBM System Storage Archive Manager
- IBM DB2 Content Manager family of products
- IBM DB2 CommonStore family of products

More information can be found at:

http://www-03.ibm.com/servers/storage/solutions/ilm/index.html

ILM is also addressed in the IBM Redbook:

- *ILM Library: Techniques with Tivoli Storage and IBM TotalStorage Products*, SG24-7030

A

SAN standards and organizations

The success and adoption of any new technology, and any improvement to existing technology, is greatly influenced by *standards*. Standards are the basis for the interoperability of hardware and software from different, and often rival, vendors. Although de facto standards bodies and organizations such as the Internet Engineering Task Force (IETF), American National Standards Institute (ANSI), and International Organization for Standardization (ISO) publish these formal standards, other organizations and industry associations, such as the Storage Networking Industry Association (SNIA) and the Fibre Channel Industry Association (FCIA), play a significant role in defining the standards and market development and direction.

> **Important:** Compliance with the standards does not guarantee interoperability, nor does interoperability guarantee compliance with the standards. It is up to each vendor *how* they implement the standards, and their interpretation of the guidelines into their product design. Two vendors may both comply with the standards, but may not interoperate. However, vendors spend a lot of time creating, testing, and ratifying multivendor solutions that they will support.

The major vendors in the SAN industry recognize the need for standards, especially in the areas of interoperability interfaces, management information base (MIB), application programming interfaces (API), Common Information Model (CIM), and so on, as these are significant for the basis for the wide acceptance of SANs. Standards such as these will allow customers a greater breadth of choice, and will lead to the deployment of cross-platform, mixed-protocol, multivendor, enterprise-wide SAN solutions. SAN technology has a number of industry associations and standard bodies evolving, developing, and publishing the SAN standards.

As you might expect, IBM actively participates in most of these organizations. The roles of these associations and bodies fall into three categories.

- Market development

  These associations are architecture development organizations that are formed early in the product life cycle, have a marketing focus, and perform the market development, gather the requirements, provide customer education, arrange user conferences, and so on. This includes organizations such as SNIA, FCIA, and the SCSI Trade Association (STA). Some of these organizations, such as SNIA, also help define the defacto industry standards, and thus have multiple roles.

- Defacto standards

  These organizations and bodies tend to be formed from two sources. They include working groups within the market development organizations, such as SNIA and the FCIA. Others are partnerships between groups of companies in the industry, such as Jini™ and Fibre Alliance, which work as pressure groups towards defacto industry standards.

  They offer architectural definitions, write white papers, arrange technical conferences, and may reference implementations based on developments by their own partner companies. They may submit these specifications for formal standards acceptance and approval.

- Formal standards

  These are the formal standards organizations such as the IETF, IEEE, and ANSI, which are in place to review for approval, and publish standards defined and submitted by the preceding two categories of organizations.

A number of industry associations, alliances, consortium, and formal standards bodies are involved in the SAN standards; these include SNIA, FCIA, STA, INCITS, IETF, ANSI, and IEEE. A brief description of the roles of some of these organizations are described in the following topics.

Storage Networking Industry Association

The Storage Networking Industry Association (SNIA) is an international computer system industry forum of developers, integrators, and IT professionals who evolve and promote storage networking technology and solutions. SNIA was formed to ensure that storage networks become efficient, complete, and trusted solutions across the IT community. IBM is one of the founding members of this organization. SNIA is uniquely committed to networking solutions into a broader market.

SNIA is using its Storage Management Initiative (SMI) and its Storage Management Initiative Specification (SMI-S) to create and promote adoption of a highly functional interoperable management interface for multivendor storage networking products. SMI-S makes multivendor storage networks simpler to implement, and easier to manage. IBM has led the industry in not only supporting the SMI-S initiative, but also using it across its hardware and software product lines. The specification covers fundamental operations of communications between management console clients and devices, auto-discovery, access, security, the ability to provision volumes and disk resources, LUN mapping and masking, and other management operations.

For additional information about the various activities of SNIA, see its Web site:

http://www.snia.org

Fibre Channel Industry Association

The Fibre Channel Industry Association (FCIA) is organized as a not-for-profit, mutual benefit corporation. The FCIA mission is to nurture and help develop the broadest market for Fibre Channel products. This is done through market development, education, standards monitoring, and fostering interoperability among members' products. IBM is a board member in the FCIA.

The FCIA also administers the SANmark, SANmark Qualified Test Provider programs. SANmark is a certification process designed to ensure that Fibre Channel devices, such as HBAs and switches, conform to Fibre Channel standards. The SANmark Qualified Test Provider program was established to increase the available pool of knowledgeable test providers for equipment vendors.

For additional information about the various activities of the FCIA, visit the Web site:

http://www.fibrechannel.org

For more information about SANmark, visit the Web site:

http://www.sanmark.org/

SCSI Trade Association

The SCSI Trade Association (STA) was formed to promote the use and understanding of the small computer system interface (SCSI) parallel interface technology. The STA provides a focal point for communicating SCSI benefits to the market, and influences the evolution of SCSI into the future. IBM is one of the founding members of STA. As part of its current work, and as part of its roadmap, STA has Serial Attached SCSI defined as the logical evolution of SCSI.

For additional information visit the Web site:

http://www.scsita.org

International Committee for Information Technology Standards

The International Committee for Information Technology Standards (INCITS) is the primary US focus of standardization in the field of information and communication technologies (ICT), encompassing storage, processing, transfer, display, management, organization, and retrieval of information. As such, INCITS also serves as ANSI's Technology Advisory Group for ISO/IEC Joint Technical Committee 1. JTC 1 is responsible for international standardization in the field of Information Technology.

For storage management, the draft standard defines a method for the interoperable management of heterogeneous SANs, describes an object-oriented, XML-based, messaging-based interface designed to support the specific requirements of managing devices in and through SANs.

For additional information visit the Web site:

http://www.incits.org

INCITS Technical Committee T11

The Technical Committee T11 is the committee within INCITS responsible for Device Level Interfaces. T11 has been producing interface standards for high-performance and mass storage applications since the 1970s. At the time of writing, the T11 program of work includes two current and three complete standards development projects.

Proposals for Fibre Channel transport, topology, generic services, and physical and media standards is available at Web site:

http://www.t11.org

Information Storage Industry Consortium

The Information Storage Industry Consortium (INSIC) is the research consortium for the worldwide information storage industry, whose mission is to enhance the growth and technical vitality of the information storage industry, and to advance the state of information storage technology.

INSIC membership consists of more than 65 corporations, universities, and government organizations with common interests in the field of digital information storage. IBM is a founding member of INSIC. For more information, visit the Web site:

http://www.insic.org

Internet Engineering Task Force

The Internet Engineering Task Force (IETF) is a large, open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture, and the smooth operation of the Internet. An IETF working group, the Internet and Management Support for Storage, is chartered to address the areas of IPv4 over Fibre Channel, an initial Fibre Channel Management MIB, other storage-related MIBs for other storage transports such as INCITS T10 Serial Attached SCSI (SAS), or SCSI command-specific commands. In addition, they are also responsible for iSCSI. An IBMer is the current chairman of IETF.

For additional information about the IETF, visit the Web site:

http://www.ietf.org

American National Standards Institute

The American National Standards Institute (ANSI) does not itself develop American national standards. Its mission is to enhance both the global competitiveness of U.S. business and the U.S. quality of life by promoting and facilitating voluntary consensus standards and conformity assessment systems, and safeguarding their integrity. It does this by working closely with organizations such as ISO.

It facilitates development by establishing consensus among qualified groups. IBM participates in numerous committees, including those for Fibre Channel and SANs. For more information about ANSI, visit the Web site:

http://www.ansi.org

Institute of Electrical and Electronics Engineers

The Institute of Electrical and Electronics Engineers (IEEE) is a non-profit, technical professional association of more than 365,000 individual members in 175 countries. Through its members, the IEEE is a leading authority in technical areas ranging from computer engineering, biomedical technology, and telecommunications, to electric power, aerospace, and consumer electronics, among others. It administers the Organizational Unique Identifier (OUI) list used for the addressing scheme within Fibre Channel.

For additional information about IEEE, visit the Web site:

http://ww.ieee.org

For additional information about IEEE and its standards work, visit the Web site:

http://standards.ieee.org/

Distributed Management Task Force

With more than 3,000 active participants, the Distributed Management Task Force, Inc. (DMTF) is the industry organization leading the development of management standards and integration technology for enterprise and Internet environments. DMTF standards provide common management infrastructure components for instrumentation, control, and communication in a platform-independent and technology-neutral way. DMTF technologies include common information models (CIM), communication/control protocols (WBEM), and core management services/utilities.

For more information visit the Web site:

http://www.dmtf.org/

Glossary

8b/10b A data-encoding scheme developed by IBM, translating byte-wide data to an encoded 10-bit format. Fibre Channel's FC-1 level defines this as the method to be used to encode and decode data transmissions over the Fibre Channel.

active configuration In an ESCON environment, the ESCON Director configuration determined by the status of the current set of connectivity attributes. Contrast with *saved configuration*.

Adapter A hardware unit that aggregates other I/O units, devices, or communications links to a system bus.

ADSM ADSTAR Distributed Storage Manager.

Agent (1) In the client-server model, the part of the system that performs information preparation and exchange on behalf of a client or server application. (2) In SNMP, the word agent refers to the managed system. See *Management Agent*.

Aggregation In the Storage Networking Industry Association Storage Model (SNIA), *virtualization* is known as *aggregation*. This aggregation can take place at the file level or at the level of individual blocks that are transferred to disk.

AIT Advanced Intelligent Tape - A magnetic tape format by Sony that uses 8 mm cassettes, but is only used in specific drives.

AL See *Arbitrated Loop*.

allowed In an ESCON Director, the attribute that, when set, establishes dynamic connectivity capability. Contrast with *prohibited*.

AL_PA Arbitrated Loop Physical Address.

ANSI American National Standards Institute - The primary organization for fostering the development of technology standards in the United States. The ANSI family of Fibre Channel documents provide the standards basis for the Fibre Channel architecture and technology. See *FC-PH*.

APAR See *authorized program analysis report*.

authorized program analysis report (APAR) A report of a problem caused by a suspected defect in a current, unaltered release of a program.

Arbitration The process of selecting one respondent from a collection of several candidates that request service concurrently.

Arbitrated Loop A Fibre Channel interconnection technology that allows up to 126 participating node ports and one participating fabric port to communicate.

ATL Automated Tape Library - Large scale tape storage system, which uses multiple tape drives and mechanisms to address 50 or more cassettes.

ATM Asynchronous Transfer Mode - A type of packet switching that transmits fixed-length units of data.

Backup A copy of computer data that is used to recreate data that has been lost, mislaid, corrupted, or erased. The act of creating a copy of computer data that can be used to recreate data that has been lost, mislaid, corrupted or erased.

Bandwidth Measure of the information capacity of a transmission channel.

basic mode A S/390 or zSeries central processing mode that does not use logical partitioning. Contrast with *logically partitioned (LPAR) mode*.

blocked In an ESCON and FICON Director, the attribute that, when set, removes the communication capability of a specific port. Contrast with *unblocked*.

Bridge (1) A component used to attach more than one I/O unit to a port. (2) A data communications device that connects two or more networks and forwards packets between them. The bridge may use similar or dissimilar media and signaling systems. It operates at the data link level of the OSI model. Bridges read and filter data packets and frames.

Bridge/Router A device that can provide the functions of a bridge, router or both concurrently. A bridge/router can route one or more protocols, such as TCP/IP, and bridge all other traffic. See also *Bridge, Router*.

Broadcast Sending a transmission to all N_Ports on a fabric.

byte. (1) In fibre channel, an eight-bit entity prior to encoding or after decoding, with its least significant bit denoted as bit 0, and most significant bit as bit 7. The most significant bit is shown on the left side in FC-FS unless otherwise shown. (2) In S/390 architecture or zSeries z/Architecture™ (and FICON), an eight-bit entity prior to encoding or after decoding, with its least significant bit denoted as bit 7, and most significant bit as bit 0. The most significant bit is shown on the left side in S/390 architecture and zSeries z/Architecture.

Cascaded switches The connecting of one Fibre Channel switch to another Fibre Channel switch, thereby creating a cascaded switch route between two N_Nodes connected to a fibre channel fabric.

chained In an ESCON environment, pertaining to the physical attachment of two ESCON Directors (ESCDs) to each other.

channel (1) A processor system element that controls one channel path, whose mode of operation depends on the type of hardware to which it is attached. In a channel subsystem, each channel controls an I/O interface between the channel control element and the logically attached control units. (2) In the ESA/390 or zSeries architecture (z/Architecture), the part of a channel subsystem that manages a single I/O interface between a channel subsystem and a set of controllers (control units).

channel I/O A form of I/O where request and response correlation is maintained through some form of source, destination and request identification.

channel path (CHP) A single interface between a central processor and one or more control units along which signals and data can be sent to perform I/O requests.

channel path identifier (CHPID) In a channel subsystem, a value assigned to each installed channel path of the system that uniquely identifies that path to the system.

channel subsystem (CSS) Relieves the processor of direct I/O communication tasks, and performs path management functions. Uses a collection of subchannels to direct a channel to control the flow of information between I/O devices and main storage.

channel-attached (1) Pertaining to attachment of devices directly by data channels (I/O channels) to a computer. (2) Pertaining to devices attached to a controlling unit by cables rather than by telecommunication lines.

CHPID Channel path identifier.

CIFS Common Internet File System.

cladding. In an optical cable, the region of low refractive index surrounding the core. See also *core* and *optical fiber.*

Class of Service A Fibre Channel frame delivery scheme exhibiting a specified set of delivery characteristics and attributes.

Class-1 A class of service providing dedicated connection between two ports with confirmed delivery or notification of non-deliverability.

Class-2 A class of service providing a frame switching service between two ports with confirmed delivery or notification of non-deliverability.

Class-3 A class of service providing frame switching datagram service between two ports or a multicast service between a multicast originator and one or more multicast recipients.

Class-4 A class of service providing a fractional bandwidth virtual circuit between two ports with confirmed delivery or notification of non-deliverability.

Class-6 A class of service providing a multicast connection between a multicast originator and one or more multicast recipients with confirmed delivery or notification of non-deliverability.

Client A software program used to contact and obtain data from a *server* software program on another computer -- often across a great distance. Each *client* program is designed to work specifically with one or more kinds of server programs and each server requires a specific kind of client program.

Client/Server The relationship between machines in a communications network. The client is the requesting machine, the server the supplying machine. Also used to describe the information management relationship between software components in a processing system.

Cluster A type of parallel or distributed system that consists of a collection of interconnected whole computers and is used as a single, unified computing resource.

CNC Mnemonic for an ESCON channel used to communicate to an ESCON-capable device.

configuration matrix In an ESCON environment or FICON, an array of connectivity attributes that appear as rows and columns on a display device and can be used to determine or change active and saved ESCON or FICON director configurations.

connected In an ESCON Director, the attribute that, when set, establishes a dedicated connection between two ESCON ports. Contrast with *disconnected.*

connection In an ESCON Director, an association established between two ports that provides a physical communication path between them.

connectivity attribute In an ESCON and FICON Director, the characteristic that determines a particular element of a port's status. See *allowed, prohibited, blocked, unblocked, (connected and disconnected)*.

control unit A hardware unit that controls the reading, writing, or displaying of data at one or more input/output units.

Controller A component that attaches to the system topology through a channel semantic protocol that includes some form of request/response identification.

core (1) In an optical cable, the central region of an optical fiber through which light is transmitted. (2) In an optical cable, the central region of an optical fiber that has an index of refraction greater than the surrounding cladding material. See also *cladding* and *optical fiber*.

coupler In an ESCON environment, link hardware used to join optical fiber connectors of the same type. Contrast with *adapter*.

Coaxial Cable A transmission media (cable) used for high speed transmission. It is called *coaxial* because it includes one physical channel that carries the signal surrounded (after a layer of insulation) by another concentric physical channel, both of which run along the same axis. The inner channel carries the signal and the outer channel serves as a ground.

CRC Cyclic Redundancy Check - An error-correcting code used in Fibre Channel.

CTC (1) Channel-to-channel. (2) Mnemonic for an ESCON channel attached to another ESCON channel, where one of the two ESCON channels is defined as an ESCON CTC channel and the other ESCON channel would be defined as a ESCON CNC channel (3) Mnemonic for a FICON channel supporting a CTC Control Unit function logically or physically connected to another FICON channel that also supports a CTC Control Unit function. FICON channels supporting the FICON CTC control unit function are defined as normal FICON native (FC) mode channels.

CVC Mnemonic for an ESCON channel attached to an IBM 9034 convertor. The 9034 converts from ESCON CVC signals to parallel channel interface (OEMI) communication operating in block multiplex mode (Bus and Tag). Contrast with *CBY*.

DASD Direct Access Storage Device - any online storage device: a disc, drive or CD-ROM.

DAT Digital Audio Tape - A tape media technology designed for very high quality audio recording and data backup. DAT cartridges look like audio cassettes and are often used in mechanical auto-loaders. typically, a DAT cartridge provides 2 GB of storage. But new DAT systems have much larger capacities.

Data Sharing A SAN solution in which files on a storage device are shared between multiple hosts.

Datagram Refers to the Class 3 Fibre Channel Service that allows data to be sent rapidly to multiple devices attached to the fabric, with no confirmation of delivery.

DDM See *disk drive module.*

dedicated connection In an ESCON Director, a connection between two ports that is not affected by information contained in the transmission frames. This connection, which restricts those ports from communicating with any other port, can be established or removed only as a result of actions performed by a host control program or at the ESCD console. Contrast with *dynamic connection*.

Note: The two links having a dedicated connection appear as one continuous link.

default Pertaining to an attribute, value, or option that is assumed when none is explicitly specified.

Dense Wavelength Division Multiplexing (DWDM) The concept of packing multiple signals tightly together in separate groups, and transmitting them simultaneously over a common carrier wave.

destination Any point or location, such as a node, station, or a particular terminal, to which information is to be sent. An example is a Fibre Channel fabric F_Port; when attached to a fibre channel N_port, communication to the N_port via the F_port is said to be to the F_Port destination identifier (D_ID).

device A mechanical, electrical, or electronic contrivance with a specific purpose.

device address (1) In ESA/390 architecture and zSeries z/Architecture, the field of an ESCON device-level frame that selects a specific device on a control unit image. (2) In the FICON channel FC-SB-2 architecture, the device address field in an SB-2 header that is used to select a specific device on a control unit image.

device number (1) In ESA/390 and zSeries z/Architecture, a four-hexadecimal character identifier (for example, 19A0) that you associate with a device to facilitate communication between the program and the host operator. (2) The device number that you associate with a subchannel that uniquely identifies an I/O device.

dB Decibel - a ratio measurement distinguishing the percentage of signal attenuation between the input and output power. Attenuation (loss) is expressed as dB/km.

direct access storage device (DASD) A mass storage medium on which a computer stores data.

disconnected In an ESCON Director, the attribute that, when set, removes a dedicated connection. Contrast with *connected*.

disk A mass storage medium on which a computer stores data.

disk drive module (DDM) A disk storage medium that you use for any host data that is stored within a disk subsystem.

Disk Mirroring A fault-tolerant technique that writes data simultaneously to two hard disks using the same hard disk controller.

Disk Pooling A SAN solution in which disk storage resources are pooled across multiple hosts rather than be dedicated to a specific host.

distribution panel (1) In an ESCON and FICON environment, a panel that provides a central location for the attachment of trunk and jumper cables and can be mounted in a rack, wiring closet, or on a wall.

DLT Digital Linear Tape - A magnetic tape technology originally developed by Digital Equipment Corporation (DEC) and now sold by Quantum. DLT cartridges provide storage capacities from 10 to 35 GB.

duplex Pertaining to communication in which data or control information can be sent and received at the same time, from the same node. Contrast with *half duplex.*

duplex connector In an ESCON environment, an optical fiber component that terminates both jumper cable fibers in one housing and provides physical keying for attachment to a duplex receptacle.

duplex receptacle In an ESCON environment, a fixed or stationary optical fiber component that provides a keyed attachment method for a duplex connector.

dynamic connection In an ESCON Director, a connection between two ports, established or removed by the ESCD and that, when active, appears as one continuous link. The duration of the connection depends on the protocol defined for the frames transmitted through the ports and on the state of the ports. Contrast with *dedicated connection.*

dynamic connectivity In an ESCON Director, the capability that allows connections to be established and removed at any time.

Dynamic I/O Reconfiguration A S/390 and z/Architecture function that allows I/O configuration changes to be made nondisruptively to the current operating I/O configuration.

ECL Emitter Coupled Logic - The type of transmitter used to drive copper media such as Twinax, Shielded Twisted Pair, or Coax.

ELS See *Extended Link Services.*

EMIF See *ESCON Multiple Image Facility.*

E_Port Expansion Port - a port on a switch used to link multiple switches together into a Fibre Channel switch fabric.

Enterprise Network A geographically dispersed network under the auspices of one organization.

Enterprise System Connection (ESCON) (1) An ESA/390 computer peripheral interface. The I/O interface uses ESA/390 logical protocols over a serial interface that configures attached units to a communication fabric. (2) A set of IBM products and services that provide a dynamically connected environment within an enterprise.

Enterprise Systems Architecture/390® (ESA/390) An IBM architecture for mainframe computers and peripherals. Processors that follow this architecture include the S/390 Server family of processors.

Entity In general, a real or existing thing from the Latin ens, or being, which makes the distinction between a thing's existence and it qualities. In programming, engineering and probably many other contexts, the word is used to identify units, whether concrete things or abstract ideas, that have no ready name or label.

ESA/390 See *Enterprise Systems Architecture/390*.

ESCD Enterprise Systems Connection (ESCON) Director.

ESCD console The ESCON Director display and keyboard device used to perform operator and service tasks at the ESCD.

ESCON See *Enterprise System Connection.*

ESCON channel A channel having an Enterprise Systems Connection channel-to-control-unit I/O interface that uses optical cables as a transmission medium. May operate in CBY, CNC, CTC or CVC mode. Contrast with *parallel channel.*

ESCON Director An I/O interface switch that provides the interconnection capability of multiple ESCON interfaces (or FICON Bridge (FCV) mode - 9032-5) in a distributed-star topology.

ESCON Multiple Image Facility (EMIF) In the ESA/390 architecture and zSeries z/Architecture, a function that allows LPARs to share an ESCON and FICON channel path (and other channel types) by providing each LPAR with its own channel-subsystem image.

Extended Link Services (ELS) An Extended Link Service (command) request solicits a destination port (N_Port or F_Port) to perform a function or service. Each ELS request consists of an Link Service (LS) command; the N_Port ELS commands are defined in the FC-FS architecture.

Exchange A group of sequences which share a unique identifier. All sequences within a given exchange use the same protocol. Frames from multiple sequences can be multiplexed to prevent a single exchange from consuming all the bandwidth. See also *Sequence*.

F_Node Fabric Node - a fabric attached node.

F_Port Fabric Port - a port used to attach a Node Port (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 (1) (Fibre Channel), a short form when referring to something that is part of the fibre channel standard. (2) Also used by the IBM I/O definition process when defining a FICON channel (using IOCP of HCD) that will be used in FICON native mode (using the FC-SB-2 communication protocol).

FC-FS Fibre Channel-Framing and Signaling, the term used to describe the FC-FS architecture.

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 Fibre 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. See also *Arbitrated Loop*.

FC-CT Fibre Channel common transport protocol.

FC-FG Fibre Channel Fabric Generic - A reference to the document (ANSI X3.289-1996) which defines the concepts, behavior and characteristics of the Fibre Channel Fabric along with suggested partitioning of the 24-bit address space to facilitate the routing of frames.

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 X.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.

FC Storage Director See *SAN Storage Director*.

FCA Fibre Channel Association - a Fibre Channel industry association that works to promote awareness and understanding of the Fibre Channel technology and its application and provides a means for implementers to support the standards committee activities.

FCLC Fibre Channel Loop Association - an independent working group of the Fibre Channel Association focused on the marketing aspects of the Fibre Channel Loop technology.

FCP Fibre Channel Protocol - the mapping of SCSI-3 operations to Fibre Channel.

FCS See *fibre channel standard*.

fiber See *optical fiber*.

fiber optic cable See *optical cable.*

fiber optics The branch of optical technology concerned with the transmission of radiant power through fibers made of transparent materials such as glass, fused silica, and plastic.

Note: Telecommunication applications of fiber optics use optical fibers. Either a single discrete fiber or a non-spatially aligned fiber bundle can be used for each information channel. Such fibers are often called “optical fibers” to differentiate them from fibers used in non-communication applications.

Fibre Channel A technology for transmitting data between computer devices at a data rate of up to 4 Gbps. It is especially suited for connecting computer servers to shared storage devices and for interconnecting storage controllers and drives.

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 four layers. The lower of the four layers defines the physical media and interface, the upper of the four layers defines one or more Upper Layer Protocols (ULP)—for example, FCP for SCSI command protocols and FC-SB-2 for FICON protocol supported by ESA/390 and z/Architecture. Refer to ANSI X3.230.1999x.

FICON (1) An ESA/390 and zSeries computer peripheral interface. The I/O interface uses ESA/390 and zSeries FICON protocols (FC-FS and FC-SB-2) over a Fibre Channel serial interface that configures attached units to a FICON supported Fibre Channel communication fabric. (2) An FC4 proposed standard that defines an effective mechanism for the export of the SBCCS-2 (FC-SB-2) command protocol via fibre channels.

FICON channel A channel having a Fibre Channel connection (FICON) channel-to-control-unit I/O interface that uses optical cables as a transmission medium. May operate in either FC or FCV mode.

FICON Director A Fibre Channel switch that supports the ESCON-like “control unit port” (CUP function) that is assigned a 24-bit FC port address to allow FC-SB-2 addressing of the CUP function to perform command and data transfer (in the FC world, it is a means of in-band management using a FC-4 ULP).

field replaceable unit (FRU) An assembly that is replaced in its entirety when any one of its required components fails.

FL_Port Fabric Loop Port - the access point of the fabric for physically connecting the user's Node Loop Port (NL_Port).

FLOGI See *Fabric Log In*.

Frame A linear set of transmitted bits that define the basic transport unit. The frame is the most basic element of a message in Fibre Channel communications, consisting of a 24-byte header and zero to 2112 bytes of data. See also *Sequence*.

FRU See *field replaceable unit*.

FSP Fibre Channel Service Protocol - The common FC-4 level protocol for all services, transparent to the fabric type or topology.

FSPF Fabric shortest path first - is an intelligent path selection and routing standard and is part of the Fibre Channel Protocol.

Full-Duplex A mode of communications allowing simultaneous transmission and reception of frames.

G_Port Generic Port - a generic switch port that is either a Fabric Port (F_Port) or an Expansion Port (E_Port). The function is automatically determined during login.

Gateway A node on a network that interconnects two otherwise incompatible networks.

Gbps Gigabits per second. Also sometimes referred to as Gbps. In computing terms it is approximately 1,000,000,000 bits per second. Most precisely it is 1,073,741,824 (1024 x 1024 x 1024) bits per second.

GBps Gigabytes per second. Also sometimes referred to as GBps. In computing terms it is approximately 1,000,000,000 bytes per second. Most precisely it is 1,073,741,824 (1024 x 1024 x 1024) bytes per second.

GBIC Gigabit interface converter - Industry standard transceivers for connection of Fibre Channel nodes to arbitrated loop hubs and fabric switches.

Gigabit One billion bits, or one thousand megabits.

GLM Gigabit Link Module - a generic Fibre Channel transceiver unit that integrates the key functions necessary for installation of a Fibre channel media interface on most systems.

half duplex In data communication, pertaining to transmission in only one direction at a time. Contrast with *duplex.*

hard disk drive (1) A storage media within a storage server used to maintain information that the storage server requires. (2) A mass storage medium for computers that is typically available as a fixed disk or a removable cartridge.

Hardware The mechanical, magnetic and electronic components of a system, e.g., computers, telephone switches, terminals and the like.

HBA Host Bus Adapter.

HCD Hardware Configuration Dialog.

HDA Head and disk assembly.

HDD See *hard disk drive.*

head and disk assembly The portion of an HDD associated with the medium and the read/write head.

HIPPI High Performance Parallel Interface - An ANSI standard defining a channel that transfers data between CPUs and from a CPU to disk arrays and other peripherals.

HMMP HyperMedia Management Protocol.

HMMS HyperMedia Management Schema - the definition of an implementation-independent, extensible, common data description/schema allowing data from a variety of sources to be described and accessed in real time regardless of the source of the data. See also *WEBM, HMMP*.

hop A FC frame may travel from a switch to a director, a switch to a switch, or director to a director which, in this case, is one hop.

HSM Hierarchical Storage Management - A software and hardware system that moves files from disk to slower, less expensive storage media based on rules and observation of file activity. Modern HSM systems move files from magnetic disk to optical disk to magnetic tape.

HUB A Fibre Channel device that connects nodes into a logical loop by using a physical star topology. Hubs will automatically recognize an active node and insert the node into the loop. A node that fails or is powered off is automatically removed from the loop.

HUB Topology See *Loop Topology*.

Hunt Group A set of associated Node Ports (N_Ports) attached to a single node, assigned a special identifier that allows any frames containing this identifier to be routed to any available Node Port (N_Port) in the set.

ID See *identifier.*

identifier A unique name or address that identifies things such as programs, devices or systems.

In-band Signaling This is signaling that is carried in the same channel as the information. Also referred to as in-band.

In-band virtualization An implementation in which the virtualization process takes place in the data path between servers and disk systems. The virtualization can be implemented as software running on servers or in dedicated engines.

Information Unit A unit of information defined by an FC-4 mapping. Information Units are transferred as a Fibre Channel Sequence.

initial program load (IPL) (1) The initialization procedure that causes an operating system to commence operation. (2) The process by which a configuration image is loaded into storage at the beginning of a work day, or after a system malfunction. (3) The process of loading system programs and preparing a system to run jobs.

input/output (I/O) (1) Pertaining to a device whose parts can perform an input process and an output process at the same time. (2) Pertaining to a functional unit or channel involved in an input process, output process, or both, concurrently or not, and to the data involved in such a process. (3) Pertaining to input, output, or both.

input/output configuration data set (IOCDS) The data set in the S/390 and zSeries processor (in the support element) that contains an I/O configuration definition built by the input/output configuration program (IOCP).

input/output configuration program (IOCP) A S/390 program that defines to a system the channels, I/O devices, paths to the I/O devices, and the addresses of the I/O devices.The output is normally written to a S/390 or zSeries IOCDS*.*

interface (1) A shared boundary between two functional units, defined by functional characteristics, signal characteristics, or other characteristics as appropriate. The concept includes the specification of the connection of two devices having different functions. (2) Hardware, software, or both, that links systems, programs, or devices.

Intermix A mode of service defined by Fibre Channel that reserves the full Fibre Channel bandwidth for a dedicated Class 1 connection, but also allows connection-less Class 2 traffic to share the link if the bandwidth is available.

inter-switch link A FC connection between switches and/or directors. Also known as ISL.

I/O See *input/output*.

I/O configuration The collection of channel paths, control units, and I/O devices that attaches to the processor. This may also include channel switches (for example, an ESCON Director).

IOCDS See *Input/Output configuration data set.*

IOCP See *Input/Output configuration control program.*

IODF The data set that contains the S/390 or zSeries I/O configuration definition file produced during the defining of the S/390 or zSeries I/O configuration by HCD. Used as a source for IPL, IOCP and Dynamic I/O Reconfiguration.

IPL See *initial program load*.

I/O Input/output.

IP Internet Protocol.

IPI Intelligent Peripheral Interface.

ISL See *inter-switch link.*

Isochronous Transmission Data transmission which supports network-wide timing requirements. A typical application for isochronous transmission is a broadcast environment which needs information to be delivered at a predictable time.

JBOD Just a bunch of disks.

Jukebox A device that holds multiple optical disks and one or more disk drives, and can swap disks in and out of the drive as needed.

jumper cable In an ESCON and FICON environment, an optical cable having two conductors that provides physical attachment between a channel and a distribution panel or an ESCON/FICON Director port or a control unit/device, or between an ESCON/FICON Director port and a distribution panel or a control unit/device, or between a control unit/device and a distribution panel. Contrast with *trunk cable.*

laser A device that produces optical radiation using a population inversion to provide *light amplification by stimulated emission of radiation* and (generally) an optical resonant cavity to provide positive feedback. Laser radiation can be highly coherent temporally, or spatially, or both.

L_Port Loop Port - A node or fabric port capable of performing Arbitrated Loop functions and protocols. NL_Ports and FL_Ports are loop-capable ports.

LAN A network covering a relatively small geographic area (usually not larger than a floor or small building). Transmissions within a Local Area Network are mostly digital, carrying data among stations at rates usually above one megabit/s.

Latency A measurement of the time it takes to send a frame between two locations.

LC Lucent Connector. A registered trademark of Lucent Technologies.

LCU *See Logical Control Unit.*

LED See *light emitting diode*.

licensed internal code (LIC) Microcode that IBM does not sell as part of a machine, but instead, licenses it 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 alternate to hard-wire circuitry.

light-emitting diode (LED) A semiconductor chip that gives off visible or infrared light when activated. Contrast with *Laser*.

link (1) In an ESCON environment or FICON environment (fibre channel environment), the physical connection and transmission medium used between an optical transmitter and an optical receiver. A link consists of two conductors, one used for sending and the other for receiving, thereby providing a duplex communication path. (2) In an ESCON I/O interface, the physical connection and transmission medium used between a channel and a control unit, a channel and an ESCD, a control unit and an ESCD, or, at times, between two ESCDs. (3) In a FICON I/O interface, the physical connection and transmission medium used between a channel and a control unit, a channel and a FICON Director, a control unit and a fibre channel FICON Director, or, at times, between two fibre channels switches.

link address (1) 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. See also *port address.* (2) On a FICON interface, the port address (1-byte link address), or domain and port address (2-byte link address) portion of a source (S_ID) or destination address (D_ID) in a fibre channel frame that the fibre channel switch uses to route a frame through a fibre channel switch or fibre channel switch fabric. See also *port address.*

Link_Control_Facility A termination card that handles the logical and physical control of the Fibre Channel link for each mode of use.

LIP A Loop Initialization Primitive sequence is a special Fibre Channel sequence that is used to start loop initialization. Allows ports to establish their port addresses.

local area network (LAN) A computer network located in a user’s premises within a limited geographic area.

logical control unit (LCU) A separately addressable control unit function within a physical control unit. Usually a physical control unit that supports several LCUs. For ESCON, the maximum number of LCUs that can be in a control unit (and addressed from the same ESCON fiber link) is 16; they are addressed from x’0’ to x’F’. For FICON architecture, the maximum number of LCUs that can be in a control unit (and addressed from the same FICON fibre link) is 256; they are addressed from x’00’ to x’FF’. For both ESCON and FICON, the actual number supported, and the LCU address value, is both processor- and control unit implementation-dependent.

logical partition (LPAR) A set of functions that create a programming environment that is defined by the ESA/390 architecture or zSeries z/Architecture. ESA/390 architecture or zSeries z/Architecture uses the term LPAR when more than one logical partition 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, LPAR does not depend on an operating system to create the virtual machine environment.

logical switch number (LSN) A two-digit number used by the I/O Configuration Program (IOCP) to identify a specific ESCON or FICON Director. (This number is separate from the director’s “switch device number” and, for FICON, it is separate from the director’s “FC switch address”).

l**ogically partitioned (LPAR) mode** A central processor mode, available on the Configuration frame when using the PR/SM™ facility, that allows an operator to allocate processor hardware resources among logical partitions. Contrast with *basic mode.*

Login Server Entity within the Fibre Channel fabric that receives and responds to login requests.

Loop Circuit A temporary point-to-point like path that allows bi-directional communications between loop-capable ports.

Loop Topology An interconnection structure in which each point has physical links to two neighbors resulting in a closed circuit. In a loop topology, the available bandwidth is shared.

LPAR See *logical partition.*

LVD Low Voltage Differential.

Management Agent A process that exchanges a managed node's information with a management station.

Managed Node A managed node is a computer, a storage system, a gateway, a media device such as a switch or hub, a control instrument, a software product such as an operating system or an accounting package, or a machine on a factory floor, such as a robot.

Managed Object A variable of a managed node. This variable contains one piece of information about the node. Each node can have several objects.

Management Station A host system that runs the management software.

MAR Media Access Rules. Enable systems to self-configure themselves is a SAN environment.

Mbps Megabits per second. Also sometimes referred to as Mbps. In computing terms it is approximately 1,000,000 bits per second. Most precisely it is 1,048,576 (1024 x 1024) bits per second.

MBps Megabytes per second. Also sometimes referred to as MBps. In computing terms it is approximately 1,000,000 bytes per second. Most precisely it is 1,048,576 (1024 x 1024) bytes per second.

Metadata server In Storage Tank™, servers that maintain information ("metadata") about the data files and grant permission for application servers to communicate directly with disk systems.

Meter 39.37 inches, or just slightly larger than a yard (36 inches).

Media Plural of medium. The physical environment through which transmission signals pass. Common media include copper and fiber optic cable.

Media Access Rules (MAR).

MIA Media Interface Adapter - MIAs enable optic-based adapters to interface to copper-based devices, including adapters, hubs, and switches.

MIB Management Information Block - A formal description of a set of network objects that can be managed using the Simple Network Management Protocol (SNMP). The format of the MIB is defined as part of SNMP and is a hierarchical structure of information relevant to a specific device, defined in object oriented terminology as a collection of objects, relations, and operations among objects.

Mirroring The process of writing data to two separate physical devices simultaneously.

MM Multi-Mode - See *Multi-Mode Fiber*.

MMF See *Multi-Mode Fiber* - In optical fiber technology, an optical fiber that is designed to carry multiple light rays or modes concurrently, each at a slightly different reflection angle within the optical core. Multi-Mode fiber transmission is used for relatively short distances because the modes tend to disperse over longer distances. See also *Single-Mode Fiber, SMF*.

Multicast Sending a copy of the same transmission from a single source device to multiple destination devices on a fabric. This includes sending to all N_Ports on a fabric (broadcast) or to only a subset of the N_Ports on a fabric (multicast).

Multi-Mode Fiber (MMF) In optical fiber technology, an optical fiber that is designed to carry multiple light rays or modes concurrently, each at a slightly different reflection angle within the optical core. Multi-Mode fiber transmission is used for relatively short distances because the modes tend to disperse over longer distances. See also *Single-Mode Fiber*.

Multiplex The ability to intersperse data from multiple sources and destinations onto a single transmission medium. Refers to delivering a single transmission to multiple destination Node Ports (N_Ports).

N_Port Node Port - A Fibre Channel-defined hardware entity at the end of a link which provides the mechanisms necessary to transport information units to or from another node.

N_Port Login N_Port Login (PLOGI) allows two N_Ports to establish a session and exchange identities and service parameters. It is performed following completion of the fabric login process and prior to the FC-4 level operations with the destination port. N_Port Login may be either explicit or implicit.

Name Server Provides translation from a given node name to one or more associated N_Port identifiers.

NAS Network Attached Storage - a term used to describe a technology where an integrated storage system is attached to a messaging network that uses common communications protocols, such as TCP/IP.

ND See *node descriptor*.

NDMP Network Data Management Protocol.

NED See *node-element descriptor.*

Network An aggregation of interconnected nodes, workstations, file servers, and/or peripherals, with its own protocol that supports interaction.

Network Topology Physical arrangement of nodes and interconnecting communications links in networks based on application requirements and geographical distribution of users.

NFS Network File System - A distributed file system in UNIX developed by Sun Microsystems™ which allows a set of computers to cooperatively access each other's files in a transparent manner.

NL_Port Node Loop Port - a node port that supports Arbitrated Loop devices.

NMS Network Management System - A system responsible for managing at least part of a network. NMSs communicate with agents to help keep track of network statistics and resources.

Node An entity with one or more N_Ports or NL_Ports.

node descriptor In an ESCON and FICON environment, a node descriptor (ND) is a 32-byte field that describes a node, channel, ESCON Director port or a FICON Director port, or a control unit.

node-element descriptor In an ESCON and FICON environment, a node-element descriptor (NED) is a 32-byte field that describes a node element, such as a disk (DASD) device.

Non-Blocking A term used to indicate that the capabilities of a switch are such that the total number of available transmission paths is equal to the number of ports. Therefore, all ports can have simultaneous access through the switch.

Non-L_Port A Node or Fabric port that is not capable of performing the Arbitrated Loop functions and protocols. N_Ports and F_Ports are not loop-capable ports.

OEMI See *original equipment manufacturers information.*

open system A system whose characteristics comply with standards made available throughout the industry and that therefore can be connected to other systems complying with the same standards.

Operation A term defined in FC-2 that refers to one of the Fibre Channel *building blocks* composed of one or more, possibly concurrent, exchanges.

optical cable A fiber, multiple fibers, or a fiber bundle in a structure built to meet optical, mechanical, and environmental specifications. See also *jumper cable, optical cable assembly*, and *trunk cable.*

optical cable assembly An optical cable that is connector-terminated. Generally, an optical cable that has been connector-terminated by a manufacturer and is ready for installation. See also *jumper cable* and *optical cable*.

optical fiber Any filament made of dialectic materials that guides light, regardless of its ability to send signals. See also *fiber optics* and *optical waveguide*.

optical fiber connector A hardware component that transfers optical power between two optical fibers or bundles and is designed to be repeatedly connected and disconnected.

optical waveguide (1) A structure capable of guiding optical power. (2) In optical communications, generally a fiber designed to transmit optical signals. See *optical fiber.*

Ordered Set A Fibre Channel term referring to four 10 -bit characters (a combination of data and special characters) providing low-level link functions, such as frame demarcation and signaling between two ends of a link.

original equipment manufacturer information (OEMI) A reference to an IBM guideline for a computer peripheral interface. More specifically, it refers to IBM S/360™ and S/370™ Channel to Control Unit Original Equipment Manufacturer Information. The interface uses ESA/390 logical protocols over an I/O interface that configures attached units in a multi-drop bus environment. This OEMI interface is also supported by the zSeries 900 processors.

Originator A Fibre Channel term referring to the initiating device.

Out of Band Signaling This is signaling that is separated from the channel carrying the information. Also referred to as out-of-band.

Out-of-band virtualization An alternative type of virtualization in which servers communicate directly with disk systems under control of a virtualization function that is not involved in the data transfer.

parallel channel A channel having a System/360™ and System/370™ channel-to-control-unit I/O interface that uses bus and tag cables as a transmission medium. Contrast with *ESCON channel*.

path In a channel or communication network, any route between any two nodes. For ESCON and FICON this would be the route between the channel and the control unit/device, or sometimes from the operating system control block for the device and the device itself.

path group The ESA/390 and zSeries architecture (z/Architecture) term for a set of channel paths that are defined to a controller as being associated with a single S/390 image. The channel paths are in a group state and are online to the host.

path-group identifier The ESA/390 and zSeries architecture (z/Architecture) term for the identifier that uniquely identifies a given LPAR. The path-group identifier is used in communication between the system image program and a device. The identifier associates the path-group with one or more channel paths, thereby defining these paths to the control unit as being associated with the same system image.

Peripheral Any computer device that is not part of the essential computer (the processor, memory and data paths) but is situated relatively close by. A near synonym is input/output (I/O) device.

Petard A device that is small and sometimes explosive.

PLDA Private Loop Direct Attach - A technical report which defines a subset of the relevant standards suitable for the operation of peripheral devices such as disks and tapes on a private loop.

PCICC (IBM) PCI Cryptographic Coprocessor.

PLOGI See *N_Port Login*.

Point-to-Point Topology An interconnection structure in which each point has physical links to only one neighbor resulting in a closed circuit. In point-to-point topology, the available bandwidth is dedicated.

Policy-based management Management of data on the basis of business policies (for example, "all production database data must be backed up every day"), rather than technological considerations (for example, "all data stored on this disk system is protected by remote copy").

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. (3) See also *duplex receptacle*.

port address (1) In an ESCON Director, an address used to specify port connectivity parameters and to assign link addresses for attached channels and control units. See also *link address.* (2) In a FICON director or Fibre Channel switch, it is the middle 8 bits of the full 24-bit FC port address. This field is also referred to as the “area field” in the 24-bit FC port address. See also *link address.*

Port Bypass Circuit A circuit used in hubs and disk enclosures to automatically open or close the loop to add or remove nodes on the loop.

port card In an ESCON and FICON environment, a field-replaceable hardware component that provides the optomechanical attachment method for jumper cables and performs specific device-dependent logic functions.

port name In an ESCON or FICON Director, a user-defined symbolic name of 24 characters or less that identifies a particular port.

Private NL_Port An NL_Port which does not attempt login with the fabric and only communicates with other NL Ports on the same loop.

processor complex A system configuration that consists of all the machines required for operation; for example, a processor unit, a processor controller, a system display, a service support display, and a power and coolant distribution unit.

program temporary fix (PTF) A temporary solution or bypass of a problem diagnosed by IBM in a current unaltered release of a program.

prohibited In an ESCON or FICON Director, the attribute that, when set, removes dynamic connectivity capability. Contrast with *allowed.*

protocol (1) A set of semantic and syntactic rules that determines the behavior of functional units in achieving communication. (2) In fibre channel, the meanings of and the sequencing rules for requests and responses used for managing the switch or switch fabric, transferring data, and synchronizing the states of fibre channel fabric components. (3) A specification for the format and relative timing of information exchanged between communicating parties.

PTF See *program temporary fix*.

Public NL_Port An NL_Port that attempts login with the fabric and can observe the rules of either public or private loop behavior. A public NL_Port may communicate with both private and public NL_Ports.

Quality of Service (QoS) A set of communications characteristics required by an application. Each QoS defines a specific transmission priority, level of route reliability, and security level.

Quick Loop is a unique fibre-channel topology that combines arbitrated loop and fabric topologies. It is an optional licensed product that allows arbitrated loops with private devices to be attached to a fabric.

RAID Redundant Array of Inexpensive or Independent Disks. A method of configuring multiple disk drives in a storage subsystem for high availability and high performance.

Raid 0 Level 0 RAID support - Striping, no redundancy.

Raid 1 Level 1 RAID support - mirroring, complete redundancy.

Raid 5 Level 5 RAID support, Striping with parity.

Repeater A device that receives a signal on an electromagnetic or optical transmission medium, amplifies the signal, and then retransmits it along the next leg of the medium.

Responder A Fibre Channel term referring to the answering device.

route The path that an ESCON frame takes from a channel through an ESCD to a control unit/device.

Router (1) A device that can decide which of several paths network traffic will follow based on some optimal metric. Routers forward packets from one network to another based on network-layer information. (2) A dedicated computer hardware and/or software package which manages the connection between two or more networks. See also *Bridge, Bridge/Router*.

SAF-TE SCSI Accessed Fault-Tolerant Enclosures.

SAN A storage area network (SAN) is a dedicated, centrally managed, secure information infrastructure, which enables any-to-any interconnection of servers and storage systems.

SAN System Area Network - term originally used to describe a particular symmetric multiprocessing (SMP) architecture in which a switched interconnect is used in place of a shared bus. Server Area Network - refers to a switched interconnect between multiple SMPs.

SANSymphony In-band block-level virtualization software made by DataCore Software Corporation and resold by IBM.

saved configuration In an ESCON or FICON Director environment, a stored set of connectivity attributes whose values determine a configuration that can be used to replace all or part of the ESCD's or FICON's active configuration. Contrast with *active configuration*.

SC Connector A fiber optic connector standardized by ANSI TIA/EIA-568A for use in structured wiring installations.

Scalability The ability of a computer application or product (hardware or software) to continue to function well as it (or its context) is changed in size or volume. For example, the ability to retain performance levels when adding additional processors, memory and/or storage.

SCSI Small Computer System Interface - A set of evolving ANSI standard electronic interfaces that allow personal computers to communicate with peripheral hardware such as disk drives, tape drives, CD_ROM drives, printers and scanners faster and more flexibly than previous interfaces. The table below identifies the major characteristics of the different SCSI version.

SCSI Version	Signal Rate MHz	BusWidth (bits)	Max. DTR (MBps)	Max. Num. Devices	Max. Cable Length (m)
SCSI-1	5	8	5	7	6
SCSI-2	5	8	5	7	6
Wide SCSI-2	5	16	10	15	6
Fast SCSI-2	10	8	10	7	6
Fast Wide SCSI-2	10	16	20	15	6
Ultra SCSI	20	8	20	7	1.5
Ultra SCSI-2	20	16	40	7	12
Ultra2 LVD SCSI	40	16	80	15	12

SCSI-3 SCSI-3 consists of a set of primary commands and additional specialized command sets to meet the needs of specific device types. The SCSI-3 command sets are used not only for the SCSI-3 parallel interface but for additional parallel and serial protocols, including Fibre Channel, Serial Bus Protocol (used with IEEE 1394 Firewire physical protocol) and the Serial Storage Protocol (SSP).

SCSI-FCP The term used to refer to the ANSI Fibre Channel Protocol for SCSI document (X3.269-199x) that describes the FC-4 protocol mappings and the definition of how the SCSI protocol and command set are transported using a Fibre Channel interface.

Sequence A series of frames strung together in numbered order which can be transmitted over a Fibre Channel connection as a single operation. See also *Exchange*.

service element (SE) A dedicated service processing unit used to service a S/390 machine (processor).

SERDES Serializer Deserializer.

Server A computer which is dedicated to one task.

SES SCSI Enclosure Services - ANSI SCSI-3 proposal that defines a command set for soliciting basic device status (temperature, fan speed, power supply status, etc.) from a storage enclosures.

Single-Mode Fiber In optical fiber technology, an optical fiber that is designed for the transmission of a single ray or mode of light as a carrier. It is a single light path used for long-distance signal transmission. See also *Multi-Mode Fiber*.

Small Computer System Interface (SCSI) (1) An ANSI standard for a logical interface to computer peripherals and for a computer peripheral interface. The interface uses an SCSI logical protocol over an I/O interface that configures attached targets and initiators in a multi-drop bus topology. (2) A standard hardware interface that enables a variety of peripheral devices to communicate with one another.

SMART Self Monitoring and Reporting Technology.

SM Single Mode - See *Single-Mode Fiber*.

SMF Single-Mode Fiber - In optical fiber technology, an optical fiber that is designed for the transmission of a single ray or mode of light as a carrier. It is a single light path used for long-distance signal transmission. See *MMF*.

SNIA Storage Networking Industry Association. A non-profit organization comprised of more than 77 companies and individuals in the storage industry.

SN Storage Network. Also see *SAN*.

SNMP Simple Network Management Protocol - The Internet network management protocol which provides a means to monitor and set network configuration and run-time parameters.

SNMWG Storage Network Management Working Group is chartered to identify, define and support open standards needed to address the increased management requirements imposed by storage area network environments.

SSA Serial Storage Architecture - A high speed serial loop-based interface developed as a high speed point-to-point connection for peripherals, particularly high speed storage arrays, RAID and CD-ROM storage by IBM.

Star The physical configuration used with hubs in which each user is connected by communications links radiating out of a central hub that handles all communications.

Storage Tank An IBM file aggregation project that enables a pool of storage, and even individual files, to be shared by servers of different types. In this way, Storage Tank can greatly improve storage utilization and enables data sharing.

StorWatch Expert These are StorWatch applications that employ a 3 tiered architecture that includes a management interface, a StorWatch manager and agents that run on the storage resource(s) being managed. Expert products employ a StorWatch data base that can be used for saving key management data (e.g. capacity or performance metrics). Expert products use the agents as well as analysis of storage data saved in the data base to perform higher value functions including -- reporting of capacity, performance, etc. over time (trends), configuration of multiple devices based on policies, monitoring of capacity and performance, automated responses to events or conditions, and storage related data mining.

StorWatch Specialist A StorWatch interface for managing an individual fibre Channel device or a limited number of like devices (that can be viewed as a single group). StorWatch specialists typically provide simple, point-in-time management functions such as configuration, reporting on asset and status information, simple device and event monitoring, and perhaps some service utilities.

Striping A method for achieving higher bandwidth using multiple N_Ports in parallel to transmit a single information unit across multiple levels.

STP Shielded Twisted Pair.

Storage Media The physical device itself, onto which data is recorded. Magnetic tape, optical disks, floppy disks are all storage media.

subchannel A logical function of a channel subsystem associated with the management of a single device.

subsystem (1) A secondary or subordinate system, or programming support, usually capable of operating independently of or asynchronously with a controlling system.

SWCH In ESCON Manager, the mnemonic used to represent an ESCON Director.

Switch A component with multiple entry/exit points (ports) that provides dynamic connection between any two of these points.

Switch Topology An interconnection structure in which any entry point can be dynamically connected to any exit point. In a switch topology, the available bandwidth is scalable.

T11 A technical committee of the National Committee for Information Technology Standards, titled T11 I/O Interfaces. It is tasked with developing standards for moving data in and out of computers.

Tape Backup Making magnetic tape copies of hard disk and optical disc files for disaster recovery.

Tape Pooling A SAN solution in which tape resources are pooled and shared across multiple hosts rather than being dedicated to a specific host.

TCP Transmission Control Protocol - a reliable, full duplex, connection-oriented end-to-end transport protocol running on top of IP.

TCP/IP Transmission Control Protocol/ Internet Protocol - a set of communications protocols that support peer-to-peer connectivity functions for both local and wide area networks.

Time Server A Fibre Channel-defined service function that allows for the management of all timers used within a Fibre Channel system.

Topology An interconnection scheme that allows multiple Fibre Channel ports to communicate. For example, point-to-point, Arbitrated Loop, and switched fabric are all Fibre Channel topologies.

T_Port An ISL port more commonly known as an E_Port, referred to as a Trunk port and used by INRANGE.

TL_Port A private to public bridging of switches or directors, referred to as Translative Loop.

trunk cable In an ESCON and FICON environment, a cable consisting of multiple fiber pairs that do not directly attach to an active device. This cable usually exists between distribution panels (or sometimes between a set processor channels and a distribution panel) and can be located within, or external to, a building. Contrast with *jumper cable*.

Twinax A transmission media (cable) consisting of two insulated central conducting leads of coaxial cable.

Twisted Pair A transmission media (cable) consisting of two insulated copper wires twisted around each other to reduce the induction (thus interference) from one wire to another. The twists, or lays, are varied in length to reduce the potential for signal interference between pairs. Several sets of twisted pair wires may be enclosed in a single cable. This is the most common type of transmission media.

ULP Upper Level Protocols.

unblocked In an ESCON and FICON Director, the attribute that, when set, establishes communication capability for a specific port. Contrast with *blocked.*

unit address The ESA/390 and zSeries term for the address associated with a device on a given controller. On ESCON and FICON 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.

UTC Under-The-Covers, a term used to characterize a subsystem in which a small number of hard drives are mounted inside a higher function unit. The power and cooling are obtained from the system unit. Connection is by parallel copper ribbon cable or pluggable backplane, using IDE or SCSI protocols.

UTP Unshielded Twisted Pair.

Virtual Circuit A unidirectional path between two communicating N_Ports that permits fractional bandwidth.

Virtualization An abstraction of storage where the representation of a storage unit to the operating system and applications on a server is divorced from the actual physical storage where the information is contained.

Virtualization engine Dedicated hardware and software that is used to implement virtualization.

WAN Wide Area Network - A network which encompasses inter-connectivity between devices over a wide geographic area. A wide area network may be privately owned or rented, but the term usually connotes the inclusion of public (shared) networks.

WDM Wave Division Multiplexing - A technology that puts data from different sources together on an optical fiber, with each signal carried on its own separate light wavelength. Using WDM, up to 80 (and theoretically more) separate wavelengths or channels of data can be multiplexed into a stream of light transmitted on a single optical fiber.

WEBM Web-Based Enterprise Management - A consortium working on the development of a series of standards to enable active management and monitoring of network-based elements.

Zoning In Fibre Channel environments, the grouping together of multiple ports to form a virtual private storage network. Ports that are members of a group or zone can communicate with each other but are isolated from ports in other zones.

z/Architecture An IBM architecture for mainframe computers and peripherals. Processors that follow this architecture include the zSeries family of processors.

zSeries A family of IBM mainframe servers that support high performance, availability, connectivity, security and integrity.

Related publications

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

IBM Redbooks

- *IBM TotalStorage: SAN Product, Design, and Optimization Guide*, SG24-6384
- *IBM TotalStorage SAN Volume Controller*, SG24-6423
- *Implementing an Open IBM SAN,* SG24-6116
- *Implementing the Cisco MDS 9000 in an Intermix FCP, FCIP, and FICON Environment*, SG24-6397
- *Introduction to SAN Distance Solutions*, SG24-6408
- *Introducing Hosts to the SAN Fabric*, SG24-6411
- *IP Storage Networking: IBM NAS and iSCSI Solutions*, SG24-6240
- *The IBM TotalStorage NAS Integration Guide*, SG24-6505
- *Implementing the IBM TotalStorage NAS 300G: High Speed Cross Platform Storage and Tivoli SANergy!*, SG24-6278
- *Using iSCSI Solutions' Planning and Implementation*, SG24-6291
- *IBM Storage Solutions for Server Consolidation*, SG24-5355
- *Implementing the Enterprise Storage Server in Your Environment*, SG24-5420
- *Implementing Linux with IBM Disk Storage*, SG24-6261
- *IBM Tape Solutions for Storage Area Networks and FICON*, SG24-5474
- *IBM Enterprise Storage Server*, SG24-5465
- *The IBM TotalStorage Solutions Handbook*, SG24-5250

Referenced Web sites

These Web sites are also relevant as further information sources:

- IBM TotalStorage hardware, software and solutions:

  http://www.storage.ibm.com
- IBM TotalStorage Storage Area Networks:

  http://www-03.ibm.com/servers/storage/san/
- Brocade:

  http://www.brocade.com
- Cisco:

  http://www.cisco.com
- McDATA:

  http://www.mcdata.com
- QLogic:

  http://www.qlogic.com
- Emulex:

  http://www.emulex.com
- Finisar:

  http://www.finisar.co
- Veritas:

  http://www.veritas.co
- Tivoli:

  http://www.tivoli.com
- JNI:

  http://www.Jni.com
- IEEE:

  http://www.ieee.org
- Storage Networking Industry Association:

  http://www.snia.org
- Fibre Channel Industry Association:

  http://www.fibrechannel.com
- SCSI Trade Association:

  http://www.scsita.org

- Internet Engineering Task Force:

  http://www.ietf.org

- American National Standards Institute:

  http://www.ansi.org

- Technical Committee T10:

  http://www.t10.org

- Technical Committee T11:

  http://www.t11.org

- zSeries FICON connectivity:

  http://www-1.ibm.com/servers/eserver/zseries/connectivity/

- iSeries TotalStorage products:

  http://www-1.ibm.com/servers/storage/product/products_iseries.html

- iSeries SAN:

  http://www.ibm.com/servers/eserver/iseries/hardware/storage/san.html

- IBM TotalStorage DS Series portfolio:

  http://www-1.ibm.com/servers/storage/disk/index.html

- IBM Network Attached Storage (NAS) portfolio:

  http://www-1.ibm.com/servers/storage/nas/index.html

How to get IBM Redbooks

You can search for, view, or download Redbooks, Redpapers, Hints and Tips, draft publications and Additional materials, as well as order hardcopy Redbooks or CD-ROMs, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

IBM

Redbooks

Introduction to Storage Area Networks

Introduction to Storage Area Networks

Learn basic SAN terminology and component uses

Introduce yourself to the benefits a SAN can bring

Discover the IBM TotalStorage SAN portfolio

The explosion of data created by the businesses of today is making storage a strategic investment priority for companies of all sizes. As storage takes precedence, three major initiatives have emerged:

- Infrastructure simplification: Consolidation, virtualization, and automated management with IBM TotalStorage can help simplify the infrastructure and ensure an organization meets its business goals.
- Information lifecycle management: Managing business data through its life cycle from conception until disposal in a manner that optimizes storage and access at the lowest cost.
- Business continuity: Maintaining access to data at all times, protecting critical business assets, and aligning recovery costs based on business risk and information value.

Storage is no longer an afterthought. Too much is at stake. Companies are searching for more ways to efficiently manage expanding volumes of data, and to make that data accessible throughout the enterprise; this is propelling the move of storage into the network. Also, the increasing complexity of managing large numbers of storage devices and vast amounts of data is driving greater business value into software and services.

With current estimates of data to be managed and made available increasing at 60 percent per annum, this is where a storage area network (SAN) enters the arena. Simply put, SANs are the leading storage infrastructure for the global economy of today. SANs offer simplified storage management, scalability, flexibility, availability, and improved data access, movement, and backup.

This IBM Redbook gives an introduction to the SAN. It illustrates where SANs are today, who are the main industry organizations and standard bodies active in the SAN world, and it positions IBM's comprehensive, best-of-breed approach of enabling SANs with its products and services. It introduces some of the most commonly encountered terminology and features present in a SAN.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-5470-03 ISBN 0738495565