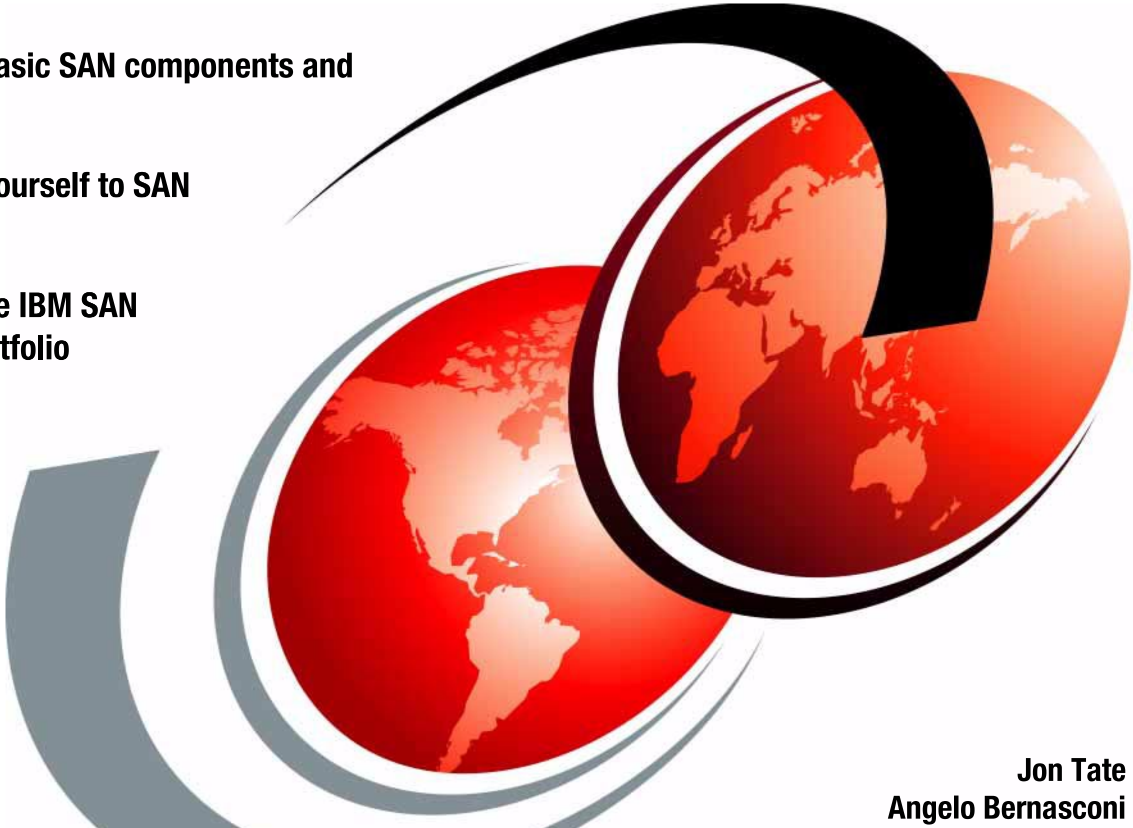
Introduction to Storage Area Networks

- Learn the basic SAN components and their use
- Introduce yourself to SAN benefits
- Discover the IBM SAN product portfolio



Jon Tate
Angelo Bernasconi
Peter Mescher
Fred Scholten

Redbooks

ibm.com/redbooks

International Technical Support Organization

Introduction to Storage Area Networks

March 2003

SG24-5470-01

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

Second Edition (March 2003)

This edition applies to the software and hardware that is in the IBM TotalStorage portfolio, although it may include descriptions and overviews of products and technology that is not.

Contents

Figures

Tables

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:

Redbooks (logo)™
Redbooks™
IBM eServer™
AFS®
AIX®
AS/400®
DB2®
DYNIX®
e-business on demand™
ECKD™
Enterprise Storage Server™
ESCON®
FICON™
FlashCopy™
GDPS™
IBM®
iSeries™
Magstar®
MVS™
NUMA-Q®
OS/390®
OS/400®
Parallel Sysplex®
Perform™
pSeries™
Redbooks™
RS/6000®
S/390®
SANergy™
Sequent®
SP™
StorWatch™
System/390®
Tivoli®
TotalStorage™
Wave®
xSeries™
z/OS™
zSeries™
Lotus®
Lotus Notes®
Notes®
Word Pro®

The following terms are trademarks of other companies:

ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States, other countries, or both.

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

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

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

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

SET, SET Secure Electronic Transaction, and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.

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

Preface

The explosion of data created by e-business is making storage a strategic investment priority for companies of all sizes. As storage takes precedence, two major concerns have emerged: business continuity and business efficiency.

- Business continuity requires storage that supports data availability, so that employees, customers, and trading partners can access data 24x7 through reliable, disaster-tolerant systems.
- Business efficiency, where storage is concerned, is the need for investment protection, reduced total cost of ownership, and high performance and manageability.

Storage cannot be an afterthought anymore. Too much is at stake. Two new trends in storage are helping to drive new investments. First, companies are searching for more ways to efficiently manage expanding volumes of data and how to make that data accessible throughout the enterprise; this is propelling the move of storage into the network. Second, the increasing complexity of managing large numbers of storage devices and vast amounts of data is driving greater business value into software and services.

This is where a Storage Area Network (SAN) enters the arena. Simply put, SANs are the leading storage infrastructure for the world of e-business. SANs offer simplified storage management, scalability, flexibility, availability, and improved data access, movement, and backup.

This IBM Redbook is written for people marketing, planning for, or implementing Storage Area Networks, such as IBM system engineers, IBM Business Partners, system administrators, system programmers, storage administrators, and other technical support and operations managers and staff.

This redbook gives an introduction to SAN. It illustrates where SANs are today, who are the main industry organizations and standard bodies active in SANs, and it positions IBM's approach of enabling SANs with its products and services.

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.

The authors are shown in Figure 0-1.

Figure 0-1 L-R Jon, Peter, Fred, and Angelo

Jon Tate is a Project Manager for SAN TotalStorage 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 17 years of experience in storage software and management, services and support, and is both an IBM Certified IT Specialist and an IBM Certified SAN Specialist.

Angelo Bernasconi is an Advisory ITS Storage and SAN Software Specialist in Italy. He has 16 years of experience in the delivery of maintenance and professional services for IBM Enterprise customers in z/OS and open systems. He holds a degree in Electronics from I.T.I.S. Ernesto Breda (Sesto S. Giovanni (MI) Italy). His areas of expertise include storage hardware products, Storage Area Network products, the design and implementation of SAN solutions, and TSM and Veritas NetBackUp implementation.

Peter Mescher is a Product Engineer for the IBM SAN Central Support Team in Research Triangle Park, North Carolina. He has 3 years of experience in the Storage Area Networking field and has worked at IBM for 4 years. He holds a degree in Computer Engineering from Lehigh University. His areas of expertise include SAN problem determination, switching/routing protocols, and host configuration. He is one of the co-authors of the Storage Networking Industry Association Level 3 (Specialist) certification exam.

Fred Scholten is a Product Engineer. He started at IBM in 1979 as a Field Engineer repairing data recording products. In 1984 he became a Senior Engineer for mainframes at Philips in the Netherlands. In 1995 he went on assignment to Montpellier, France where he worked for 3 years as a Mainframe Specialist for EMEA. He installed some of the first SAN environments in the Netherlands, and is currently on assignment to Mainz, Germany in the IBM SAN Central Solution Team for EMEA. His areas of expertise include defect and non defect problem determination in SAN environments, FICON, and other vendor products such as Brocade, INRANGE, and McDATA.

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

Scott Drummond
IBM Storage Systems Group

The previous authors of this redbook:
Kjell E. Nyström
Ravi Kumar Khattar
Mark S. Murphy
Giulio John Tarella

Mary Lovelace
Tom Cady
Maritza Marie Dubec
Emma Jacobs
Yvonne Lyon
Barry Mellish
Deanna Polm
Sokkieng Wang
International Technical Support Organization, San Jose Center

Jim Baldyga
Brian Steffler
Brocade Communications Systems

Dave Burchwell
INRANGE Technologies Corporation

Barbara Sewell
Larry Shane
McDATA Corporation

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. QXXE Building 80-E2
  650 Harry Road
  San Jose, California 95120-6099

1

Why and what is a SAN?

Computing is based on data. Data is the underlying resource on which all computing processes are based; it is a company asset. Data is stored on storage media, and is accessed by applications executing on a server. Often the data is a unique company asset. You cannot buy your data on the market, but rather you must create and acquire it day by day.

To ensure that business processes deliver the expected results, they must have access to the data. Management and protection of business data is vital for the availability of business processes. Management covers aspects such as configuration, performance, and protection, which ranges from what to do if media fails, to complete disaster recovery procedures.

In the mainframe environments, the management of storage is centralized. Storage devices are connected to the host, and managed directly by the IT department where a system programmer (storage administrator) is completely dedicated to this task. It is relatively straightforward and easy to manage storage in this manner.

The advent of client/server computing created a new set of problems, such as escalating management costs for the desktop, as well as new storage management problems. The information that was centralized in a mainframe environment is now dispersed across the network and is often poorly managed and controlled. Storage devices are dispersed and connected to individual machines; capacity increases must be planned machine by machine; storage acquired for one operating system platform often cannot be used on other platforms. We can visualize this environment as *islands of information*, as shown in Figure 1-1. Information in one island is often hard to access from other islands.

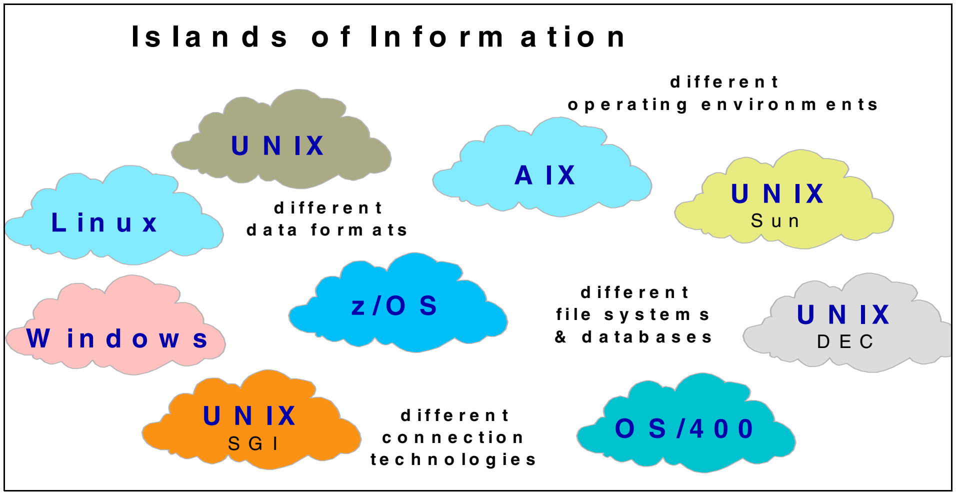


Figure 1-1 Islands of information

The industry has recognized for decades the split between presentation, processing, and data storage. Client/server architecture is based on this three tiered model. In this approach, we can divide the computer network into tiers:

- The top tier uses the desktop for data presentation. The desktop is usually based on Personal Computers (PC).
- The middle tier, application servers, does the processing. Application servers are accessed by the desktop and use data stored on the bottom tier.
- The bottom tier consists of storage devices containing the data.

1.1 The Storage Area Network

In today's Storage Area Network (SAN) environment, the storage devices in the bottom tier are centralized and interconnected, which represents, in effect, a move back to the central storage model of the host or mainframe. A SAN is a high-speed network that allows the establishment of direct connections between storage devices and processors (servers) within the distance supported by Fibre Channel.

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 1-2 shows a tiered overview of a SAN connecting multiple servers to multiple storage systems.

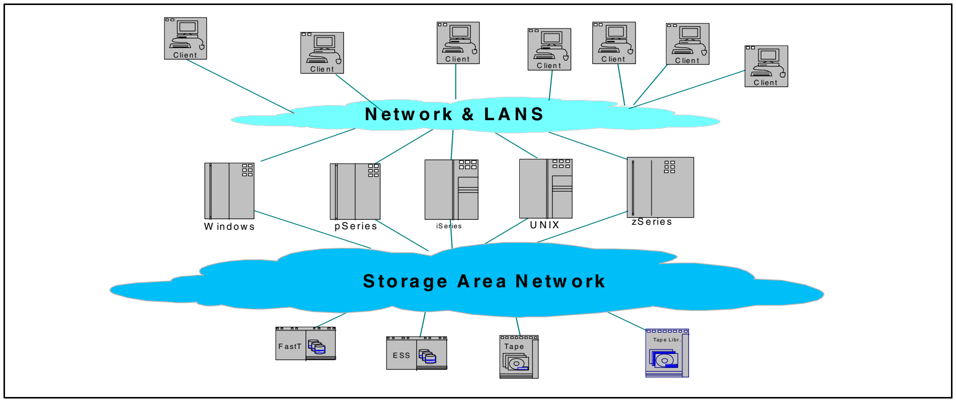


Figure 1-2 What is a SAN?

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. They can interconnect mainframe disk or tape to mainframe servers where the SAN devices allow the intermixing of open systems (such as Windows, AIX) and mainframe traffic.

We discuss this further in Chapter 4, "SAN fabrics and connectivity" on page 53.

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 source to 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.

1.1.1 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.

For more information see 3.6, “Network Attached Storage” on page 50.

1.2 The evolution of storage device connectivity

With the introduction of mainframes, computer scientists began working with various architectures to speed up I/O performance in order to keep pace with increasing server performance and the resulting demand for higher I/O and data throughputs.

In the SAN world, to gauge data throughput when reading a device’s specifications, we usually find a definition using gigabit (Gb) or gigabyte (GB) as the unit of measure. Similarly, we may also find megabit (Mb) and megabyte (MB). For the purpose of this redbook we will use:

- 1 Gb/s = 100 MB/s
- 2 Gb/s = 200 MB/s

1.2.1 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.

1.2.2 Fibre Channel

The Fibre Channel (FC) interface is a serial interface (usually implemented with I fiber-optic cable). It is the primary architecture for most SANs at this time. There are many vendors in the market place producing Fibre Channel adapters. The maximum length for a Fibre Channel cable is dependant on many factors such as the fiber-optic cable size and its mode. With a Fibre Channel connection, we can have up to 200 MB/s data transfer at a 100 km distance.

1.2.3 FICON

The z/OS 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 for transportation. FICON channels are capable of data rates up to 200 MB/s full duplex, 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.

The architectures discussed above are used in z/OS environments and are discussed in z/OS terms. Slightly different approaches were taken on other platforms, particularly in the UNIX and PC worlds where there are different connectivity methods.

1.2.4 SCSI

The Small Computer System Interface (SCSI) is a parallel interface. The SCSI protocols can be used on Fibre Channel (where they are then called FCP). The 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 protocol 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 MB/s.

1.2.5 Ethernet interface

Ethernet adapters are typically used for networking connections over the TCP/IP protocol and can be used today to share storage devices. With an Ethernet adapter we can have up to 10 Gb/s of data transferred.

In the following sections we will present an overview of the basic SAN storage concepts and building blocks, which enable the vision stated above to become a reality.

1.3 SAN definition and evolution

The Storage Network Industry Association (SNIA) defines 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.

It 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.

1.3.1 Fibre Channel architecture

Today, Fibre Channel is the architecture on which most SAN implementations are built. Fibre Channel is a technology standard that allows data to be transferred from one network node to another at very high speeds. Current implementations transfer data at 1 Gb/s or 2 Gb/s. The 10 Gb/s data rates have already been tested and many companies have products in development that will support this. 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 after the fact.

Note: Be aware that the word *Fibre* in Fibre Channel is spelled in the French way rather than the American way. This is because the interconnections between nodes are not necessarily based on fiber optics, but can also be based on copper cables. It is also the ANSI X3T11 technical committee’s preferred spelling. This is the standards organization responsible for Fibre Channel (and certain other standards) for moving electronic data in and out of computers. Even though copper-cable based SANs are rare, the spelling has remained.

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.

In the following sections, we introduce some basic Fibre Channel concepts, starting with the physical and upper layers, topologies, and we proceed to define the classes of service that are offered.

In Chapter 4, “SAN fabrics and connectivity” on page 53, we will introduce some of the other concepts associated with Fibre Channel, namely port types, addressing, fabric services, and Fabric Shortest Path First (FSPF).

Physical layers

Fibre Channel is structured in independent layers, as are other networking protocols. There are five layers, where 0 is the lowest layer. The physical layers are 0 to 2.

In Figure 1-3 we show the upper and physical layers.

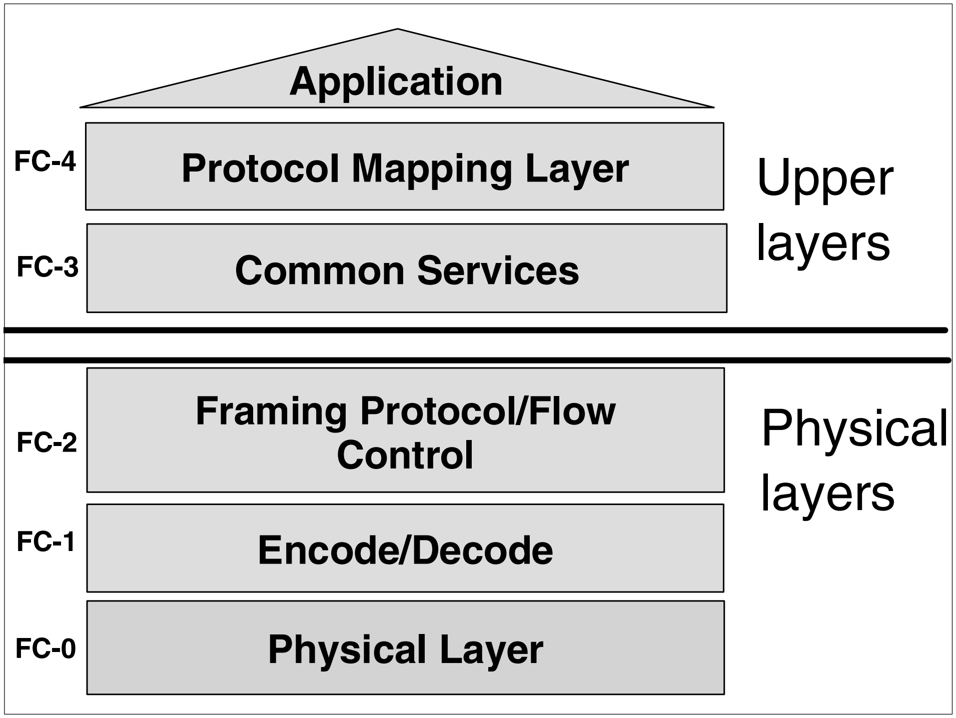


Figure 1-3 Upper and physical layers

They are as follows:

- **FC-0** defines physical media and transmission rates. These include cables and connectors, drivers, transmitters, and receivers.
- **FC-1** defines encoding schemes. These are used to synchronize data for transmission.
- **FC-2** defines the framing protocol and flow control. This protocol is self-configuring and supports point-to-point, arbitrated loop, and switched topologies.

Upper layers

Fibre Channel is a transport service that moves data quickly and reliably between nodes. The two upper layers enhance the functionality of Fibre Channel and provide common implementations for interoperability:

- **FC-3** defines common services for nodes. One defined service is multicast, to deliver one transmission to multiple destinations.
- **FC-4** defines upper layer protocol mapping. Protocols such as FCP (SCSI), FICON, and IP can be mapped to the Fibre Channel transport service.

Topologies

This is the logical layout of the components of a computer system or network, and their interconnections. Topology, as we use the term, deals with questions of *what* components are directly connected to other components from the standpoint of being able to communicate. It does not deal with questions about the physical location of components or interconnecting cables.

Fibre Channel interconnects nodes using three physical topologies that can have variants. These three topologies are:

- *Point-to-point* — The point-to-point topology consists of a single connection between two nodes. All the bandwidth is dedicated to these two nodes.
- *Loop* — In the loop topology, the bandwidth is shared between all the nodes connected to the loop. The loop can be wired node-to-node; however, if a node fails or is not powered on, the loop is out of operation. This is overcome by using a hub. A hub opens the loop when a new node is connected, and closes it when a node disconnects.
- *Switched or fabric* — A switch allows multiple concurrent connections between nodes. There can be two types of switches: circuit switches and frame switches. Circuit switches establish a dedicated connection between two nodes, where as frame switches route frames between nodes and establish the connection only when needed. This is also known as switched fabric.

Classes of service

A mechanism for managing traffic in a network by specifying message or packet priority is needed. The classes of service are the characteristics and guarantees of the transport layer of a Fibre Channel circuit. Different classes of service may simultaneously exist in a fabric.

Fibre Channel provides a logical system of communication for the class of service that is allocated by various login protocols. The following five classes of service are defined:

- **Class 1** — Acknowledged Connection Service

  A connection-oriented class of communication service in which the entire bandwidth of the link between two ports is dedicated for communication between the ports, and not used for other purposes. Also known as dedicated connection service. Class 1 service is not widely implemented.

- **Class 2** — Acknowledged Connectionless Service

  A connectionless Fibre Channel communication service, which multiplexes frames from one or more N_Ports or NL_Ports. Class 2 frames are explicitly acknowledged by the receiver, and notification of delivery failure is provided. This class of service includes end-to-end flow control.

- **Class 3** — Unacknowledged Connectionless Service

  A connectionless Fibre Channel communication service, which multiplexes frames to or from one or more N_Ports or NL_Ports. Class 3 frames are datagrams, that is they are not explicitly acknowledged, and delivery is on a “best effort” basis.

- **Class 4** — Fractional Bandwidth Connection Oriented Service

  A connection-oriented class of communication service in which a fraction of the bandwidth of the link between two ports is dedicated for communication between the ports. The remaining bandwidth may be used for other purposes. Class 4 service supports bounds on the maximum time to deliver a frame from sender to receiver. Also known as fractional service. Class 4 service is not widely implemented.

- **Class 6** — Simplex Connection Service

  A connection-oriented class of communication service between two Fibre Channel ports that provides dedicated unidirectional connections for reliable multicast. Also known as uni-directional dedicated connection service. Class 6 service is not widely implemented.

1.4 SAN components

As mentioned earlier, Fibre Channel is the 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 1-4.

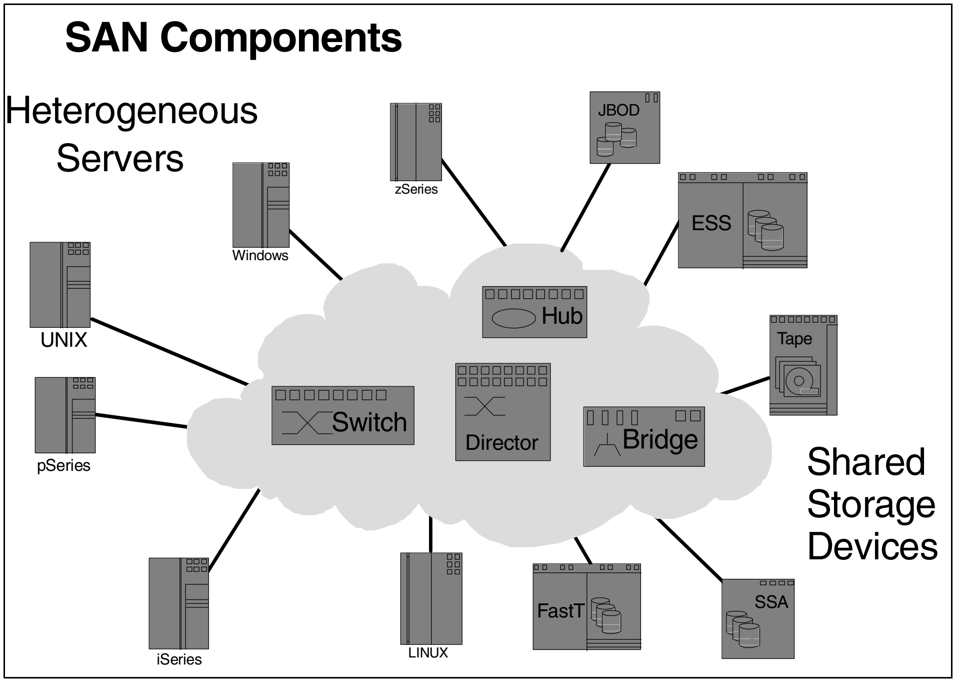


Figure 1-4 SAN components

1.4.1 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.

1.4.2 SAN storage

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. 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.

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).

1.4.3 SAN interconnects

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 below have typically been used for LAN and WAN implementations. SANs, like LANs, interconnect the storage interfaces together into many network configurations and across long 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.

In this section we will take a brief look at some of the components that are commonly encountered.

Cables and connectors

As with parallel SCSI and traditional networking, there are different types of cables with various lengths for use in a Fibre Channel configuration. Fiber-optic cables come in two distinct types: Multi-Mode fiber (MMF) for short distances (up to 2 km), and Single-Mode Fiber (SMF) for longer distances (up to 100 km).

In Figure 1-5 we show the two types of optical connectors.

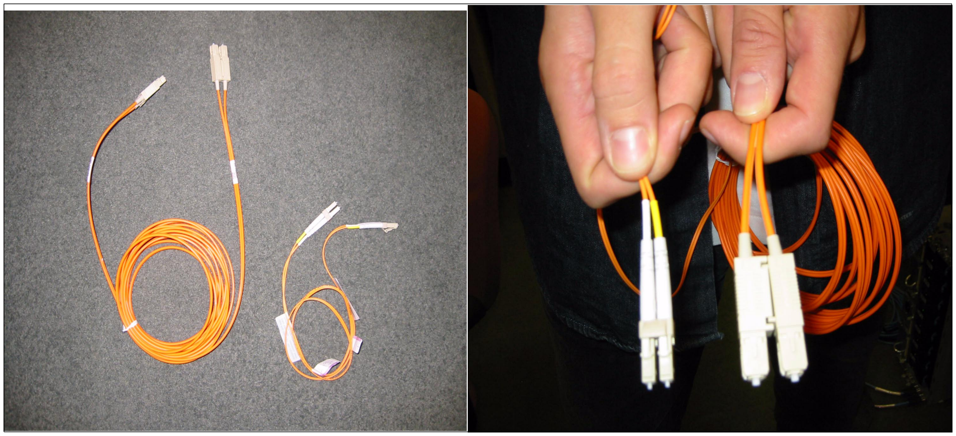


Figure 1-5 LC and SC connectors

IBM will support the following distances for FCP depending on the fiber-optic size and the data transfer speed. We show this in Table 1-1.

Table 1-1 Speeds and feeds

		1 Gb/s	2 Gb/s	10 Gb/s
Fiber-optic size(microns)	**Wavelength modal bandwidth**	**Distance**	**Distance**	**Distance**
9 um SM	1300 nm LX	10Km	10Km	<=292m
50 um MM	850 nm SX	500m	300m	<=280m
62.5 um MM	850 nm SX	250m	120m	<=33m
62.5 um MM	1300 nm LX	550m	NA	N/A

While the FC standards list 4 Gb/s specifications, the table does not show 4 Gb/s characteristics, because it is likely that 4 Gb/s connections will be bypassed and that manufacturers will make the leap directly to 10 Gb/s. The 10 Gb/s characteristics are preliminary numbers specified in the draft standards, and may change before deployment.

In addition, connectors have been developed that allow the interconnection of fiber-optic based adapters and are described in the topics that follow.

Gigabit Interface Converters

Gigabit Interface Converters (GBICs) are laser-based, hot-pluggable, data communications transceivers. They are suitable for a wide range of networking applications requiring high data rates. Designed for ease of configuration and replacement, they are shown in Figure 1-6.

Figure 1-6 Golf ball, SFP GBIC (middle) and regular GBIC

Host bus adapters

Host bus adapters (HBAs) are devices that connect to a server or storage device and control the electrical protocol for communications. There are many vendors in the market place producing HBAs including QLogic, Emulex, JNI, and IBM. They can be single or dual ported, with this likely to increase in the future. We show an example of what an HBA looks like in Figure 1-7.

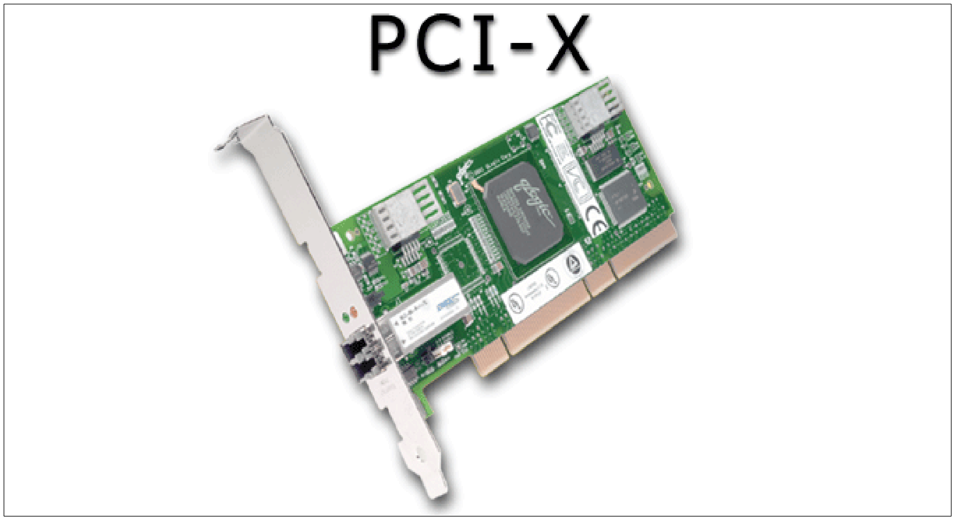


Figure 1-7 QLogic QL2300 HBA

Extenders

Extenders are used to facilitate longer distances between nodes that exceed the theoretical maximum. We show an example of a channel extender in Figure 1-8.

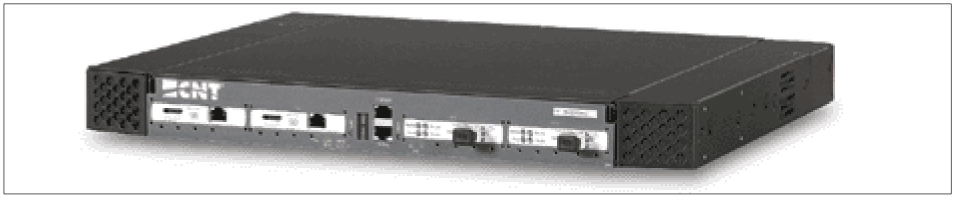


Figure 1-8 Channel extender

Hubs

Hubs are typically used in a SAN to attach devices or servers which do not support switched fabrics, but only FC-AL. They may be either unmanaged or managed, and may or may not be upgradeable to a switched fabric device. We show a picture of a hub in Figure 1-9.

Figure 1-9 IBM managed hub

Gateways

A gateway (also referred to as a bridge or a router) is a fabric device used to interconnect one or more storage devices with different protocol support, such as SCSI to FC, or FC to SCSI devices. We show a picture of a gateway in Figure 1-10.

Figure 1-10 IBM SAN Data Gateway

Switches

Switches are among the highest performing devices available for interconnecting large numbers of devices, increasing bandwidth, reducing congestion, and providing high aggregate throughput. The Fibre Channel protocol was designed specifically by the computer industry to remove the barriers of performance with legacy channels and networks. When a Fibre Channel switch is implemented in a SAN, the network is referred to as a fabric, or switched fabric. Each device connected to a port on the switch can potentially access any other device connected to any other port on the switch, enabling an on-demand connection to every connected device. Various FC switch offerings support both switched fabric and/or loop connections. As the number of devices increases, multiple switches can be cascaded for expanded access (fanout).

As switches allow any-to-any connections, the switch and management software can restrict the other ports to which a specific port can connect to. This is called port zoning. In Figure 1-11 we show a picture of a switch.

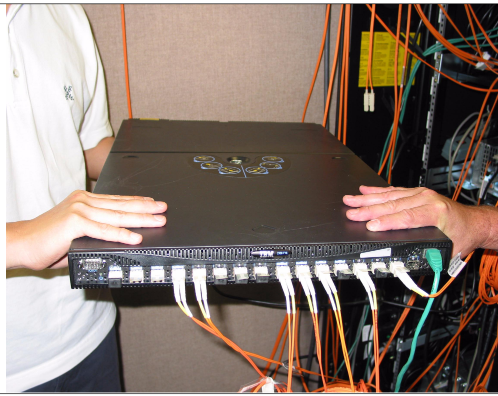


Figure 1-11 IBM 2109-F16 switch

Directors

Directors are SAN devices, similar in functionality to switches, but owing to the redundancy of many of the hardware components, they can supply a higher level of reliability, availability, and serviceability with a smaller footprint. Today, many directors can be used to connect FICON or FC devices at the same time. In Figure 1-12 we show four directors.

Figure 1-12 Directors

1.4.4 SAN applications

SANs enable a number of applications that provide enhanced performance, manageability, and scalability to the IT infrastructure. These applications are being driven in part by technological capabilities, and as technology and products mature over time, we are likely to see more and more applications in the future that are less parochial. They will become true SAN-wide applications that are not bound to one particular vendor's SAN products or SAN implementation. This may be achieved, for example, by adhering to standards, by the use of APIs, or by more intelligent interconnect technology using interconnection agents.

Shared repository and data sharing

SANs enable storage to be externalized from the server and centralized elsewhere, and in so doing, allow data to be shared among multiple host servers without impacting system performance. The term *data sharing* describes the access of common data for processing by multiple computer platforms or servers. Data sharing can be between platforms that are similar or different; this is also referred to as homogeneous and heterogeneous data sharing. We show an example of the concepts in Figure 1-13.

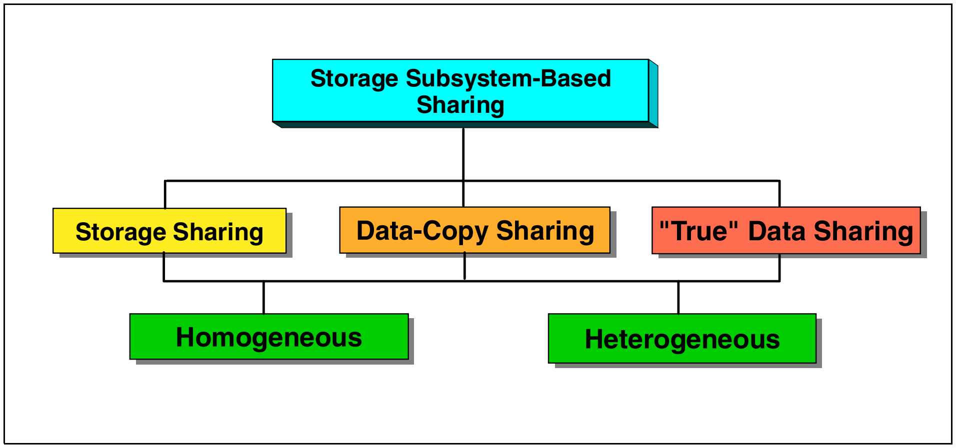


Figure 1-13 Data sharing concepts and types

Storage sharing

With storage sharing, two or more homogeneous or heterogeneous servers share a single storage subsystem whose capacity has been physically partitioned so that each attached server can access only the units allocated to it. Multiple servers can own the same partition, but this is possible only with homogeneous servers.

Data-copy sharing

Data-copy sharing allows different platforms to access the same data by sending a copy of the data from one platform to the other.

We discuss this further in 6.4.1, “From storage partitioning to data sharing” on page 113.

"True" data sharing

In “true” data sharing, only one copy of the data is accessed by multiple platforms, whether homogeneous or heterogeneous. Every platform attached has read and write access to the single copy of data.

In today’s computing environment, true data sharing exists only on homogeneous platforms, for example, in a z/OS Parallel Sysplex configuration. Another example of true data sharing is a RS/6000 cluster configuration with shared-disk architecture using software such as DB2 Parallel Edition to guarantee data integrity.

Data vaulting and data backup

SANs enable multiple copy, data vaulting, and data backup operations on servers to be faster and independent of the primary network (LAN), which has led to the delivery of data movement applications such as LAN-free backup and server-less backup.

Clustering

Clustering is usually thought of as a server process providing failover to a redundant server, or as scalable processing using multiple servers in parallel. In a cluster environment, SAN provides the data pipe, allowing storage to be shared.

Data protection and disaster recovery

The highest level of application availability requires the avoidance of traditional recovery techniques, such as recovery from tape. Instead, new techniques that duplicate systems and data must be architected so that, upon the event of a failure or a disaster being declared, another system is ready to go. Techniques to duplicate the data portion include remote copy and warm standby techniques.

Data protection in environments with the highest level of availability is best achieved by creating secondary redundant copies of the data by storage mirroring, remote cluster storage, Peer-to-Peer Remote Copy (PPRC), and other High Availability (HA) data protection solutions. These are then used for disaster recovery situations. SANs any-to-any connectivity enables these redundant data/storage solutions to be dynamic, and with minimal impact to the primary network and servers, including serialization and coherency control.

For more information on the copy services mentioned see:

- *Implementing ESS Copy Services on UNIX and Windows NT/2000*, SG24-5757
- *IBM TotalStorage Enterprise Storage Server Model 800*, SG24-6424

Subsystem local copy services, such as FlashCopy, assist in creating copies in high availability environments and for traditional backup, and therefore are not directly or strictly applicable to disaster recovery as part of a SAN.

1.4.5 SAN management in general

In order to achieve the various benefits and features of SANs, such as performance, availability, cost, scalability, and interoperability, the infrastructure (switches, directors, and so on) of the SANs, as well as the attached storage systems, must be effectively managed. To simplify SAN management, SAN vendors typically develop their own management software and tools. The major challenge facing vendors and implementers is to ensure that all components will interoperate and work with the various management software packages. This is in order to avoid different management applications that could overwhelm the people who manage the SANs. For more information on SAN management refer to Chapter 5, “SAN management” on page 85.

1.5 Where we were and where we are heading

In this section we discuss the SAN environment as it exists today, and where we think that it is likely to evolve to in time.

1.5.1 Server attached storage

SANs today are a fact, and the promise of an any-to-any nirvana is compelling to executives and technicians alike. However, as most analysts agree and history bears witness to this truth, both technological maturity and mainstream adoption are an evolutionary process. With that in mind, it is best to see where we have come from in order to better understand where we are today, and where we are heading in the future.

Figure 1-14 highlights the limitations of server attached storage.

.

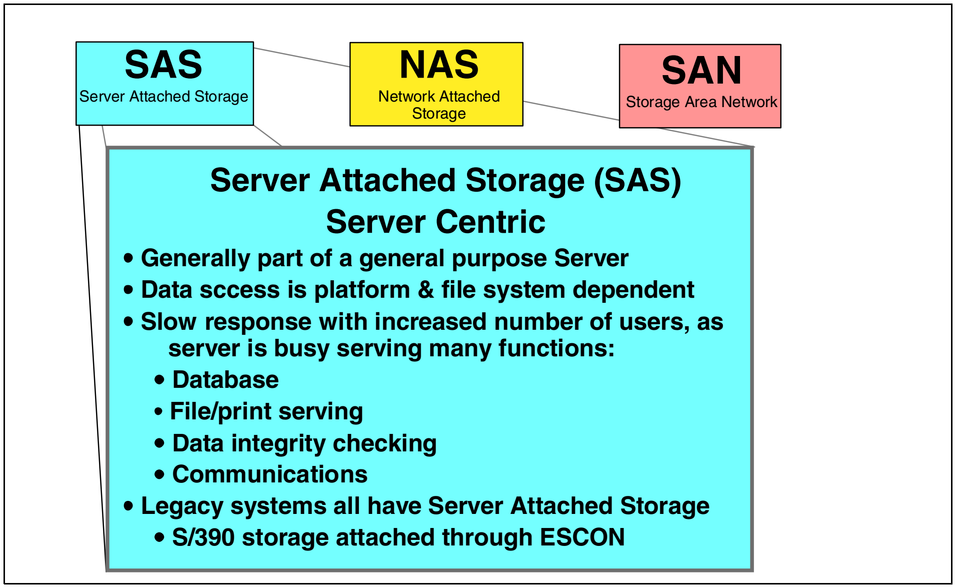


Figure 1-14 Server-centric approach

1.5.2 Network Attached Storage

All of the promises of a SAN — any-to-any connectivity, scalability, availability, performance, and so on — have led customers to begin shopping for SAN solutions today. For customers looking to begin an early SAN-like implementation, there are NAS solutions that allow them to “test the waters” without making a huge investment, and without implementing an entirely new infrastructure.

NAS solutions have evolved over time. Early NAS implementations used a standard UNIX or NT server with NFS or CIFS software to operate as a remote file server. Clients and other application servers access the files stored on the remote file server as though the files are located on their local disks. The location of the file is transparent to the user. Several hundred users could work on information stored on the file server, each one unaware that the data is located on another system. The file server has to manage I/O requests accurately, queuing as necessary, fulfilling the request, and returning the information to the correct client. The NAS server handles all aspects of security and lock management.

Figure 1-15 highlights the peculiarities of Network Attached Storage.

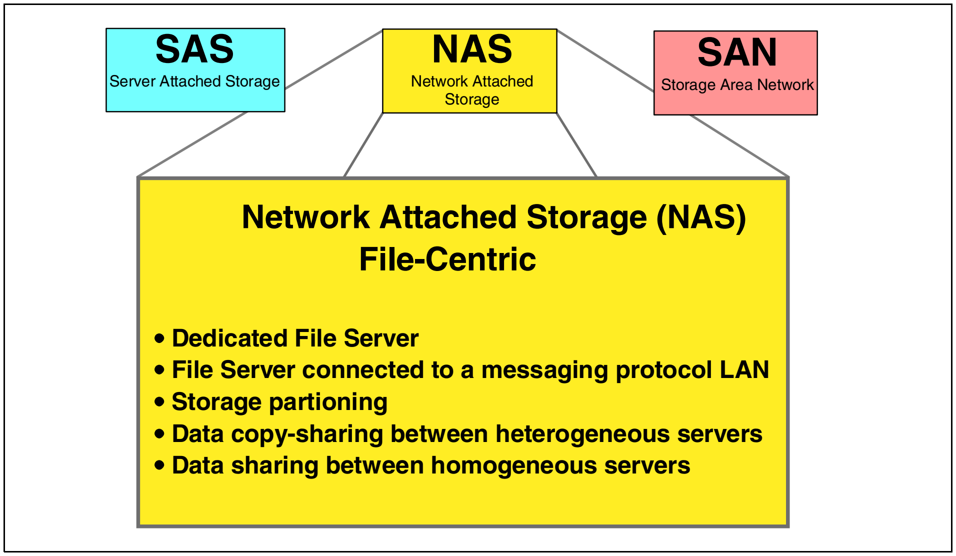


Figure 1-15 File-centric approach

1.5.3 Storage Area Networks

As described throughout this chapter, we are able to summarize the benefits that are inherent in a SAN by using Figure 1-16.

.

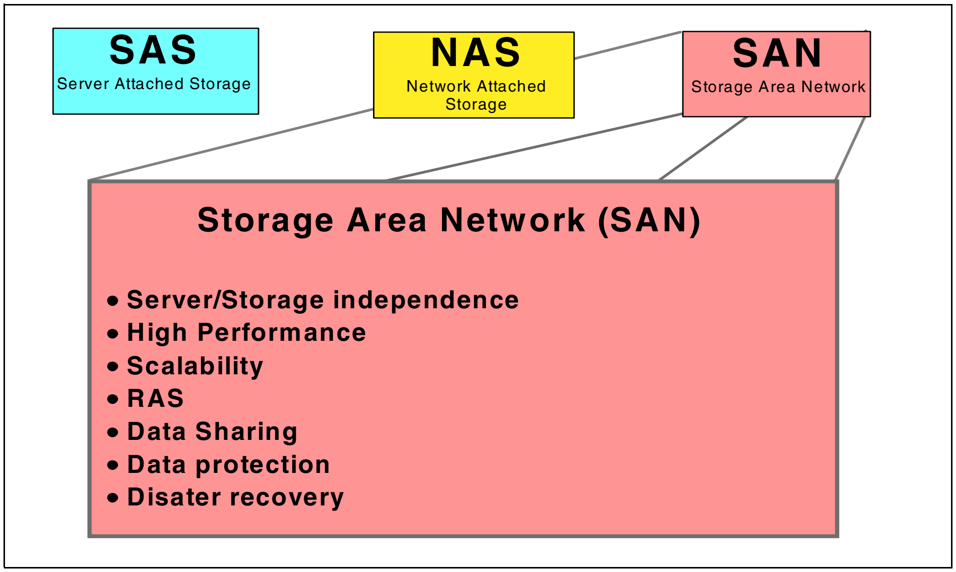


Figure 1-16 Storage Area Network benefits

1.5.4 Storage virtualization in the SAN

Although beyond the scope of this redbook, 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."*

Or, to put 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 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. The IBM storage strategy introduces ways to eliminate these problems with more details available at:

http://publib-b.boulder.ibm.com/Redbooks.nsf/RedpaperAbstracts/redp3633.html

1.6 Standards

One other area that is not related to the future per se, but is instrumental in readying the customer for adoption of any new technology, is the existence of standards. Accordingly, standards bodies are becoming as essential as the referee in a fight to ensure that all sides best interests are served.

1.6.1 The importance of general standards

Because of the presence of many vendors in the market place, and the different ways to implement a SAN using different devices, transport support and protocols, the customer who wants to begin shopping for a SAN has to be certain that their investment is not only the correct one, but protected.

Any kind of device, transport support and protocol has to be approved by the SAN standardization organizations. We discuss some of these august bodies in Chapter 7, “SAN standardization organizations” on page 125.

But more than this, any kind of connection between devices that have been supplied by different vendors, work on different platforms with different software applications, must be certified by the vendors themselves, and have to be supported in order by the vendors in case of problems.

A useful place to find more information about what will work with what is in the interoperability matrix published at any reputable vendor's Web site, or by asking vendors directly for a written guarantee of interoperability. If they are unwilling to supply or underwrite such a guarantee, it might be best to reconsider their claims and take your business elsewhere.

1.7 Conclusion

The nature of the data we put on computers these days has changed. Computer systems do not just run your payroll, keep your inventory, or provide you with a set of graphs for the quarterly results. These systems should drive your results and your payroll, because in today's business climate, they are your shop window and your sales desks. People become knowledge workers, storing data is storing and capturing the intellectual capital of your company. Computer storage also contains much of your private data, none of which you can afford to lose from a legal and business angle. Sharing data and making more efficient use of information with uninterrupted access is facilitated by the introduction of storage networking, whether it is SAN or NAS.

2

The IBM SAN solution

The evolution of information technology (IT), coupled with the explosion of both the intranet and Internet, has led to a storage capacity need that doubles every quarter. This goes on year by year and leads to the need for flexible and fast access to the servers, and the data stored on fast, accessible storage. To achieve this, we need a network that is fast, multipathing, easy to manage, and available 24x7, 365 days a year. The answer to this is a SAN. A SAN will enable host servers to make use of any storage device (disk or tape) attached to the SAN. SANs promise an expansion of openness by allowing heterogeneous storage and server attachment to the SAN. In theory, the user will have the freedom to attach storage to the SAN independent of server type or storage vendor. SAN management, via an abstraction of storage that is called virtualization, will further break the traditional server/storage ownership relationship.

The elements of a SAN can be summarized into the following:

- Server operating systems and host bus adapters
- Fibre Channel fabric interconnect components
- Storage subsystems and/or devices
- Storage abstraction by appliance or software solution
- SAN management software
- Software to exploit SAN capabilities

2.1 e-business on demand

IBM is looking towards the future by looking beyond today's business climate to the next stage of the e-business evolution, and by moving beyond the transformation of individual processes to the transformation of the entire businesses, by looking beyond the integration of the back-office with the front-office:

- To integrate the whole enterprise through entire selling cycles, throughout entire industries
- To use the Internet and computing standards not just for e-business, but for ad hoc e-business, and e-business when you need it

This is e-business on demand.

In such a vision, on-demand storage networks will rapidly adjust to spikes in usage, or to disasters such as fires or floods in order to keep businesses up and running. IBM will use standards to maintain interoperability between different varieties of hardware and software. It is envisaged that, in time, this will lead to solutions that will give customers more flexibility in making their payments and choosing services rendered.

By adopting the on-demand concept, customers can save money and gain competitive advantage by gaining access to more responsive computing capabilities. To quote Sam Palmisano, IBM President and Chief Executive Officer:

> *"Based on everything we see happening in the world of business and technology, we believe we're on the cusp of a fundamental change in computing, which we call the era of "***e-business on demand***."This new model will require serious innovation, but it is more than a technical road map. It includes important new alternatives for how enterprises will acquire and pay for IT, but this goes beyond utility-based computing. At its core, on demand is a way to talk about and understand the full meaning of the networked world and its impact on business and society."*

2.1.1 Server on demand

IBM introduced a service that for the first time enables corporations to access large-scale computing infrastructure on-demand over the Internet. The new IBM business on demand service connects customers to IBM hosting centers that provide managed server processing, and storage and networking capacity on an on-demand basis. Instead of the physical Web, database, and application servers they rely on now, customers tap into virtual servers. At this moment it is only possible on IBM zSeries mainframes in a secure hosting environment. We will see in the near future that this will be available on other platforms as well. The user pays only for the computing power and capacity they require.

2.1.2 Storage on demand

In the same way as server on demand, IBM introduced a service called storage on demand. This introduction marks a major expansion of IBM's strategy to deliver storage. By providing instant access to computing infrastructure without the up front expense of buying the physical hardware, IBM has created an entirely new way for customers to obtain information technology. The idea behind server and storage on demand is that you will plug into a socket on the wall, and you will receive the processing power and/or storage capacity the same way as you do currently out of an outlet.

2.1.3 Pervasive computing

Pervasive computing is computing power freed from the desktop and embedded in wireless handheld devices, automobile telematics systems, home appliances, and commercial tools-of-the-trade. IBM pervasive computing software helps manage information and reduces complexity for a mobile workforce and a mobile society. In the enterprise, it extends timely business data to workers in the field, and drives down costs. In our personal lives, it expands our freedom to exchange information anytime, anywhere. The pervasive computing solution makes it easy for broadband service providers to deliver new services such as streaming video, gaming, backup storage, and much more.

2.2 Storage technology trends driving SAN solutions

There are many trends driving the evolution of storage and storage management. Heterogeneous IT environments are here to stay. Companies and IT departments buy applications and often accept the application vendor's suggested platform. This leads to having many different platforms in the IT shop.

This leads to a second trend, centralization of server and storage assets. This trend is driven by rightsizing and rehosting trends, as opposed to the pure downsizing practiced throughout the 1990s. Storage capacity is also growing, and increasing amounts of data are stored in distributed environments. This leads to issues related to appropriate management and control of the underlying storage resources.

The advent of the Internet and 24x7 accessibility requirements have led to greater emphasis being put on the availability of storage resources, and the reduction of windows in which to perform necessary storage management operations.

Another important trend is the advent of the *virtual* resource. A virtual resource is an appliance solution that performs some kind of function or service transparently. The system using the virtual resource continues to function as if it were using a real resource, but the underlying technology in the appliance solution may be completely different from the original technology. This storage virtualization offers the following benefits:

- Integrates with any storage from any vendor
- Simplifies SAN management and reduces costs
- Improves ROI on storage assets
- Reduces storage related application downtime

All of these trends, and more, are driving SAN implementations.

2.2.1 What is in the SAN today?

The storage infrastructure needs to support the business objectives and new emerging business models. The requirements of today's SAN are:

- Unlimited and just-in-time scalability. Businesses require the capability to flexibly adapt to rapidly changing demands for storage resources.
- Flexible and heterogeneous connectivity. The storage resource must be able to support whatever platforms are within the SAN 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.
- Secure transactions and data transfer. This is a security and integrity requirement aiming to guarantee 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.
- 24x7 response and availability. This is an availability requirement that implies both protection against media failure (that are possibly using RAID technologies) as well as ease of data migration between devices, without interrupting application processing.
- Storage consolidation. This attaches storage to multiple servers concurrently, and leverages investments.
- Storage sharing. Once storage is attached to multiple servers, it can be partitioned or shared between servers that operate with the same operating system.

- Data sharing. This can be enabled if storage is shared and the software to provide the necessary locking and synchronization functions exists.
- Improvements to backup and recovery processes. Attaching disk and tape devices to the same SAN allows for fast data movement between devices, which provides enhanced backup and recovery capabilities. Such as:
  – Serverless backup. This is the possibility to backup 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. One IBM product to realize this is Peer to Peer Remote Copy (PPRC).
  – 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 dataflow. IBM products to realize this is Peer to Peer Remote Copy over eXtended Distance (PPRC-XD) and eXtended Remote Copy (XRC).
- Disaster tolerance. The ability to continue operations without interruption, even in the case of a disaster at one location, is enabled by remote mirroring of data.
- Higher availability. SAN any-to-any connectivity provides higher availability by enabling mirroring, multipathing and alternate pathing, and other methods.
- Improved performance. Enhanced performance is enabled because of more efficient transport mechanisms such as Fibre Channel.
- Selection of *best in class* products. Products from multiple vendors can operate together, and storage purchase decisions can be made independent of servers.
- Simplified migration to new technologies. This facilitates both data and storage subsystem migration from one system to another without interrupting service.

2.2.2 SAN standards

Standards are the base for interoperability of devices and software from different vendors. The 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 allow for heterogeneous, cross-platform, multi-vendor 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.

Figure 2-1 illustrates a SAN to which multiple systems can connect. The SAN is the key to moving digital information from operating system islands to universal access.

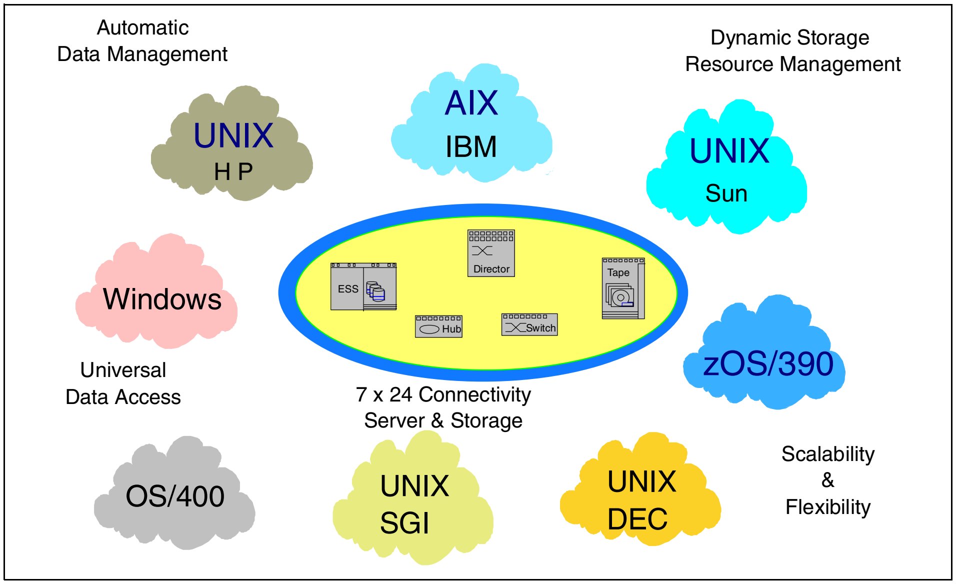


Figure 2-1 Storage with a SAN

Having discussed the trends and issues driving SANs, in the following section we will discuss some of the areas in which IBM is responding to the challenge.

Lowering costs with improved administration

Storage consolidation and centralized management, which uses a common set of processes and procedures, can offer cost savings by reducing the number of people needed to manage a given environment.

Any-to-any connectivity and advanced load balancing can greatly improve storage resource utilization, with return on investment (ROI) increasing in step with deployment, and benefits improving with the development of advanced management functions. At the same time, improved administration in the storage network results in lower TCO.

Improving service levels and availability

The availability of a storage subsystem is generally increased by implementing component redundancy. Availability of storage in a SAN is improved by exploiting a SAN’s any-to-any connectivity, and building in redundancy to eliminate single points of failure. Fault isolation and automated path-to-data selection are also key.

This type of deployment typically offers a short-term ROI. The fault-tolerant components are built and supplied by the storage vendor. Once deployed, the solution is fault-tolerant.

Increased business flexibility

Increased business flexibility is achieved with true data sharing. True data sharing across heterogeneous environments at SAN speeds allows data to be accessed and updated by multiple users and different operating systems. This leads to reduced data movement, and less data transformation between multiple formats. A SAN:

- Reduces or eliminates the expense of redundant storage by eliminating multiple copies and different formats of data
- Increases productivity by providing the data where you need it, when you need it, and in the format you need it, at SAN speeds
- Simplifies workgroup operations by ensuring that data is always at its most current level
- Reduces infrastructure costs by reducing network congestion and overhead on the LAN, particularly for applications such as streaming digital multi-media, and large-file, multi-server workflow

True data sharing supports today’s e-business environment in which on-line or *connected* applications enable a business to seamlessly expand its reach to include customers and suppliers. This value proposition probably has the highest ROI, but can be the most difficult to measure.

2.2.3 IBM and industry standards organizations

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’s participation are described in Chapter 7, “SAN standardization organizations” on page 125.

2.3 IBM TotalStorage. Connected. Protected. Complete.

The IBM TotalStorage family of products is designed to deliver the quality and performance that is expected from IBM, while leading the industry in open storage networking. The mandate is to provide solutions that are based on open standards, and that are interoperable in today's heterogeneous environments.

IBM has a heritage of technology leadership. IBM pioneered the storage business and remains the foundry of invention in the industry with year after year of patent leadership.

As the industry shifts to open storage networking, organizations will be looking for a total solutions approach. IBM has a unique ability to deliver those integrated solutions. IBM TotalStorage has a strategy and portfolio that is connected, protected, and complete.

2.3.1 Connected: Improving data access

IBM storage solutions — based on open standards and storage networking technology — can help improve business efficiency by providing access to information across your organization. By consolidating resources with storage solutions, you can often improve hardware and software utilization and storage management while helping to reduce costs and provide enhanced service. IBM storage networking solutions are designed to provide cross-platform data sharing and any-to-any access, so you can reduce duplication, enhance information availability, and increase network response time and utilization. IBM has proven experience with industry-specialized solutions and applications, and can help optimize solutions for your particular environment.

The IBM SAN solutions can grow quickly, provide exceptional configurational flexibility with any-to-any device connectivity, and help transform the storage infrastructure into an easy-to-manage, highly scalable system. NAS solutions feature high performance storage appliances that can provide efficient pooled storage, allowing data to be shared by multiple clients and servers in a heterogeneous environment. Additionally, IBM and IBM Business Partners have a worldwide network of testing and integration centers, including IBM TotalStorage Solution Centers (TSSC), which provide local venues for hands-on test drives of IBM SAN and NAS solutions, as well as TotalStorage disk, tape, and software products.

Furthermore, the IBM TotalStorage Proven program can help identify third-party storage solutions and configurations that have been tested for interoperability with IBM storage products. IBM Tivoli storage management software offers policy-based automated storage solutions that include backup, restore, data protection, and disaster recovery management. These solutions provide a business-centric approach for information management that spans the entire enterprise while helping leverage the information technology infrastructure.

2.3.2 Protected: Safeguarding your information

The IBM storage solutions address the need for business continuity through data protection and disaster tolerance. A good backup and recovery strategy protects vital data and reduces the backup window. It enables more efficient storage usage, frees up server cycles, and provides faster, more effective recovery. The IBM storage products are designed to provide reliable technology that avoids single points of failure, offers automated backup and recovery processes, and incorporates comprehensive storage management software. They use powerful RAID disk subsystems that are designed to deliver fast access to stored information, without compromising data integrity.

IBM RAID technology, found in products including the IBM Enterprise Storage Server and Virtual Tape Server, is intended to promote high availability and data protection, and help create a disaster-tolerant environment. By leveraging advanced IBM Linear Tape-Open (LTO) technology, the IBM Ultrium family of products can provide a cost-effective storage solution for handling a wide range of backup, restore, archive, and disaster recovery data storage needs. Products such as Tivoli Storage Manager provide centralized storage management; automatic and routine data backup for workstations, servers, and desktop machines; archive and retrieval capabilities; and high-speed, policy-based disaster recovery.

2.3.3 Complete: Providing comprehensive solutions

IBM is one of the only vendors that can provide storage solutions that encompass a full spectrum of products and services: tape, disk, SAN, NAS, storage software, consulting services, and financing. IBM offers one of the broadest portfolios of data storage services — from assessment, design and integration to storage consolidation, data migration, testing and implementation — including fully managed storage services such as capacity, management, and backup and restore services available as e-business on demand. IBM Global Services brings best practices from its thousands of customer engagements.

IBM Global Financing provides flexible payment options at highly attractive rates, and is the world’s premier single-source provider of multi-vendor financing solutions. IBM Global Financing has helped businesses in more than 40 countries realize tremendous savings on systems, software, and services.

IBM TotalStorage provides connected, protected, and complete storage solutions designed for specific requirements — helping to make the storage environment easier to manage, helping to lower costs, and providing enhanced business efficiency and business continuity. IBM has solutions for companies of all sizes in all types of industries. With its technology leadership, integrated solutions, and extensive experience, IBM is uniquely qualified to help design and implement a storage strategy that provides a competitive edge.

For more information visit the Web site:

http://www.ibm.com/totalstorage

For additional information, see the IBM SAN Web site at:

http://www.storage.ibm.com/snetwork/index.html

3

SAN servers and storage

The foundation of a SAN is that the storage and servers are connected. This chapter will present and discuss servers and storage, and how different types of server and storage are used in a SAN environment.

The growth of data, and the increasing dependency of businesses on this data, has resulted in various developments in the field of data storage. For example:

- Different data storage media types suitable for different applications, like disk, tape, and optical media
- Higher data storage capacity: increased capacity of disk drives, from a few megabytes to many gigabytes per drive
- Increased storage performance; for example, 5 MB/s for SCSI-1 to 200 MB/s for Fibre Channel
- Many different storage device and server interconnects; for example, SCSI, SSA, FC-AL, FCP, ESCON, FICON, and others
- Data protection, using solutions such as mirroring and remote copy
- Storage reliability, using various RAID implementations
- Data and storage security and manageability
- Storage pooling by using virtualization products

In Figure 3-1 we illustrate the evolution of storage architecture in relation to the era, or phase, of computing. From centralized computing with controller-based dedicated storage, to the client/server model with distributed data, and finally to the current networked era with its requirement for universal access to data, robust software tools, and data management solutions.

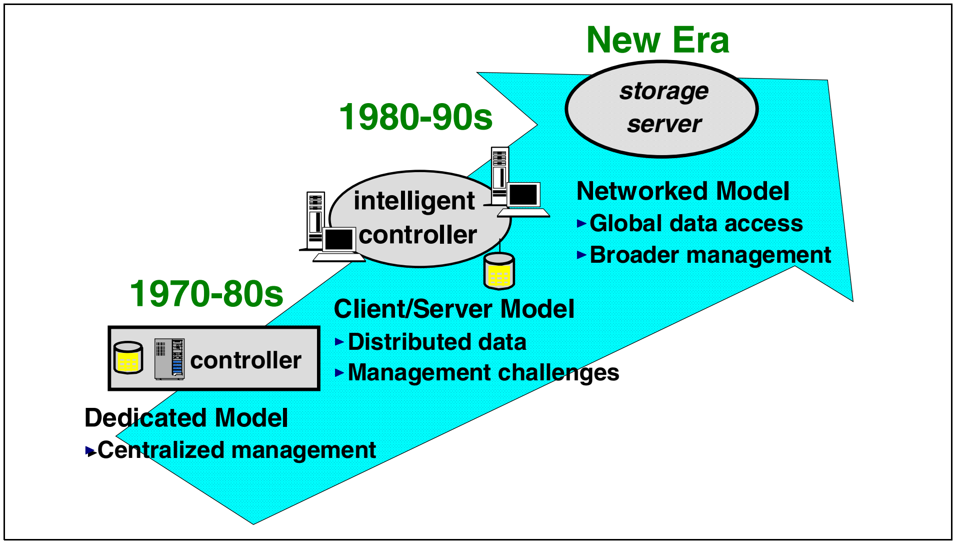


Figure 3-1 The evolution of storage architecture

3.1 Servers and storage environments

Before the advent of SANs, NAS, and modern storage systems, each server in an enterprise needed its own storage. In today's environment, many IT organizations are either planning for, or are actually in the process of implementing server, disk, and/or tape storage consolidation. These organizations have experienced the escalating costs associated with managing information residing on servers and storage that is distributed about their enterprise. This is not to mention the difficulty in implementing and enforcing business recovery procedures, and providing consistent access securely to the information. The objective is to move from an environment of islands of information on distributed servers, with multiple copies of the same data and varying storage management processes, to a consolidated storage management environment, where a single copy of the data can be managed in a central repository. The ultimate goal is to provide transparent access to enterprise information to all users.

3.1.1 The challenge

The real challenge today is to implement true heterogeneous storage and data consolidation environments across different hardware and operating systems platforms; for example, disk and tape sharing across z/OS, OS/400, UNIX, and Windows.

Figure 3-2 illustrates the consolidation movement from the distributed islands of information to a single heterogeneous configuration.

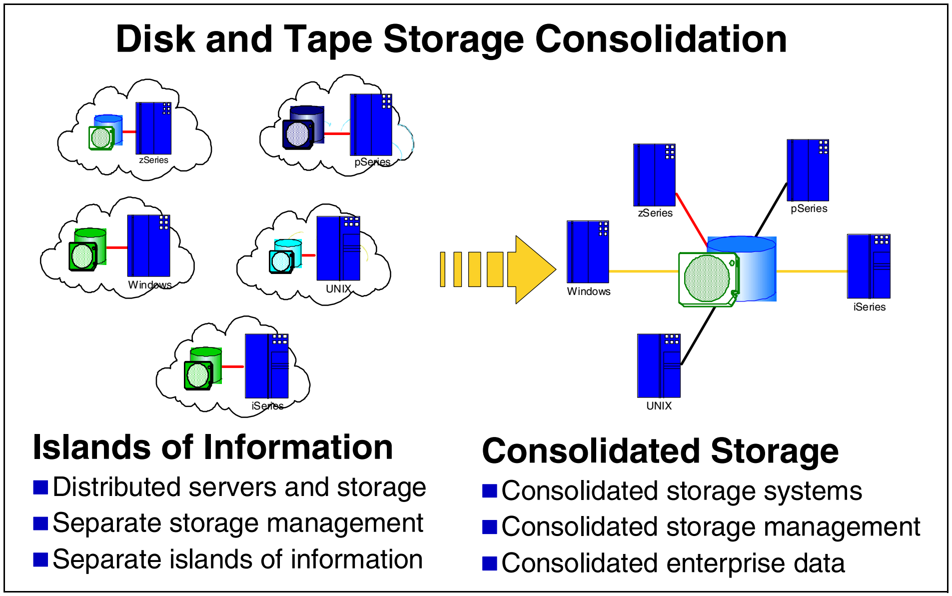


Figure 3-2 Disk and tape storage consolidation

For the past several years, IBM has been working to develop products to meet this ambitious challenge. These products combine technologies from inside and outside IBM to create complete storage solutions. A SAN solution is just that — it is a solution, not a product. In Figure 3-3 we show the IBM strategy for enterprise storage in the network.

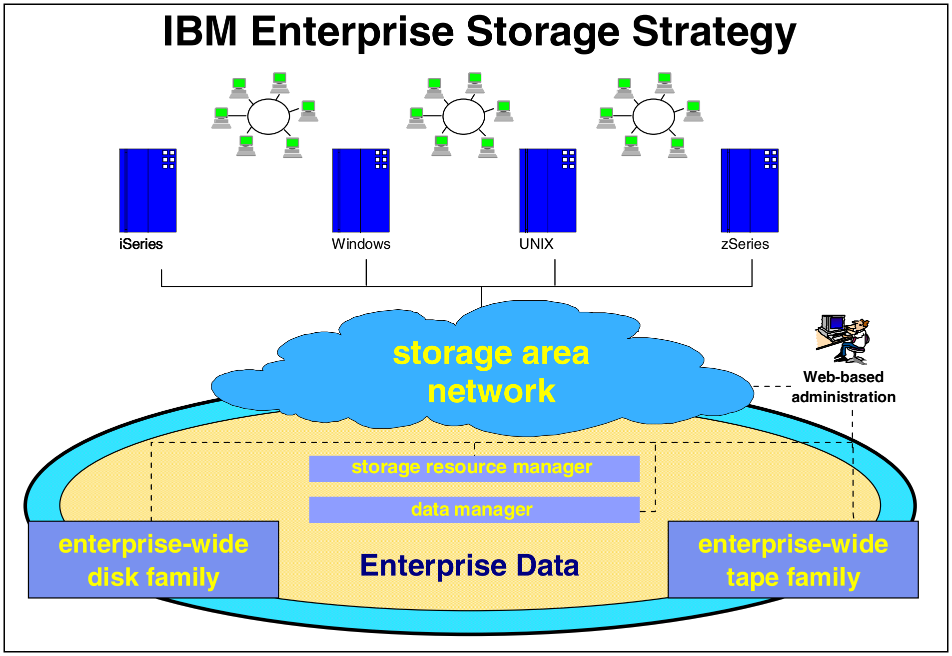


Figure 3-3 IBM storage strategy

Storage consolidation is not a simple task. 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 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), 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 3-4 is a brief summary of these differences for several different systems.

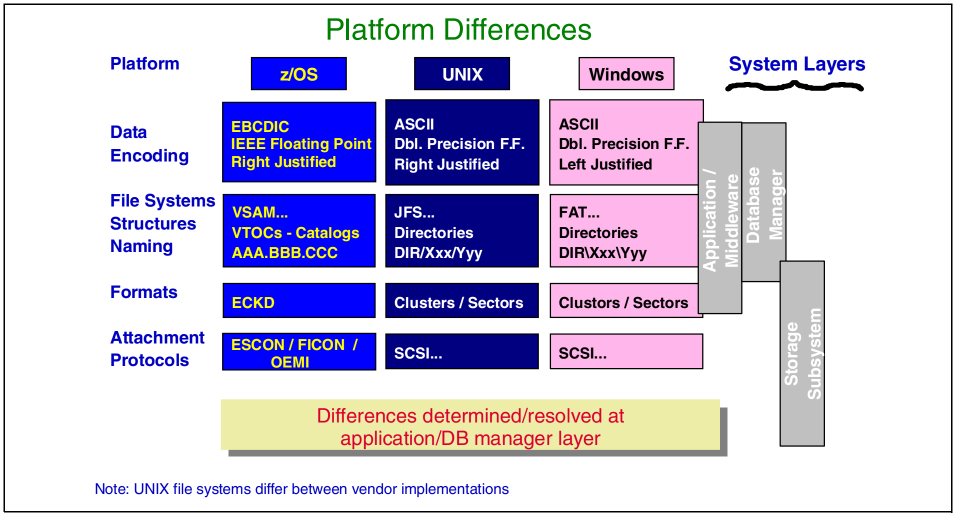


Figure 3-4 Hardware and operating systems differences

3.2 Server environments

Each of the different server platforms (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 sections below explore these solutions and the different implementations on each of the platforms.

3.2.1 zSeries servers

The zSeries (formerly known as S/390) is a processor(s) and operating system 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 interconnections has also evolved from a bus and tag interface, to ESCON channels, and now to FICON channels. Figure 3-5 shows the various processor to storage interfaces on the S/390.

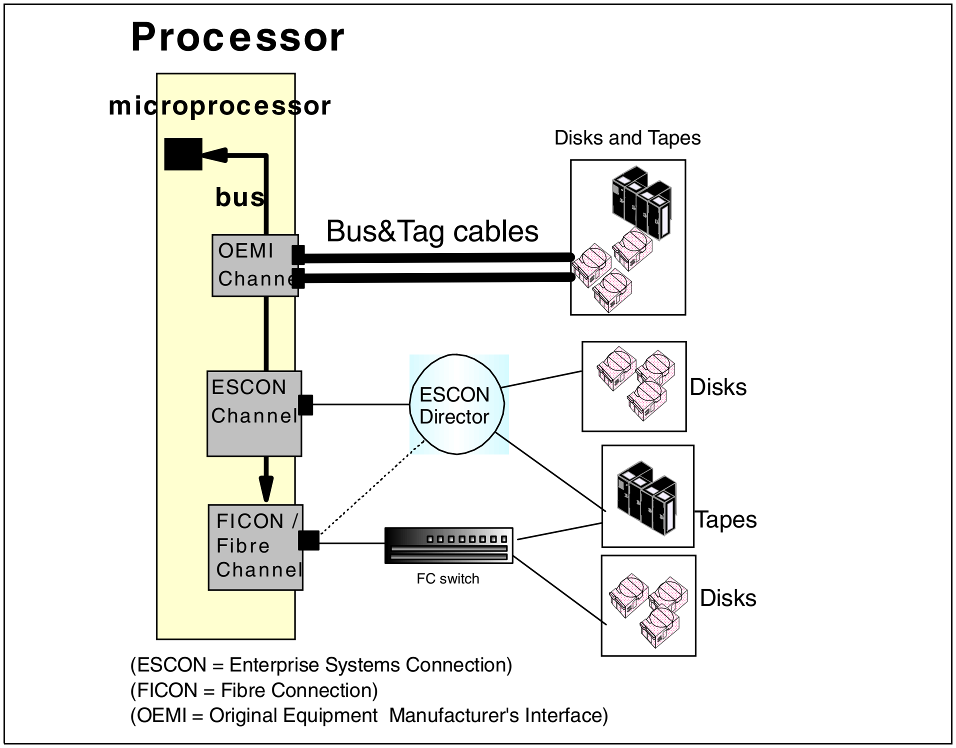


Figure 3-5 S/390 processor to storage interface connections

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 very nearly caught up with FCP SANs.

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 is in the process of developing standard Fibre Channel adapters for use with zSeries servers that implement Linux.

3.2.2 pSeries servers

The IBM 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 3-6 shows the various processor to storage interconnect options for the pSeries family.

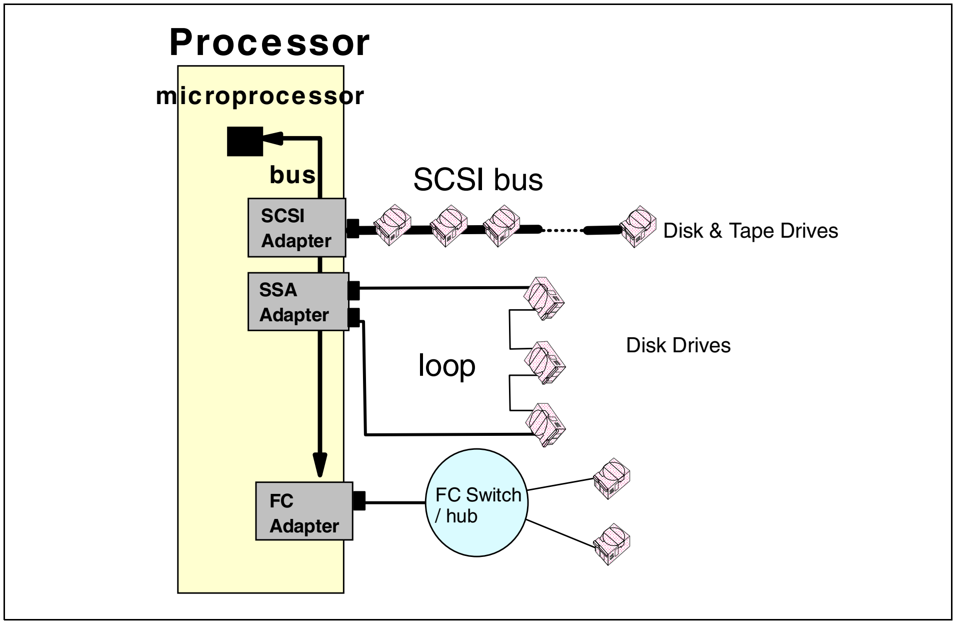


Figure 3-6 pSeries processor to storage interconnections

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) and the Andrew File System (AFS). The server to storage interconnect is similar to pSeries, as shown in Figure 3-6.

3.2.3 iSeries 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 3-7 summarizes the various modules of an iSeries hardware architecture.

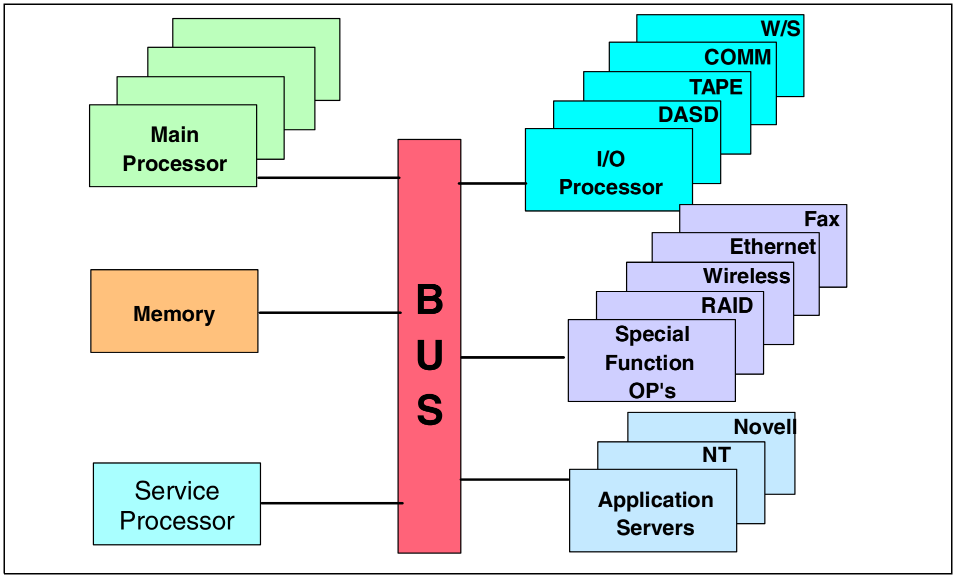


Figure 3-7 iSeries hardware design

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

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 3-8 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 of 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.

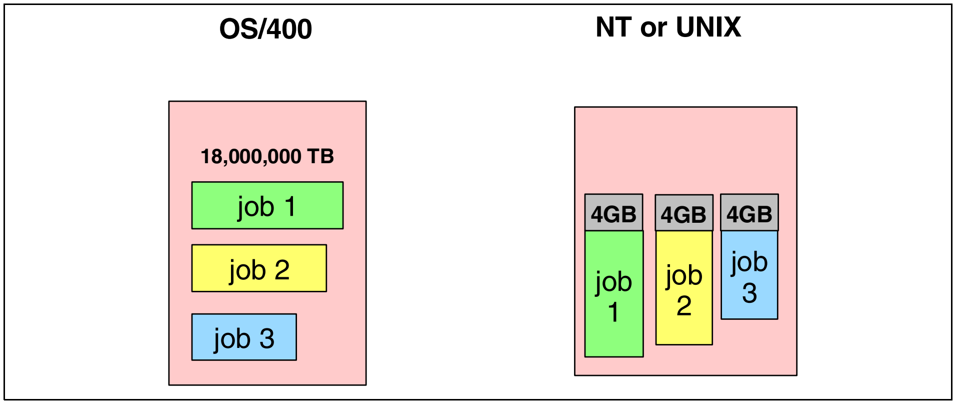


Figure 3-8 OS/400 versus NT or UNIX storage addressing

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

For more information, see the iSeries SAN Web site:

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

3.2.4 xSeries 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, and remote copy.

The processor to storage interfaces on xSeries servers (IBM's 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.pc.ibm.com/ww/eserver/xseries/fa_san.html

3.3 IBM storage products

IBM, as the industry's only complete storage solution provider, sells a complete set of SAN attached storage devices. These products lead the industry in features, openness, performance, and versatility. As a general SAN tutorial, this redbook provides brief descriptions of IBM products. If you would like more in-depth information on the products in the IBM TotalStorage portfolio, refer to *The IBM TotalStorage Solutions Handbook*, SG24-5420.

3.3.1 Open standards and interoperability

Naturally, IBM is also committed to open standards, and as a company recognizes that on occasion customers may choose to implement products from other vendors. In such cases, IBM will work with other vendors in order to solve any storage problems. For instance, IBM is a member of the Technical Support Alliance, which is a network of agreements between most of the major companies in the computer industry. It exists to provide contacts within the support organization in order to solve problems that cross vendor lines. As of this writing, TSA has 123 member companies, including IBM, Sun, Hitachi, Microsoft, Veritas, HP, Dell, and many others.

3.4 IBM TotalStorage SAN-attach disk products

In this topic we will detail the IBM disk products that are SAN-attachable components.

3.4.1 IBM TotalStorage Enterprise Storage Server

The IBM TotalStorage Enterprise Storage Server (ESS) is currently IBM's top-of-the-line disk storage product. The ESS is very scalable and can be configured to store anywhere from a few hundred gigabytes of data, up to many terabytes. The ESS supports attachment to many different operating systems, including Windows, LINUX, AIX, Solaris, OS/400, z/OS, DG-UX, HP-UX, OpenVMS, Tru64, DYNIX, and Netware.

It is equipped with features necessary for enterprise SANs, such as:

- FlashCopy - Creates duplicates of partitions in virtually no time at all. This is very useful in backup applications
- PPRC - Provides real-time remote mirroring of disk data
- Versatile connection options - 16 I/O slots, which can accommodate SCSI, ESCON, FICON, and Fibre Channel adapters
- Extremely fault-tolerant design - The ESS has many hot-swappable parts and can have its code upgraded concurrently, for minimum downtime.

For more information refer to the following IBM Redbooks:

- *IBM TotalStorage Enterprise Storage Server Model 800*, SG24-6424

- *IBM Enterprise Storage Server*, SG24-5465

The product homepage for the ESS is at Web site:

http://www.storage.ibm.com/hardsoft/products/ess/index.html

3.4.2 IBM TotalStorage FAStT Storage Servers

For installations with some less stringent storage requirements, IBM offers the FAStT line of storage servers:

- The FAStT 200 is meant for installations of modest capacity and connectivity needs, using 1 Gb Fibre Channel connections. It supports up to 66 drives, and features 128 MB of battery-backed cache, and can perform FlashCopy operations.
- The FAStT 500 is a Fibre Channel attached unit, using 1 Gb Fibre Channel connections and can support up to 224 disk drives. It features from 256 MB to 1 GB of battery-backed cache, and can perform FlashCopy and remote mirroring operations.
- The FAStT 700 has 2 Gb Fibre Channel interfaces and can support up to 224 drives. It includes 2 GB of battery-backed cache, and can perform FlashCopy and remote mirroring operations.
- The FAStT 900 has 2 Gb Fibre Channel interfaces and can attach to 224 drives. It includes 2 GB of battery-backed cache, and can perform FlashCopy and remote mirroring operations.

For host interfaces, all FAStT support both the arbitrated loop and switched fabric standards.

For more information refer to the following IBM Redbooks:

- *IBM TotalStorage FAStT700 and Copy Services*, SG24-6808
- *Fibre Array Storage Technology A FAStT Introduction*, SG24-6246

3.5 IBM TotalStorage SAN-attach tape products

IBM is an active participant in the high-performance tape market, and is always producing solutions with ever-higher capacity and speed. IBM carries two major lines of SAN-attach tape products. Linear Tape Open (LTO) and the 3590 Magstar line.

3.5.1 IBM TotalStorage Linear Tape Open

The Linear Tape Open (LTO) standard is an open standard for tape jointly developed by IBM, Hewlett Packard, and Seagate. It is a high-capacity, high-performance technology with a strong roadmap for the future.

With Generation 1 LTO technology, an LTO cartridge will hold 100 GB of data in raw format, and around 200 GB if the built-in hardware compression routines are used (of course total compressed data capacity depends on the data being compressed). Generation 2 LTO has been defined by the group of companies developing LTO, and IBM has started shipping Generation 2 LTO drives. Generation 2 LTO will feature cartridges with a capacity of 200 GB native, and up to 400 GB with 2:1 compression.

Generation 1 LTO drives have a raw write speed of 15 MB/s, which can be increased up to around 30 MB/s with 2:1 compression. With Generation 2 LTO, up to 35 MB/s data rates (70 MB/s at 2:1 compression) can be achieved.

LTO drives are available to attach to both Fibre Channel SANs and to SCSI. They will attach to servers running any of the following operating systems: AIX, Windows, Solaris, Linux and OS/400.

For more information on LTO, refer to the IBM Redbook:

- *The IBM LTO Ultrium Tape Libraries Guide*, SG24-5946

While there is a model of LTO available for stand-alone use (the 3580 series), most SAN customers will use an LTO drive inside an automated tape library. These libraries have a wide range of sizes and abilities:

- 3581 LTO Ultrium Tape Autoloader: This unit can accommodate a single LTO drive, and up to seven LTO cartridges. With Generation 1 LTO technology, this gives the unit up to 1.4 terabytes of capacity (with 2:1 compression). This unit is equipped with a SCSI interface that can be attached to a SAN through the SAN Data Gateway or Router. Future models are likely to have a Fibre Channel interface.
- 3583 LTO Ultrium Scalable Tape Library: This unit is available in configurations with anywhere from one to six drives and slots to hold from 18 to 72 cartridges. With compression, this library can hold up to 14.4 terabytes of data with Generation 1 LTO cartridges. Fibre Channel attachment is available as a feature with this library.
- 3584 LTO UltraScalable Tape Library: This is IBM's enterprise-class Generation 1 and 2 LTO library. It can consist of up to six frames, each of which can hold up to 12 tape drives. The library can be equipped with either one or two tape grippers, for both speed and redundancy purposes.

This library can be equipped with either SCSI or Fibre Channel attached drives.

The 3584 is IBM's highest capacity LTO library. Depending on how many drives are in a frame, a single frame (filled to capacity with slots, and using Generation 1 technology) can hold anywhere from 27.6 to 56.2 terabytes (compressed at 2:1). With the minimum number of drives, this gives the library a maximum Generation 1 capacity of 496.2 terabytes (with 2:1 compression).

IBM intends to further introduce and integrate the IBM TotalStorage Ultrium 2 Tape Drive (Generation 2 LTO) technology in additional or alternative IBM LTO Ultrium 358x products.

IBM also offers an external drive for use outside of a library:

- Ultrium External Tape Drive 3580: The Ultrium 2 models of the IBM TotalStorage have a capacity of up to 400 GB with compression (2:1) with the use of the new IBM TotalStorage LTO Ultrium 200 GB Data Cartridge. IBM Ultrium 2 Tape Drives can read and write first generation LTO Ultrium Data Cartridges at original capacities providing protection investment, and with improved performance.

3.5.2 IBM TotalStorage Enterprise Tape System 3590

For years, IBM's Magstar line of tape drives have earned a reputation for both performance and reliability. They can be attached to servers running z/OS, AIX, Windows, OS/400, Linux, Solaris, and HP-UX.

The 3590 drives are currently available with three different tape capacities: The "B" models store 20 GB per cartridge at a speed of up to 9 MB/s (these numbers are without considering the effect of the built in compression). The "E" models will store 40 GB per cartridge at 14 MB/s. The "H" models will store 60 GB, also at 14 MB/s. It is likely that IBM will release further enhancements to the technology in the future.

For a general overview of the 3590 drives, refer to the IBM Redbook, *IBM Magstar Tape Products Family: A Practical Guide*, SG24-4632.

The 3590 Drives are available from IBM in three configurations, each of which can attach to either ESCON, FICON (done through an ESCON/FICON bridge), SCSI, and Fibre Channel. ESCON/FICON attachment must be run through a separate control unit, which in turn is cabled to the drives.

- Stand-alone: If desired, the 3590 drive can be loaded manually by the user without any automatic aids.

- Autoloader: The 3590 drive may have a 10-cartridge changer attached to the front of the unit.
- 3494 Enterprise Tape Library: This is IBM’s enterprise-class Magstar library. It can accommodate up to 32 of any of the three different size 3590 drives (in addition to the older 3490 drives, if desired). In its maximum configuration of a 16 frame library, it can hold 6,240 cartridges for a maximum capacity of 1,122 terabytes of storage.

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.

3.6 Network Attached Storage

IBM sells several Network Attached Storage units (NAS). 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.

Each type of storage has advantages and disadvantages. SAN-attach storage generally has better performance, and more solid security. NAS storage is less expensive for hosts to implement (Ethernet adapters are much less expensive than Fibre Channel adapters), the same data can be more easily shared between hosts, and virtually any end-user on an Ethernet network can obtain files from a NAS box. The security of the data can also be implemented at a far more granular level (each file can be assigned to specific users).

In an effort to bridge the two worlds, and to open up new configuration options for customers, IBM also sells a NAS unit that acts as a gateway between IP-based users, and SAN-attached storage. This allows you to connect 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. 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.

For more information on IBM’s rapidly evolving NAS portfolio, refer to the Web site:

`http://www.storage.ibm.com/snetwork/nas/index.html`

4

SAN fabrics and connectivity

A Fibre Channel, SAN employs a fabric to connect devices. A fabric can be as simple as a single cable connecting two devices. 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 4-2.

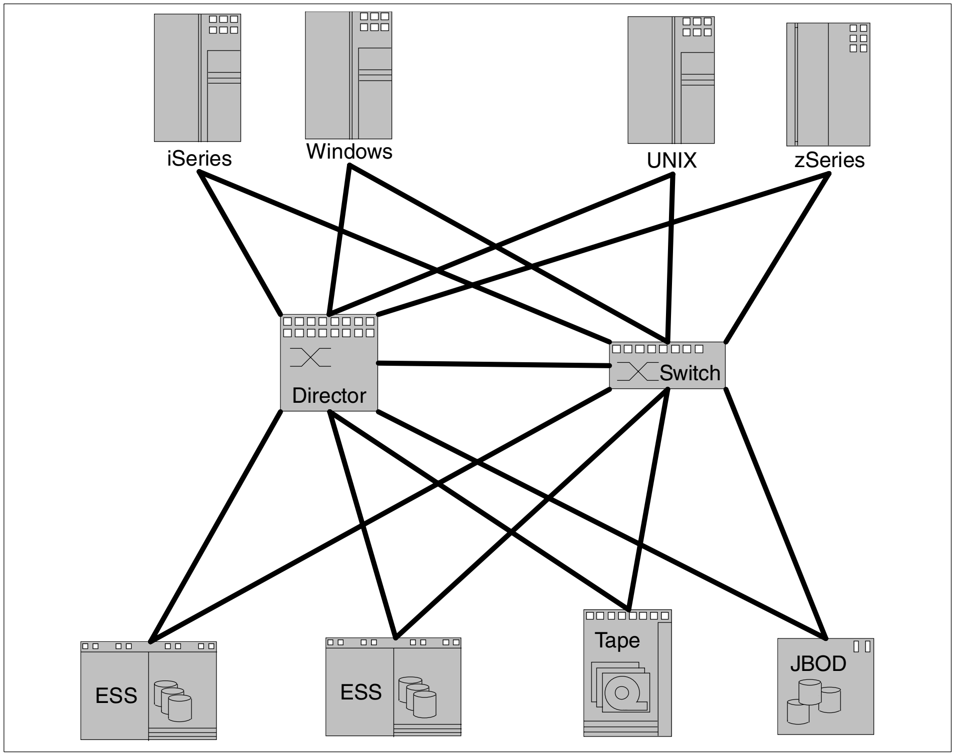


Figure 4-1 High level view of a fabric

4.1 The SAN environment

Historically, interfaces to storage consisted of parallel bus architectures (such as SCSI and IBM's 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 in the order of 200 MB/s, and enhancements to the Fibre Channel standard will support 10 Gb/s in the near future. This will be mostly used for inter switch links (ISL) between switches and directors.

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.

A Fibre Channel network may be composed of many different types of interconnect entities, including directors, switches, hubs, and bridges.

Different types of interconnect entities 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.

4.1.1 The Storage Area Network

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 4-2.

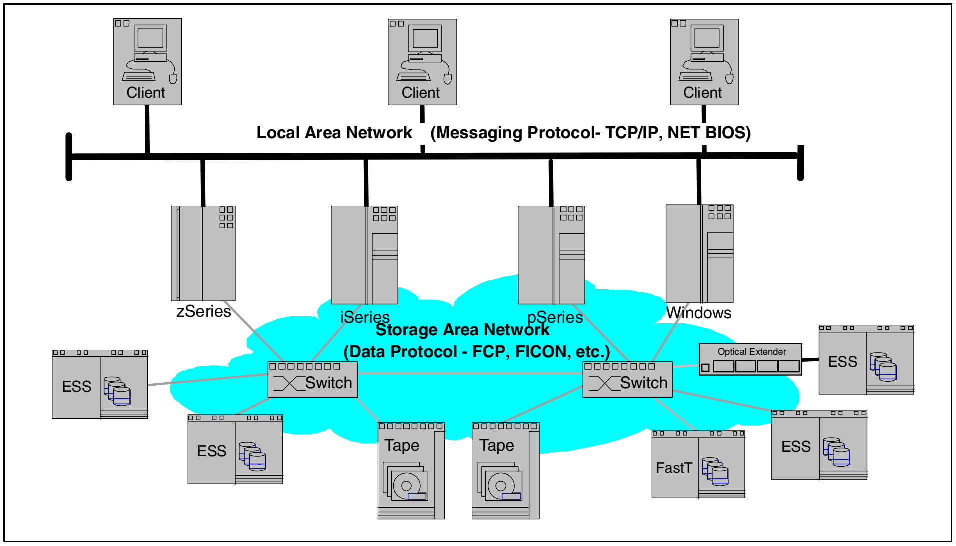


Figure 4-2 The SAN

A Fibre Channel SAN has a high performance because of the speed, which is now up to 200 MB/s, and because of the unique packaging of data which only consumes about 5% of overhead. This leads to an effective use for one connection of approximately 190 MB/s, bidirectionally. The network technology components are directors, switches, and gateways. It is even possible to go over 100 km by using extenders, dark fiber, or Metropolitan Area Networks (MANs).

Unlike NAS products, SAN products do not function like a server or file server. Instead, the SAN product processes different kind of protocols, such as FCP (SCSI),FICON, iSCSI, etc., and the way it does this is transparent to the server or the storage.

4.2 Fibre Channel

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.

4.2.1 Point-to-point

Point-to-point is the easiest Fibre Channel configuration to implement, and it is also the easiest to administer. Concurrent transmission and reception of data on a single link is known as full duplex mode. This simple link can be used to provide a high-speed interconnect between two nodes, as shown in Figure 4-3.

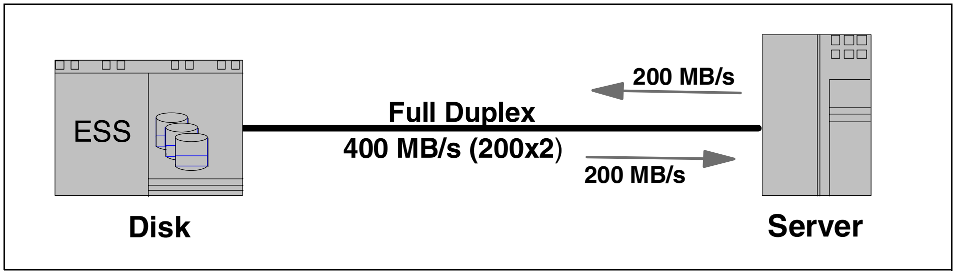


Figure 4-3 Fibre Channel point-to-point topology

Possible implementations of point-to-point connectivity may include the following connections:

- Between central processing units
- From a workstation to a specialized graphics processor or simulation accelerator
- From a file server to a disk array or optical jukebox
- From a disk subsystem to a tape subsystem for serverless backup

When greater connectivity and performance is required, each device can be connected to a fabric without incurring any additional expense beyond the cost of the fabric itself.

4.2.2 Loops and hubs for connectivity

Fibre Channel arbitrated loop (FC-AL) offers relatively high bandwidth and connectivity at a low cost. For a node to transfer data, it must first arbitrate to win control of the loop. Once the node has control, it is now free to establish a point-to-point (virtual) connection with another node on the loop. After this point-to-point (virtual) connection is established, the two nodes consume all of the loop's bandwidth until the data transfer operation is complete. Once the transfer is complete, any node on the loop can now arbitrate to win control of the loop. The characteristics which make Fibre Channel arbitrated loop so popular include:

- Support of up to 126 devices is possible on a single loop, but the more devices that are on a single loop, the more competition to win arbitration will take place.
- Devices can be hot swapped with the implementation of hubs and bypass ports.
- A loop is self-discovering (it finds out who is on the loop and tells everyone else).
- Logic in the port allows a failed node to be isolated from the loop without interfering with other data transfers.
- Virtual point-to-point communications are possible.
- A loop can be interconnected to other loops to essentially form its own fabric.
- A loop can be connected to a Fibre Channel switch to create fanout, or the ability to increase the size of the fabric even more.

More advanced FC hub devices support FC loop connections while offering some of the benefits of switches. Figure 4-4 on page 58 shows an FC loop using a hub.

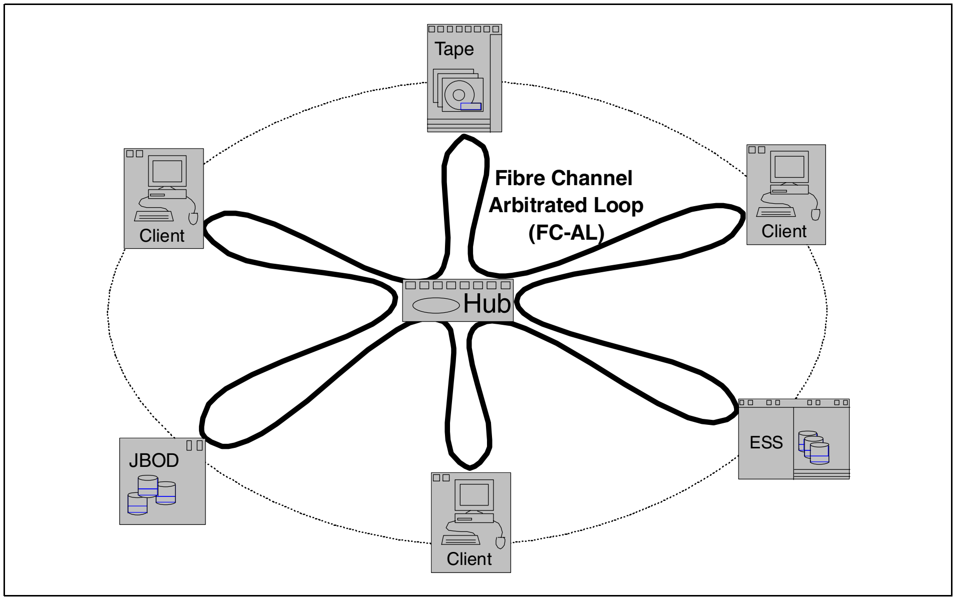


Figure 4-4 Fibre Channel loop topology

To be able to connect a hub to a fabric, the devices that want to communicate with another device in the fabric must be able to support public loop. When you connect a hub to a fabric, the switch looks for the Arbitrated Loop Physical Address (AL_PA) of the devices on the loop. If a device does not have an AL_PA, it is not a public loop, but a private loop device. For example, until recently, the iSeries was only able to run as a private loop device.

To get a device with a private loop HBA to communicate with FC- AL storage devices through IBM TotalStorage SAN switches, you need a special function in the switch called *QuickLoop*.

QuickLoop

Peculiar to the IBM 2109 family of switches (except the IBM TotalStorage SAN Swich M12), QuickLoop is a unique FC topology that combines arbitrated loop and fabric topologies. It is an optionally licensed product that allows arbitrated loops to be attached to a fabric. An arbitrated loop supports communication between devices that do not recognize fabrics. Such devices are called *private devices*, and arbitrated loops are sometimes called *private loops*. Without modifying the private devices on the arbitrated loop, they can be accessed by public or private hosts elsewhere on the fabric. A QuickLoop consists of multiple private arbitrated looplets (a set of devices connected to a single port) that are connected by a fabric. All devices in a QuickLoop share a single AL_PA bit map, and act as if they are in one loop. This allows private devices to communicate with other devices over the fabric, provided they are in the same QuickLoop. Other vendors may also provide the means to perform a similar function in their products. They should be consulted as to the methods they use to allow arbitrated loops to connect to fabrics.

4.2.3 Fabrics for scalability, performance, and availability

Amongst the elite of the FC SAN world are switches and directors. They offer the highest availability, number of ports in a single footprint, and a lot of other features and benefits.

4.2.4 Switches and directors

Fibre Channel switches and directors function in a manner similar to traditional network switches to provide increased bandwidth, scalability and performance, devices, and most importantly, increased redundancy. Fibre Channel switches and directors vary in the number of ports and media types they support. Multiple switches can be connected to form a switched fabric capable of supporting a large number of system hosts and storage subsystems as shown in Figure 4-5. When switches or directors are connected, each switch configuration information has to be copied (cascaded) into all the other participating switches and directors in the fabric.

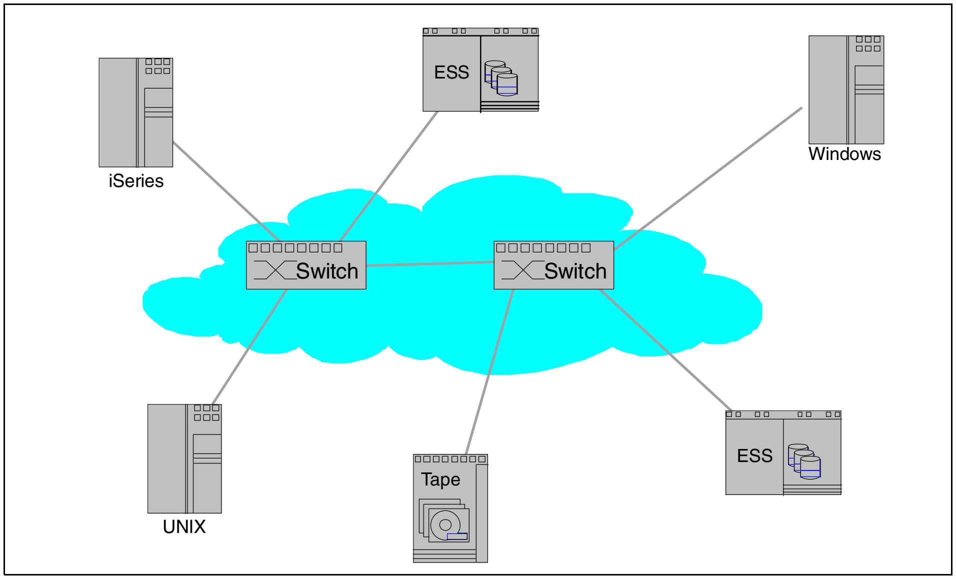


Figure 4-5 Fibre Channel switched topology

4.2.5 Multiple fabric connections

Topologies where there is no ISL between two directors or switches are called multi-fabric connections. This supports those configurations required by very large systems for FICON connections.

This topology is also used in larger SANs, so if there is a catastrophic error, a complete fabric goes down the servers and the storage devices still have their connectivity over the other fabric. This configuration can be set up to allow every system to have access to every fabric, and every controller to be connected to at least two fabrics. This allows any system to get to any controller/device, and it allows for continuous operation (with degraded performance) in the event that a fabric fails. An example of a multiple fabric connection is shown in Figure 4-6.

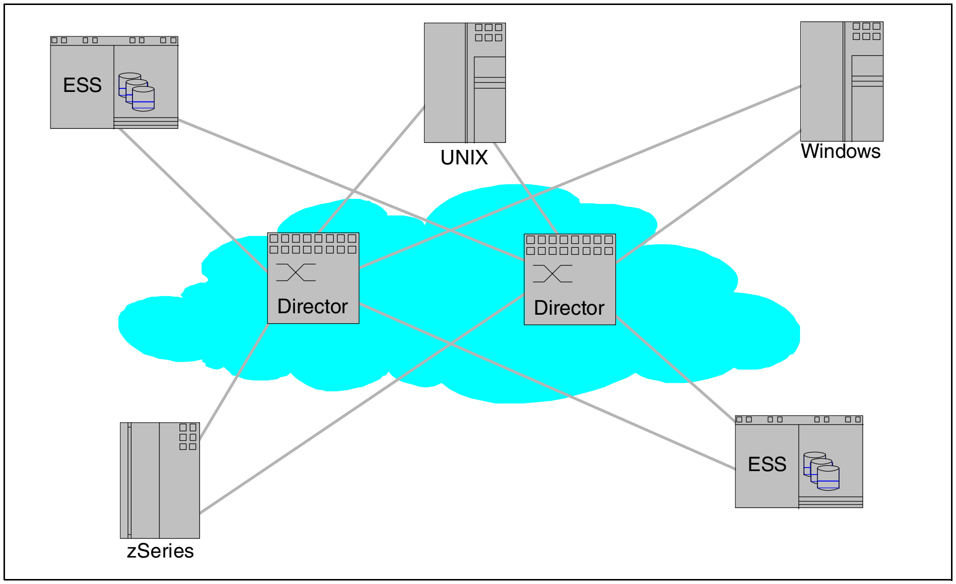


Figure 4-6 Two separate fabrics, no cascading

4.2.6 Switched fabric with cascading

A switched fabric with cascading provides for interconnections between switches, where the collection of switches looks like one large, any-to-any switch. Management becomes more extensive than basic switched point-to-point configurations. ISLs can fail and must be identified. Traffic can be routed in many ways. For technical, security, or other reasons, various levels of zoning (specifying access authority to connected ports or devices) or other mechanisms may be used to restrict the any-to-any access. Performance monitoring and configuration changes (upgrades) that are needed to keep the network performing adequately are more complex. The primary advantage of a switched cascaded fabric is that it looks like a very large logical switch, where a single connection provides access to any other port on the total set of switches, as shown in Figure 4-7.

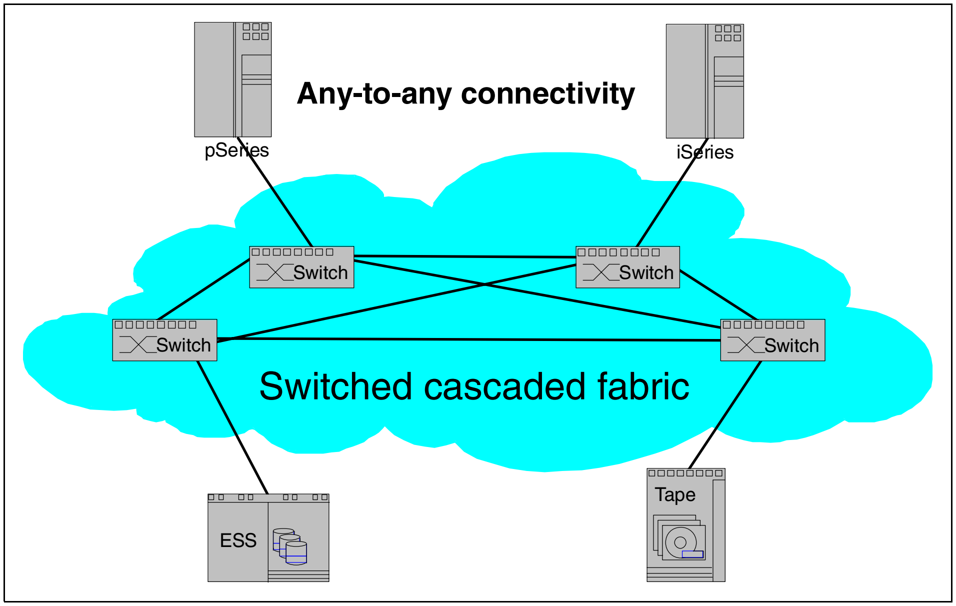


Figure 4-7 Fibre Channel switched topology with cascading

4.2.7 Hubs versus switches

An FC loop can address 126 nodes (HBAs, disk drives, disk controllers, and so on), while a switched fabric can (in theory) address up to sixteen million nodes. A loop supports a total bandwidth of 200 MB/s, where only two devices can talk at the same time, while a switch can support 200 MB/s between any two ports, with many ports transmitting or receiving data concurrently. Like a hub, a switch can connect multiple servers to one FC port on a storage subsystem. However, the connection is shared by switching the connections rapidly from one server to the next as buffered data transfers are completed, rather than by arbitrating for control of a shared loop. A hub can be thought of as a blocking device, whereas a switch is a non-blocking device.

Finally, switches enhance availability. In a pure loop configuration, a host reset or device failure, or a newly added device joining the loop can cause a loop initialization primitive (LIP) that resets all the other devices on the loop; this interrupts the loop's availability to other servers and storage devices. A switch is not affected by a LIP.

4.2.8 Mixed switch and loop topologies

To mix a switched fabric protocol and an arbitrated loop protocol, the switch must be able to detect and act on the correct protocol. This leads to the fact that hubs are only used in small environments. An example of this intermix is shown in Figure 4-8.

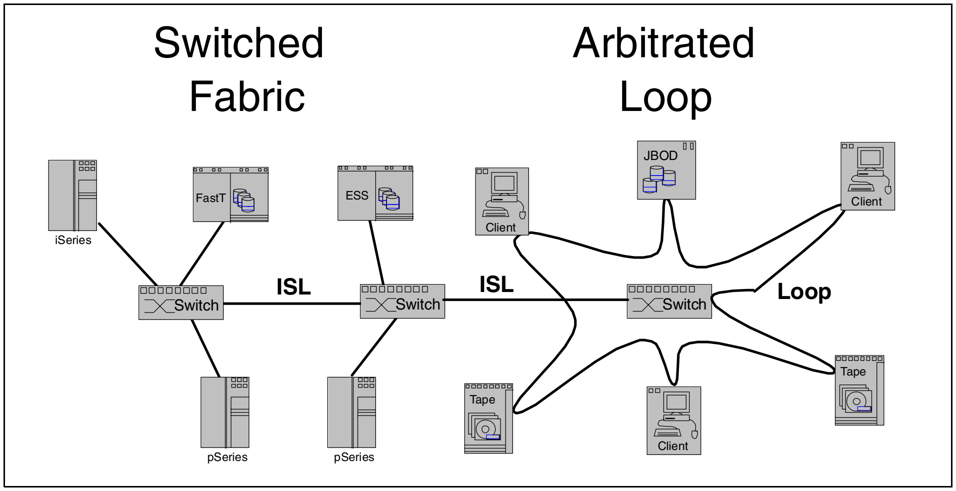


Figure 4-8 Mixed switch and loop protocols

Fibre Channel's Media Access Rules (MAR) enable systems to self-configure themselves to achieve maximum performance (even in a mixed media, copper, and optical fiber environment). A Fibre Channel port only has to manage the simple point-to-point connection between itself and the fabric. Whether the Fibre Channel topology consists of a point-to-point, arbitrated loop or switch is irrelevant, because the station management issues related to topology are not part of the Fibre Channel ports, but the responsibility of the fabric. Because Fibre Channel delegates the responsibility for station management to the fabric and employs consistent data encoding and media access rules, topology independence is achieved. It also means that Fibre Channel ports are much less complex than typical network interfaces, and as a result, less costly.

4.3 Port types

In various discussions, we will hear of different kinds of Fibre Channel ports. It is, therefore, important to understand what is meant by these different types of ports.

E_Port

An E_Port is an expansion port. A port is designated an E_Port when it is used as an inter-switch expansion port to connect to the E_Port of another switch, in order to build a larger switched fabric. These ports are found in Fibre Channel switched fabrics, and are used to interconnect the individual switch or routing elements. They are not the source or destination of information units (IU), but instead function like F_Ports and FL_Ports to relay the IUs from one switch or routing element to another. E_Ports can only attach to other E_Ports. An isolated E_Port is a port that is online, but not operational between switches due to overlapping domain ID or non-identical parameters. It is also known as an ISL port.

F_Port

An F_Port is a fabric port that is not loop capable. It is used to connect a N_Port to a switch. These ports are found in Fibre Channel switched fabrics. They are not the source or destination of IUs, but instead function only as a “middle-man” to relay the IUs from the sender to the receiver. F_Ports can only be attached to N_Ports.

FL_Port

An FL_Port is a fabric port that is loop capable. It is used to connect NL_Ports to the switch in a loop configuration. These ports are just like the F_Ports described above, except that they connect to an FC-AL topology. FL_Ports can only attach to NL_Ports.

G_Port

A G_Port is a generic port that can operate as either an E_Port or an F_Port. A port is defined as a G_Port when it is not yet connected, or has not yet assumed a specific function in the fabric.

L_Port

An L_Port is a loop capable fabric port or node. This is a basic port in a Fibre Channel Arbitrated Loop (FC-AL) topology. If a N_Port is operating on a loop, it is referred to as a NL_Port. If a fabric port is on a loop, it is known as an FL_Port. To draw the distinction, throughout this book we will always qualify L_Ports as either NL_Ports or FL_Ports.

N_Port

A N_Port is a node port that is not loop capable. It is used to connect an equipment port to the fabric. These ports are found in Fibre Channel nodes, which are defined to be the source or destination of IUs. I/O devices and host systems interconnected in point-to-point or switched topologies use N_Ports for their connection. N_Ports can only attach to other N_Ports or to F_Ports.

NL_Port

A NL_Port 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 FL_Port. These ports are just like the N_Port described above, except that they connect to a Fibre Channel Arbitrated Loop (FC-AL) topology. NL_Ports can only attach to other NL_Ports or to FL_Ports.

U_Port

A U_Port is a universal port. A generic switch port that 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.

In addition to these Fibre Channel port types, the following port types are only used in the INRANGE products.

T_Port (INRANGE specific)

A T_Port is an ISL port more commonly known as an E_Port.

TL_Port (INRANGE specific)

A TL_Port is a private to public bridging of switches or directors.

The following port type is used only in the McDATA products:

B_Port (McDATA specific)

A B_Port is a bridge port that provides fabric connectivity by attaching to the E_Port of a director. This B_Port connection forms an ISL through which a fabric device can communicate with a public loop device.

In the newest products of INRANGE and McDATA do not use their specific port types anymore. They only use port types that are in the standards.

For more detailed information about the port types, read the following IBM Redbook, *Designing and Optimizing an IBM Storage Area Network,* SG24-6419.

4.4 Addressing

All participants 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.

4.4.1 World Wide Name

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

Each N_Port will have its own WWN, but it is 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 adapters 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 for 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.

4.4.2 Port address

Because of this potential impact on routing performance, there is another addressing scheme used in Fibre Channel networks. This scheme is used to address the ports in the switched fabric. Each port in the switched fabric has its own unique 24-bit address. With this 24-bit addressing 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 fabric 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 sections that follow.

The 24-bit address scheme also 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 time of manufacture, similar to naming a child at birth. 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.

4.4.3 24-bit port addresses

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 on that port. This function of the switch is implemented by using a 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 a fabric-attached device registers their values.

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

4.4.4 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 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. This number is based on the requirements of 8b/10b running disparity between frames.

4.5 Fabric services

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

- Management services
- Time services
- Name services
- Login services
- Registered State Change Notification (RSCN)

The 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 service

This is an in-band fabric service which allows data to be passed from devices 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.

4.5.2 Time service

This is defined, but has not yet been implemented at the time of writing.

4.5.3 Name services

Fabric switches and directors implement a concept known as the name server, or Simple Name Server, or SNS. All switches and directors in the fabric keep the name server updated, and are therefore aware of all other devices in the name server. After a node has successfully logged into the fabric, it registers itself and passes on critical information such as class of service parameters, its WWN, and the Upper Layer Protocols which it can support.

4.5.4 Login service

In order to do a fabric login, a node communicates with the login server.

4.5.5 Registered State Change Notification

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

4.6 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. Certainly, all of the switches and directors in the IBM portfolio implement and utilize FSPF.

4.6.1 What is FSPF?

FSPF keeps track of the links on all switches and directors in the fabric, and associates a cost with each link. At the time of writing, the cost is always calculated as being directly proportional to the number of hops.

The protocol computes paths from a switch to all the other switches and directors in the fabric by adding the cost of all links traversed by the path, and choosing the path that minimizes the cost.

4.7 FICON connections using directors

Since both INRANGE and McDATA support FICON on their directors, we can use the 2032 and 2042 for FICON connections. Non-S/390 servers and storage systems can also be attached. This is known as a *mixed* environment — the current restriction for FICON connections are that you cannot cascade directors. This is because of the addressing mechanism used on the mainframe. From the the first quarter of 2003, FICON over more than one director will be announced.

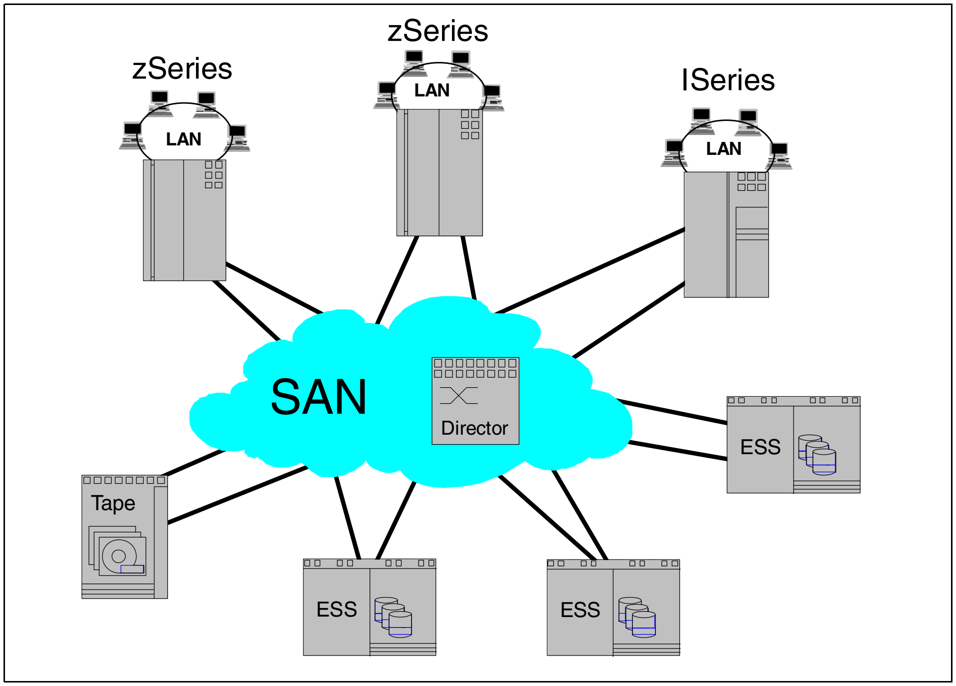


Figure 4-9 FICON switched SAN

4.7.1 Routers and bridges

Fibre Channel routers and bridges provide functions similar to traditional (Ethernet, TCP/IP, and FDDI) network routers and bridges. They are needed primarily as migration aids to the Fibre Channel SAN environment. For example, they enable existing parallel SCSI devices to be accessed from Fibre Channel networks.

Routers provide connectivity for many existing I/O busses to one or more Fibre Channel interfaces, whereas bridges provide an interface from a single I/O bus to a Fibre Channel interface. For host systems lacking Fibre Channel interfaces, bridges and routers enable access to Fibre Channel arbitrated loop environments, or point-to-point environments.

An example of this would be the IBM SAN Data Gateway component which allows up to four Ultra SCSI back-end interfaces, and up to six front-end Fibre Channel interfaces, as shown in Figure 4-10.

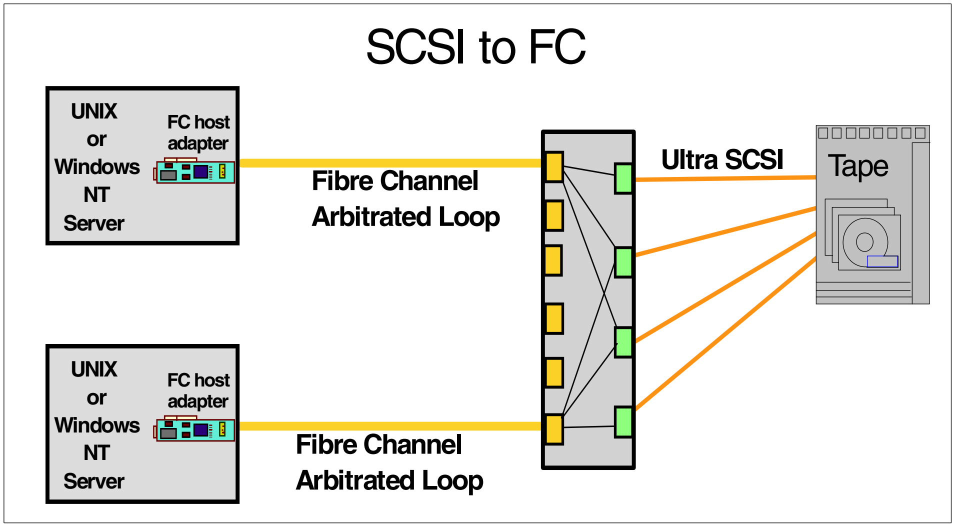


Figure 4-10 IBM SAN Data Gateway

SAN Data Gateway for serial disk

To connect serial disk (SSA) to a SAN, you need a special gateway that transforms the FC protocol into the SSA protocol, and vice versa. The IBM TotalStorage SAN Controller 160 will do this as shown in Figure 4-11.

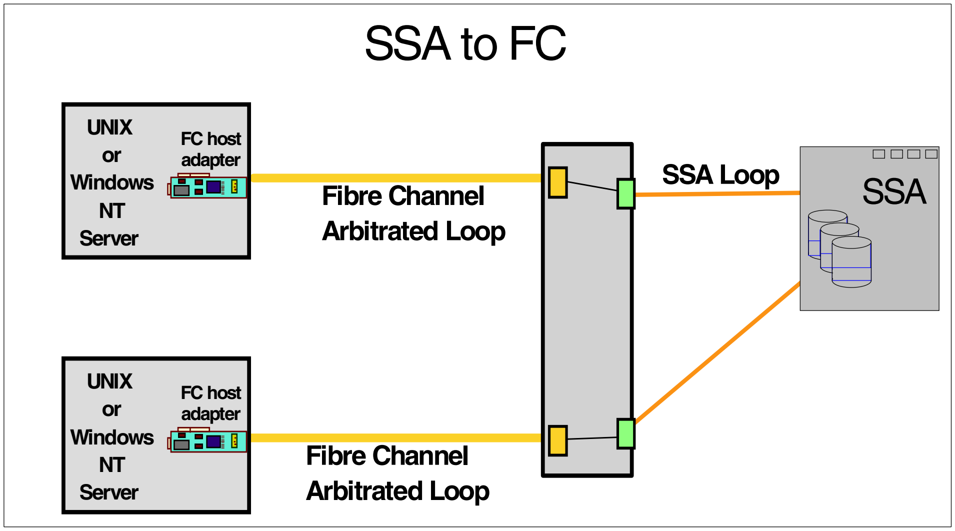


Figure 4-11 IBM TotalStorage SAN Controller 160

Extended distance gateways

In the SAN environment, an extended distance gateway component (for example, an optical extender or a Dense Wave Division Multiplexing (DWDM) multiplexer), can connect two different remote SANs with each other over a wide area network (WAN). Other terms for WAN are metropolitan area network (MAN), or dark fiber, but they all mean the same thing. Cable capacity can be bought from a cable company to go over long distances to your remote site.

An example of this is shown in Figure 4-12. A typical application requiring extended distance gateways would be remote copy solutions.

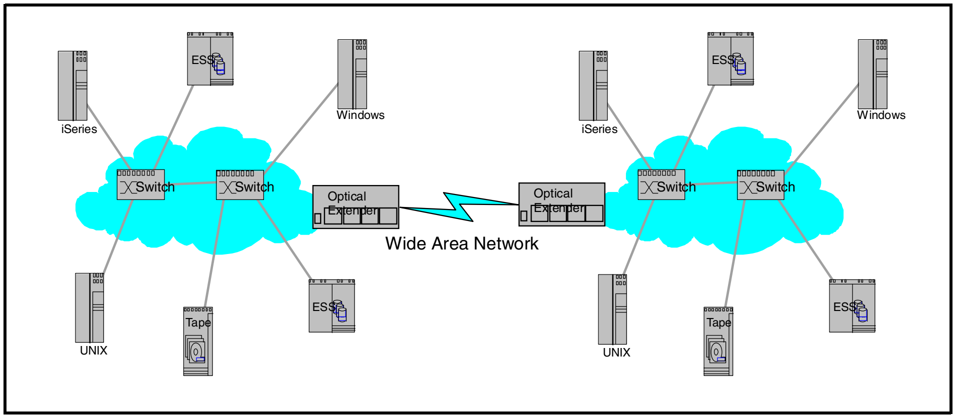


Figure 4-12 Extenders connecting remote SANs

Multiplexing signals over long distance

The cost of an extended distance solution can be very high, depending on the country and area you are in. You can buy the bandwidth and the type of protocol from a dark fiber provider, but it is also possible to lease a fiber-optic cable from one site to another. Besides using it for your SAN solution, you might also use it for the other protocols you are using such as ATM, a-sync, gigabit Ethernet, and so on. The solution for this is called DWDM.

Some of the benefits of DWDM are:

- Reduce the number of fiber-optic strands required for connectivity
- Provides channel extension up to 100 km distance range
- Up to 32 channels at 2.5 Gb/s per channel today
- Mix and match any protocol: ESCON, FICON, Fibre Channel, ATM, gigabit Ethernet, Parallel Sysplex, and so on (also known as Protocol Independent or Protocol Agnostic)
- Lower cost, more flexibility

- Multiple channels transmitted over a single fiber-optic pair

In Figure 4-13 we show the DWDM concept.

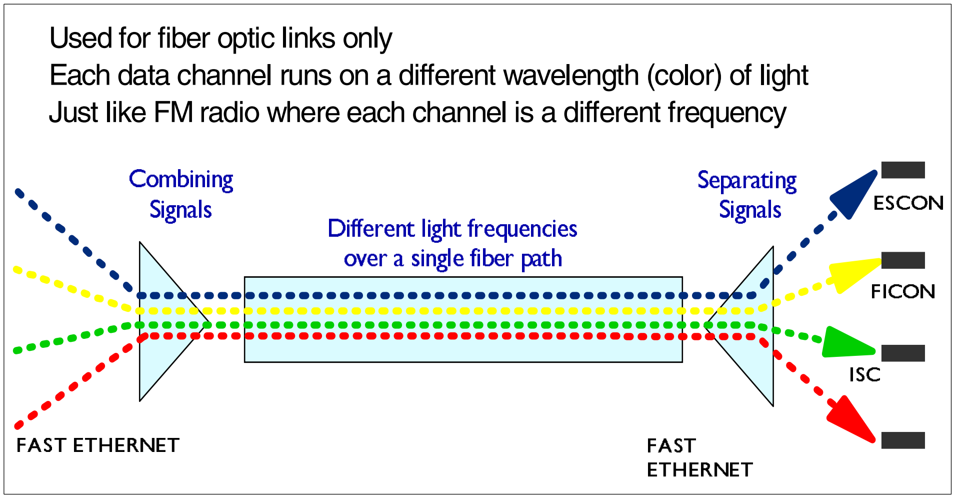


Figure 4-13 DWDM concepts

4.8 Cables and connections

The specifications for the modules, connectors, and the cables are described in the SAN standard. For 1 Gb/s, the modules in the switches and directors are called GBICs, and for 2 Gb/s the modules are called Small Form factor Plugable (SFP). There are several types of GBICs and SFPs available. There are shortwave, longwave, and extended distance modules. These modules all use their own specific kind of cable. See 1.4.3, “SAN interconnects” on page 12 for the specifications and details of the cables and connections.

4.9 Interoperability and certification

There are a lot of vendors on the SAN market that provide products like HBAs, switches, directors, extenders, gateways, hubs, disk, and tape subsystems. In IBM we have several interoperability labs where we test all kinds of combinations. A typical list of combinations to test for an ESS are:

- Server hardware
- Operating systems
- HBAs
- HBA drivers

- Switches, directors, and hubs
- Microcode levels of the SAN components
- Microcode levels of the ESS

After meeting the requirements, the tested component will be added to the supported list for the ESS. Each time that an item changes in the list (hardware or microcode) these tests must be run again to keep the list up to date. This list is always available and up-to-date on the Web at:

http://www.storage.ibm.com

If a SAN environment contains only supported items, then it is fully certified for support.

Not supported does not mean that it does not work. It means that it is not tested by the interoperability lab. If you want a combination tested and certified for support, you can ask IBM through the sales channel for a Request for Price Quotation (RPQ) to support your particular requirements.

4.10 The IBM TotalStorage SAN portfolio

In this section, we will overview the IBM TotalStorage SAN components that IBM either OEMs, or has a reseller agreement for. We include some products that have been withdrawn from marketing, as it is likely that they will still be encountered.

4.10.1 Entry level Fibre Channel switches and hubs

In this section we give a brief introduction to these components.

IBM TotalStorage SAN Switch F08

This entry-level fabric switch provides 8-port, 2 Gb/s fabric switching for Windows NT/2000 and UNIX server clustering, LAN-free backup, storage consolidation, and remote disk mirroring. The entry-fabric switch does not offer zoning, which is required for participation in core/edge fabrics.

This is shown in Figure 4-14.

Figure 4-14 IBM TotalStorage SAN Switch F08

McDATA ES-1000 Loop Switch

Provides one 1 Gb/s fabric bridge port and 8-port FC-AL attachment for IBM 3590 and LTO tape devices to McDATA Directors and McDATA Fabric Switches.

This is shown in Figure 4-15.

Figure 4-15 McDATA ES-1000 Loop Switch

IBM Fibre Channel Storage Hub

Provides 7-port, 1 Gb/s loop connectivity for configuring entry-level homogeneous Windows NT/2000 server and storage server solutions.

This is shown in Figure 4-16.

Figure 4-16 IBM Fibre Channel Storage Hub

4.10.2 Gateways and routers

In this section we give a brief introduction to these components.

IBM TotalStorage SAN Controller 160

The IBM TotalStorage SAN Controller 160, enables all IBM 7133, 7131, and 3527 SSA Serial Disk Systems to attach to host systems using Fibre Channel host adapters and drivers. This allows you to protect your investment in SSA disk, while being able to create and build a SAN infrastructure.

This is shown in Figure 4-17.

Figure 4-17 IBM TotalStorage SAN Controller 160 (top)

IBM TotalStorage SAN Data Gateway

This enables the attachment of SCSI and Ultra SCSI-attached tape storage systems to Fibre Channel-enabled UNIX-based servers.

This is shown in Figure 4-18.

Figure 4-18 IBM TotalStorage SAN Data Gateway

IBM TotalStorage SAN Data Gateway Router

This enables lower cost attachment of SCSI and Ultra SCSI-attached tape storage systems to Fibre Channel-enabled UNIX-based servers.

This is shown in Figure 4-19.

Figure 4-19 IBM TotalStorage SAN Data Gateway Router

4.10.3 Midrange switches

In this section we give a brief introduction to these components.

IBM TotalStorage SAN Switch F08

With its full-fabric switch feature enabled, this provides 8-port, 2 Gb/s switching for high availability with dual redundant fabrics, and zoning required for core/edge SAN fabrics. The F08 provides UNIX and Intel-based servers and storage with fabric scalability from small workgroups to large enterprise SANs with a common management system.

This is shown in Figure 4-20.

Figure 4-20 IBM TotalStorage SAN Switch F08

IBM TotalStorage SAN Switch F16

This provides 16-port, 2 Gb/s, high availability core-to-edge SAN fabrics for UNIX and Intel-based servers and storage, scalable from small workgroups to large enterprise SANs with a common management system.

This is shown in Figure 4-21.

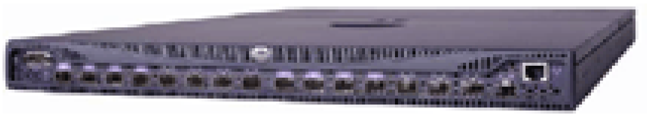


Figure 4-21 IBM TotalStorage SAN Switch F16

IBM TotalStorage SAN Switch F32

This provides 32-port, 2 Gb/s midrange and enterprise solutions designed with a common architecture and an integrated enterprise SAN management capability.

This is shown in Figure 4-22.

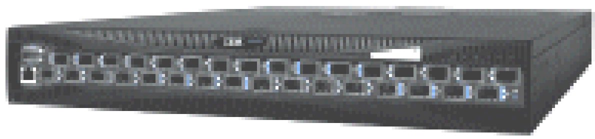


Figure 4-22 IBM TotalStorage SAN Switch F32

Cisco MDS 9216 Multilayer Fabric Switch

This is a fabric switch providing 16 to 48, 1 or 2 Gb/s auto-sensing Fibre Channel ports, with hot-swappable modular components. Virtual SAN (VSAN) hardware is available for SAN environments to provide added scalability and resilience.

This is shown in Figure 4-23.

Figure 4-23 Cisco MDS 9216 Multilayer Fabric Switch

McDATA Sphereon 4500 Fabric Switch

This is a 24-port, 2 Gb/s, scalable and affordable modular switch for departmental, core or edge switching requirements. It is McDATA's first fabric switch to provide support for IBM loop tape devices. It provides UNIX and Intel-based servers and storage with highly scalable, core-to-edge fabrics for large enterprise SANs with a common management system.

This is shown in Figure 4-24.

Figure 4-24 McDATA Sphereon 4500 Fabric Switch

McDATA Sphereon 3216 and 3232 Fabric Switches

These provide 16 and 32-port, 2 Gb/s, scalable, high availability departmental Fibre Channel switching for UNIX and Intel-based servers and storage. When combined with McDATA Directors, they provide a 2 Gb/s, high availability enterprise-to-edge switched fabric with a common management system.

These are shown in Figure 4-25.

Figure 4-25 McDATA Sphereon 3216 (top) and 3232

4.10.4 Enterprise core fabric switch and directors

In this section we give a brief introduction to these components.

IBM TotalStorage SAN Switch M12 - Core Fabric Switch

This provides one or two 64-port, 2 Gb/s switch(es) for high availability dual redundant core/edge SAN fabrics. Provides UNIX and Intel-based servers and storage with highly scalable large enterprise SANs with a common management system. The Fabric Manager feature simplifies management of complex fabrics.

This is shown in Figure 4-26.

Figure 4-26 IBM TotalStorage SAN Switch M12

Cisco MDS 9509 Multilayer Director

The Cisco MDS 9509 Director has 32 to 224 1 or 2 Gb/s auto- sensing Fibre Channel ports, director availability, and service. Virtual SAN (VSAN) hardware is available for SAN environments with added scalability and resilience.

This is shown in Figure 4-27.

Figure 4-27 Cisco MDS 9509 Multilayer Director

INRANGE IN-VSN FC/9000 Fibre Channel Director family

This provides 64, 128 and 256-port, 2 Gb/s data center class Fibre Channel switching for: IBM S/390 G5/G6 and zSeries 900 servers and IBM storage with FICON adapters; IBM and non-IBM UNIX-based and Intel-based servers, and IBM and non-IBM storage with Fibre Channel adapters; IBM FC-AL attached tape devices and IBM Ultra-SCSI-attached tape devices with gateways and routers. INRANGE Directors feature ultra scalability from 24 to 64 and 48 to 256 ports; upgrades from 64 to 128 to 256 port models; and upgrades from 1 to 2 Gb/s switching and intermix of 1 and 2 Gb/s port modules in the same director.

This is shown in Figure 4-28.

Figure 4-28 INRANGE IN-VSN FC/9000

McDATA Intrepid 6140

This provides 140-port, 2 Gb/s high availability, data center class Fibre Channel switching with 2 Gb/s director full interoperability with other members of the McDATA Director and Switch family. Supports: IBM S/390 G5/G6 and zSeries 900 servers and IBM storage with FICON adapters; IBM and non-IBM UNIX-based and Intel-based servers and IBM and non-IBM storage with Fibre Channel adapters; IBM FC-AL attached tape devices with the McDATA ES-1000 Loop Switch; and IBM Ultra-SCSI-attached tape devices with gateways and routers.

This is shown in Figure 4-29.

Figure 4-29 McDATA Intrepid 6140

McDATA Intrepid 6064

This provides 64-port, 2 Gb/s high availability, data center class Fibre Channel switching with 1 to 2 Gb/s director and switch blade upgrades. Supports: IBM S/390 G5/G6 and zSeries 900 servers and IBM storage with FICON adapters; IBM and non-IBM UNIX-based and Intel-based servers and IBM and non-IBM storage with Fibre Channel adapters; IBM FC-AL attached tape devices with the McDATA ES-1000 Loop Switch; and IBM Ultra-SCSI-attached tape devices with gateways and routers, a-SCSI-attached tape devices with gateways and routers.

This is shown in Figure 4-30.

Figure 4-30 McDATA Intrepid 6064

McDATA Fabricenter FC-512

This provides spacing saving cabinet for McDATA Directors and Switches. It provides redundant power distribution for high availability directors and space, cooling and cabling flexibility to support up to 512 Fibre Channel ports.

This is shown in Figure 4-31.

Figure 4-31 McDATA Fabricenter FC-512

> **Note:** For detailed information relating to the IBM SAN portfolio, go to the following Web site:
>
> http://www.storage.ibm.com/ibmsan/index.html

4.11 New products and trends

The SAN environment is growing fast, so new products have to be developed and implemented. The next step in speed will be 10 Gb/s. This speed will first be found in ISL connections and the forecast for limited implementation is the second half of 2003. In time it is expected that servers and disk subsystems will be able to handle up to 10 Gb/s speed. It is unlikely that tape subsystems will move up to that speed. At this moment a single tape drive can handle 30 MB/s maximum so 200 MB/s will be quick enough for the next years.

Beside the speed there is also the need for other protocols to use the SAN environment. In the near cards, or modules, will be provided for switches and directors that will support TCP/IP, fast Ethernet, Gigabit Ethernet and iSCSI. These protocols will also un in an intermixed environment with FCP and FICON.

In the first quarter of 2003 the mainframe will be ready with their FICON addressing changes to be able to use a cascading environment with SAN directors.

One of the next steps will be storage virtualization. This will transfer more intelligence into the SAN itself. The servers will not know where the data is stored, nor will they be able to directly access storage any longer; the intelligence, which includes the ability to access any kind of storage will move into the SAN. This will be achieved by devices whose sole purpose is to act as a SAN volume control engine.

Some of the intelligence that will move into the storage network are:

- Storage partitioning and LUN masking
- Dynamic volume extension
- Single point of management for JBODs
- Box-to-box Copy Services like FlashCopy, remote mirroring and PPRC
- Large R/W Cache
- Enterprise reliability

5

SAN management

SAN management has a critical role in any SAN solution, and can also be a deciding factor in the selection of a vendor's SAN solution implementation.

5.1 Is SAN management required?

Management capabilities for any device require additional circuitry, microcode and application software, which raises the cost, and therefore necessitates a trade-off between cost and functionality. The more complex a SAN configuration, the greater the need for manageability. But, the cost of hardware and software is not the issue; it is the total cost of management, cost for human resources, and the difficulty to achieve a sufficient skill level in order to give an efficient and quick response to any event that occurs in a SAN.

For example, a large 10-to-50 node SAN (which includes hosts, SAN devices and storage devices) may have HBAs, SAN devices and storage arrays from a variety of vendors. The storage network may be running a server cluster application. The cabling infrastructure may include a mix of copper as well as multi-mode and single-mode fiber. The FC connections may be transporting IP, FCP, and FICON protocols. All these factors add complexity to the SAN, and therefore, make management an essential part of a SAN vendor selection.

One of the biggest challenges facing SAN vendors is to create and agree upon a common standard for SAN management that enables the ultimate goal of an any-to-any, centrally managed, heterogeneous SAN environment.

5.2 SAN management, standards and definitions

The SAN management architecture can be divided into three distinct levels at which to manage. These are the:

- SAN storage level
- SAN network level
- Enterprise systems level

In Figure 5-1 we illustrate the three management levels in a SAN solution.

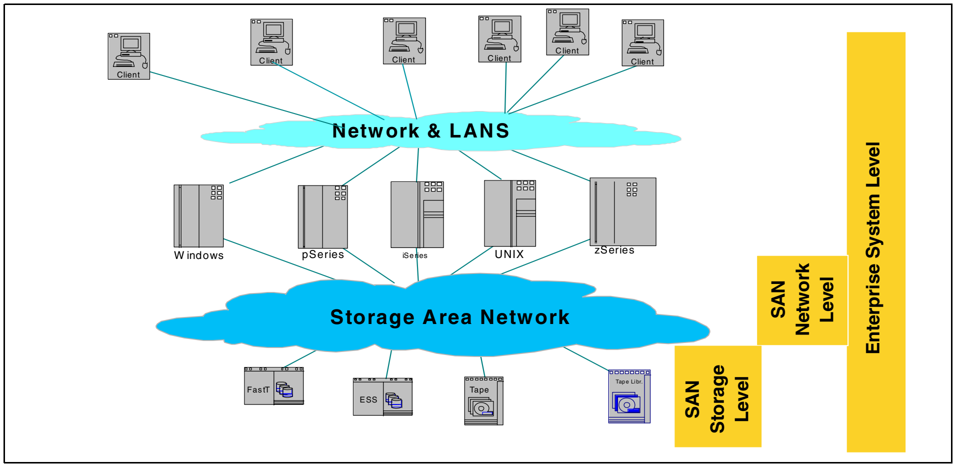


Figure 5-1 SAN management levels

5.2.1 SAN storage level

The SAN storage level is composed of various storage devices 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 spans both storage resources and servers as shown in Figure 5-1.

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.

SCSI Enclosure Services

In SCSI legacy 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.

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. For more information about SNMP see "Simple Network Management Protocol" on page 89.

5.2.2 SAN storage management tools

In the sections that follow, we describe some of the SAN storage device management and reporting tools.

IBM StorWatch Enterprise Storage Server Specialist

The IBM TotalStorage Enterprise Storage Server Specialist allows administrators to configure, monitor, and centrally manage the IBM Enterprise Storage Server remotely through a Web interface.

IBM TotalStorage Enterprise Tape 3494 Library Specialist

The IBM TotalStorage Enterprise Tape 3494 Library Specialist allows administrators to monitor and centrally manage the IBM Enterprise Storage

Server remotely through a Web interface including the TotalStorage Virtual Tape Server in a P2P configuration.

IBM TotalStorage LTO Tape library Specialist

The IBM TotalStorage LTO Tape Library Specialist allows administrators to monitor, and centrally manage all of the libraries in the LTO Tape Library Family remotely through a Web interface.

IBM TotalStorage ETL Expert

The IBM TotalStorage ETL Expert:

- Provides a real-time status on key measurements concerning the Tape Library and Virtual Tape Server
- Easy to access performance metrics that include information on managing Tape Volume Cache, logical volume fragmenting, and throttling
- Monitor and track the performance of Enterprise Storage Servers
- LUN to host disk mapping providing complete information on the logical volumes including the physical disks and adapters
- Customization and enablement capability for threshold events relating to ESS performance and utilization
- SNMP alerts for ESS Expert exception events that exceed threshold values

IBM StorWatch Enterprise Storage Server Specialist Expert

The IBM StorWatch Enterprise Storage Server Specialist Expert:

- Helps optimize the performance of IBM tape and disk subsystems when used in conjunction with the IBM TotalStorage Enterprise Tape Library Expert
- Hosts disk mapping via a LUN, providing complete information on logical volumes including physical disks and adapters
- Customizes and enables threshold events relating to ESS performance and utilization
- Uses SNMP alerts for ESS exception events that exceed threshold values
- Provides statistics to help manage the performance of the IBM TotalStorage Enterprise Storage Server
- Tracks total capacity, used capacity and free capacity for all connected ESS disk storage in the enterprise
- Provides information about systems, such as z/OS and UNIX that share volumes within the ESS.
- Helps centralize management of multiple remote systems located throughout the enterprise

FAStT Storage Manager

The FAStT Storage Manager provides:

- Intuitive centralized administration/management tool
- Manage multiple FAStT from a single console
- Managing and configuring FlashCopy and mirroring
- Dynamic volume expansion and configuration
- Error reporting and diagnostic tool for subsystem components
- Performance monitoring
- Storage partitioning

5.2.3 SAN network level

The SAN network level is composed of all the various components that provide connectivity, like the SAN cables, SAN hubs, SAN switches, inter switch links, SAN gateways, and HBAs.

This SAN network level becomes closely tied to inter-networking infrastructures such as those seen in local area network (LAN) and wide area network (WAN) solutions. The hubs, switches, gateways, and cabling, used in LAN/WAN solutions, use the SNMP support feature to provide management of all these networking components. Most SAN solution vendors also mandate SNMP support for all of their SAN networking components, which is then used and exploited by SNMP based network management applications to report on the various disciplines of SAN management.

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 wishes to include additional device information that is not specified in a standard MIB, then that is usually done through MIB extensions.

Application Program Interface

As we know, there are many SAN devices that come from many different vendors, and every one 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.

APIs (Application Program Interface) are a good help about it. Some vendors make the API of their product available for other vendors in order to make it possible for common management in the SAN.

Fabric monitoring and management is an area where a great deal of standards work is being focused. Two management techniques are in use: in-band and out-of-band management.

In-band management

Device communications to the network management facility are most commonly done directly across the Fibre Channel transport, using SES. This is known as in-band management. It is simple to implement, requires no LAN connections, and has inherent advantages, such as the ability for a switch to initiate a SAN topology map by means of SES 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 developments: 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 make additional TCP/IP connections.

Out-of-band management

Out-of-band management means that device management data are gathered over a TCP/IP connection such as Ethernet. Commands and queries can be sent using SNMP, Telnet (a text-only command line interface), or a Web browser Hyper Text Transfer Protocol (HTTP). Telnet and HTTP implementations are more suited to small networks.

Out-of-band management does not rely on the Fibre Channel 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, especially by using SNMP. However, unlike in-band management, it cannot automatically provide SAN topology mapping.

- **Management Information Base (MIB):** A management information base (MIB) organizes the statistics provided. The MIB runs on the SNMP management workstation, and also on the managed device. A number of industry standard MIBs have been defined for the LAN/WAN environment. Special MIBs for SANs are being built by SNIA. When these are defined and adopted, multi-vendor SANs can be managed by common commands and queries.
- **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.
- **SNMP:** This protocol is widely supported by LAN/WAN routers, gateways, hubs, and switches, and is the predominant protocol used for multi-vendor 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. This usually runs on a UNIX or NT workstation attached to the network. 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, and messages can be sent to the network operator.

Element management is concerned with providing a framework to centralize and automate the management of heterogeneous elements, and to align this management with application or business policy.

One of the big differences in SAN management tools is their possibility to operate either in-band or out-of-band, or both.

In-band advantages

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 advantages

Out-of-band management using Ethernet has three 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.

In a SAN, we will typically encounter both of these methods.

5.2.4 SAN network management tools

In the following sections we describe some of the SAN network management tools available today.

IBM TotalStorage SAN Data Gateway Specialist

Based on a Web interface, the StorWatch SAN Data Gateway Specialist provides configuration, gateway discovery, asset management, device mirroring, disk copy management, and device monitoring for IBM SAN Data Gateways for serial disk.

IBM TotalStorage SAN Switch Specialist

Based on a Web interface, the StorWatch SAN Switch Specialist provides configuration, switch and devices discovery, asset management, traffic, alerts and device monitoring for IBM SAN 2109 family switches.

Fabric Manager

Fabric Manager provides a graphical interface that allows the administrator to monitor and manage a fabric from a standard workstation. Fabric Manager can be used to manage fabrics containing the IBM 2109 and 3534 family of switches.

Fabric Manager provides high-level information about all switches in the fabric, and launching the IBM TotalStorage SAN Switch Specialist application when more detailed information is required. The launching is transparent, providing a seamless user interface. Fabric Manager provides improved performance over the IBM TotalStorage Specialist alone.

Fabric Manager provides the following information and capabilities:

- Monitoring and management of the entire fabric.
- The status of all switches in the fabric.
- Access to event logs for the entire fabric.
- Zoning functions.
- Loop diagnostics and query and control of loop interfaces to aid in locating faulty devices.
- Ability to name and zone QuickLoops.
- Access to the name server table.
- Telnet functions.
- Switch beaconing for rapid identification in large fabric environments.

For more information about this product go to:

http://www.brocade.com

Cisco manager

Cisco MDS family provides three principal modes of management: the Cisco MDS 9000 Family command-line interface (CLI), Cisco Fabric Manager and integration with third-party storage management tools.

Cisco presents the user with a consistent, logical CLI. Adhering to the syntax of the widely known Cisco IOS CLI, the Cisco MDS 9000 Family CLI is easy to learn and has broad functionality. It is an extremely efficient and direct interface, designed to provide optimal functionality to administrators in enterprise environments.

Cisco Fabric Manager is a responsive, easy-to-use Java application that simplifies management across multiple switches and fabrics. It enables administrators to perform vital tasks such as topology discovery; fabric configuration; and verification, provisioning, monitoring, and fault resolution. All functions are available through an interface, which enables remote management from any location.

Cisco Fabric Manager may be used independently or in conjunction with third-party management applications. Cisco provides an extensive application programming interface (API) for integration with third-party and user-developed management tools.

For more information about this product go to:

http:// www.cisco.com/go/ibm/storage

INRANGE IN-VSN Enterprise Manager

The IN-VSN Enterprise Manager (management system) centralizes the management of multiple distributed directors of the IBM 2042 FICON or Fibre Channel directors in an enterprise-wide Fibre Channel fabric backbone and continuously monitors their operations. Designed for usability, this flexible management system helps simplify the addition of managed directors as the enterprise Fibre Channel fabric grows. Utilizing a Java application graphical user interface (GUI), operators can use the IN-VSN Enterprise Manager locally or remotely from LAN-attached IN-VSN Enterprise Manager Clients.

The IN-VSN Enterprise Manager provides a flexible architecture consisting of the IN-VSN Enterprise Manager Server with the IN-VSN Enterprise Manager application and multiple IN-VSN Enterprise Manager Clients. One or more directors can be controlled from the same control node if they are connected through Fibre Channel to the director that is in turn connected to the IN-VSN Management System via Ethernet.

The Enterprise Manager application provides continuous director monitoring, logging and alerting; centralizes log files, configuration databases, and firmware distribution; and supports centralized *call-home*, pager alert, service and support operations.

For more information about this product go to:

`http://www.inrange.com`

or

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

McDATA Enterprise Fabric Connectivity Management

The MCDATA Enterprise Fabric Connectivity Management (EFC) Management software is designed to centralize the management of multiple, distributed switches and directors, including the IBM/McDATA 6064 and 6140 family in an enterprise-wide Fibre Channel fabric. Designed for usability, this flexible software simplifies the addition of managed directors as the enterprise Fibre Channel fabric grows. Simplified operation of multiple FICON directors and Fibre Channel directors is provided by this integrated switch management system. Leveraging Java technology, operators can use the EFC Management software remotely from anywhere.

EFC management provides a scalable, modular architecture consisting of the EFC Server PC, the EFC Management Software, and the EFC Product Manager application.

For more information about this product go to:

`http://www.mcdata.com`

SANsurfer Tool Kit

The SANsurfer Tool Kit, included at no additional charge with all QLogic HBAs, gives you everything you need to configure and manage a QLogic SAN fabric from a single interface. SANsurfer Tool Kit handles every aspect of SANbox switch and SANblade HBA management including drivers, firmware installation and upgrade.

For more information about this product go to:

`http://www.qlogic.com`

Emulex HBAnyware Management Software

This is a centralized host bus adapter management suite. HBAnyware is a centralized HBA management suite that simplifies SAN management and lowers TCO. HBAnyware incorporates driver-based technology to enable complete management of Emulex HBAs, including the ability to upgrade firmware anywhere in a Fibre Channel or iSCSI SAN from a single console. Emulex's HBAnyware complements third party management applications and enables cost-effective and scalable management that is critical to the efficient operation of enterprise-class SANs. Additionally, HBAnyware leverages Emulex's unique architectural capabilities, including firmware upgradeability and driver compatibility across product generations, to further reduce planned downtime and improve IT management productivity.

For more information about this product go to `http://www.emulex.com`

EZ Fibre

EZ FIbre is a configuration utility for JNI Host Bus Adapters. JNI's EZ Fibre Configuration Utility is the fastest and easiest way to install and configure JNI HBAs in Windows, Mac, or Solaris environments. Featuring a point-and-click graphical user interface, EZ Fibre allows a network administrator to go through every step of the installation and configuration process while avoiding complicated operating system database or configuration file editing.

For more information about this product go to `http://www.jni.com`

FINISAR SAN performance tools

Building on a history of leading-edge innovation and expertise, Finisar increases network performance through intelligent test and monitoring tools. Finisar engineers flexible solutions that bridge the gap from SAN to LAN.

For more information about this product go to http://www.finisar.com

5.2.5 Enterprise systems level

The enterprise systems level essentially ensures the ability to have a single management view and console.

Enterprise systems management applications have to integrate the management data, and then present it in an easy-to-understand format. This management data comes from various layers of the enterprise infrastructure, including storage, networks, servers, and desktops, often with each of them running its own management application.

In addition, the rapid deployment of Internet-based solutions in the enterprise has necessitated Web-based management tools. Various industry-wide initiatives, like 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 some level of management standardization, and many SAN solutions vendors are deploying these standards in their products.

In heterogeneous management solutions, CIM enables data integration of management applications.

We overview some of those initiatives in the following sections.

Web-Based Enterprise Management

Web-Based Enterprise Management (WBEM) is focused on enterprise management. It is being developed by the Desktop Management Task Force (DMTF), in which Sun participates. WBEM is based on an information model called the Common Information Model (CIM), and on a protocol based on the eXtensible Markup Language (XML). This protocol is still being defined. WBEM does not define APIs, and thus does not facilitate implementation.

Common Information Model

The Common Interface Model (CIM) enables distributed management of systems and networks. It facilitates a common understanding of management data across different management systems, and also facilitates the integration of management information from different sources.

Application Program Interface

An application program interface (API), sometimes referred to as an *application programming interface*, is the specific method prescribed by a computer operating system or by an application program by which a programmer writing an application program can make requests of the operating system, or another application.

In other words, an API is a sort of a “go-between”, or tool by which a programmer can accomplish their task, without having specific knowledge of the way a computer works.

Java Management API

Java Management API (JMAPI) provides a specification for the management extension of the Java platform. This specification is being drawn up by a number of selected management experts according to the new Java open standardization process.

JMAPI addresses the management of all Java computing environments; not only in the enterprise, but also in the telco, consumer and datacom markets. JMAPI is protocol- and information model-independent. It provides APIs to facilitate implementation. It is based on Java, which makes portability a non-issue.

JMAPI provides for integration with existing management systems such as SNMP. It can interoperate with CIM/WBEM and can provide APIs for CIM/WBEM.

Desktop Management Interface

The Desktop Management Interface (DMI) is a set of interfaces and a service provider that mediate between management applications and components residing in a system.The DMI is a free-standing interface that is not tied to any particular operating system or management process.

By using DMI, you may manage various elements within most systems (for example, PCs, workstations, routers, hubs, and other network objects).

5.2.6 SAN enterprise level management tools

In the following sections we describe some of the SAN enterprise level management tools available today.

IBM Tivoli Storage Area Network Manager

IBM Tivoli Storage Area Network Manager is a solution that discovers, monitors and manages SAN fabric components. Tivoli SAN Manager is architected to ANSI SAN standards, allowing you to choose best of breed products for the storage infrastructure.

Some of the areas that IBM Tivoli Storage Area Network Manager addresses are:

- Reduce storage administration costs
- Reduce administrative workloads
- Maintain high availability
- Minimize downtime

For more information about this product go to:
http://www.tivoli.com

VERITAS SANPoint Control

With SANPoint Control, a SAN administrator has a single, centralized, consistent storage management interface to simplify the complex tasks involved in deploying, managing, and growing a multi-vendor networked storage environment. SANPoint Control seamlessly integrates performance and policy management, storage provisioning and zoning capabilities. In addition, SANPoint Control offers customizable policy-based management to automate notification, recovery, and other user-definable actions.

The three areas that SANPoint Control addresses are:

- Centralized storage resource and infrastructure management
- Integrated enterprise and infrastructure application awareness
- Real-time performance and capacity management with integrated reports

For many information about this product go to:
http://www.veritas.com

VIXEL SAN InSite Software

Based on Java, it provides graphical visibility and control over the SAN by allowing the operator to manage and control local and remote SANs. Proactive SNMP management functions allow the operator to deploy Fibre Channel enhanced servers and storage.

SAN Insite's SNMP traps support alerts, system managers, and other SNMP management software packages produced by Tivoli, Veritas, HP, CA, and Legato.

For more information about this product go to:
http://www.vixel.com

5.3 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 obtain potentially better performance, and to aid recovery in case of adapter, cable, switch or GBIC failure.

In Figure 5-2 we show a typical configuration in a core-edge SAN environment from a high level view.

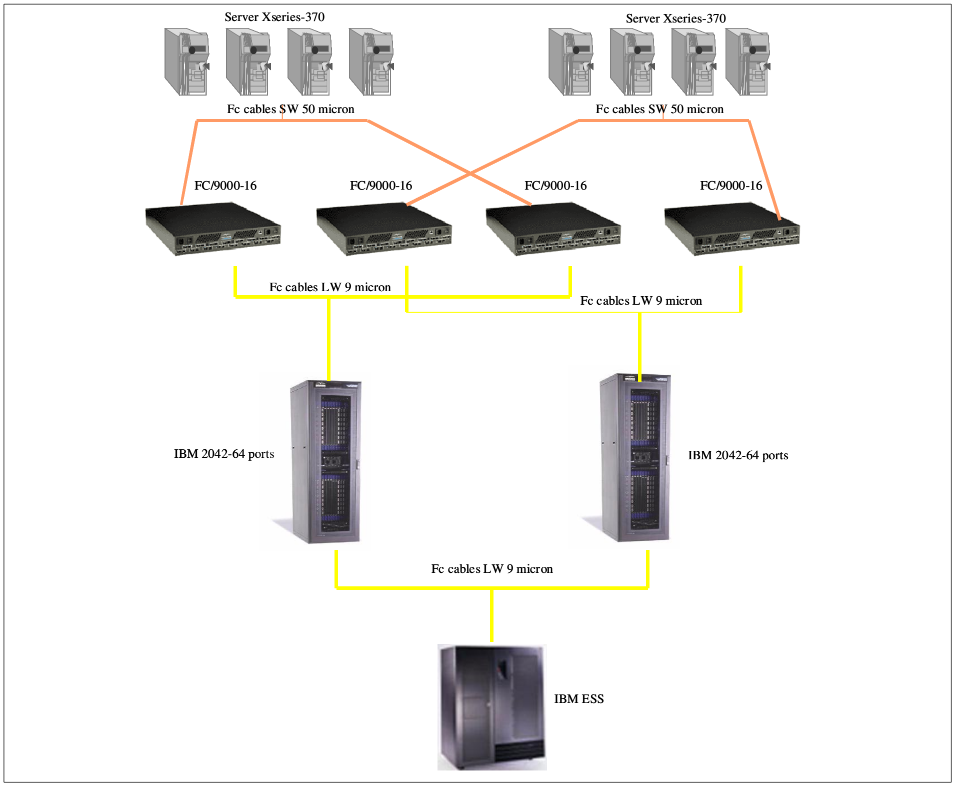


Figure 5-2 Core-edge SAN environment

In Figure 5-3 we show this in more detail.

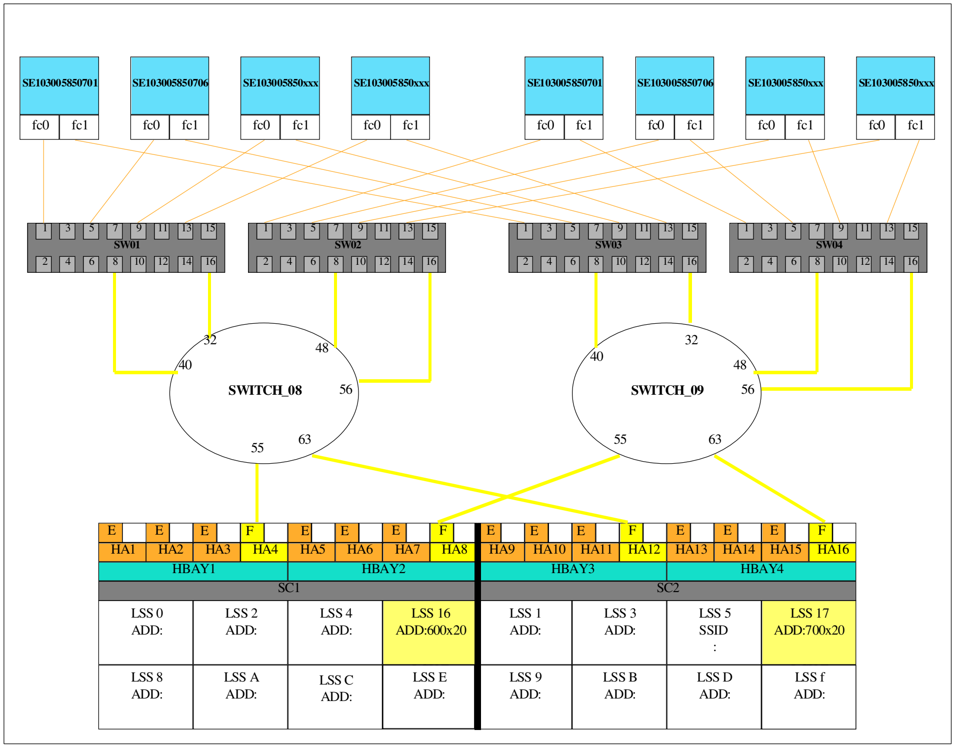


Figure 5-3 Core-edge SAN environment details

In this case, the same 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 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 coming from different vendors, and the following sections represent some of them:

IBM Data Path Optimizer

The IIBM Data Path Optimizer (DPO) provides the ability to dynamically switch paths, if multiple paths are assigned, in Windows and UNIX environments providing high availability, and also providing a load balancing algorithm that can potentially enhance performance and throughput. It was used in SCSI environments and has been superseded by the IBM Subsystem Device Driver (SDD) for all ESS models.

IBM Subsystem Device Driver

The IBM Subsystem Device Driver (SDD) is a pseudo device driver designed to support the multipath configuration environments in the IBM ESS. It resides in a host system with the native disk device driver, and provides the following functions:

- Enhanced data availability
- Dynamic I/O load-balancing across multiple paths
- Automatic path failover protection
- Concurrent download of licensed internal code
- Path-selection policies for the host system

IBM Redundant Disk Array Controller

The IBM Redundant Disk Array Controller (RDAC) is a Fibre Channel driver that implements path failover and load balancing on many platforms, and it is used in conjunction with the IBM TotalStorage FAStT Server Family.

VERITAS Volume Manager™ & Dynamic Multipathing

Volume Manager is a management tool for heterogeneous enterprise environments, which reduces planned and unplanned downtime. Volume Manager ensures high availability of data, optimized I/O performance, and offers freedom of choice in storage hardware investments. Volume Manager supports Solaris, HP-UX, AIX, Linux, Windows.

Dynamic Multipathing (DMP) support eliminates a single point of failure by providing multiple I/O paths into the array.

For more information about this product go to:
`http://www.Veritas.com`

5.4 SAN security

When designing or managing a SAN, both physical and logical security is one of the most important issues.

We do not intend to cover the physical issues associated with creating a secure environment. In this redbook, which serves as an introduction to security, we divide security in two branches:

- Access control security
- Data security

Access control security

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 competent 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.

Data security

This is a security and integrity requirement aiming to guarantee that data from one application or system does not become overlaid or corrupted by other applications or systems. This may involve the authorization and ability to fence off one system’s data to other 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. True cross-platform data sharing solutions, as opposed to data partitioning solutions, are also a requirement. Security and access control needs to be improved to guarantee data integrity.

In a SAN environment, data integrity and security is guaranteed by zoning and LUN masking.

Zoning

Zoning allows for finer segmentation of the switched fabric. Zoning can be used to instigate a barrier between different environments. Only members of the same zone can communicate within that zone, and all other attempts from outside are rejected. 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 can easily separate your test environment from your production bandwidth allocation on the same fabric using zoning.

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.

5.5 Troubleshooting

As SANs become more and more complex, troubleshooting can become an issue in a large fabric. Or even a small one.

5.5.1 Problem determination and problem source identification

Problem determination and problem source identification (PD/PSI) is an issue that needs to be considered when, or even before, a SAN is implemented.

There are many tools to collect the necessary data in order to perform a PD/PSI. These tools are typically the same management tools already discussed in 5.2.2, “SAN storage management tools” on page 87.

Often these tools, especially at the SAN network level, are as manyfold as there are devices and vendors.

At a minimum, a SAN troubleshooter should know how to:

- Interrogate the fabric
- Find devices or LUNs that have gone missing
- Identify failing cables
- Interpret message and error logs
- Know where to start troubleshooting from
- How to record information in an orderly fashion
- Create a diagram of the SAN
- Access the name server
- Access port and GBIC information
- Understand switch LEDs
- Use vendor specific diagnostic tools
- Find error messages on the host or storage devices
- Check for zoning conflicts

Advanced troubleshooters may require access to a Fibre Channel analyzer.

Most troubleshooters would like to have only one tool that is capable of collecting all the necessary information. This would be done automatically in a fast and simple way, and then formatted into an easy to understand mode for analysis. However, until such a tool exists, it is much more of a manual process.

The management tools discussed previously will help, are good, but they may not be as comprehensive as some users would like. It is likely that in the future the SAN components will be as “easy” to debug as it is on the mainframe today — by that we mean that performance and error datum is typically automatically stored for later analysis, or simple traps can be set to prompt for action.

Some of the best advice has already been stated and that is to remember that a good SAN begins with a good design and is well documented. This is undoubtedly of considerable help when any form of error occurs, and will shorten the time to problem resolution.

In terms of PD/PSI, what we mean by a good design is that configuration design information is understandable, available at any support level, and is always updated with the latest configuration. There is also a data base kept with respect to information about connections, naming conventions, device serial numbers, WWN, zoning, system applications, and so on, These should be recorded and kept up to date.

6

SAN exploitation and 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 the IBM Redbook *Designing and Optimizing an IBM Storage Area Network*, SG24-6419.

6.1 The SAN toolbox

A SAN is often depicted as a cloud with connecting lines from the cloud to servers and storage devices. This is a high level representation of a SAN. We need a way to represent the components that make up the cloud. A simple and effective way to represent the various components of a SAN configuration is shown in Figure 6-1, the SAN toolbox.

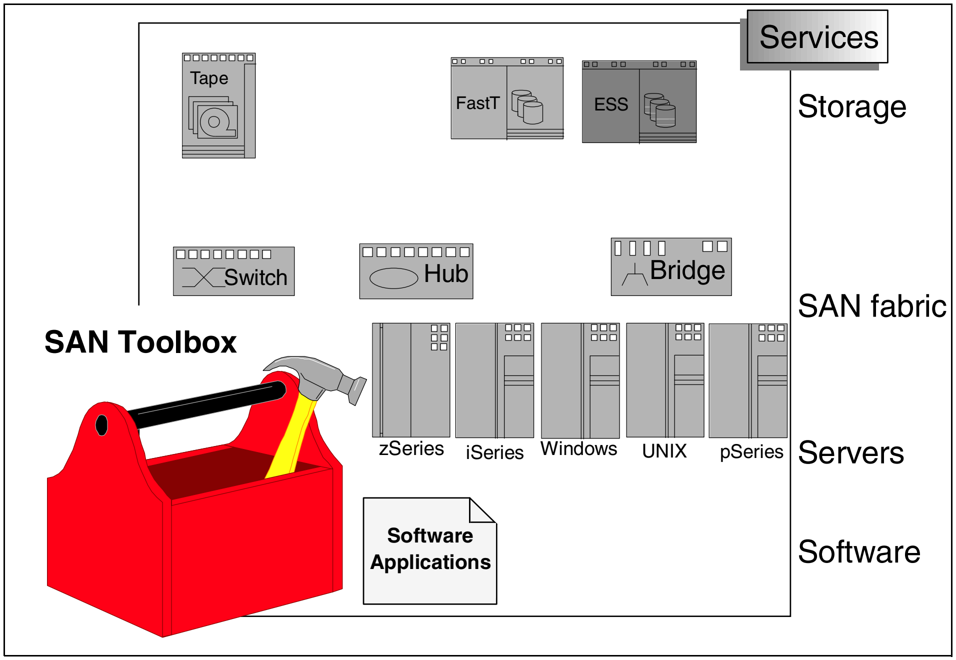


Figure 6-1 SAN toolbox

The toolbox contains all the necessary building blocks, or tools, to construct a SAN. We will use the icons shown in the toolbox in most of the figures in this chapter.

There are five basic SAN building blocks.

- *Services*: Services is a term that can be applied to support, education, consultation, and so on.
- *Storage elements*: The storage elements include tape drives and libraries, disk storage, optical storage, and intelligent storage servers (defined as storage subsystems, each having a storage control processor, cache storage, and cache management algorithms).
- *SAN fabric*: The SAN fabric is built from interconnecting elements such as FC hubs, FC switches, routers, bridges, and gateways. These components transfer Fibre Channel packets from server to storage, server to server, and storage to storage, depending on the configurations supported and the functions used.
- *Servers*: The servers are connected to the fabric. Applications on these servers can take advantage of the SAN's benefits.

- *Software*: There are two kinds of storage management applications: fabric management applications used to configure, manage, and control the SAN fabric; and applications that exploit the SAN functions to bring business benefits such as improved backup/recovery and remote mirroring.

6.2 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.

6.3 Resource pooling solutions

Before SANs, the concept of 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. When hardware is installed or needs to be moved, you can be assured that your physical infrastructure already supports it.

6.3.1 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.

6.3.2 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 NT, 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 6-2 shows an example of disk storage pooling across two servers.

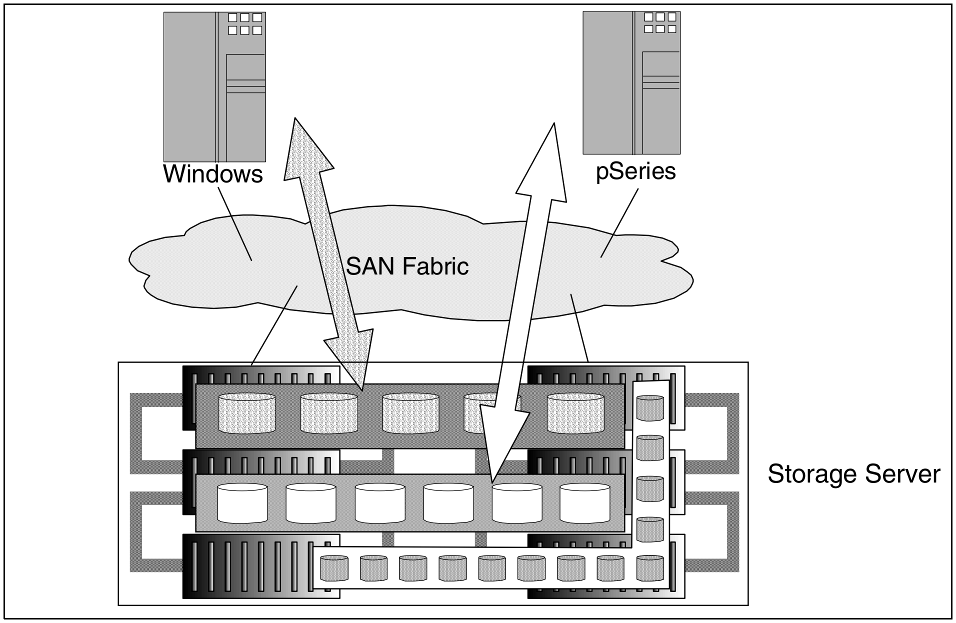


Figure 6-2 Disk pooling

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.

6.3.3 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 if 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 6-3.

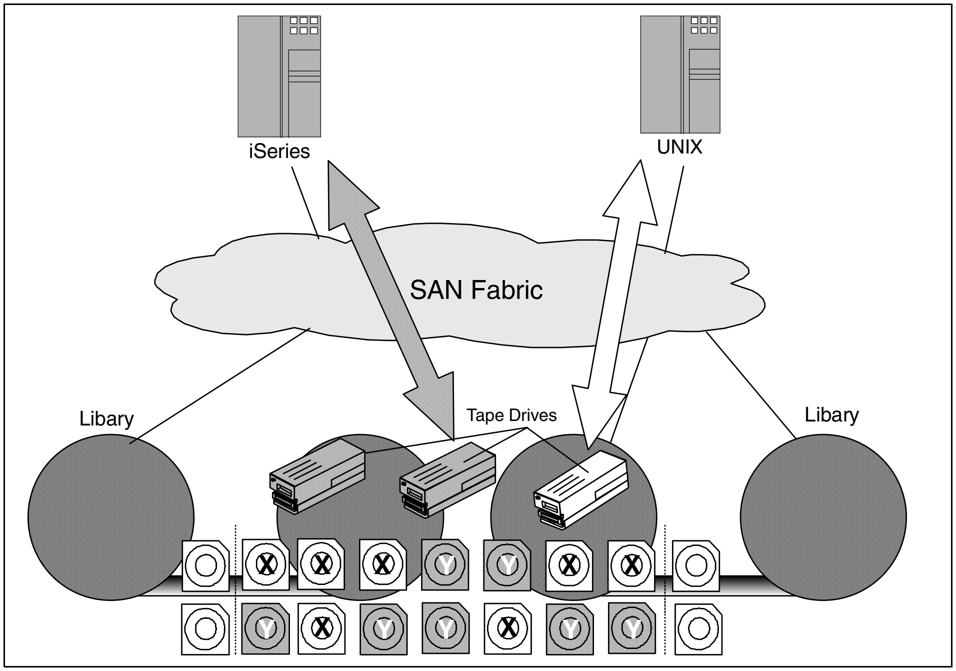


Figure 6-3 Tape pooling

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.

6.3.4 Server clustering

SAN architecture naturally lends itself to scalable 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 6-4 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.

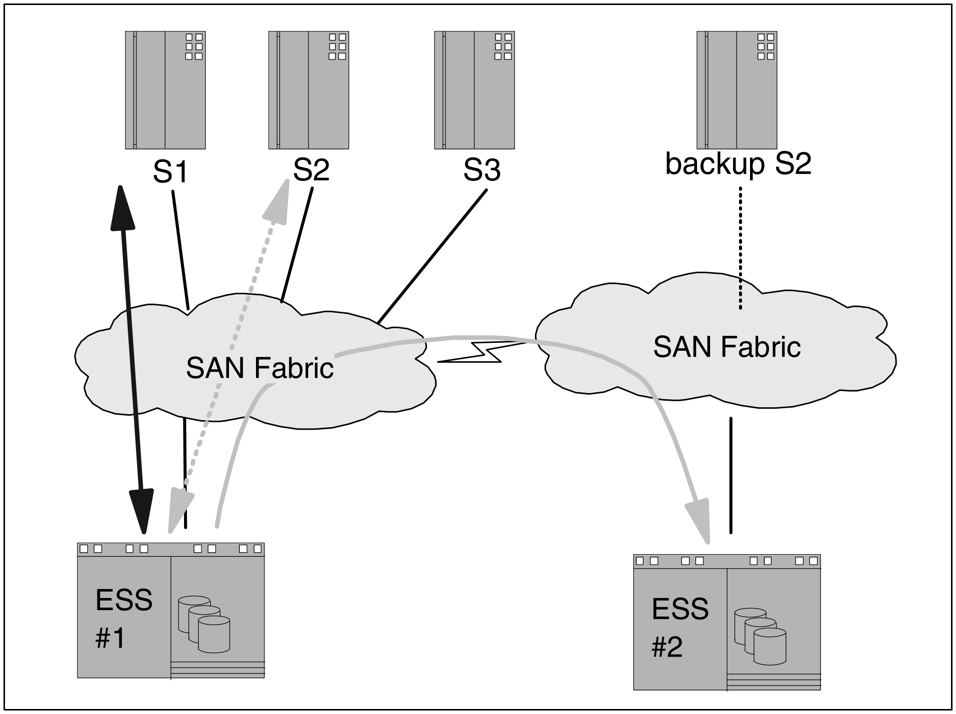


Figure 6-4 Server clustering

6.4 Pooling solutions, storage, and data sharing

The term *data sharing,* refers to accessing the same data from multiple systems and servers, as described in "Shared repository and data sharing" on page 18. 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.

6.4.1 From storage partitioning to data sharing

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 6-5 shows multiple servers accessing logical volumes created using the different alternatives mentioned above. (The logical volume *Another volume* is not assigned to any server.)

.

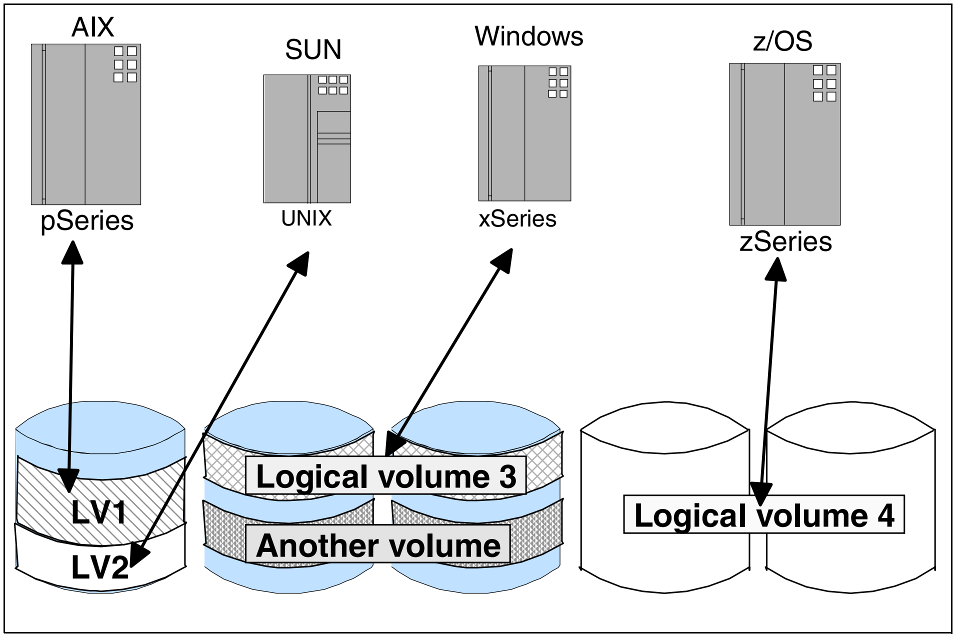


Figure 6-5   Logical volume partitioning

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 if they were stored on a shared physical or logical disk.

Figure 6-6 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.

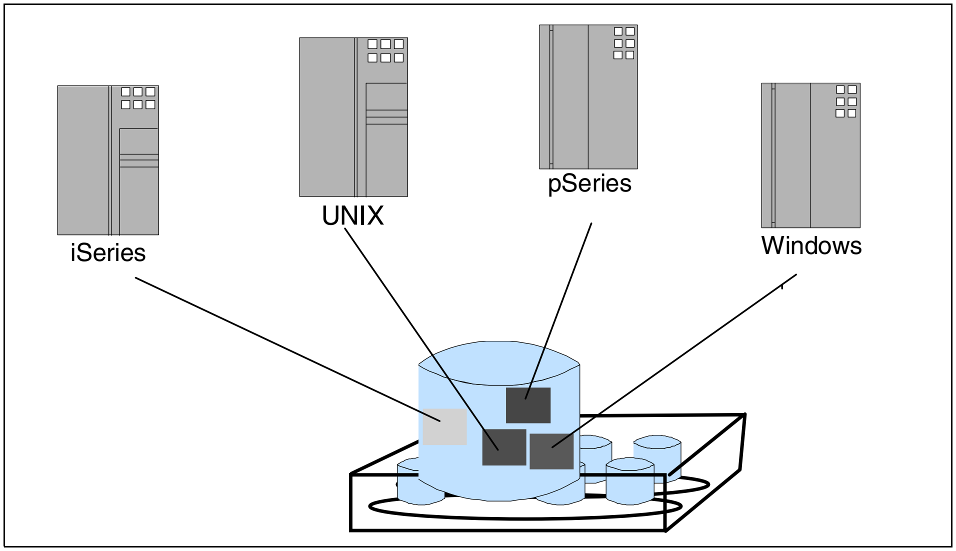


Figure 6-6 File pooling

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 data base.

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 6-7 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.

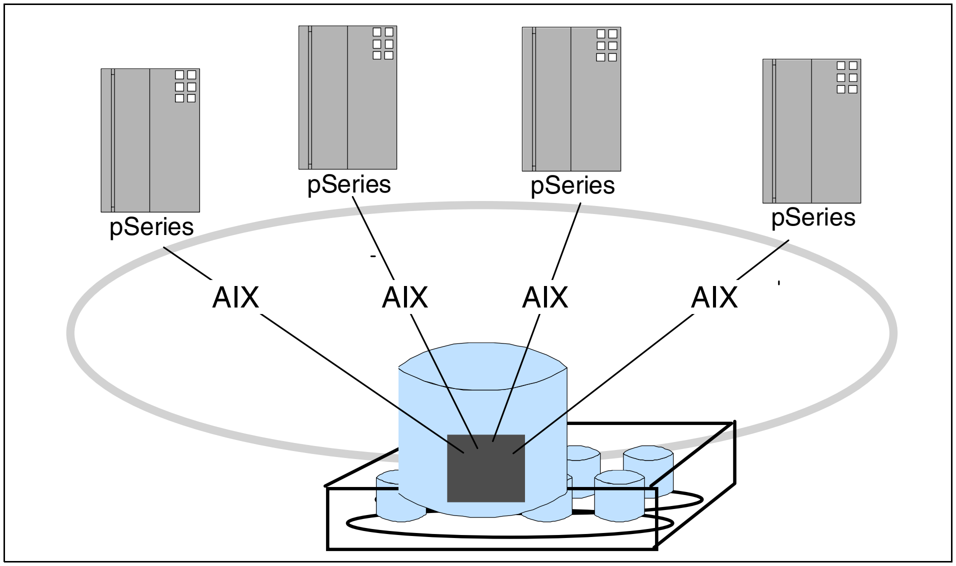


Figure 6-7 Homogeneous data sharing

Heterogeneous data sharing

In heterogeneous data sharing (as illustrated in Figure 6-8) 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 other differences described in 3.1.1 “The challenge” on page 39. Thus, we have the requirement for a SAN distributed file system. With technology such as IBM’s Storage Tank, it will be possible to provide access to the files.

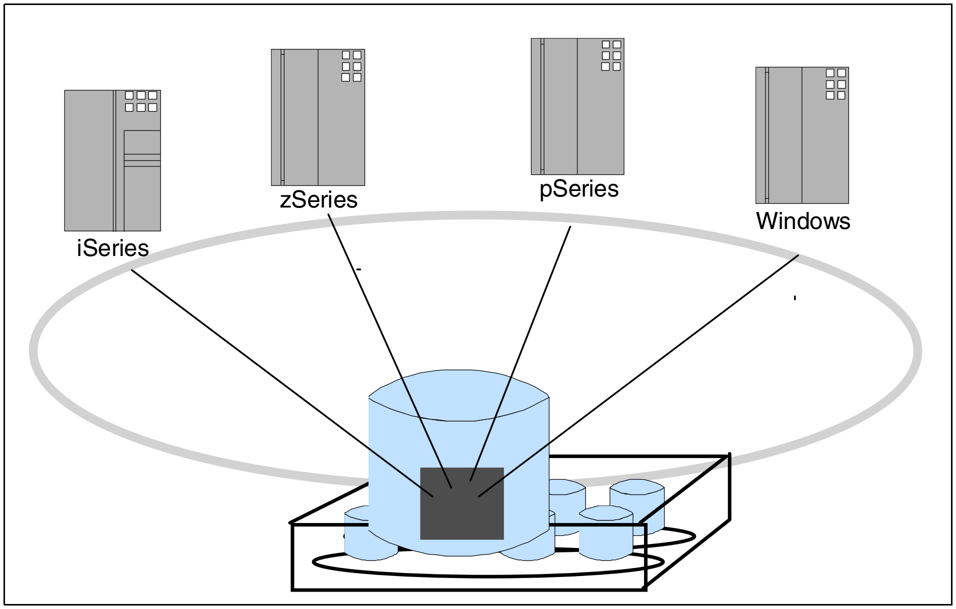


Figure 6-8 Heterogeneous data sharing

6.5 Data movement solutions

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, maybe 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.

6.5.1 Data copy services

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 (data bases, 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 extendability 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 data base 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 (like IBM's PPRC) or asynchronous (like IBM's XRC), 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.

6.6 Backup and recovery solutions

Today, data protection of multiple network-attached servers are 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 will discuss these approaches in more detail.

6.6.1 LAN-free data movement

The network backup and recovery paradigm implies that data flows from the backup and recovery client (usually a file or data base 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 6.3.3 “Tape pooling” on page 109, between backup and recovery servers, and between a backup and recovery server, and its backup and recovery client (usually a file or data base 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 6-9 shows an example of tape drive or tape library sharing.

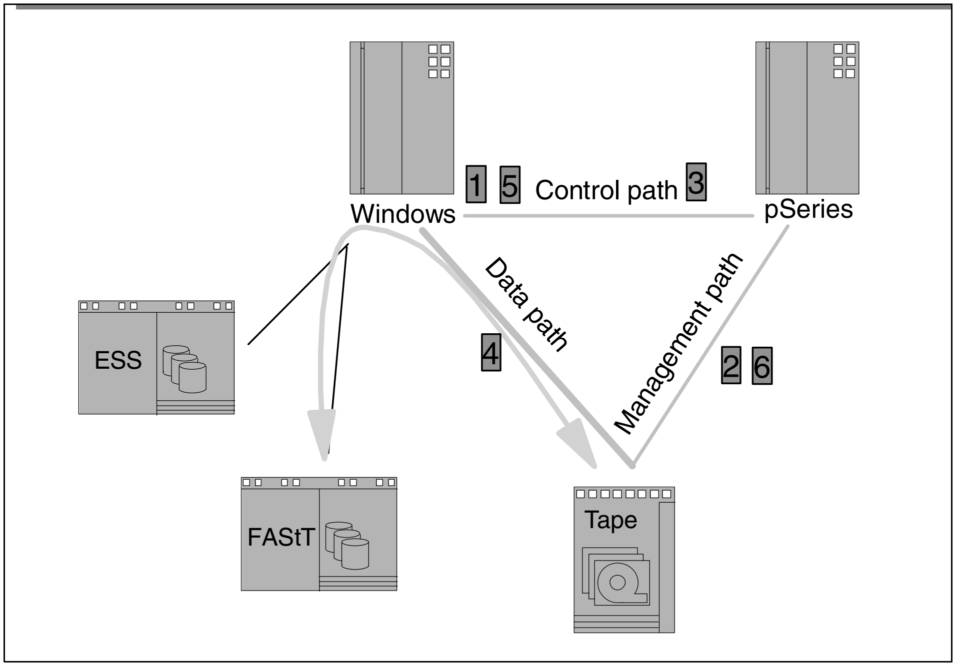


Figure 6-9 LAN-less backup and recovery

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.

6.6.2 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 6-10 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.

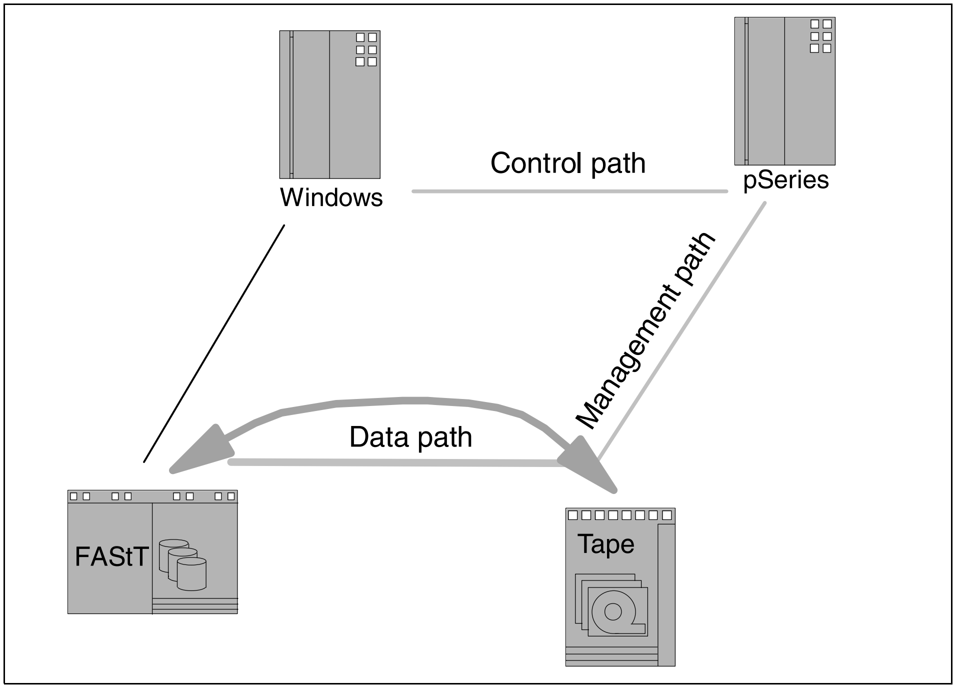


Figure 6-10 Server-free data movement for backup and recovery

In the example shown in Figure 6-10, 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 data base 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 data base 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 6-11.

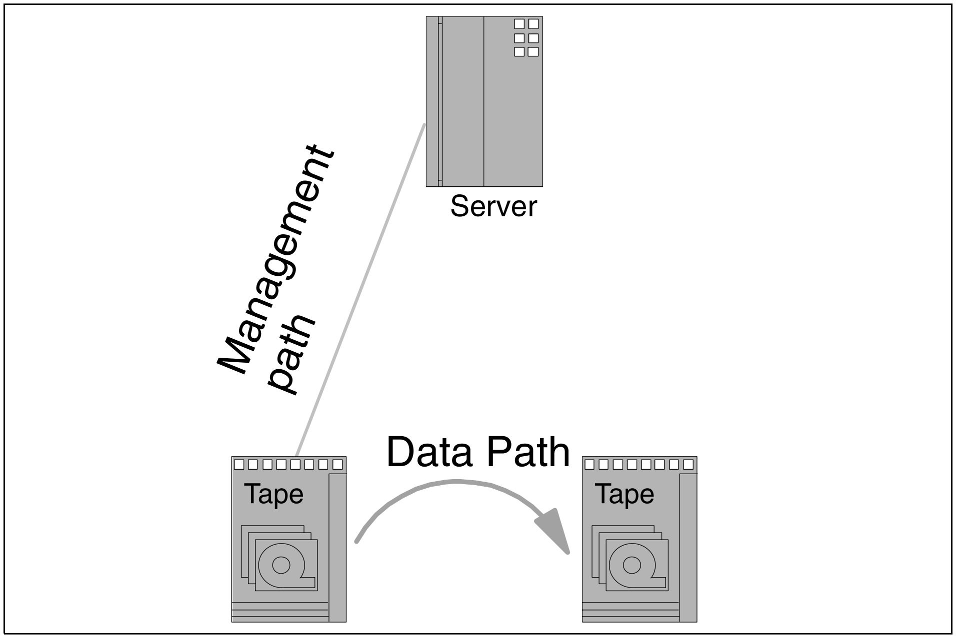


Figure 6-11 Server-free data movement for tape reclamation

6.6.3 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 6-12.

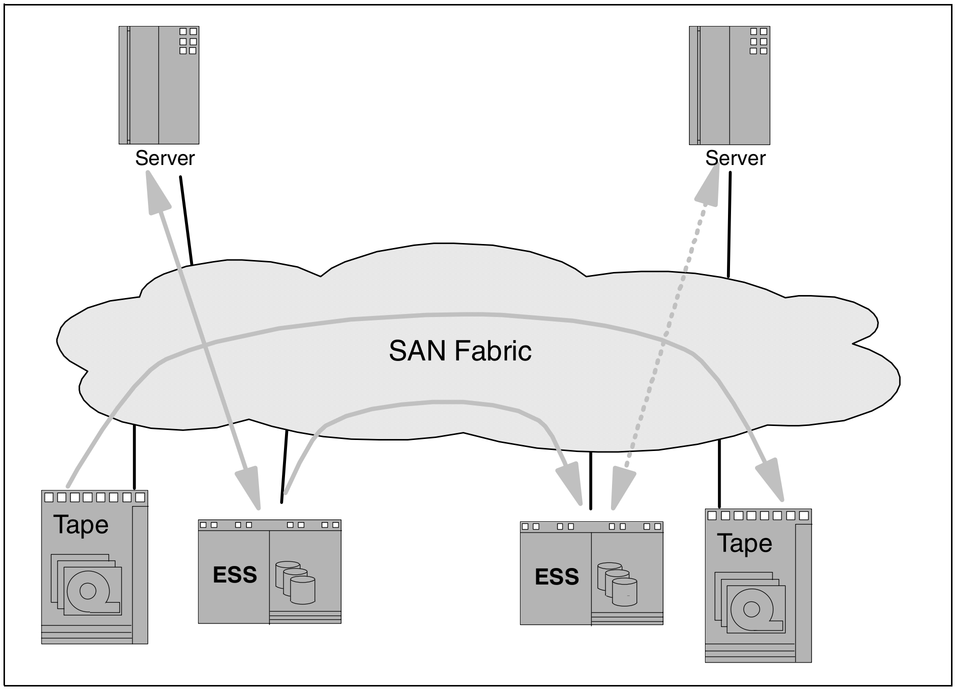


Figure 6-12 Disaster backup and recovery

When longer distances are required, SANs must be connected using gateways and WANs, similar to the situation discussed in 6.3.4 "Server clustering" on page 110, and shown in Figure 6-4 on page 112.

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.

7

SAN standardization organizations

Most importantly, the success and adoption of any new technology is greatly influenced by the standards. Although defacto standards bodies and organizations such as the Internet Engineering Task Force (IETF), American National Standards Institute (ANSI), and International Standards Institute (ISO), publish these formal standards; other organizations and industry associations, such as the SNIA, and the Fibre Channel Industry Association (FCIA) play a very significant role in defining the standards and market development.

SAN technology has a number of industry associations and standards bodies evolving, developing, and publishing the SAN standards. Figure 7-1 contains a summary of the various associations and bodies linked with SAN standards and their role.

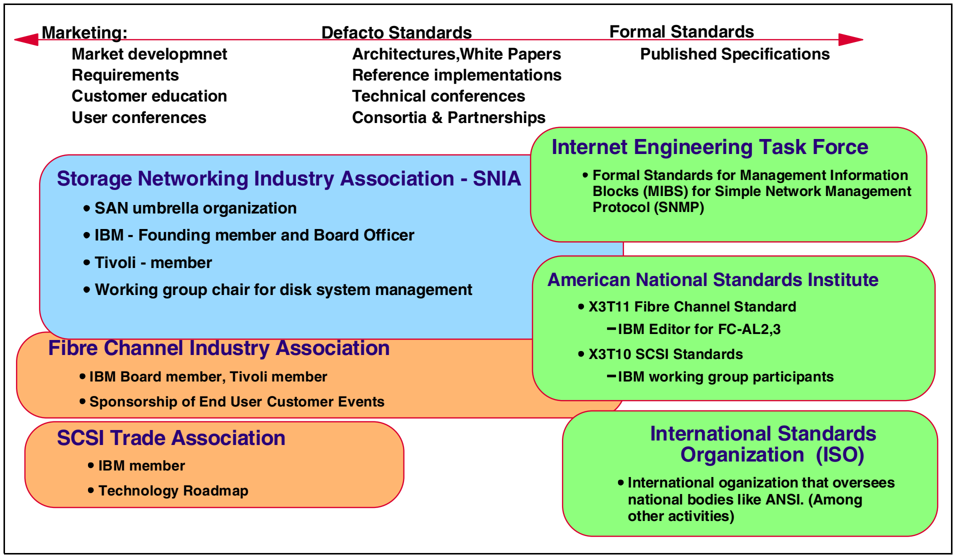


Figure 7-1   SAN Industry associations and standards bodies

IBM actively participates in most of these organizations. Notice the positioning of these organizations. We start from the left; these are the *marketing oriented organizations*. In the middle are the *defacto standards organizations*, which are often industry partnerships formed to work on defining standards. To the far right are the *formal standards organizations.* The roles of these associations and bodies fall into three categories:

- *Marketing oriented* (to the left in Figure 7-1)
  These associations are architecture development organizations that are formed early in the product life-cycle, have a marketing focus, and do the market development, requirements, customer education, 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 oriented* (in the middle in Figure 7-1) These organizations and bodies, often industry partnerships, evolve the defacto industry standards by defining the architectures, writing white papers, arranging technical conferences, and referencing implementations based on their association partners developments. They then submit these specifications for formal standard acceptance and approval. They include organizations such as SNIA and FCIA.
- *Formal standards oriented*: (to the far right in Figure 7-1) These are the formal standards organizations like 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.

7.1 SAN industry associations and organizations

A number of industry associations, alliances, consortium and formal standards bodies are involved in the SAN standards; these include SNIA, FCIA, STA, IETF, ANSI, and IEEE. A brief description of the roles of some of these organizations are given below.

7.1.1 Storage Networking Industry Association

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. SNIA can be classified as the SAN umbrella organization with over 100+ computer industry vendors as its members. IBM is one of the founding members of this organization. SNIA is uniquely committed to delivering architectures, education, and services that will propel storage networking solutions into a broader market.

SNIA is using its Storage Management Initiative (SMI) to create and promote adoption of a highly functional interoperable management interface for multi-vendor storage networking products. The SNIA strategic imperative is to have all storage managed by the SMI interface by 2005. 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 addition, they also provide a vendor-neutral certification program to help identify SAN professionals with various levels of expertise.

For additional information on the various activities of SNIA, see its Web site at:

`http://www.snia.org`

7.1.2 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 program. SANMark is a certification process designed to ensure that Fibre Channel devices, such as HBAs and switches, conform to Fibre Channel standards.

For additional information on the various activities of the FCIA, see its Web site at:

`http://www.fibrechannel.com`

7.1.3 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. For additional information see its Web site at:

`http://www.scsita.org`

7.1.4 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. It is responsible for the formal standards for the Management Information Blocks (MIB) for SNMP for SAN management. In addition, they are also responsible for iSCSI. For additional information on the IETF, see its Web site at:

`http://www.ietf.org`

7.1.5 American National Standards Institute

The American National Standards Institute (ANSI) does not itself develop American national standards. It facilitates development by establishing consensus among qualified groups. For Fibre Channel and Storage Area Networks, two of the important committees are: X3T11 Fibre Channel Standard and X3T10 SCSI Standard. IBM actively participates in both the T11 and T10 committee. For more information on ANSI, and the ANSI standards concerning Fibre Channel, FICON, and SCSI, refer to:

ANSI:
`http://www.ansi.org/`

T10 (SCSI and FCP):
`http://www.t10.org`

T11 (Fibre Channel and FICON):
http://www.t11.org

7.1.6 Institute of Electrical and Electronics Engineers

The Institute of Electrical and Electronics Engineers (IEEE) is a non-profit, technical professional association of more than 377,000 individual members in 150 countries. The organization is most popularly known and referred to as “Eye-triple-E”. 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 Organizationally Unique Identifier (OUI) list used for the addressing scheme within Fibre Channel.

For additional information on IEEE, see its Web site at:

http://www.ieee.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.

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 also: 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

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.

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.

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.

Channel A point-to-point link, the main task of which is to transport data from one point to another.

Channel I/O A form of I/O where request and response correlation is maintained through some form of source, destination, and request identification.

CIFS Common Internet File System

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**.**

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.

Controller A component that attaches to the system topology through a channel semantic protocol that includes some form of request/response identification.

CRC Cyclic Redundancy Check - An error-correcting code used in Fibre Channel.

DASD Direct Access Storage Device - any on-line 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.

dB Decibel - A ratio measurement distinguishing the percentage of signal attenuation between the input and output power. Attenuation (loss) is expressed as dB/km.

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.

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.

E_Port Expansion Port - a port on a switch used to link multiple switches together into a Fibre Channel switch fabric.

ECL Emitter Coupled Logic - The type of transmitter used to drive copper media such as Twinax, Shielded Twisted Pair, or Coax.

Enterprise Network A geographically dispersed network under the auspices of one organization.

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.

ESCON Enterprise System Connection

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 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.

Fiber Optic Refers to the medium and the technology associated with the transmission of information along a glass or plastic wire or fiber.

Fibre Channel A technology for transmitting data between computer devices at a data rate of up to 4 Gb/s. It is especially suited for connecting computer servers to shared storage devices, and for interconnecting storage controllers and drives.

FICON Fibre Connection - A next-generation I/O solution for IBM S/390 parallel enterprise server.

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.

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.

Gb/s 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.

GB/s 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 A mode of communications allowing either transmission or reception of frames at any point in time, but not both (other than link control frames which are always permitted).

Hardware The mechanical, magnetic and electronic components of a system, e.g., computers, telephone switches, terminals and the like.

HBA Host Bus Adapter

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.

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.

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 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.

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 See Local Area Network - 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.

Link A connection between two Fibre Channel ports consisting of a transmit fibre and a receive fibre.

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 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.

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.

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

Mb/s 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.

MB/s 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.

NDMP Network Data Management Protocol

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.

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.

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 Disk A storage device that is written and **read by laser light.**

Optical Fiber A medium and the technology associated with the transmission of information as light pulses along a glass or plastic wire or 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.

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.

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.

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 The hardware entity within a node that performs data communications over the Fibre Channel.

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.

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.

Protocol A data transmission convention encompassing timing, control, formatting and data representation.

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.

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.

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 (MB/s)	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

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

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 also: 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. See also: 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.

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.

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

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.

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

- *Designing and Optimizing an IBM Storage Area Network*, SG24-6419
- *Designing and Optimizing an IBM Storage Area Network Featuring the IBM 2109 and 3534*, SG24-6426
- *Designing and Optimizing an IBM Storage Area Network Featuring the INRANGE Portfolio*, SG24-6427
- *Designing and Optimizing an IBM Storage Area Network Featuring the McDATA Portfolio*, SG24-6428
- *IBM SAN Survival Guide*, SG24-6143
- *IBM SAN Survival Guide Featuring the IBM 2109*, SG24-6127
- *IBM SAN Survival Guide Featuring the McDATA Portfolio*, SG24-6149
- *IBM SAN Survival Guide Featuring the INRANGE Portfolio*, SG24-6150
- *Designing an IBM Storage Area Network*, SG24-5758
- *Introduction to SAN Distance Solutions*, SG24-6408
- *Introducing Hosts to the SAN fabric*, SG24-6411
- *Implementing an Open IBM SAN,* SG24-6116
- *Implementing an Open IBM SAN Featuring the IBM 2109, 3534-1RU, 2103-H07*, SG24-6412
- *Implementing an Open IBM SAN Featuring the INRANGE Portfolio*, SG24-6413
- *Implementing an Open IBM SAN Featuring the McDATA Portfolio*, SG24-6414
- *Introduction to Storage Area Network, SAN*, SG24-5470
- *IP Storage Networking: IBM NAS and iSCSI Solutions*, SG24-6240
- *The IBM TotalStorage NAS 200 and 300 Integration Guide*, SG24-6505
- *Implementing the IBM TotalStorage NAS 300G: High Speed Cross Platform Storage and Tivoli SANergy!*, SG24-6278

- *iSCSI Performance Testing & Tuning*, SG24-6531
- *Using iSCSI Solutions' Planning and Implementation*, SG24-6291
- *Storage Networking Virtualization: What's it all about?*, SG24-6210
- *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
- *Storage Area Networks: Tape Future In Fabrics*, SG24-5474
- *IBM Enterprise Storage Server*, SG24-5465

Other resources

These publications are also relevant as further information sources:

- *Building Storage Networks*, ISBN 0072120509

These IBM publications are also relevant as further information sources:

- *ESS Web Interface User's Guide for ESS Specialist and ESS Copy Services*, SC26-7346
- *IBM Storage Area Network Data Gateway Installation and User's Guide*, SC26-7304
- *IBM Enterprise Storage Server Configuration Planner*, SC26-7353
- *IBM Enterprise Storage Server Quick Configuration Guide*, SC26-7354
- *IBM SAN Fibre Channel Managed Hub 3534 Service Guide,* SY27-7616
- *IBM Enterprise Storage Server Introduction and Planning Guide, 2105 Models E10, E20, F10 and F20*, GC26-7294
- *IBM Enterprise Storage Server User's Guide, 2105 Models E10, E20, F10 and F20*, SC26-7295
- *IBM Enterprise Storage Server Host Systems Attachment Guide, 2105 Models E10, E20, F10 and F20*, SC26-7296
- *IBM Enterprise Storage Server SCSI Command Reference, 2105 Models E10, E20, F10 and F20*, SC26-7297
- *IBM Enterprise Storage Server System/390 Command Reference, 2105 Models E10, E20, F10 and F20*, SC26-7298
- *IBM Storage Solutions Safety Notices*, GC26-7229
- *PCI Adapter Placement Reference, SA38-0583*
- *Translated External Devices/Safety Information*, SA26-7003

- *Electrical Safety for IBM Customer Engineers*, S229-8124
- *SLIC Router Installation and Users Guide*, 310-605759
- *SLIC Manager Installation and User Guide*, 310-605807

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 Networking

  `http://www.storage.ibm.com/snetwork/index.html`
- Brocade

  `http://www.brocade.com`
- Cisco

  `http://www.cisco.com`
- INRANGE

  `http://www.inrange.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.com`
- Vixel

  `http://www.vixel.com`
- 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
- eServer xSeries 430 and NUMA-Q Information Center

  http://webdocs.numaq.ibm.com

How to get IBM Redbooks

You can order hardcopy Redbooks, as well as view, download, or search for Redbooks at the following Web site:

ibm.com/redbooks

You can also download additional materials (code samples or diskette/CD-ROM images) from that site.

IBM Redbooks collections

Redbooks are also available on CD-ROMs. Click the CD-ROMs button on the Redbooks Web site for information about all the CD-ROMs offered, as well as updates and formats.

Index

E

F

G

H

I

J

L

X

Z

IBM Redbooks

Introduction to Storage Area Networks

(0.2”spine)
0.17”<->0.473”
90<->249 pages

Introduction to Storage Area Networks

Learn the basic SAN components and their use

Introduce yourself to SAN benefits

Discover the IBM SAN product portfolio

The explosion of data created by e-business is making storage a strategic investment priority for companies of all sizes. As storage takes precedence, two major concerns have emerged: business continuity and business efficiency:
-Business continuity requires storage that supports data availability so employees, customers, and trading partners can access data 24x7 through reliable, disaster-tolerant systems.
-Business efficiency, where storage is concerned, is the need for investment protection, reduced total cost of ownership, and high performance and manageability.

Storage cannot be an afterthought anymore. Too much is at stake. Two new trends in storage are helping to drive new investments. First, companies are searching for more ways to efficiently manage expanding volumes of data and make that data accessible throughout the enterprise — this is propelling the move of storage into the network. Second, the increasing complexity of managing large numbers of storage devices and vast amounts of data is driving greater business value into software and services.

This is where a Storage Area Network (SAN) enters the arena. Simply put, SANs are the leading storage infrastructure for the world of e-business. SANs offer simplified storage management, scalability, flexibility, availability, and improved data access, movement, and backup.

This IBM Redbook is written for people marketing, planning for, or implementing Storage Area Networks, such as IBM system engineers, IBM Business Partners, system administrators, system programmers, storage administrators, and other technical support and operations managers and staff. This redbook gives an introduction to SAN. It illustrates where SANs are today, who are the main industry organizations and standard bodies active in SANs, and it positions IBM's approach to enabling SANs with its products and services.

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

SG24-5470-01 ISBN 0738428345