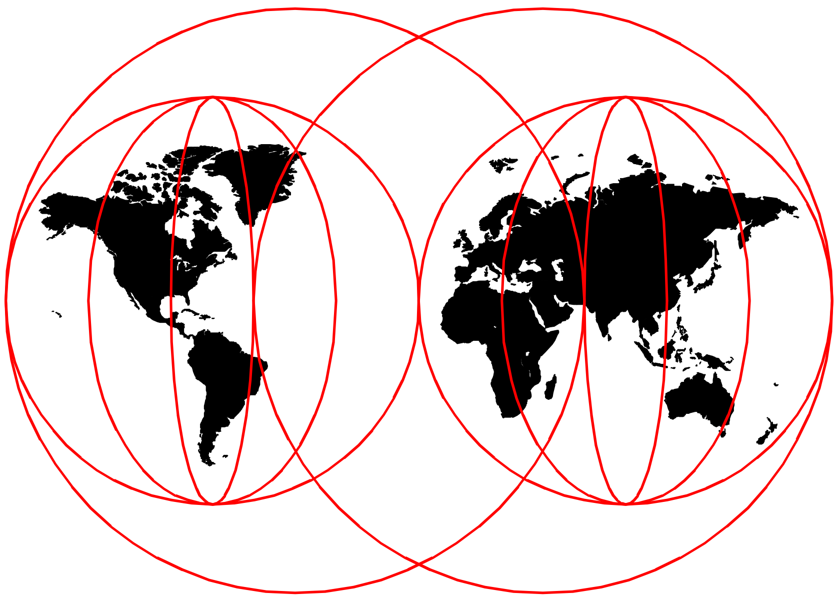
Introduction to Storage Area Network, SAN

Ravi Kumar Khattar, Mark S. Murphy, Giulio John Tarella,
Kjell E. Nystrom



International Technical Support Organization

www.redbooks.ibm.com

SG24-5470-00

SG24-5470-00

International Technical Support Organization

Introduction to Storage Area Network, SAN

August 1999

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix B, “Special Notices” on page 111.

First Edition (August 1999)

This edition is based on information available in July, 1999.

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

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

Contents

Figures

Preface

Storage Area Network (SAN) offers simplified storage management, scalability, flexibility, availability, and improved data access, movement, and backup.

This 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 in 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.

Kjell E. Nystrom is a Cross Storage Solutions Specialist at the International Technical Support Organization, San Jose Center. He writes extensively and teaches IBM classes worldwide on all areas of Storage Solutions. Before joining the ITSO in 1997, he worked for IBM Storage Systems Nordic Division, Sweden, as Storage Software Products Manager/Storage Marketing Support Specialist. Kjell has over 30 years of experience in S/390 and Storage Systems for all platforms. His areas of expertise include all major hardware and software products for the S/390 platform, Storage Systems, data disaster recovery, and data sharing. Kjell can be reached at `kjellen@us.ibm.com`

Ravi Kumar Khattar is a Storage Systems Consultant with QuantM Systems, India. He has 15 years of experience in the Information Technology field, both in India and outside india in various capacities. He holds a degree in Electronics Engineering and Is also a Microsoft Certified Systems Engineer (MCSE). His areas of expertise include Open Storage Solutions and Data Management solutions on Heterogeneous platforms. Ravi can be reached at `ravikkh@musys.net`

Mark S. Murphy is a Senior Education Specialist in the Learning Services Organization, Boulder Center. He has 24 years of experience in the Information Technology field. He has been with IBM for 2 months. Before joining IBM in March of 1999, he worked for StorageTek Corporation as a Senior Software Engineer/Systems Engineer and Senior Education Specialist. His areas of expertise include Software Engineering, RAID Technology, Open System's and Mainframe Networking. He was one of the original Software Engineers on the RAMAC Virtual Storage Array (RVA) subsystem. Mark can be reached at `msmurph@us.ibm.com`

Giulio John Tarella is a Consulting IT Specialist in Milan, Italy. He has 11 years of experience in storage and performance management on mainframes, servers, and workstations. Giulio holds a Ph.D. in seismic structural engineering from Politecnico di Milano. His areas of expertise include distributed data protection, ADSM exploitation and performance, and S/390 storage and performance management. Giulio has written extensively on DFSMS and ADSM implementation. Giulio can be reached at `gj_tarella@it.ibm.com`

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

Clod Barrera
IBM SSD, San Jose

Cynthia Behrens
IBM SSD, San Jose

Larry Connoy
IBM AS/400, Rochester

Mark Doumas
IBM SSD, Tucson

Scott Drummond
IBM SSD, San Jose

Reed A. Hancock
IBM SSD, Tucson

Lisa Haut-Mikkelsen
IBM SSD, San Jose

Dale Head
Tivoli Systems, San Jose

Peter Holmes
IBM S/390, Poughkeepsie

Keiko Kurita
Tivoli Systems, San Jose

Mike Kaczmarski
IBM SSD, Tucson

Allan S. Meritt
IBM S/390, Poughkeepsie

Alison Pate
IBM ITSO, San Jose

Bob Rees
IBM Research, Almaden

David Sacks
IBM Chicago

Greg Tevis
IBM SSD, Tucson

Matthias Werner
IBM ITSO, San Jose

Comments Welcome

Your comments are important to us!

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

- Fax the evaluation form found in “ITSO Redbook Evaluation” on page 137 to the fax number shown on the form.
- Use the online evaluation form found at `http://www.redbooks.ibm.com/`
- Send your comments in an internet note to `redbook@us.ibm.com`

Chapter 1. The Market Perception of Storage Area Networks

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 these media are accessed by applications executing in a server. Often the data is a unique company asset. You cannot buy your data on the market, but you 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 and performance; protection ranges from what to do if media fails, to complete disaster recovery procedures.

In host-centric legacy environments, the management of storage is centralized. Storage devices are connected directly (for S/390 with ESCON) to the host, and managed directly by the IT department. It is relatively straightforward and easy to manage storage in this manner.

The advent of Client/Server computing, together with downsizing and rightsizing trends, 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 host-centric 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 on page 2. Information in one island is often hard to access from other islands.

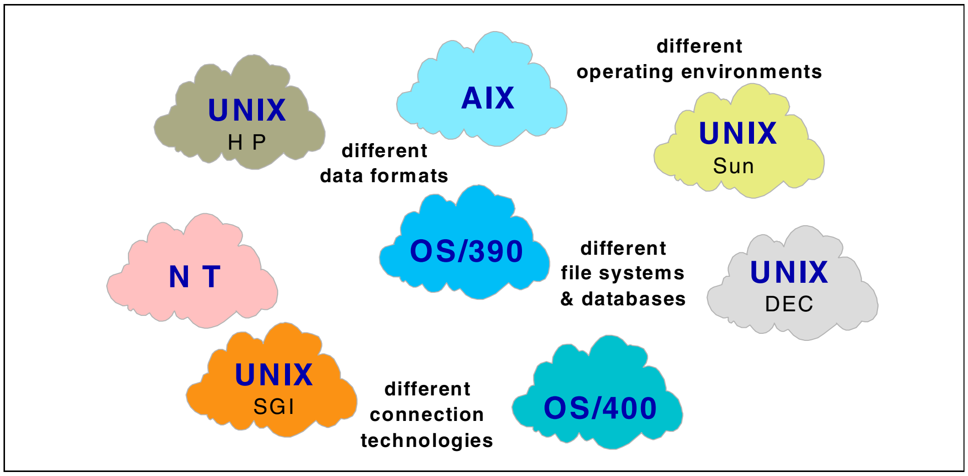


Figure 1. Islands of Information

Our 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 in the following manner:

- The top tier uses the desktop for data presentation. The desktop is based on Personal Computers (PC) or Network Computers (NC).
- 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.

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. One definition of SAN is a high-speed network, comparable to a LAN, that allows the establishment of direct connections between storage devices and processors (servers) centralized to the extent supported by the distance of Fibre Channel. The SAN can be viewed as an extension to the storage bus concept that 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 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. SAN interfaces can be Enterprise Systems Connection (ESCON), Small Computer Systems Interface (SCSI), Serial Storage Architecture (SSA), High Performance Parallel Interface (HIPPI), Fibre Channel (FC) or whatever new physical connectivity emerges.

The diagram in Figure 2 on page 3 shows a schematic overview of a SAN connecting multiple servers to multiple storage systems.

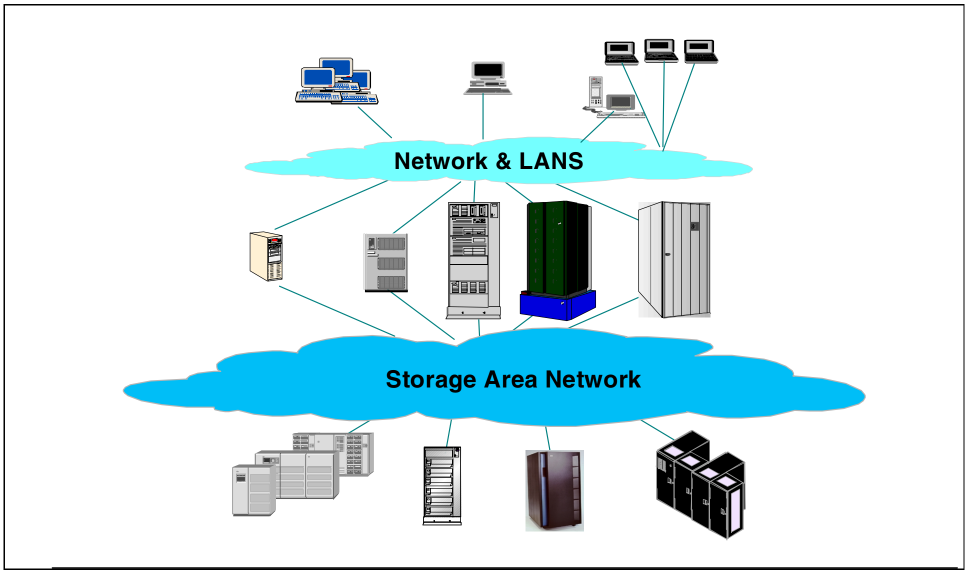


Figure 2. What is a SAN?

SANs create new methods of attaching storage to servers. These new methods promise great improvements in both availability and performance. Today's SANs are used to connect shared storage arrays to multiple servers, and are used by clustered servers for failover. They can interconnect mainframe disk or tape to network servers or clients, and can create parallel data paths for high bandwidth computing environments. A SAN is another network that differs from traditional networks because it is constructed from storage interfaces. Often it is referred to as the network behind the server.

A SAN can be used to bypass traditional network bottlenecks. It supports direct, high speed data transfers between servers and storage devices in 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 without any 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 independent of applications and accessible through alternate data paths.
- Higher application performance: Storage processing 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.
- Simplified centralized management: Single image of storage media simplifies management.

Often we hear the term Network Attached Storage (NAS) or network appliance. A NAS is basically a LAN attached file server that serves files using a network protocol such as Network File System (NFS) or Simple Network Management Protocol (SNMP). It is not a SAN device.

1.1 Evolution of Storage Devices 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 throughputs.

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 building, issues the low-level device commands, and listens for 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.

With server-attached storage, part of the processor cycles are used up to perform I/O. The next step was to develop off-load processors, such as the channel subsystem introduced by Multiple Virtual Storage/Extended Architecture (MVS/XA). The channel subsystem directs the flow of information between I/O devices and main storage. It relieves the processor of the task of communicating directly with I/O devices and permits data processing to proceed concurrently with I/O processing.

In S/390 environments, connections between I/O devices and the processor used to be established using bus and tag cables. These cables were bulky, limited to low data rates, at most 4.5 MB/sec, and to a maximum of 400 feet in length. In 1991, IBM introduced a new form of connectivity between S/390 servers and storage devices called Enterprise System CONnection (ESCON). ESCON channels are based on optical fiber. They are capable of data rates up to 17 MB/sec half-duplex, and are much less bulky than the old bus and tag cables, thus reducing cabling complexity and space required, and increasing the distance covered to multiple kilometers (up to 43 km). In S/390 an I/O device can be attached using multiple channels for higher performance and availability.

The ESCON architecture has been further enhanced with the introduction of ESCON Directors. These are essentially high speed switches that provide dynamic connection capability between attached units: servers and storage devices. With an ESCON Director's ability to create connections dynamically, a single channel can communicate with many control units, and a single control unit can communicate with many channels on one or more host servers. This can reduce the number of channels and control units required to support an equivalent configuration that uses direct paths (a configuration without ESCON Directors). ESCON channels, coupled with ESCON Directors, represent the first SAN architecture. The ESCON Director is managed by the software package ESCON Manager.

The S/390 FICON architecture is an enhancement of, rather than a replacement for, the existing ESCON architecture. As a SAN is Fibre Channel based, FICON is a prerequisite for S/390 to fully participate in a heterogeneous SAN. FICON is a protocol that uses Fibre Channel for transportation and is mapped in layer FC-4 (1.2.1.2, “Upper Layers” on page 9) of Fibre Channel. FICON channels are capable of data rates up to 100 MB/sec 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. FICON architecture is fully compatible with existing S/390 channel commands (CCWs) and programs.

FICON is currently in the standardization process as *FC-SB-2*. The current state of what has been disclosed can be found at `www.t11.org` (Click on *Docs*, then on *Search*, and enter `FC-SB-2` as the search argument). There are numerous occurrences representing different sections and/or presentations.

The architectures discussed above are used in S/390 environments and are discussed in S/390 terms. Slightly different approaches were taken on other platforms, particularly in the UNIX and PC worlds.

The UNIX and PC world developed different connectivity standards. In the PC world, the AT Attachment (ATA) interface was one of the first common standards.

The ATA interface includes Integrated Drive Electronics (IDE), Enhanced IDE (EIDE), and Ultra ATA (UltraATA). These were evolutions of the ATA standard that added speed and functionality, up to 66 MB/sec with UltraATA. Although the interface is backward compatible, there are many limitations, such as a maximum of four devices per interface, a maximum cable length of 10 inches, and many different types of read and write commands. The ATA interface is popular because it is cheap, it is supported on all PC motherboards, and it is backward compatible.

The Small Computer System Interface (SCSI) is a parallel interface. Although SCSI protocols can be used on Fibre Channel (then called FCP) and SSA devices, most people mean the parallel interface when they say SCSI. 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.

Serial Storage technology was introduced by IBM in 1991. Today, IBM's third-generation Serial Storage Architecture (SSA) solutions provide an open storage interface designed specifically to meet the high performance demands of network computing. SSA is primarily an architecture for device attachment within subsystems, but is also used connect devices to servers. When compared to other technologies, such as parallel SCSI, SSA offers superior performance and data availability with greater flexibility and value. IBM leads the industry in serial storage technology, and SSA has achieved extraordinary market acceptance, with over 5 petabytes shipped to customers worldwide in the last 3 years.

SSA uses a loop rather than a bus. The SSA architecture enables two simultaneous read and two simultaneous write paths, with each of the four paths having 40 MB/sec bandwidth. With arbitrated subsystems, such as SCSI and Fiber Channel-Arbitrated Loop (FC-AL), arbitration occurs before and after every transfer. This significantly reduces the amount of data that can be transferred. Even 100 MB/sec FC-AL can sustain significantly less than 50% of the available bandwidth for transaction processing systems. SSA has no arbitration and can utilize the available bandwidth significantly better than SCSI.

With the SSA loop design, a single cable failure will not cause loss of access to data. If there is a failure on the loop, the SSA adapter will automatically continue accessing the devices in a non-loop configuration. Once the path is restored, the adapter will automatically reconfigure to resume the normal mode of operation. If there is a disk failure, the hot-swappable disks can be removed without loss of communication between the adapter and the other disks on the loop. The current architecture allows for up to 127 drives per loop and will scale to multiple terabytes. Add disk drives and, depending on access patterns and rates, performance increases; add adapters, and performance increases; segment into additional loops, and performance increases again. These types of performance increases do not occur in SCSI architectures.

1.2 Evolution to SAN

Another definition of a SAN is a centrally managed, high-speed storage network consisting of multi-vendor storage systems, storage management software, application servers and network hardware that allows you to exploit the value of your business information.

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

1.2.1 Fibre Channel Architecture

Today, the industry considers Fibre Channel (FC) as the architecture on which most SAN implementations will be 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 100 MB/sec. The 200 MB/sec and 400 MB/sec data rates have already been tested. This standard is backed by a consortium of industry vendors and has been accredited by the American National Standards Institute (ANSI). Many products are now on the market that take advantage of FC's high-speed, high-availability characteristics.

Note 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.

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 at the same level on the standard FC transport. An analogy could be ethernet, where IP, NetBOIS, 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 expected to be the standard protocol for S/390, and FCP is the expected standard protocol for the non-S/390 systems, both using Fibre Channel architecture to carry the traffic.

In the following sections, we will introduce some basic Fibre Channel concepts, starting with the physical layer and proceeding to define the services offered.

1.2.1.1 Physical Layer

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.

- **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.

1.2.1.2 Upper Layers

Fibre Channel is a transport service that moves data fast 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.

1.2.1.3 Topologies

Fibre Channel interconnects nodes using three physical topologies that can have variants. Topologies include:

- ***Point-to-point*** — The point-to-point topology consists of a single connection between two nodes. All the bandwidth is dedicated for 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*** — 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, whereas frame switches route frames between nodes and establish the connection only when needed. A switch can handle all protocols as it does not look at the Fibre Channel layer FC-4.

1.2.1.4 Classes of Service

Fibre Channel provides a logical system of communication called Class of Service that is allocated by various Login protocols. The following five classes of service are defined:

- **Class 1** — Acknowledged Connection Service

  Dedicates connections through fabric equivalent of a dedicated physical link and delivers frames with acknowledgment in the same order as transmitted.

- **Class 2** — Acknowledged Connectionless Service

  Multiplexes frames from multiple sources with acknowledgment. Frame order is not guaranteed.

- **Class 3** — Unacknowledged Connectionless Service

  As class 2, without frame acknowledgment. Flow has to be controlled at buffer level.

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

  As class 1, but with only a minimum of bandwidth guaranteed. If sufficient bandwidth is available, class 2 and 3 frames will share connections.

- **Class 6** — Simplex Connection Service

  As class 1, but also provides multicast and preemption.

1.3 SAN Components

As mentioned earlier, the industry considers Fibre Channel as the architecture on which most SAN implementations will be built, with FICON as the standard protocol for S/390 systems, and FCP as the standard protocol for non-S/390 systems. Based on this perception, the SAN components described in the following sections will be Fibre Channel based (see Figure 3 on page 11).

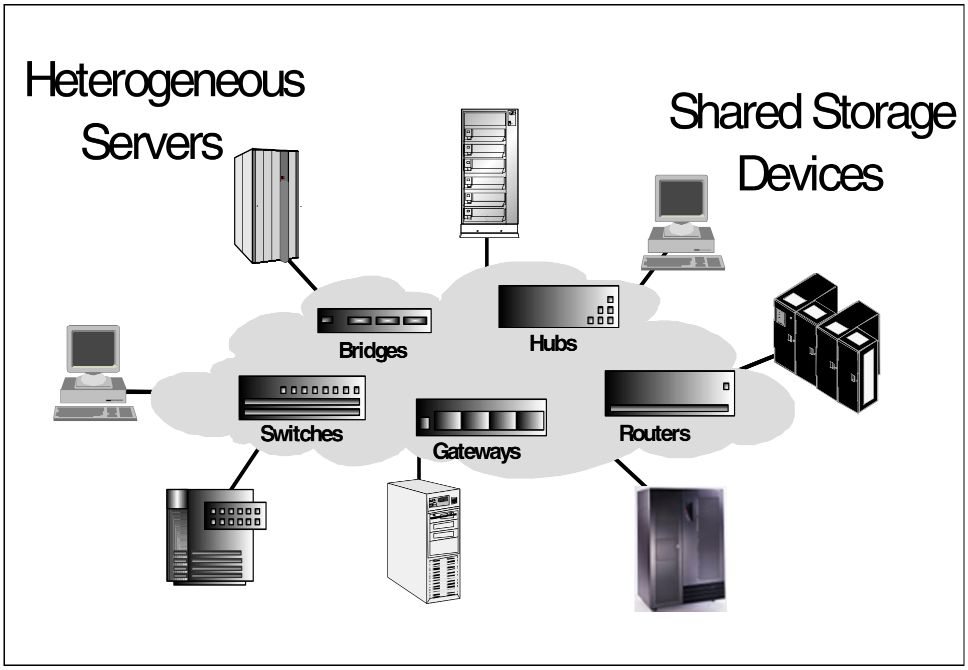


Figure 3. SAN Components

1.3.1 SAN Servers

The server Infrastructure is the underlying reason for all SAN solutions. This infrastructure includes a mix of server platforms such as Windows NT, UNIX (various flavors) and OS/390. With initiatives such as Server Consolidation and e-Business, the need for SAN will increase. Although the early SAN solutions only supported homogeneous environments, SAN will evolve into a truly heterogeneous environment.

1.3.2 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 new kind of infrastructure is needed, one that provides more enhanced network availability, data accessibility, and system manageability than is provided by today's infrastructure.

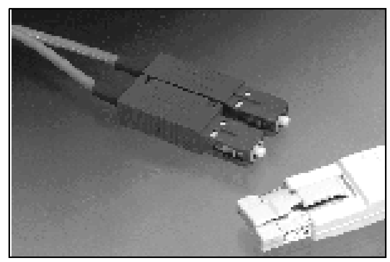
The SAN meets this challenge. The SAN liberates the storage device, so it is not on a particular server bus, and attaches it directly to the network. In other words, storage is externalized and functionally distributed across the organization. The SAN also enables the centralizing of storage devices and the clustering of servers, which makes for easier and less expensive administration.

1.3.3 SAN Interconnects

The first element that must be considered in any SAN implementation is the connectivity of storage and server components using technologies such as 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 origin 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. For example, after reading the sections 1.3.3.10, “Bridges” on page 15 and 1.3.3.11, “Gateways” on page 15, you will find that the IBM SAN Data Gateway is really a bridge, not a gateway.

1.3.3.1 Cables and Connectors



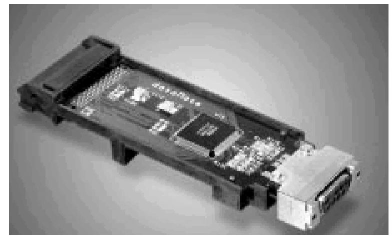
As with parallel SCSI and traditional networking, there are different types of cables of various lengths for use in a Fibre Channel configuration. Two types of cables are supported: copper and optical (fiber). Copper cables are used for short distances (up to 30m) and can be identified by their DB9 (9-pin) connector. Fiber 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 10km). IBM will support the following distances for FCP:

Diameter(Microns)	Mode	Laser type	Distance
9	Single-mode	Longwave	<=10 km
50	Multi-mode	Shortwave	<=500 m
62.5	Multi-mode	Shortwave	<=175 m

In addition, connectors (see 1.3.3.3, “Gigabit Interface Converters (GBIC)” on page 13) and cable adapters (see 1.3.3.4, “Media Interface Adapters (MIA)” on page 13) have been developed that allow the interconnection of fiber optic based adapters with copper based devices.

1.3.3.2 Gigabit Link Model (GLM)

Gigabit Link Models are generic Fibre Channel transceiver units that integrate the key functions necessary for installation of a Fibre Channel media interface on most systems.



1.3.3.3 Gigabit Interface Converters (GBIC)

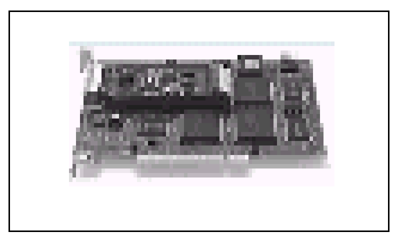
Gigabit Interface Converters are typically used with hubs and switches, and allow both copper and fiber optics to connect up to the same hub or switch. This works well in an environment where the components to be attached may be both fiber optics and copper, depending on cost and distance requirements.

1.3.3.4 Media Interface Adapters (MIA)

Media Interface Adapters can be used to facilitate conversion of an optical interface connection to copper and vice versa. Typically, MIAs are attached to host bus adaptors, but can also be used with switches and hubs. When a hub or switch only supports copper or optical connections, MIAs can be used to convert the signal to the appropriate media type (copper or optical).

1.3.3.5 Adapters

Adapters are devices that connect to a network, server, or storage device and control the electrical protocol for communications. Adapters are also referred to as Network Interface Cards (NIC), Enterprise Systems Connection (ESCON) adapters, Host Bus Adapters (HBA), and SCSI host bus adapters.



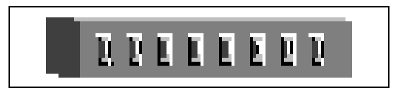
1.3.3.6 Extenders

Extenders are used to facilitate longer distances between nodes that exceed the theoretical maximum.

1.3.3.7 Multiplexors

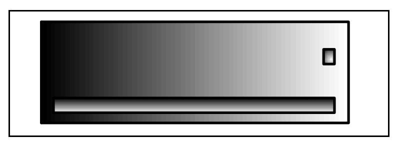
Multiplexors provide for more effective utilization of high speed bandwidth resources by interleaving data from multiple sources onto a single Link. An example of this would be the FIbre CONnection Channel (FICON) Bridge, which allows up to eight separate ESCON paths to be multiplexed over a single FICON Link. The FICON bridge is not a separate box; as shown on the picture, it is a card in the ESCON Director. Multiplexors are also becoming increasingly efficient in terms of data compression, error correction, transmission speed and multi-drop capabilities.

1.3.3.8 Hubs



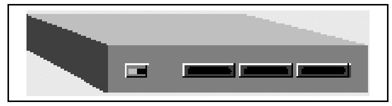
Fibre Channel hubs are used to connect up to 126 nodes into a logical loop. All connected nodes share the bandwidth of this one logical loop. Fibre Channel Arbitrated Loop (FC-AL) protocol is the most widely accepted, cost-effective alternative. Each port on a hub contains a Port Bypass Circuit (PBC) to automatically open and close the loop to support hot pluggability. Multiple hubs and links can be implemented to provide alternate path failover capability for high availability server environments. Intelligent hubs are currently being offered that provide features such as dynamic loop configuration and some of the benefits of switches.

1.3.3.9 Routers



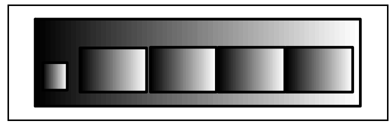
Storage routing is a new technology based on the old concept of routing as it is understood by the data communications industry. Storage routers differ from network routers in that the data being routed uses storage protocols like FCP (SCSI) instead of messaging protocols such as TCP/IP. The data path used to transfer storage data may be the same as that used for messaging traffic, but the content of the data itself contain imbedded storage protocol information. This is similar to the concept of tunneling protocols used in the broader market. For example, a storage router could encapsulate SCSI information in TCP/IP packets for transmission over an intervening ethernet network. The term *routing* implies data transfers over differing transmission media and addressing schemes. As a combination of communications and storage channel capabilities, Fibre Channel represents the first opportunity to apply communication techniques, such as routing, to storage traffic.

1.3.3.10 Bridges



Bridges facilitate communication between LAN/SAN segments and/or networks with dissimilar protocols. An example of this would be a FICON bridge, which allows ESCON protocol to be tunneled over Fibre Channel protocol. FICON Bridges reduce the requirements of ESCON connections, ESCON channels, ESCON Director ports, and so on; they support large and small block multiplexing.

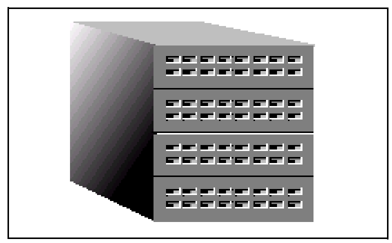
1.3.3.11 Gateways



A gateway is a network station used to interconnect two or more dissimilar networks or devices, and may or may not perform protocol conversion. These boxes are typically used to provide access to WAN from a LAN. With gateways, SANs can be extended across a WAN.

Note: The IBM SAN Data Gateway is used to attach SCSI devices like tape libraries, disk subsystems at one end, and Fibre Channel connections at the other end, which makes it a router, not a gateway.

1.3.3.12 Switches



Switches are among the highest performing devices available for interconnecting large numbers of devices, increasing bandwidth, reducing congestion and providing aggregate throughput. 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 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 which other ports a specific port can connect to. This is called port zoning.

1.3.3.13 Directors

When ESCON was announced, the switches were named *directors*, as they included high availability and other features; for example, dual power supply, not included in traditional switches.

ESCON

ESCON transfers data in half duplex mode at a transfer rate of 17 MB/sec. The directors provide dynamic or static connections between attached channels, control units, devices, and other ESCON directors. ESCON Directors provide single-link distances up to 3 km for multi-mode fiber and 20 km for single-mode fiber.

The ESCON director supports the FICON bridge card mentioned in 1.3.3.10, “Bridges” on page 15. Using the FICON bridge, an S/390 system can, through a FICON channel, communicate with ESCON control units and devices attached to the director.

FICON

FICON transfers data in full duplex mode at a rate of 100 MB/sec. The FICON director is a regular Fibre Channel switch with some additional features. It provides connections between FICON channels and control units or devices with FICON interfaces. FICON Directors provide single-link distances up to 550 meters for multi-mode fiber (50 or 62.5 micron) and 10 km (20 km with RPQ) for single-mode fiber (9 micron).

1.3.4 SAN Applications

Storage Area Networks (SANs) enable a number of applications that provide enhanced performance, manageability, and scalability to IT infrastructures. These applications are being driven by parts technology capabilities, and as technology matures over time, we are likely to see more and more applications in the future.

Figure 4 on page 17 Illustrates the various SAN Applications.

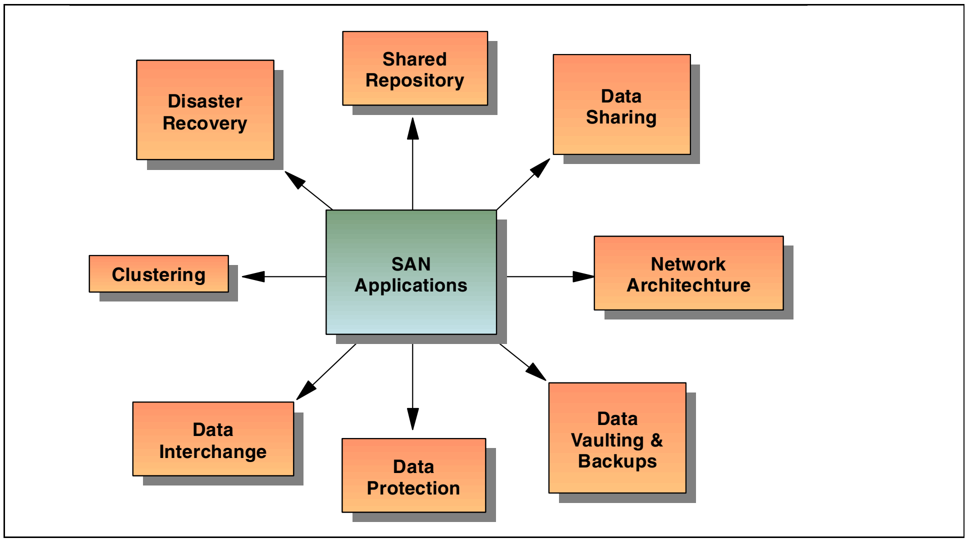


Figure 4. SAN Applications

1.3.4.1 Shared Repository and Data Sharing

SANs enable storage to be externalized from the server and centralized, 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.

Figure 5 on page 18 shows the data sharing concepts and types of data sharing.

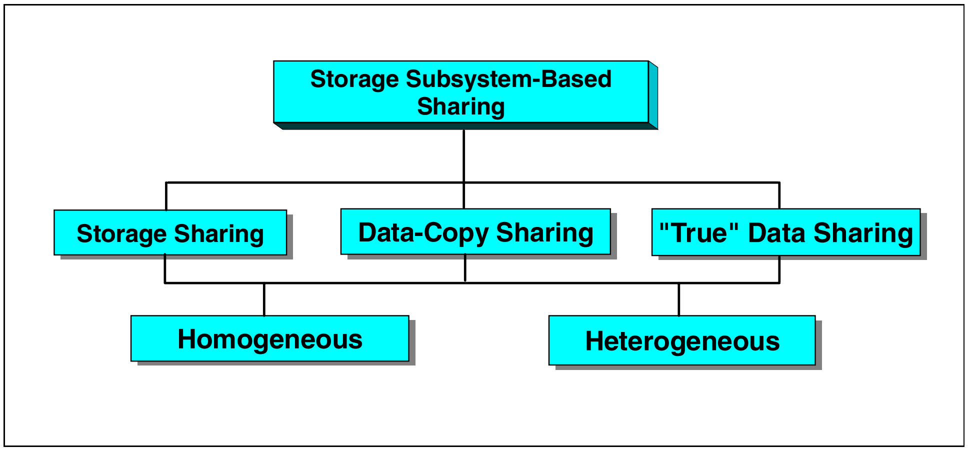


Figure 5. Data Sharing Concepts and Types

Storage Sharing

With storage sharing, two or more homogeneous or heterogeneous servers share a single storage subsystem whose capacity has been physically partioned 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 such as in a S/390 Parallel Sysplex configuration.

For more information on storage sharing, see, for example, *IBM Magstar 3494 Tape Libraries: A Practical Guide*, SG24-4632; *IBM Versatile Storage Server*, SG24-2221; and *Implementing the Enterprise Storage Server in Your Environment*, SG24-5420.

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. There are two approaches to data-copy sharing between platforms: flat file transfer and piping.

For more information on data-copy sharing, see *Data Sharing: Implementing Copy-Sharing Solutions Among Heterogeneous Systems*, SG24-4120.

"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 an OS/390 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 or Oracle Parallel server to guarantee data integrity.

1.3.4.2 Network Architecture

The messaging network like an Ethernet, is often defined as the primary network, in which case the SAN can be defined as the secondary network or, as some vendors describe it, *the network behind the server*. This network architecture enables centralized storage by using similar interconnect technologies used by LANs and WANs: routers, hubs, switches and gateways — and allows the SAN to be local or remote, shared or dedicated. This provides significant improvements in availability, performance, and scalability of the storage resources.

1.3.4.3 Data Vaulting and Data Backup

In most present day scenarios of data vaulting and data backups to near-line or off-line storage, the primary network, LAN or WAN, is the medium used for transferring both the server, file and database server, or end-user client data to the storage media. SANs enable data vaulting and data backup operations on servers to be faster and independent of the primary network, which has led to the delivery of new data movement applications like LAN-less backup and server-free backup.

1.3.4.4 Data Interchange

Today, data interchange primarily involves moving data from one storage system to another. However, the SAN roadmaps of many vendors show data interchange between different heterogeneous systems as one of the goals for SAN. This is primarily because different platforms store and access data using different methods for data encoding and file structures. True heterogeneous data interchange can only be achieved when these different file systems allow for transparent data access.

1.3.4.5 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.

1.3.4.6 Data Protection and Disaster Recovery

The highest level of application availability requires avoiding 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, 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 second redundant copies of the data by storage mirroring, remote cluster storage, Peer-to-Peer Remote Copy (PPRC) and Extended Remote Copy (XRC), Concurrent Copy, and other High Availability (HA) data protection solutions. These are then used for disaster recovery situations. SAN any-to-any connectivity enables these redundant data/storage solutions to be dynamic and not impact the primary network and servers, including serialization and coherency control.

For more information on the copy services mentioned above, see *Planning for IBM Remote Copy*, SG24-2595; and *IBM Enterprise Storage Server*, SG24-5465.

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

1.3.5 Managing the SAN

In order to achieve the various benefits and features of SANs, like performance, availability, cost, scalability, and interoperability, the infrastructure (switches, routers, and so on) of the SANs, as well as the attached storage systems, must be effectively managed. To simplify SAN management, SAN vendors need to adapt Simple Network Management Protocol (SNMP), Web Based Enterprise Management (WBEM) and Enterprise Storage Resources Management (ESRM) type standards to consistently monitor, alert and manage all components of the SAN. There is also a need for managing partitions of the SAN from a central console. The big challenge is to ensure that all components are interoperable and work with the various management software packages.

Management of the SAN can be divided into various disciplines defined in Enterprise Storage Resources Management (ESRM). These disciplines should be implemented across all heterogeneous resources connected to the SAN, with the objective of providing one common user interface across all resources.

Figure 6 on page 21 highlights the IBM's ESRM disciplines.

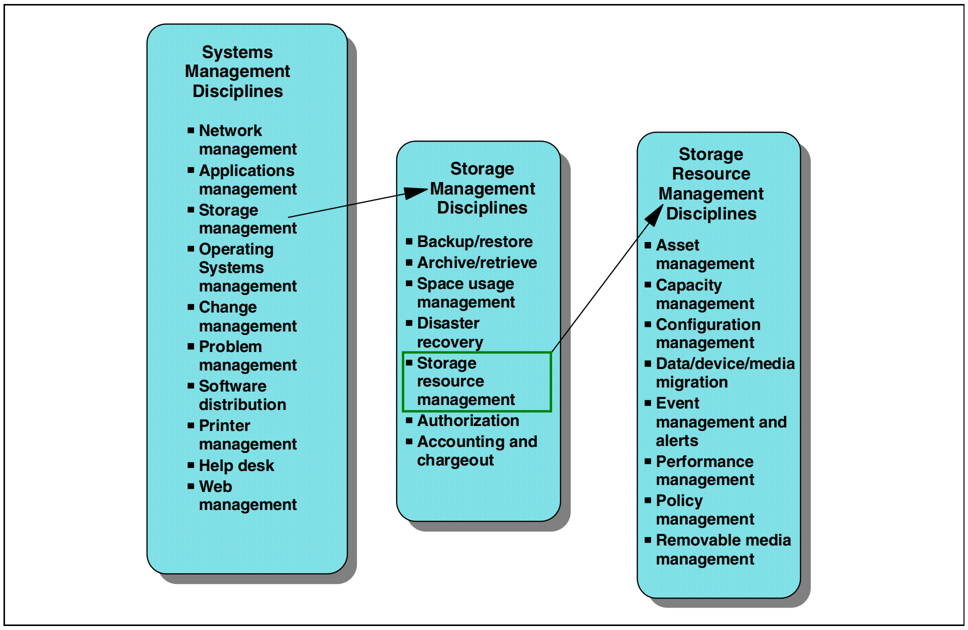


Figure 6. IBM's ESRM Disciplines

1.3.5.1 Asset Management

Asset management addresses the need to discover resources, recognize the resources, and tie them into the rest of the topology. The output is a complete inventory list of the assets that include manufacturer, model information, software and licence information.

1.3.5.2 Capacity Management

Capacity management addresses the sizing of the SAN, for example, how many switches of which size are needed. It also addresses the need to know about free space/slots, unassigned volumes, free space in the assigned volumes, number of backups, number of tapes, percent utilization, and percentage free/scratch.

1.3.5.3 Configuration Management

Configuration management addresses the need for current logical and physical configuration data, ports utilization data, and device driver data to support the ability to set SAN configurations based on business requirements of high availability and connectivity. It also addresses the need of integrating the configuration of storage resources with configuration of the server's logical view of them; for example, whoever configures an Enterprise Storage Server affects what must ultimately be configured at the server.

1.3.5.4 Performance Management

Performance management addresses the need to improve performance of the SAN and do problem isolation at all levels — device hardware and software interfaces, application, and even file level. This approach requires common platform-independent access standards across all SAN solutions.

1.3.5.5 Availability Management

Availability management addresses the need to prevent failure, correct problems as they happen, and provide warning of the key events long before they become critical. For example, in the event of a path failure, an availability management function might be to determine a link or other component failed, assign an alternate path, and page the engineer to repair the failing component, keeping the systems up through the entire process.

1.4 Where We Are Today

In the following sections we discuss the SAN scene today.

1.4.1 Server Attached Storage

SAN excitement is rampant; after all, the promise of an any-to-any nirvana is pretty compelling. However, most analysts agree, and history bears witness to this truth, that both technology 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 7 on page 23 highlights the limitations of Server Attached Storage.

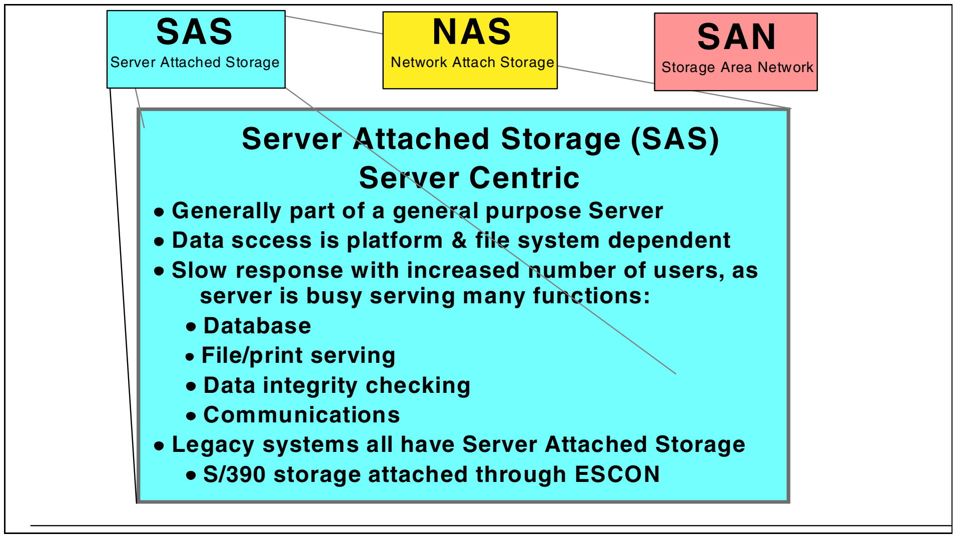


Figure 7. Server-Centric

1.4.2 Network Attached Storage

All of the promises of 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 early SAN implementation, there are solutions that allow them to *test the waters* without making a huge investment and without implementing an entirely new infrastructure.

IBM laid the groundwork for SAN in 1991 with the introduction of ESCON, the industry's first SAN. ESCON started out as a serial, fiber optic, point-to-point switched network connecting disk, tape, and printing devices to S/390 (mainframe) servers.

Although its initial entry into the market was as a high-end solution, ESCON has responded to customer demands for open connectivity and has matured to become truly heterogeneous, supporting ether native ESCON attachment or ESCON attachment through converters or gateways to multiple systems, including UNIX servers for IBM, HP, SUN, DEC, and Sequent, as well as servers running Windows NT. These networks deliver many of the benefits of a SAN deployment using proven mature technology, and will provide a migration path to Fibre Channel as it evolves. In addition, bridges (gateways) are being delivered to the market that allow connection of existing storage resources to a SAN FC fabric. This will provide customers with a Fibre Channel migration path.

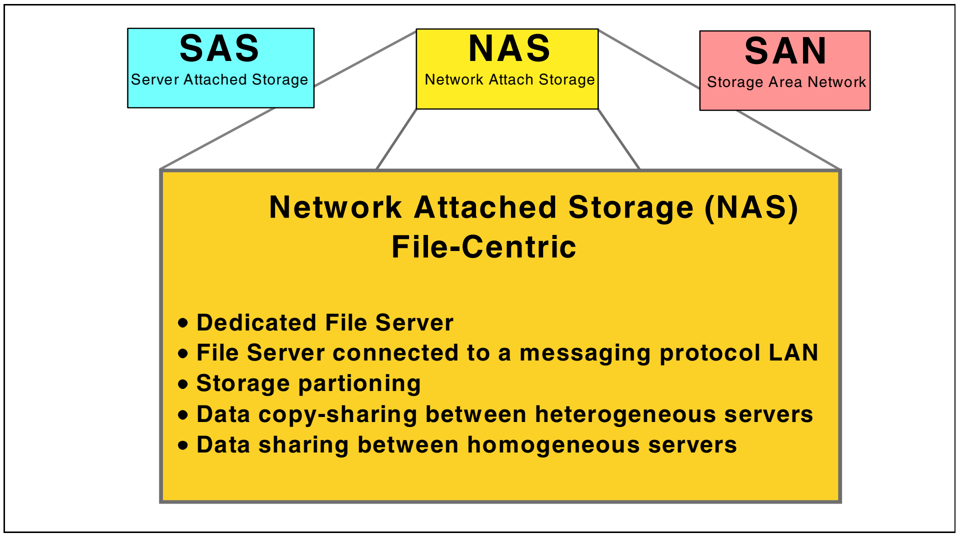


Figure 8. File-Centric

1.4.3 Data-Centric

As depicted in Figure 9 on page 25, there are some major obstacles which need to be overcome to reach a true Data Centric Storage Area Network. Today, IBM is aggressively working with the industry to use its years of experience and expertise to deliver the next generation SAN. IBM's accomplishments include the following:

- Invented System Managed Storage (SMS) in the 1970's
- Gained SAN management expertise with ESCON Manager
- Has extensive experience with Storage Management software (ADSM)
- Was involved in the initial definition of the Fibre Channel initiative, and is active in the SAN standards organizations, including the Fibre Channel Association and the Storage Networking Industry Association.
- Is aggressively working with the industry to leverage its years of experience, expertise, and intellectual property in first-generation SAN technology and SAN management in the S/390 world to deliver second-generation SANs, enabling all the benefits of a switched network topology in today's open environment.
- Has already been established in the intelligent server market (Versatile Storage Server and Enterprise Storage server)
- Has an Enterprise Storage Resource Management framework (StorWatch)

• Has a world-leading research and development (R&D) organization (IBM Research)

• Has a long record of experience and leadership in storage, servers, networking, systems management, and services

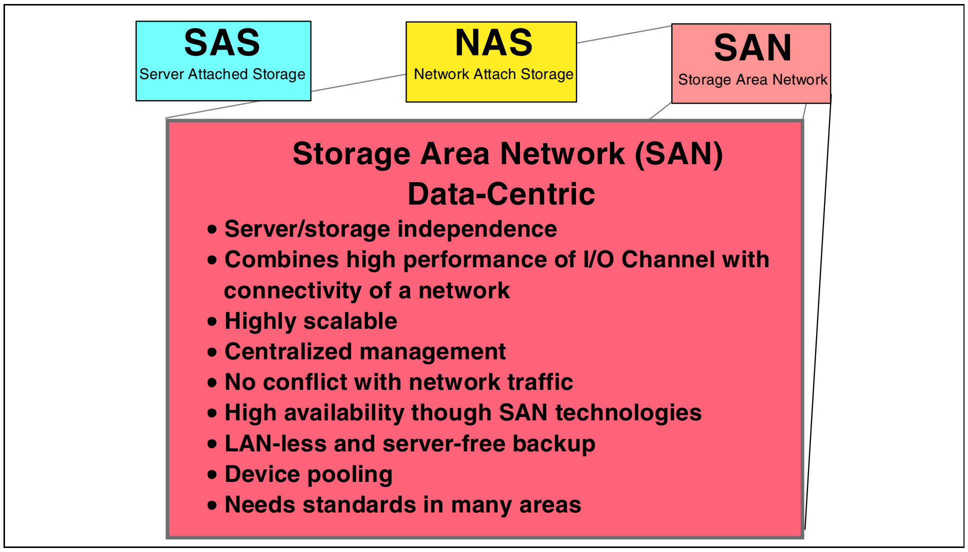


Figure 9. Data-Centric

Chapter 2. The IBM SAN Initiative

The evolution of Information Technology (IT), coupled with the explosion of the Internet as a means for individuals and companies to communicate and exchange information, has led to the rise of a new business model that is commonly referred to as e-business, and in some cases e-commerce.

2.1 New Paradigms: e-Business

e-business is becoming a strategic imperative for many companies worldwide. The driving force behind this is the need to cut costs, thereby improving profits. A second but equally important reason is the need and desire to expand a company's reach. e-business allows you to communicate easily and cheaply with customers anytime, anywhere.

e-business is changing traditional business models because it is network-centric, global, and data-intensive. The foundation of e-business is the network itself, and the most widely used network today is the Internet. The Internet is global in reach as it can be accessed from anywhere around the globe. The term data-intensive describes the fact that all electronic transactions can be stored for processing and subsequent analysis. Today we are faced with an explosion of data generated by e-business applications. This data must be stored and managed appropriately. There is also a requirement to distill information out of raw data using data mining applications.

These industry trends are driving the requirement for more flexible and reliable access to data. The data resides on storage devices. This is driving the requirement for SANs. Enterprise Resource Planning (ERP), data mining and decision support applications are also driving the requirement for SANs because the data involved has to be accessed from or copied between heterogeneous environments.

In the following sections we will discuss the technology trends driving storage evolution and the development of SANs.

2.1.1 Storage Technology Trends

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, recentralization of server and storage assets. This trend is driven by the 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 a black box 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 black box may be completely different from the original technology.

The above issues lead to the trends that are driving SANs, as discussed in the following section.

2.1.2 Trends Driving SANs

The trends driving the evolution of storage are also the trends that are driving the evolution of SANs. Data is the basis for most of the business value created by information technology, and the storage infrastructure is the foundation upon which data relies.

The storage infrastructure needs to support business objectives and new emerging business models. Some of the requirements of today's companies 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 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 require the ability to fence off one system's data to other systems.

- 24x7 response and availability. This is an availability requirement that implies both protection against media failure, possibly using RAID technologies, as well as ease of data migration between devices without interrupting application processing.

Current technologies, especially SCSI technology, have limitations in terms of distance, addressability, and performance. Typically, SCSI devices can be connected to a limited number of systems, and reconnecting devices to different systems is a manual operation that requires unplugging and replugging SCSIs to different machines.

SANs mirror the e-business model for storage. They hold the following promises:

- Storage consolidation. Today, storage is mostly server-attached and cannot be easily moved. Storage consolidation holds the promise of attaching storage to multiple servers concurrently and leveraging investments.
- Storage sharing. Once storage is attached to multiple servers, it can be partitioned or shared between servers.
- Data sharing. 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.
- 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, alternate pathing, etc.
- 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. Facilitate both data and storage subsystem migration from one system to another without interrupting service.
- Centralized management. The ability to manage multiple heterogeneous boxes from one single point of control.

Most of the functions described above are not available today, as we will see in the following section.

2.1.3 SAN Status Today

Today we are early in the SAN technology life cycle. The technology is still relatively immature and there is also a lack of accepted industrial standards. Data integrity is an important issue. Today, many current implementations rely on the physical wiring only for security, but other security schemes are being developed. For example, some vendors are developing a logical partitioning approach called zoning.

Most of the SAN solutions on the market today are limited in scope and restricted to specific applications. Interoperability is not possible in many of these solutions. Some of the currently available solutions are:

- Backup and recovery
- Storage consolidation
- Data replication
- Virtual storage resources

Most of the SAN management packages available today are limited in scope to the management of one particular element or aspect of the SAN. Most of these solutions are proprietary and are lacking in interoperability.

We will now discuss what the industry requires for SANs to enter the mainstream of computing and deliver on their true promise.

2.1.4 SAN Evolution

The first requirement is the definition of standards. Standards are the base for interoperability of devices and software from different vendors. The SNIA organization is active in defining these standards. Most of the players in the SAN industry recognize the need for standards, as these are the basis for wide acceptance of SANs. Widely accepted standards allow for heterogeneous, cross-platform, multi-vendor deployment of SAN solutions.

There is a requirement for the expansion of the scope of SANs to enterprise-wide environments, with 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 need to be improved to guarantee data integrity. There is a requirement for complex, end-to-end SAN management from a single point. Last, but probably most important of all, SANs should evolve to support an open, standards-based environment. Figure 10 on page 31 illustrates a SAN to which multiple systems can connect. SAN is the key to moving digital information from operating system islands to universal access.

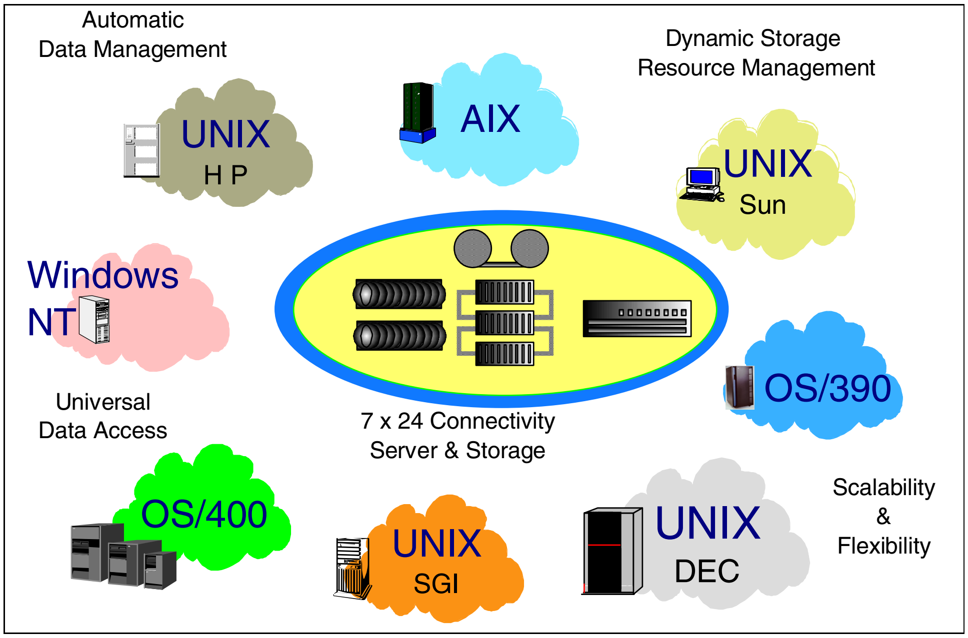


Figure 10. Storage with a SAN

Having discussed the market requirements, in the following section we will discuss how IBM is responding to the challenge.

2.2 The IBM Response

IBM: ***A storage area network (SAN) is a dedicated, centrally managed, secure information infrastructure, which enables any-to-any interconnection of servers and storage systems***.
Source: `www.storage.ibm.com/ibmsan/basics.htm`

IBM has recognized industry needs and is addressing them with the ***IBM SAN Initiative***. This initiative draws upon decades of IBM leadership in enterprise computing across all platforms. IBM has established a cross-division task force to coordinate and leverage the experiences across all divisions. IBM also closely cooperates with standards bodies to define industry standards and promote interoperability across multiple vendor devices.

IBM will deliver interoperability-tested, certified value-add solutions that provide universal access to data. IBM SAN solutions will consist of best-of-breed IBM and non-IBM software, servers, SAN fabric components and storage. IBM Global Services will develop and implement the best practices required to support the new generation of SAN technology.

Figure 11 on page 32 outlines IBM's organizational approach to the SAN initiative. The IBM SAN initiative is an IBM corporate commitment to provide the complete range of services, infrastructure, and technology required for businesses to successfully implement a SAN. The SAN value-add solutions layer covers the solutions that provide the true customer value of SAN technology. Software, servers, fabric and storage are the building blocks used to construct, use and manage the SAN. The services and system integration components are critical to the success of the SAN initiative because of the complexity of heterogeneous, multi-platform environments.

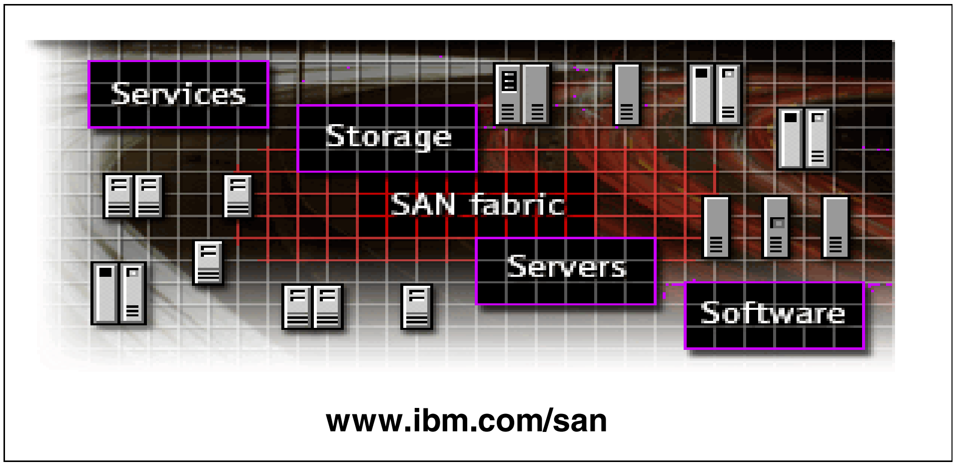


Figure 11. IBM Organizational Approach to SANs

In the following section we will discuss the promise of value for the IBM SAN initiative.

2.2.1 The Promise of Value

IBM is orchestrating its knowledge, expertise, and relationships in computing to provide the complete range of services, infrastructure, and technology required for its customers to capitalize on the power of Storage Area Networks by:

- Combine reach of network computing with manageability of a data center
- Enhance security and data integrity of network computing architectures

- Provide anywhere/anytime universal access to data regardless of the platform, source, format, or application in which the data is archived

Figure 12 on page 33 illustrates the following concepts:

- The bottom layer is the hardware, both ***servers*** and ***storage*** devices, that need to be interconnected and enabled for SAN connectivity. IBM will support a phased implementation of FC technology through both tactical and strategic moves. Tactically, IBM will deliver bridges (gateways) that enable FC connectivity to current devices and systems. Strategically, IBM will implement ESCON to FICON and SCSI/SSA to FC migrations.

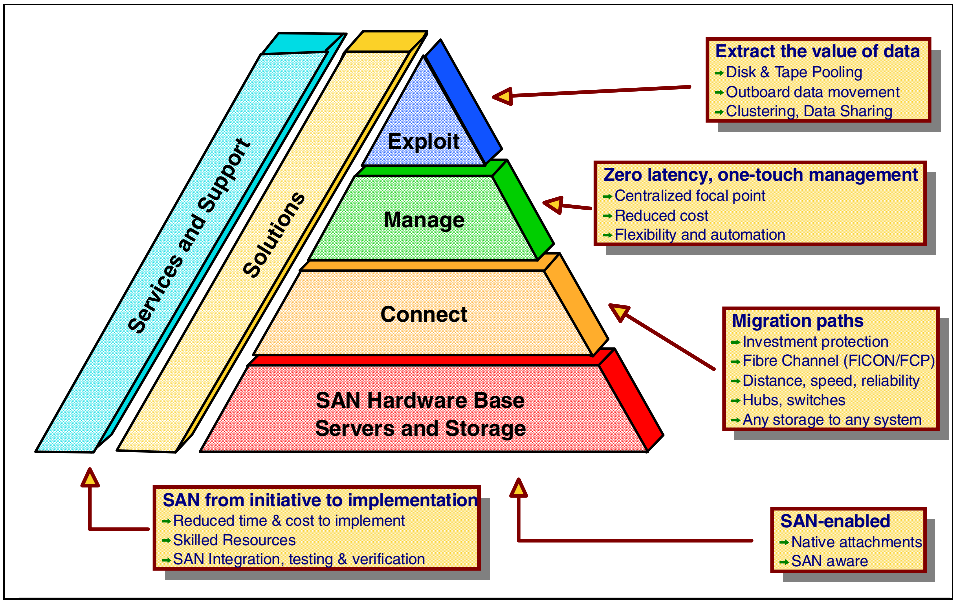


Figure 12. Outline of IBM SAN Promise of Value

- The next layer, ***connectivity***, is the physical storage network. IBM will provide the necessary storage network hardware to build SANs. IBM will also deliver frameworks and tools to build platform-independent networks. IBM will provide a flexible and scalable building block infrastructure for plug-and-play interoperability across different vendors and platforms.
- The third layer is ***management***. IBM will port mature and proven management technologies to the open systems world. Some of these technologies are first generation (ESCON) SAN switch fabric management and system managed storage concepts from OS/390 systems. IBM is also working with Tivoli to develop a zero-latency, one-touch management

infrastructure that will allow all elements in the SAN to be centrally monitored and managed.

- The fourth layer is ***exploitation***. This is the fundamental reason for the introduction of a SAN. IBM intends to deliver functions such as disk and tape pooling, data migration, backup, and recovery that utilize outboard data movement, as well as clustering and data sharing solutions.

Across all layers lie ***services and support***. Implementing a SAN requires an understanding of how to deploy the components of this emerging technology. SAN's any-to-any connectivity, increased capacity, and complexity highlight the critical nature of system integration and services in any SAN strategy. As a result of this, IBM is gearing up its IBM Global Services division to provide support, services, and education needed to support end-to-end SAN ***solutions***.

What business value does a SAN provide? There are various value propositions that may be followed, depending on your requirements.

2.2.1.1 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.

2.2.1.2 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 so as 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.

2.2.1.3 Increased Business Flexibility

Increased business flexibility is achieved with true data sharing, where data can by accessed by multiple systems running different operating systems. This leads to reduced data movement and less data transformation between multiple formats. 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 its customers and suppliers. This value proposition probably has the highest ROI, but is the most difficult to measure.

2.2.2 IBM Global Services

The deployment of a SAN solution in a multi-platform, heterogeneous environment can involve quite a large effort. IBM Global Services (IGS) has the global reach breadth of skills, and state-of-the-art comprehensive range of services to assist you in harnessing the power offered by SANs. IGS is setting up worldwide interoperability centers to enable customers to test and validate SAN solutions using both IBM and non-IBM components.

The SAN promise of any-to-any connectivity and the complexity it introduces highlight the critical nature of both systems integration and services in any SAN strategy. IBM Global Services, in combination with service provider partners, will offer the services and education required to plan and support end-to-end SAN solutions.

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

`www.ibm.com/san`

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 defined and in place for SANs to enter 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 97.

Chapter 3. SAN Servers and Storage

The whole basis of SAN evolves around storage and the connected servers. Figure 13 on page 37 illustrates this base layer, which forms the basis for the SAN solutions. This chapter will present and discuss storage: its evolution, its importance, and how different server platforms address their storage needs.

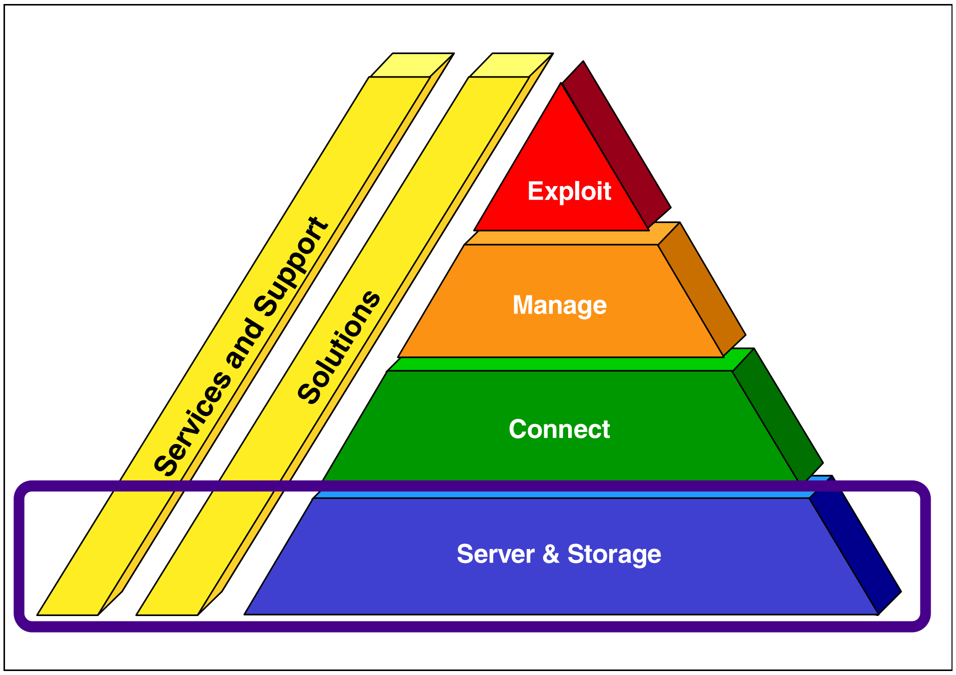


Figure 13. Outline of IBM SAN Initiative

The growth of data, and the increasing dependency of businesses on this data, have 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 disks drives, from a few megabytes (MB) to multi-gigabytes (GB) per disk drive.
- Increased storage performance; for example, 5 MB/sec for SCSI-1 to 80 MB/sec for SCSI-3, together with faster disk and tape devices.
- Different storage device and server interconnects; for example, SCSI, SSA, FC-AL, FCP, ESCON, FICON, and others.

- Data protection, using solutions like mirroring, remote copy, etc.
- Storage reliability, using various RAID implementations.
- Data and storage security and manageability.

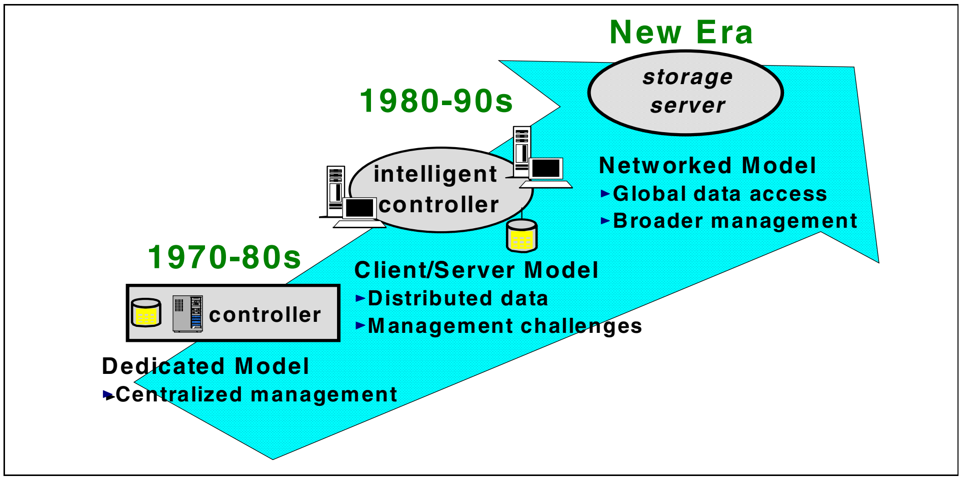


Figure 14. Evolutions in Storage Architecture

Figure 14 on page 38 illustrates the evolution of storage architecture in relation to era or phase of computing, from the centralized computing with controller-based dedicated storage, to the Client/Server model with distributed data, and finally to the current networked era calling with its requirement for universal access to data, robust software tools, and data management solutions.

3.1 Servers and Storage Environments

Figure 15 on page 39 illustrates the different islands of information or data as they exist due to the different operating systems and hardware platforms.

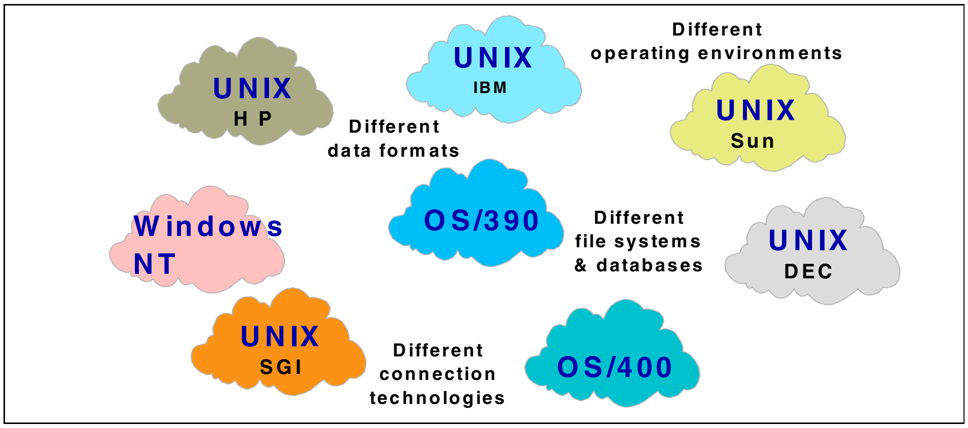


Figure 15. Islands of Distributed Information

In today’s environment, many UNIX, Windows NT, and OS/390 users are either planning for, or actually in the process of, implementing server, disk, and/or tape storage consolidation. These customers have experienced the escalating costs associated with managing information residing on servers and storage distributed about their enterprise, not to mention the difficulty in implementing and enforcing business recovery procedures and providing consistent access security 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.

PC and UNIX servers with SCSI channels face many limitations in achieving this goal: bulky cables, shared bus architecture that limits performance due to the bus arbitration required for the half-duplex, one-at-a-time I/O operations; limited distance of 25 meters; and limited addressing of up to 15 storage devices per channel.

Many PC and UNIX server users have already migrated to the IBM Serial Storage Architecture (SSA) with the IBM 7133 serial disk storage. SSA, which supports serial loops with up to 128 devices and extended distances of up to 10 kilometers with fiber optic extenders, has eliminated many of the SCSI limitations.

S/390 servers have been using the ESCON channels technology for almost 9 years to implement SAN solutions like remote mirroring. ESCON is a serial transfer, switched point-to-point Fibre Channel technology that replaced mainframe parallel channels.

S/390 FICON channels are a technology upgrade to the S/390 ESCON channel, using industry standard Fibre Channel technology. FICON major benefits are higher performance and more extensive addressing. Since S/390 FICON is an FC-4 protocol, it can share the same infrastructure with other Fibre Channel protocols, such as FCP.

Fibre Channel Arbitrated Loop FC-AL provides many benefits, including smaller cables and connectors, faster data transfers, and a wide range of Fibre Channel connection options, including point-to-point, loop, and switched fabric. Distances are up to 10 kilometers, and loop addressing supports up to 127 storage devices.

3.1.1 The Challenge

The real challenge today is to implement a true heterogeneous storage and data consolidation across different hardware and operating systems platforms; for example, disk and tape sharing across OS/390, UNIX, and Windows NT.

Figure 16 on page 40 illustrates the consolidation movement from the distributed islands of information to a single heterogeneous configuration.

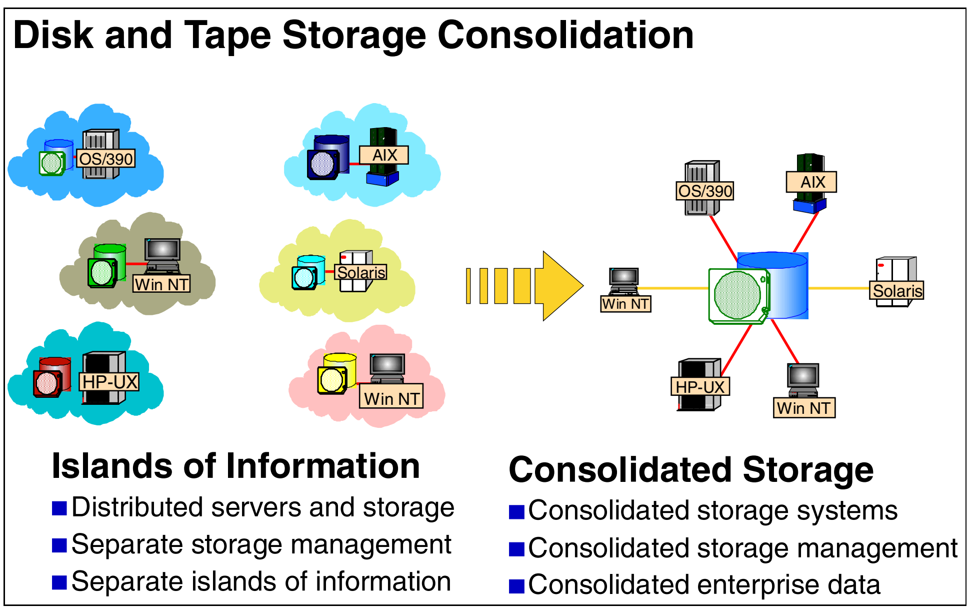


Figure 16. Disk and Tape Storage and Data Consolidation

IBM's Seascape architecture addresses this movement from islands of distributed data to a single heterogeneous SAN. For more information on storage consolidation, see *IBM Storage Solutions for Server Consolidation*, SG24-5355. Figure 17 on page 41 shows the underlying principles of the IBM storage strategy. The Seascape architecture — with its concept of integrating leading technologies from IBM to provide storage solutions for all platforms, together with SAN, consolidation of servers, storage, applications, and data — is a part of that strategy.

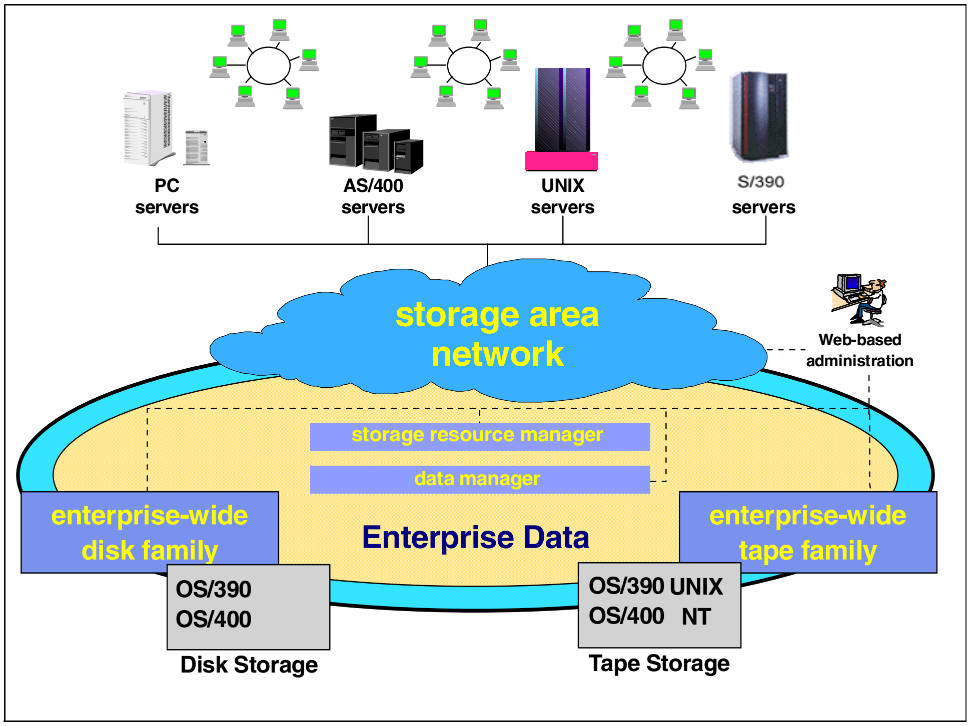


Figure 17. IBM Storage Strategy

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 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.

Figure 18 on page 42 is a brief summary of these differences and the associated system architecture layers, excluding OS/400 for simplistic reasons.

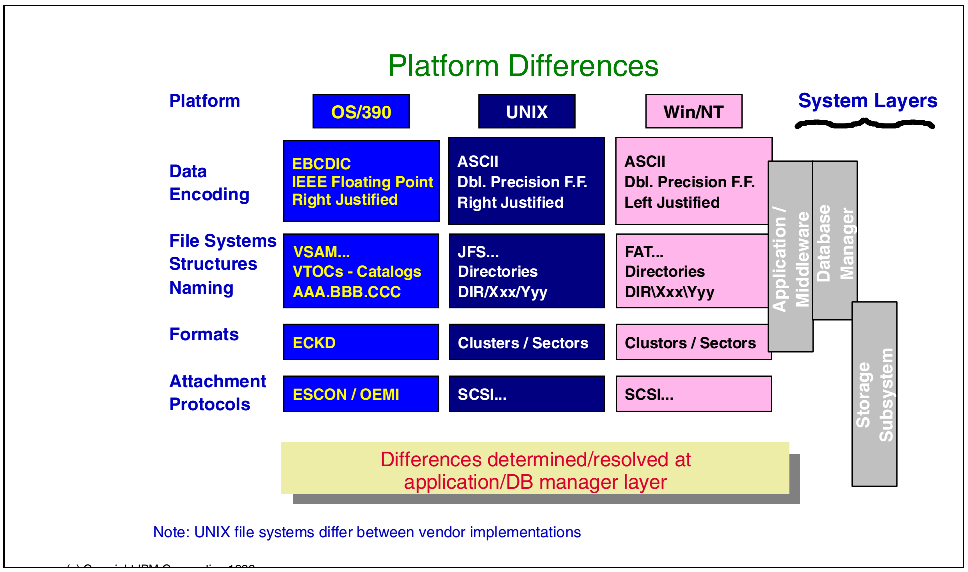


Figure 18. Hardware and Operating Systems Differences

3.2 Server Environments

Each of the different server platforms — OS/390, UNIX (AIX, HP, SUN, and others), OS/400, and Windows NT (PC Servers) — have implemented SAN-like 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 S/390 Servers

The System/390 (S/390) is a processor(s) and Operating System set. Historically, S/390 servers have supported many different operating systems, such as OS/390, VM, VSE, and TPF, that have been enhanced over the years. The processor to storage device interconnections have also evolved from a Bus and Tag interface, to ESCON channels, and now to FICON channels. Figure 19 on page 43 shows the various processor to storage interfaces on the S/390.

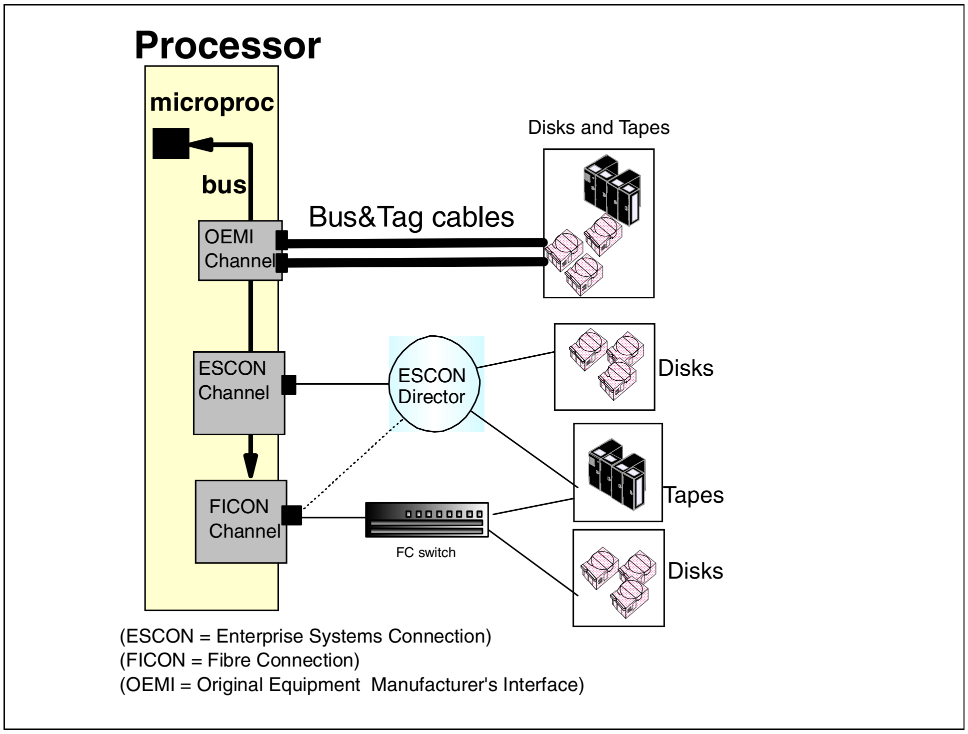


Figure 19. S/390 Processor to Storage Interface Connections

S/390 participation in the evolving Storage Area Networks (SANs) is a logical extension of existing S/390 environments. For more information, see the S/390 Web site at:

`www.s390.ibm.com/san/`

3.2.2 RS/6000 Servers and UNIX Servers

The IBM RS/6000, running a UNIX operating system called AIX, offers various processor to storage interfaces, including SCSI, SSA, and FC-AL. The SSA interconnection has primarily been used for disk storage. The FC-AL interface currently being deployed for disk storage only. Today, tape devices connect through Fibre Channel gateways to the FC-AL adapters, but direct FC attached tape devices are likely to be announced very soon.

Figure 20 on page 44 shows the various processor to storage interconnect options for the RS/6000 family.

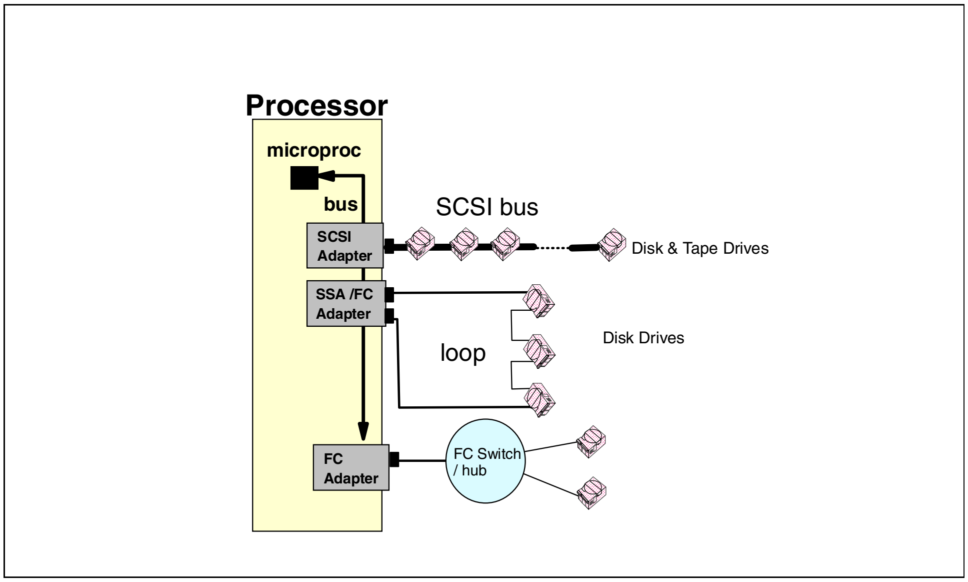


Figure 20. RS6000 Processor to Storage Interconnections

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 RS/6000, as shown in Figure 20 on page 44.

RS/6000 is committed to delivering world-class SAN solutions based on Fibre Channel so that our customers can have the type of information access and manageability that doing business in the next century will require. For more information, see the RS/6000 Web site at:

`www.rs6000.ibm.com/hardware/san`

3.2.3 AS/400 Servers

The AS/400 system architecture is defined by a high-level machine interface, referred to as Technology Independent Machine Interface (TIMI), that 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). Additionally, the AS/400 implements I/O processors that are architected for *end user* technologies, including Windows NT and Novell. Figure 21 on page 45 summarizes the various modules of an AS/400 hardware architecture.

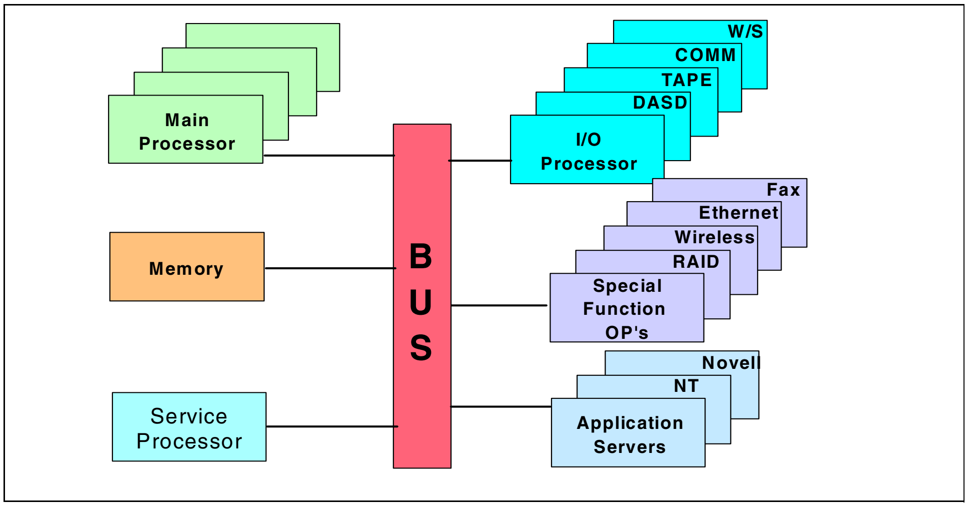


Figure 21. AS400 Hardware Design

Several architectural features of the AS/400 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

3.2.3.1 Single Level Storage

Single-level storage is probably the most significant differentiator in a SAN solution implementation on AS/400, as compared to other systems like OS/390, UNIX, and Windows NT. In OS/400, both the main storage (memory) and the secondary storage (disks) are treated as a very large virtual address space known as Single Level Storage (SLS).

Figure 22 on page 46 compares the OS/400 SLS addressing with the way Windows NT 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.

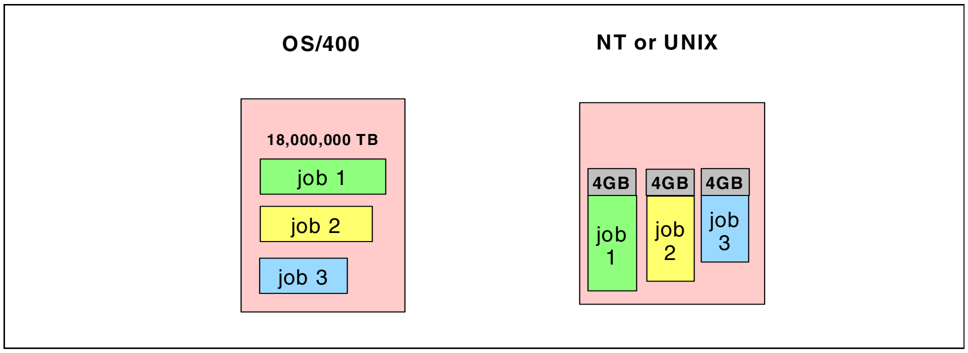


Figure 22. OS/400 versus NT or UNIX Storage Addressing

AS/400 intends to support SAN based storage solutions in the future. These solutions will initially involve Fibre Channel technology, hubs, external IBM storage subsystems, and high performance tape drives.

For more information, see the AS/400 Web site at:

`www.as400.ibm.com/periph/san.htm`

3.2.4 Windows NT and Netfinity Servers

Based on the reports of various analysts regarding growth in the Windows NT server market — both in number and size of Windows NT servers — Windows NT will become the largest market for SAN solution deployment. More and more Windows NT 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 Netfinity servers (IBM's INTEL-based processors that support the Microsoft Windows NT Operating System) are quite similar to those supported on UNIX servers, like SCSI, SSA, and FC-AL. Figure 23 on page 47 illustrates the various processor to storage interconnects.

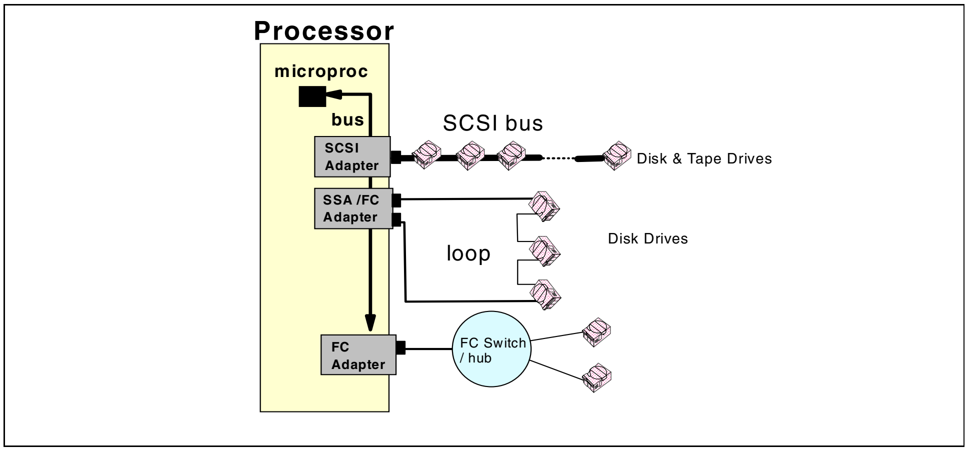


Figure 23. Netfinity Processor to Storage Interconnections

For more information, see the Netfinity Web site at:

`www.pc.ibm.com/us/netfinity/storesvr.html`

3.3 Disclaimer Statement

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Chapter 4. SAN Fabrics and Connectivity

4.1 Fabric

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 utilizing hubs, switches and gateways.

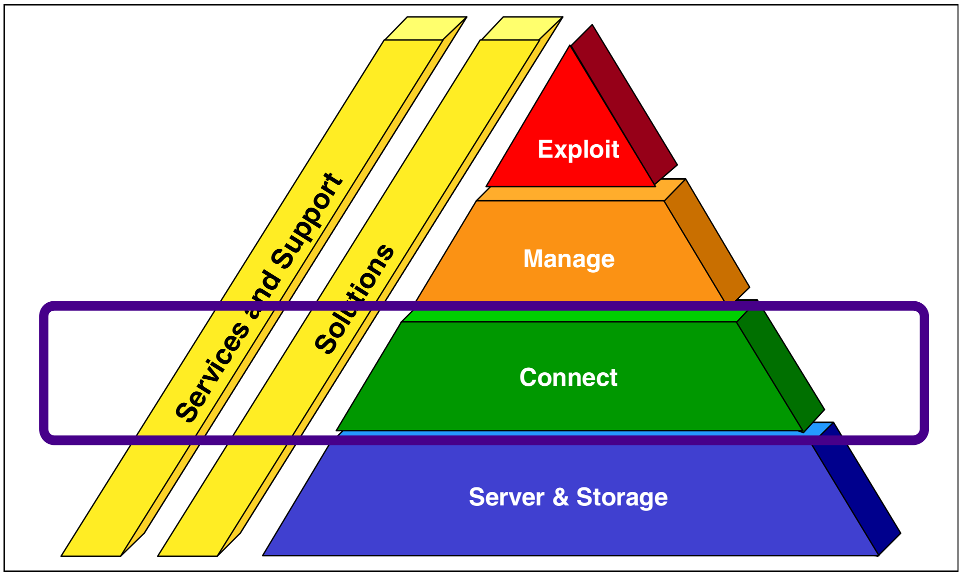


Figure 24. Outline of IBM SAN Initiative

4.2 Storage Area Network 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 area networks that may consist of hundreds or thousands of devices. Fibre Channel Storage Area Networks support high bandwidth storage traffic on the order of 100 MB/sec, and enhancements to the Fibre Channel standard will support even higher bandwidth in the near future.

Storage subsystems, storage devices, and server systems can be attached to a Fibre Channel Storage Area Network. Depending on the implementation, several different components can be used to build a Storage Area Network.

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

Different types of interconnect entities allow Fibre Channel networks of varying scale to be built. In smaller SAN environments, Fibre Channel arbitrated loop topologies employ hub and bridge products. As SANs increase in size and complexity, Fibre Channel switches 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.

4.2.1 Server-Attached Storage (SAS)

Server-attached or bus-attached storage operates through the server. Early on, central host mainframe processors could access a variety of storage devices through *channels*. The central processor managed the processing load, while the I/O processor, also within the mainframe, sent and received data to and from storage devices asynchronously. This was the first stage of introducing intelligence into the data transfer process. Server-attached storage architectures dominated the scene for several years, from mainframe processor channels to PC server bus slots and adapters. See Figure 25 on page 50.

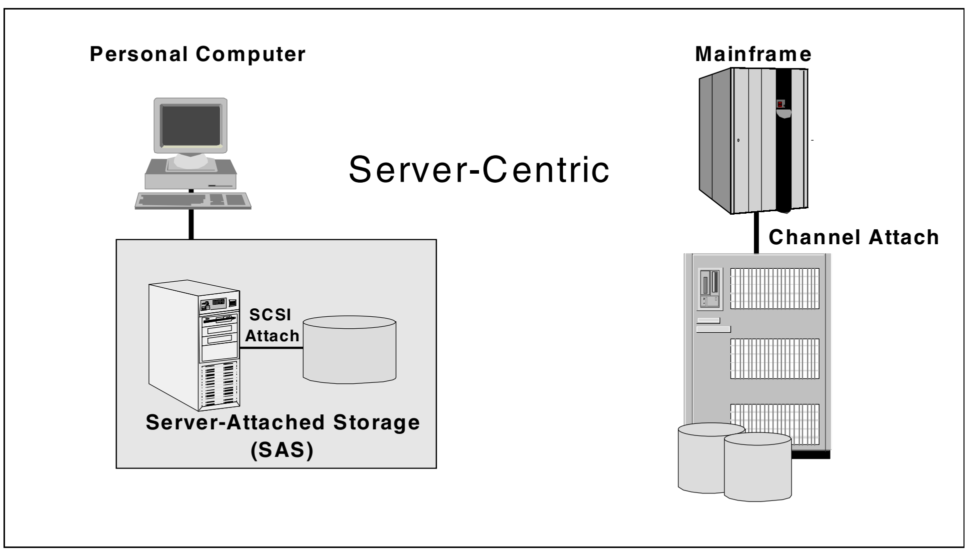


Figure 25. Server-Attached Storage (SAS)

4.2.2 Network-Attached Storage (NAS)

Local Area Networks came along next, and helped in the sharing of data files among groups of desktop microcomputers. Soon it became clear that LANs would be a significant step toward distributed, client/server systems. Large-scale client/server systems were then constructed, tying sizable numbers of LANs together through Wide Area Networks (WANs). The idea was to leverage cheap microcomputers and cheap disk storage to replace expensive (but reliable) central computers. See Figure 26 on page 51.

NAS describes technology in which an integrated storage system, such as a disk array or tape device, connects directly to a messaging network through a LAN interface, such as Ethernet, using messaging communications protocols like TCP/IP. The storage system functions as a server in a client/server relationship. It has a processor and an operating system or micro-kernel, and it processes file I/O protocols such as Network File System (NFS) to manage the transfer of data between itself and its clients.

An example of network-attached storage is the IBM 3466 Network Storage Manager (NSM). For more information on NSM, see the Web site:

`www.storage.ibm.com/nsm/nsmhome.htm`

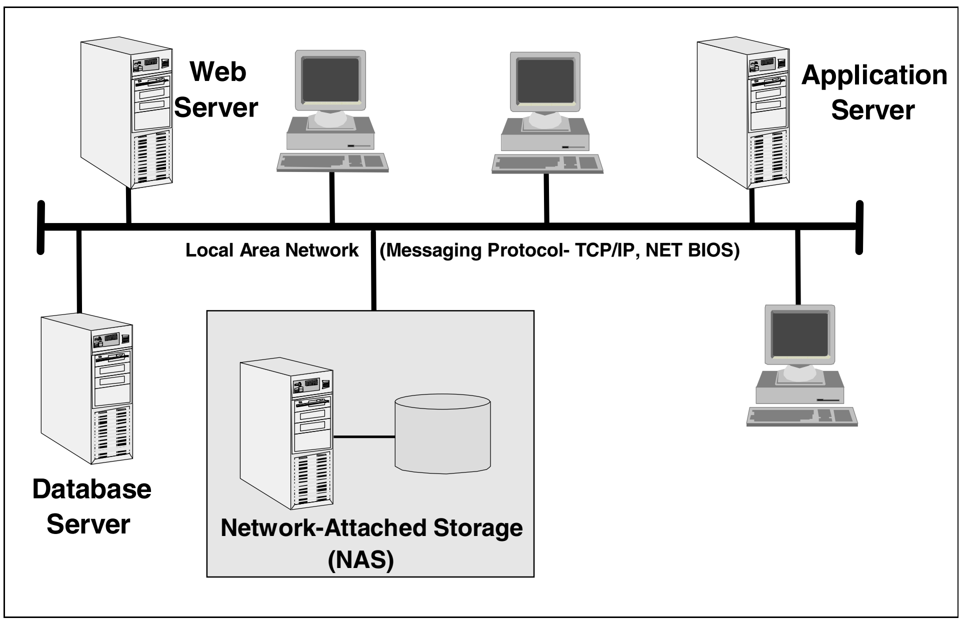


Figure 26. Network Attached Storage (NAS)

4.2.3 Storage Area Network (SAN)

A SAN is a dedicated high performance network to move data between heterogeneous servers and storage resources. Unlike Network Attached Storage (NAS), a separate dedicated network is provided which avoids any traffic conflicts between clients and servers on the traditional messaging network. See Figure 27 on page 52.

A Fibre Channel based SAN combines the high performance of an I/O channel and the advantages of a network (connectivity and distance of a network) using similar network technology components like routers, switches and gateways.

Unlike NAS products, SAN products do not function like a server. Instead, the SAN product processes block I/O protocols, such as FCP (SCSI) or FICON, for some other system, quite possibly a server.

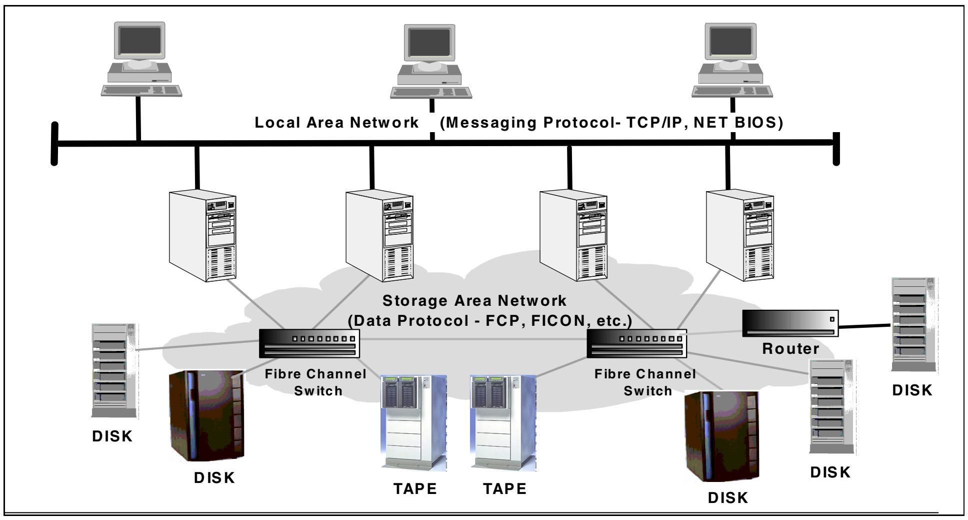


Figure 27. Storage Area Network (SAN)

4.3 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, loop (arbitrated), and star (switched). These can be stand-alone or interconnected to form a fabric.

4.3.1 Point-to-Point

Point-to-point is the easiest Fibre Channel configuration to implement, and it is also the easiest to administer. This simple link can be used to provide a high-speed interconnect between two nodes, as shown in Figure 28 on page 53.

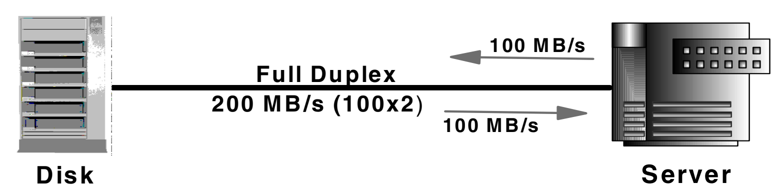


Figure 28. Fibre Channel Point-to-Point Topology

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

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.

4.3.2 Loops and Hubs for Connectivity

The Fibre Channel arbitrated loop 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.
- 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 29 on page 54 shows an FC loop using a hub.

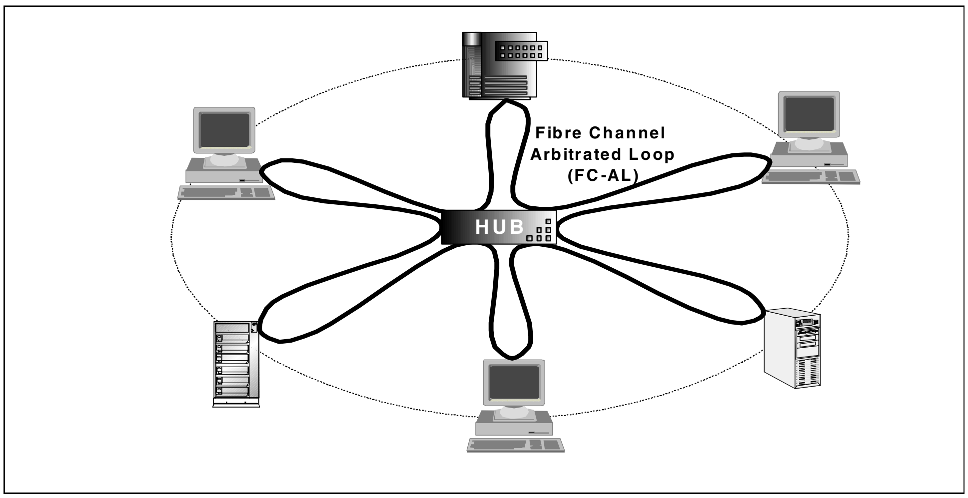


Figure 29. Fibre Channel Loop Topology

4.3.3 Switches for Scalability, Performance and Availability

Fibre Channel switches function in a manner similar to traditional network switches to provide increased bandwidth, scalable performance, increased number of devices, and in some cases, increased redundancy. Fibre Channel switches vary in the number of ports and media types they support. Multiple switches can be connected to form a switch fabric capable of supporting a large number of system hosts and storage subsystems as shown in Figure 30 on page 55. When switches are connected, each switch's configuration information has to be copied (cascaded) into all other participating switches.

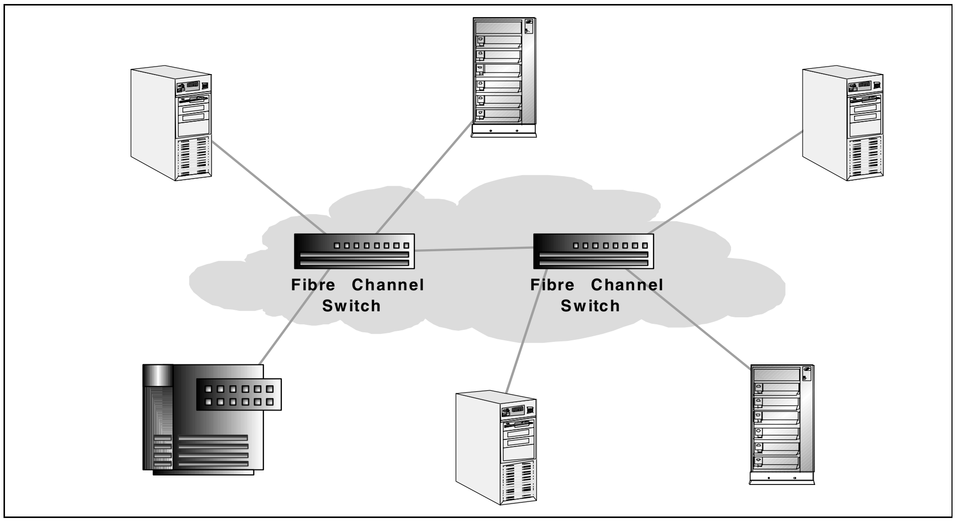


Figure 30. Fibre Channel Switched Topology

4.3.4 Switched Fabric without Cascading

Switched point-to-point (without cascading) can support configurations required by very large systems (IBM's ESCON directors, for example). The configuration can be set up to allow every system to have access to every switch, and every controller to be connected to at least two switches. This allows any system to get to any controller/device, and it allows for continuous operation (with degraded performance) in the event that a switch fails. An example of a fabric with non-cascaded switches is shown in Figure 31 on page 56.

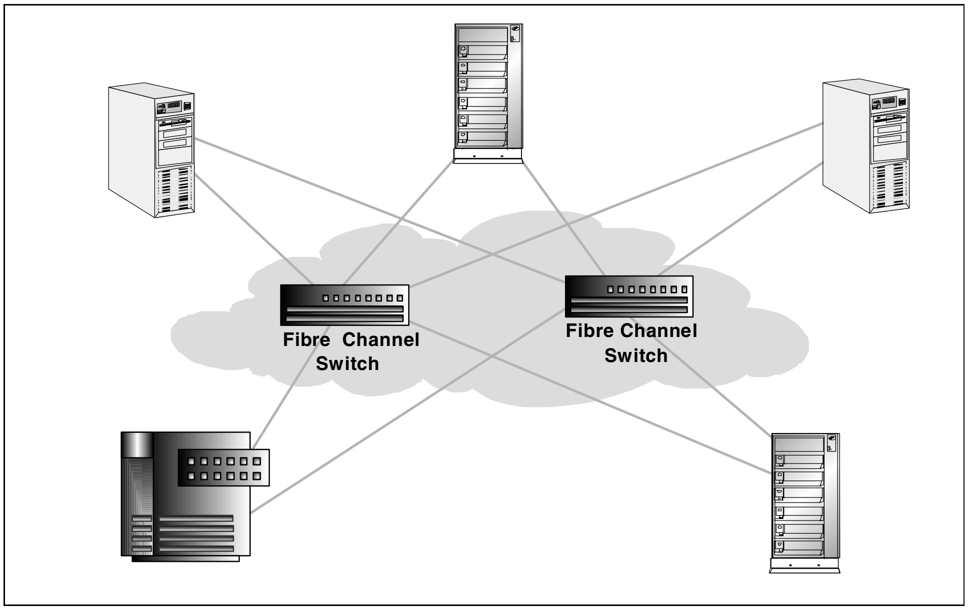


Figure 31. Fibre Channel Switched Topology (Non-Cascaded Switches)

4.3.5 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. Inter-switch links can fail and must be identified (many switch vendors do not yet support any reporting on inter-switch links). 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) 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 32 on page 57.

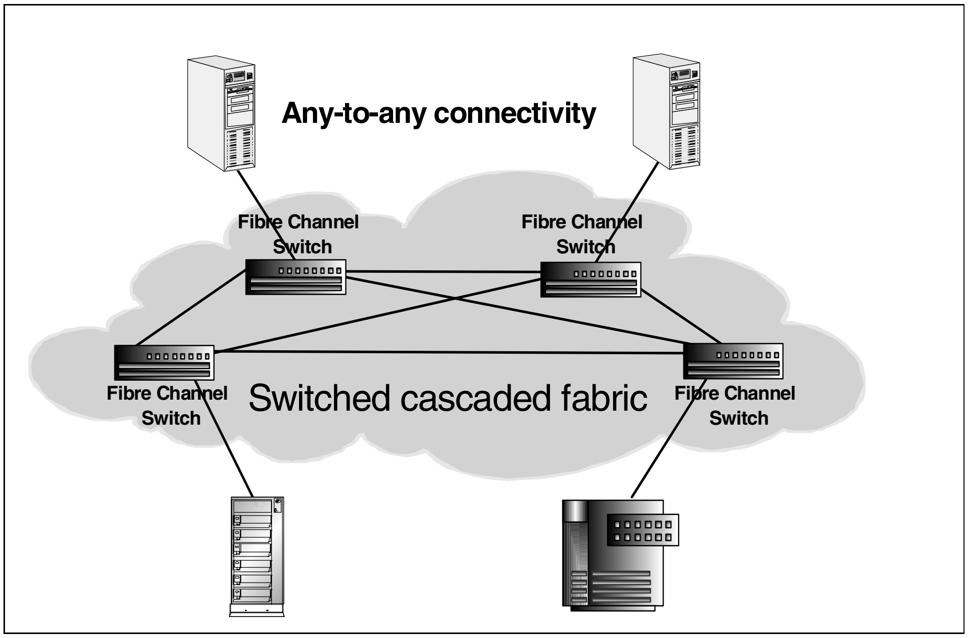


Figure 32. Fibre Channel Switched Topology (Cascaded Switches)

4.3.6 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 100 MB/sec, shared among all the connections, while a switch can support 100 MB/sec between any two ports. 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.

Finally, switches enhance availability. In a pure loop configuration, a host reset or device failure can cause a loop initialization that resets all the other devices on the loop; this interrupts the loop's availability to other servers and storage devices. A switch can prevent this kind of error propagation and can also enable rapid isolation and repair of hardware device failures.

4.3.7 Mixed Switch and Loop Topologies

To mix Fibre Channel Arbitrated Loop with Fibre Channel switches, the switch must have a specific Fiber Loop (FL) port to attach to the loop. Discovery becomes more complex, and management software must be able to deal with the combination of hubs and switches in the configuration, for example, for the configuration shown in Figure 33 on page 58.

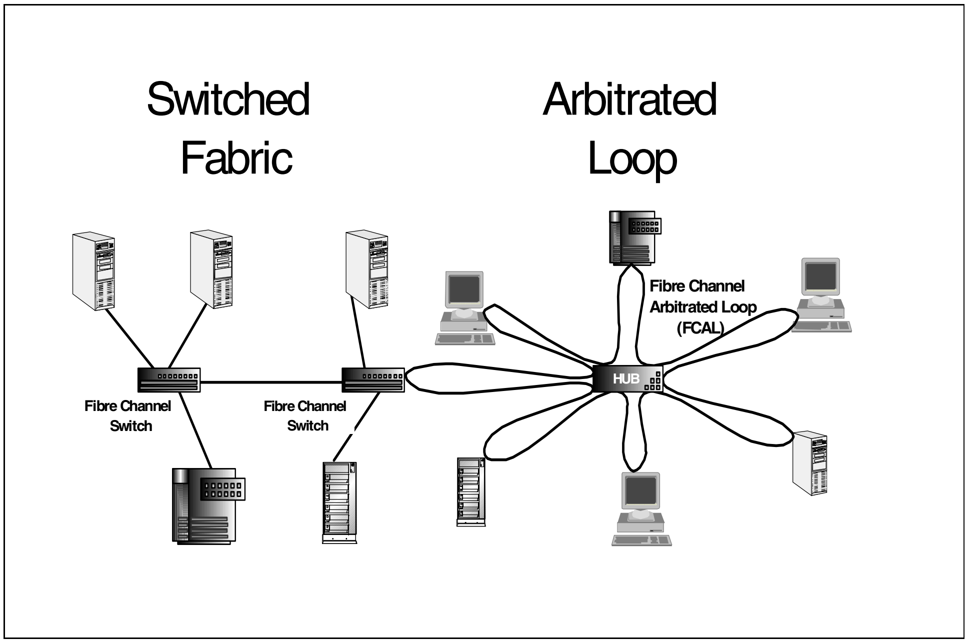


Figure 33. Switch Attached Arbitrated Loop

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.

4.3.8 ESCON Directors

Using switched technology, ESCON Directors provided the first true homogeneous switched Storage Area Network. ESCON Directors provide a completely separate network dedicated to move data. Data can be transferred at a rate of 17 MB/sec, and dynamic connections can be established between attached ESCON control units and channels. See Figure 34 on page 59.

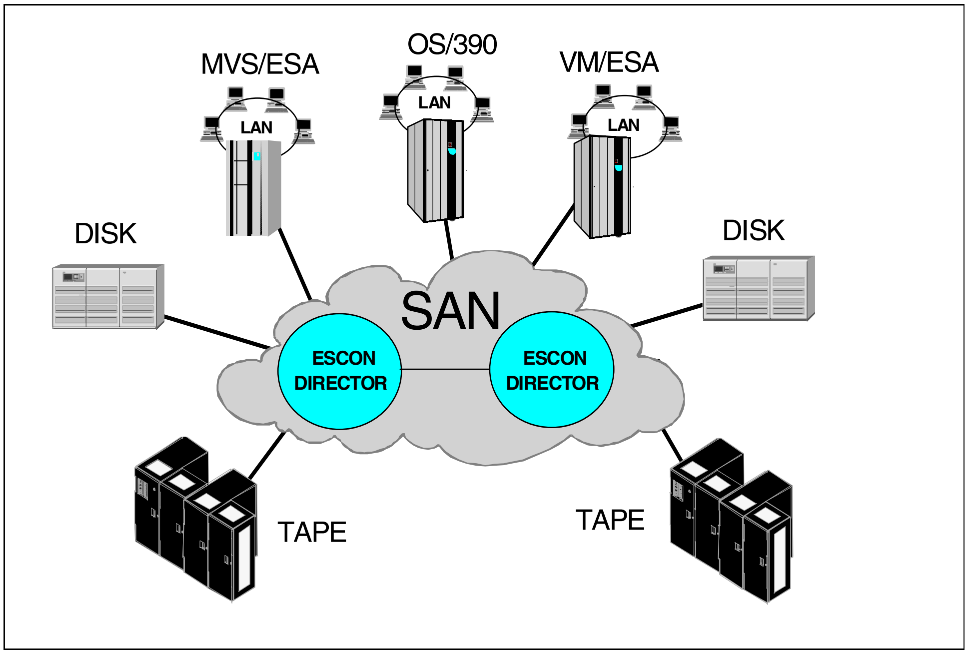


Figure 34. ESCON Storage Area Network

4.3.9 FICON Directors

FICON Directors are standard Fibre Channels switches with additional function to provide management from OS/390. They provide dynamic connections between FICON channels and control units with FICON interfaces, as shown in Figure 35 on page 60.

Since the FICON Director is a standard Fibre Channel switch, non-S/390 servers and storage systems can also be attached.

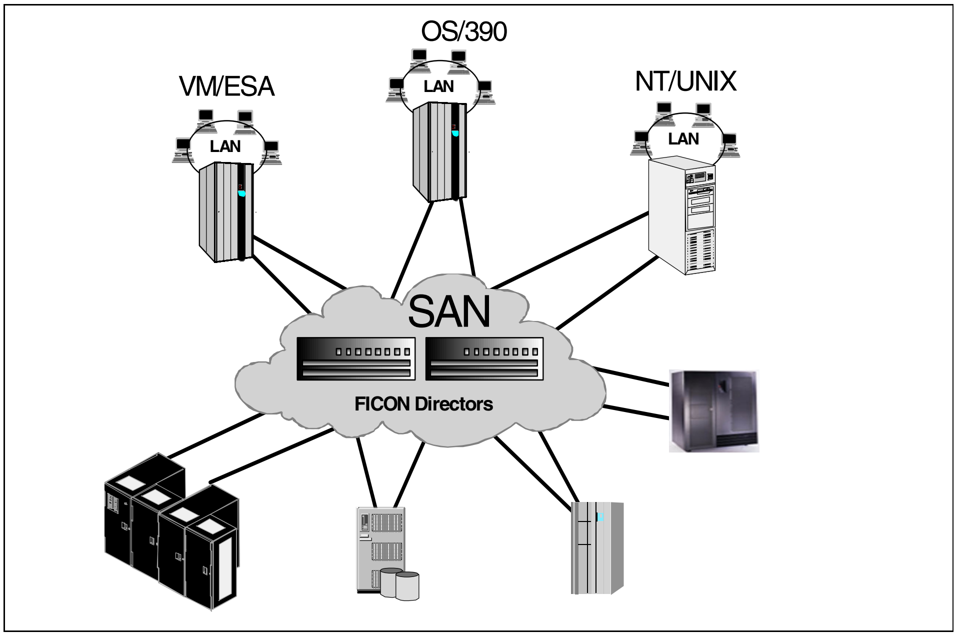


Figure 35. FICON Switched Storage Area Network

4.3.10 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, which is actually a router (despite its name). This router (named Gateway) component allows up to four Ultra SCSI back-end interfaces and two front end Fibre Channel interfaces, as shown in Figure 36 on page 61.

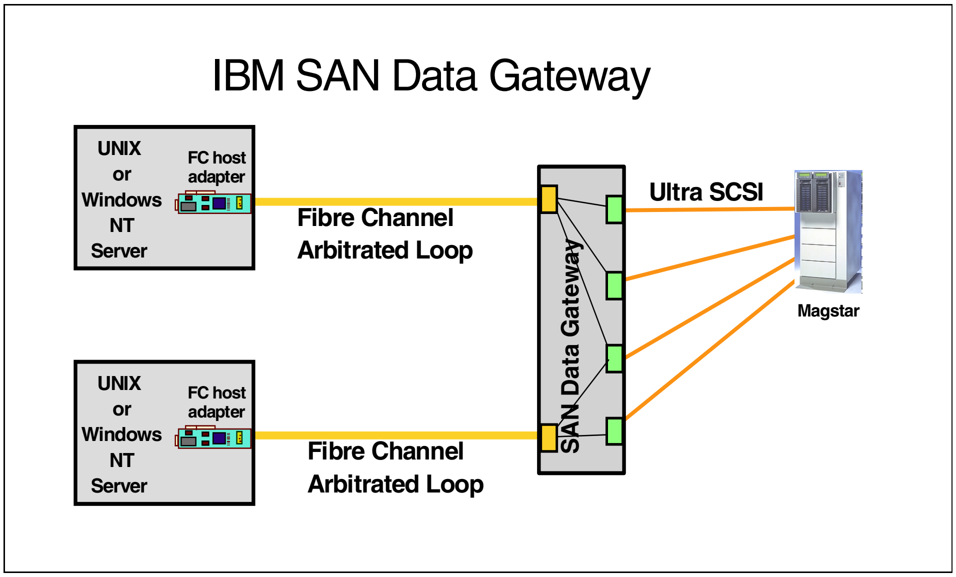


Figure 36. IBM SAN Data Gateway

4.3.11 Gateways

In the SAN environment a gateway component can, for example, connect two different remote SANs with each other over a Wide Area Network, as shown in Figure 37 on page 61. A typical application requiring gateways would be remote copy solutions.

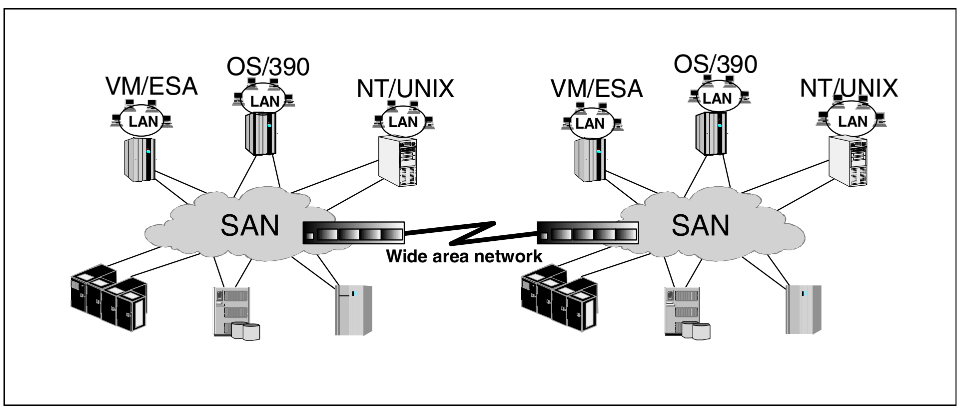


Figure 37. SAN Gateway Connecting Remote SANs

Chapter 5. SAN Management

SAN Management has a very critical role in any SAN solution and can also be a deciding factor in the selection of a vendor's SAN solution implementation. SAN Management can be seen in the IBM SAN Initiative pyramid shown in Figure 38 on page 63. This figure refers to management of both the SAN hardware base as well as the connectivity, which is key to being able to fully exploit the SAN application and solution.

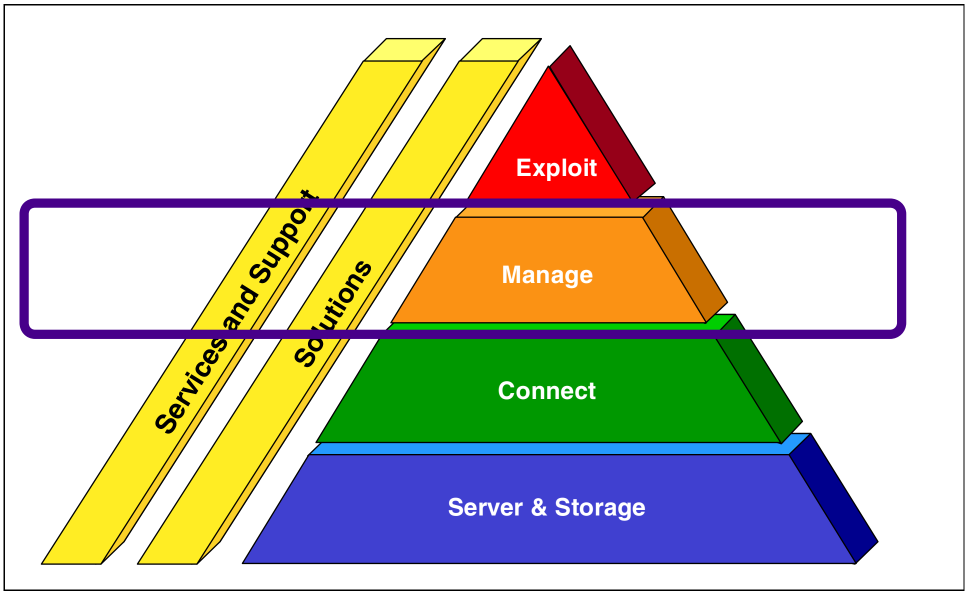


Figure 38. IBM SAN Initiative

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 tradeoff 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 so on. SAN management reduces total cost of management, for example, by managing the two lower layers, connectivity, servers, and storage from a variety of vendors, from one single point using the same tool.

For example, a large, 10-to-50 node arbitrated loop may have Host Bus Adaptors (HBAs) 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 loop topology may be transporting IP and SCSI-3 (FCP) protocols. All these factors add complexity to the SAN, and 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 Evolving Standards and Definitions

SAN Management architecture can be divided into three distinct levels — SAN storage level, SAN network level, and enterprise systems level. Figure 39 on page 64 illustrates the three management levels in a SAN solution.

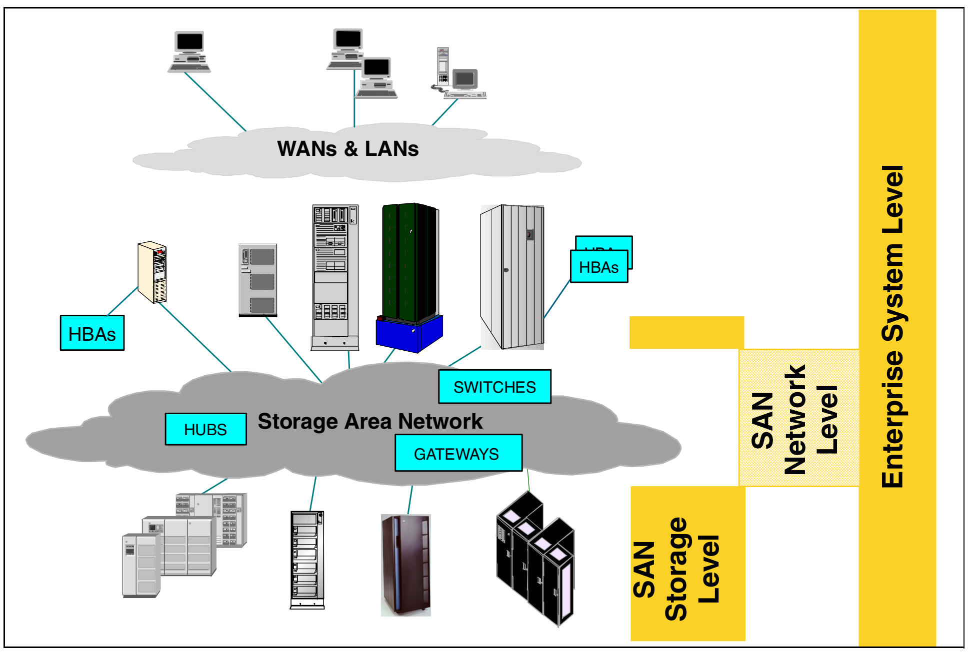


Figure 39. SAN Management Levels

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 configuration of a storage resource must be integrated with configuration of the server's logical view of them, the SAN storage level management spans both storage resources and servers in Figure 38 on page 63.

Today, SAN storage predominantly uses the Small Computer Systems Interface (SCSI) protocol. The ANSI SCSI-3 serial protocol is used over the Fibre Channel 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.

5.2.1.1 SCSI Enclosure Services (SES)

In SCSI legacy systems, 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 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 an impact on Fibre Channel - Arbitrated Loop (FC-AL) management, since it is a potential source of overhead with loop data traffic. For example, if an FC-AL hub or switch supports SES queries, it becomes a participating node in the loop, and the management traffic generated on the loop would then effectively reduce the server to storage data transfer throughput. Most SAN vendors deploy SAN management strategies using Simple Network Management Protocol (SNMP) based and non-SNMP based (SES) protocols.

5.2.2 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 Host Bus Adapters (HBAs).

This SAN network infrastructure becomes closely tied to the inter-networking infrastructure, as seen in Local Area Network (LAN) and Wide Area Network solutions (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 the SAN networking components, which is then used and exploited by SNMP based network management applications to report on the various disciplines of SAN management.

5.2.2.1 Simple Network Management Protocol (SNMP)

SNMP, an IP-based protocol, has a set of commands for getting 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 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.

The Storage Networking Industry Association (SNIA) very recently announced formation of a task force to define standard MIBs for the SAN industry using the Fibre Channel interface that will enable SAN management applications to query the SAN fabric components (like switches) to obtain general status information using Fabric Element (FE) MIBs and Fibre Channel InterConnect (FC-IC) MIBs on supporting devices.

Figure 40 on page 67 illustrates the various elements of an SNMP-based management architecture.

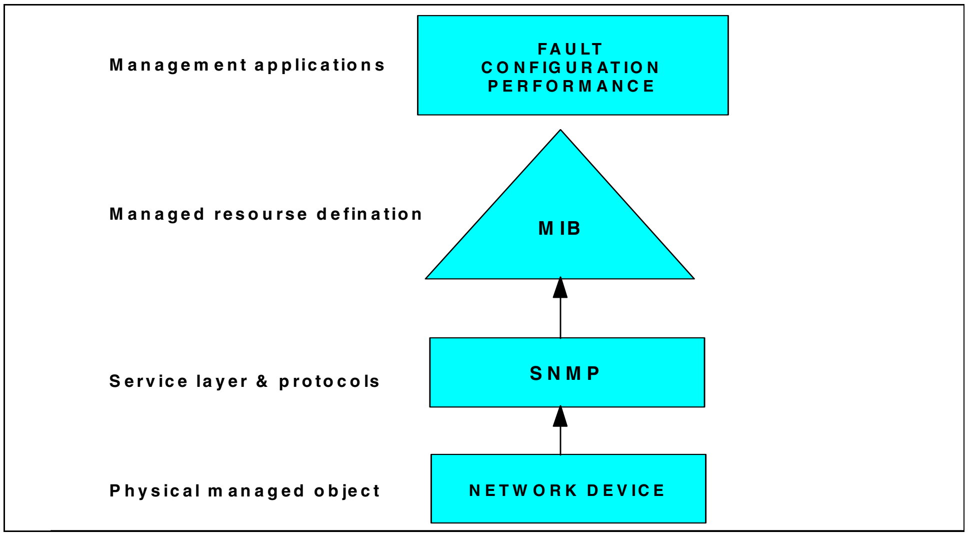


Figure 40. Elements of SNMP-Based Management Architecture

5.2.3 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 Interface 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 and WBEM provides Web access to the managed elements and managed data.

5.2.3.1 Web Based Enterprise Management (WBEM)

WBEM enables the industry to deliver a well-integrated set of standards-based management tools that leverage the emerging Web technologies and unify management interfaces.

5.2.3.2 Common Interface Model (CIM)

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.

5.3 SAN Management Strategy and Disciplines

There are many aspects to total SAN Management. Some of the simpler functions include:

- Device installation, configuration and monitoring
- Inventory of resources on the SAN
- Reporting utilities
- Assistance in SAN planning

Some of the more advanced, highly valued functions include:

- Automated component and fabric topology discovery
- Management of the fabric configuration, including zoning configurations
- Name services
- Security management
- Health and performance monitoring
- Workload management/balancing of SAN activities based on application priorities set by business objectives (for example, backup copy creation has a lower priority than business transactions)

5.3.1 SAN Management Strategy

A comprehensive SAN Management strategy covers the following general topics:

- Device management
- Problem detection
- Problem isolation
- Problem diagnosis
- Problem recovery
- Predictive analysis

Unlike many other management products available today, IBM offers a comprehensive SAN management solution family of products, called *StorWatch*. The StorWatch products are based on the IBM Enterprise Storage Resource Management (ESRM) disciplines. Figure 41 on page 69 illustrates IBM's ESRM strategy.

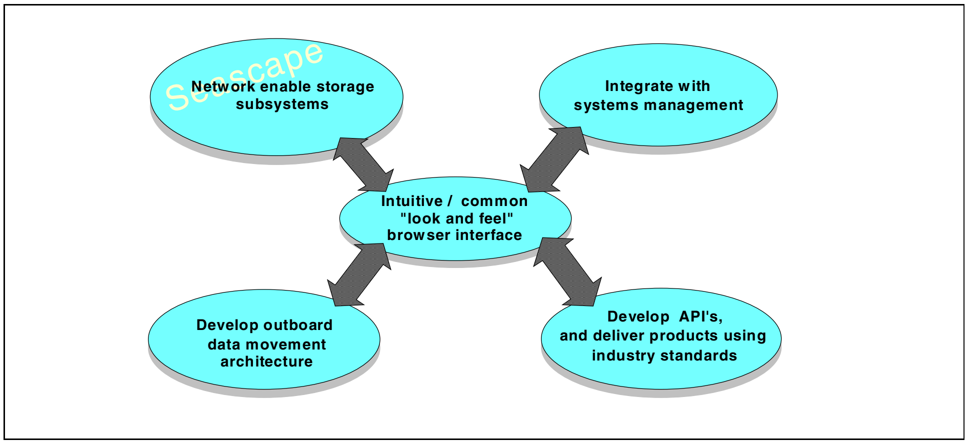


Figure 41. IBM ESRM Strategy

ESRM strategy defines the various disciplines of storage management, such as asset management, capacity management, configuration management, performance management, and availability management — described earlier in 1.3.5, "Managing the SAN" on page 20 — that need to be implemented across the heterogeneous storage resources connected to the SAN.

For more information on enterprise storage management, see *Principles of Enterprise Storage Management*, SG24-5250.

5.4 General Issues in Fibre Channel Arbitrated Loop Management

The Fibre Channel Arbitrated Loop (FC-AL) in a storage network is analogous to shared media LAN technologies like FDDI and Token Ring, but with some unique features and definitions that pose a challenge to implementing a comprehensive Fibre Channel management strategy. The following section explores some of these features.

5.4.1 In-Band and Out-of-Band Management

In traditional LAN/WAN technology, the terms *In-Band* and *Out-of-Band* refer to the flow of the management traffic, either through the primary LAN/WAN interface (in-band) or through an alternate data path like the serial RS-232 or SLIP connection (out-of-band). Typically, in the LAN/WAN environments, in-band management access methods are preferred and used to perform normal device configurations, status checks, and monitoring of network information from management stations anywhere on the network. In this case, the management traffic (SNMP commands using IP) intermixes with data traffic along the primary interface (Ethernet, Token Ring, Frame Relay, and so on). LAN/WAN environments use the out-of-band method only for worst-case situations, for example, when the primary interface is down and SNMP management is not possible through the network.

In FC-AL, in-band management in the arbitrated loop incurs traffic overhead, since the initiating manager has to arbitrate, open, and close repeatedly to solicit information from the loop's nodes. In a network, repeated management traffic can drastically reduce the effective throughput of the link. So, although there are some valid functions that could be performed by in-band management in an arbitrated loop (for example, inquiry of SCSI enclosure data, in-band management is not the preferred access method and is usually deployed sparingly by the vendors.

In FC-AL, out-of-band management occurs through an Ethernet interface or a serial console interface on the switch or hub. Out-of-band management using Ethernet has three main advantages:

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

This is similar to the in-band advantages in LAN/WAN environments.

5.5 Troubleshooting

In any network environment setup, precise problem identification is very critical in order to reduce the effective downtime.

5.5.1 Arbitrated Loops

In an FC-AL environment, transactions on the loop may not always be visible to all the nodes. An arbitrated loop sends a data frame from the source node to the destination node, but the destination node removes the frame from the loop, and the other nodes are unaware of the transaction. Thus, troubleshooting an arbitrated loop may require a data trace at each of the suspected ports. This is not only expensive to implement, but it is also difficult to implement and still maintain the integrity of the loop.

In many FC-AL implementations, only two nodes are attached, for example, an application server and a storage server. This makes the configuration look like point-to-point, and therefore much easier to manage.

5.5.2 Switched Environments

In a switched environment, the frame does not pass any other node as in a loop environment. The source node and destination node are the only two nodes involved. This makes it easy to identify and isolate failing components, and thus reduce the downtime.

5.6 IBM StorWatch

StorWatch is both a family of software products that implement IBM's Enterprise Storage Resource Management (ESRM) solution as well as a set of configuration and setup tools for managing storage resources. The goal of the StorWatch family is to make IBM and non-IBM storage subsystems easy to install, configure, monitor, manage and maintain. To this end, StorWatch will provide advanced storage management functions for a variety of storage resources, including:

- Logical and physical disk devices and disk subsystems
- Tape drives and tape subsystems
- Removable media libraries
- Fibre Channel, SSA, and SAN fabric components (switches, hubs, and gateways)
- Storage management software systems (like ADSM and DFSMS)

Storage management functions and features implemented by StorWatch products will include:

- Advanced Web-browser based user interfaces
- Asset management
- Capacity management
- Configuration management
- Data/device/media migration and replication
- Device and event monitoring and alerts
- Performance management
- Service management
- Storage policy management
- Storage resource discovery and launch

StorWatch will deploy a number of products, each providing advanced storage management functions for particular storage resources. Typically, storage and Fibre Channel hardware products will provide their own Web or stand-alone interfaces for basic single unit management and configuration. These interfaces will be StorWatch branded and integrated. StorWatch will, in addition, provide value-add management functions (monitoring, capacity usage over time, reporting, performance, and so on) for managing multiple instances of resources. StorWatch will also provide a set of core services (database, security, reporting, agent) which will provide a technology base for most other StorWatch products.

The following sections describe the StorWatch products that are either available today or that will be available in the near future.

5.6.1 StorWatch Serial Storage Expert

StorWatch Serial Storage Expert simplifies the management of SSA storage networks by providing capabilities to help administrators plan SSA disk networks, centrally manage the configuration of existing SSA networks, and monitor SSA network devices for errors and events.

5.6.2 StorWatch Versatile Storage Specialist

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

5.6.3 StorWatch Enterprise Storage Server Specialist

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

5.6.4 StorWatch Enterprise Storage Server Expert

StorWatch Enterprise Storage Server Expert helps storage administrators:

- Manage the performance of the new Enterprise Storage Server (ESS).
- Keep track of the total capacity, assigned capacity, and free capacity of the disk storage on all the ESSs anywhere in the enterprise.
- Identify which SCSI-attached servers are sharing which volumes within the ESS.

The StorWatch ESS Expert provides a common look-and-feel with the configuration interface for the Enterprise Storage Server Specialist, and with the StorWatch Reporter.

5.6.5 StorWatch DFSMShsm Monitor

StorWatch DFSMShsm Monitor monitors DFSMShsm and automates the DFSMShsm environment.

5.6.6 StorWatch Reporter

StorWatch Reporter provides asset, capacity and utilization management for the major server platforms (OS/390, Windows NT, AIX, HP, Sun, OS/2, Netware) as well as discovery and launch of Web interfaces for ADSM, Enterprise Storage Server, Versatile Storage Server, and Network Storage Manager)

5.6.7 StorWatch Fiber Channel RAID Specialist

StorWatch Fiber Channel RAID Specialist provides configuration, performance monitoring, error reporting and management operations for IBM Fibre Channel RAID Storage Servers.

5.6.8 StorWatch SAN Data Gateway Specialist

StorWatch SAN Data Gateway Specialist provides configuration, gateway discovery, asset management, and device monitoring for IBM SAN Data Gateways for SCSI-attached disk and tape storage.

5.6.9 StorWatch SAN Data Gateway S20 Specialist

StorWatch SAN Data Gateway S20 Specialist provides configuration, gateway discovery, asset management, device mirroring, disk copy management, and device monitoring for IBM SAN Data Gateways for serial disk.

5.6.10 StorWatch SAN Fibre Channel Switch Specialist

StorWatch SAN Fibre Channel Switch Specialist provides a comprehensive set of management tools that support a Web browser interface for flexible, easy-to-use integration into existing enterprise storage management structures. The specialist provides security and data integrity by limiting (zoning) host system attachment to specific storage systems and devices.

5.6.11 StorWatch Data Path Optimizer

StorWatch Data Path Optimizer provides the ability to dynamically switch paths, if multiple paths are assigned, in an NT and UNIX environment providing high availability and also providing a load balancing algorithm that can potentially enhance the performance and throughput.

For more information on StorWatch products, see the Web site:

`www.storage.ibm.com/software/storwatch/swhome.htm`

Chapter 6. SAN Exploitation and Solutions

The added value of a SAN lies in the exploitation of the technology to provide tangible benefits to the business. As illustrated in Figure 42 on page 75, exploitation is at the top of the SAN pyramid, and is enabled by all the underlying layers. Benefits range from increased availability and flexibility to additional functionality that can reduce application downtime. This chapter will also present and discuss various solutions that are enabled by a SAN.

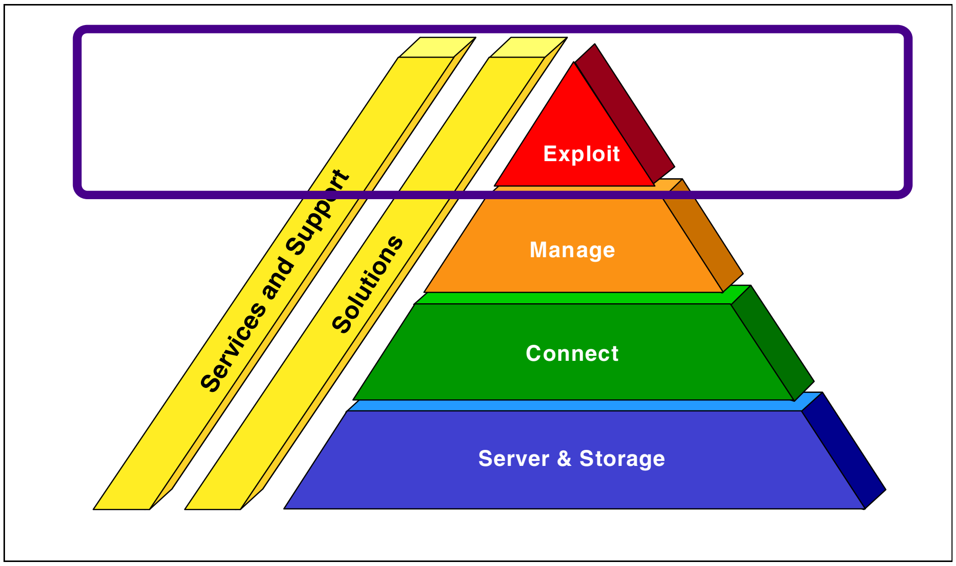


Figure 42. Outline of IBM SAN Initiative: Exploitation

We will start with a discussion of the SAN building blocks.

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 describe 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 43 on page 76, the SAN Toolbox. The toolbox contains all the necessary building blocks to construct a SAN. We will use the icons shown in the toolbox in most of the figures in this chapter.

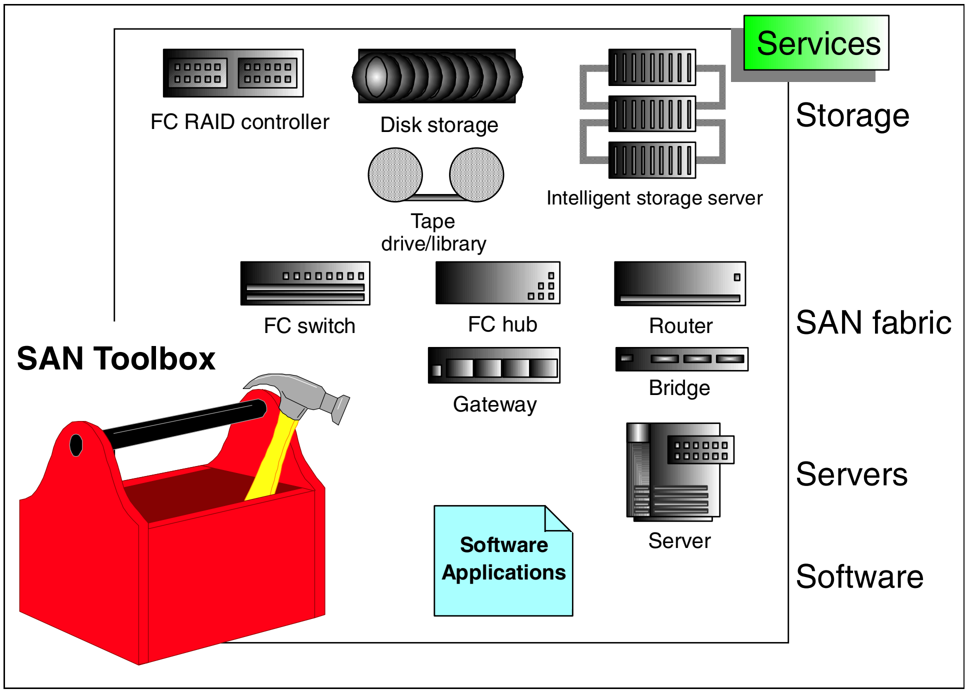


Figure 43. SAN Toolbox

There are five basic SAN building blocks.

Services — This building block, which includes support and education, is key in implementing a SAN using the other building blocks.

Storage — 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 (either through bridges or native attachment) have the following benefits:

- Running SCSI over Fibre Channel for improved performance (sometimes called SCSI Accelerator)
- Extended connection distances (sometimes called storage at a distance)
- Enhanced addressability

Many of the first 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 permits the common handling of the interconnection cables to storage and 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.

One of the most prevalent storage protocols in today’s distributed environment is SCSI. Although SCSI is commonly used, it 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 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 and at the same time providing investment protection by using bridges.

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 goes away. Hardware interconnections become common across all servers and devices.

The advantages are similar to those brought to the S/390 world with the introduction of ESCON in 1991. For example, common trunk cables can be used for all boxes. 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. Adding storage capacity to Windows NT systems that use FAT or NTFS file systems requires a re-boot of the machine to make the new storage available. Operating systems such as AIX and OS/390 however allow for non-disruptive addition and removal of storage devices. They do not require a re-boot or IPL.

If new storage devices are attached to a section of a SAN with loop topology, the loop initialization protocol (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. 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 *directly attached* to all the servers, and efficiencies of scalability result from consolidation of storage capacity. The semantics of data access from the server, in this case, are block level requests to a specific LUN (Logical Unit) or S/390 device address.

When storage is added, port zoning will restrict access to added capacity. As many devices (or LUNs) can be attached to a single port, access can be more 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 44 on page 80 shows an example of disk storage pooling across two servers. 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.

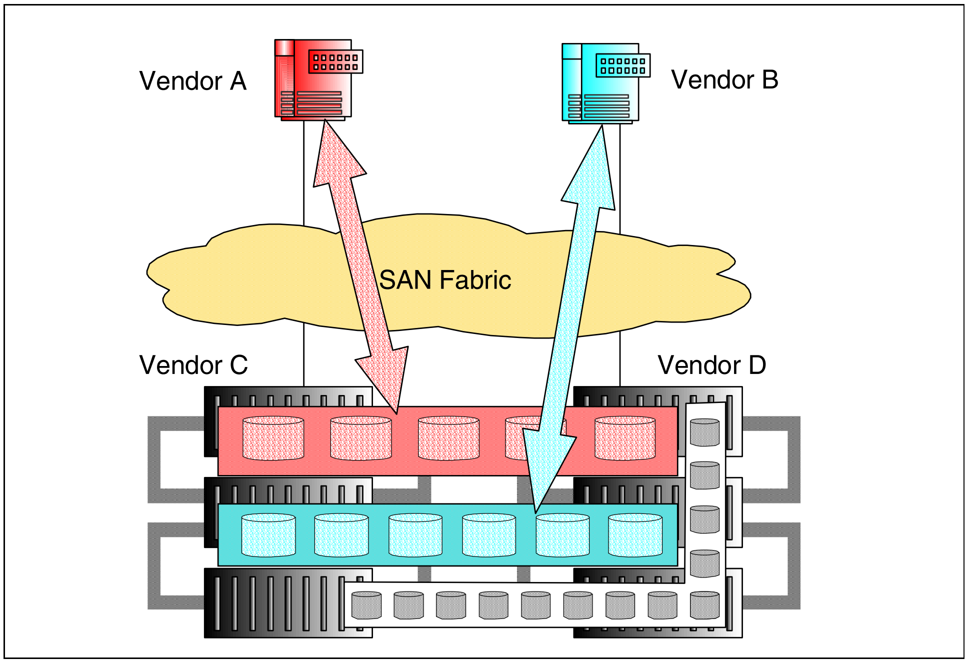


Figure 44. Disk Pooling

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. Current 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 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. A SAN even opens up the possibility of sharing of tape resources among heterogeneous systems that include OS/390.

At any particular instant in time, a tape drive can be owned by one system, as shown in Figure 45 on page 81.

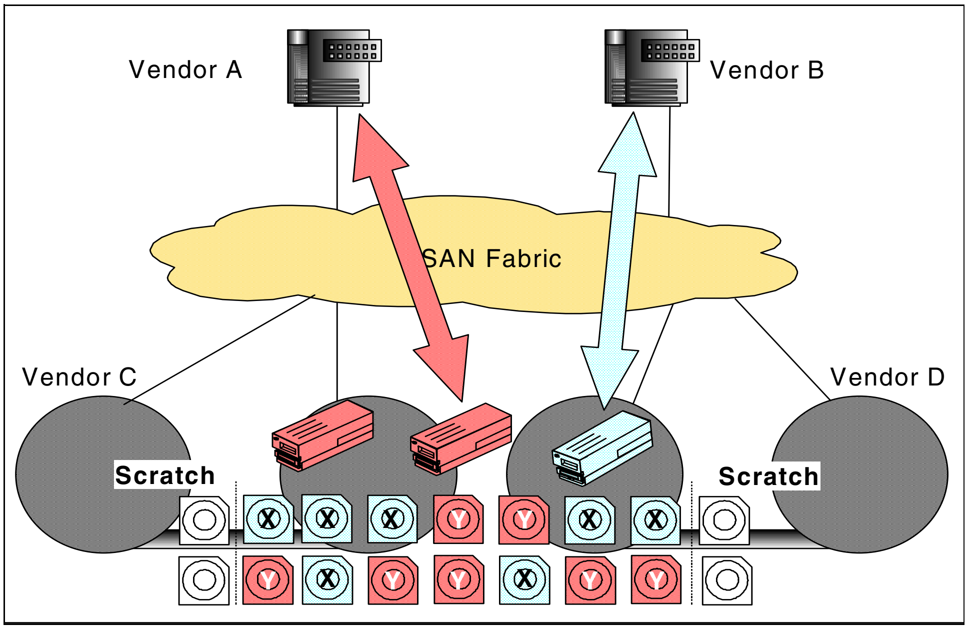


Figure 45. Tape Pooling

In this example, the server to the left currently has two tape drives assigned, and the server to the right 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 assignable to servers (in scratch groups) if not yet used or 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 is possible today with SCSI using multiple pathing, scalability is an issue because of the distance constraints of SCSI.

SCSI allows for distances of 25 meters, and the size of SCSI connectors limits the number of connections to server or subsystem.

A SAN allows for efficient load balancing in distributed processing applications 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 facility 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 46 on page 83 shows an example of clustering. Servers S1 and S2 share intelligent storage servers (ISS) #1 and 2. If S1 fails, S2 can access the data on ISS #1. The example also shows that ISS #1 is mirrored to a remote site on ISS #2. Moving the standby server, S2, to the remote, WAN connected, site would allow for operations to continue even in case of a disaster.

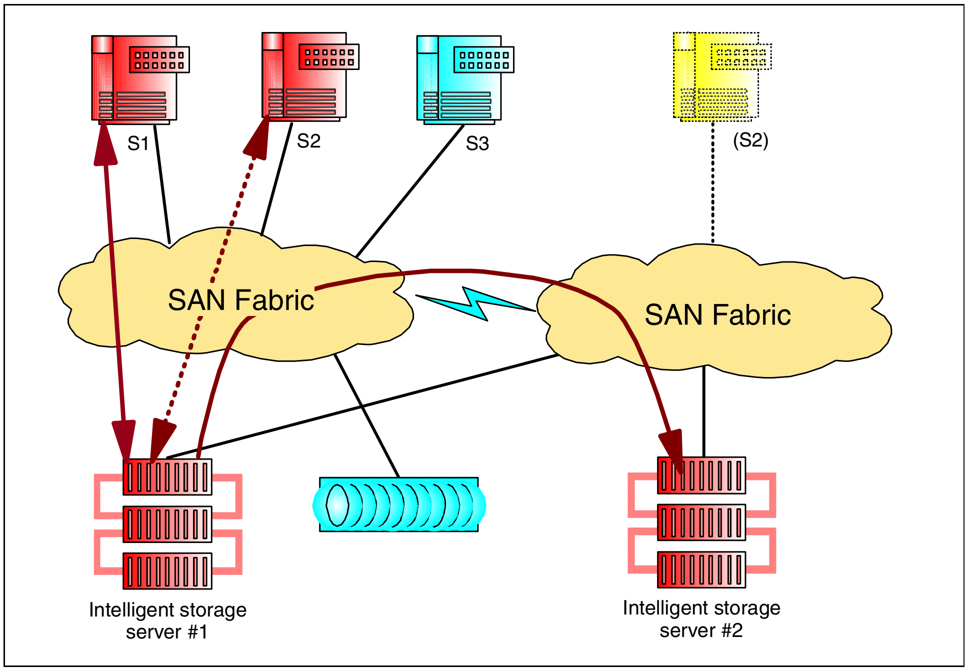


Figure 46. Server Clustering

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 1.3.4.1, “Shared Repository and Data Sharing” on page 17. 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. Operating systems such as OS/390 have supported data sharing since 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, as is often the case today. SAN connectivity has the potential to provide sharing services to heterogeneous hosts, including UNIX, Windows NT, and OS/390.

6.4.1 From Storage Partitioning to Data Sharing

Storage partitioning does not represent a true data sharing solution. It is essentially a storage-subsystem-sharing or storage-volume-partitioning type of solution. The storage subsystems are connected to multiple servers, and storage capacity is partitioned among the various subsystems.

Storage partioning 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:

- Physical volume partitioning
- Logical volume partitioning
- File pooling
- Data copy sharing
- True data sharing

6.4.1.1 Physical Volume Partitioning

In the simplest form, disk volumes are assigned from a pool to servers on a hardware addressable basis. Once assigned to a server, the disk is used exclusively by this system and cannot be accessed by other systems. The capacity of the assigned disk volumes is the capacity of the physical disk volumes. Figure 47 on page 84 shows an example of four heterogeneous systems sharing one storage subsystem.

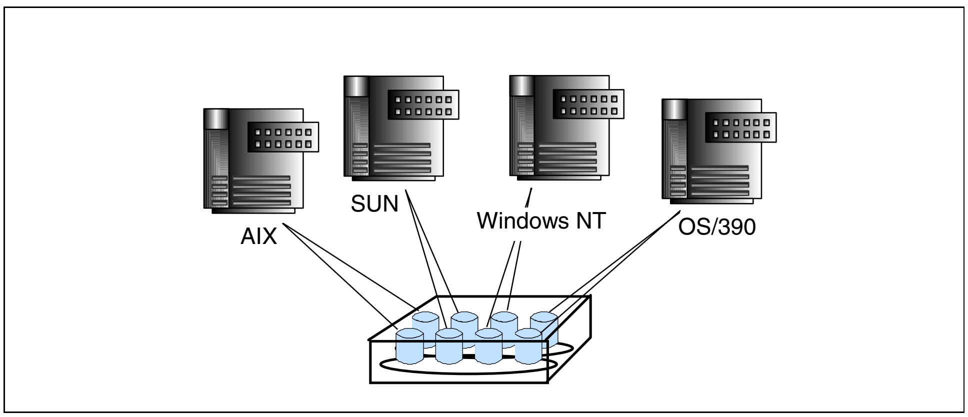


Figure 47. Physical Volume Partitioning

6.4.1.2 Logical Volume Partitioning

This is another storage partitioning solution. 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, 5 GB and 15 GB) may be created from a single 30 GB hardware addressable disk, with each being assigned to a different server, leaving 10 GB of unassigned capacity. A single 200 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 security to the data.

Figure 48 on page 85 shows multiple servers accessing logical volumes created using the different alternatives mentioned above. (The logical volume *Another volume* is not assigned to any server.)

.

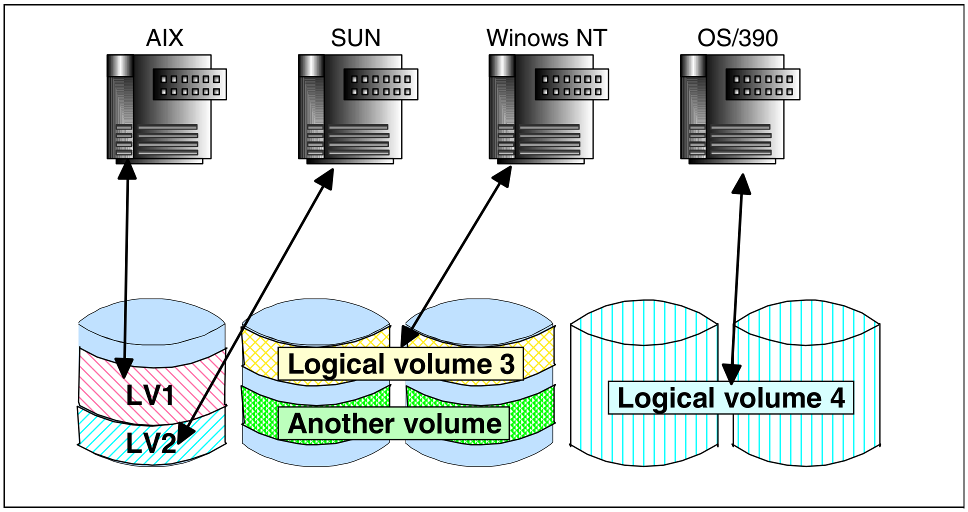


Figure 48. Logical Volume Partitioning

6.4.1.3 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 and manually or using the operating system functions (as in OS/390, 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 49 on page 86 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.

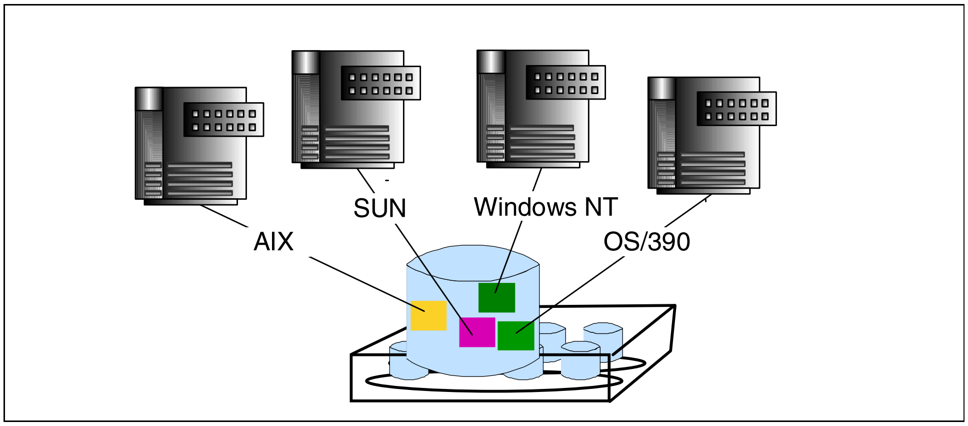


Figure 49. File Pooling

6.4.2 Data Copy-Sharing

In data copy-sharing, a copy is made of the data and the copy is assigned to the same server or a different server on either a like or unlike platform. There are two forms of data copy-sharing:

- Complete file copy-sharing — A complete file or database is copied on a regular basis. This form is most common in archive and backup/recovery applications
- Incremental file copy-sharing — An initial copy of the complete or database is done, followed by incremental updates, where changes are extracted from the source and copied into the target. This form is most common in data warehouse types of applications.

The copy can be used and updated independently from the original; it may diverge from the original.

6.4.3 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 of 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 disk in an orderly way. Serialization mechanisms may be defined to serialize from a group of physical or logical disk volumes to an individual physical block of data within a file or database.

6.4.3.1 Homogeneous Data Sharing

Figure 50 on page 88 shows multiple hosts accessing (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.

Access to the data is similar to what we have today with NFS. The main difference from NFS is that the access occurs through channel commands and not through network commands and protocols (IP).

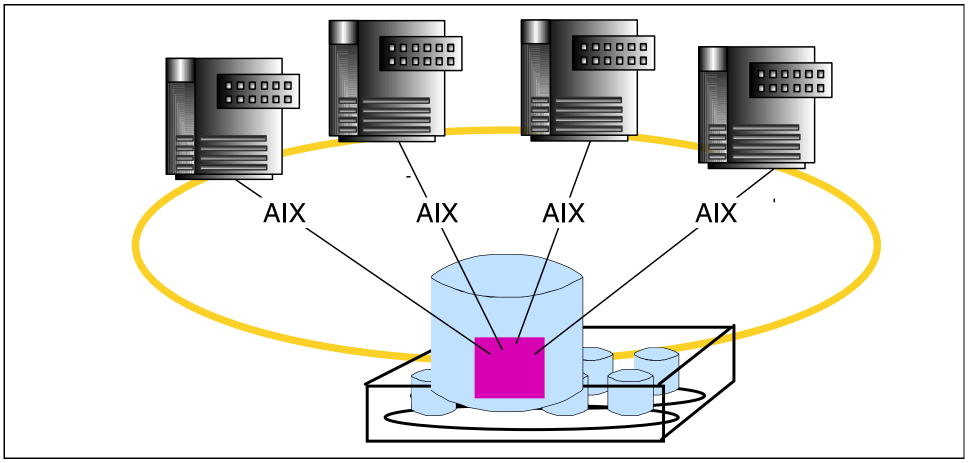


Figure 50. Homogeneous Data Sharing

6.4.3.2 Heterogeneous Data Sharing

In heterogeneous data sharing, as illustrated in Figure 51 on page 88, 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 40. Thus, we have the requirement for a SAN distributed file system.

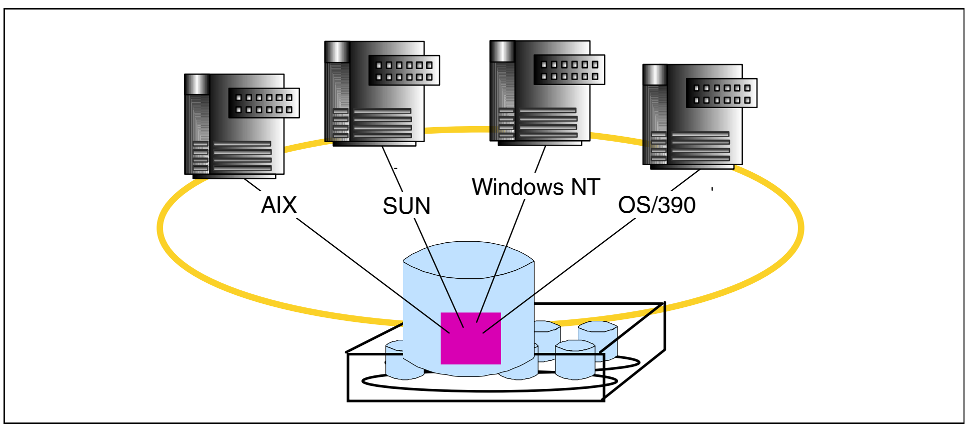


Figure 51. Heterogeneous Data Sharing

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 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-less), 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 Media Migration

The first data movement solution is media migration. It is used when you need to move data from one device to another without using a server. The process is logically carried out according to the following steps:

1. Define a target volume.
2. Copy all the data from the source to the target volume. Updates to the source volume must be applied concurrently to the target volume.
3. Redirect all requests from the source volume to the target volume, which now becomes the new source volume.
4. Release the old source volume.

This kind of solution allows for transparent media migration. The source and the target volume may be in the same, or different storage subsystem. The application on the server is unaware that there has been a device change. This type of solution is available today, for example, by using the IBM SAN Data Gateway for Serial Disk.

6.5.2 Data Migration and Recall

Data migration is, in certain ways, similar to media migration. The primary difference is that the unit of storage is no longer a volume, be it physical or logical, but rather a file or other unit of data. When a file is migrated, the data is moved down a hierarchy to a lower cost media (usually tape) or recalled, moved back up the hierarchy. Today, this type of movement is done through servers by archive, or hierarchical storage management software. This type of implementation calls for a SAN outboard data movement implementation.

6.5.3 Data Copy Services

6.5.3.1 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 IBM's ADSM, during the reclamation or recycle process. With SAN outboard data movement, traditional copy will be performed server-free, therefore making it easier to plan, and faster to execute.

6.5.3.2 T-0 Copy

Another outboard copy service enabled by Fibre Channel technology is T-0 (time=zero) copy. This is the process of taking a snapshot, or freezing the data (databases, files, or volumes) at a certain time and then allowing applications to update the original data while the frozen copy is duplicated. With the flexibility and extendability that Fiber Channel brings, these snapshot copies can be made to local or remote devices. The requirement for this type of function is driven by the need for 24x7 availability of key database systems.

6.5.3.3 Remote Copy

Remote copy is a business requirement in order to protect data from disasters, or to migrate data from one location to avoid application downtime for planned stops like 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 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 support tape and 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 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-Less Data Movement

The network backup and recovery paradigm implies that data flows from the backup and recovery client (usually a file or database server) to the backup and recovery server, or between backup and recovery servers, over a network connection. The same is true for archive or hierarchical storage management applications. Often the network connection is the bottleneck for data throughput. This is due to network connection bandwidth limitations. The SAN can be used instead of the LAN as a transport network.

6.6.1.1 Tape Drive and Tape Library Sharing

A basic requirement for LAN-less 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 80, between backup and recovery servers, and between a backup and recovery server and its backup and recovery client (usually a file or database server). Network-attached end-user backup and recovery clients will still use the network for data transportation.

In the tape drive and tape library sharing approach, the backup and recovery server or client that requests a backup copy to be copied to or from tape will read or write the data directly to the tape device using SCSI commands. This approach bypasses the network transport's latency and network protocol path length, therefore it can offer improved backup and recovery speeds in cases where the network is the constraint. The data is read from the source device and written directly to the destination device.

Figure 52 on page 92 shows an example of tape drive or tape library sharing: (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.

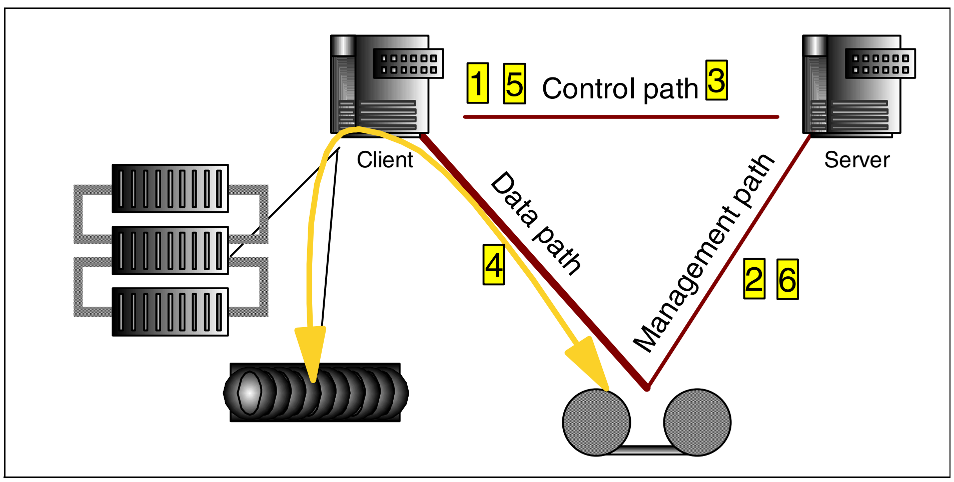


Figure 52. LAN-Less Backup and Recovery

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 by utilizing the SCSI-3 third party copy function.

Figure 53 on page 93 illustrates this approach. Management of the tape drives and cartridges is handled as in the proceeding 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.

In the example, shown in Figure 53 on page 93, 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 command on request from the client, using disk pooling.

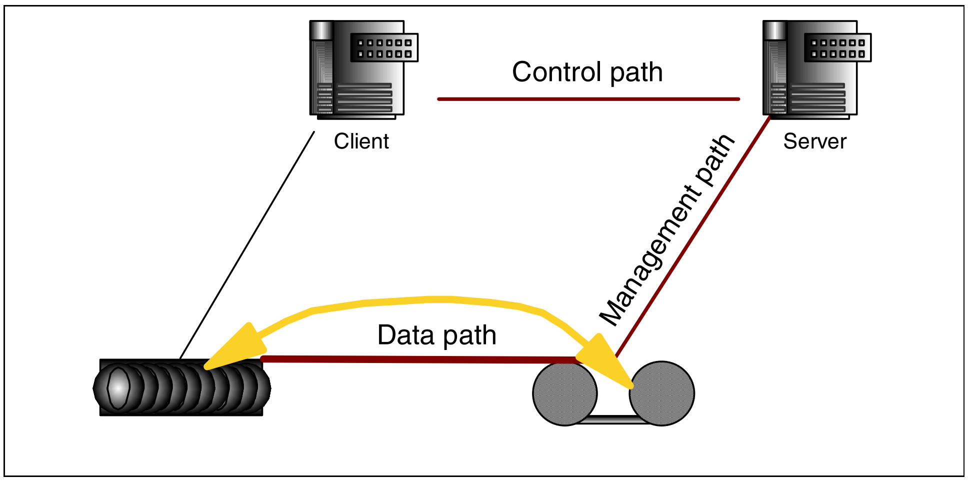


Figure 53. Server-Free Data Movement (Backup/Recovery)

The third party copy SCSI-3 command defines block-level operations, as is the case for all SCSI commands. The SCSI protocol is not aware of the file system or database structures. Using third part copy for file level data movement requires file systems to provide mapping functions between file system files and device block addresses. This mapping is a first step towards sophisticated database backup and recovery, log archiving, etc.

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 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 any server processor cycles for data movement, as shown in Figure 54 on page 94.

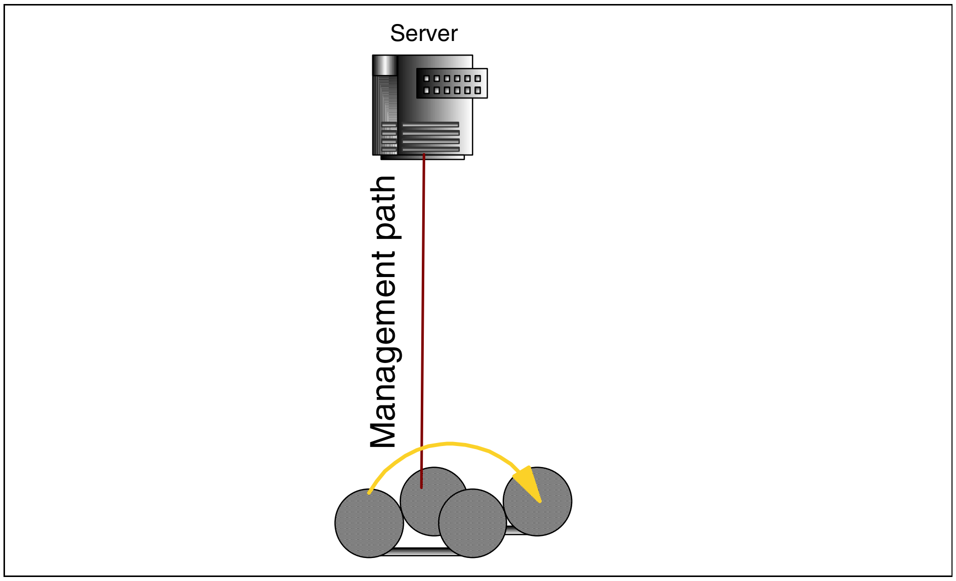


Figure 54. Server-Free Data Movement (Reclamation)

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. It is possible, using a SAN infrastructure, to perform extended distance backups for disaster recovery within a campus or city, as shown in Figure 55 on page 95.

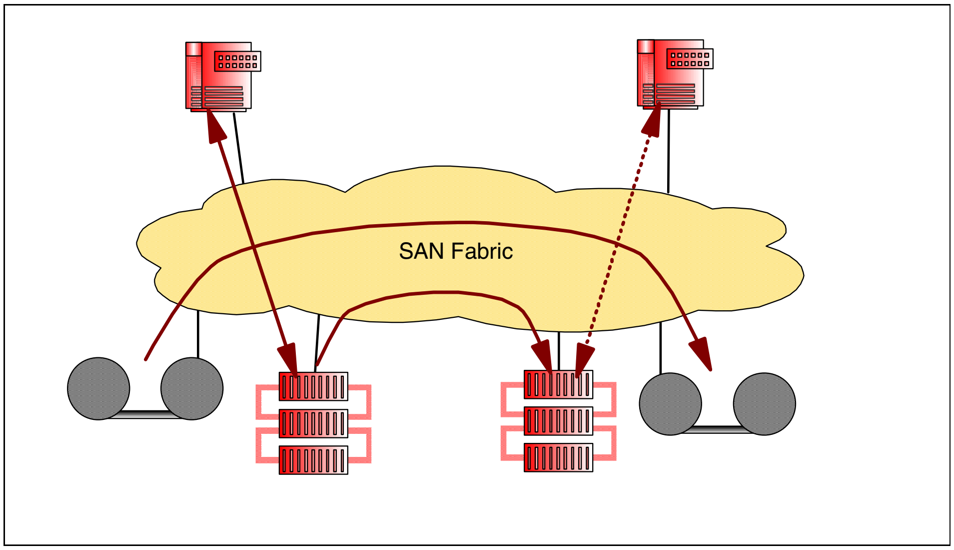


Figure 55. Disaster Backup and Recovery

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 81 and shown in Figure 46 on page 83.

Depending on business requirements, disaster protection implementations will 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 S/390, available today from IBM Global Services — will be required to monitor and manage these environments.

Chapter 7. SAN Standardization Organizations

The success of any new technology is greatly influenced by the standards. Although defacto Standards Bodies/Organizations such as 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 Storage Networking Industry Association (SNIA), and Fibre Channel Association (FCA) play a very significant role of 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 56 on page 97 contains a summary of the various associations and bodies linked with SAN standards and their role.

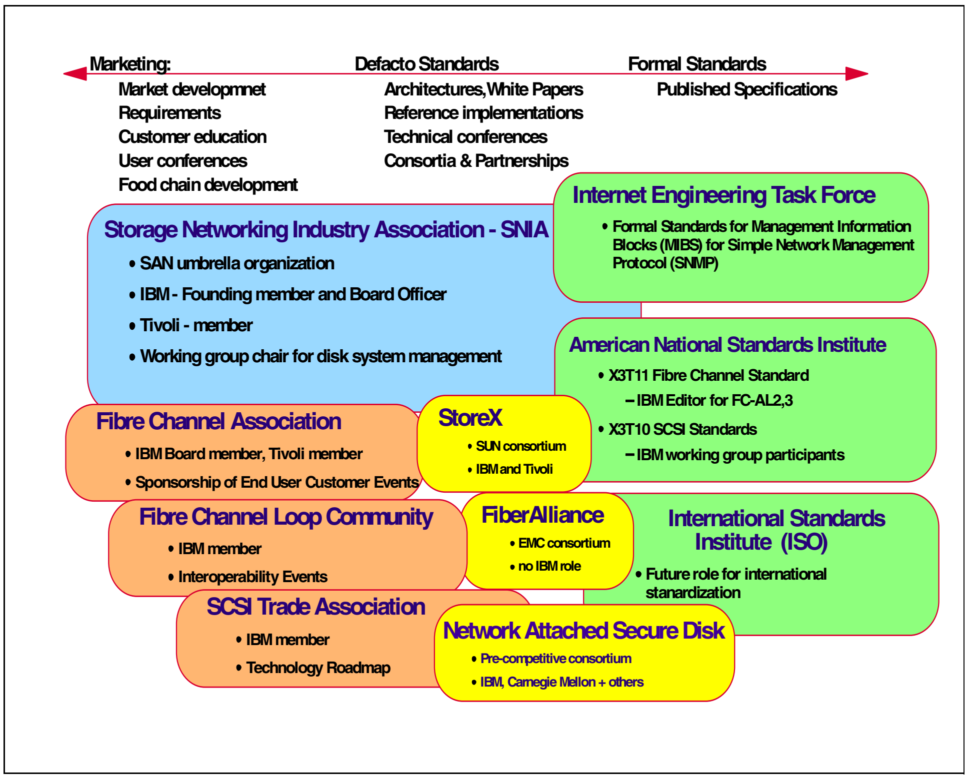


Figure 56. SAN Industry Associations and Standards Bodies

IBM is actively participating 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* that 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 56 on page 97.)
  These associations architecture development organizations that are formed early in the product life-cycle, have marketing focus, and do the market development, requirements, customer education, user conferences, and so on. This includes organizations are SNIA, FCA, Fibre Channel Loop Community (FCLC) and SCSI Trade Association (STA). Some of these organizations also help define the defacto industry standards and thus have multiple roles.
- ***Defacto Standards oriented*** — (In the middle in Figure 56 on page 97.)
  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 organization such as SNIA, FCA, FCLC, StoreX, FibreAlliance.
- ***Formal standards oriented*** — (To the far right in Figure 56 on page 97.)
  These are the formal standards organizations like IETF, ANSI, and ISO, 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, FCA, FCLC, STA, StoreX, FibreAlliance, IETF, ANSI, and ISO. A brief description of the roles of some of these organization is given below.

7.1.1 Storage Networking Industry Association — SNIA

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

`www.snia.org`

7.1.2 Fibre Channel Association — FCA

The Fibre Channel Association (FCA) is organized as a not-for-profit, mutual benefit corporation. The FCA 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 FCA. For additional information on the various activities of FCA, see its Web site at:

`www.fibrechannel.com`

7.1.3 Fibre Channel Community — FC/C

The Fibre Channel Community (FC/C) is a corporation consisting of more than 100 computer industry-related companies. Like other working groups, the FC/C complements activities of the various standards committees. Its goal is to provide marketing support, exhibits, and trade shows for system integrators, peripheral manufacturers, software developers, component manufacturers, communications companies, consultants, and computer service providers. The initial focus of the FC/C was on storage attachment migrating to network attached storage and solutions of FC-AL in networking environments. FC is the foundation upon which broader implementation of Fibre Channel interconnect schemes, such as SANs, will be achieved. IBM is a member of FC/C. For additional information on FC/C, see its Web sites at:

`www.fccommunity.org`

`www.fcloop.org`

7.1.4 The SCSI Trade Association — STA

The SCSI Trade Association (STA) was formed to promote the use and understanding of 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:

`www.scsita.org`

7.1.5 Future I/O

Future I/O is a new I/O organization or consortium. Standards are being developed to maximize data transfer between next-generation high-performance servers and peripheral subsystems. The server performance will no longer be I/O bound. The standard will enable resilient and scalable, distributed computing through inter-processor clustering. For additional information, see its Web site at:

`www.futureio.org`

7.1.6 StoreX

StoreX is a SUN Microsystems initiative to evolve a standard open platform for storage software development using Java and Jini. IBM is a member of this consortium. For additional information on StoreX, see its Web site at:

`java.sun.com/products/storex`

7.1.7 Internet Engineering Task Force - IETF

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 Simple Network Management Protocol (SNMP) for SAN management. For additional information on IETF, see its Web site a:

`ietf.org`

7.1.8 American National Standards Organization - ANSI

American National Standards Organization (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 is an editor of FC-AL 2,3 under X3T11, and IBM is also a working group participant in X3T10. For more information on ANSI, see its Web site at:

`www.ansi.org`.

Appendix A. Comparison of OS/390 and Open System Storage

The goal of this appendix is to explain the similarities and differences between S/390 and Open Systems. We will introduce the different terminologies used in the two environments. The UNIX operating system is used in most examples. Intel platforms are similar in their implementation to the UNIX environment and will not be treated separately.

We have made a number of simplifications. Not all UNIX systems are the same; some of the examples are drawn from AIX. Not all topics are covered. We ignore low level channel programming, security, performance management, allocation and space management, and other topics. The focus in the examples is disk storage and not tape or other media.

The term S/390 refers to both a processor and product set. OS/390 is an S/390 operating system. OS/390 and its predecessor MVS generally lead UNIX support of similar function by many years. Parallel Sysplex is S/390 processor clustering with failover and data sharing at the record level.

Figure 57 on page 101 illustrates the conceptual structure for accessing data on I/O devices.

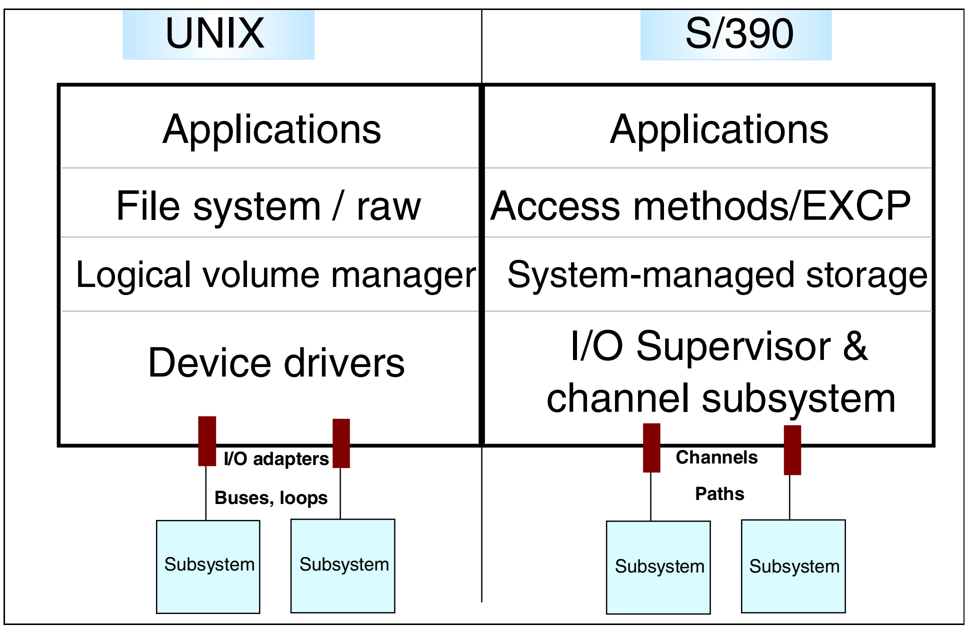


Figure 57. Conceptual I/O Structure

Applications see files mapped to volumes. These volumes are accessed by device drivers using low level I/O protocols. The Logical Volume Manager (LVM) to system-managed storage (SMS) mapping shown above is not perfect; many LVM functions are available only with S/390 storage subsystems.

OS/390 UNIX Systems Services offers basically the same file structure and facilities as described for the UNIX environment, but will not be covered.

7.2 Managing Files

We will now look at each component in more detail. Figure 58 on page 102 shows the similarities in file navigation: how files are located.

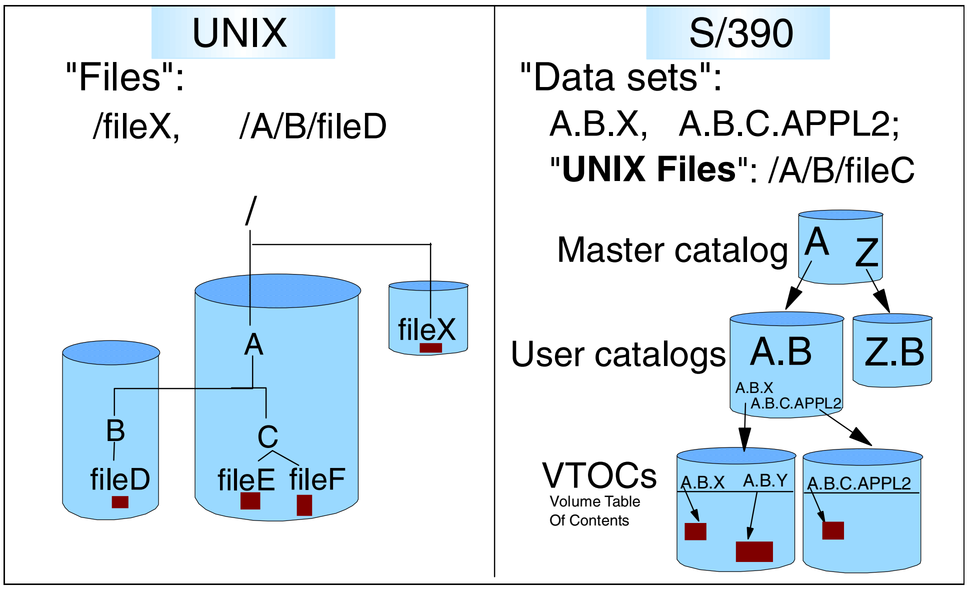


Figure 58. File Navigation

The OS/390 equivalent of a file is a data set. The term *data set* probably comes from complex files made from multiple physical files, for example, data and indexes. Each file has a unique file name. The S/390 name is, at most, 44 characters long, whereas UNIX names can often grow to 1024 characters or more. In UNIX the file name depends on the location of the file, whereas the name is location independent in OS/390.

In UNIX files are accessed by navigating down the path name. In OS/390 files are accessed through a catalog structure consisting of a master catalog and one or more user catalogs. Catalogs tell you which volume a data set is on. Volume Tables Of Content (VTOCs) on the volumes contain the cylinder/track location of the data set on the volume. OS/390 also supports UNIX files within a Hierarchical File System (HFS). This allows for UNIX applications to be ported to OS/390 with minimum changes.

Figure 59 on page 103 illustrate how files are accessed in OS/390 and UNIX.

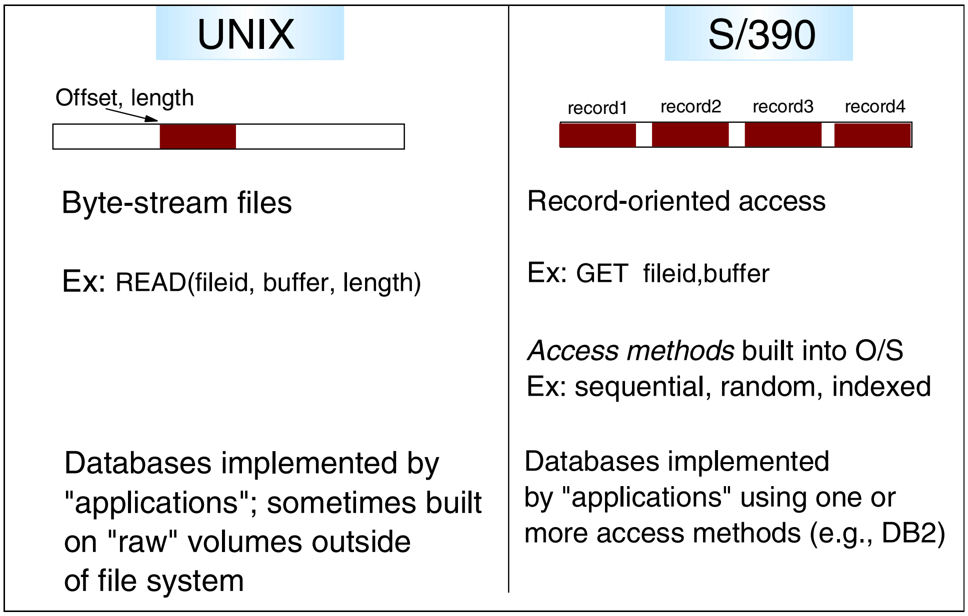


Figure 59. File Access

UNIX accesses files as a byte-stream of data, whereas OS/390 has record oriented access. With the byte-stream access model, the application must specify the offset to the required data and the amount to be retrieved. With record oriented access methods, the application requires a record at a time.

OS/390 supplies programs called access methods to simplify application access to data. Many access methods are supplied; each one is specialized for a particular use. OS/390 allows raw access to a device through the EXCP access method. This function is seldom used by applications any more.

7.3 Managing Volumes

Figure 60 on page 104 illustrates the difference in volume structures between S/390 and UNIX.

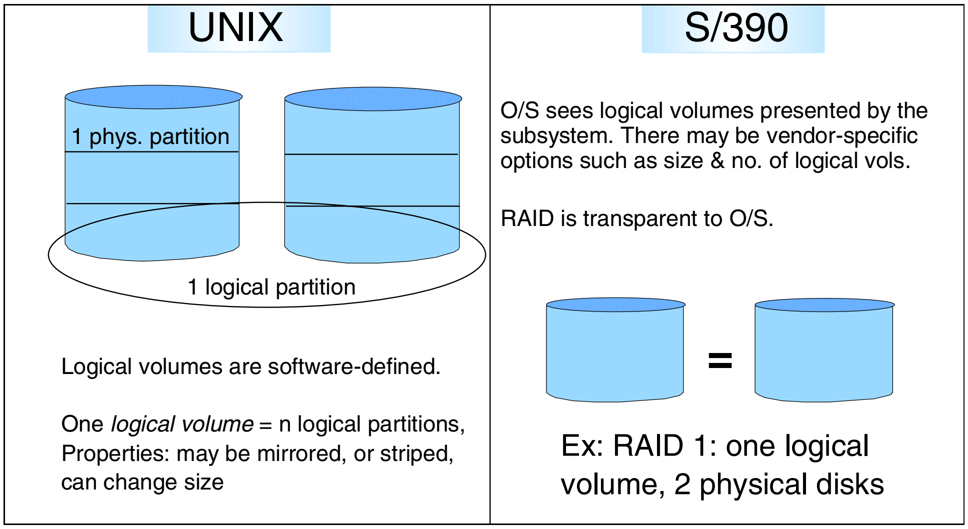


Figure 60. Volume Structure

In UNIX physical volumes represent the physical media you buy. Logical volumes, once defined, are what applications see. Not all UNIX systems support logical volumes. AIX, HP-UX, and Veritas define and use logical volumes. The UNIX term physical partition refers to a part of a physical volume. The UNIX term logical partition refers to a collection of physical partitions. A logical volume is made from a collection of logical partitions. A logical volume may be striped, mirrored, and can change in size.

In S/390 logical volumes are what is presented by the storage subsystem to the operating system and applications. Functions like RAID 1 are transparent to the operating system and are implemented at the storage subsystem level, inside the storage controller. Logical volumes are mapped to physical volumes by the storage subsystem.

Figure 61 on page 105 illustrates where basic operating system storage management functions are carried out.

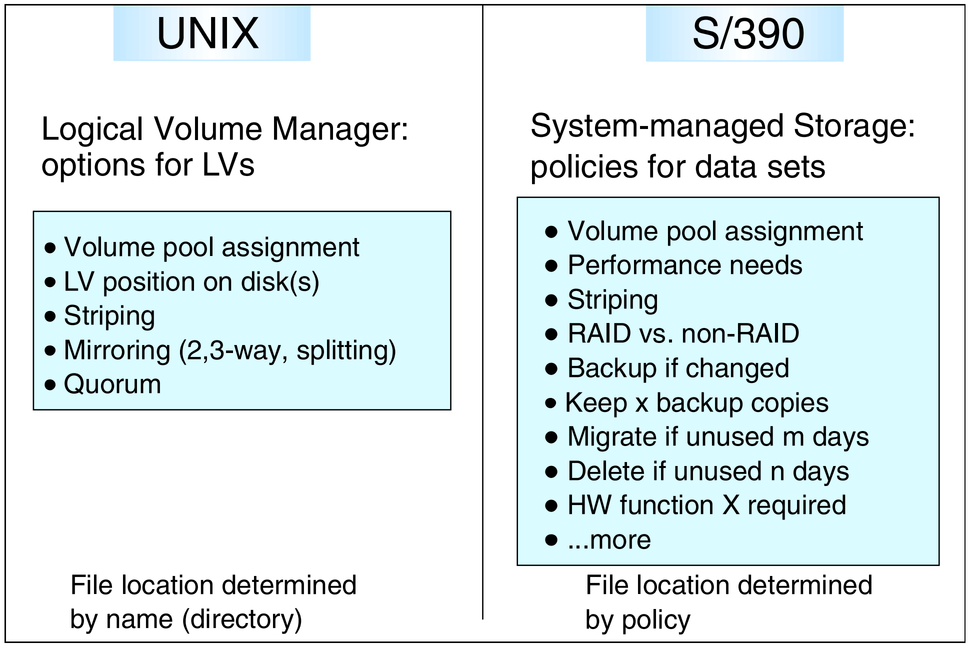


Figure 61. Basic O/S Storage Management

In UNIX, controls and parameters are set manually for the logical volumes at the logical volume manager level. The location of a file is determined by the name of the file, specifically the directory part of the name.

OS/390 offers system-managed storage (SMS) for data sets. SMS presents a logical view of storage that is separate from the physical device configuration and attributes. SMS uses policies based on logic to automatically assign storage management policies to data sets. Additionally, IBM or Independent Software Vendor (ISV) products are required to implement the backup, migration, and deletion policies.

Figure 62 on page 106 illustrates the differences in the low level software interface to devices.

UNIX	S/390
I/O Drivers - standard in kernel - others from vendors	I/O Supervisor - standard in OS/390 - vendors make hardware match OS "device types" (fixed track, cyl attributes)
SCSI command set	S/390 I/O commands (ECKD: Extended Count Key Data)
Queuing in device (SCSI Command Tag Queuing)	Queuing in host memory

Figure 62. Low Level SW Interface to Device

In S/390 a *device type* refers to a product specification. Examples are 3380-K or 3990-3. Each device type has a defined track size, maximum number of cylinders, and error status information. S/390 devices are handled by the I/O supervisor. Devices manufactured by other vendors emulate one of the defined S/390 device types. The most common device type is 3390-3.

In UNIX, devices are handled by I/O drivers sometimes referred to as device drivers. Some drivers come standard in the operating system kernel, but individual hardware vendors can and do write their own.

S/390 uses extended count key and data (ECKD) I/O commands. UNIX uses the SCSI command set. The SCSI command set addresses data by logical block number, whereas the ECKD commands use cylinder/track/block address.

S/390 storage subsystems can use FBA SCSI disk drives; in this case, the storage subsystem maps the ECKD commands to SCSI commands.

OS/390 issues one I/O per logical volume; other I/Os are queued in processor memory. With the Enterprise Storage Server in OS/390 environments, I/Os are queued in the storage servers, and up to 8 concurrent I/Os requests can be executed in parallel for a logical volume. In SCSI devices queueing is performed at the device.

7.4 Storage Device Attachment

Figure 63 on page 107 outlines how the processor to storage interaction occurs.

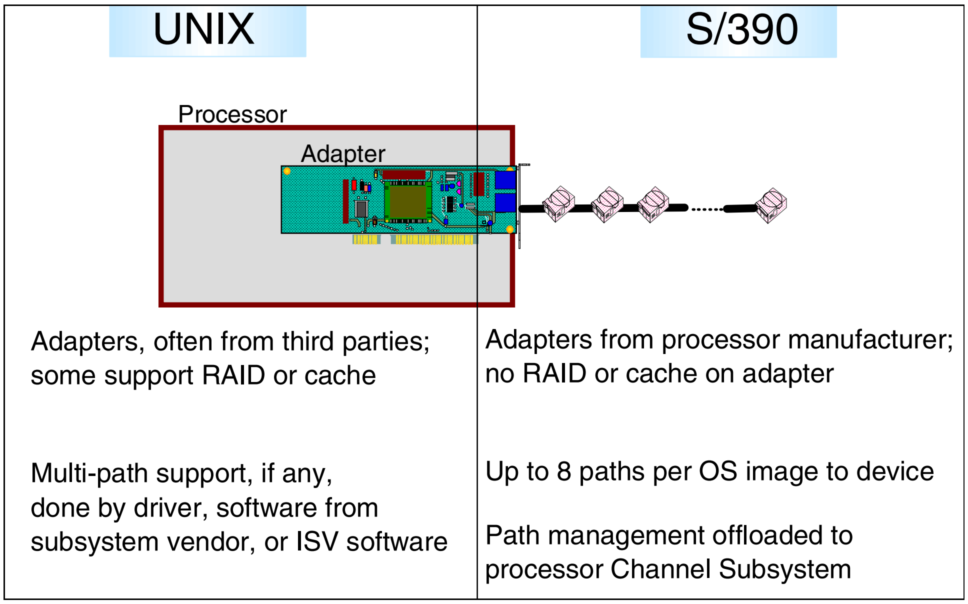


Figure 63. Processor to Storage Interface

In S/390, storage adapters are supplied by the processor manufacturer. The adapters do not support RAID or cache. In UNIX, storage adapters are supplied by the processor manufacturer or third parties. Some adapters do support caching and RAID.

S/390 supports up to 8 paths per OS image to each device, and path management is off-loaded to the channel subsystem. In UNIX environments, multi path support, if any, is performed by additional software.

Figure 64 on page 108 shows the various ways in which a device can be connected to the device adapter on the processor.

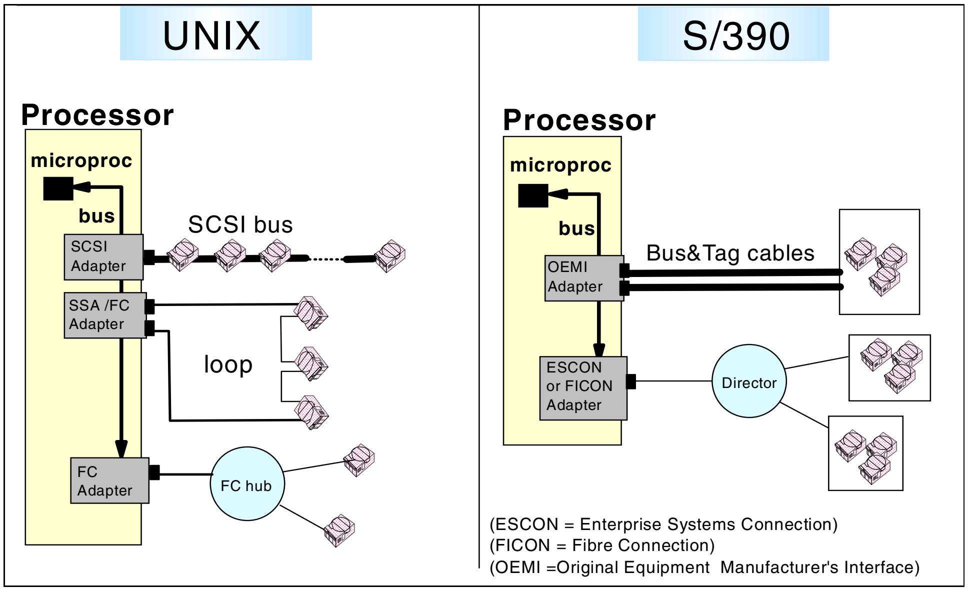


Figure 64. I/O Device Attachment

S/390 uses the Original Equipment Manufacturer's Interface (OEMI) adapter. OEMI allowed data transfer rates of up to 4.5 MB/sec. ESCON was introduced in 1991 and allows data rates of up to 17 MB/sec half duplex. The recent introduction of FICON allows for data rates of up to 100 MB/sec full duplex. The ESCON or FICON storage director is a switch and allows full bandwidth along each connection.

The SCSI protocol is based on and evolved from the OEMI standard. SCSI protocol has evolved over the years, and current implementation (SCSI-3) allows for SCSI commands to be sent over Fibre Channel (FC) connections.

S/390 disks are connected behind storage controllers. These controllers show external interfaces to the processor that are different from the internal interfaces. In UNIX systems often the disks are Just a Bunch Of Disks (JBOD) with direct attach to SCSI adapters. Today, storage controllers such as the ESS, with different upper and lower interfaces, are becoming more common.

7.5 Volume Sharing

Figure 65 on page 109 shows the main differences in how the sharing of volumes is managed.

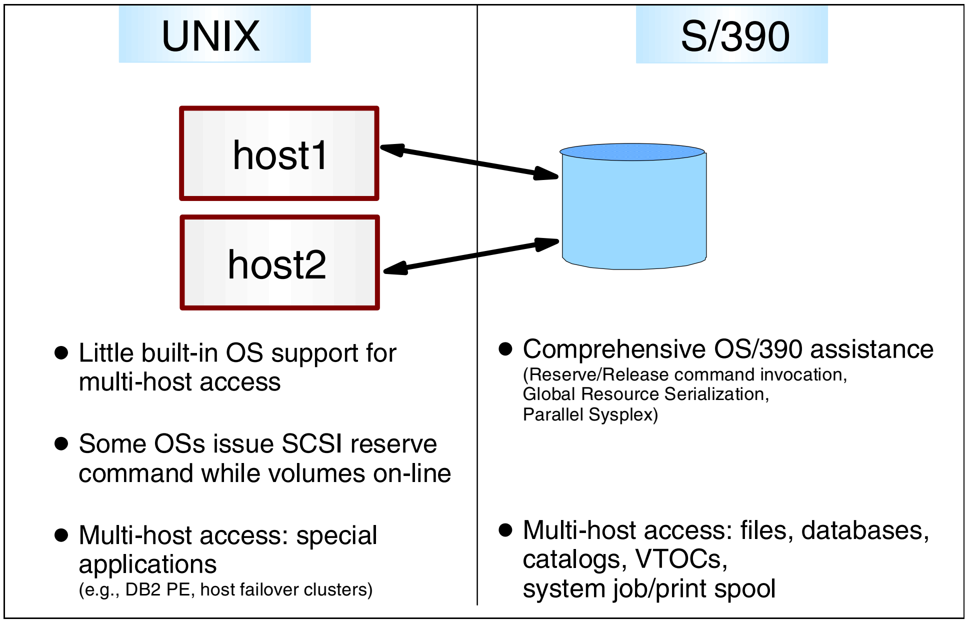


Figure 65. Multi-host Shared Disk

For many years OS/390 has supplied a comprehensive set of functions for sharing data and volumes across multiple OS/390 system images.

UNIX systems supply limited built-in support for multi access and volume sharing across systems. Some applications will reserve a volume using the SCSI reserve command. Various failover products such as HACMP support the takeover of a volume from a failing processor.

The most commonly used method of file sharing in UNIX environments is the Network File System (NFS). NFS supports sharing by having a single host own the data. Other hosts connect to the owning host over the network.

7.6 Caching

Figure 66 on page 110 illustrates the main differences in caching between S/390 and UNIX environments.

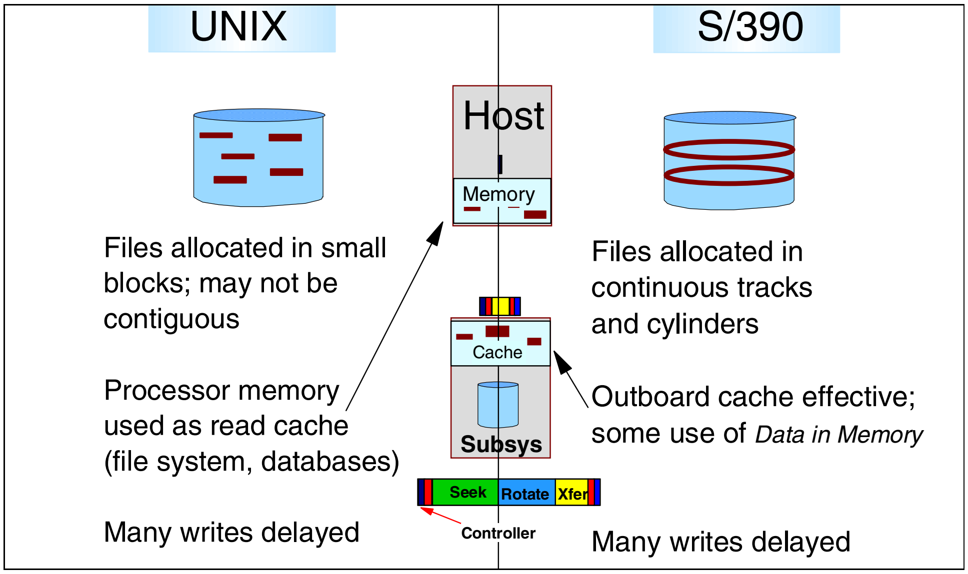


Figure 66. Cache Considerations

In S/390, files are allocated in contiguous tracks and cylinders. Outboard cache storage is used extensively in this environment because data is generally stored continuously. Sequential prefetch can be performed effectively by the storage subsystem because the data is stored continuously.

In UNIX, systems files are allocated in small blocks and these blocks may be non-contiguous. UNIX knows the location of non-contiguous blocks, and so it can perform its own sequential prefetch. UNIX uses processor memory for read cache operations.

Because OS/390 is designed for multi-system data sharing, use of memory as a read-cache may require extensive cross-system coordination. This is generally limited to catalogs and Parallel Sysplex databases and applications.

In both environments writes can be delayed by using non-volatile storage in the storage subsystem.

Appendix B. Special Notices

This publication is intended to help storage administrators, system programmers, system administrators, and other technical support staffs to understand the concept of Storage Area Networks and what functionality and benefits it gives. The information in this publication is not intended as the specification of any programming interfaces that are provided by described products. See the PUBLICATIONS section of the IBM Programming Announcement for each described product for more information about what publications are considered to be product documentation.

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program or service.

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

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

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

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

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

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

This document contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples contain 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.

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

AIX	AS/400
AT	CT
DB2	DFSMS
DFSMShsm	ECKD
ESCON	IBM ®
Magstar	MVS/XA
Netfinity	OS/2
OS/390	OS/400
Parallel Sysplec	RAMAC
RS/6000	S/390
Seascape	SP
StorWatch	System/390
Versatile Storage Server	400

The following terms are trademarks of other companies:

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

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

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

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

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

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

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

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

Appendix C. Related Publications

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

C.1 International Technical Support Organization Publications

For information on ordering these ITSO publications see "How to Get ITSO Redbooks" on page 117.

- *Data Sharing: Implementing Copy-Sharing Solutions Among Heterogeneous Systems*, SG24-4120
- *Storage Area Networks: Tape Future In Fabrics*, SG24-5474
- *IBM Magstar 3494 Tape Libraries: A Practical Guide*, SG24-4632
- *IBM Versatile Storage Server*, SG24-2221
- *Implementing the Enterprise Storage Server in Your Environment*, SG24-5420
- *Planning for IBM Remote Copy*, SG24-2595
- *IBM Enterprise Storage Server*, SG24-5465
- *IBM Storage Solutions for Server Consolidation*, SG24-5355
- *Principles of Enterprise Storage Management,* SG24-5250

C.2 Redbooks on CD-ROMs

Redbooks are also available on the following CD-ROMs. Click the CD-ROMs button at `www.redbooks.ibm.com` for information about all the CD-ROMs offered, updates and formats.

CD-ROM Title	Collection Kit Number
System/390 Redbooks Collection	SK2T-2177
IBM Enterprise Storage and Systems Management Solutions	SK3T-3694
Networking and Systems Management Redbooks Collection	SK2T-6022
Transaction Processing and Data Management Redbooks Collection	SK2T-8038
Lotus Redbooks Collection	SK2T-8039
Tivoli Redbooks Collection	SK2T-8044
AS/400 Redbooks Collection	SK2T-2849
Netfinity Hardware and Software Redbooks Collection	SK2T-8046
RS/6000 Redbooks Collection (BkMgr)	SK2T-8040
RS/6000 Redbooks Collection (PDF Format)	SK2T-8043
Application Development Redbooks Collection	SK2T-8037

How to Get ITSO Redbooks

This section explains how both customers and IBM employees can find out about ITSO redbooks, redpieces, and CD-ROMs. A form for ordering books and CD-ROMs by fax or e-mail is also provided.

- **Redbooks Web Site** `http://www.redbooks.ibm.com/`

  Search for, view, download, or order hardcopy/CD-ROM redbooks from the redbooks Web site. Also read redpieces and download additional materials (code samples or diskette/CD-ROM images) from this redbooks site.

  Redpieces are redbooks in progress; not all redbooks become redpieces and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

- **E-mail Orders**

  Send orders by e-mail including information from the redbooks fax order form to:

	e-mail address
In United States	usib6fpl@ibmmail.com
Outside North America	Contact information is in the “How to Order” section at this site: `http://www.elink.ibmlink.ibm.com/pbl/pbl/`

- **Telephone Orders**

United States (toll free)	1-800-879-2755
Canada (toll free)	1-800-IBM-4YOU
Outside North America	Country coordinator phone number is in the “How to Order” section at this site: `http://www.elink.ibmlink.ibm.com/pbl/pbl/`

- **Fax Orders**

United States (toll free)	1-800-445-9269
Canada	1-403-267-4455
Outside North America	Fax phone number is in the “How to Order” section at this site: `http://www.elink.ibmlink.ibm.com/pbl/pbl/`

This information was current at the time of publication, but is continually subject to change. The latest information may be found at the redbooks Web site.

IBM Intranet for Employees

IBM employees may register for information on workshops, residencies, and redbooks by accessing the IBM Intranet Web site at `http://w3.itso.ibm.com/` and clicking the ITSO Mailing List button. Look in the Materials repository for workshops, presentations, papers, and Web pages developed and written by the ITSO technical professionals; click the Additional Materials button. Employees may access `MyNews` at `http://w3.ibm.com/` for redbook, residency, and workshop announcements.

IBM Redbook Fax Order Form

Please send me the following:

Title	**Order Number**	**Quantity**

First name　　Last name

Company

Address

City　　Postal code　　Country

Telephone number　　Telefax number　　VAT number

☐ Invoice to customer number

☐ Credit card number

Credit card expiration date　　Card issued to　　Signature

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

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

AIT Advanced Intelligent Tape - A magnetic tape format by Sony that uses 8mm cassettes, but is only used in specific drives.

AL See Arbitrated Loop

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 2GB 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 35GB.

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 Fiber Channel Association.

FC-AL Fibre Channel Arbitrated Loop - A reference to the Fibre Channel Arbitrated Loop standard, a shared gigabit media for up to 127 nodes, one of which may be attached to a switch fabric. 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.

Full-Duplex A mode of communications allowing simultaneous transmission and reception of frames.

G_Port Generic Port - a generic switch port that either a Fabric Port (F_Port) or an Expansion Port (E_Port) function is automatically determined during login.

Gateway A node on a network that interconnects two otherwise incompatible networks.

GBIC GigaBit/Global/Generic 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 Gigabaudit Linking 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

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 Signaling that is carried in the same channel as the information.

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.

I/O Input/output

IP Internet Protocol

IPI Intelligent Peripheral Interface

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.

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.

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.

Meter 39.37 inches, or just slightly larger than a yard

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 Signaling that is separated from the channel carrying the information.

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.

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 the is dedicated

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.

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.

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	Bus-Width (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
Ultra 2 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 - See 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.

Storewatch Expert These are Storewatch applications that employ a 3 tiered architecture that includes a management interface, a Storewatch manager and agents that run on the storage resource(s) being managed. Expert products employ a Storewatch 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.

StoreWatch Specialist A Storewatch 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.

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 Unschielded Twisted Pair

Virtual Circuit A unidirectional path between two communicating N_Ports that permits fractional bandwidth.

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.

List of Abbreviations

ADSM	Adstar Distributed Storage Manager
AFS	Andrews File System
ANSI	American National Standards Institute
ASCII	American Standard Code for Information Interchange
ATA	AT Attachment
CCW	Channel Command Word
CIM	Common Interface Model
DFSMS	Data Facility Storage Management Subsystem
DMI	Desktop Management Interface
EBCDIC	Extended Binary Communication Data Interchange Code
ECKD	Extended Count key and Data
EIDE	Enhanced Integrated Drive Electronics
ERP	Enterprise Resource Planning
ESCON	Enterprise Systems Connection
ESRM	Enterprise Storage Resource Management
FC	Fibre Channel
FCA	Fibre channel Association
FC-AL	Fibre Channel Arbitrated Loop
FCIC	Fibre Channel Loop Community
FCIC	Fibre Channel Inter-Connect
FE	Fibre Element
FICON	Fibre Connection
FL	Fibre Loop
GB	GigaByte
Gb	GigaBit
GBIC	Gigabit Interface Converters
GLM	Giga Bit Link Module
HA	High Availability
HBA	Host Bus Adaptor
HFS	Hierarchical File System
HIPPI	High Performance Parallel Interface
I/O	Input Output
IBM	International Business Machines Corporation
IDE	Integrated Drive Electronics
IETF	Internet Engineering Task Force
IIS	Intelligent Storage Server
IP	Internet protocol
ISO	International Standards Organization
ISV	Independent Software Vendor
IT	Information Technology
ITSO	International Technical Support Organization
JBOD	Just a bunch of Disks
JFS	Journalling File System

JMAPI	Java Management Application Programming Interface
LAN	Local Area Network
LIP	Loop Initialization Protocol
LVM	Logical Volume Manager
MAR	Media Access Rules
MB	Mega Byte
MIA	Media Interface Adapters
MIB	Management Information Block
MMF	Multi Mode Fibre
NAS	Network Attached Storage
NC	Network Computer
NFS	Network File System
NTFS	NT File System
PBC	Port Bypass Circuit
PPRC	Peer to peer Remote Copy
RAID	Redundant Array of Inexpensive Disks
ROI	Return on Investment
S/390	System 390
SAN	Storage Area Network
SCSI	Small Computer Systems Interface
SES	SCSI Enclosure Services
SLS	Single Level Storage
SMB	Server Message Block
SMF	Single Mode Fibre
SMS	System Managed Storage
SNIA	Storage Networking Industry Association
SNMP	Simple Network Management Protocol
SPD	System Product Division
SSA	Serial Storage Architecture
SSD	Storage systems Division
STA	SCSI Trade Association
TCP/IP	Transmission Control Protocol / Internet Protocol
TIMI	Technology Independent Machine Interface
ULTRA ATA	Ultra AT Attachment
VSAM	Virtual Storage Access Method
VTOC	Volume Table of Contents
WAN	Wide Area Network
WEBM	WEB Management
XRC	Extended Remote Copy

Index

ITSO Redbook Evaluation

Introduction to Storage Area Network, SAN
SG24-5470-00

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

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

Which of the following best describes you?
_ Customer _ Business Partner _ Solution Developer _ IBM employee
_ None of the above

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

Overall Satisfaction __________

Please answer the following questions:

Was this redbook published in time for your needs? Yes___ No___

If no, please explain:

What other redbooks would you like to see published?

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

SG24-5470-00
Printed in the U.S.A.

Introduction to Storage Area Network, SAN

SG24-5470-00