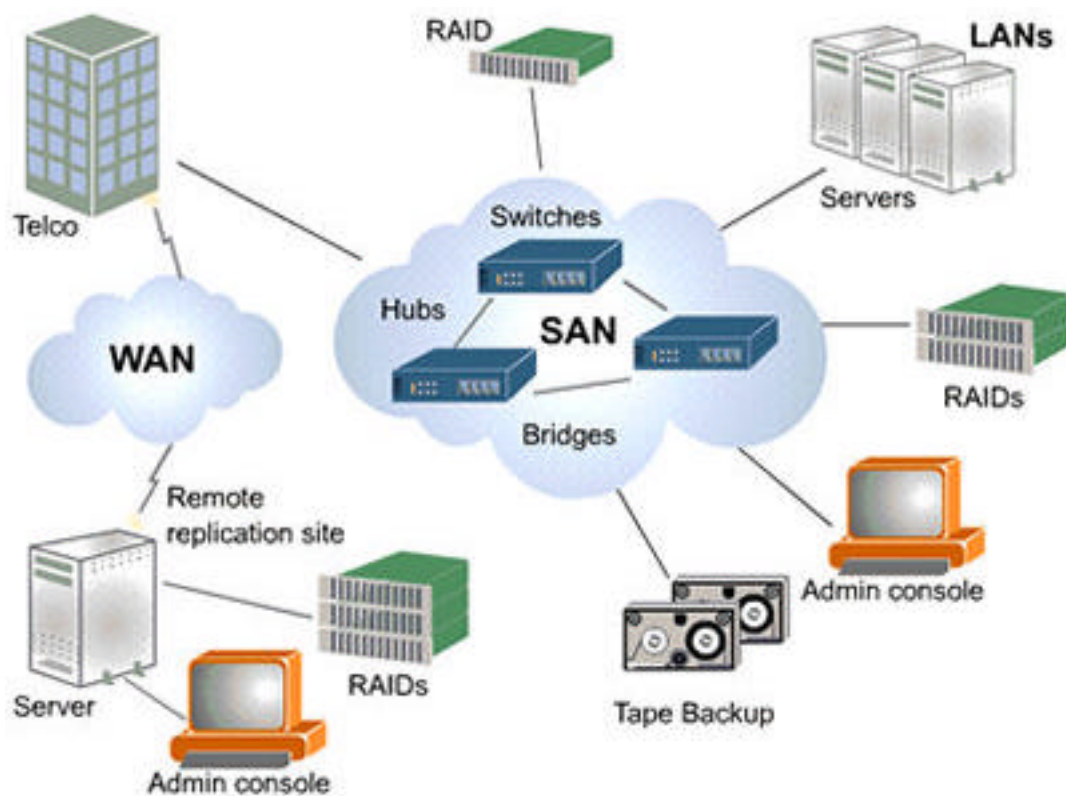# SAN TUTORIAL

by

Manhattan Skyline GmbH

1. What is a SAN?

A SAN, or storage area network, is a dedicated network that is separate from LANs and WANs. It generally serves to interconnect the storage-related resources that are connected to one or more servers. It is often characterized by its high interconnection data rates (Gigabits/sec) between member storage peripherals and by its highly scalable architecture. Though typically spoken of in terms of hardware, SANs very often include specialized software for their management, monitoring and configuration.

SANs can provide many benefits. Centralizing data storage operations and their management is certainly one of the chief reasons that SANs are being specified and developed today. Administrating all the storage resources in high-growth and mission-critical environments can be daunting and very expensive. SANs can dramatically reduce the management costs and complexity of these environments while providing significant technical advantages.

SANs can be based upon several different types of high-speed interfaces. In fact, many SANs today use a combination of different interfaces. Currently, Fibre Channel serves as the de facto standard being used in most SANs. Fibre Channel is an industry-standard interconnect and high-performance serial I/O protocol that is media independent and supports simultaneous transfer of many different protocols. Additionally, SCSI interfaces are frequently used as sub-interfaces between internal components of SAN members, such as between raw storage disks and a RAID controller.
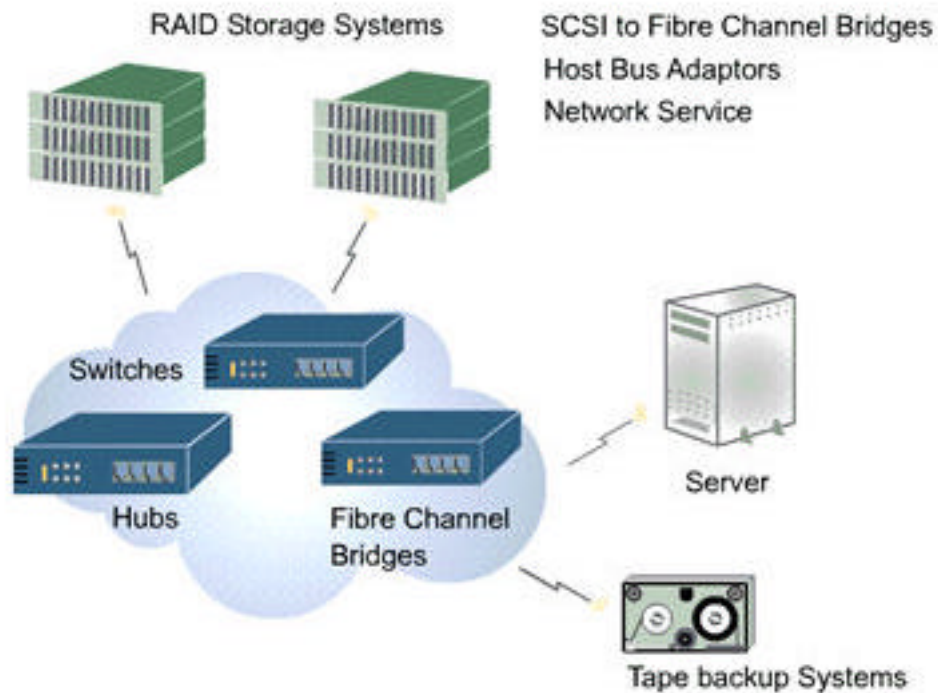
Provding large increases in storage performance, state-of-the-art reliability and scalability are primary SAN benefits. Storage performance of a SAN can be much higher than traditional direct attached storage, largely because of the very high data transfer rates of the electrical interfaces used to connect devices in a SAN (such as Fibre Channel). Additionally, performance gains can come from opportunities provided by a SAN's flexible architecture, such as load balancing and LAN-free backup. Even storage reliability can be greatly enhanced by special features made possible within a SAN. Options like redundant I/O paths, server clustering, and run-time data replication (local and/or remote) can ensure data and application availability. Adding storage capacity and other storage resources can be accomplished easily within a SAN, often without the need to shut down or even quiese the server(s) or their client networks. These features can quickly add up to large cost savings, fewer network outages, painless storage expansion, and reduced network loading.

By providing these dedicated and "very high speed" networks for storage and backup operations. SANs can quickly justify their implementation. Offloading tasks, such as backup, from LANs and WANs is vital in today's IT environments where networks loads and bandwidth availability are critical metrics by which organizations measure their own performance and even profits. Backup windows have shrunken dramatically and some environments have no backup windows at all since entire data networks and applications often require 24x365 availability.

As with many IT technologies, SANs depend on new and developing standards to ensure seamless interoperability between their member components. SAN hardware components such as Fibre Channel hubs, switches, host bus adapters, bridges and RAID storage systems rely on many adopted standards for their connectivity. SAN software, every bit as important its hardware, often provides many of the features and benefits that SANs have come to be known for. SAN software can provide or enable foundation features and capabilities, including:

- SAN Management
- SAN Monitoring (including "phone home" notification features)
- SAN Configuration
- Redundant I/O Path Management
- LUN Masking and Assignment
- Serverless Backup
- Data Replication (both local and remote)
- Shared Storage (including support for heterogeneous platform environments)

2. SAN Hardware

RAID Storage Systems   SCSI to Fibre Channel Bridges
                          Host Bus Adaptors
                          Network Service

Switches

Hubs   Fibre Channel
   Bridges

Server

Tape backup Systems

SANs are built up from unique hardware components. These components are configured together to form the physical SAN itself and usually include a variety of equipment. RAID storage systems, hubs, switches, bridges, servers, backup devices, interface cards and cabling all come together to form a storage system that provides the resources that facilitate the policies of an IT organization.

It is very important to select the hardware devices (and their configuration) for a SAN with care and consideration. Many of the "standards" that are involved with SANs are concerned with interoperability. Some of these standards are still evolving and haven't been equally adopted by all manufacturers of equipment used in SANs. This can lead to difficulties when matching up devices from different vendors and suppliers. Since SANs are typically just as dependent upon software for their proper operation, it can be vital to secure the latest version information about software (and firmware) and potential compatibilty issues.

Working with Manhattan Skyline you choose a company that specialize in the design, integration and implementation of SAN systems. MSKL can bring you the lastest software and hardware and can speed the process of successfully deploying SAN technology. By working with other vendors and manufacturers, MSKL can help ensure that the promised benefits are realized and successfully integrated into new or existing IT infrastructures.

3. RAID Systems
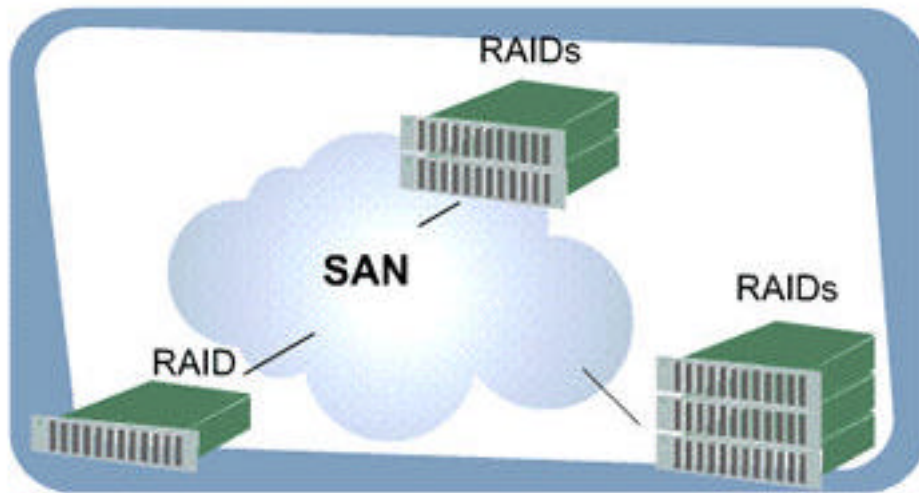
Most contemporary SANs include RAID systems as their primary data storage devices. These systems have become highly evolved and offer the foundation features that have come to be expected in a modern SAN. First and foremost, RAID systems offer data protection, or fault tolerance, in the event of a component or I/O path failure. This is true even if fundamental elements, such as disk drives, fail in the system. Additionally, by way of numerous data striping techniques (described below),and controller configurations, today's RAID systems offer very high performance, storage capacity, scalability, and survivability. Other reliability features available in today's RAID systems include redundant cooling systems, power supplies, controllers and even monitoring circuitry. These, and other features and characteristics, contribute dramatically to high data availability in a SAN. Modern RAID systems can even permit the direct connection of backup equipment, thus facilitating LAN-free and even serverless data backup and replication.

Background

The rrots of RAID technology can be traced back to 1987, when Patterson, Gibson and Katz at University of California at Berkeley, published a paper entitled "A Case for Redundant Arrays of Inexpensive Disks (RAID)". The ideas presented and explained in the paper involved combining multiple small, inexpensive disk drives into arrays in order to provide features that single drives alone couldn't supply. These new features centered around improving I/O Performance and automatically preserving the contents of drives during, and after, drive or component failures.

These drive arrays are presented to a computer as a single logical storage unit (LUN) or drive. Additional benefits of drive arrays include the ability to make these arrays fault-tolerant by redundantly storing data in various ways. Five of the array architectures, RAID levels 1 through 5, were defined by the Berkeley paper as providing disk fault-tolerance with each offering various trade-offs in features and performance. Overall, the idea was to improve the reliability of the storage system by significantly increasing the Mean Time Between Failure (MTBF) for the array and to dramatically improve the storage system's performance.

A sixth common type of RAID architecture, RAID 0, has subsequently been defined that can substantially improve the I/O performance of an array but it provides no data protection should a hardware component fail. The performance gains possible with RAID 0 arrays can be very dramatic. RAID 0 arrays are ideal for applications that demand the highest possible data throughput. Note that these applications must be able to tolerate possible data loss, and service interruption, if a drive or other component in the array fails.

Data Striping

Fundamental to the RAID concept is "striping", a method of dividing and distributing data among the arrayed drives and effectively joining multiple drives into one logical storage unit. Striping involves partitioning each drive's storage space into stripes that may be as small as one block (512 bytes) or as large as several megabytes. These stripes are then interleaved in a round robin fashion, so that the combined space is composed of joined stripes from each drive. In most instances, the application environment determines the suitability of larger vs. smaller stripe sizes.

Most contemporary multi-user operating systems like UNIX, Solaris, NT and Netware support overlapping disk I/O operations across multiple drives. However, in order to maximize throughput for a combined disk subsystem, its I/O load must be balanced between all of its member drives so that each drive can be kept as active as possible. High parallelism during I/O operations generally translates into much greater performance.

In multiple drive systems without stripping, the disk I/O load is almost never perfectly balanced. Some drives will contain data that is frequently accessed while other drives will only occasionally be accessed. During intense I/O operations, performance can be optimized by striping the drives in the array with stripes large enough so that each record potentially falls entirely within one stripe segment. This helps insure that data and I/O operations are evenly distributed across the arrayed drives, thus allowing each drive to work on separate I/O operations at the same time, thereby maximizing the number of simultaneous I/O operations that can be serviced by the array. By contrast, in data-intensive applications that access large records, smaller stripe sizes can be used so that each record will span across many, or all, of the drives in an array with each drive storing only part of a record's data. This can allow long record accesses to be performed faster, since the data transfers can occur in parallel on multiple drives in the array. Applications such as digital video editing, audio/video on demand, imaging and data acquisition that employ long record accesses are examples of applications that often achieve optimum performance with smaller stripe sizes. Unfortunately, smaller stripe sizes typically rule out multiple overlapping I/O operations since each I/O will typically involve all of the drives.

RAID 0
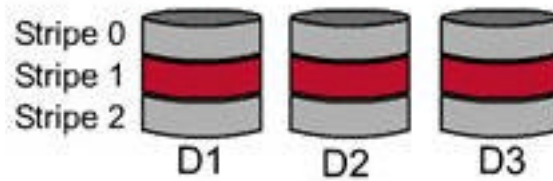
Data is striped across a group of hard disk drives (HDDs) without generating parity or redundant data. Improved performance is achieved by reading and writing to the HDDs in parallel. If an HDD fails, data is loss since there is no data redundancy. RAID 0 is employed when maximum performance is more important than possible data loss.

RAID 1
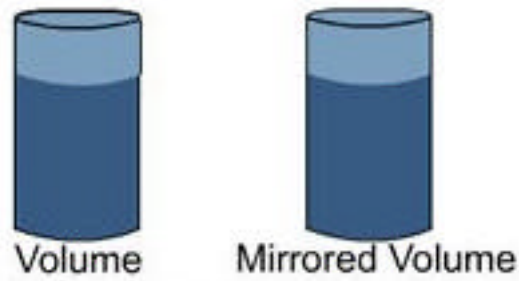
HDDs, in the array, are divided into two groups. When writing, identical data is stored onto HDDs in both groups. Data can be read from either group. No parity data is generated. This approach is popular for its simplicity and is a frequently used method. Its shortcomings are that it halves the total available array capacity (since a full copy of the data is maintained) and this method provides no improvement in speed.
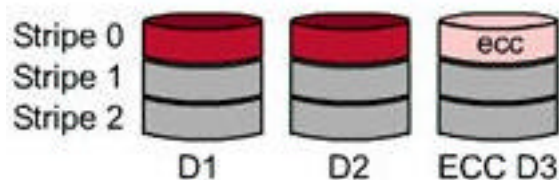
RAID 2

Data is byte, or bit, striped across groups of drives and employs an error correction code (ECC) technique widely employed with random access memory (RAM). After generating ECC codes or hamming codes, the controller writes data onto multiple HDDs at the byte or bit level with some drives assigned to storing ECC information. If any of the HDDs fail, this redundancy scheme can restore the original data in real-time. The ECC overhead generated by this method is large compared to RAID 3-5 and commercial products rarely employ it.

RAID 3

A simplified version of RAID 2. This scheme consists of an array of HDDs for data and one unit for parity. Data is byte, or bit, striped across groups of drives, with one drive dedicated to storing error correction code (ECC) information. Instead of the multiple ECC bits found in RAID 2, single bit parity is used. It is not the most efficient method for accessing small amounts of data, but it is well suited for high speed access to large data blocks.
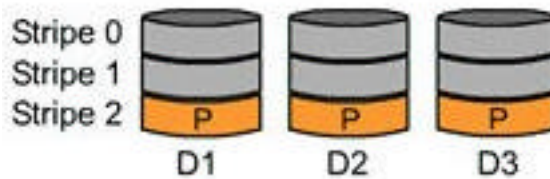
(RAID 4)
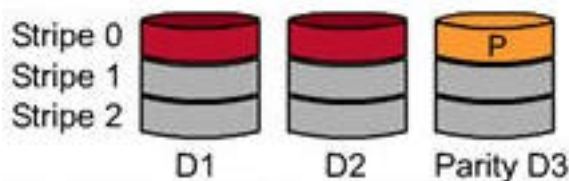
Identical to RAID 3 except that data is dispersed by blocks (or sectors) instead of bytes or bits. Data can be read from any individual drive in the array (except the parity drive), allowing read operations to be overlapped. HDDs in the array can operate independently. Where RAID 3 constantly accesses all the HDDs in the array, RAID 4 only does so if necessary. Since stored data and parity must be retrieved before new data is written, processing time is longer than for RAID 3 when writing. Commercial products rarely employ RAID 4.

RAID 5

This technique avoids the write bottleneck caused by the single dedicated parity drive in RAID 4. In RAID 5, parity is dispersed and stored on all HDDs. This approach was developed to resolve the fact that the dedicated parity HDD becomes a bottleneck for performance in RAID 4. RAID 5 is very popular and is commonly employed in storage products today.

Summary

RAID 0 is the fastest and most efficient array type but offers no data redundancy or fault-tolerance.

RAID 1 is the array technique of choice for performance-critical, fault-tolerant environments and is the only choice for fault-tolerance if no more than two drives are available.

RAID 3 is a popular choice for data-intensive or single-user applications that access long sequential records. However, it does not typically allow multiple I/O operations to be overlapped.

RAID 4 offers no practical advantages over RAID 5 and does not typically support multiple simultaneous write operations.

RAID 5 is generally the best choice for multi-user environments that are not particularly sensitive to write performance. At least three, and typically five or more drives are required to build a RAID 5 array.

Many contemporary RAID systems incorporate a newer serial interface standard known as Fibre Channel-Arbitrated Loop (FC-AL). It is yet another interface option for RAID subsystems, and is currently very popluar. FC-AL is capable of providing data throughputs up to 200 Mbytes/sec (in dual loop configurations) while allowing RAID subsystems, or other connected peripherals, to be located up to 10 kilometers from host. This interface also supports up to 126 disk drives, or other devices, on a single controller (compared to seven, or fifteen, devices using conventional SCSI). FC-AL can be operated in either single- or dual-loop configurations. The dual loop architecture provides an added level of I/O path redundancy by supporting two separate I/O paths for each connected device.

4. Switches, Hubs and Bridges
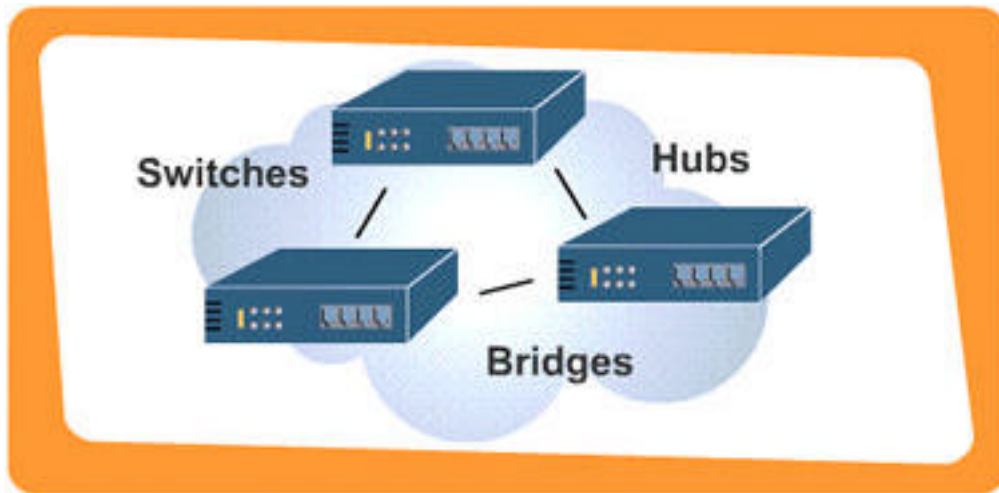
More and more, the design and deployment of SAN technology involves incorporating specialized interconnection equipment. This category of devices often includes Fibre Channel Hubs, Switches and Bridges. This hardware is generally responsible for linking together the data storage peripherals, such as RAID systems, tape backup units and servers within a SAN.

These interconnection devices are somewhat analogous to their LAN related counterparts. They perform functions such as data frame routing, media and interface conversion (i.e. copper to optical, Fibre Channel to SCSI), network expansion, bandwidth enhancement, zoning, and they allow concurrent data traffic. Just as customers today are more involved in the design and implementation of their LANs and WANs, they are also looking at these building blocks of SANs to create their own SAN solutions.

Fibre Channel HBAs, hubs, switches, and FC/SCSI bridges are some of the building block components with which IT administrators can develop SAN based backup solutions, server clusters, enhanced bandwidth, extended distance and other application driven challenges. Selecting the appropriate pieces to address these issues requires an understanding of what each component can do. When, for example, is a fabric switch a better than a hub? When should hubs and switches be used in combination? There are no universal answers to these questions, but understanding the architecture and capabilities of switches, hubs and bridges provides a basis for making appropriate choices for SAN designs.

FC Arbitrated Loop Topology
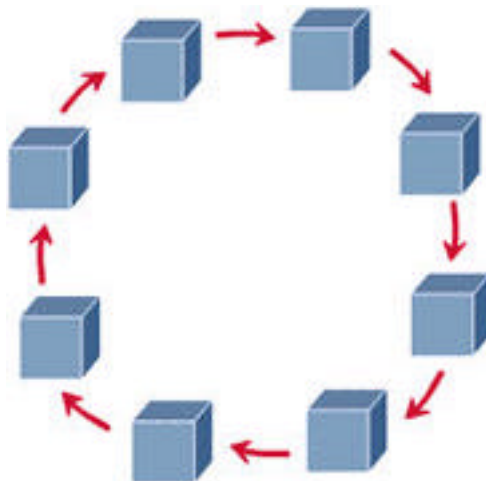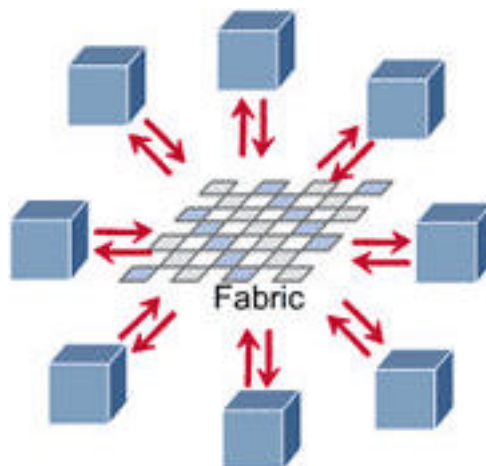

FC Switched Topology

Figure 1. FC-AL and Non-Blocking cross-point switch topologies

Fibre Channel Hubs

Similar in function to Ethernet or Token Ring hubs, an Arbitrated Loop hub is a wiring concentrator. Hubs were engineered in response to problems that arose when Arbitrated Loops were build simply connecting the transmit lines to the receive lines between multiple devices.

A hand-build daisy chain of transmit/receive links between three or mode devices allows for a circular data path or loop to be created, but poses significant problems for troubleshooting and adding or removing devices. In order to add a new device, for example, the entire loop must be brought down as new links are added. If a fiber optic cable breaks or a transceiver falls, all cables and connectors between all devices must be examined to identify the offending link.

Hubs resolve these problems by collapsing the loop technology into a star configuration. Since each device is connected to a central hub, the hub becomes the focal point of adds/moves or changes to the network. Arbitrated Loop hubs provide port bypass circuitry that automatically reconfigures the loop if a device is removed, added or malfunctions. Before a new device is allowed to be inserted into a loop, the hub will, at a minimum, verify and validate its signal quality. Devices with poor signal quality, or an inappropriate clock speed, will be left in bypass mode and allow other devices on the loop to continue operating without disruption. Hubs typically provide LEDs for each port that provide "at a glance" status of insertion, bypass or bad-link state. These features enable a much more dynamic environment where problems can be more readily identified, particularly since devices can be hot-plugged or removed with no physical layer disruption.

A hub port can be designed to accept either electrical or optical I/O. This capability is very useful in designing a network or configuring it. For instance, if it were desirable to locate the hub some distance from the server, an optical connection (long wave or short wave) could be used between the server and hub while copper connections could be used between the hub and local controllers. Hubs can be cascaded to provide additional ports for even more connectivity.

Fibre Channel Switches

Fibre Channel fabric switches are considerably more complex than loop hubs in both design and functionality. While a hub is simply a wiring concentrator for a shared 100MB/sec segment, a switch provides high-speed routing engine and 100MB/sec data rates for each and every port. Apart from custom management functions, hubs do not typically participate in Fibre Channel activity at the protocol layer. A fabric switch, by contrast, is a very active participant in Fibre Channel conversations, both for overseeing the flow of frames between initiators and targets (buffer-to-buffer credit, fabric loop support, etc.) at each port. Providing fabric services, 100MB/sec. Per port performance and the advanced logic required for routing initially kept the per port cost of first generation fabric switches quite high. Second generation, ASIC (Application Specific Integrated Circuit) based, fabric switches have effectively cut the per port cost by more than half. This brings Fibre Channel fabric switches within reach of medium to large enterprise networks.

Fibre Channel Arbitrated Loops (FC-AL) are serial interfaces that create logical point-to-point connections between ports with the minimum number of transceivers and without a centralized switching function. FC-AL therefore provides a lower cost solution. However, the total bandwidth of a Fibre Channel Arbitrated Loop is shared by all ports on the loop. Additionally, only a single pair of ports on the loop can communicate at one time, while the other ports on the loop act as repeaters.

Fibre Channel to SCSI Bridges

Fibre Channel to SCSI bridges provide conversion between these two different electrical interfaces and therefore allow IT managers to leverage investments in existing SCSI storage devices, while taking full advantage of the inherent benefits of Fibre Channel technology. These devices are commonly used to connect Fibre Channel networks to legacy SCSI peripherals, such as tape abckup systems.

5. Backup Solutions

One of the most valuable time- and cost-saving features of a SAN architecture is its ability to offload backup operations from a LAN and/or backup servers. This capability can significantly increase the amount of LAN bandwidth available to network clients and end users during backup operations. When backup servers are relieved from the "data mover" tole, they become more available for other productive tasks.

LAN-free and serverless backup solutions optimize backup operations by offloading backup data traffic from a LAN, thereby increasing the amount of LAN bandwidth available to end users. Serverless backup extends these performance gains by removing more than 90 percent of the backup administration overhead that is usually placed on a backup server as backups are performed. This is achieved by incorporating some of the backup intelligence into the data storage or connectivity peripherals themselves. This can significantly free up backup servers by releasing them from large portions of a backup operation's administration and data moving chores. Using these SAN based backup solutions lets administrators optimize network and server utilization.

Background

Traditional backup operations place the application server, the backup server and the LAN all in the data path. Consequently, as the amount of storage grows, the amount of time and network resources needed to back it up grows. Now that businesses and organozations have moved towards 24 x 365 operation, backup tasks are competing with critical business applications for server time and network resources. Invariably, this causes network congestion and can result in business slowdowns.

For "serverless" backup operations, host computers (servers) do not "handle" or touch the backup data itself. Instead, these hosts merely direct and monitor the backup without actually moving the data. The backup data is copied directly from disk to tape, or disk to disk, by the storage peripherals themselves using intelligence that is incorporated into them. Optionally, this intelligence can even be placed inside of other SAN components, such as Fibre Channel switches or hubs. Freed from the routine data transport burden, server resources can be put back to more productive uses. Or, in other words, the backup or tape server is delegated the role of "backup

coordinator", rather than data mover. Serverless backup takes LAN-free a step further since it removes backup traffic from both the LAN and the backup server.

By contrast, with simple "LAN-free" backup operations, the backup and restore data (traffic) is removed from the LAN but still flows through the administrating server as it moves between data storage and backup devices. The benefit here is still valuable, since backup traffic is taken off of the LAN, reducing LAN congestion. While both serverless and LAN-free backup keep backup data off of the LAN, only the serverless backup fres up the administrating server as well – placing the data movement tasks onto the smart peripherals.

Specifically, smarter peripherals can now perform much of their own backup by supporting newer technologies and APIs – such as the "extended copy command", a Storage Networking Industry Association specification that lets data be moved between storage devices on different buses. The backup server issues the command to a data mover in the SAN, and then removes itself from the data path. This way only the source, destination and SAN devices are involved. The constraints related to the memory, I/O and CPU performance of the backup server itself are eliminated as the data moves through a high-performance copy device or agent that is optimized for data movement. The logic here is fairly obvious since this frees up the backup server for other business-critical applications, and supports server consolidation. The dedicated backup server is no longer needed. Additionally, backups can complete much more quickly over higher speed networks – such as Fibre Channel.

Serverless backup systems can also provide additional cost savings by eliminating expensive, high-end servers. Another advantage unique to the serverless backup architecture is its ability to stream the same data to several tape libraries or other targets simultaneously, even if they are geographically separated, without the need for copying and moving the actual tapes – an important advantage in disaster recovery plans.

Backup Hardware

Mechanically, backup equipment used in SANs is typically the same as that used in conventional configurations. What is different, however, is how these devices are interfaced to their host servers and client storage systems. Since most contemporary SANs are connected together using Fibre Channel, and since many backup devices use SCSI interfaces, some type of bridge is often required.

These bridges perform the electrical, and any protocol, conversions required between the disparate buses or channels. There are many bridge manufacturers that supply these units, but it is vital to confirm compatibility with the selected backup device(s) before attempting to configure the units together or specifying units for purchase. This SAN topic has many caveats and is often an area that benefits from direct experience. MSKL is a great source for this type of compatibility informations.

If a serverless backup solution is being considered or designed, it is important to note that some of these bridge manufacturers offer "smart" units that include built-in copy functions. As mentioned above, this set of features is generally required for true serverless backup. In practice, small applications running on a selected server activate and instruct these copy agents remotely, then sit back and monitor progress while the smart bridge, or similar device, moves the data.

These software, or firmware, copy agents can even be found in certain Fibre Channel switches and hubs. There are some software companies that have even developed special programs that can be "downloaded" into these units to give them these independent copy capabilities. With all of these options avilable, one can see the importance of selecting components very carefully. First time SAN adopters may want to consider consulting with SAN specialists before purchasing this type of equipment for deployment in a backup solution.

Through SAN backup solutions typically employ a tape server, tape library, and disk-based storage attached together with a Fibre Channel infrastructure, it is becoming fairly common for backup solutions to include disk to disk copies. With today's backup windows shrinking and IT policies calling for remote site replication, backup can now mean much more than just making tapes. Backup can now include data replication to local or remote disk sites via WANs, disk to disk backups that accommodate offline data processing and short backup windows, or all of these at the same time.

6. SAN Software
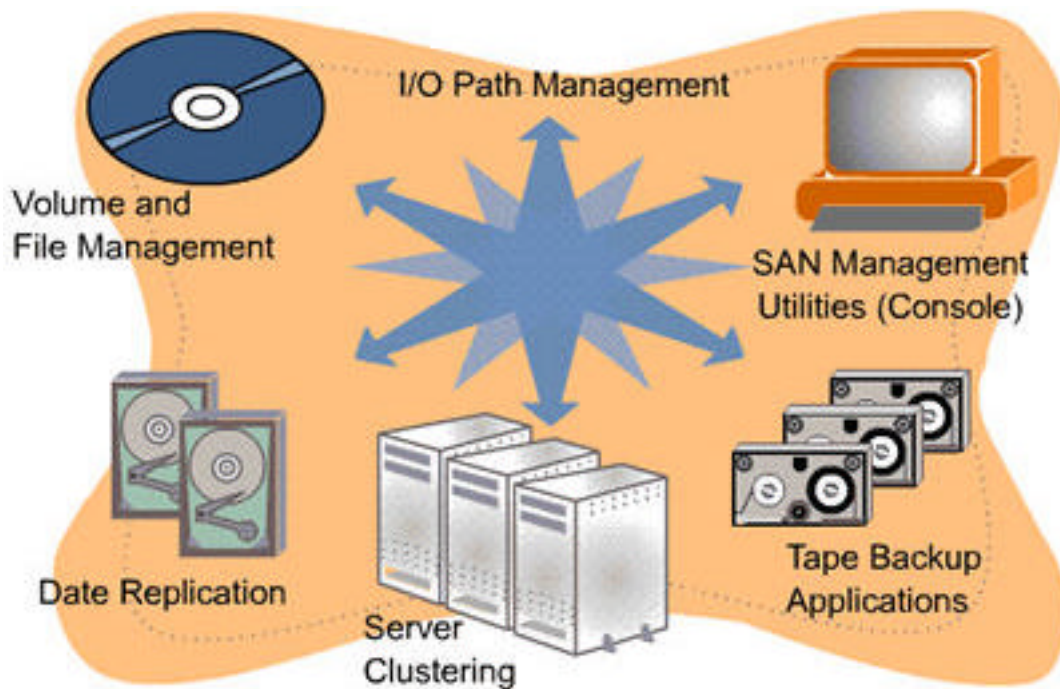
More than ever before, software is playing a vital role in the successful deployment of SANs. Much of the technology, and many of the features, provided by SANs are actually embedded in its software. From volume management to serverless backup, choosing and configuring the software components is very important and should be done with care.

Many Companies offer a wide variety of software products and solutions that are specifically designed to enhance the performance, data availability and manageability of SANs. Some of these solutions have been custom developed by these companies for lines of data storage systems. Other offerings are more universal or "open" and address a very broad range of customer requirements and equipment.
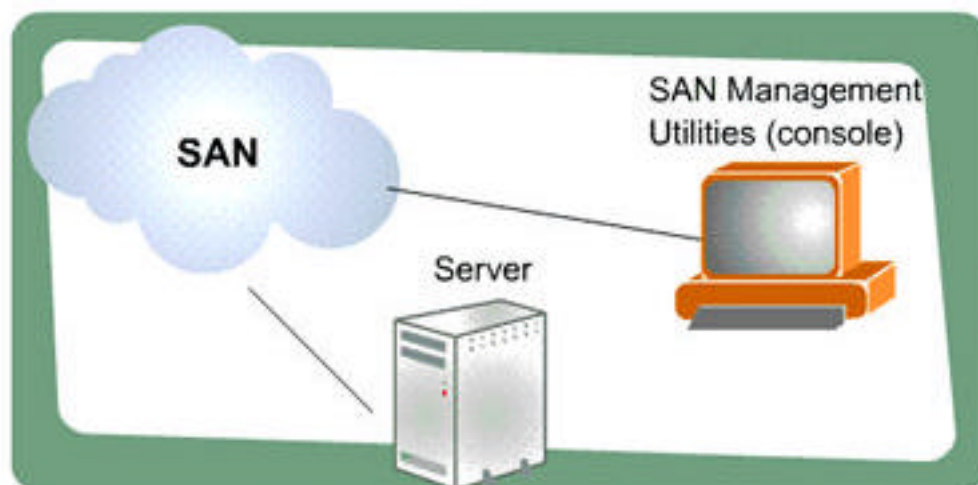
SANs today can become rather complex in both their design and implementation. Adding to this issues relating to their configuration, resource allocation and monitoring. These tasks and concerns have led to a need to proactively manage SANs, their client servers and their combined resources. These needs have led to this new category of software that has been specifically developed to perfrom these functions and more. Though somewhat recent in its development, SAN management software borrows heavily from the ideas, functions and benefits that are mature and available for traditional LANs and WANs.

Ideally, SAN management software would be universal and work with any SAN. But in today's multi-vendor and hardware-diverse SAN environments, this management software is often proprietary or tied to certain products and vendors. While this is beginning to change, it still means that SAN management software must be selected with great care, and consideration given to the SAN equipment manufacturers, OS platforms, firmware revisions, HBA drivers, client applications, and even other software that may be running on the SAN. Until such time as SAN management software becomes very universal, it will be quite important, and even vital, to work closely with product providers in order to achieve the best results in obtaining SAN management goals and benefits. When selecting SAN management software, it is particularly important to ask lots of questions. Inquire about supported OS platforms, compatibility issues with other vendors, minimum revision levels of the components that are (or will be) in the SAN, and any feature restrictions that may be imposed in certain environments.

Even when obtaining the entire SAN from a single provider, it is still prudent to ask these types of questions (especially those related to compatibility.

What is SAN management?

Management often means different things to different people. A network administrator, for example, is generally concerned with data transport, or moving user information reliably from one point to another. A network administrator is therefore concerned with such management issues as bandwidth utilization, provisioning redundant links in a meshed topology to guarantee alternate data paths, support for multiple protocols, error-free delivery, and so on. In short, the network administrator is concerned with getting data from A to B, but typically is not aware of what happens to the data once it arrives at its destination.

A storage administrator, on the other hand, is less concerned about data transport than about the organization and placement of data once it arrives at its destination. LUN mapping, RAID levels, file integrity, tape backup, disk utilization, and so on are the day by day preoccupations of the storage administrator. Storage management assumes that the data arrived at B intact and then decides how it is to be written onto disk or tape.

These different views of management converge in a SAN, since the proper operation of storage area networks requires both management of data transport and management of data placement. By introducing networking between servers and storage, a SAN forces traditional storage management to broaden its scope to include network administration, and encourages traditional network management to extend its reach to data placement and organization. Consequently, network management frameworks such as CA Unicenter™, Tivoli™and HP OpenView™ are incorporating storage management utilities for the SAN, while storage software providers, such as VERITAS™and Legato™, are including modules to monitor the Fibre Channel network transport.

The integration of storage and networking management functions is a uniqueb product of SAN evolution and confirms the shift from a server-centric to a data-centric model in the enterprise.

When is SAN management required?

Ideally, the decision to deploy a managed or unmanaged SAN solution should be determined by the application. If a user's application can withstand potential disruptions due to network outages, there is less requirement for advanced management features. In many early adopter SAN installations, however, unmanaged SAN configurations were installed by default, either due to a lack of manageable products at the time or the assumption that small homogeneous SANs would not require management. Some of these early market installations continue to run mission-critical applications, and attempt to address the lack of management by provisioning redundant data paths. However, this solution is not a substitution for SAN management. Without management visibility, the failure of a backup path may go unnoticed, resulting in system failure if the primary path subsequently goes down.

Enterprise and information networks are so dependent on continuous data access that even critical tape backup schedules or planned changes to the network are difficult to accommodate. Often in small single-vendor departmental installations, the time spent identifying and correcting a simple problem may represent an unacceptable disruption to the end users.

Awareness of SAN management surfaces almost instantly whenever failures occur. Without tools to quickly identify and isolate problems, the mean time to repair slowly expands to fill the gap between initial symptoms and all the remedial processes required to track down and finally fix the offending cable, transceiver, host bus adapter or application error. This is normally not the desired method for gaining appreciation of management capabilities.

7. Server Clustering
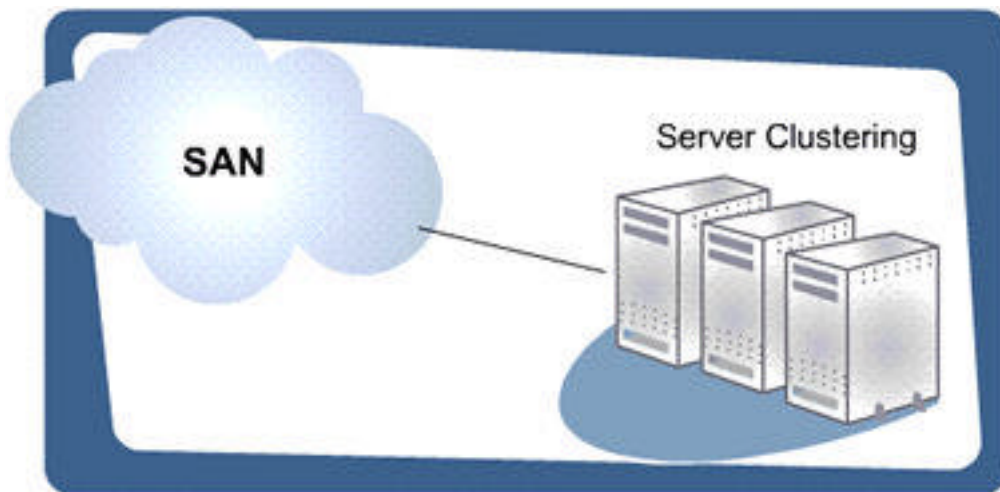
In a SAN context, server clustering generally refers to the grouping together of servers for the purpose of enhancing their performance and/or providing failover protection in the event that a member server malfunctions. Uninterrupted and seamless availability of data and applications during and after a server failure is a primary benfit of a server cluster architecture within a SAN.

Though servers can be clustered together outside of a SAN environment, there are many benefits associated with clustering them together as part of a SAN. These benefits include shared access to resilient disk and tape backup systems, higher performance data replication options, improved storage scalability, and enhanced resource availability through the inherent advantages of SAN based storage systems.

In many cases, the specialized software involved in server clustering can even fail back a server in the event it is repaired or begins working properly again. Other software options can even divide application tasks among servers in a cluster in order to dramatically improve its response time and performance.

Server clusters can be very valuable when considering disaster recovery options. In these cases, the clustered servers might be joined together with WAN connections instead of direct SAN interfaces in order to provide enough geographic distance.

However, these remotely connected servers might still be connected to local SANs at their location and these SANs could even include additional clustered servers.

Background

The term clustering has several meanings in information technology. A generic definition of clustering covers any situation in which multiple servers share a task or data.

In today's business and e-business environments, more and more customers are becoming critically dependent upon their information technology resources. As a result, they demand that these resources always be available. Any outage could have serious business implications – including lost revenue and lost business. At the extreme, an extended system outage can cause a business to be permanently closed. The cost of one hour of system downtime can range from tens of thousand of dollars to several million dollars, depending upon the nature of the business. Therefore, many users require that their system services be contiuously available, that is that the services be available 24 hours a day, 365 days a year. Technologies that support increased computer system availavility have become critical to many business.

A key technology that enables continuous data and application availability is clustering. A cluster is a collection of one or more complete systems that work together to provide a single, unified computing capability. The perspective from the end user is that the cluster operates as though it were a single system. Any single outage (planned or unplanned) in the cluster will not disrupt the services provided to the client. Client services can even be relocated from system to system within the cluster in a relatively transparent fashion.

SAN architectures can also accommodate multi-dimensional growth. Capacity management techniques can be used to ensure that new storage can be added continuously, so server applications always have the storage capacity they need. If more processing power is needed, SANs also facilitate adding more servers to the SAN, with each new server having shared access to stored data. For higher performance access to data, multiple copies of data can be created on a SAN, thus eliminating bottlenecks to a single disk.

8. Data Replication
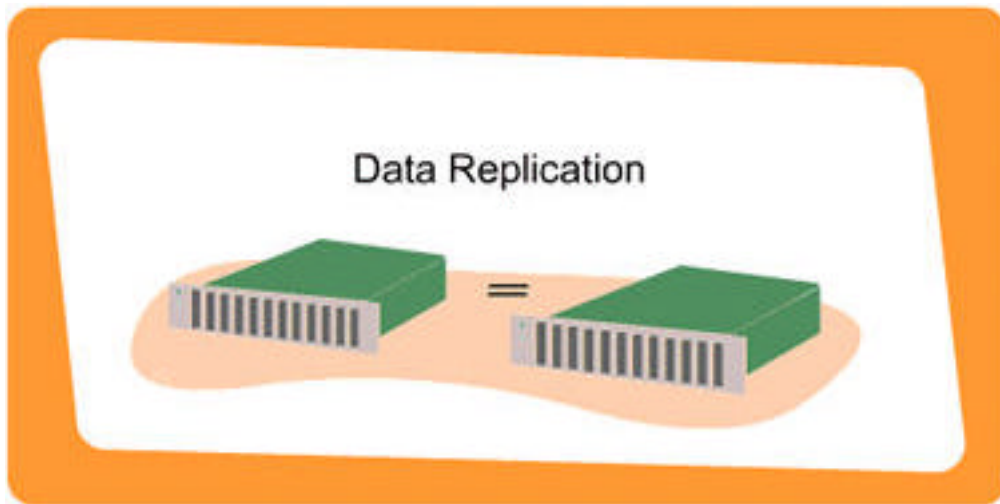
Data Replication provides many benefits in today's IT environments. For example, it can enable system administrators to create and manage multiple copies of business-critical information across a global enterprise. This can maximize business continuity, enabling diaster recovery solutions, Internet distribution of file server content, and improve host processing efficiency by moving data sets onto secondary servers for backup operations. These applications of data replication are frequently enhanced by the inherent "high data availability" features provided by today's SAN architectures.

Copying data from one server to another, or from one data storage system to one or more others, may be achieved in a number of ways. Traditionally, organizations have used tape-based technologies to distribute information. However, for many organizations that have built their business on an information infrastructure, the demand for instant access to information is increasing. While tape-based disaster recovery and content distribution solutions are robust, they do not support an instant information access model. Many organizations are supplementing, or replacing, their existing disaster recovery and content distribution solutions with online replication.

The replication of information is typically achieved with one of two basic strategies:

Storage replication is focused on the bulk data transfer of the files or blocks under an application from one server to one or more other servers. Storage replication is independent of the applications it is replicating, meaning that multiple applications can be running on a server while they are being replicated to a secondary server.

Application level replication is specific to an application such as a database or web server, and is typically done at transaction level (whether a table, row, or field) by the application itself. If multiple applications are being used on the same server, an application-specific replication solution must be used for each individually.

Replication types

Remote storage replication can be implemented at either the data storage array or host level. Array-based (or hardware) storage replication is typically homogeneous, meaning that data is copied from one disk array to another of the same make and model. A dedicated channel such as ESCON (Enterprise System Connection) is commonly required to link the two arrays, and can also require a number of other pieces of proprietary hardware such as storage bridges. Host level storage replication is implemented in software at the CPU level, and is independent of the disk array used. This replication is done using standard protocols such as TCP/IP across an existing network infrastructure such as ATM.

Wether implemented at the array or host level, storage replication works in one of two modes – synchronous and asynchronous. Whichever replication type is chosen has a number of implications:

Synchronous replication – In a synchronous replication environment, data must be written to the target before the write operation completes on the host system. This assures the highest possible level of data currency for the target system – at any point in time, it will have the exact same data as the source. However, synchronous replication can introduce performance delays on the source system, particularly if the network connection between the system is slow. Some solutions combine synchronous and asynchronous operations, switching to asynchronous replication dynamically when there are problems, then reverting to synchronous when the communication problems are resolved.

Asynchronous replication – Using asynchronous replication, the source system does not wait for a confirmation from the target systems before proceeding. Products may queue (or cache) data and send batches of changes between the systems during periods of network availability.

Uses of Replication

Storage replication serves many purposes, and the chosen solution depends upon an organization's replication needs. These typically fall into one of two generic stamps:

- Disaster Recovery/Off-host Processing
  Disaster recovery – maintaining an up-to-the-moment copy of critical data at an alternate site. These sites are generally geographically distant from one another and employ various types of WANs for connectivity.

  Off-host processing – movin critical data to a replicated server, from which to perform backups, reporting, testing, or, other operations that require production data but would affect the performance of the production system.

- Content Distribution
  Data sharing – giving multiple servers dynamic update capabilities to a shared pool of data.

  Data consolidation – copying data from distributed sites to a central site (such as regional sales servers to corporate headquarters).

  Follow-the-sun processing – maintaining local copies of critical data to support the global enterprise.

  Most solutions address one or more of these applications very well – few can support all easily.

9. Volume and File Management
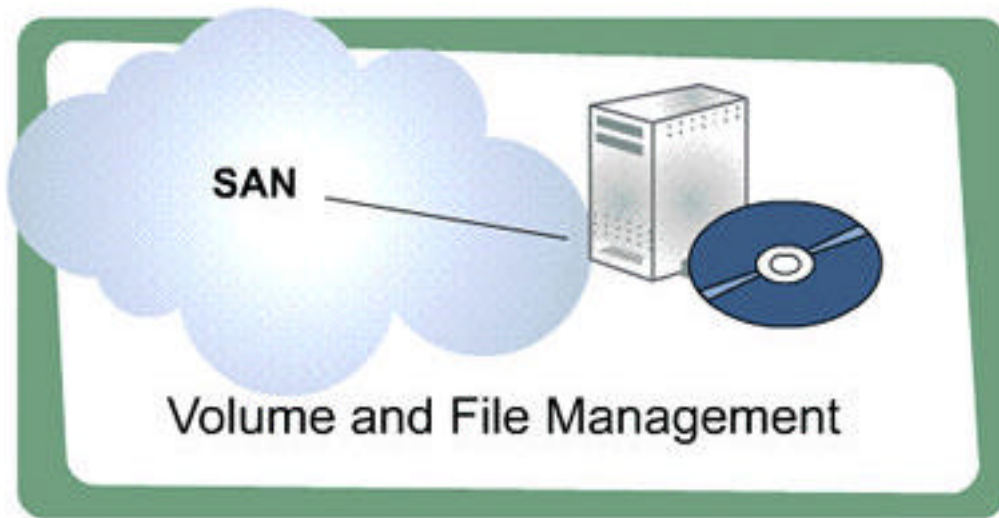
One of the many useful, and often touted, features of most SANs is their easy and inherent scalability. And while it is true that SAN technology makes adding physical devices fairly simple, it is important to understand that merely adding raw hardware resources to a SAN doesn't mean that its clients and their operating systems, or other SAN members, will automatically recognize them. This is particularly true when it comes to adding disk storage capacity.

Most SANs incorporate some type of data storage hardware. Generally, this equipment is in the form of RAID storage systems and JBODs. Adding additional, or higher capacity, disk drives to these RAID and JBOD systems in a SAN is a very common way to increase the SAN's raw storage capacity. And, while the RAID systems will often auto detect these additional disk drives, they won't typically expand the file systems that were created on them before the additional disk capacity was added. In fact, most of a SAN's client servers and their operating systems will not automatically do anything about the extra added capacity. The simple reason for this is that the "raw" storage capacity was increased by adding additional hardware (disk drives), the file systems that existed on the storage systems – before the additional disks were added – won't magically expand to include the new disk spce without some help from specialized volume and file management software. This is especially true when expanding JBODs.

Once raw capacity is increased, either via expanded RAID systems or JBODs, an administrator must manually use operating system utilities running on the SAN's client servers to either prepare and(or create new file systems on the added devices (JBOD) or additional LUN space (RAID). But a problem arises with many, if not most, unmodified operating systems when it is desired to "simply add or append" additional new capacity to pre-existing volumes or file systems. Generally, pre-existing file systems need to be erased and then re-created in order to incorporate new raw capacity.

However, software is available that can seamlessly add new raw storage capacity to both volumes and file systems. In fact, this software can even perform this expansion transparently and without the need to unmount files systems from their operating systems or shutting down servers.

Another inherent feature of many SANs is the ability to permit the sharing of storage capacity between multiple servers, even servers running different operating systems (heterogeneous server environments). This means that a single storage system's raw capacity can be divided among, and exclusively assigned to, different servers. But please note that this sharing relates to the raw storage capacity available on a storage system and not to any data contained on it. Here too, in order to share data on a shared storage system, special software is almost always required – particularly if the data is to be shared between servers running different operating systems.

Even in cases where storage systems are used without file systems, such as with certain database applications, it is still possible to add storage without quiesing or shutting down the servers. All that is needed is the proper management software for the appropriate applications.

Along with the features cited above, volume and file system management software can provide numerous other capabilities within a SAN environment. Some of these offerings include creating software RAID volumes on JBODs, changing RAID levels "on-the-fly", spanning disk drives or RAID systems to form larger contiguous logical volumes, file system journaling for higher efficiency and performance, and open file management.