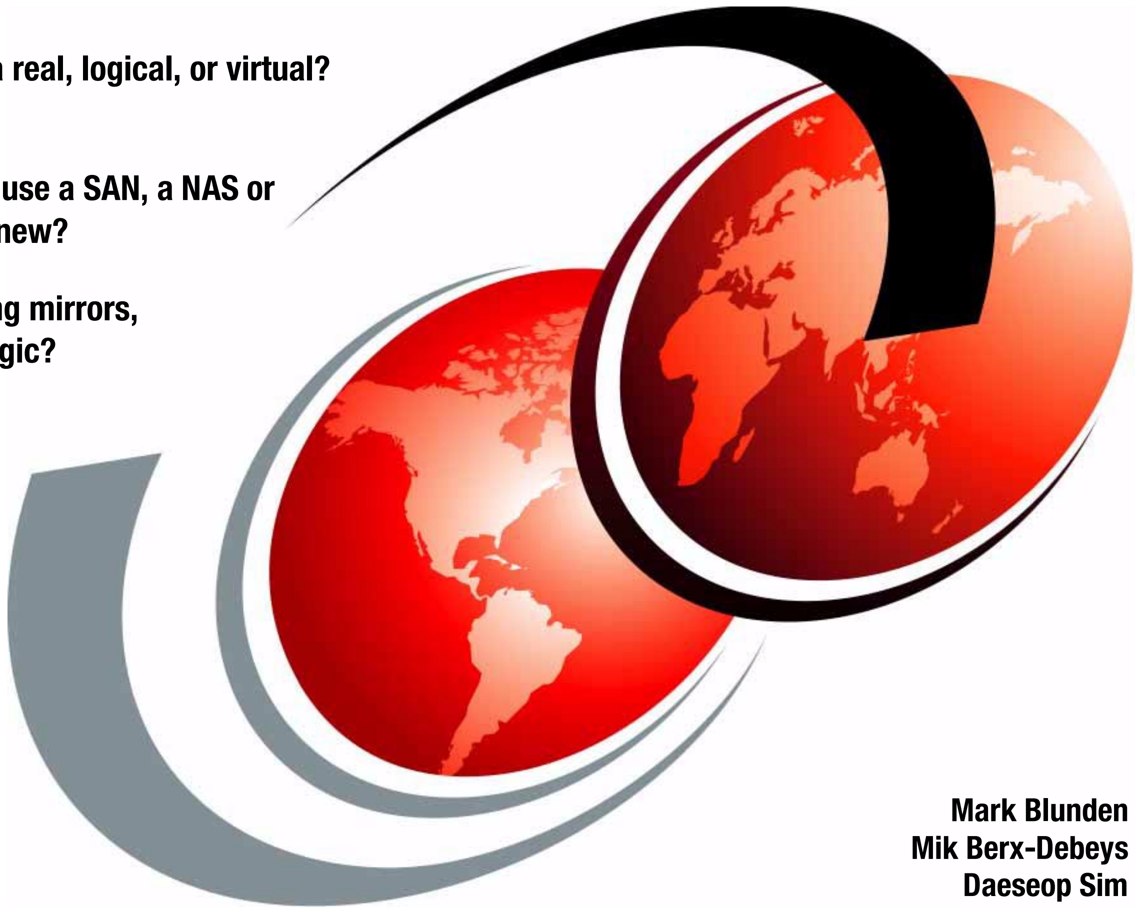
Storage Networking Virtualization

What's it all about?

Is your data real, logical, or virtual?

Should you use a SAN, a NAS or something new?

Are we using mirrors, myth or magic?



Mark Blunden
Mik Berx-Debeys
Daeseop Sim

Redbooks

ibm.com/redbooks

SG24-6210-00

International Technical Support Organization

Storage Networking Virtualization
What's it all about?

December 2000

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix A, “Special notices” on page 95.

First Edition (December 2000)

This edition applies to the concept of storage networking virtualization as it is defined at publishing time.

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. 471 Building 80-E2
650 Harry Road
San Jose, California 95120-6099

Contents

Figures

Tables

Preface

This IBM Redbook gives a broad understanding of a new architecture for storage networks, using virtual mapping of data files onto a network of storage devices. The concept and reality of virtual data has been around for many years in the mainframe environment, but now we are entering a new era of virtualization with the design, architecture, and implementation of a new set of virtual products to help all installations make efficient use of IT storage resources. This new approach will revolutionize the way data is stored and managed in the UNIX/NT environment, similar to the way that systems managed storage revolutionized storage management for the mainframe.

These emerging products integrate into existing configurations to raise efficiency and utilization.

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.

Mark Blunden is the project leader for Storage Systems Disk at the International Technical Support Organization, San Jose Center. He has co-authored nine previous redbooks and comanaged six other redbooks. Mark also teaches IBM classes worldwide on all areas of storage. He has worked for IBM for 20 years in many areas of the IT business. Before joining the ITSO in 1998, Mark worked in Sydney, Australia, as an Advisory Storage Specialist.

Mik Berx-Debeys is an IT Specialist in Belgium. She holds a degree in Electro-Mechanical Engineering from the K.U.Leuven, but has been working as Systems Engineer and Sales Specialist for IBM for 25 years. Her areas of expertise include disk, tape and optical products, and the software to manage them. She teaches on storage products and presently works as Technical Sales Support for heterogeneous customer accounts.

Daeseop Sim is a System Engineer in Korea. He joined IBM in 1996. Daeseop’s areas of expertise include open storage systems, AIX, SP, HACMP and ADSM. He has a bachelor’s degree in Computer Engineering.

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

David Pease, IBM Almaden Research Center

Marc Roskow, IBM SSG

Ron Riffe, Tivoli Systems

Pratt Parrish, IBM SSG

Randy Kearnes, Evaluator Group

Dean Underwood, IBM SSG

Phil Ricketts, IBM Dallas

Comments welcome

Your comments are important to us!

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

- Fax the evaluation form found in "IBM Redbooks review" on page 111 to the fax number shown on the form.
- Use the on-line evaluation form found at **ibm.com**/redbooks
- Send your comments in an Internet note to redbook@us.ibm.com

Figure 1. Daeseop, Mark, Mik—the authors of this redbook

Part 1. Virtualization principles

This part describes the concept of data virtualization as it applies in a Storage Networking environment. Storage Networking is a term used to cover both Storage Area Networks (SANs), and Network Attached Storage (NAS) devices. The following chapters provide an introduction to storage networking virtualization, and the benefits, services, and functionality it provides. Part 2 of this book talks more specifically about some product offerings.

Chapter 1. Introduction

In this chapter we introduce the concept of storage virtualization, and the different levels of abstraction that can be made to make storage management more efficient and transparent to the users. Virtualization is separation of the logical view of storage from its physical implementation enabling physical changes to be made without changing the logical view.

1.1 The need for virtualization

Many authors these days state that storage has become a commodity over the past few years. This implies that people want to be able to just use storage without limitations or worries, to completely disregard its whereabouts and management, yet always be sure of its abundance and availability. At the same time that storage cost has been steadily going down, because of technology evolution and new ways of connecting storage devices have been evaluated and implemented, the volume of data storage required in daily life and business has exploded. The capacity is growing by 60% and hardware cost is decreasing by 30% per year, but availability requirements are approaching 100%. Three hundred million Internet users are driving two petabytes of data traffic per month. Users are mobile, access patterns unpredictable, the content of data becomes more and more interactive, so a new level of bandwidth and load balancing are required.

Managing these high volumes of critical data has become impossible as a manual task, and automatization of the processes has been one of the major efforts in the industry in the last decades. Especially in the Open Systems environment (UNIX and NT), this has introduced new policies which used to be the discipline in the mainframe glasshouses before.

Sharing of data between heterogeneous platforms also has become a prime concern, since the communication networks cannot cope with the massive replications, downloads and copying that are required in today’s industry.

The development of Storage Area Networks (SANs) has passed its experimental phase and SAN has become a mature technology that allows businesses to implement storage pooling and sharing today. SANs are high speed switched networks designed to allow multiple computers to have shared access to many storage devices.

Yet, implementing a SAN does not in itself solve the basic problems of meeting the requirements of the ever growing explosion of data.

Sophisticated software architectures will be needed to handle them and take advantage of the hardware.

Similar to the introduction of virtual memory in computer science that has solved all the problems of managing a limited resource in application programs, the virtualization of storage can give users a more intuitive usage of storage, while it is being taken care of transparently by the software that manages the storage network.

1.2 A definition of virtualization

A few years ago, “virtual storage” was a synonym used for the virtual memory techniques used in operating systems. Consultants and industry have begun to use the term storage virtualization to describe the abstraction from physical volumes of data storage to a logical level recently, and the new buzzword has been rapidly accepted by the IT profession. Even though it is a word that does not exist in any English dictionary, we will continue to use it since the industry has embraced the term. In proper English it could be defined as vulgarization of storage, meaning wide diffusion and popularization of storage.

This abstraction can be made on several levels of the components of storage networks and is not limited to the mere disk subsystem. Virtualization separates the representation of storage to the operating system and its users from the actual physical components. This has been represented in the mainframe environment for many years by methodologies such as system-managed storage and products like DFSMS.

1.2.1 Levels of virtualization

We will define the different levels at which virtualization can be achieved in a storage network, as illustrated in Figure 2.

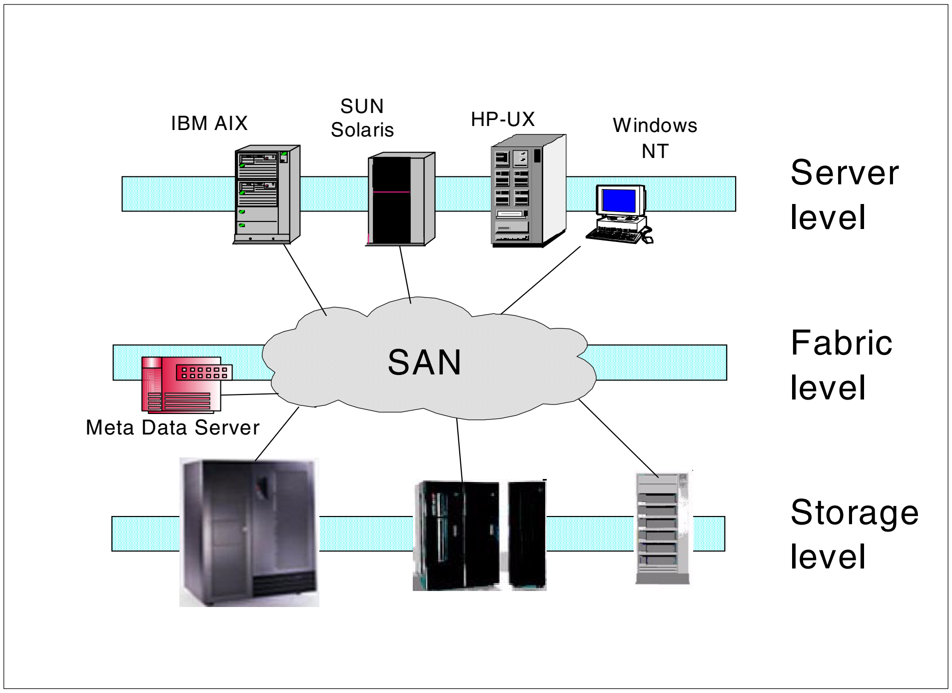


Figure 2. Levels at which virtualization can be applied

Server level

Abstraction at the server level is by means of the volume management of the operating systems on the servers. Some operating systems, like OS/390 with DFSMS, have a high level of abstraction, whereas others still present to the user volumes that are directly related to physical disks. At first sight, increasing the level of abstraction on the server seems well suited for environments without storage networks, but this can be vitally important in storage networks too, if we look at the security and sharing issues.

Fabric level

On the SAN Fabric level, virtualization can enable the independence of storage pools from heterogeneous servers. A single management interface can be implemented to manage disparate storage systems without impacting the servers. Several approaches are possible, and we will go into more detail on this.

Storage level

At the storage device level, several virtualization techniques have already been tried, ranging from striping and rudundant array of independent disks

(RAID) to sophisticated volume management in the Enterprise Storage Server (ESS) or Log Structured Arrays (LSAs), such as the RAMAC Virtual Array (RVA). The Virtual Tape Server (VTS) is another example of virtualization at the device level. Future developments can implement virtualization at this level to enhance functions like cloning or snapshots and allow storage usage beyond the physically installed capacity.

We mainly address storage virtualization on SANs in this book, but the Network Attached Storage (NAS) products can play a role in this arena as well. They provide easy to manage storage attached directly on the communications network, and basically act like more or less sophisticated file servers. Actually, they are virtual disk servers on a separate network, the Internet/intranet or the LAN/WAN.

File system level

The highest level of abstraction, to the file system level, will result in the greatest benefit, since what we want to share, allocate and protect are not volumes, but data sets or files. Therefore, a model that can make abstraction to the file level or even to a higher granularity, like to a block, will yield the best results. This may seem to go against the trend to consolidate storage at the device level, but it introduces the global data space across heterogeneous networked storage devices. Both SAN and NAS storage products can be part of this space.

Scalable, managed distributed file systems must be developed, without performance penalty from client/server interactions and with integrated security, data integrity and continuous availability.

1.3 The purpose of virtualization

Storage network virtualization addresses the increasing complexity of managing storage and will reduce the associated costs dramatically.

Its main purpose is the full exploitation of the benefits promised by a SAN, like data sharing, higher availability, disaster tolerance, improved performance, consolidation of resources, and many more, which do not automatically result from the SAN hardware components and which were so far only partially addressed by the available software that are often limited to a subset of supported platforms.

The objectives are:

- To remove the management of the storage devices from the server platforms

- To allow management by high level attributes rather than blunt physical characteristics
- Allocation and management in accordance to the Quality of Service (QoS) associated with the data
- To effectively pool or group heterogeneous storage devices as a single resource
- To adapt the size of the pools to varying capacity requirements
- To migrate data within the pools to match its profile in performance, availability, frequency of access, and so on, as pools expand or retract
- To allow for load balancing and avoid introducing hot spots in the data path
- To efficiently assure the availability of data by using mirrors or snapshots to speed up recovery and backup
- To present one single management interface to the administrator(s) which is easy to learn and use
- To be completely transparent to various applications and heterogeneous file systems
- To provide continuous availability of the virtualization system by failover
- To coexist with non virtualized storage on the same SAN and evolve with the introduction of new technology

By realizing these objectives, storage virtualization can reduce costs associated with storage management:

- Better usage of the available storage capacity in a true open SAN environment
- Better usage and development of the skills available to manage the SAN allowing administrators to manage larger amounts of storage
- Reduced human error by automatization
- Enforcement of rules concerning storage usage

Translated to business value, this signifies:

- Ability to anticipate trends
- Get to market faster
- Take advantage of new opportunities
- Minimize business risks and exposures
- Gain competitive advantage

- Exploit technology change
- Handle growth
- Reduce cost and improve profit

1.4 Virtualization models

Symmetric and asymmetric models can be drawn for storage virtualization, as illustrated in Figure 3.

File systems can be distributed in either way. When you look at Figure 3, you may instinctively conclude that asymmetrical is better than symmetrical. Nothing is less true. Both have advantages which we will illustrate in this section.

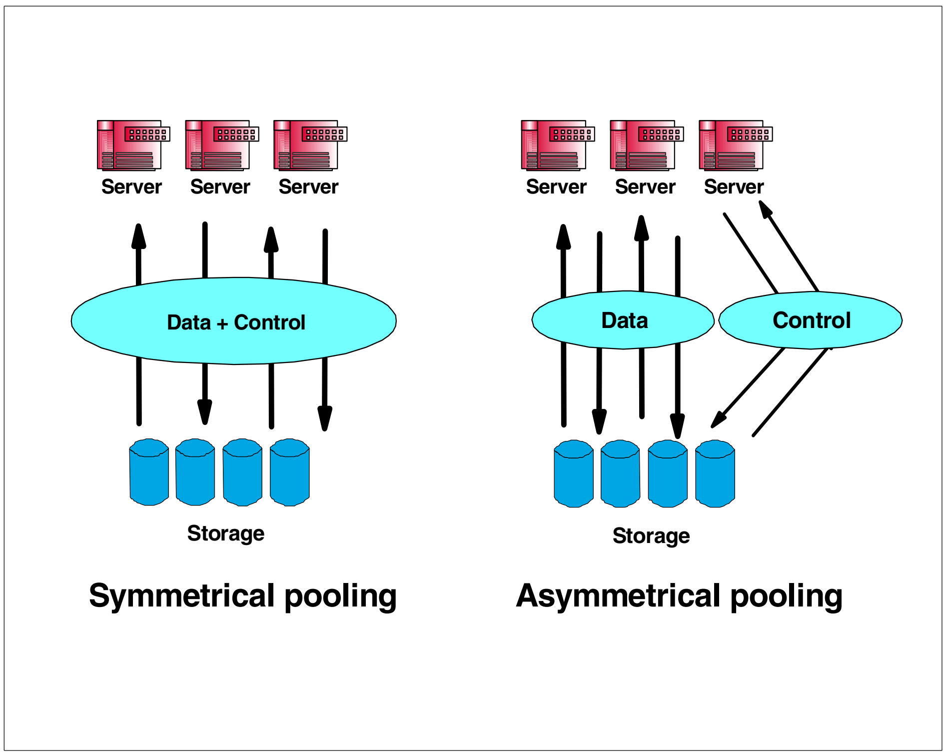


Figure 3. Symmetrical compared to asymmetrical pooling

1.4.1 Terminology

Before we continue with this discussion, let us clarify some terms. Since we are used to the terms *client* and *server* in application development, things can get very confusing if we use them in another context. The notions of client, server, controller, agent and so on are usually associated with a certain hierarchy of service execution. Depending on the paper you read, you may see different terms or acronyms describing the same things. Some people call it a *metadata server*, while others mean a similar thing when they write *storage domain server* or *storage domain controller*.

In this book we use the following terms:

host — Any server that is using storage and runs applications for its clients.

agent — A hardware or software component installed on a host to virtualize its view of storage or to execute a specific storage management function for that particular host.

server — A component of the storage network that provides a service to its clients. This can be a host server or a metadata server.

client — Something requesting a service from a server. This can be a client to an application on a host or a host when requesting service from a virtualization server.

controller — A hardware or software component that controls access and usage of storage resources it manages.

Now to save you from total confusion, let us agree that client and server are interchangeable terms and that the relationship is defined by the interaction at a particular time. We will use illustrations to clarify these interactions and qualify the terms, where appropriate.

1.4.2 Symmetrical virtualization

When we implement a virtual storage network in a symmetrical way, both data and control flow over the same path. Levels of abstraction exist in the data path, and storage can be pooled under the control of a domain manager. Logical to physical mapping is done by the domain manager or metadata controller. Some authors will tell you that the only advantage of this approach is that it is simple, but the disadvantages lie in the single points of failure or performance bottlenecks that it can introduce. Also, it may be more difficult to move data across pools if it is controlled by different managers.

On the contrary, locking and data sharing integrity can be implemented in a simple fashion, using zoning or masking techniques that are available today.

Symmetrical virtualization allows you to build a kind of fire wall in the storage network. Also the metadata controllers or domain managers can implement security, caching, snapshots, remote copy and many other functions, since they directly control access to the storage and do not only grant access to it. Figure 4 illustrates these characteristics of symmetrical virtualization.

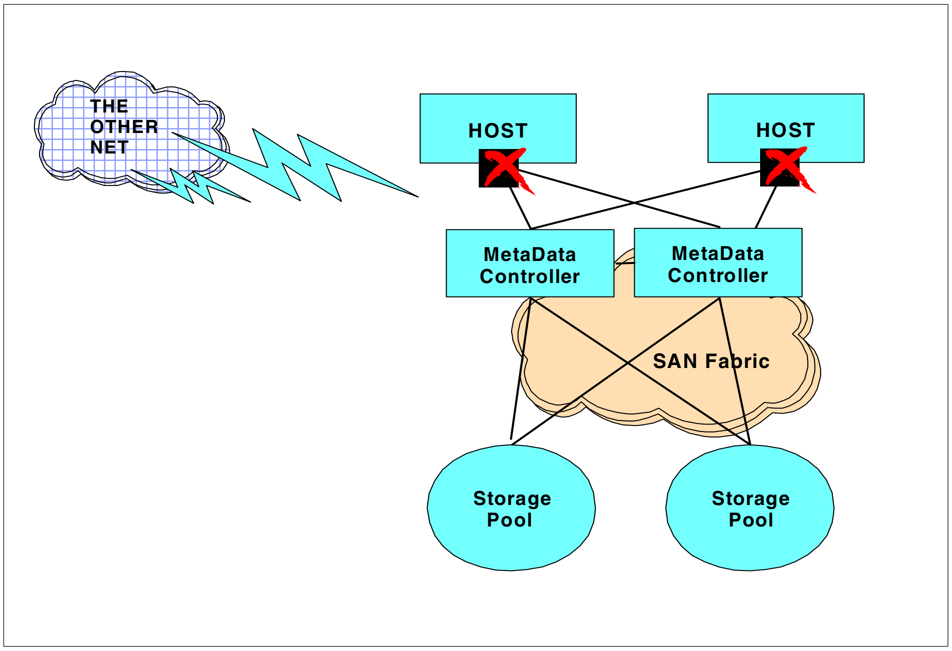


Figure 4. Symmetrical virtualization

When it comes to performance, it is all a matter of scalability. Who said there should be only one single metadata controller (MDC) in the SAN clustered with a failover? For all we know, there could be any number of cooperating domain controllers, in an n-way cluster. Adding processing power, cache memory and adapter bandwidth can be scaled to achieve the performance level required. And, this cache and processing power can be used to implement functions like snapshots and caching on devices like “just a bunch of disks” (JBODs) or RAID arrays that do not have these functions.

As the cross marks in Figure 4 indicate, if the host bus adapter (HBA) is sufficiently fenced, a firewall against unsolicited access is easily built at this level.

1.4.3 Asymmetrical virtualization

In asymmetrical virtual storage networks, the data flow is separated from the control flow. This is achieved by storing data only on the storage devices, and moving all mapping and locking tables to a separate server that contains the metadata of the files. This approach is certainly more complex to implement, but has its strong points as well. Figure 5 illustrates asymmetrical virtualization in a different way.

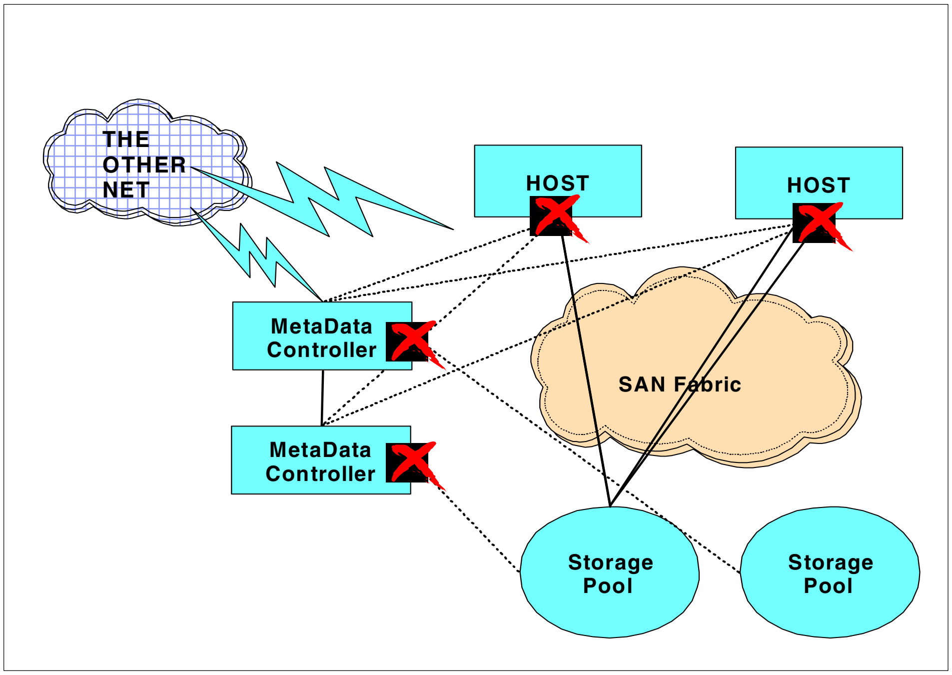


Figure 5. Asymmetrical virtualization

Separating the flow of control and data allows the IO to use the full bandwidth of the SAN, while control could go over a separate network or over circuits in the SAN that are isolated for this purpose. For many operations, the metadata controller does not even intervene. Once a client has obtained access to a file, all IO will go directly over the SAN to the storage devices. However, note the increase of security exposures this introduces — a separate control network should also be firewalled.

1.4.4 Combinations and variations on these themes

We can take this one step further and separate the SAN into two separate networks, an upper one to connect the hosts to the domain controllers and a lower one to connect the domain controllers to the storage devices, as shown in Figure 6.

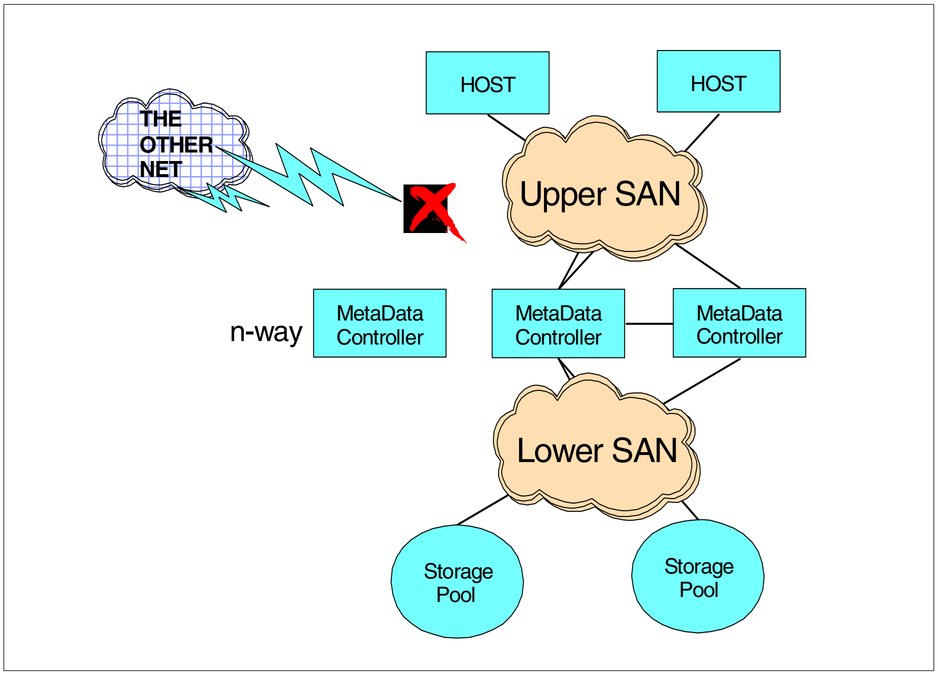


Figure 6. Two-layered storage network

In this way, we can scale performance, bandwidth, processing power, caching in an n-way cluster of domain controllers, who do virtualization both at the fabric and at the storage level, and the security firewall can be set up in a symmetrical way.

Taking another view, why would we not combine asymmetrical and symmetrical virtualization? Not everyone uses remote copy, FlashCopy, snapshots, or caching on every piece of storage equipment they bought. After all, virtualization is trying to set up different QoS levels on heterogeneous storage devices. So we can draw Figure 7 with a combination of symmetrical virtualization through a domain server that has high performance, high availability functions for one storage group, and through an asymmetrical virtualization engine for basic QoS management. If functionality is needed, both could communicate extent lists to invoke each others functions for the files they manage.

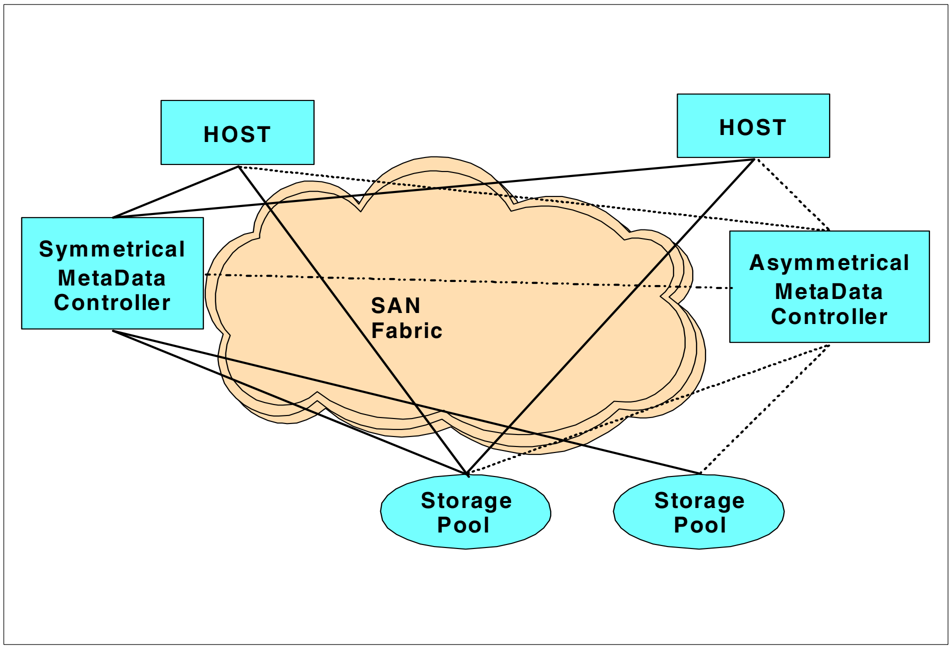


Figure 7. Combining both approaches

We will come back to these possible combinations and variations after an overview of the technology that will become available soon.

1.4.5 IBM SAN network software model

Whether symmetrical or asymmetrical virtualization is chosen, software is the prime element. IBM/Tivoli's storage network software model is illustrated in Figure 8 as a five layer model encompassing all aspects of systems management.

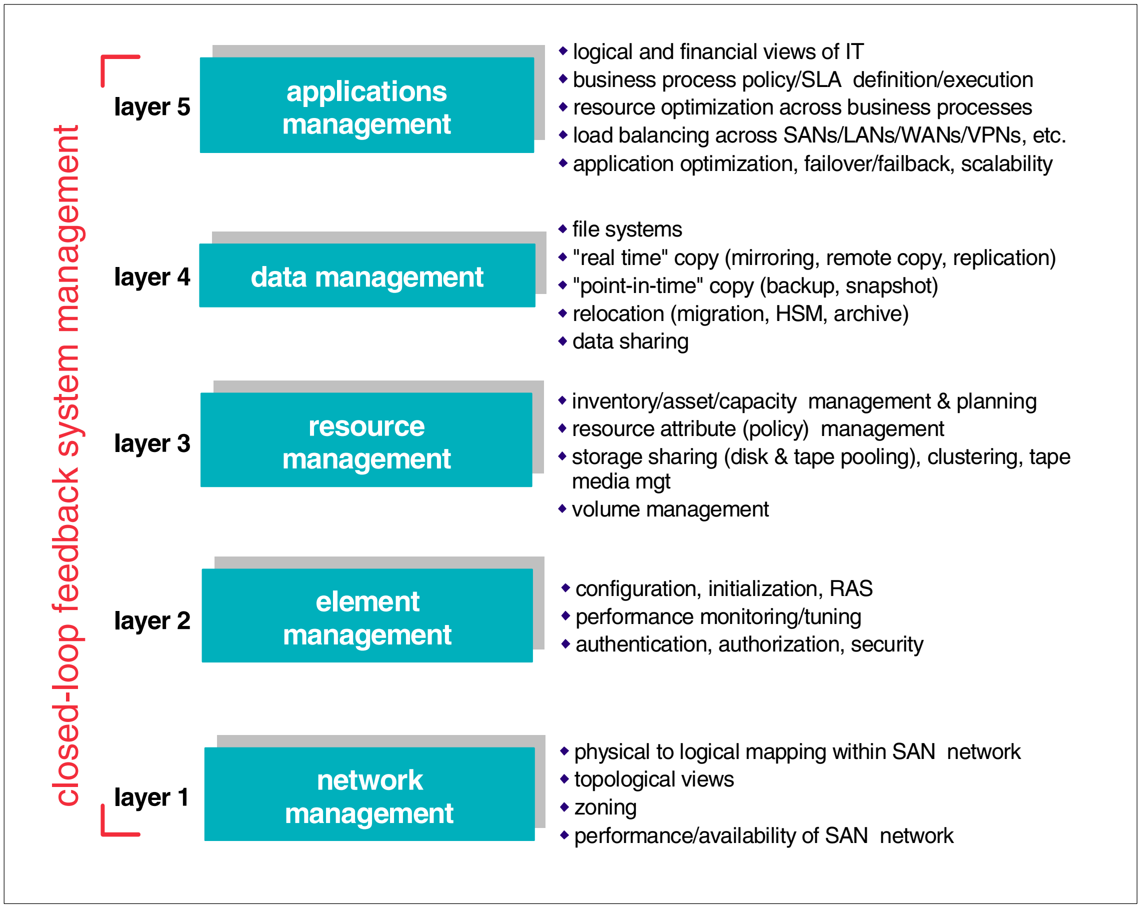


Figure 8. IBM/Tivoli SAN network software model

These functional layers allow us to implement a storage software system structure that isolates the application from the underlying storage hardware, and it can be used for file type applications as well as for database applications, as is illustrated in Figure 9.

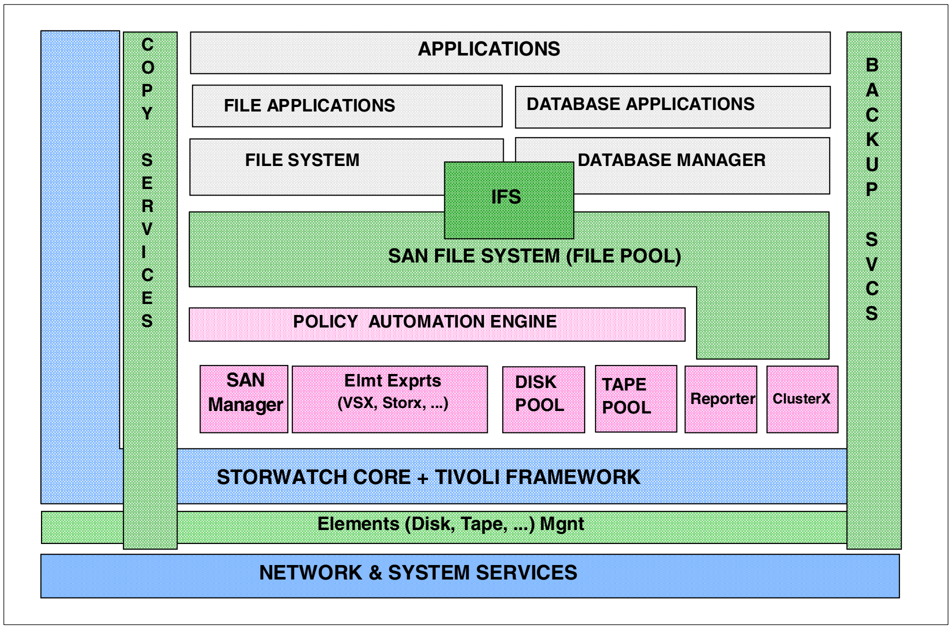


Figure 9. SAN software system structure

Note that in this model, both file system and raw device level access should be possible. This means virtualization to a higher level of granularity for some applications, whereas others continue to use volumes, even if they are managed by the virtualization software. The applications and the operating systems on the host servers are isolated from direct volume management by an Installable File System (IFS) which can either supplement or completely replace the volume management of the host server (UNIX, Linux, NT).

1.5 Current status

At this moment, several virtualization systems have been designed at various levels (server, fabric and storage). Most have been described in white papers and research reports, and some are close to being turned into new products or new releases.

In Chapter 6, “Tivoli Open System-managed Storage” on page 71 and Chapter 8, “Tivoli Storage Network Manager” on page 85, the IBM/Tivoli products for SAN virtualization will be covered in more detail. In this introduction we want to briefly cover these as illustrations of different levels of abstraction that are under development today.

Under the IBM-Compaq agreement, IBM has announced that it will endorse the VersaStor technology originally developed by Compaq. This will be described in more detail in Chapter 7, “VersaStor” on page 81.

To position the different products, we have tried to put some of these in the illustration in Figure 10.

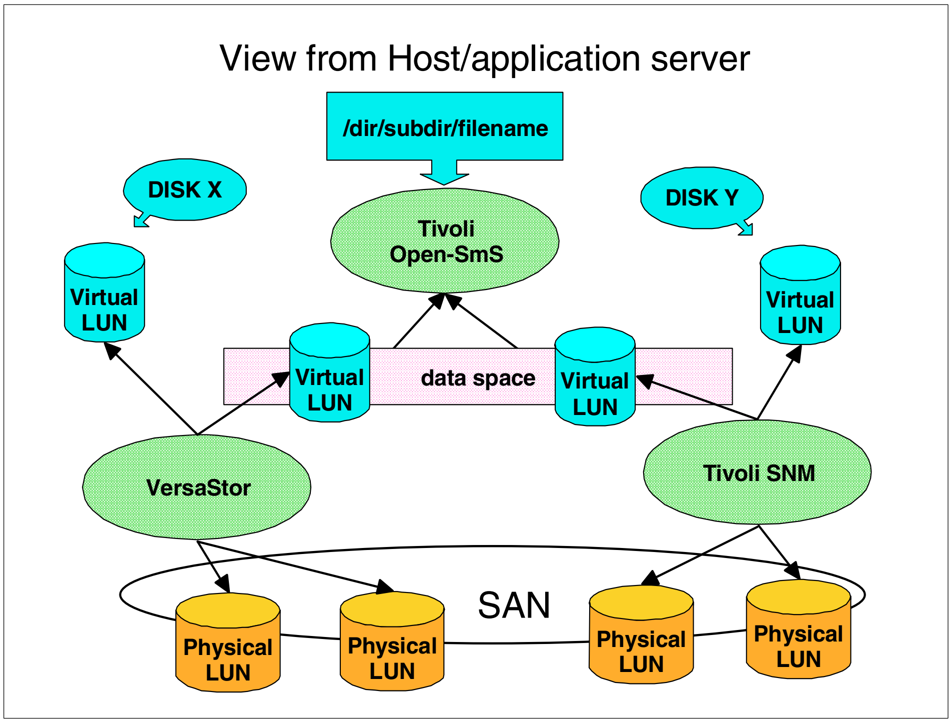


Figure 10. Positioning of architectures and products

We should discern between virtualization models that come in at the highest level, the file system, and virtualization appliances that only present volumes to the operating system, which can then use these in its volume management.

Tivoli Open System-managed Storage (SmS)

Tivoli Open SmS is a new storage management model that implements virtualization in a heterogeneous SAN environment, independent of hardware. It comes in at the highest level and sets up one single distributed file system that can be shared and used by all the servers that plug in to the storage utility. Tivoli Open SmS sets up storage groups spanning pools that have been defined across storage devices on the SAN.

Because it is hardware independent, it could use any storage from JBODs, RAID disks to ESS or MSS-like subsystems, and even virtual volumes that are managed by another virtualization technology, such as VersaStor, or a SAN disk manager, such as Tivoli Storage Network Manager.

The hosts see storage groups that can be segmented according to policies for QoS. One example could be: high, medium and low performance or availability storage groups. Or in another installation: project A, project B, project C could each have their allocations in subsets of the storage groups. Who actually can access a particular file in the shared groups is a matter of security, and Tivoli Open SmS manages this via a POSIX compliant access control (Figure 11).

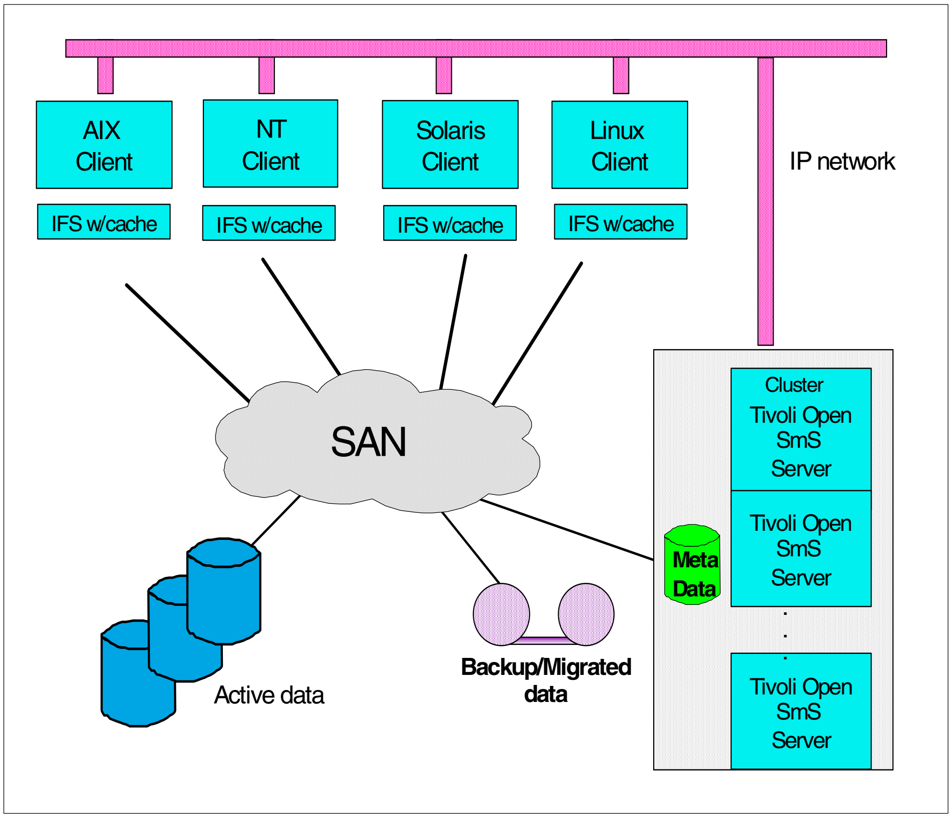


Figure 11. Tivoli Open SmS

VersaStor

VersaStor is a storage virtualization technology that uses both software and specific hardware components and virtualizes volumes at the SAN fabric level.

Between the hosts at the top and the storage devices at the bottom, it creates a set of logical volumes out of the storage volumes it finds at the bottom, and manages these via an asymmetric control unit.

VersaStor separates the traditional control unit function from the storage device and sets it on a separate control unit on the SAN, therefore allowing virtualization across theoretically any kind of disks.

It hands the volumes it defines over the SAN to the attached host servers, who then manage it as a volume in their volume management, but cannot share it with other host servers, as such. VersaStor does not handle SAN topology or resource management. It does implement virtual volumes and it isolates the host servers not only by the software, but also by using a specific VersaStor enabled host bus adapter.

Tivoli Storage Network Manager

This SAN software suite encompasses the products previously called:

- SAN Manager
- Disk Manager
- Storage Automation

Tivoli Network Storage Manager provides the vital SAN management that virtualization systems rely on. It leverages Tivoli's Common Services and uses presentation services based on a Netview interface. The administrators can launch the different element managers from the common console to manage SAN fabric elements, such as switches or hubs and storage elements like disks, tapes, optical and robotics.

SAN topology discovery and monitoring are provided both in-band and out-band to assist in SAN infrastructure maintenance and improve availability for the applications.

Disk resources can be identified and allocated across the SAN in a secure and simple way on heterogeneous disk subsystems.

Information on the elements is stored in a database which can be used in data mining by the Tivoli Storage Resource Manager for trend analysis.

1.6 Network Attached Storage (NAS)

NAS devices do file virtualization on the communications network. As illustrated in Figure 12, they come in two flavors:

- NAS Appliances who contain their own disk storage, protected by RAID
- NAS Gateways to other disk subsystems, which can be on a SAN

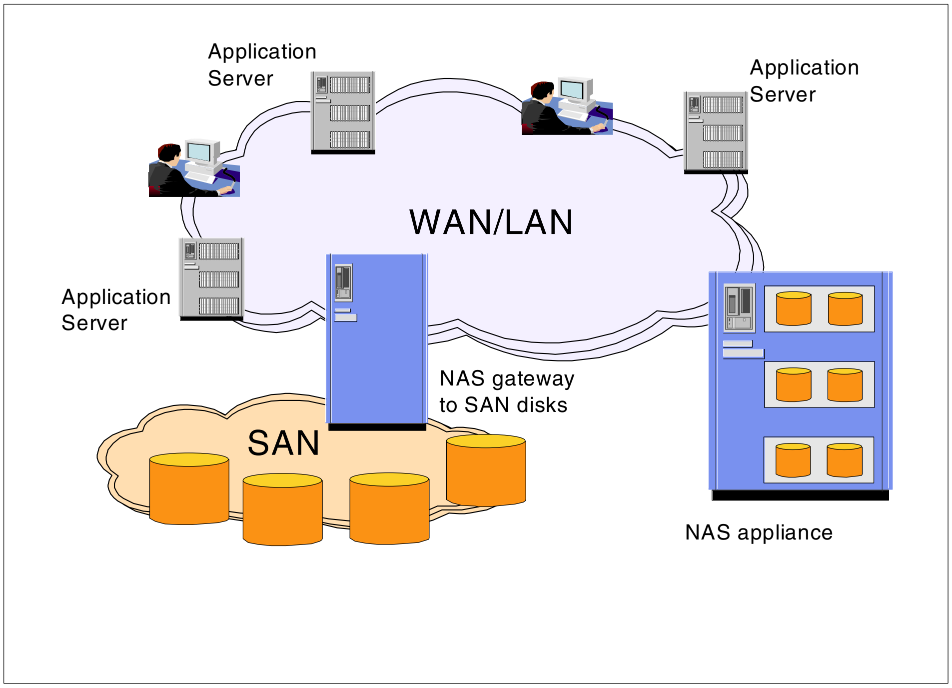


Figure 12. NAS storage devices

A NAS device is a virtual file server, and supports NFS-like protocols. To position NAS versus SAN devices, we consider several arguments, as listed in Table 1.

Table 1. NAS, SAN, and direct attached storage

Direct attach	SAN	NAS
Low configuration flexibility	Highly scalable	Low scalability
High administration costs	High IT skills needed	Lower IT skills needed
No data sharing	Limited data sharing unless virtualized	File sharing

Direct attach	SAN	NAS
	High performance	Limited performance
	Scales UP	Scales DOWN
	Lengthy, expensive initial deployment	Quick, inexpensive initial deployment
	Limited reach	WAN reach

NAS and SAN use different protocols, but they will probably converge one day. The gateways are a first step. For many installations, NAS will be a first step in storage networking.

You may not want to deploy a SAN in each location, so NAS could be a complementary solution for remote locations, even if that creates new islands of information. Most virtualization products promise a future integration of NAS.

1.7 iSCSI - SCSI over TCP/IP networks

Since the bandwidths of both SANs and Ethernet networks have gone over the gigabit limit, many people are beginning to consider using cheaper Ethernet interfaces rather than the expensive Fibre Channel interfaces. Another argument is that every server will need an Ethernet (or other TCP/IP) adapter anyway, so why not use it for other purposes? And, skills to manage an IP network are more readily found than skills to design and implement a SAN.

This could provide a new way to bridge distances by using the communications network for storage networking.

However, communications networking is based on design considerations different from those of storage networking. Is it possible to merge the two and yet provide the performance of a specialized storage protocol? The challenges are many. The TCP/IP protocol is software-based and geared towards unsolicited packets. On the other hand, storage protocols are hardware-based and are based on solicited packets. Storage wants acknowledgment of all commands or data sent — communications will simply resend if no answer comes back. Data integrity must be handled with care.

A storage networking protocol would have to interface with TCP/IP without change (it's a standard) and yet achieve high performance. Storage block IO could be slowed down by sending it over a different type of network. And, is it really a good idea to mix traffic on a single network? One of the first aims of SAN is to free up LAN bandwidth, remember?

On the other hand, much of the IP functionality already has moved to the adapter level, which frees up processor time on the servers. And the IP network could be split in separate communications and storage IP networks, or different priorities and class of service for both types of traffic can be used within a single network.

Figure 13 illustrates the possibilities of iSCSI, by itself, or integrated with SAN and NAS products.

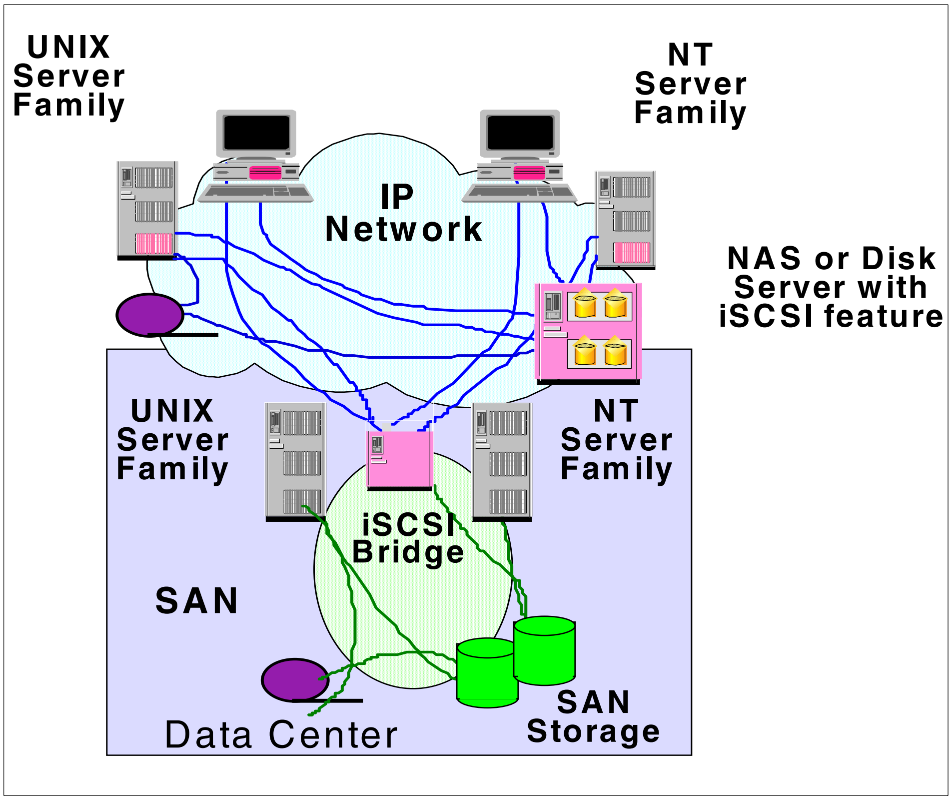


Figure 13. iSCSI in storage networking

IBM is investing in research on iSCSI to address these challenges and make it a reality, and to develop products (iSCSI interfaces on storage devices, device drivers) and management software for this new environment.

1.8 The future

When deciding on a virtualization approach, you should consider the ability of the model to evolve with new technology and the associated appliances and elements, whether it's hardware or software. If you choose for a model that will lock you into a platform limitation or even a single vendor, then going to another virtualization approach in the future may prove difficult.

If you choose a model that does not put limitations on the hardware and can use other software as an engine for certain functionality, like specific device configuration management, then that will set you on a path that allows you to integrate future technologies.

Like in basic SAN component decisions, openness and standardization are key factors. Ultimately what we really want is a universal file system. We list the standards in Chapter 5, "Standards" on page 67.

1.9 Benefits

In this section we point out some of the benefits of virtualization.

Reducing cost

Over the last several years, it has been widely documented and understood that the cost of managing storage will easily outweigh the cost of purchasing the storage hardware.

Saving time

As corporate data assets continue to grow equating to an ever-increasing necessity for data storage resources, the time demands for planning, upkeep and maintenance by the IT staff are juggled and stretched to the breaking point. One of the most time-consuming operations in storage is managing the necessary addition of new storage hardware, the transfer of information from old to new, and the removal of old or obsolete hardware.

Managing your data storage resources

Most businesses are concerned with the efficient utilization of their data storage resources. The entire Storage Resource Management industry exists to enable IT organizations with answering usage questions.

How much storage do I own? How much has been assigned to hosts? Of that, how much has been formatted by a System Administrator? After formatting, how much usable space do I have left? How much of the formatted space has been given to a specific file system. How much of that space are the individual file systems actually using? Where am I running low? Where am I over allocated? The questions that IT organizations are asking point out that in open system environments, the basic model of a host system owning storage leads to fragmentation, inefficient usage and difficulty in tracking. The problem is that even when the questions are answered, technology is not readily available to solve the problems.

In sharp contrast, the complete virtualization of disk storage into QoS pools of capacity provided by System-managed Storage eliminates the kind of fragmentation and inefficient usage found in open system environments. All participating host systems share the capacity. Capacity is only used as new data assets are created. Out-of-space conditions are virtually eliminated through the use of intelligent overflow capacity and the implementation of capacity-on-demand.

System-managed Storage utilizes data storage resources much more efficiently on OS/390 when compared to current open system volume managers and file systems. Storage virtualization changes this by introducing to open system environments the validated data storage techniques that have been available for years in OS/390 environments.

1.10 Deployment

In this section we will highlight some considerations related to the implementation of virtualization.

1.10.1 Time frame

Clearly, many virtualization technologies are in the phase of moving from research into development. True virtualization products can be expected to release a first version in the year 2001.

This should not stop you from moving in the direction of virtualization, since the most important job is to clearly define your company's policies on data management, and bring law and order in your SAN.

1.10.2 Configuration considerations

Let us briefly discuss some possible implications.

1.10.2.1 HBAs

A virtualization system may depend on specific hardware on the servers that plug in to it. This is true for VersaStor, which will rely on specific HBAs to enable it. Other systems, like Tivoli Open SmS, can be implemented without any specific hardware requirements.

1.10.2.2 File systems

The ultimate purpose of virtualization is to provide true universal data sharing. This may not be possible today across all existing platforms. The file systems in heterogeneous systems differ. Even when servers address NAS, which acts as a network file server, the protocols used can vary: Windows NT uses Common Internet File System (CIFS) and UNIX uses Network File System (NFS). The access control and locking that the metadata server needs to provide is based on Access Control Lists (ACLs), which are different between NT and UNIX.

Using an Installable File System (IFS) to mask these differences in operating systems, is a very open approach to solving this.

1.10.2.3 Availability

One of the major concerns is to implement a system that truly enhances the availability of data. At a workgroup or department level, this can mean 99.9% availability (meaning an average downtime of 8.45 hours per year). At the enterprise scale, we need systems that have a 99.999% availability with less than five minutes planned or unplanned outage. Designing the storage network for either situation will have different implications and cost.

Failover

Any new server or appliance that is introduced in the storage network to implement virtualization, manages metadata, pools or volumes should never become a single point of failure. Dual active servers clustering or other rapid failover is a prime requirement.

Scalability

Adding storage capacity or reconfiguring it must never imply an outage of the storage already in use. The virtualization system should be designed for easy growth, introduction of new technology and easy migration.

Dynamic growth and reduction

As the requirements change, the system should adapt itself, by moving data transparently from one device or pool to another. Storage groups are able to expand and shrink as devices are added or removed. Cloning and snapshots can be used to assist in data movement across the pools. If these functions are not available on some of the storage systems in the configuration, then the virtualization software can initiate and execute such copies by itself. Isolating the servers and applications from the management of the storage allows the virtualization system to change allocations and adapt the metadata without impact on the users.

Data movement

The virtualization system should be able to move data across the different pools and associated devices, for performance reasons or for migration purposes.

1.10.2.4 Performance

Managing the location of the physical data is very important in a virtual file system. The software must be able to decide to move data to a different location if a bottleneck comes up, or even make multiple synchronous copies of data to distribute access paths. This requires sophisticated performance monitoring and automatic tuning actions. Since the purpose of virtualization is also to give administrators an easier job, tools should be available, allowing them to visualize and monitor the overall file system, yet at the same time some artificial intelligence in the virtualization system could automatically perform the necessary actions to solve performance problems.

1.10.2.5 Security

Access to data should be protected by a security system that controls access at a file level. This requires serious consideration, since the servers will be heterogeneous and different security Access Control Lists exist for UNIX and NT, for instance. The virtualization should include global security independent of the clients' operating systems.

1.11 Summary

The nature of the data we put on computers these days has changed. Computer systems do not just run your payroll, keep your inventory or give you a set of graphs of quarterly results. They drive your results, and your payroll, because they are your shop window and your sales desks, in today's businesses. People become knowledge workers, storing data is storing the intellectual capital of your company. Computer storage also contains much of your private data, none of which you can afford to lose.

Sharing data and making more efficient usage of information with uninterrupted access is facilitated by the introduction of storage networking, whether it is SAN or NAS. Virtualization can make them more transparent to manage and exploit.

Chapter 2. Policies

In this chapter we explain how to set up Quality of Service (QoS) policies in storage network virtualization and how they influence migration, backup, and performance within storage groups or pools. We will also describe the hardware and the software requirements for the storage network virtualization.

2.1 Policy

To manage your data storage you will need to set up storage policies or rules covering decisions, such as:

- When and how to back up data
- Which data to migrate when more primary space is needed
- How long to keep data that is no longer being used
- Which data is to be placed on the higher performance devices, such as cached devices
- Which data is vital and mission-critical and should be placed on high availability storage subsystems with disaster tolerance

The policies should take into account the overall needs of an organization to most effectively utilize its limited resources, such as disk space and processor time. This is one of the main arguments against the second flavor of decentralized storage management that was common in most open systems environments until fairly recently — user groups owning and managing their own disks. This can lead to gross imbalances in space utilization and performance between each group’s volumes. For example, the space that one group is not using, or is using for inactive data, may be much more profitably used by another group which otherwise has to waste system resources migrating and recalling relatively active data.

If you have data which may need to be handled in a special manner because that data is more sensitive to performance, or the owners of that data need to have a higher priority than normal users, then rules, or policies, can be set up to accomodate those requirements.

The next step is to have a system to automate the application and execution of these rules. Make it transparent to the user by letting him only handle files or data sets, while the virtualization process does the actual allocation and execution of the storage management.

2.2 Migration

Your I/T environment forces you to do many storage related jobs. Adding new storage, deleting unnecessary storage, moving a large data set to the new storage, backing up the data, restoring the data — there are so many kinds of tasks, and migrating data is one of them.

Usually it was time-consuming, dangerous and difficult, because it involved planning, meeting, scheduling, stopping service, backing up, migration, setting up the environment, testing the service, starting service, and so on.

But in a storage virtualization environment, it is easier since the system can handle it for you, once you have set up the rules and the storage pools (Figure 14).

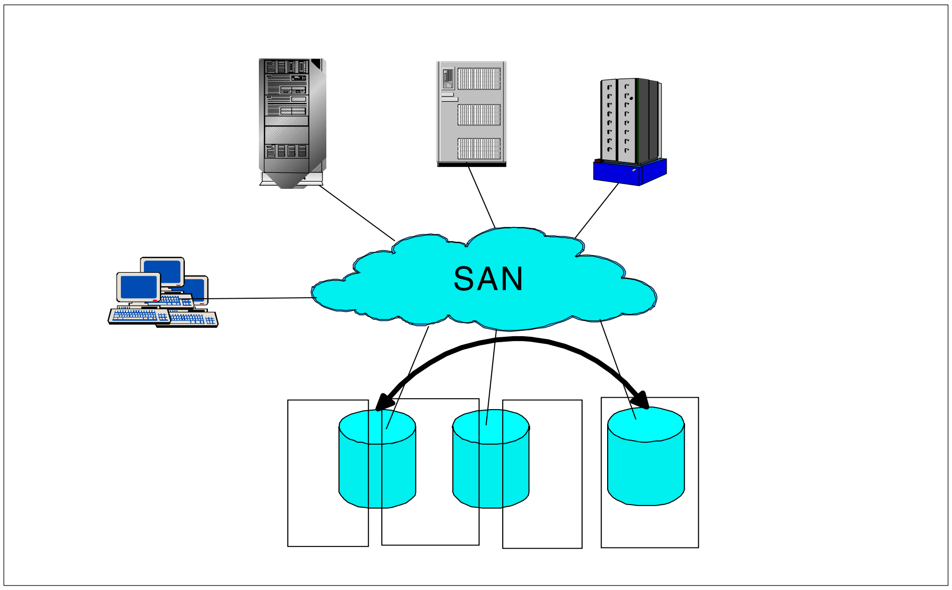


Figure 14. Migration in storage network virtualization

Storage virtualization combines the heterogeneous storage systems into a single resource within the SAN environment. With separation of physical view and logical view, migration can be done without affecting applications.

The policy based management will detect data sets being left unused for a time exceeding what you have set as an acceptable time, and move it to a lower level in the hierarchy. You can set up these lower level storage pools on slower devices or on automated tape, for example.

When a user references the data set again, it will be brought back to the appropriate storage pool by the system. All the user might see is a longer time to wait before the data set opens, if it has to be recalled from a slow device.

The virtualization system should manage its own pools so that when devices are added to the pools, they are automatically put in use, and the pool is spread over them. If devices are to be removed, it should be able to empty a device by migrating the data on the device out of the pool. Good housekeeping rules apply here too.

2.3 Backup

Backup is the most important job in storage management.

Most installations make a daily backup of either all their data or just the changed data. Before storage networking, the server used direct attached storage tape devices for backup, or invoked network attached storage with backup software, like an IBM Network Storage Manager. But SANs allow the servers to share the storage resources. Every tape device resource attached to the SAN can be used by every server attached to the SAN. So every server can direct its backup data directly to a tape device, and send the metadata describing them to the backup server who stores them in its data base. One of the first benefits is LAN-free backup or ultimately server-free backup, which is being implemented in many backup/restore utilities.

A true virtualization system will handle this automatically for the different files it handles in the storage network, based on the QoS associated with the file (see Figure 15).

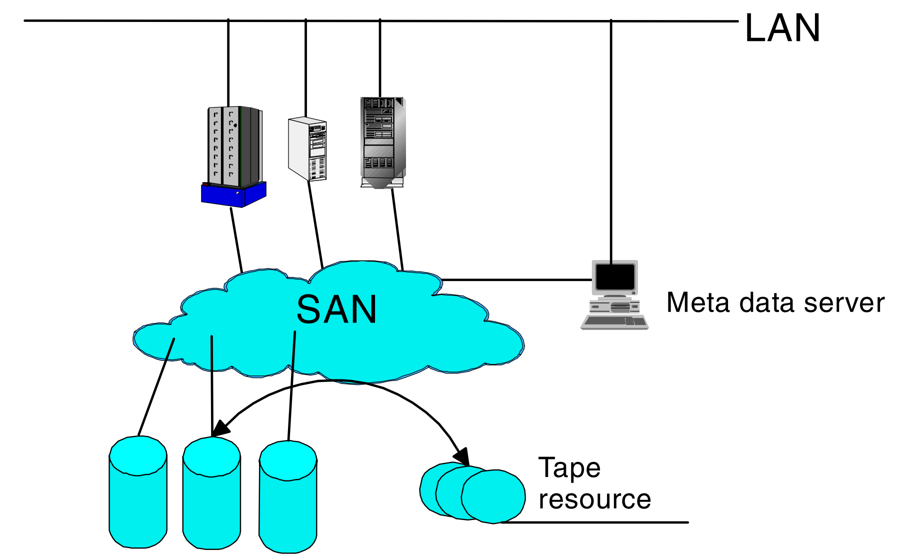


Figure 15. Backup using virtualization

Backup using virtualization has several possible advantages:

- First, the management can be centralized as part of the functions that administrators launch or monitor from a centralized console, with automated problem detection and follow-up. So the responsibility for data can be given to skilled people who have the appropriate resources at their disposal, and this frees the users of having to meddle with media and files they may not even be aware of as being part of their work environment
- Second, backup devices can be located in separate, vaulted rooms at a greater distance from the application servers or the storage pools where the active data reside, for higher security and availability.
- The volume of data to be backed up can be greatly reduced by only making incremental backups of the data that has changed since the last backup cycle, and compressing data while it is written to the backup pools.
- Mobile users can have their data automatically backed up when they log in to the network.
- Resources for backup can be optimized for performance and reliability, by centralizing them on the SAN.

- Backup policies can be set up differently for critical data and for trivial data. Frequency of backup, number of copies to maintain, vaulting, priority and window can be adapted to the profile of user groups or data characteristics.

2.4 Performance

For data intensive applications, performance is critical. Other data types can easily be used with a longer response time or at a lower throughput. So when defining data management policies, these differences can be another criterion for the way the files will be allocated and handled afterwards.

Files can be stored in a storage pool that contains devices which fit their performance requirements. For example, JBODs will give lower performance than an Enterprise Storage Server or a Modular Storage Server (MSS) who have large caches and intelligence to manage their algorithms.

If the storage domain controller is in the data path (symmetrical model) and caches data for its clients, it can use different cache algorithms for files that are in a high performance policy than for files that have lower performance requirements. It could also prioritize IO based on the policy for each particular file.

If the storage domain controller is not in the data path (asymmetrical model), then the choice of device that the file goes to will be the main influence on performance, and it would be up to the device itself to distinguish between I/O types, if it has such capabilities.

2.5 Storage groups

Most installations have a variety of storage products installed: servers from several vendors, several vendor storage products, several models of storage even though one vendor's product is used. You want to get the maximum return on the investment made, by exploiting the functions and features of every device to the fullest. Wasting space costs money, having poor performance aggravates users, and running out of space makes some servers inoperable.

The servers' storage requirements will vary according to business changes, working application changes, backup policy changes. You want to allocate storage to servers as requirements change. The storage networking virtualization technology combines the heterogeneous storage to a virtual storage pool and then allocates it to the servers' requirements (see Figure 16).

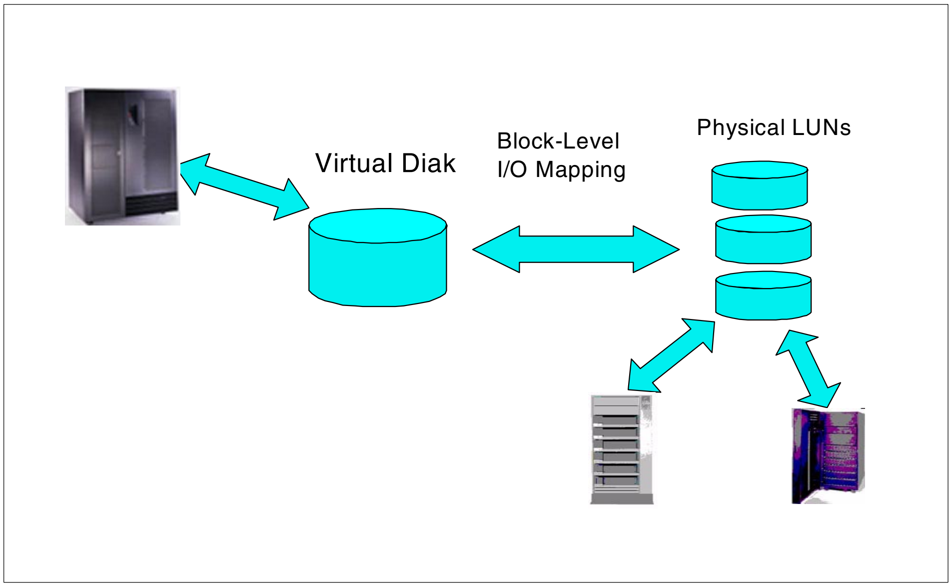


Figure 16. Storage virtual pool

Setting up these storage groups requires some thought, and inventorying both the data and the storage on the SAN in a different light.

For the data grouping, we can consider separating files based on:

- Performance requirements
- Availability requirements
- Size
- Content
- Sharing requirements
- Longevity
- Ownership (system versus user)
- Access pattern (sequential, direct, or a small temporary file?)
- Aggregation (data that absolutely must be available together)

After the virtualization is in place, users no longer will be able to choose where their data will go. But, we need to give them some way of indicating the purpose of the file, so that the system can take this into account while it manages the file.

The last thing we want is an individual policy for every user, so it's a good idea to group users in some way, so that general defaults can be set for a department, for different job contents, or for different locations, to name a few examples. Next, the properties of the files come into use: system data, data bases (shared data), and the criteria mentioned in the bullets above.

Setting up the storage devices in storage groups or pools to accommodate these requirements, you might want to separate volumes and devices depending on:

- Performance: cache size and management, disk speed, disk type (SSA versus SCSI), number of concurrent paths
- Availability: RAID, mirroring, Remote Copy, concurrent maintenance
- Accessibility: flashcopy, cloning
- Cost

So, taking another look at the disk storage devices that you have at your disposal is a second exercise you will have to make. Critically review your tape systems as well, in the context of migration and backup, to see how they fit in the virtual storage pools.

2.6 Requirements

Because the objective of the storage network virtualization is to use the heterogeneous storage in any server according to requirements, you could think that there are no hardware and software requirements to implement. However, we will list some considerations.

2.6.1 Hardware requirements

Usually the virtualization is hardware independent. Disks like JBOD, disk arrays and high function virtual disk array implementations are all managed. This includes recent Storage Area Network (SAN) based symmetrical and asymmetrical disk virtualization technologies.

You will of course need a SAN in the first place, with redundant meshed components like switches or hubs, and the host bus adapters to connect the hosts.

The Meta Data Controllers will require specific hardware to run on, and depending on the level of caching they will implement, will require more memory and processing power, as well as HBAs to connect to the SAN. If they communicate the metadata on a separate IP network, you will have to plan this network as well. Some HBAs may be specially designed for the virtualization — VersaStor is a good example of this.

2.6.2 Software requirements

In the virtualization system, software will drive the technology to manage, communicate, and share data among the heterogeneous servers.

A first requirement is a well managed SAN infrastructure, so a SAN management software must be in place, like Tivoli Storage Network Manager.

This should include topology management, and element management, so a disk manager, removable media manager, and SAN fabric manager are necessary, beside reporting and resource managers.

The main requirement is the virtualization software itself, with its agents or clients to be installed on every server platform that needs to participate. This will be an Installable File System, or another agent that blinds the device drivers on the host.

2.6.3 Platforms

IBM intends to support all platforms (S/390 and zSeries, RS/6000 and pSeries, AS/400 and iSeries, Netfinity and xSeries) and also support non-IBM servers (HP, Compaq, SUN).

Standards are required for interoperability. The first virtualization systems based on present SANs will use SCSI LUN semantics and file systems data trees, so they probably will be able to include most open operating systems. Standards available in storage networking will be reviewed in Chapter 5, “Standards” on page 67.

The arena in which virtualization will break through is the open systems, since they have the highest need for it. S/390 and zSeries already has a very deep system managed storage imbedded, and the AS/400 and iSeries has in its object oriented data structure a true virtual storage system. The main reason to get these systems to participate in a global data space is universal data sharing, and for that they will have to have a part of their data accessed in file systems like UFS.

2.6.4 Working with people — redefining a storage administrator’s job

If you are a storage administrator, do you feel like everyone’s always yelling at you or that you’re the one they dump on, in their rage against the machine?

Well, setting up an environment like a SAN, implementing a NAS and going to storage virtualization is going to change the scope of your job. First of all there is the new technology aspect, but once that has been decided upon, centrally managing data will mean working closely with people.

It will make you work with people more than with devices, since understanding their needs, establishing installation standards, negotiating Quality of Service, and educating them to a new environment is going to be an ongoing task.

Chapter 3. Management

In this chapter we describe data management in a virtualization environment, which consists of SAN management and data management.

We review some of the key issues in storage networking to illustrate the benefits of virtualization.

3.1 Management

Here are the main areas of control to improve the quality and value of a storage network.

3.1.1 Storage management policies

Policies bring discipline to data management. They enforce decisions made across an enterprise.

Chapter 2 discussed policies that allow an administrator to specify storage management rules that determine where data is stored, how often it is backed up, and the conditions under which to migrate data from high-performance storage devices to slower or more long-term storage media such as tape.

An administrator also sets up quality-of-service storage pools that are available to all clients and in which the policies are automatically driving data movement.

3.1.2 System-managed storage

The storage network virtualization architecture allows for centralized, automated storage management in a distributed environment. This centralized management makes it possible to provide mainframe-like, system-managed storage capabilities to the systems attached to the network storage.

3.1.3 Performance

How fast does access to data need to be? This is also linked to frequency of access. Data which is accessed infrequently usually does not need such rapid response as data that is used on a regular basis.

These decisions govern the type of media that the data should be stored on. High frequency, high response time data should be stored on fast disk storage. Lower frequency, lower response time data may be stored on optical disk or tape. Each host server can use suitable storage in the virtual storage pool. The allocation can be managed easily by virtual storage software.

3.1.4 Availability

How critical is immediate access to the information? This will depend, to a large extent, upon the business. Some businesses can last longer than others without access to information. A distribution operation, for example, may be unable to locate items or schedule transport without continuous access to the information, particularly if the inventory, dispatch and stock location are all automated. Business on the Internet is stopped when vital data are not accessible.

These decisions govern both the type of media and the access type. Very high availability would probably require some sort of fault-tolerant RAID array with disaster tolerant remote copy.

Since most companies have become utterly dependent on the data they use on computer systems, availability has become a prime management focal point.

3.1.5 Space Management

It is inevitable that the amount of space required by user applications and data will grow to meet, and then exceed, the amount of space available. With good space management, running into problems can be reduced or avoided.

Space management comprises monitoring information in user storage, and, based on a set of predefined criteria, moving data from one storage level to another level of the storage hierarchy. A pointer of some sort is left in the user storage space, and the information is still there, though the space it occupied is now available for other data. Criteria include things like the age of the file, last access, size, and user storage utilization thresholds.

Managing space means to avoid wasting capacity, but also can be a factor for higher availability. Some operating systems will still suspend themselves if they meet a low disk space availability threshold.

3.1.6 Reduced administrator costs and better administrators

Storage administration is easier, because all storage is centralized and available to all clients. Storage resources are not fragmented, where some storage devices belong only to a specific workstation or only to a specific server. An administrator does not have to move storage devices or reallocate storage among servers manually after they have set up the rules that govern system-managed storage. Data administrators can develop their skills and focus on the important part of their job: managing the information of the company and using tools that are consistent and allow them to control the environment from a single point of control.

3.1.7 Scalability

A SAN is designed to implement large-scale systems. An administrator can add storage devices to the Fibre Channel network as necessary. Each new device is available to all clients on the SAN. Transparent scaling of capacity to varying needs is one of the new benefits of SANs. This also implies that the SAN management components must scale up to growing complexity or workload.

3.2 Storage management in open environment

In the open systems environment, storage management had many restrictions, mainly because of lack of connectivity. Usually the Small Computer Systems Interface (SCSI) was used in open client/server environment. But the storage capacity has grown rapidly, and it becomes necessary to link many users of multi-vendor heterogeneous systems to multi-vendor shared storage resources and to allow those users to access common data, wherever it is located in the enterprise. The answers to connectivity issues are the SAN Fabric and its components, and storage network virtualization addresses the data sharing issues.

3.2.1 SCSI limitations

Usually SCSI hardware was used for storage before SAN configurations existed. But SCSI has limitations. One is the distance limitation; the maximum tolerated length of cabling is only 30 meters, and less if higher speed ultra-SCSI interfaces are used. Another limitation is the inability to address more than a few devices per host bus adapter. This makes disks slaves to individual servers and makes cabling for shared access too complex. For example, in Figure 17 the middle server has exhausted its disk space, while adjacent servers have surplus space. But these resources are bound to specific hosts. There are restrictions to reallocating their spare capacity. Normally, only a single server can access data on a specific disk by means of a SCSI bus. In a shared bus environment, it is clear that all devices cannot transfer data at the same time to multiple hosts.

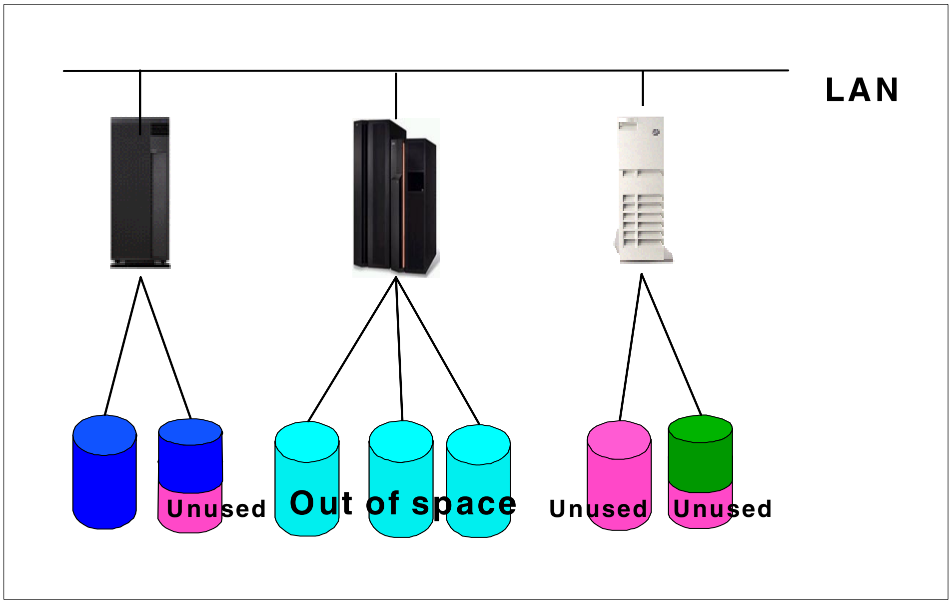


Figure 17. Limited allocation

Disks are addressed as logical unit numbers (LUNs). The SAN industry tends to use LUN to mean a logical unit. But a logical unit may be visible on several paths, and have a different LUN on each path. We use LUN to represent a logical unit, representing a logical volume.

3.2.2 Fibre Channel and SAN

A SAN differs from traditional networks, because it is constructed from storage interfaces. SAN solutions utilize a dedicated network behind the servers, based primarily (though, not necessarily) on a Fibre Channel architecture. Fibre Channel provides a highly scalable bandwidth over long distances, and with the ability to provide full redundancy, including switched, parallel data paths to deliver high availability and high performance.

A SAN enables disk storage capacity to be consolidated. Data from disparate storage subsystems can be combined on to a large enterprise class shared disk array, which may be located at some distance from the servers.

However, several technical issues remain, related to the underlying control of inherently SCSI devices. These issues are:

1. Hosts communicate with physical storage devices using SCSI protocol over Fibre Channel. The destination address is a combination of a target ID and a LUN.
2. Some hosts take ownership of any visible LUNs on attached cables or on the SAN.
3. Two or more hosts connected to the same LUN will unknowingly compete for the same storage device.

To solve these problems, Fibre Channel switches are often zoned to block other servers from connecting to the same storage device. Another solution is LUN masking. This method puts the responsibility for ignoring neighboring servers' access to LUNs that are not theirs on the fabric components, such as switches, hubs, routers and storage servers. We will discuss this in more detail later in this chapter.

Therefore, this inability to share storage assets prevents servers from taking advantage of each other's surplus capacity and leads to application downtime.

3.2.3 Virtualization

Virtualization is a solution for solving the sharing problem, and if it implements security protocols, it can help also solve the SAN LUN access problem (see Figure 18).

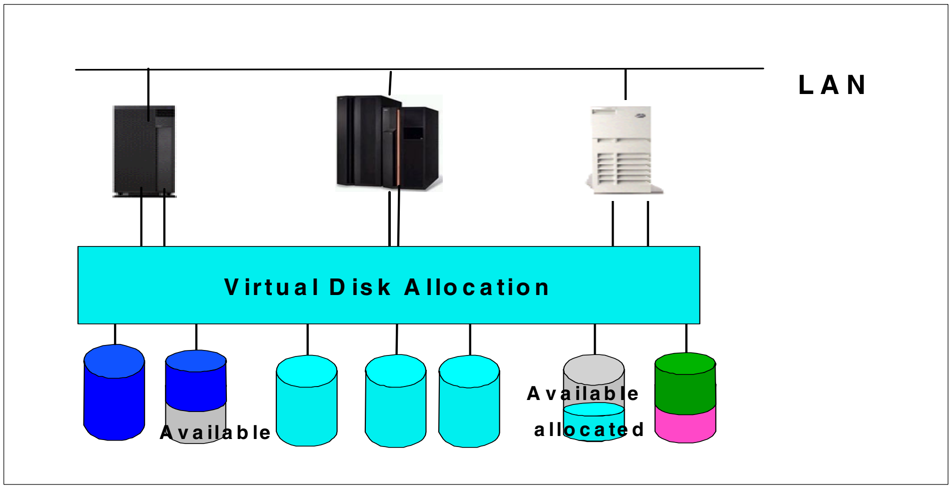


Figure 18. Virtual storage pooling

Physical disks are first virtualized into logical volumes to abstract their properties and ownership. This abstraction enables many important new capabilities that simplify the management and deployment of storage in an open SAN configuration. Storage virtualization provides real-time translation of virtual disk block addresses to physical LUN block addresses, which provides the storage abstraction function for host servers.

3.3 SAN Management

A SAN is simply a new network for storage, which becomes equally vital as Internet/intranet or LAN/WAN for business continuity. As a host cannot function without the TCP/IP network, a host system cannot function when the SAN is down.

Because configuration mistakes lead to downtime and reduced performance, a powerful management tool is needed. Lets look at some of the features required from such a tool.

3.3.1 Topology mapping

In order to manage a SAN, it is first necessary to identify the network topology: which resources are physically connected, what are the available paths, what are the addresses, and so on.

One feature of a network is the ability to create logical subsets in the network. In other words, a physical connection between the host and the SAN does not mean that there is a logical connection between SAN fabric and storage. The SAN management software should present a real-time view of the configuration with easy interfaces to modify the status of every component.

Several views, like server centric, or storage centric, should be presented in a map format and in detail format, with the possibility to activate functions like element managers from the graphical interface.

3.3.2 Event monitoring

Furthermore, an event monitoring feature must be added to detect modifications of state or error conditions. This function collects and consolidates information about events, as well as provides multi-vendor, multi-component fault isolation and problem determination in the SAN. When a problem in the SAN exceeds tolerances or experiences a failure, alerts are generated.

3.3.3 Performance analysis

This function provides the necessary information on throughput rates by fabric device, the average daily throughput, the average hourly throughput, and object size effect.

With correct data, you can tune the performance, make decisions on particular components, and balance the network. Ideally, the management software does this automatically.

3.3.4 Capacity analysis

This function should provide us with data on the average daily transfer data, the average hourly transfer data, SAN usage growth rate, SAN usage by zone, and peak time for data transfer, in order to plan for extensions.

3.4 Data management

System-managed storage (SMS) is a concept that grew out of customer interactions at mostly mainframe user group conferences. The SMS concept is the management of all storage resources and data that is done automatically by system resource — hardware and software — rather than manually. This automated management is installation policy-driven and flexible to meet the needs of the individual business.

This provides applications and users with a view of unlimited storage that meets administrator defined service levels for performance, availability and reliability, as well as platform independent data sharing.

3.4.1 Policy-based data placement

In recent times, it is common to find storage hardware from multiple vendors being implemented side-by-side. Each of the storage devices has different characteristics. With SANs, we have enabled all of this wide variety of storage to be physically networked and seen by each host attached to the SAN. The challenge for administrators is to decide which storage is to be used for which data.

Using a set of customized policy definitions, administrators can define rules that govern the placement of data as it is created.

The policies are defined to the metadata server and may operate on any of the metadata associated with a file. At allocation time, the agent (IFS, for example) communicates the file metadata to the metadata server where the policies determine placement within the data store.

3.4.2 Policy-based life cycle management

Using a set of customized policy definitions, administrators can define rules that govern the detection of life cycle changes and the actions that should be taken when these changes happen. For example, an administrator might want all files in a shared work directory to be monitored for access patterns. When one of these files has not been accessed in eight calendar days, it might be moved from the cached subsystem on which it was initially allocated to a less expensive JBOD, and then, after 21 days of not being accessed, it might be moved to an automated tape system.

3.4.3 Server-free data movement

Objects may be moved around in the data store without notifying the hosts that are accessing them.The data is moved from source, through the SAN fabric and to the target without any interaction with the hosts attached to the SAN.

3.4.4 Data sharing

Another major yearning is cross platform data sharing.

Let us look at the implications of sharing and find the potential of virtualization by looking at it step by step.

Suppose a single server uses one SCSI disk, as illustrated in Figure 19. That will not be a problem. There is no special requirement for software to manage the disk, as the device drivers that are provided with the HBA are sufficient to interface with the operating system.

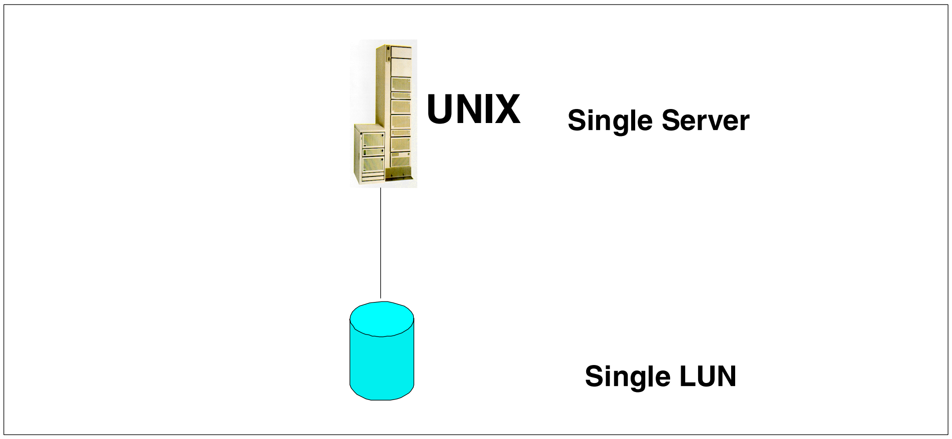


Figure 19. Single server, single LUN

Most servers however, use more than one physical disk. So next, on a single server we want to map multiple disks to one volume, so we introduce a new level of abstraction, the Logical Volume Manager (LVM). Many operating systems now have a built-in LVM to control mapping physical disks to logical disks, as illustrated in Figure 20. If not present in the base operating system, some additional LVM software can be installed on top of it. The server applications will use just logical disks.

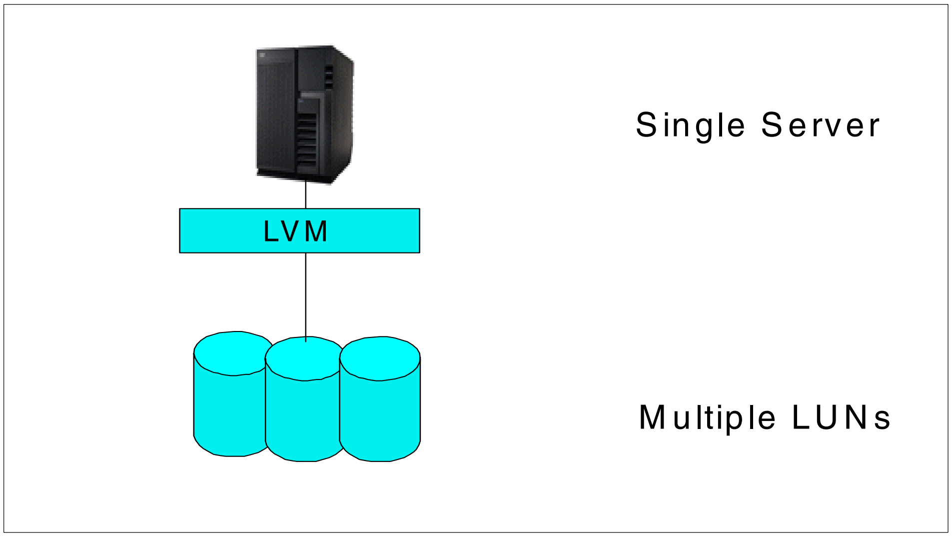


Figure 20. Single server, multiple LUNs

But most of the time, two servers need to control the same disks. For more powerful performance or higher availability, you may want to cluster the servers, and the same data is used by two or more servers. This is manageable between homogeneous systems, where all the servers use the same operating system. A Concurrent Logical Volume Manager (CLVM) will enable that function with lock control, as shown in Figure 21. Clustering software almost always comes with its own CVLM.

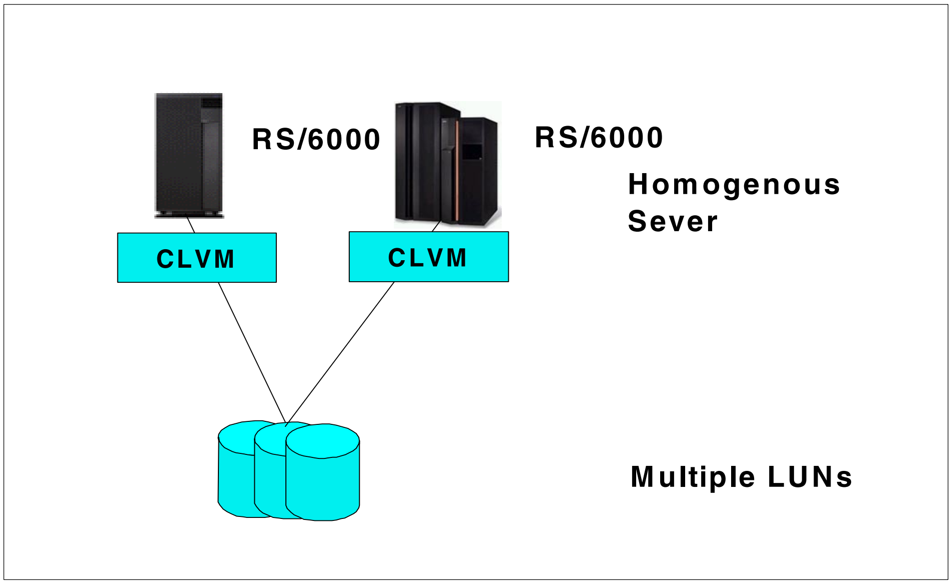


Figure 21. Homogenous server, multiple LUNs

But usually life is more complex. If different servers want to use the same data, what happens? Ideally, they all would use a common CLVM. But some of the servers that you use do not have a CLVM, like NT. Additionally, we haven't made much progress on the convergence of the CLVMs of different UNIX systems like AIX, Solaris, HP/UX, and others (see Figure 22).

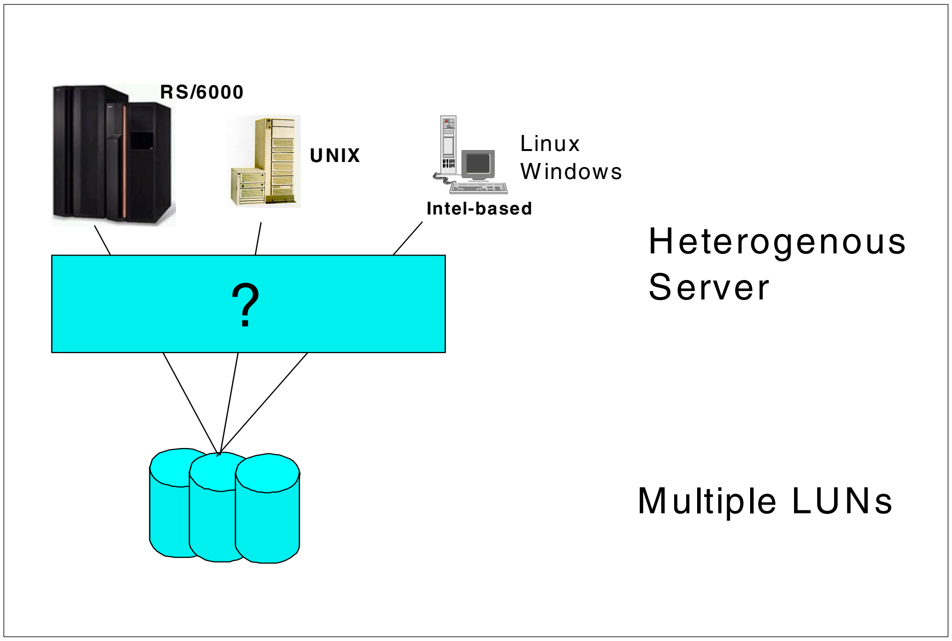


Figure 22. Heterogeneous servers, multiple LUNs

But in a storage network virtualization environment, it may be possible.

We need a higher level of abstraction than mere volume management. We need an agent on the application server to interface with a virtualization system that handles the volume mapping and file mapping for all the connected clients, as illustrated in Figure 23.

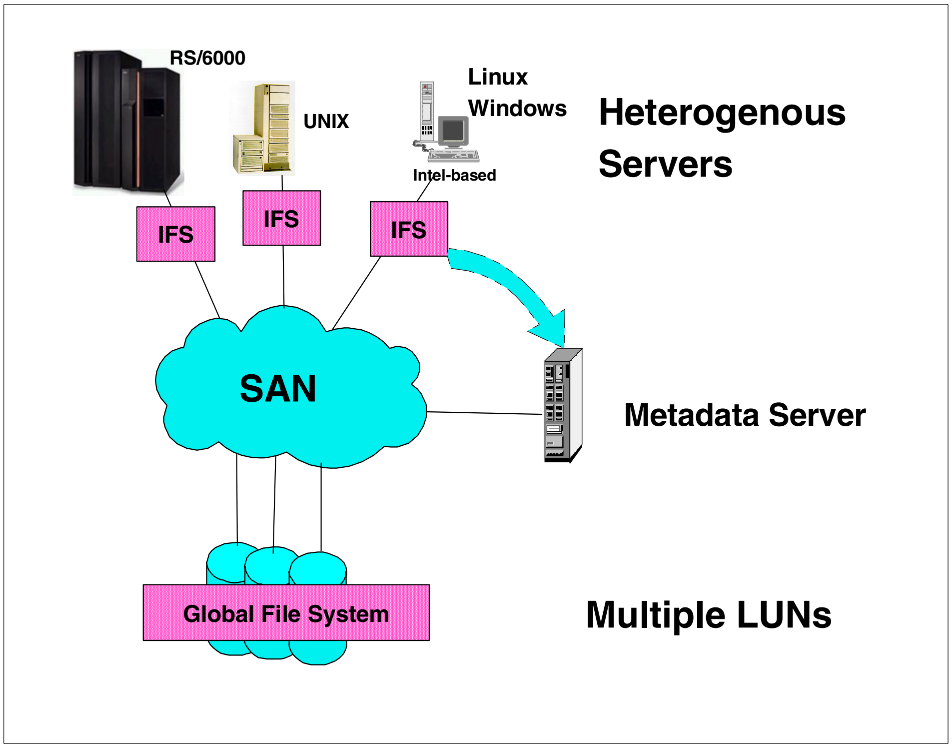


Figure 23. Virtual file system

This agent, as discussed in Chapter 1, could be implemented at different levels, such as:

- On the HBA (the VersaStor approach, which will be detailed in Chapter 7, “VersaStor” on page 81)
- On the file system of each client, as an Installable File System

IBM/Tivoli Open SmS will use the IFS. The LUNs on which the Open SmS file system are spread, can be either physical LUNS managed on the SAN, or virtual LUNs, managed by VersaStor.

An IFS is a subsystem of a client’s file system that directs all metadata operations to a metadata server, and directs all data operations to storage devices attached to the Fibre Channel network. It makes the metadata that is visible to the client’s operations system, and any application that the client runs, look identical to metadata read from a native, locally-attached file system. This means that all hosts see the same directory structure. The IFS on each host provides local presentation of the data in a format needed by the host, and the metadata server handles the locks.

If a Windows NT server wants to open a file, the IFS asks the metadata server for location and authentication, and it is granted a read lock. While the Windows NT server is still reading the file, suppose a participating Sun server wants to open the same file for read. The same process is followed and both servers are allowed to read the same file. If a third participating host, Linux in this case, wanted to open the same file for update, it would not be granted a lock and would have to wait.

So basically, we stick to file system and LUN semantics in the virtualization of storage networks. Only the LUNs that the operating systems see are no longer under their control. They are managed, allocated, zoned, and masked by the software that manages the virtual SAN. The file systems that the operating systems see, are no longer their own file systems alone, but the file system managed by the metadata server.

3.5 Security

Storage devices did not have an inherent security function like Kerberos until now. So managing security is particularly critical in a heterogeneous platform environment. Server level security is essential for data integrity and configuration change management.

In a virtual storage network, security can be controlled at the SAN level and at the metadata server level.

If the metadata server had to control all security, this could become a complex task and overburden the metadata server. The metadata server will control access to file systems by means of access control lists and locking mechanisms.

Managing the SAN level security can be done by the SAN management software. It may be essential to provide security by means of zoning and/or LUN masking at the SAN level. LUN masking at the storage device level (including ESS, SAN Data Gateways or Routers) provides the highest level of security, because it binds storage volumes to specific servers. This ensures that each server can only access its own data, just as though the storage was directly attached to the server. But if sharing is what we want to achieve, these techniques cannot always be used.

For more detailed description about zoning and LUN masking, refer to *Designing an IBM Storage Area Network*, SG24-5758, and *Planning and Implementing an IBM SAN*, SG24-6116.

3.5.1 Zoning

The purpose of zoning is to make a barrier between two different groups. Only the members of the same zone can communicate within that zone and all other attempts from outside are rejected.

Zoning can be implemented in hardware or software (see Figure 24).

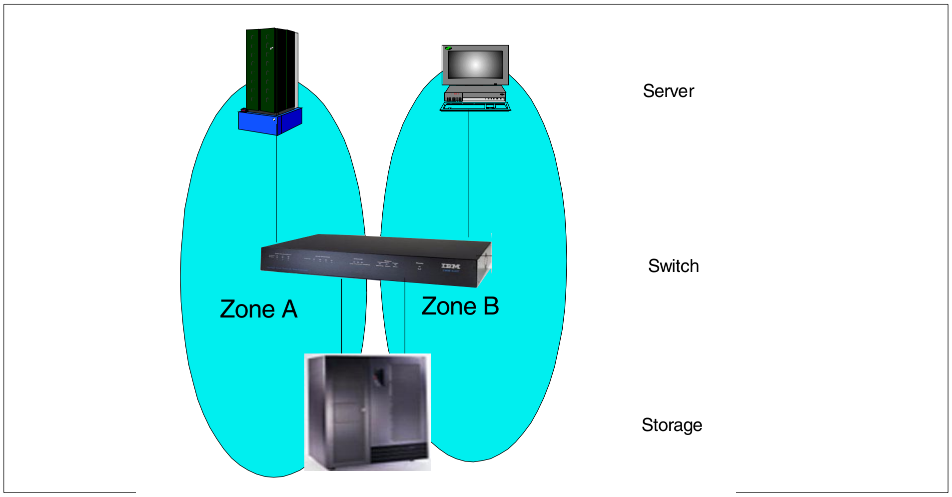


Figure 24. Zoning Configuration

3.5.1.1 Hardware zoning

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

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

A single port can also belong to multiple zones. We show an example of hardware zoning in Figure 25.

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

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

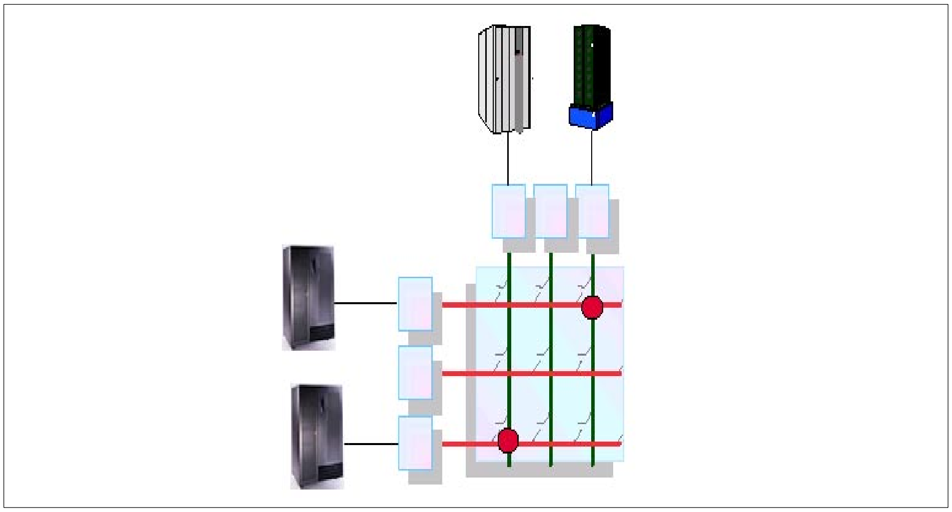


Figure 25. Hardware zoning

3.5.1.2 Software zoning

Software zoning is implemented within the Simple Name Server (SNS) running inside the fabric switch. When using software zoning, the members of the zone can be defined with:

- Node WWN (World Wide Name)
- Port WWN (World Wide Name)

Figure 26 shows the WWN format that is used for software zoning.

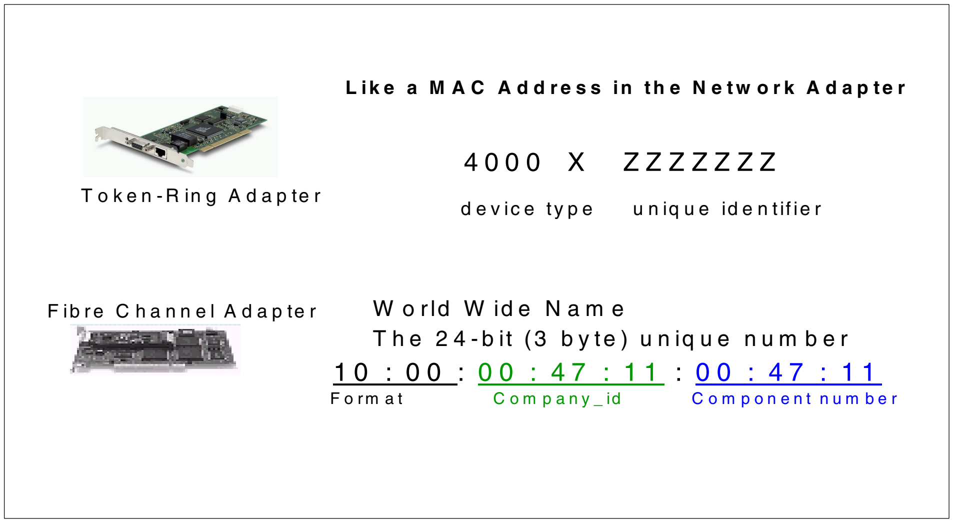


Figure 26. World Wide Name

Usually zoning software also allows you to create symbolic names for the zone members and for the zones themselves.

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

There is a potential security leak with software zoning when a specific host logs into the fabric and asks for available storage devices. The SNS will look into the software zoning table to see which storage devices are allowable for that host. The host will only see the storage devices defined in the software zoning table. But the host can also make a direct connection to the storage device while doing device discovery, without asking SNS for the information it has. This is typical for Windows NT, for instance.

3.5.2 LUN masking

Another level of security can also be introduced. Every storage device offers its resource to the hosts by means of a LUN. Each partition of storage device has its own LUN. If the host wants to access the storage, it needs to request access to the LUN in the storage device. LUN masking will control access to the LUNs. The storage device itself accepts or rejects access requests from different hosts depending on whether or not they can “see” the LUN. An example of a view is given in Figure 27.

The administrator defines which hosts can access which LUN by means of the storage device control program, an element controller that is launched from the storage network manager software. Whenever the host accesses a particular LUN, the storage device will check its access list for that LUN, and it will allow or reject access to the LUN.

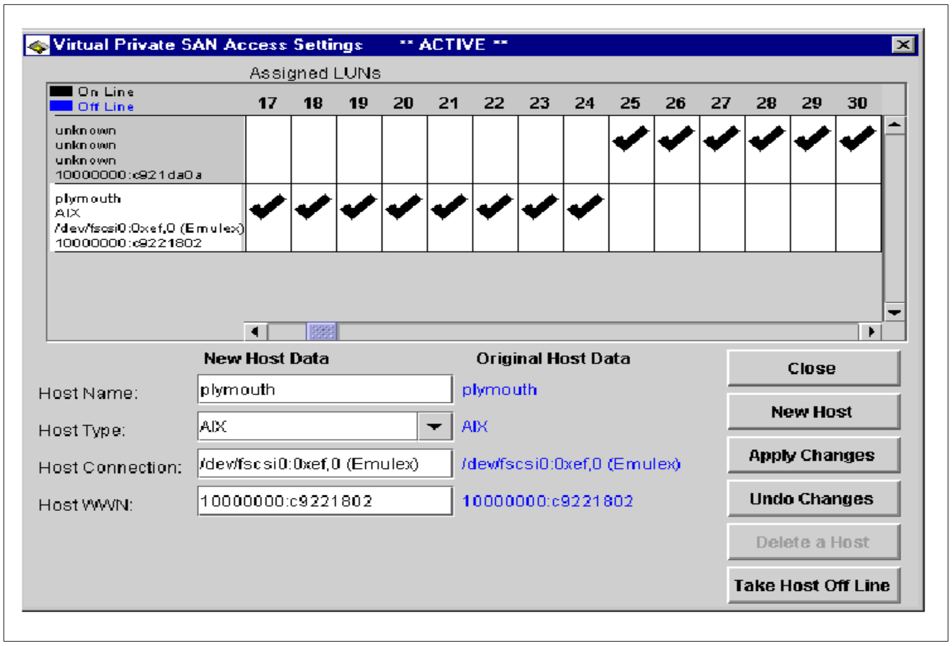


Figure 27. LUN masking example

3.5.3 Locks

Locks can be used in file sharing mode. With storage network virtualization, it is possible for several clients to access the same storage. If a client wants to open a file for read, it requests a read lock and metadata from the metadata server. The metadata server provides metadata and does the read lock operation. And, if another client wants to update the same file, the metadata server will reject client’s request or put him on hold.

3.5.4 Catalogs

The metadata server creates and manages data describing the attributes of a file. That includes lock information, file attributes, device location, mapping of logical blocks to physical blocks, extent lists, status information about existing snapshots for the file, and more. When a metadata server receives a request from a host for read or write, it refers this information to make a decision of accepting or denying a request. It then updates the lock status and sends metadata describing the file's extents and attributes to the requesting host.

Open Systems do not have a catalog system like mainframes. They use directory trees to inventory and locate files.

But all this information in the metadata servers strongly resembles the contents of the catalogs that mainframe operating systems have built to maintain extended information about their files. Attributes describing the policies governing the backup, migration, access lists and availability or performance attributes of a particular file need to be stored, to implement system managed storage.

So the metadata servers in a virtual storage network can be looked at as maintaining catalog information that goes beyond the information normally stored about files. This is interesting for future exploitation: actual file content, statistics about usage — an infinite variety of new information about files could be maintained and exploited at this level.

3.5.5 Security in a virtual storage network

It is obvious that if we want to access all storage from all hosts and leave the access control to the metadata server, that LUN masking and zoning cannot be used. They will only be used to fence out the nonparticipating elements on the SAN, like hosts that have no access to the shared storage pools and storage devices that do not participate in the virtual storage pools. All access within the virtual space will be controlled by the metadata servers and the IFS agent on the hosts.

3.6 Tivoli Storage Network Management

The Tivoli Storage Network Management illustrated in Figure 28 contains:

- SAN Manager that manages the SAN fabric components.
- Disk Manager assigns/unassigns SAN storage to managed hosts.
- Automation extends file system space on hosts managed by Disk Manager when their allocated space is used up.

It provides for all SAN management tasks described in this chapter. For more product detail, see Chapter 8, “Tivoli Storage Network Manager” on page 85.

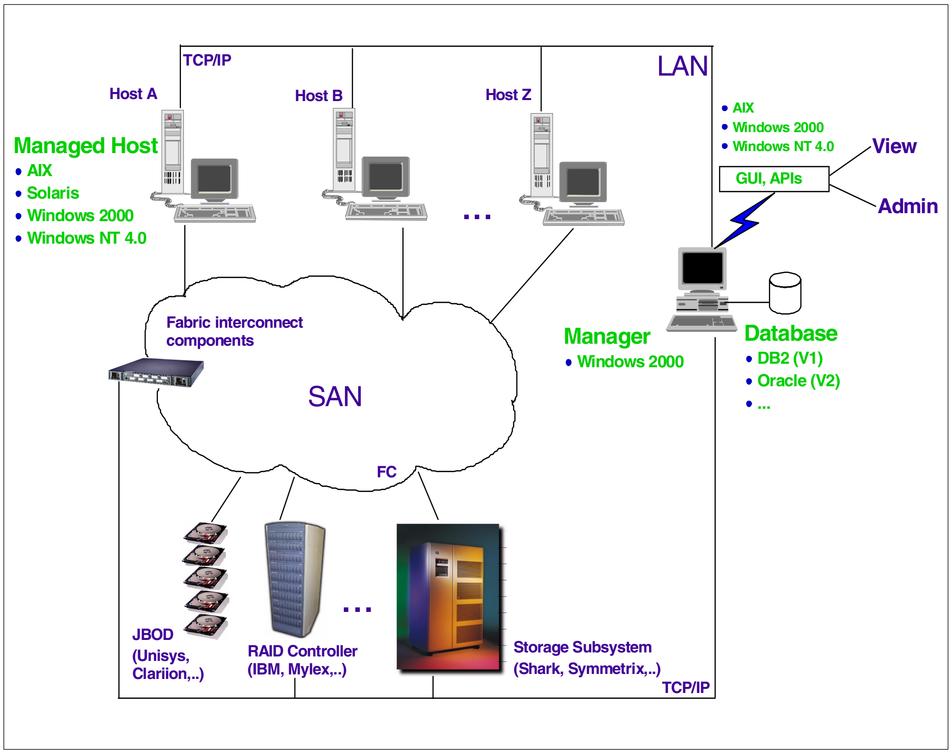


Figure 28. Tivoli Storage Network Manager management environment

Chapter 4. Running a virtual storage network

In this chapter we illustrate how a virtualization system can exploit the functionality of SANs and storage devices, while off-loading server workload, sharing devices and providing multi-platform access to data.

Implementing storage virtualization helps to solve your data management problems, but it may not solve your SAN concerns. These should be addressed by SAN management software, which is described in Chapter 8, “Tivoli Storage Network Manager” on page 85.

4.1 Device recognition

In the virtual storage network, new devices and servers must be recognized as they appear — since many users may be mobile and not always present on the network, and storage devices may be reconfigured in-flight. In-band and out-band discovery of these elements must be provided by the software that manages the virtual storage network via its remote agents or via Simple Network Management Protocol (SNMP). Administrators add new storage or remove storage from the pools as the environment changes. This task should be simple, and automated tools with a single console interface should assist in this job.

Configuration management functions, such as Tivoli Storage Network Manager, provide this task. Once the devices are added to the configuration and LUNs are mapped to them, they can be added to the virtual space managed by a higher virtualization system like Tivoli Open SmS.

Likewise, failures must be detected, so that recovery can take place. Storage device failure can lead to recovery of data either from backup pools, or from a clone or mirror copy if one exists. Client failure detection leads to the freeing of locks and cached data. Server failure detection leads to the regrouping of servers or failover to a clustered server. At all times the disappearance of files, clients, storage, fabric or servers must be detected and treated to ensure data integrity and availability.

Device functionality can include cloning, (remote) mirroring capability, or performance attributes, which can make them candidates for a different QoS pool.

Authentication is important, since it assures that no unauthorized access to the storage network is possible. At this moment, storage devices do not have the possibility to authenticate themselves, but in the future this may become possible. When all components of a storage network can participate in a global security scheme, it may no longer be necessary to isolate them in locked and secure machine rooms, and then we'll have a truly dynamic network.

4.2 SMS policies

When setting up the storage groups or pools, we take the following device characteristics into account:

- Performance characteristics: cache, high speed disks
- Availability enhancing features: RAID, dual controllers, path fault tolerance
- Copy features: cloning, FlashCopy, snapshots

When Quality of Service (QoS) levels are set, these characteristics will determine the administrator's choice of whether to add a set of volumes to a particular pool.

The service level agreements for files that are set up are both static (backup frequency and recovery criteria) and dynamic (performance and availability).

4.2.1 Let virtualization do what storage subsystems cannot do

Since the software model is open and can fit in many building blocks, the underlying hardware devices will not in all installations have the advanced functions that enhance availability and performance.

Some of these functions can be implemented at the software level. Volume management in many operating systems has already included mapping of a single logical volume to multiple physical volumes, allowing for mirror copies. If the software detects that insufficient functionality is present at the hardware level, it can handle special requirements, specifically:

- Caching
- Real-time copying by mirroring, remote copy, replication
- Point-in-time copying for backup or as a FlashCopy (snapshot)
- Compression of data

If a software implementation is used, it takes into account the attributes of a file, that the hardware is not be able to distinguish, let alone manage.

Let us look at an example: FlashCopy or snapshot, meaning point-in-time copies of data. We cover this function in more detail in 4.4.1, “FlashCopy or snapshot” on page 63.

Hardware implementations of snapshots are easy to do at a volume level, since that is what makes up a disk storage server. Even if within the disk storage server there is a volume mapping (RAID, LUN striping, LUN size, not matching physical volume size), it emulates volumes to the hosts. But it has no understanding as to the usage of the data that the hosts are putting on those volumes.

In 3.5.4, “Catalogs” on page 54, we have concluded that the information that is kept about files in a virtualization environment for Windows, UNIX and other Open operating systems, resembles what is in the catalogs of mainframe systems. Access to this information is what makes data set level snapshots possible on a RVA in an OS/390 and zSeries environment. Without catalogs, only volume level snapshots are possible. The copy software on mainframes uses this information to invoke snapshot, specifically for the extents of a data set.

Bringing this sort of functionality to an open environment, where almost every NT, Novell, Linux, or UNIX system has its own file system implementation may be very complicated indeed. It means many versions of utilities are adapted to each file system implementation. So they use snapshot functions at a volume level only: talk SCSI LUN semantics to a disk storage server, and it will get your point.

But a global virtualization, using an IFS that plugs into common metadata for the file systems, makes data set (or file) level snapshots possible. Since the metadata server manages the snapshots, it can serialize access to multiple copies of the files, between many servers.

So software snapshots are more efficient: VersaStor can do this, Tivoli Open SmS can do this. A NAS server can do snapshots, because it manages files, not volumes.

SAN Appliances or SAN managers will be limited to handling SCSI LUNs. Underlying disk hardware can be JBOD. If it can accelerate copy operations at a volume level with functions like flashcopy in an ESS, all the better.

Compression at the software level could also be implemented by the virtualization system to reduce traffic of data on the SAN.

4.3 Coexistence

Obviously, not all data will be managed in the virtual data space, and servers will access data on other servers and storage devices while concurrently they plug in to the virtualization system.

We do not recommend that the same devices are used by the system managed storage software and other allocation software at the same time. This could only lead to hot spots out of the control of the virtualization software.

We also do not recommend that two or more sets of virtual storage servers share the same physical devices, since that could lead to running out of space conditions if the separate storage management software does not communicate with each other. One way of avoiding this conflict is to zone off devices to each virtual storage network, or to mask LUNs within a single storage device.

Maybe you cannot afford SAN attachments for all your application servers. After all, Fibre Channel HBAs are rather expensive, as are ports on switches and hubs. So when you want to do storage consolidation for a high number of NT or UNIX servers you may want to combine the best of both worlds. High performance, super scalability for some, and NAS storage for the consolidating storage for other workgroups. NAS Gateways could plug into the SAN disk configuration for larger departments.

This allows for the coexistence of several solutions within one installation, as illustrated in Figure 29.

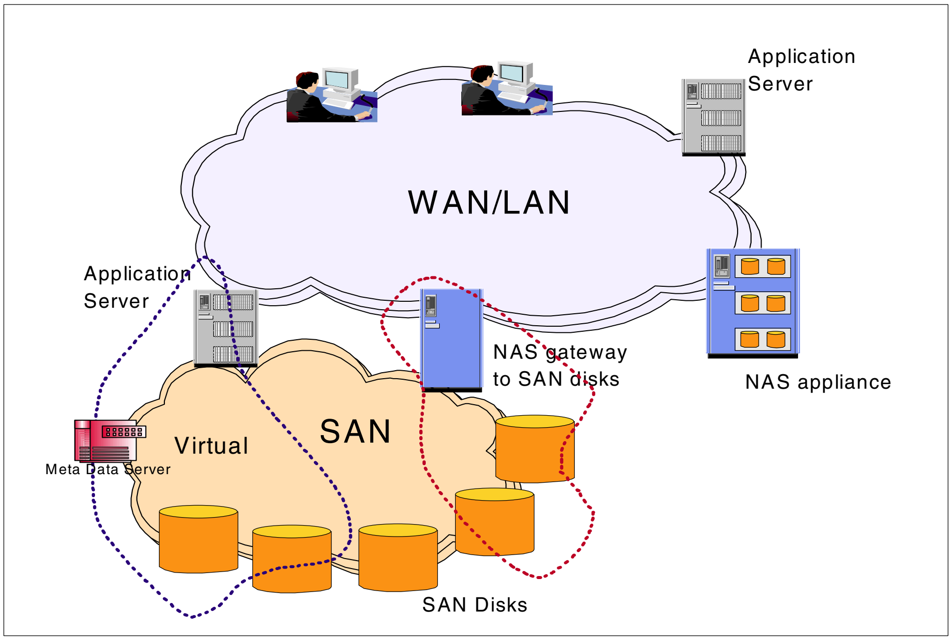


Figure 29. NAS and SAN coexist

Another way to look at storage pooling is by disk functions required on different application servers.

Database applications typically want raw device access, and they don't like file systems in their way, even when they can function on file systems. Since Data Base Management Systems (DBMS) do their own data caching in memory and have logging facilities for recovery, they can handle availability and performance issues by themselves, and generally only want the shortest and fastest route to the physical device once they require it. So putting yet another layer of virtualization under a DBMS could hinder rather than help their response time. At the same time, your other nondatabase applications could be greatly helped if their data was secured and cached on virtual storage. This is illustrated in Figure 30.

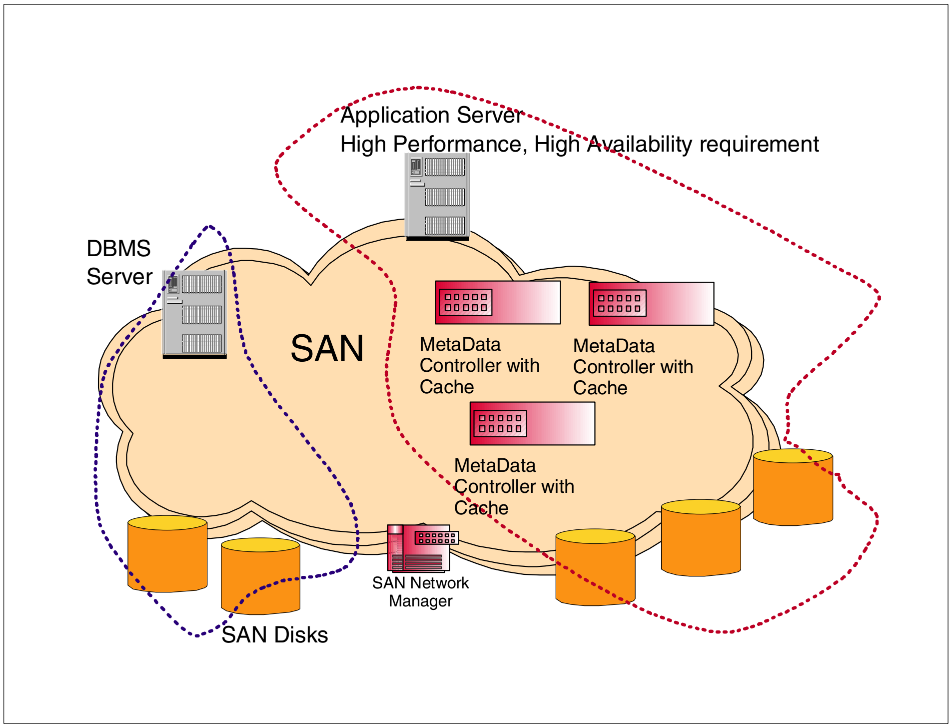


Figure 30. Combination of two virtualization levels

The DBMS server uses virtual LUNs that are managed by the SAN Network Manager. Other applications have their files under control of a higher level virtualization system.

4.3.1 Fallback

Once you have made the step into a virtual data space, why would you want to go back? It's obvious that taking such a step and the planning, design, pilot project and implementation efforts that come with it are no one-night decisions, and that after the migration there should be no reason to go back to physical volume management.

That makes those planned phases of installation of the utmost importance. Before you hand control to the auto-pilot, you make sure it is set up correctly, and that means that the policies for data management are clearly outlined and set up for execution by the virtualization system. It also implies that the SMS servers have been tested for fault tolerance and failover. And that the backups are in place, as with every migration, to allow crude data restoration to traditional physical volumes.

Also, with all the emerging options, redeployment of resources will become more flexible.

4.3.2 Different types of implementation

Tivoli Open SmS will provide an open, standards compliant virtualization of data. Some other software will be developed by other vendors with the same purpose. So when you choose one particular brand of virtualization, it is very important to evaluate its openness and coexistence with other software, and the scope of possible clients, since one IFS is not necessarily compatible with another.

4.4 Copy functions

Virtualization is about avoiding unnecessary copies. So minimization of duplicate data is a prime objective. On the other hand, if we want transparent backup or multi-user concurrent read and write in the storage network, then copies of data cannot be avoided. Migrating data is basically copying them to a new location. The virtualization software will have to be able to make these copies even if the functionality is not present in all hardware storage devices that are in use. Also, it will be able to do this more efficiently than hardware, as indicated in 4.2.1, “Let virtualization do what storage subsystems cannot do” on page 58.

4.4.1 FlashCopy or snapshot

A FlashCopy or snapshot makes a copy of a volume or, if possible, of a single file, at a point in time. It then maintains this copy separately from the original. The copy remains in the original state, unless it is opened for update. The original continues its own life. This is also called cloning. The terms “flash” and “snap” refer to the speed at which this is done. The speed results from the fact that, at the moment that the copy is set up, not all the data is physically duplicated. The tables mapping the data are copied, and once that is done, both copies are available for use.

This means that the time, which applications that are using the original file in update mode have to be stopped, is reduced dramatically.

It then boils down to duplicating data that changes in either copy to a different place, managing duplicate mapping tables of the data in the metadata controllers, domain controllers, or at the storage level, and running a full copy of the file in the background, if required, without affecting applications. Figure 31 illustrates this.

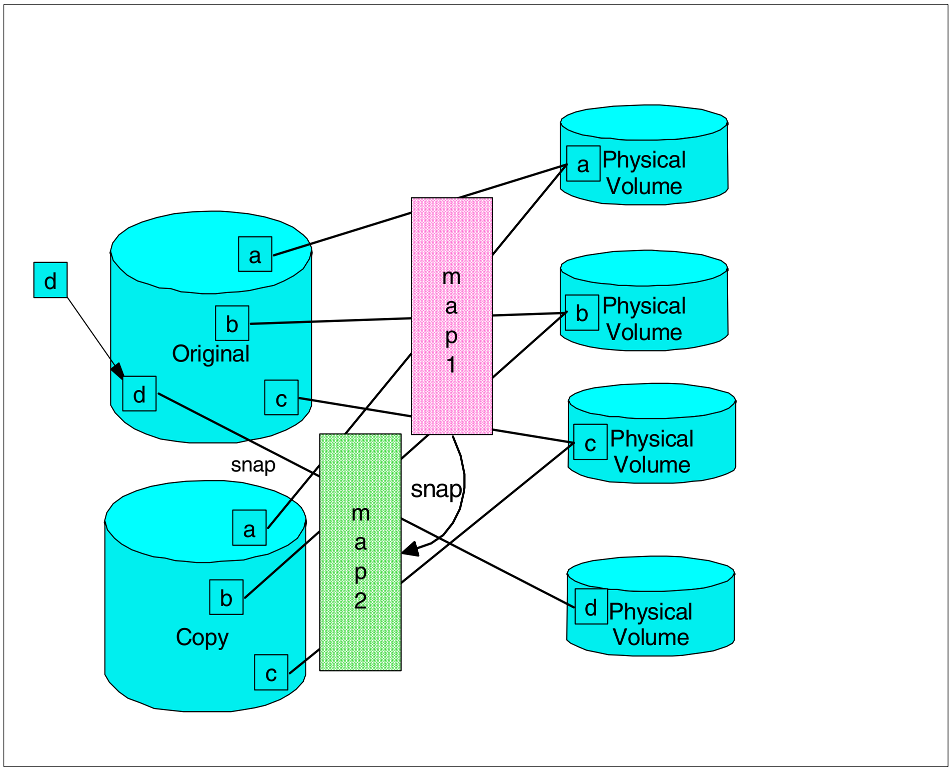


Figure 31. Flashcopy or snapshot principle

There are several possible ways to implement this: with or without physical copy of the data, and at volume level or at a higher granularity, such as at a file or block.

In a virtual storage system, the best way to implement this would be at the metadata controller level, so all host servers can access the copies. While the flash is being set up, the metadata controller will have to manage the incoming requests from various clients to the data being flashed, and hopefully it will not get swamped. Before starting the flashcopy, the MDC will need to stop clients from accessing the data, and relieve this restraint after the flash copy is made. This is a performance issue: the MDC will have to be very scalable in cache, processing power and bandwidth, or be limited in the number of clients it can handle.

If the actual flashcopy can be delegated to a hardware controller, this will solve part of the problem, but will probably fence the copies within the same hardware box.

The application of flashcopies or snapshots is mainly for backup purposes: they can constitute online disk versions of the file as its changes to be retrieved, when a user messes up the file, for example. Or they can be used to make tape backup copies from them while the application uses the original again for further updates. Some SAN disk storage servers or NAS appliances available now provide some sort of flash copy, but at the volume level, not at the file level, and most of them can only maintain a reduced number of copies simultaneously. Many installations will prefer to have more than two copies of a file, and not be forced to restore at the volume level when one user loses his files, impacting all the others. Virtualization software can improve on this situation.

4.4.2 Remote copy

Making a remote copy set up is either a special form of flashcopy, or maintains a synchronous set of files, such as PPRC. Managing this will require extension of the SAN over distance, and using storage servers which have this capability.

4.5 Performance

In a distributed environment, a client/server access model is used. The file server accesses data from storage devices, and then sends the files to the clients.

This has an additional limitation of using conventional network bandwidth to transfer the data. While these systems allow users to share data, they do not provide the high performance required for data-intensive applications.

But in a storage network virtualization environment, the clients obtain only metadata from the storage domain server, and can then access the data directly at high-bandwidth provided by a Fibre Channel or other high-speed network. Direct data access eliminates server bottlenecks and provides the performance necessary for data-intensive applications.

Chapter 5. Standards

IBM and Tivoli Systems are committed to enabling an interoperable, manageable Storage Area Network (SAN) community. This requires standardization. Fortunately, many of these standards already exist.

For customers, the following section is provided as a reference. Use it to drive your selected Fibre Channel infrastructure hardware vendors and management software vendors to standardization. The results will be more functionality and choice for you.

For Fibre Channel infrastructure vendors, these are the requirements that are branded Tivoli Ready for SAN Management.

5.1 Switches

For switches to be be standardized, they need to meet the following levels:

- Name Server as defined in ANSI FC-GS-2 version 5.3
- Management Server support as defined in ANSI FC-GS-3 version 6.4 or later
- Registered-State-Change-Notification (RSCN) and Registered-Link-Incident-Record (RLIR) / Link-Incident-Record-Registration (LIRR) event reporting as defined in ANSI FC-FS version 0.2 or later

Optionally, they can also provide:

- IETF FC Management MIB version 2.2 or later

5.2 Hubs

To be Tivoli Ready, hubs must support at least one of the following:

- Management Server support as defined in FC-GS-3 version 6.4 or later
- Respond to Request Node-Identification Data (RNID) and Request Topology Information (RTIN) Extended Link Services (ELS) requests as defined in ANSI FC-FS version 0.2 or later
- IETF FC Management MIB version 2.2 or later

5.3 Bridges, Gateways and Routers

Bridges, gateways and routers must be able to:

- Respond to RNID ELS requests
- Support RSCN event reporting

And support at least one of the following:

- Respond to RTIN ELS requests
- IETF FC Management MIB version 2.2 or later

5.4 End Point Devices including Host Bus Adapters (HBAs)

To provide a standard level of functionality, HBAs must be able to:

- Respond to RNID ELS requests
- Support RSCN event reporting
- Register with Management Server (as defined in FC-GS-3 version 6.4 or later):
  - Platform Name
  - Platform Type
  - Management Addresses

5.5 HBA Device Driver

The device driver for the installed HBA must provide a mechanism for performing the following:

- Issue Name Server and Management Server Queries
- Issue RNID/RTIN ELS frames to a specified target port
- Notification for RSCN and RLIR/LIRR events

Part 2. Products and offerings

This part discloses some specific products, offerings or plans that are currently under development and that have already been mentioned or announced in the public arena.

Chapter 6. Tivoli Open System-managed Storage

The Tivoli Open SmS research project at Almaden takes a global approach to the virtualization of storage. The purpose is to bring the function known on the mainframe as System-Managed-Storage to the Open Systems environment. The product that is being developed by Tivoli, based on this research project, is called Tivoli Open System-managed Storage (Tivoli Open SmS) and is planned to be available in the year 2001.

Compaq will support Tivoli's Open System-managed Storage model and its supporting technologies, under the agreement between both companies announced on July 6th, 2000. Both IBM and Compaq are working on products for a true Open SAN.

6.1 Overview

Tivoli Open SmS is a distributed object storage system that provides a complete storage management solution. It is designed to provide, in a distributed environment, performance comparable to that of file systems built on bus-attached, high-performance storage. In addition, it provides high availability and increased scalability and manageability. Tivoli Open SmS also provides integrated storage management facilities, such as backup and data migration.

Figure 32 shows that Tivoli Open SmS clients communicate with Tivoli Open SmS servers over an enterprise's existing IP network using the Tivoli Open SmS protocol. It also shows that Tivoli Open SmS clients, servers, and storage devices are all connected to a Storage Area Network (SAN) on a high-speed, Fibre Channel network.

An installable file system (IFS) is installed on each of the heterogeneous clients supported by Tivoli Open SmS. The IFS directs requests for metadata and locks to a Tivoli Open SmS server, and requests for data to storage devices on the SAN. Tivoli Open SmS clients can access data directly from any storage device attached to the SAN.

Tivoli Open SmS clients aggressively cache files, as well as metadata and locks that they obtain from a Tivoli Open SmS server, in memory. They do not cache files to disk.

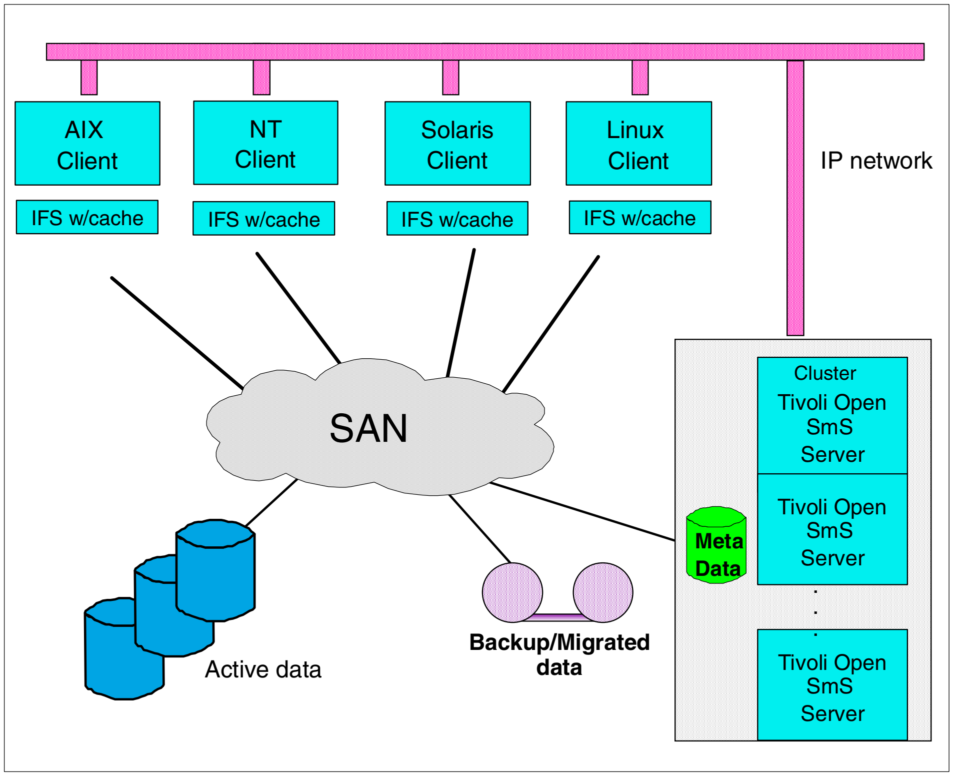


Figure 32. Tivoli Open System-managed Storage

An enterprise can use one Tivoli Open SmS server, a cluster of Tivoli Open SmS servers, or multiple clusters of Tivoli Open SmS servers. Clustered servers provide load balancing, fail-over processing, and increased scalability. A cluster of Tivoli Open SmS servers are interconnected on their own high-speed network or on the same IP network they use to communicate with Tivoli Open SmS clients. The private server storage that contains the metadata managed by Tivoli Open SmS servers can be attached to a private network that is connected only to the cluster of servers, or it can be attached to the Tivoli Open SmS SAN.

Within a server cluster is a storage management server. A storage management server is a logical server that issues commands to back up and migrate files directly over the Fibre Channel network from one storage device to another. No client involvement is required to perform these tasks.

6.2 Tivoli Open SmS protocol

The Tivoli Open SmS protocol is a locking and data consistency model that allows the Tivoli Open SmS distributed object storage system to look and behave exactly like a local file system. Using the Tivoli Open SmS protocol, clients and servers work together to ensure strong data consistency in a distributed environment.

The Tivoli Open SmS protocol provides locks that enable file sharing among Tivoli Open SmS clients, and, when necessary, provides locks that allow clients to have exclusive access to files. A Tivoli Open SmS server grants the locks to clients. With the Tivoli Open SmS protocol, when a client reads data from a particular file, it always reads the last data written to that file from anywhere in the Tivoli Open SmS distributed object storage system.

To open a file in the Tivoli Open SmS distributed object storage system, a client does the following:

- Contacts a Tivoli Open SmS server to obtain metadata and locks.

  Metadata supplies the client with information about a file, such as its attributes and location on storage device(s).

  Locks supply the client with the privileges it needs to open a file and read or write data. The Tivoli Open SmS locking scheme ensures strong data consistency.

- Accesses the data for the file directly from a shared storage device attached to the Tivoli Open SmS high-performance, Fibre Channel network.

6.3 Tivoli Open SmS clients

Tivoli Open SmS enables full, transparent data sharing of files among heterogeneous clients. Clients are currently available on Windows NT, AIX, Solaris, and Linux operating systems.

All Tivoli Open SmS clients can access the same data using the Tivoli Open SmS uniform global namespace. A uniform global namespace allows all clients to have a consistent view of the Tivoli Open SmS name tree at all times.

6.3.1 File server support

A file server, such as an NFS, CIFS, or HTTP server, can also be a Tivoli Open SmS client. For these clients, Tivoli Open SmS provides:

- Scalability: A file server that is a Tivoli Open SmS client can access all the files in the Tivoli Open SmS distributed object storage system. Tivoli Open SmS is highly scalable and can, therefore, provide a file server access to a vast amount of data.
- Reliability and fail-over processing: Because many servers can be Tivoli Open SmS clients and can export the same files, requests from clients of a failed server can be transferred to another server using any technique, such as high availability cluster multi-processing (HACMP) or IP address stealing, supported by the file server.

6.3.2 Installable File Systems

Tivoli Open SmS requires an Installable File System for each client operating system it supports. IFS driver software is available through a Web interface to a Tivoli Open SmS server and is easy to install.

An IFS is a subsystem of a Tivoli Open SmS client's file system that directs all metadata operations to a Tivoli Open SmS metadata server, and directs all data operations to storage devices attached to the Fibre Channel network. It makes the metadata, that is visible to the client's operating system and any application that the client runs, look identical to metadata read from a native, locally-attached file system.

Note that special purpose applications, such as digital libraries and databases, can also access data from the Tivoli Open SmS distributed object storage system by using an application programming interface (API) to the Tivoli Open SmS protocol. Because these applications do not use the file system to access their data, the clients on which they run do not need to have the Tivoli Open SmS IFS installed.

6.3.3 Tivoli Open SmS client cache

Tivoli Open SmS client cache is used to achieve low-latency access to metadata and data. A Tivoli Open SmS client can cache:

- Data: Caching data allows a client to perform reads and writes for smaller files locally, eliminating I/O operations to storage devices attached to the SAN.
- Metadata: Caching metadata allows a client to perform multiple metadata reads locally without contacting a Tivoli Open SmS server.

Note that all metadata writes are done by a Tivoli Open SmS server.

- Locks: Caching locks allows a client to grant multiple opens to a file locally without contacting a Tivoli Open SmS server.

A Tivoli Open SmS client performs all data caching in memory. If there is not enough space in the client's cache for all of the data in a file, the client simply reads the data from the shared storage device on which the file is stored. Data access is fast because the client has direct access to all storage devices attached to the Tivoli Open SmS high-speed, Fibre Channel network. There is no need for a client to cache data to a private disk.

6.4 Tivoli Open SmS Servers

Tivoli Open SmS servers are currently available on AIX and Linux operating systems, and can easily be ported to other operating systems, such as Windows NT. Support for multiple operating systems allows an administrator to choose from a wide range of workstations on which to install the server programs. This allows the administrator to provide the appropriate level of performance for an enterprise. For example, an administrator can choose to install the server programs on workstations built on Intel processors for cost-effective scalability or on IBM SP2 supercomputers for high-end scalability.

Tivoli Open SmS includes these three types of logical servers:

- Metadata server
- Administrative server
- Storage management server

These logical servers can all reside on the same physical server or on separate physical servers.

6.4.1 Metadata servers

A metadata server performs these functions:

- Manages allocation and placement of data on storage devices
- Authenticates clients
- Performs metadata writes
- Serves file system metadata to clients
- Grants file and data locks to clients

An enterprise can use a single metadata server, a cluster of metadata servers, or multiple clusters of metadata servers. An administrator can move data between clusters using standard file system commands.

Using metadata servers in a cluster configuration has these benefits:

- Load balancing: The workload and data structures for the Tivoli Open SmS distributed object storage system are partitioned and allotted to the servers in the cluster.

  This is a continuous process that keeps the cluster workload balanced at all times.
- Fail-over processing: If one metadata server fails, its load is distributed evenly among the remaining metadata servers in the cluster.
- Scalability: An administrator can add more metadata servers to a cluster or add more metadata server clusters to the SAN to serve more data and more clients. Note that multiple metadata server clusters cooperate to maintain the Tivoli Open SmS uniform global namespace.

The metadata servers in a specific cluster must all be of the same type. However, an installation can have multiple clusters of different types. For example, an enterprise might have one server cluster in which all the servers run AIX, and another server cluster in which all the servers run Linux.

Metadata servers in a cluster can be interconnected on their own high-speed network or on the same IP network through which the servers communicate with Tivoli Open SmS clients. The metadata managed by a cluster of metadata servers resides on private server storage that is shared among the metadata servers in the cluster.

6.4.2 Administrative server

There is only one administrative server per Tivoli Open SmS server cluster. One of the systems in a Tivoli Open SmS server cluster (or the only Tivoli Open SmS server in an installation) is elected to perform the role of administrative server.

An administrator uses a Tivoli Open SmS administrative server to:

- Manage physical devices, for example, by adding or removing disks and server nodes
- Create quality-of-service storage pools for reliability (for example, storage pools that consist of RAID or striped storage devices) and performance guarantees (for example, storage pools that consist of hierarchically-managed, random or sequential access, or low-latency storage devices)
- Manage quotas that specify the amount of storage in a specific storage pool that can be used by a specific user or group

- Manage groups that are a collection of users to which an administrator can grant specific rights and permissions
- Perform authentication for users and groups
- Take snapshots of the Tivoli Open SmS metadata

  A snapshot creates a point-in-time view of the entire Tivoli Open SmS uniform global namespace or any portion of it. Tivoli Open SmS clients support metadata snapshots by implementing a copy-on-write facility. Any data a client writes to a file between snapshots is written to a different location in storage.

  The data that existed when the last snapshot was taken remains in the original location. Data mining applications and backup utilities can access the point-in-time view of Tivoli Open SmS data without interrupting normal data operations on the Tivoli Open SmS SAN. While a snapshot is being taken, all data remains online and can be read and written to by clients.

Note that all administrative tasks can be performed online with no interruption in service to clients.

6.4.3 Storage management server

The storage management server performs all storage management services for the Tivoli Open SmS distributed object storage system. It has direct access to all of the data and metadata in the storage system, and, therefore, can perform services, such as backing up and migrating data based on storage management policies set up by an administrator. It performs all of these tasks across the SAN with no client involvement.

6.5 Tivoli Open SmS shared storage

An administrator can choose to use various types of storage for the Tivoli Open SmS SAN. For example, an administrator can attach JBOD, RAID, and hierarchically-managed storage devices to the SAN. An administrator can also attach tape devices for backups and long-term storage.

All storage devices attached to the Tivoli Open SmS SAN can be accessed by all clients that are part of the Tivoli Open SmS system. This enables data sharing among the heterogeneous clients supported by Tivoli Open SmS.

6.6 Tivoli Open SmS security

Storage devices that are capable of performing authentication are not yet available. To protect data on storage devices that lack security protocols, the Tivoli Open SmS SAN must be set up in a secure location. Unauthorized clients must be prevented from attaching to the Fibre Channel network to which Tivoli Open SmS clients, servers, and storage devices are connected.

When storage devices that are capable of performing authentication become available, Tivoli Open SmS will be able to take advantage of those devices to increase the security of the Tivoli Open SmS environment.

Currently, Tivoli Open SmS does the following:

- Uses Kerberos authentication protocols to identify Tivoli Open SmS clients and valid users when communicating with clients over existing, untrustworthy TCP/IP networks
- Optionally encrypts lock and metadata messages to conceal the contents of client and server actions

Note that data traffic that occurs on a secured SAN does not need to be authenticated or encrypted.

6.7 Summary

Tivoli Open SmS's data access model, in which a client obtains only metadata and locks from a server and reads data directly from storage devices, improves the performance of data sharing when compared to a traditional distributed environment, in which servers obtain data from storage devices, and then send the data to clients. The improved data access model and the SAN and Fibre Channel network technologies used by Tivoli Open SmS provide heterogeneous data sharing in a distributed environment with performance that is indistinguishable from local file system performance.

Centralized storage and server clustering increase the scalability of the Tivoli Open SmS SAN. An administrator can add more storage devices as needed to serve all clients, and add more servers to a server cluster or add more server clusters to the SAN to manage more data and serve more clients. Tivoli Open SmS server clusters also increase the availability of data to clients by performing load balancing and fail-over processing.

In addition, Tivoli Open SmS is easy to manage. An administrator can perform SAN management tasks, such as adding or removing disks, creating quality-of-service storage pools, and taking snapshots of the Tivoli Open

SmS data tree without interrupting service to clients and with a minimal number of commands.

Finally, because the Tivoli Open SmS architecture makes it possible to realize the advantages of open system-managed storage, Tivoli Open SmS provides a complete storage management solution.

Chapter 7. VersaStor

In this chapter we look into the VersaStor technology, that is under development by Compaq and endorsed by IBM, under the agreement between both companies announced on July 6th, 2000. Both IBM and Compaq are working on products for a true Open SAN.

No VersaStor products are available at the time of this writing, but the technology will be made available to system, software and host bus adapter manufacturers so that products can be expected for delivery in the year 2001.

7.1 VersaStor technology

Compaq's VersaStor technology is based on the Asymmetrical Pooling model and is a complete and functionally rich implementation of SAN-wide storage virtualization at the fabric level. The elements of VersaStor Technology include:

- VersaStor Management Software
- VersaStor Appliances
- VersaStor Agent Technology

The management software, which resides on VersaStor Appliances, manages the overall storage virtualization process. It controls the VersaStor storage pool and migrates data within the pool independent of host servers. It also builds virtual disk mapping tables based on high level attributes, permanently stores these mapping tables, and loads them selectively to host bus adapters or systems that are VersaStor Ready. VersaStor Ready adapters or systems provide real-time translation of virtual disk block addresses to physical LUN block addresses, which provides the storage abstraction function for host servers. The management software also creates and manages virtual disk mirrors and snapshots according to user requests, and provides error recovery for the VersaStor system.

VersaStor Technology provides SAN-wide storage pooling across heterogeneous array controllers from multiple vendors to simplify the deployment of Open SANs. It provides automatic, policy-driven data migration within the VersaStor storage pool and does automatic load redistribution when new capacity is added to the pool.

VersaStor Technology supports a storage pool hierarchy that makes it easy to organize and manage virtual disks and physical data placement. The hierarchy can be changed at any time, and data will be migrated automatically to accommodate the change. The VersaStor storage pool hierarchy greatly simplifies storage management by elevating it to a single storage model.

VersaStor Technology also provides selective presentation of virtual disks to host servers so virtual disk security can be maintained in an Open SAN environment. Virtual disks can be dynamically expanded in size, and point-in-time copies of virtual disks can be made using mirrors or snapshots that only consume physical storage capacity as blocks are updated.

VersaStor Technology provides automatic discovery of VersaStor components for automatic VersaStor system formation. Simultaneous operation of multiple appliances with distributed workload and management is supported for scalability and availability.

Automatic fail over between appliances is provided, along with online appliance code updates. The architecture is designed to eliminate single points of failure, although failure of a VersaStor Appliance will not prevent virtual disk IO from occurring under normal conditions since the appliance is not in the direct SAN data path.

The benefits of VersaStor Technology are:

VersaStor Technology erases the boundaries between heterogeneous storage devices and protects customer investments in current storage systems. It also simplifies large-scale Open SAN deployment by obsoleting physical LUN management and replacing it with attribute-based virtual disk management that is location-transparent. This increases flexibility in tailoring the storage environment while simultaneously simplifying storage management.

Examples of virtual disk attributes include capacity, RAID protection level, server visibility, storage pool hierarchy location, performance characteristics, etc.

By providing a single point of control for the heterogeneous storage pool that can be accessed essentially anytime, anywhere, VersaStor Technology provides a consistent interface for creating, deleting, and managing virtual disks, mirrors, and snapshots. It also eliminates manual data placement tasks and allows for automated, policy-based management. As a result, complexity is reduced and administrator training requirements are minimized.

VersaStor Technology enables more efficient use of physical storage assets by eliminating stranded capacity. It also provides for increased data availability by eliminating downtime when adding capacity or migrating data, eliminating backup windows with mirrors and snapshots, and minimizing recovery time by allowing data to be recovered from snapshots rather than tapes.

7.2 VersaStor in perspective

VersaStor promises a virtualization technology that is implemented at the HBA level of the servers. It presents virtual LUNs to the host servers, and keeps the extent lists up to date. All physical LUNs at the disposal of the metadata servers are managed for extension by VersaStor. The model is asymmetrical, and illustrated in Figure 33.

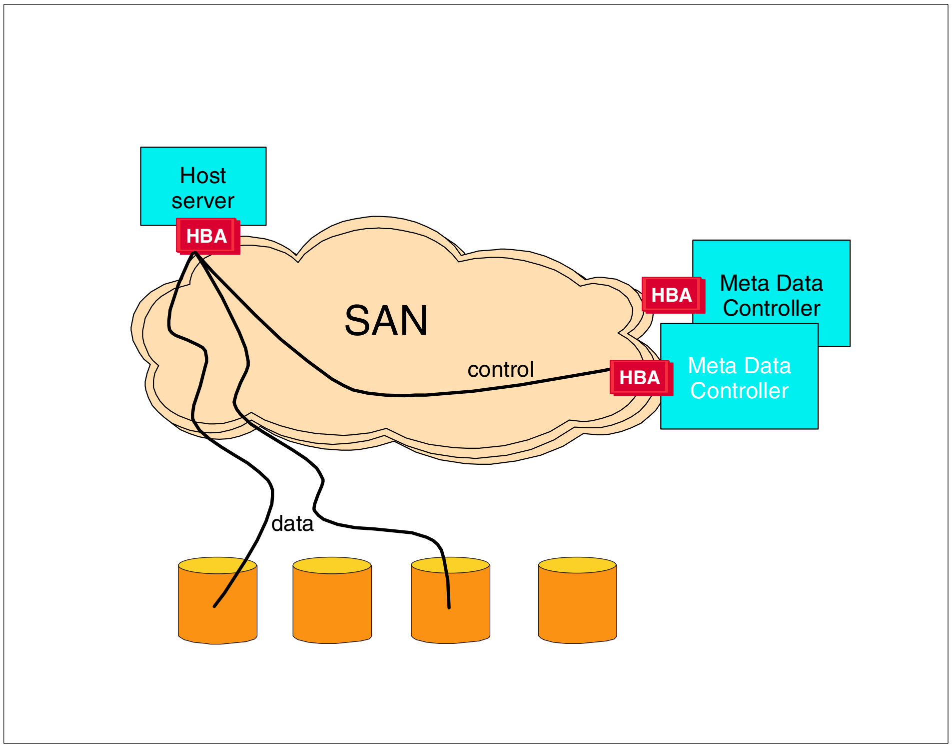


Figure 33. VersaStor technology

The metadata servers are clustered for high availability, and manage the disk configuration. When a host joins the configuration, the HBA will communicate with the metadata server to obtain information about its volumes. The agent instantiated on the HBA of the client host then maps logical LUNs to physical storage on the SAN, and it acts like a fire wall to device drives and LVM on the host. The agent on the HBA keeps the extent lists of the volumes it is allowed to see in a non-persistent way. This means that if a client reboots, it has to request his access again, and will only get those volumes that the VersaStor metadata server allows it to see. This solves the problem of one host taking all the disks in a SAN configuration, like NT is inclined to do. The protocol that runs on the client HBA insures that when a host extends its capacity, or joins the configuration, the information it uses is consistent with the metadata server and therefore with the other hosts. VersaStor eliminates the need for LVM at the host server level. It can use QoS to assign different types of storage (RAID, JBOD) to different hosts, and within those QoS definitions provide cloning or snapshot facilities for the volumes.

VersaStor will provide a good, consolidated SAN disk configuration management. The SAN fabric management is left to other software or appliances. VersaStor does not address data file sharing, however, and it doesn't do caching, since the metadata server is not in the data path. So the performance it provides for different types of data are dependent on the hardware characteristics of the disk devices it has at its disposal.

VersaStor can be working together with file system virtualization systems like Tivoli Open Sms.

Chapter 8. Tivoli Storage Network Manager

Tivoli Storage Network Manager provides the interface for managing your Storage Area Network infrastructure. Tivoli Storage Network Manager is a scalable solution utilizing open SAN standards. This comprehensive solution discovers, allocates, monitors, automates, and manages SAN fabric components and attached disk storage resources.

Tivoli Network Storage Manager is composed of the following components:

- SAN Manager will manage the SAN Fabric.
- Disk Manager assigns and unassigns SAN storage to the hosts it manages.
- File System Manager Automation will automatically extend the file system space on the hosts managed by Disk Manager.

8.1 SAN Manager

Tivoli Storage Network Manager provides you with physical and logical SAN discovery, monitoring and management.

It discovers and presents the SAN physical and logical topology, both inband and outband, and gives the administrators topology views, logical views and zone views. It does continuous real-time monitoring and fault isolation, and presents device and element configuration and state information. Tivoli Network Storage Manager can launch vendor provided management applications to operate on specific SAN elements. Tivoli Network Storage Manager stores its information in a database that can be data mined for Tivoli Decision Support, for instance.

8.1.1 SAN topology discovery

The first step in SAN management is to perform a SAN topology and disk resource discovery, producing a virtual topology map of the physical and logical components and storage resources across the SAN.

8.1.2 Components supported

Tivoli Storage Network Manager supports compliant vendors' Fibre Channel (FC) components by utilizing in-band and out-band SAN industry-standard discovery techniques including, but not limited to:

- SNMP FC management Management Information Base (MIB)

- Extended link services such as Request Node Identification (RNID) and Request Topology Information (RTIN)
- Name server queries
- Management server queries
- Selected vendor-specific interfaces

8.1.3 Continuous monitoring

Once the SAN elements have been discovered, storage resources assigned, and policies established for automation, the final step in SAN management is to do continuous monitoring of all the components within the discovered SAN topology.

8.1.4 Capacity planning and performance tuning

Capacity planning, service level planning, and performance tuning will be driven by data that is captured by Tivoli Storage Network Manager.

8.1.5 Administration

Events and data from the SAN are continuously captured and processed, providing information, alerts, and notification to administrators for problem resolution.

8.1.6 Launch applications

You can launch, from within Tivoli Storage Network Manager, the specific SAN component element management software to assist in closure of problems. This feature provides you with an easy navigational tool to launch the specific application needed to perform device and configuration management functions during initial SAN setup.

8.1.7 Single console

Tivoli Storage Network Manager integrates with Tivoli NetView allowing you to monitor and control your SAN infrastructure and devices from the same interface you use to manage your LAN and WAN. These customer networks can now be viewed from the same console.

Figure 34 shows the basic San Manager graphical view.

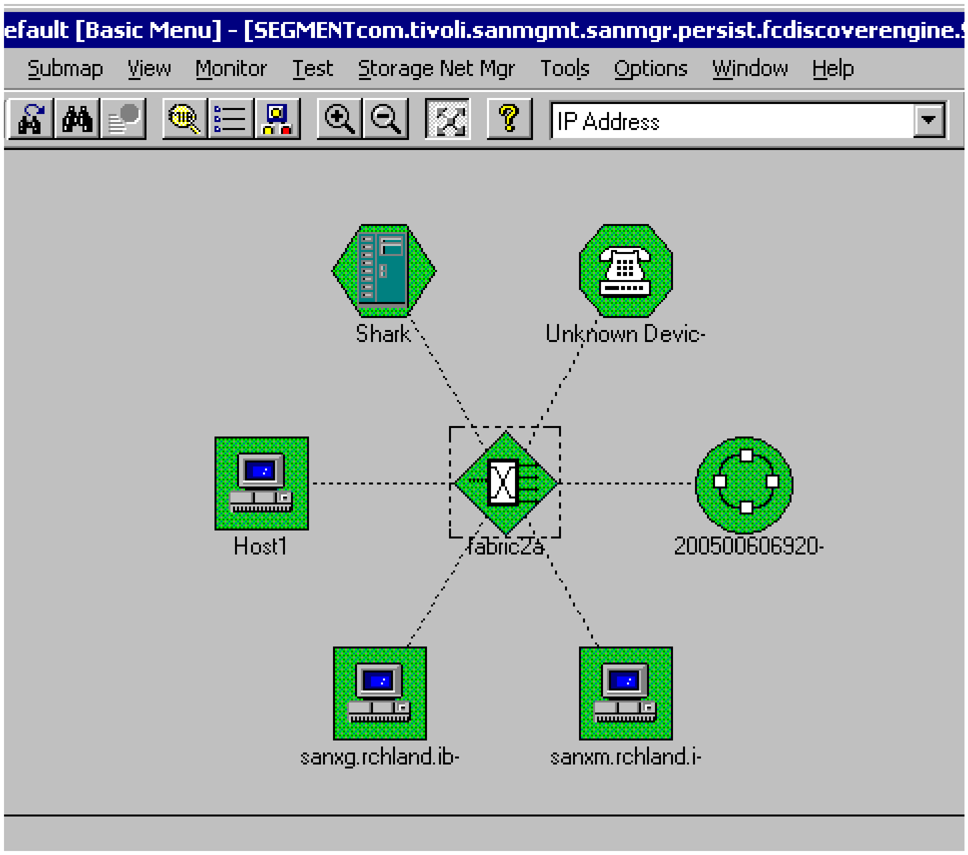


Figure 34. SAN Manager graphical view

8.2 Tivoli Disk Manager

In a SAN environment, without any kind of management, host servers via their HBAs have unlimited access to storage devices (LUNs), and most solutions that exist are vendor specific.

Tivoli Network Storage Manager LUN management discovers and attaches open system disks volumes (LUNs) in the SAN to one or more host systems. Hosts cannot access or see disks that are not attached to them. To accomplish this task, Tivoli Network Storage Manager employs multiple “LUN masking” techniques, like subsystem LUN masking (see Figure 35).

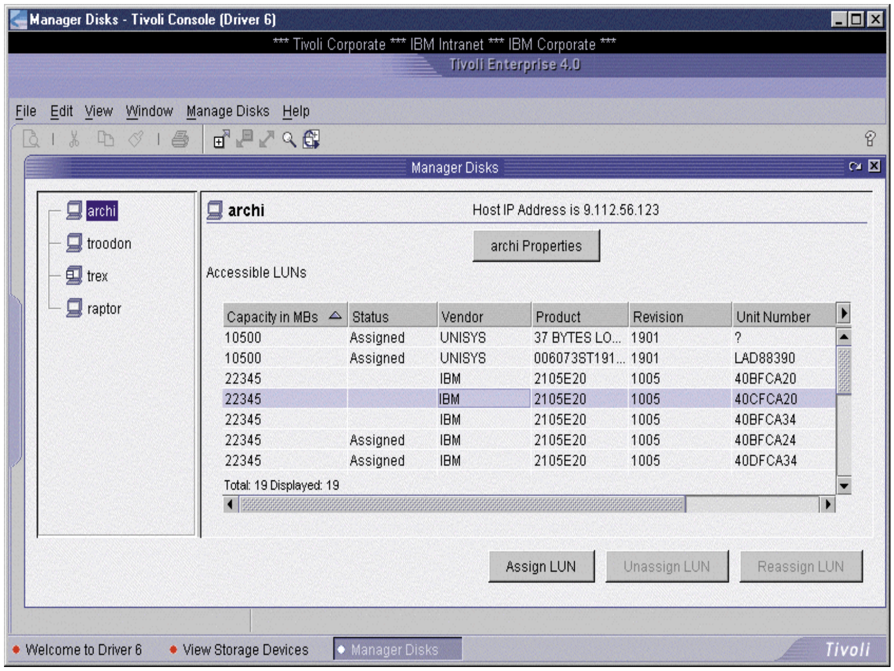


Figure 35. Disk Manager graphical view

8.2.1 Storage resource allocation

After initial discovery of the SAN topology and storage resources, the second step in SAN management is to securely allocate the discovered storage resources to the appropriate host systems for application processing. Tivoli Storage Network Manager allows you to easily assign disk storage resources or LUNs from the heterogeneous storage subsystems behind a SAN to specific computers connected to the SAN.

8.2.2 Storage resource sharing

The LUNs can then be formatted with the file system of choice and used exclusively by the assigned computer. Tivoli Storage Network Manager effectively allows multiple computers to share the same SAN resources and the same storage subsystems.

8.3 Tivoli File System Manager

This component extends the functions of Disk Manager, by providing an automated way to assign additional SAN storage to a host systems, driven by policies and based on a storage hierarchy (see Figure 36).

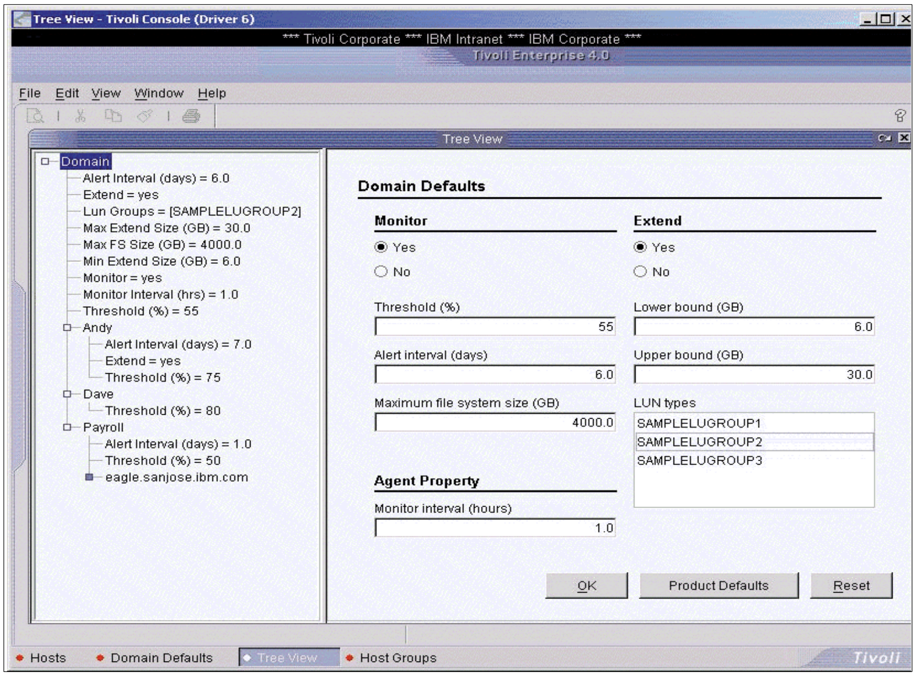


Figure 36. File Manager graphical view

8.3.1 Set policies

After discovery and assignment of storage resources, the third step in SAN management or deployment is applying policy-driven automation to the assigned disk resources. Tivoli Storage Network Manager allows you to set policies across an entire SAN, host group, specific hosts and their specific file systems.

8.3.2 Threshold monitoring

Tivoli Storage Network Manager will continuously monitor the previously assigned resources as they approach a policy-defined threshold.

8.3.3 Policy-base Automation

When the policy defined threshold is exceeded, Tivoli Storage Network Manager will automatically grow those monitored volumes by identifying and allocating an appropriate unassigned disk LUN. It will then extend the supported file system over this additional allocated space.

This unique automation capability of allocating additional LUNs and extending supported file systems can greatly reduce redundant administrative tasks and maintain continuous application processing

8.4 Event reporting

Events and data from the SAN are continuously captured and processed, providing information, alerts, and notification to administrators for problem resolution. SAN-related events are forwarded to SNMP and/or Tivoli Event Console (TEC) to report on SAN activities to the designated administrator or management console (see Figure 37).

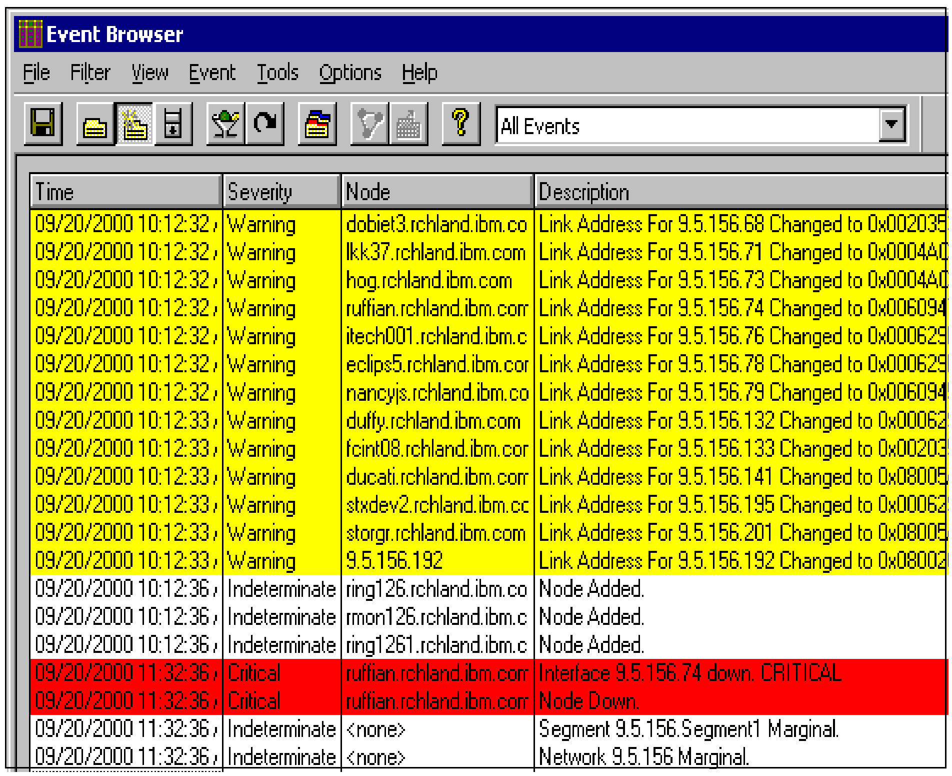


Figure 37. Event report graphical view

8.5 Others

Tivoli Storage Network Manager is a key component of the overall Tivoli Storage Solution portfolio, providing comprehensive SAN and storage management. Tivoli Storage Network Manager can operate stand-alone or integrate with: Tivoli NetView; Tivoli Enterprise Console; Tivoli Decision Support for SAN Resource Management.

8.5.1 Tivoli Netview

When used with Tivoli NetView, you can monitor and control your SAN infrastructure and devices from the same console that you use to manage your LAN and WAN.

8.5.2 Tivoli Enterprise Console

SAN-related events are forwarded to Tivoli Enterprise Console and/or through SNMP to SNMP management software.

8.5.3 Tivoli Decision Support for SAN Resource Management

Tivoli intends to provide a Decision Support Guide for reporting and analysis of data from the Tivoli Storage Network Manager. Decision support guides are a set of “best practices" guides that can be used to analyze and display data about applications and products. The information is presented in a variety of text and graphical formats, allowing a user to drill down to get details of a particular aspect of an environment. Tivoli Decision Support for SAN Resource Management will allow its users to make business decisions based on the inventory, capacity and usage of SAN resources, as well as, threshold monitoring. This statement of direction is based on Tivoli's current development plans, and is subject to change without prior notice.

8.6 Product requirements and devices supported

The Tivoli Storage Network Manager Server is supported on Windows 2000. The managed host platforms are supported on Windows NT, Windows 2000, IBM AIX, and Sun Solaris.

8.6.1 Product requirements

Summarized below are the general requirements in order to run Tivoli Storage Network Manager. The console and manager can be run on the same machine or on separate machines (Table 2).

Table 2. Product requirements

	Console	Manager	Managed Host
Machine	Minimum Pentium III 400	Minimum Pentium III 600	
Minimum disk space required to run	See the product README file	See the product README file	See the product README file
Operating system	Windows 2000 Professional, Server, or Advanced Server Editions	Windows 2000 Advanced Server Edition	Windows NT 4.0 Server, Windows 2000 Professional, Server, or Advanced Editions, AIX 4.3.3, and Solaris 2.7
Memory	Minimum 256 MB	Minimum 1 GB	Minimum 100 MB

8.6.2 Devices supported

Tivoli Storage Network Manager supports standards-compliant vendors' FC components.

Several Tivoli Ready Business Partners are working to extend the capabilities of Tivoli Storage Network Manager to customers by providing integrated tools and products.

8.7 Tivoli SAN Integrated Partners

Tivoli Storage Network Manager supports standards-compliant vendors' FC components.

Several Tivoli Ready Business Partners are working to extend the capabilities of Tivoli Storage Network Manager to customers by providing integrated tools and products.

The following companies are currently working with the Tivoli Ready certification lab and plan to announce integrated solutions in the fourth quarter of the year 2000:

- Crossroads Systems
- INRANGE Technologies
- McDATA
- Pathlight
- Qlogic
- Emulex

Each of these companies are developing third-party products that will meet the standards set forth in Tivoli Storage Network Manager and will offer easier and faster deployment times to Tivoli customers.

Appendix A. Special notices

This publication is intended to help storage personnel to help understand the concept of storage networking virtualization. The information in this publication is not intended as the specification of any programming interfaces that are provided by IBM. See the PUBLICATIONS section of the IBM Programming Announcement for products mentioned in this redbook 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 use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

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

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

IBM ®	Redbooks
RS/6000	Redbooks Logo
S/390	Tivoli Open SmS
DFSMS	e (logo)®
System 390	

The following terms are trademarks of other companies:

Tivoli, Manage. Anything. Anywhere.,The Power To Manage., Anything. Anywhere.,TME, NetView, Cross-Site, Tivoli Ready, Tivoli Certified, Planet Tivoli, and Tivoli Enterprise are trademarks or registered trademarks of Tivoli Systems Inc., an IBM company, in the United States, other countries, or both. In Denmark, Tivoli is a trademark licensed from Kjøbenhavns Sommer - Tivoli A/S.

VersaStor is a trademark of Compaq in the United States and/or other countries.

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

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

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

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

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

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

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

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

Appendix B. Related publications

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

B.1 IBM Redbooks

For information on ordering these publications see "How to get IBM Redbooks" on page 101.

- *Designing an IBM Storage Area Network,* SG24-5758
- *Introduction to Storage Area Network, SAN,* SG24-5470
- *Planning and Implementing an IBM SAN,* SG24-6116
- *AIX Storage Management,* GG24-4484

B.2 IBM Redbooks collections

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

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

B.3 Referenced Web sites

These Web sites are also relevant as further information sources:

- `http://www.ibm.com/san/`
- `http://www.storage.ibm.com/ibmsan/products.htm#software`
- `http://www.tivoli.com/products/solutions/san/`
- `http://www.datacore.com/`

How to get IBM Redbooks

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

- **Redbooks Web Site** **ibm.com**/redbooks

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

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

- **E-mail Orders**

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

	e-mail address
In United States or Canada	pubscan@us.ibm.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 Redbooks 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

agent. A hardware or software component installed on a host to virtualize its view of storage or to execute a specific storage management function for that particular host.

allocation. The entire process of obtaining a volume and unit of external storage, and setting aside space on that storage for a data set.

array. An arrangement of related disk drive modules that you have assigned to a group.

backup. A copy od 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 can be used to recreate data that has been lost, mislaid, corrupted or erased erased.

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.

client. (1)A function that requests services from a server, and makes them available to the user. (2)A term used in an environment to identify a machine that uses the resources of the network.

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.

CLVM. Concurrent Logical Volume Manager.

Data Sharing. A SAN solution in which files on a storage device are shared between multiple hosts.

DFSMS. Data Facility Storage Management Subsystem

ESS. Enterprise Storage Server - Provides an intelligent disk storage subsystem for systems across the enterprise.

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.

failover. The assumption of server responsibilities by the node designated as backup server when the primary server fails

FC. Fibre Channel.

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 controller s and drives.

gateway. In the SAN environment, a gateway connects two or more different remote SANs with each other. The gateway converts server protocols to endpoint protocols, and vice versa. A gateway can also be a server on which a gateway component runs.

hardware zoning. Hardware zoning is base on physical ports. The members of a zone are physical ports on the fabric switch. It can be implemented in the following configuration: one to one, one to many, and many to many.

HBA. Host Bus Adapter

host. Any system that has at least one internet address associated with it. A host with multiple network interface can have multiple internet addresses associated with it. Used interchangeably server.

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 loop.

IBM. International Business Machines Corporation.

IFS. Installable File Systems - A subsystem of Tivoli open-SmS client’s file system that directs al metadata operations to a metadata server, and directs all data operations to storage devices attached to the Fibre Channel network.

I/O. Input/Output

IP. Internet Protocol

iSCSI. The transfer of SCSI command set over I/P networks.

ITSO. International Technical Support Organization.

JBOD. Just a bunch of disks

Kerberos. A service for authenticating users in a network environment.

LAN. 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.

LSA. Log Structured Array

LUN. LUN is a SCSI TLA (Three Letter Acronym) which expands to Logical Unit Number. The SAN industry tends to use LUN to mean a Logical Unit. But a Logical Unit may be visible on several paths, and have a different Logical Unit Number on each path. We use LUN to mean a Logical Unit.

LUN masking. LUN masking allows or prevents I/O to the disk drives through the host bus adapter device or operating system device driver. Intelligent disk subsystems like the Enterprise Storage Server also provides subsystem LUN masking.

LVM. Logical Volume Manager

metadata. Data structures that contain access information about file data. These might include i-nodes, indirect blocks, and directories.

MDC. Meta Data Controller

MSS. Modular Storage Server

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.

network. An aggregation of interconnected nodes, workstations, file servers, and/or peripherals, with its own protocol that supports interaction.

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.

POSIX. Portable Operating System Interface for Computer Environment; an IEEE operating system standard, closely related to the UNIX system (software writing).

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

QoS. Quality of Service - 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.

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.

RVA. RAMAC Virtual Array

SAN. A Storage Network (SAN) is a dedicated, centrally managed, secure information infrastructure, which enables any-to-any interconnection of servers and storage systems.

scalability. The ability of a computer application or product 9 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 interface 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.

server. A computer which is dedicated to one task.

shared storage. Storage within a storage facility that is configured such that multiple homogenous or divergent hosts can concurrently access the storage. The storage has a uniform appearance to all hosts. the host programs that access the storage must have a common model for the information on a storage device.

SNS. Simple Name Server

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.

tape backup. Making magnetic tape copies of hard disk and optical disc files for disaster recovery.

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.

Tivoli Open-SmS. Tivoli Open System-managed Storage - the tivoli product that bring efficiencies and capabilities to open systems that have, until now, only been available in the highly scalable, mission critical environments hosted by OS/390.

UFS. Unix File System

VersaStor. The Compaq technology that provides SAN-wide storage virtualization

VTS. Virtual Tape Server

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.

WWN. World Wide Name

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.

Index

Z

IBM Redbooks review

Your feedback is valued by the Redbook authors. In particular we are interested in situations where a Redbook "made the difference" in a task or problem you encountered. Using one of the following methods, **please review the Redbook, addressing value, subject matter, structure, depth and quality as appropriate.**

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

Document Number **Redbook Title**	SG24-6210-00 Storage Networking Virtualization: What's it all about?
Review	
What other subjects would you like to see IBM Redbooks address?	
Please rate your overall satisfaction:	O Very Good O Good O Average O Poor
Please identify yourself as belonging to one of the following groups:	O Customer O Business Partner O Solution Developer O IBM, Lotus or Tivoli Employee O None of the above
Your email address: The data you provide here may be used to provide you with information from IBM or our business partners about our products, services or activities.	O Please do not use the information collected here for future marketing or promotional contacts or other communications beyond the scope of this transaction.
Questions about IBM's privacy policy?	The following link explains how we protect your personal information. ibm.com/privacy/yourprivacy/

IBM Redbooks

Storage Networking Virtualization: What's it all about?

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

Storage Networking Virtualization

What's it all about?

Is your data real, logical, or virtual?

Should you use a SAN, a NAS or something new?

Are we using mirrors, myth or magic?

This IBM Redbook gives a broad understanding of a new architecture for storage networks, using virtual mapping of data files onto a network of storage devices. The concept and reality of virtual data has been around for many years in the mainframe environment, but now we are entering a new era of virtualization with the design, architecture, and implementation of a new set of virtual products to help all installations make efficient use of IT storage resources.

This new approach will revolutionize the way data is stored and managed in the UNIX/NT environment, similar to the way that systems managed storage revolutionized storage management for the mainframe.

These emerging products integrate into existing configurations to raise efficiency and utilization.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-6210-00 ISBN 0738421367