

# PRIMARY STORAGE CONCEPTS AND LIBRARY ATTACHMENT METHODS

---

TECHNICAL WHITE PAPER

# TABLE OF CONTENTS

---

<b>FORWARD</b>	1
<b>DATA APPLICATIONS USING PRIMARY STORAGE</b>	2
<b>DATA ACCESS LEVELS</b>	2
<i>BLOCK LEVEL</i>	2
<i>FILE LEVEL</i>	2
<b>DISK DRIVE OVERVIEW</b>	3
<i>INTERNAL VERSUS EXTERNAL DISK</i>	4
<b>JBOD OVERVIEW</b>	4
<b>RAID OVERVIEW</b>	5
<i>STRIPING</i>	5
<i>REDUNDANCY</i>	6
<i>RAID 0</i>	7
<i>RAID 1</i>	8
<i>RAID 3 – RAID 4</i>	8
<i>RAID 5</i>	9
<i>RAID 10</i>	9
<b>HOST SYSTEM ATTACHMENT METHODS</b>	10
<i>SAS/DAS</i>	10
<i>Backup Methods</i>	10
<i>NAS</i>	11
<i>Servers</i>	11
<i>Backup Methods</i>	12
<i>LAN-free Backup</i>	13
<i>NAS Boxes</i>	13
<i>Backup Methods</i>	15
<i>NDMP</i>	17
<i>Future Overland NDMP Support</i>	17
<i>SAN</i>	18
<i>Enterprise Class Fibre Channel RAID subsystems</i>	18
<i>SAN Vision</i>	19
<i>SAN Reality</i>	20
<i>IPSAN</i>	21
<i>Server Clustering</i>	22
<i>HA (Highly Available) SANs</i>	23
<i>SAN Backup Methods</i>	24
<i>SAN Island Backup</i>	25
<i>Serverless Backup</i>	26
<i>SAN NAS CONVERGENCE</i>	28
<b>APPENDIX A – HAVING FUN WITH EXCLUSIVE OR'ING</b>	29



# FORWARD

---

**This Primary Storage Choices document is intended to provide an overview of the choices available and the tradeoffs involved for:**

- Users and Managers of Information Technology resources – particularly storage
- Overland Sales and Technical team members
- Overland Executives and Managers

# DATA APPLICATIONS USING PRIMARY STORAGE

---

*Data applications using primary storage consist of two types, Data Creation and Data Manipulation applications.*

Data applications using primary storage consist of two types, Data Creation and Data Manipulation applications.

Loosely defined, a data creation application is any program that generates data in the form of files, PowerPoint, Word, Excel, Etc. The vast amount of enterprise storage is occupied with data that has been created.

Data manipulation applications are your classic database type programs where data is rendered in a fixed table size format. If the Database program's file system is a journaling type, than it too can be considered a data manipulation and creation application.

## DATA ACCESS LEVELS

---

*Modern systems can deliver data as bricks (blocks) or something made from brick, a wall (file).*

Electronic data is data. How it is referenced and packaged determines if it is a "brick" (block) or a brick wall (file made up of one or more blocks). Modern systems can deliver data as blocks (bricks) or something made from bricks, such as a wall (file).

### **BLOCK LEVEL**

Block level storage refers to the ability to directly access a block of data without having to access previous blocks to get to the content on the target block. The term "block" is short for logical block, which refers to the smallest directly addressable storage entity within a storage subsystem. Disk drives write and read data in blocks.

### **FILE LEVEL**

A file is a block or blocks of data that have specific association with each other and are addressable under a file system as a single object. Files consist of two parts, the file data and the Meta data that describes the file (i.e. filename, creation date, etc). The "data" portion is made up of one or more logical blocks written to disk. The file system uses indexing to organize files within its structure of root directory, directories and files.

The Meta data of a file is typically stored in the directory structure to identify the file. Part of the Meta data for a file is the file length and the logical block address of where the data is stored. In essence, file systems map directories and files with their associated Meta data and content data to the storage system's logical blocks.

### **PRIMARY STORAGE ARCHITECTURE**

Primary storage for the last 30 or so years has used DASD (Direct Accesses Storage Device) or in other words, Disk Drives as primary storage. Over the years the form factors and interfaces have changed dramatically but the purpose of primary storage has remained the same; to write and to read data on

command. In the last few years we have seen the criticality of access to data grow tremendously. Availability to data and business continuance has come to mean almost the same thing. At the heart of it all are the lowly disk drives writing and reading data at our command.

## **DISK DRIVE OVERVIEW**

---

Disk drives are electromechanical storage devices that write and read data magnetically to disk platters contained within the drive. Key parameters for disk drives are:

Capacity	Interface
Rotational Speed	Form factor
Seek Time	Reliability

Capacity for disk drives continues grow at close to a 50% per 18 month rate. Pixie dust and other technical advances continue to allow disk vendors to put more and more data into the same physical space. Current technology advances will allow the capacity growth rate to continue at least for the next 3 years. State of the art 3.5 inch disk drives are at 180 Gbytes, with a single spindle expected to reach over 600 Gbytes by late 2004. When Seagate's recent HARM recording technology comes to full fruition, long term we could be looking at a 2 inch disk drive holding the content of the Library of Congress.

Rotational Speed measured in RPMs (Rotations Per Minute) has also seen a dramatic four fold increase over the last 10 years from sub 4000 RPM to the current state of the art 15,000 RPM drives. Changes in RPM rate directly effect the Mbytes per second. A 30% increase in RPM rate also offers a 30% improvement sustainable data rate. Current 15,000 RPM drives are capable of better than 45 Mbytes per second each.

Seek Times have also improved over the last 10 years from <30 milliseconds to <6 milliseconds. RAID controllers and their large caches have also reduced the impact of seek times.

Interfaces for disk drives come in three main varieties, ATA/IDE, SCSI and Fibre Channel. The ATA/IDE is an interface that is an extension of the original IBM PC's Disk controller's (WD's) register set. Current improvements bring the ATA interface up to 100 Mbytes per second burst rate capable. Parallel SCSI has also evolved to 320 Mbytes per second for a burst rate. Fibre Channel last year also boosted its burst data rate to 200 MB/second.

Form factor (for the enterprise disk drive) has standardized on the 1 inch high 3.5 inch form factor as the standard. Fibre Channel drive vendors have also standardized on location and type of connector making life simpler for the JBOD and RAID array vendors allowing them to choose between several different vendors offerings that physically are the same.

Power, cooling and vibration are all issues that storage subsystem suppliers have to deal with. Vibration and cooling are becoming critical issues for high-density applications like JBOD and RAID Arrays.

*Key parameters for disk drives are:*

*Capacity  
Interface  
Rotational Speed  
Form factor  
Seek Time  
Reliability*

# DISK DRIVE OVERVIEW (cont.)

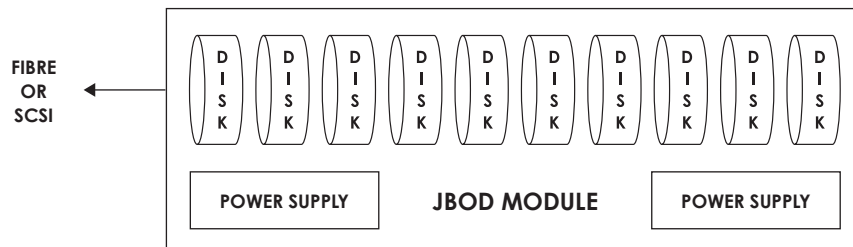
*A key trend in storage for medium sized data centers has been toward external (not within the server chassis) disk storage.*

## INTERNAL VERSUS EXTERNAL DISK

A key trend in storage for medium sized data centers has been toward external (not within the server chassis) disk storage. This trend has been developed for two reasons; first, when disks are external from the server they are typically designed for hot swap capabilities. Second, by de-coupling the server from the storage you allow the customer to negotiate the best deal for server and storage independently. Internal storage is still used today in most servers as a boot device. Systems requiring more than 4 disk drives of capacity tend toward external storage such as JBOD or RAID arrays.

# JBOD OVERVIEW

JBOD (Just a Bunch of Disk) is the true corner stone of today's enterprise class storage for both SAN and NAS. All large-scale primary storage systems use JBOD as the final resting-place for data.



*JBOD Module Block Diagram*

In JBOD configurations 10 to 14 (depending upon vendor) 3.5 inch form factor Fibre Channel or SCSI disk drives are individually mounted on sleds that are installed into the front of an JBOD enclosure. The JBOD enclosure provides daisy chain connectivity for the disk drives as well as optional redundant power and cooling.

Connectivity for the JBOD is accomplished with either LVD SCSI or Fibre Channel. In the case of Fibre Channel disk JBOD, the disk are dual ported, hence the enclosure is also dual ported. The type of Fibre Channel used at the JBOD level is FC-AL (Fibre Channel – Arbitrated Loop). The dual Fibre port design of the Fibre drives allows for redundant paths of access to improve fault tolerance overcoming single points of failure. To a system looking down the SCSI or dual Fibre pipes the disks will appear as individual disk in sequential address locations.

As the true building block of enterprise storage JBOD in low-end cases is attached directly to a server running RAID software. In other cases NAS heads provide front-end connectivity to host systems as well as RAID protection for JBOD.

ATA/IDE drives are creeping into the JBOD market where an intelligent mid-plane maps ATA/IDE drives to what appear as SCSI or Fibre Channel drives to the system.

# RAID OVERVIEW

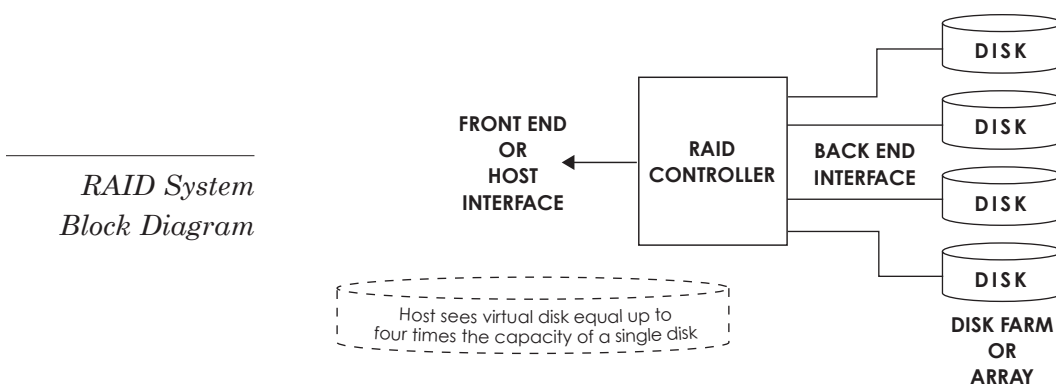
RAID (Redundant Array of Independent Disk) approaches used today are based on a research white paper originally done at UC Berkley. The purpose behind the research was to demonstrate ways in which less expensive disks could be configured to appear as a larger more reliable disk.

The initial papers defined RAID 0, 1, 2, 3, 4, and 5. Since then the RAID types have greatly expanded with several companies doing subtle twists to the basic scheme and calling it RAID 10, X, Y and Z.

The following Block diagram is a simplified block diagram of a RAID system. A detailed block diagram of an enterprise class array can be found in the Enterprise Class Fibre Channel RAID subsystems section.

The need for inexpensive, large capacity disks drove the storage industry to RAID for the virtualization aspect. By applying software or firmware RAID policies against an array of disks also called a disk farm, a RAID controller can aggregate both capacity and performance of the individual disks into what appears to the host as one very large high performance disk.

*The need for inexpensive large capacity disks drove the storage industry to RAID for the virtualization aspect.*



*RAID System Block Diagram*

Key to the RAID functionality is disk striping. In Disk Striping chunks (sequential groups of sectors) of individual disk are mapped as consecutive logical blocks of the aggregated Virtual disk. Capacity and Performance of the virtual disk are now the aggregation of the entire disk farm.

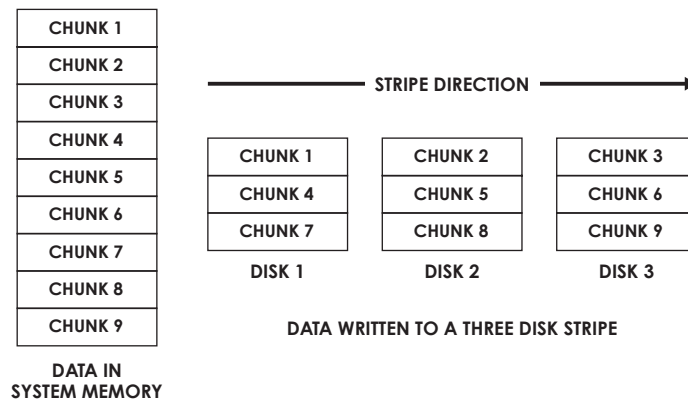
A major trend in the RAID world today is toward the use of ATA/IDE disk drives for the disk farm. Vendors and customers are beginning to recognize that most data they deal with is "created" data and not manipulated data that requires incredibly fast access times. Several companies are delivering ATA/IDE based RAID systems for this content market. Expect this trend to continue. Serial ATA will increase the flexibility of packaging for disk array vendors. Host interfaces will continue to be Fibre, SCSI and in 2003 iSCSI for block level access over Ethernet.

## STRIPING

In the following examples, three disks are striped together. When the host system writes data blocks to the RAID controller the data is distributed across the disks based upon the RAID policy and the chunk size parameters that were defined at the time of the virtual disk's bind (creation) time. Chunk size is also referred

# RAID OVERVIEW (cont.)

to as stripe size or depth and is measured in K bytes. Typical stripe sizes are 32K to 256K bytes in depth. Stripe size is one of the parameters that can be tuned on some RAID arrays to maximize performance to a given application.



In essence the RAID controller re-maps the logical blocks contained by the individual disk drives into a larger logical or "virtual" drive based on the RAID policies and presents this unified logical device to the front-end interface.

*The other key advantage that RAID brought to storage is redundancy. With the exception of RAID 0, all other RAID types offer protection against hard read errors or complete drive failures.*

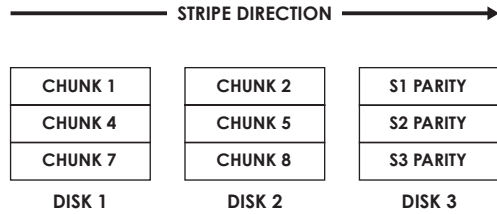
## REDUNDANCY

The other key advantage that RAID brought to storage is redundancy. With the exception of RAID 0, all other RAID types offer protection against hard read errors or complete drive failures. The need for redundancy is driven by the reliability factor of disks. Annual Failure rates for 3.5" disks have hovered around 3-4 % for the last few years. In the real world this means that for every 20 disks you have in house, the odds are high that one will fail every 18 months.

Two types of redundancy are defined in the Berkley papers. The first is simple mirroring. Any data written to one disk is also written to its partner disk. This approach is used for both RAID 1 and RAID 10.

The second type of redundancy uses what are known as parity sectors or chunks. In this architecture data written within a stripe is exclusive OR'ed together to create the parity chunks against the entire stripe. If data becomes unrecoverable by one of the physical drives within the virtual drive; the RAID controller will use the recoverable data within the stripe along with the parity data for that stripe to regenerate the unrecoverable data. This approach can be applied to a single bad sector on a disk or to an entire disk failure.

*Example Three Disk  
Stripe with Parity*



In the above 3 Disk example data chunks 1 and 2 are exclusively Or'ed together to create Stripe one's (S1) parity chunk written on disk 3. Likewise the data of chunks 4 and 5 are exclusively Or'ed to create the S2's (stripe 2) parity data and so on across all stripes with in a LUN.

Mirroring or Parity are the mechanism used by RAID technology to protect against disk failures. Both Mirroring and Parity will protect the data against single faults with in a stripe. It can not protect against multiple drives within the same LUN failing. Newer RAID types now support multi drive failure within the same LUN (at the expense of capacity).

For those of you who would like to experiment with Exclusive Or'ing, see appendix A for a further description of Exclusive Or'ing with an example.

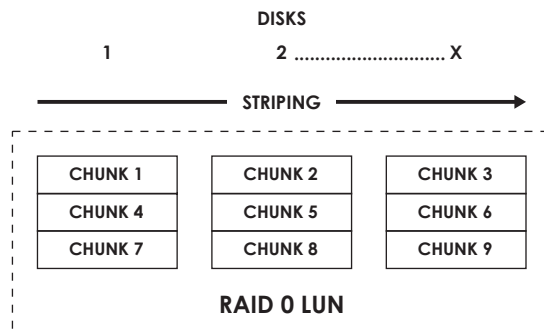
**RAID 0**

RAID 0 provides the best bandwidth by striping data across multiple drives but offers no redundancy to protect against hard read errors or drive failures. RAID 0 is used where performance is higher priority than data recoverability.

*RAID 0 is used where performance is higher priority than data recoverability.*

RAID 0 is used in temporary high speed caching applications and some video editing applications. The vulnerability to hard read errors or complete drive failure causing LUN failure limits the use of RAID 0.

$$\text{RAID 0 LUN capacity} = (\text{Disk Capacity}) \times (\# \text{ of Disk})$$



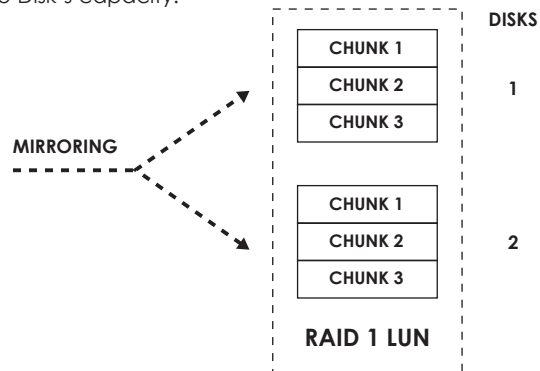
# RAID OVERVIEW (cont.)

*RAID 1 provides redundancy by simply mirroring data across two drives at the expense of capacity (50%) that is used as redundancy to protect against hard read errors or drive failure.*

## RAID 1

RAID 1 provides redundancy by simply mirroring data across two drives at the expense of capacity (50%). It is used as redundancy to protect against hard read errors or drive failure.

RAID 1 LUN capacity = one Disk's capacity.



RAID 1 is often used by entry level or host software based RAID systems. In the case of a drive fault or failure the RAID control function requests the mirrored copy of the unrecoverable data from the second drive and returns that data to the requesting application.

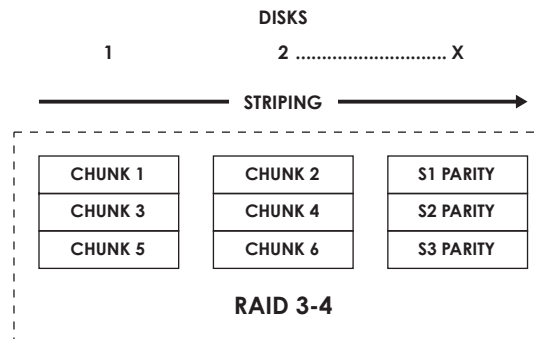
*RAID 3 & 4 provide the striped performance advantages of RAID 0 and redundancy without the capacity penalty of mirroring.*

## RAID 3 – RAID 4

RAID 3 & 4 provide the striped performance advantages of RAID 0 and redundancy without the capacity penalty of mirroring by creating Exclusive OR'ed parity chunks that provide data regeneration capabilities. The Parity drive will become a performance bottleneck for applications that are chunk rewrite intensive. Both RAID 3 & 4 are ideal for loss-less rich media and large file storage.

True RAID 3 is sub-sector level striping with a fixed parity drive. RAID 4 uses stripes at or above the sector size along with a fixed parity drive.

RAID 3 or 4 LUN capacity = (Disk Capacity X # of Disk) - 1 Disk Cap



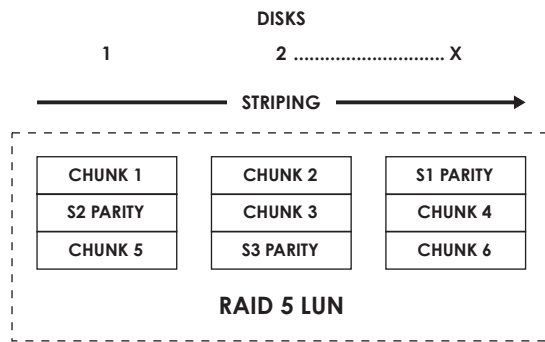
In the above illustration the blocks marked as S1, S2 and S3 are the parity chunks for each stripe 1,2 and 3.

**RAID 5**

RAID 5 provides the striped performance advantages of RAID 0 and redundancy without the capacity penalty of mirroring by creating Exclusive OR'ed parity chunks that provide data regeneration capabilities. The Parity drive bottleneck issue of RAID 3 & 4 is eased by distributing the Parity chunks across all disks. Applications that are rewrite intensive will suffer from some write latency due to Parity generation and rewrite as part of a rewrite operation. RAID 5 is the workhorse of most large databases.

$$\text{RAID 5 LUN capacity} = ((\text{Disk Capacity}) \times (\# \text{ of Disk})) - 1 \text{ Disk Cap}$$

*RAID 5 provides the striped performance advantages of RAID 0 and redundancy without the capacity penalty of mirroring.*



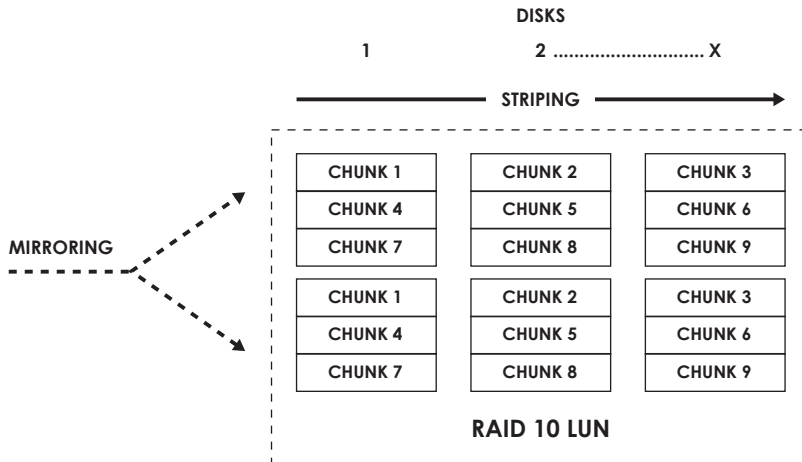
In the above illustration the blocks marked as S1, S2 and S3 are the parity chunks for each stripe 1,2 and 3.

**RAID 10**

RAID 10 provides the performance advantages of RAID 0 and the protection of RAID 1. RAID 10 is used for applications that require high bandwidth, redundancy, and low write latencies. As with mirroring a fifty-percent capacity premium is paid to have full mirroring. Very high performance databases with heavy transactional (write activity) environments are classic candidates for RAID 10's high performance, high cost market.

$$\text{RAID 10 LUN capacity} = (\text{Disk Capacity}) \times (\# \text{ of Disk}) \times 50\%$$

*RAID 10 provides the performance advantages of RAID 0 and the protection of RAID 1.*



# HOST SYSTEM ATTACHMENT METHODS

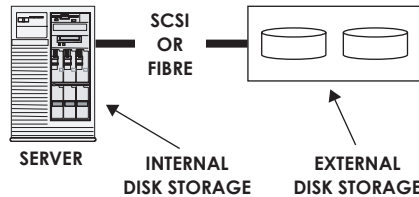
The two camps of attachment for primary storage are Server Attached Storage (SAS) and Network Storage (NS).

In the SAS configuration storage is directly attached to the host system and is made only available to the host system and its applications.

Two camps of attachment for primary storage exist today. They are SAS (Server Attached Storage—also called DAS (Direct Attach Storage)) and Networked Storage (NS). Each of attachment method has its advantages and its disadvantages. The current trend for enterprise storage is toward Networked Storage using either Ethernet or Fibre Channel to connect a storage system to multiple hosts for sharing purposes.

## SAS/DAS

SAS (Server Attached Storage) also referred to as DAS (Direct Attach Storage) has been around since the first computer. In this configuration storage is directly attached to the host system and is made only available to the host system and its applications. SAS storage as shown in the picture below can be both internal drives within the server or an external storage box.

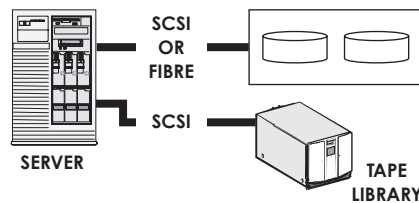


Server Attached Storage (SAS)

The key limiting factor to SAS is that data stored within the host can not be shared by other systems. The host along with its storage can be placed on a LAN for data sharing purposes but this burdens the host with file sharing request that can make overall performance less than ideal. Especially if the host has other duties beyond file sharing.

## Backup Methods

The backup method for SAS is a direct attach tape drive or library using SCSI to the host. Also see the Server section coming up for other ways to backup a host (server).



Server Attached Storage (SAS) Backup

Server Attached Storage is by far the most pervasive storage in the industry today. Over the next year SAN and NAS storage will finally exceed SAS in \$\$\$ shipments to the field.

## NETWORKED STORAGE OVERVIEW

Currently there are two main approaches to what is referred to as "networked storage." The first approach is using a server or NAS (Networked Attached Storage) head to provide file level serving functionality to an Ethernet LAN. With the advent of the NAS box or NAS "Appliance," deployment became simple and streamlined. MIS departments found it easy to just throw another NAS box at other department's storage problems. With a few boxes NAS storage was as simple as plug, and play compared to SAN based storage.

A second form of network storage hit the scene in 1996 with the introduction of the Fibre Channel to SCSI RAID systems in the enterprise storage space. Fibre Channel (a high-speed serial form of SCSI) enabled much greater distances and interconnectivity than SCSI. Now server clustering with shared disk space had plumbing capable of supporting many more (16 million) nodes than four host nodes and four storage nodes, as was the case with SCSI. With all this added capability it did not take someone long at Compaq to spin the term SAN (Storage Area Network). Hence the second form of networked storage found in the enterprise market space gained an acronym.

## NAS

Network Attached Storage has come to be defined as storage that is accessible remotely over Ethernet. A network file system such as NFS or CIFS runs on top of the Ethernet topology. In most cases the "Data" is accessed at the file level providing both completeness and context to the data. This completeness and context as well as the content (data) are a key reason why NAS storage systems offer the best path for virtualization.

NFS (Network File System) has been popularized in the Unix market space. CIFS (Common Internet File System) is popular in the Windows market space as the Ethernet protocol of choice. Both provide the basic file system functionality of reading, writing, creating and deleting files and directories.

NAS can be supported in one of two hardware/software ways. A general-purpose server can be configured to supply file level storage for a LAN using it's own direct attach storage. A more in vogue solution recently has been the introduction of the Appliance, or NAS box or Filer. Here, software is pre-installed and storage configuration is semi automatic and simple.

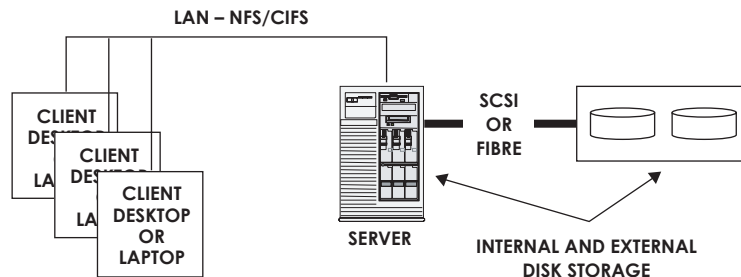
### Servers

Windows and Unix servers have been the main stay in the client / server LAN relationship for 15 plus years. A server, by definition supports at least one task. In the case of general purpose servers many task are supported. Including: File Serving, and Print Serving, In the NAS application the server provides the "file" serving capability running one or several Ethernet based file protocols such as NFS or CIFS.

*There are two main approaches to "networked storage": Network Attached Storage (NAS) and Storage Area Network (SAN).*

*Network Attached Storage (NAS) has come to be defined as storage that is accessible remotely over the Ethernet.*

*Classic Server Configuration for NAS Application*

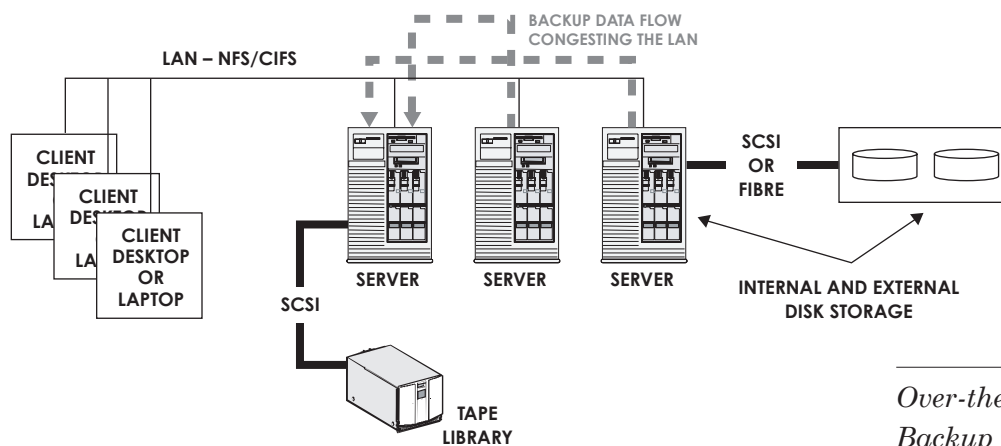


# HOST SYSTEM ATTACHMENT METHODS (cont.)

*The Over the LAN backup solution provides a cost effective means to backup two or three servers.*

## Backup Methods

The backup method of choice for this class of machine is typically a direct attach SCSI loader or library. In sites where multiple servers exist, several backup options exist including over the LAN and LAN-free configurations. The following is an example of over the LAN.



*Over-the-LAN Backup*

The Over-the-LAN backup solution provides a cost effective means to backup two or three servers. Depending on the backup criteria (MB/sec to meet backup window requirements) even a simple three-server configuration can bring a 100 BaseT network to its knees while doing backups.

To extend the life of a LAN based backup system one should consider:

- Upgrading to a switched 10/100 BaseT Environment
- Implementing a second dedicated LAN for backup traffic only
- Upgrading to a GigE backbone between the servers

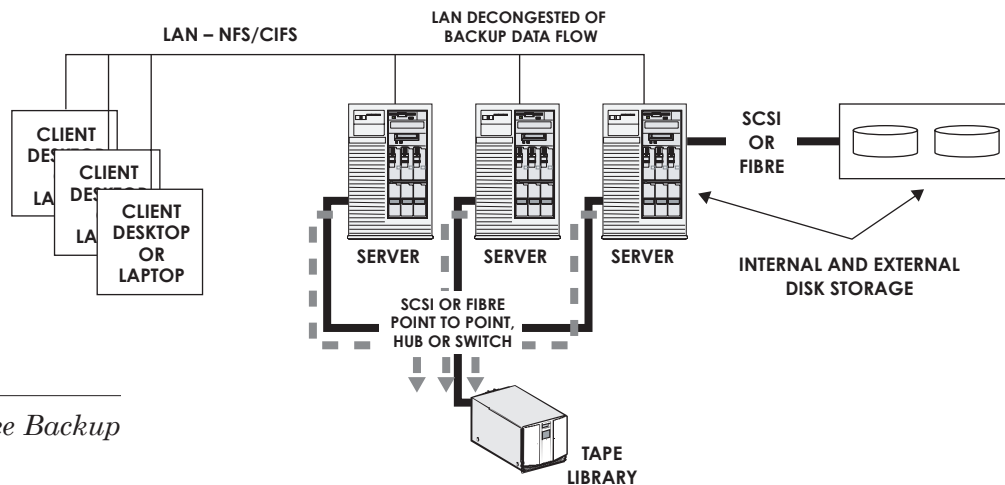
The costs for each approach will vary between \$200-\$800 dollars. Upgrading to a GigE backbone for \$800 will give you the bandwidth to support SDLT or LTO technology drives with their heavier bandwidth demands.

As backup data sets grew it became apparent to people that LAN based backup not only crippled LAN use during backup periods but the sheer volume of data could no longer be moved within the backup window due to the 8 Mbytes/second limitation of 100BaseT.

### LAN-free Backup

LAN-free backup was created to solve this very issue. With appropriate software from the backup software vendor, customers can directly attach servers to specific tape drives using SCSI or Fibre Channel. Backup data now moves over the SCSI or FC pipes freeing up LAN Bandwidth and improving backup performance. LAN-free backup has come of age. The following is an illustration of LAN-free backup.

*LAN-free backup is an excellent approach for sites with many servers or sites with large amounts of data behind the servers*



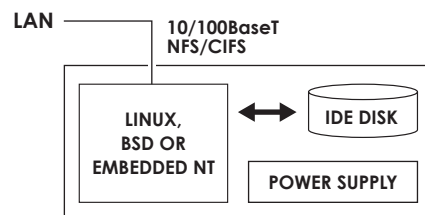
*LAN-free Backup*

LAN-free backup is an excellent approach for sites with many servers or sites with large amounts of data behind the servers. Server intensive site should consider a Fibre Channel based LAN-free configuration since this allows drive sharing providing better backup bandwidth optimization.

### NAS Boxes

NAS boxes, also referred to as filers, at the high end are stripped down servers providing file-serving functionality only. Filers come with a preinstalled operating system having the sole purpose of supporting file traffic between the filer and its LAN clients. Disk farm configurations are preset in order to simplify installation.

NAS boxes come in all capacities and price ranges. Low-end SNAP servers can be had for less than \$850 dollars and provide fundamental file server functionality. Product at this price point offers no RAID protection against disk failure. A simple block diagram of a low end NAS box follows.



*Low-End NAS Box Block Diagram*

# HOST SYSTEM ATTACHMENT METHODS (cont.)

The low-end NAS box has become the main stay for the small office environments due to its simplicity of use. Typical installs require an Ethernet connection; a power connection, a configuration tweak and you are done in less than 20 minutes.

Apple has also just recently introduced a NAS box—only 1.75 inches high—that can support CIFS for windows systems. We will use this symbol to represent the low-end (<\$20,000) NAS boxes:

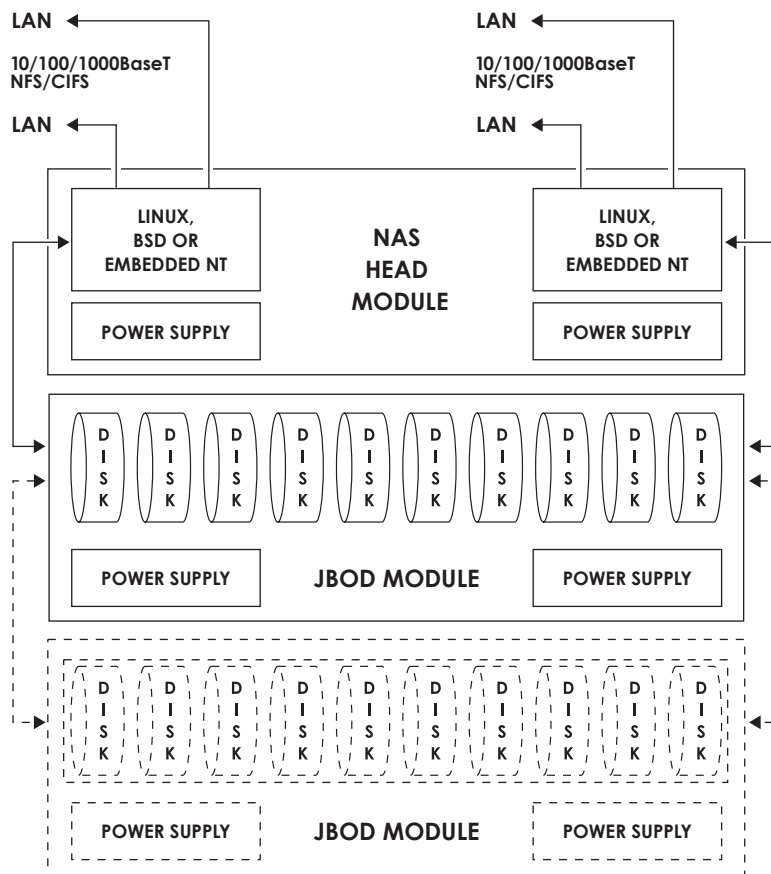
**NAS**

On the mid to high end (Network Appliances, EMC i4700) NAS heads have become file gateway front ends to disk farms that look a lot like private SANs. The Front End (LAN connectivity) for filers typically supports 2-4-or 8 10/100/1000 BaseT Ethernet connections.

High-end NAS solutions are typically modular in nature with redundant NAS controllers and power supplies in one chassis. The NAS head has backend SCSI or Fibre Channel connectivity to allow for storage expansion growth. The NAS controllers typically run hardware or software RAID on the backend disk storage to minimize the impact of disk failure.

Expansion / Scalability is achieved with the addition of JBOD storage modules. JBOD storage modules are typically SCSI or Fibre Channels based disk drives daisy chained together in an enclosure with redundant power supplies.

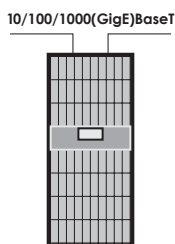
High-end filers offer additional software functionality over and above basic file storage and sharing capabilities. These options typically include snap shot copy and remote mirroring capabilities. The key to the high-end filer's success has been again the ease in which a system can be deployed, typically less than an hour as compared to the hours required to just install server software. The following is a block diagram of a high-end NAS box or filer.



*High-End NAS Box(s)  
Block Diagram*

High-end NAS boxes can typically support 100 to 120 disk drives behind a NAS head. Cost for Scalable NAS storage starts at around \$50,000 and goes up to over \$300,000 as you scale capacity and add features. .

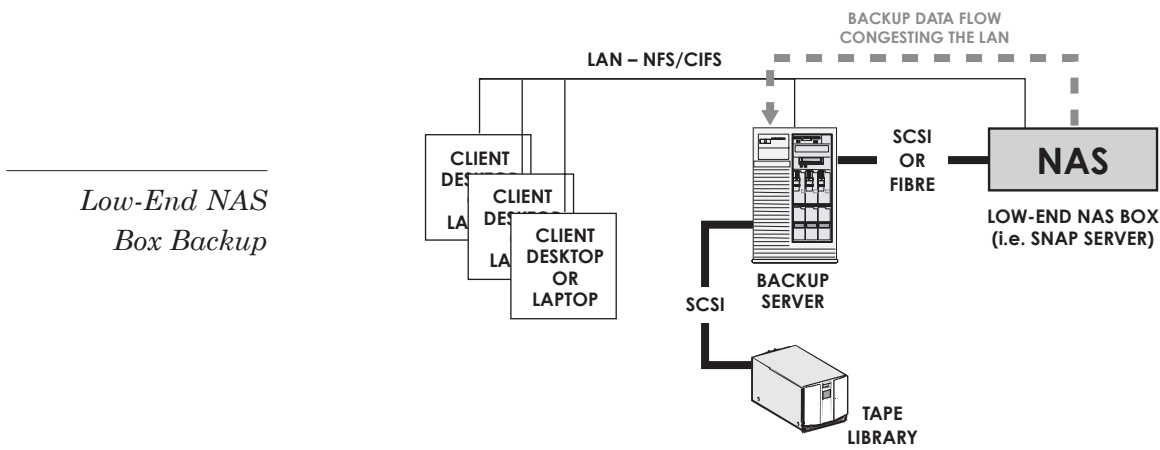
For simplicity in the future we will use this symbol to represent the mid to high-end (>\$40,000) NAS boxes:



*Backup Methods*

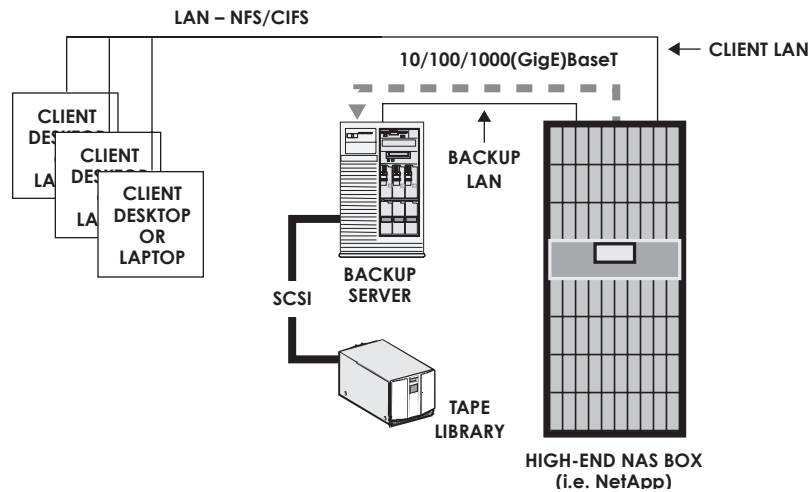
Backup methods for NAS boxes are as varied as the types of NAS boxes. Low-end NAS boxes are backed up over the LAN by a dedicated backup server with a direct connect tape drive or library.

Typically in these sites the amount of data involved and the demands put on the LAN are low enough that the LAN congestion problems may not be an issue. If LAN congestion is a problem, consider these three Over-the-LAN suggestions for improving LAN and backup performance.



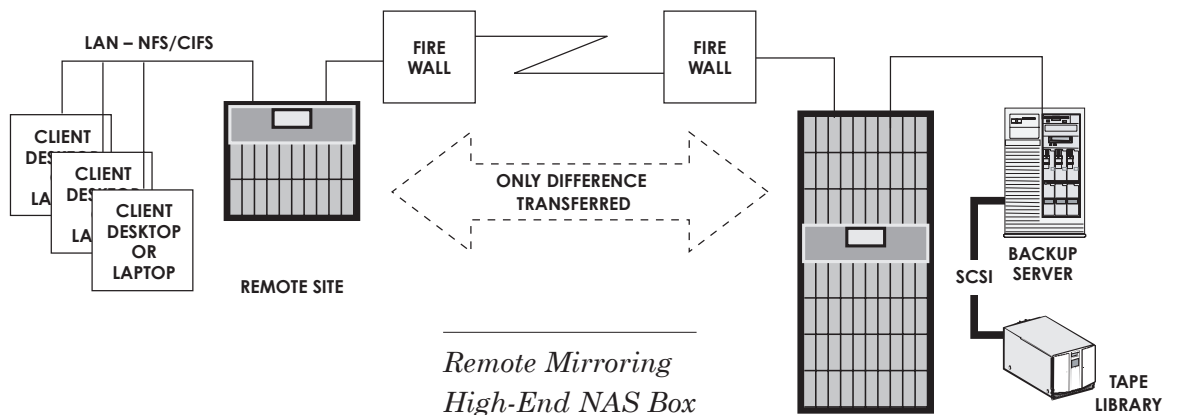
# HOST SYSTEM ATTACHMENT METHODS (cont.)

High end NAS systems are often backed up in the same manner as the low end NAS boxes but more than likely the high end NAS box will be on a dedicated backup LAN with the backup server.



*Entry-level High-End NAS Box Backup*

Another scheme that has been deployed to a limited success is the use of remote mirroring between high end NAS boxes as a means of backup. In this scenario background agents running at the remote site monitor file activity. When a new file is added or an existing file changes, the agent on the remote system asynchronously transfers only the changed data to its mirror partner.



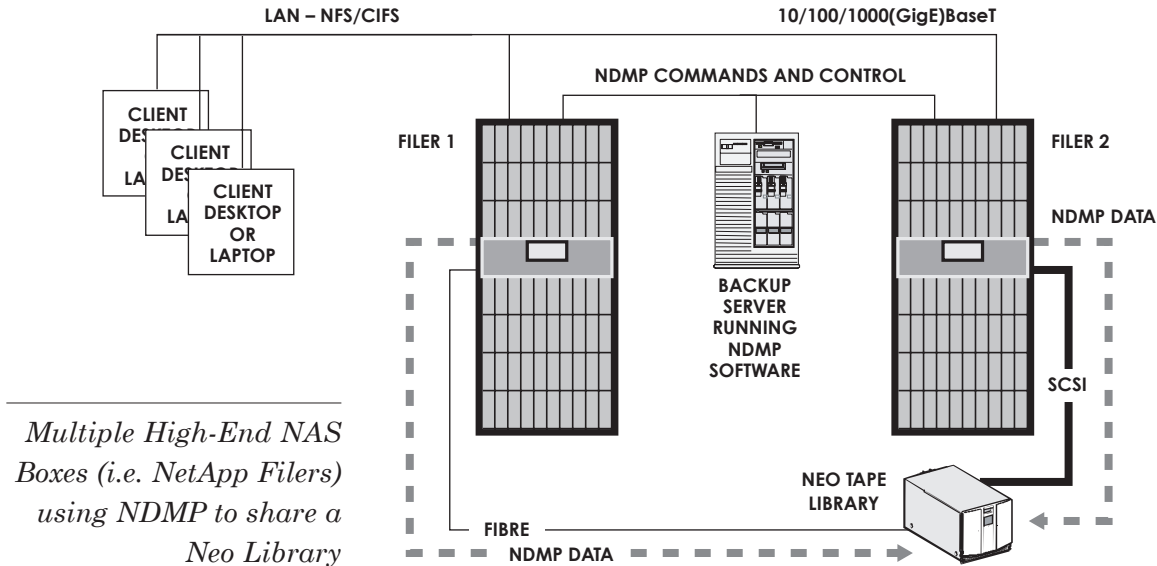
*Remote Mirroring High-End NAS Box*

Asynchronous remote mirroring solves the file backup issue for the remote NAS box, however in the case of disaster recovery or bare metal recovery a tape or tapes along with a tape drive may need to be sent out to the remote site due to slow transfer rates experienced by most remote sites.

NAS box mirroring is also vendor specific. You must have NAS boxes from the same vendor in order to mirror. The transfer agents used are intelligent and only transfer the actual change providing bandwidth savings.

## NDMP

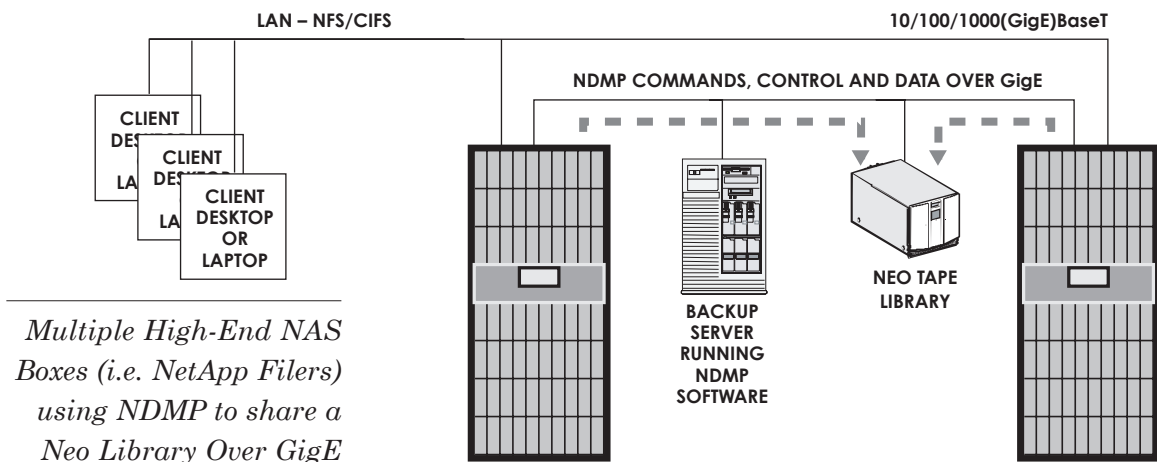
High end NAS systems often support an Ethernet protocol known as NDMP (Network Data Management Protocol) defined by Network Appliances, NDMP is a protocol interface for the purpose of backing up and restoring data to and from filers. NDMP permits the control stream to flow on a separate interface than the actual data stream.



In the example above, the backup server uses a dedicated LAN to communicate NDMP commands to the filers. The NDMP backup data flow to the library can occur over both the SCSI and Fibre Channel connections.

### Future Overland NDMP Support

In the future Overland's Neo library will be able to support NDMP functionality directly over GigE using the Neo's GigE card running NDMP tape services.



In this configuration NDMP commands and data are passed back and forth between the filers, the tape library and the backup server over a dedicated GigE LAN.

# HOST SYSTEM ATTACHMENT METHODS (cont.)

*SAN is a storage centric view where a Fibre Channel attached RAID system provides storage for many servers that are at a peer level.*

## SAN

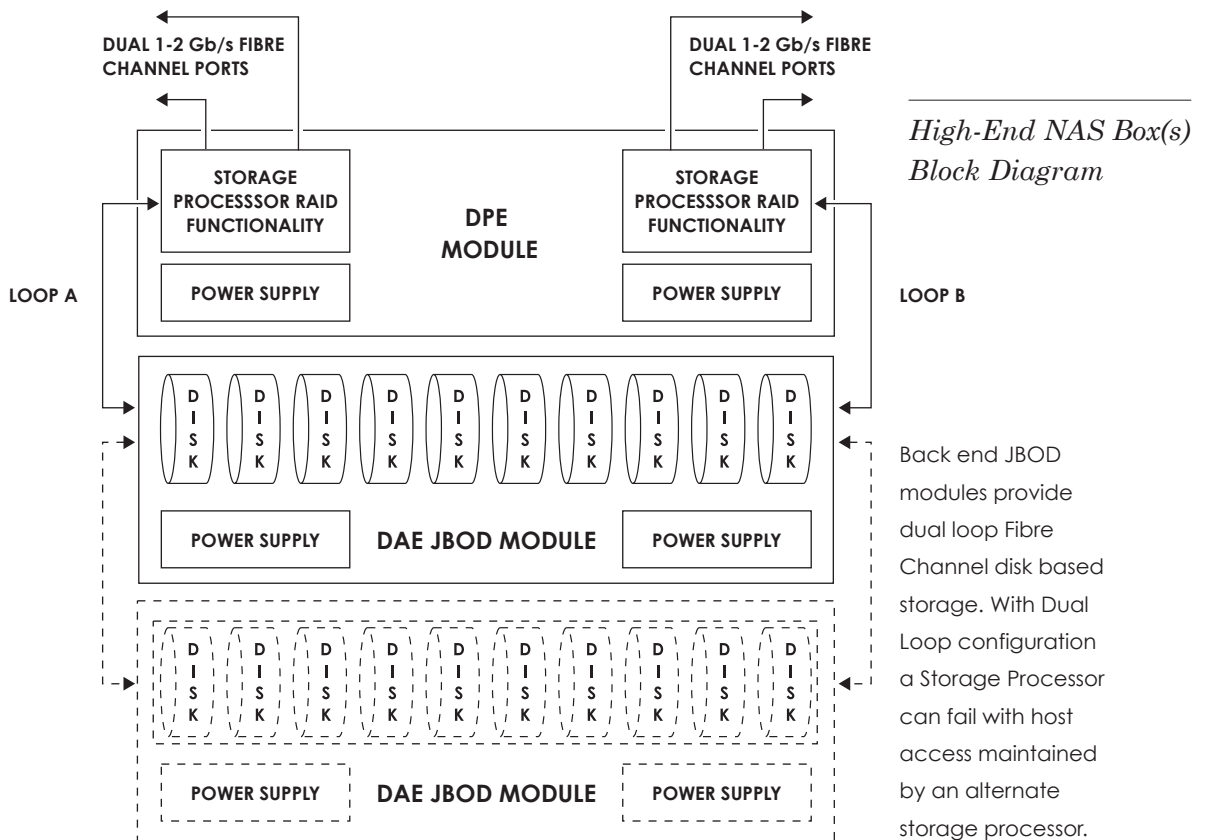
SAN is a storage centric view where a Fibre Channel attached RAID system provides storage for many servers that are at a peer level. Configuration, Access control and security are three additional burdens placed on system administrators when using SANs. These issues are addressed through vendor unique schemes for configuring Port & LUN zoning as well as LUN Masking.

### Enterprise Class Fibre Channel RAID Subsystems

Midrange enterprise class RAID systems come from two dominant RAID suppliers, EMC with the 5400 and 4700 series and LSI Logic. Disk array products in this class provide both disk redundancies through RAID as well as hardware redundancy for:

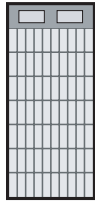
- Storage processors      Cooling
- Fibre Channel Paths      Battery backed up write Cache
- Power Supplies

RAID array subsystems are built using a DPE (Disk Processor Enclosure) and one or more DAEs (Disk Array Enclosure). The DPE contains two storage processors that provide Fibre Channel host interfaces as well as the RAIDed mapping of user data to the backend disk farm's data sectors.



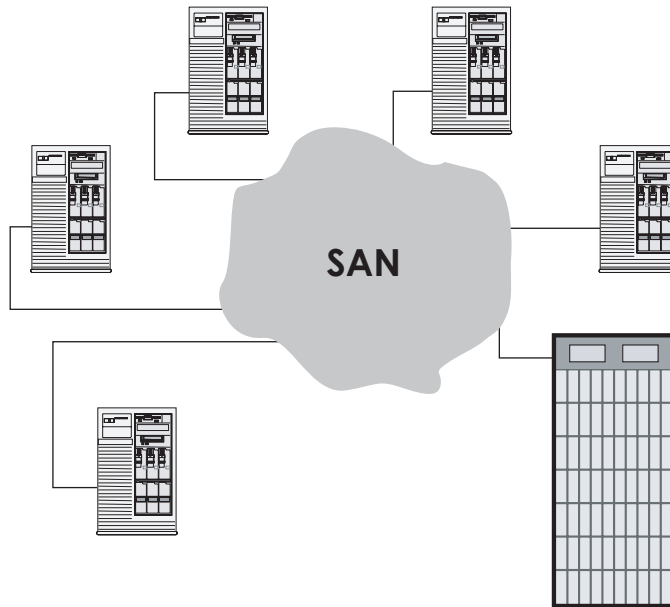
Recently IDE/ATA Disk arrays with Fibre front ends have shown up in the market. Expectations are that over the next 5 years most applications that have used Fibre Channel or SCSI Disk Drive Array storage will migrate to the IDE/ATA array for cost reasons.

We will use this symbol to represent the mid to high-end (>\$50,000) SAN RAID array:



*SAN Vision*

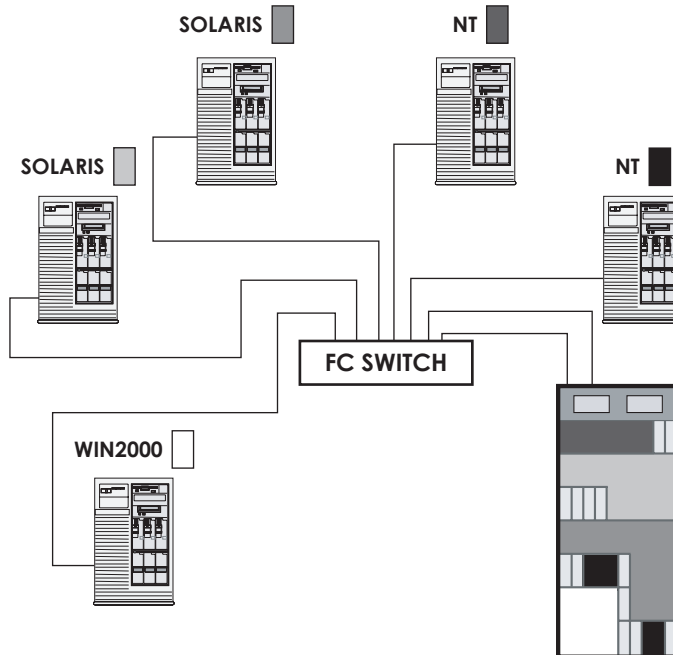
In the vision of SAN all servers and storage nodes were peers to each other. Data was shared across the SAN seamlessly. SAN was going to solve all the problems.



# HOST SYSTEM ATTACHMENT METHODS (cont.)

## *SAN Reality*

The SAN reality was much different. Heterogeneous platforms and file systems sharing the same Storage RAID array was possible using Port Zoning on the switch and LUN masking on the array but sharing data between heterogeneous file systems was not solved by SANs.



Additional interoperability issues with Host bus adapters and cascading switches from mixed vendors kept the early SAN adopters busy.

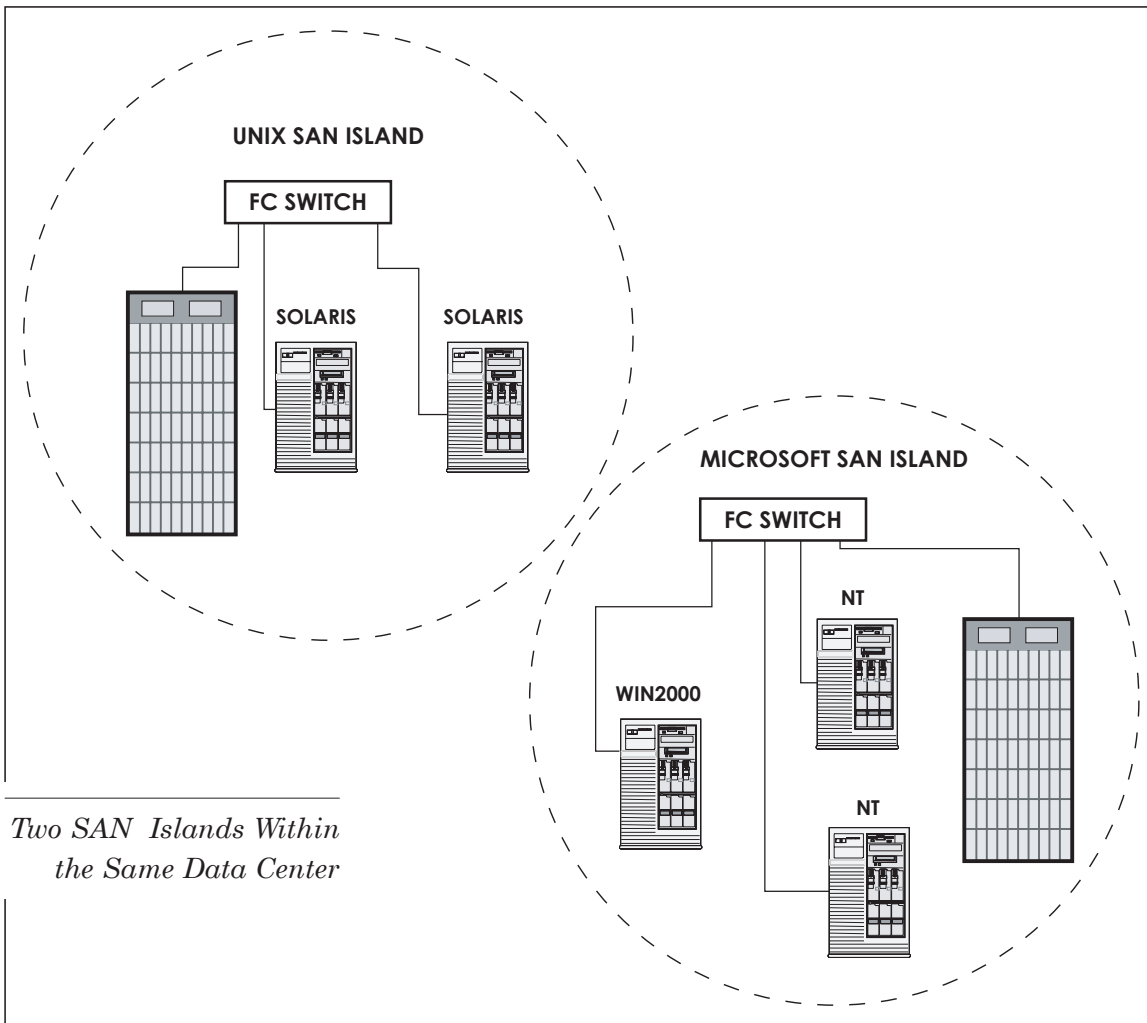
Storage management for SANs added several additional levels of complexity for the systems administrator and this slowed SAN growth. Storage administrator's duties to put a new storage system online now included:

- Determine appropriate RAID level for application
- Bind Logical Disk(s) to RAID level
- Identify bound LUNs
- Partition bound LUNs for correct OS
- Format Partitions for correct OS
- Enable access to LUN on host by host basis

The storage administrator on an ongoing basis must now monitor his data usage and grow disk space as needed.

Distributed file systems have also surfaced from IBM (Synergy) and others. These file systems run on the host and provide a level of file sharing to heterogeneous platforms that support the distributed file system.

In today's data center it is very common to find several SANs deployed. This is often done on an OS or applications basis. The following is an illustration of just such a data center with a Unix SAN island and a Microsoft SAN island.



*Two SAN Islands Within the Same Data Center*

#### IP SAN

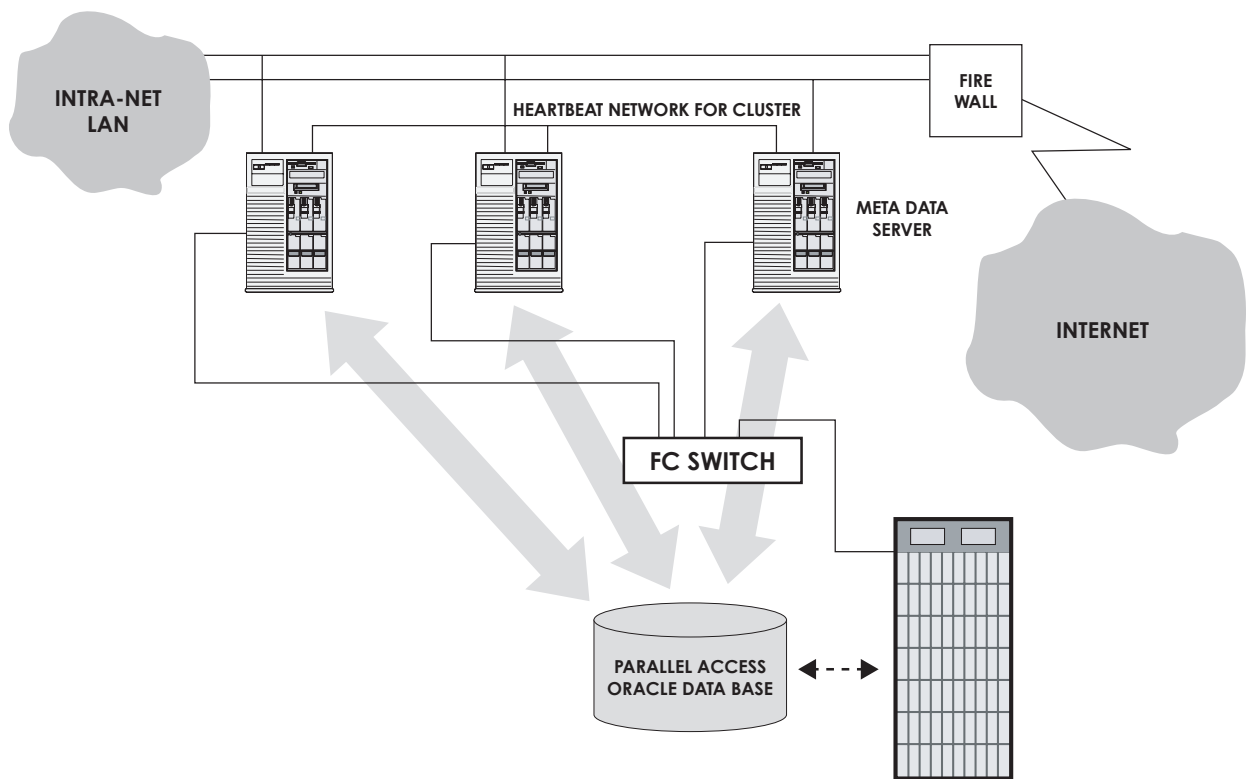
Internet Protocol (based) Storage Area Networks will be the next major areas of SAN growth. Starting in late 2002, iSCSI protocol based disk and tape will begin to ship. The advantage to iSCSI will be the cost of the GigE infrastructure (switches) versus Fibre Channel.

All of the proceeding and all of the following SAN topologies will be implement with iSCSI / Ethernet technology. Overland will have a GigE (Gigabit Ethernet) card supporting iSCSI for Neo early in 2003.

# HOST SYSTEM ATTACHMENT METHODS (cont.)

## Server Clustering

One of the key advantages to Fibre Channel is the dramatic improvement over SCSI's 25-meter total bus length limitation. With short wave optical Fibre you now have several hundred meters. An area where this added cable length plays an important role is in server clustering. In server clustering, multiple servers have access to the same logical disks and data. Disk coherency is maintained by making one of the servers the Cluster Master. The Cluster Master controls coherency by maintaining the meta data and write and read locks for the virtual disks and ensuring that write operations are carried out in an orderly fashion.



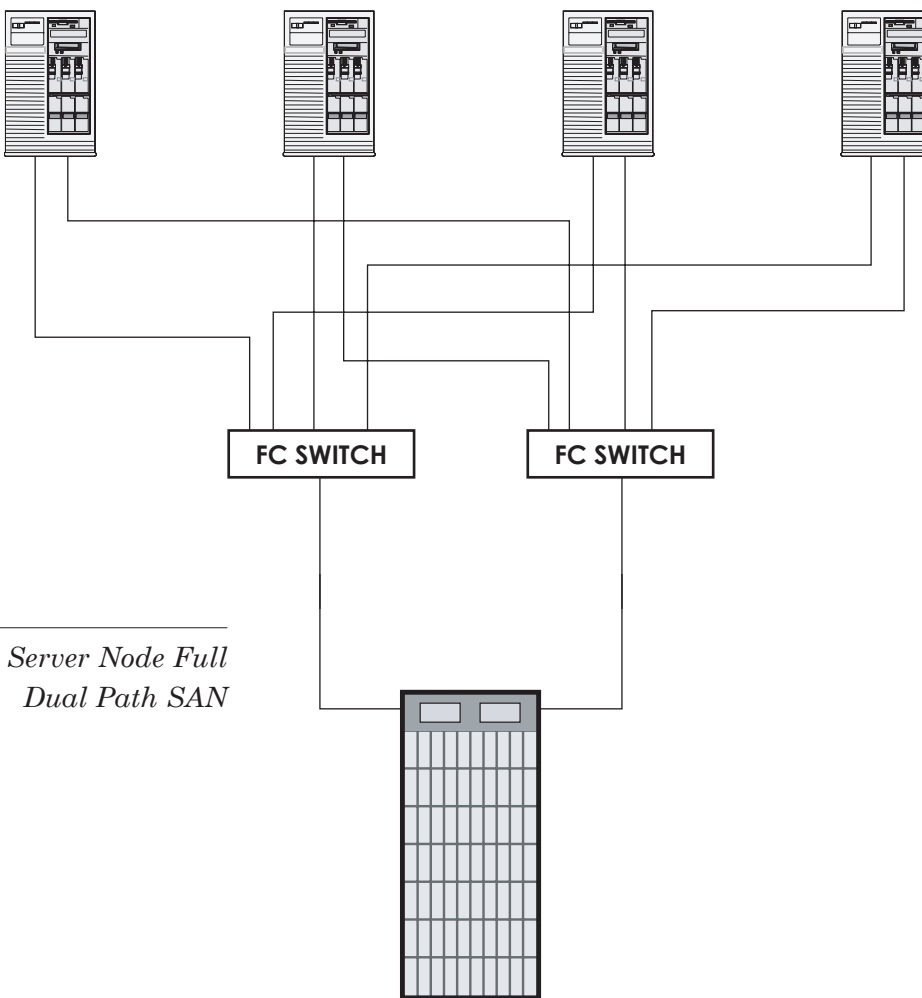
Clustering servers as shown in this illustration is a means of achieving scalability of access to the database or other shared data within the RAID array. As an example, this could be an Oracle database being used in an E-commerce OLTP (On Line Transaction Process) application. The servers are ganged together to provide transaction horsepower and redundancy. Coherency for the cluster is maintained by running a cluster control and heartbeat network between the clustered servers. The heartbeat is used to check status of all members of the cluster. If the Meta data server fails to respond to a heartbeat, predefined policies permit another member of the cluster to assume the meta data server role, this is referred to as fail over.

Sophisticated clustering software also provides means by which servers already active within the cluster can adopt the network IP address of the failed server. Clients on the Intra-LAN or on the Internet may see a brief interruption (< 10 seconds) in service but long term access to data is maintained.

*HA (Highly Available) SANs*

Highly Available SANs are designed to eliminate any single point of path failure from the SAN. In the four-system node configuration shown below each server is equipped with two Fibre Channel HBAs (Host Bus Adapter).

One of the "paired" HBA's in each server is connected to one of two independent Fibre Channel switches. Each storage processor of the RAID array unit connects to the independent Fibre Channel switches. Firmware within the array permits logical volumes to be accessible through either storage processor.



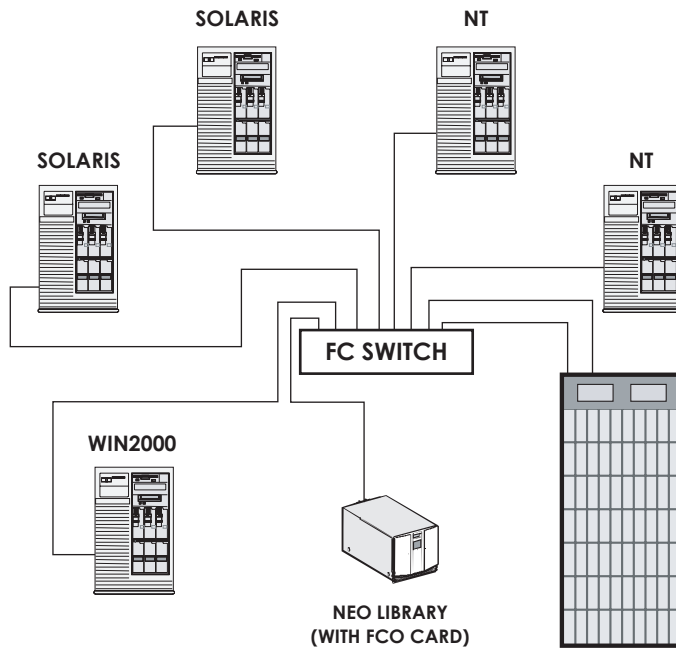
*Four Server Node Full  
Dual Path SAN*

Driver software running on the servers provides a single view of the attached dual path resources. In the event of a path failure, driver level software enables resource access down the alternate path. In sophisticated Active/Active systems application running on the servers may be unaware that a path failure has occurred and may only notice a bandwidth reduction.

# HOST SYSTEM ATTACHMENT METHODS (cont.)

## *SAN Backup Methods*

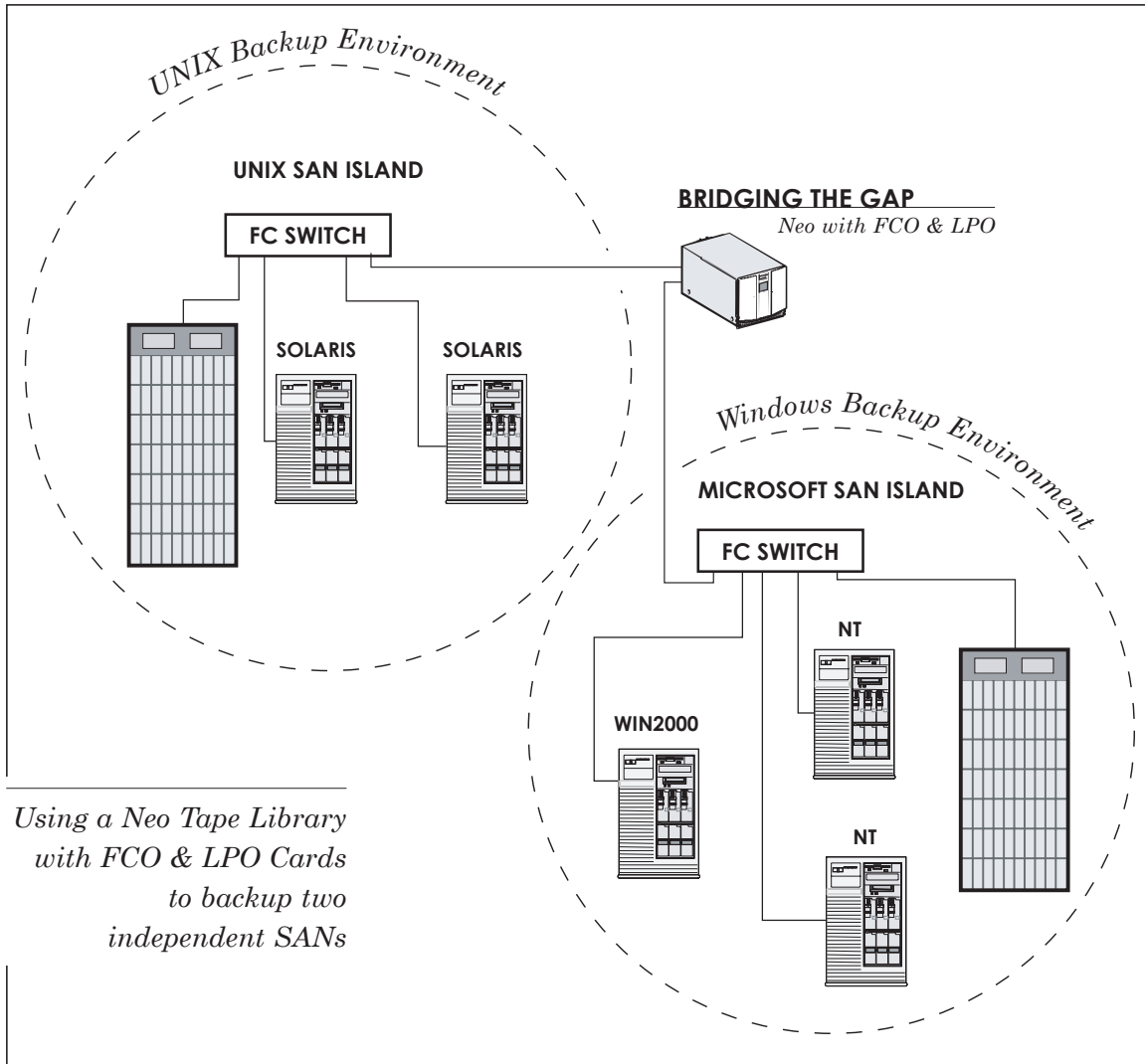
One of the key advantages of SANs is that they make library "sharing" easy from a hardware perspective. In the following example SAN, One of the Solaris servers is also the backup server running Veritas NetBackup. The Backup server manages the backup and media while backup agents on the remaining servers send the backup data over Fibre Channel to the drives within the Fibre Channel Option (FCO) equipped Neo.



As the SAN increases in size, Neo modules and FCO cards can be seamlessly added to meet the capacity and performance combination needed for high speed backup.

*SAN Island Backup*

SAN Island backup is made simple with the Neo library partitioning option (Library Partitioning Option). Independent backup software packages from different vendors can be running on each SAN Island doing concurrent backups.



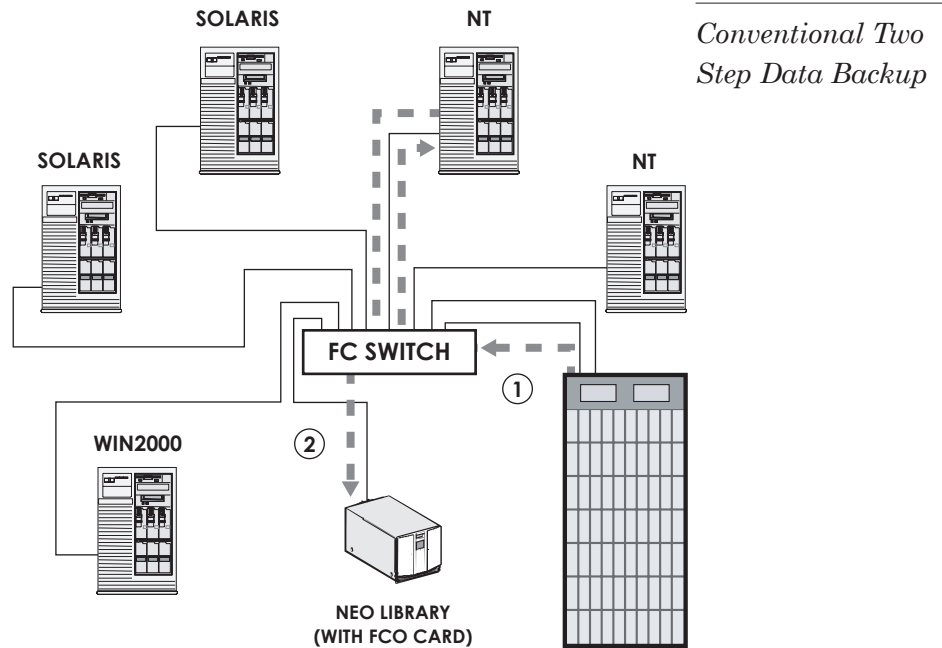
*Using a Neo Tape Library with FCO & LPO Cards to backup two independent SANs*

Independent Fibre Channel Option (FCO) cards and Library Partitioning Option (LPO) card guarantees total SAN data and command independence. An ideal solution for mixed environments shops. The LPO card allows the user to divide a Neo library into smaller virtual libraries.

# HOST SYSTEM ATTACHMENT METHODS (cont.)

## Serverless Backup

In the historical approach to backup, data was first read from the primary storage device to the server, than transferred across the SAN a second time prior to being written to tape.

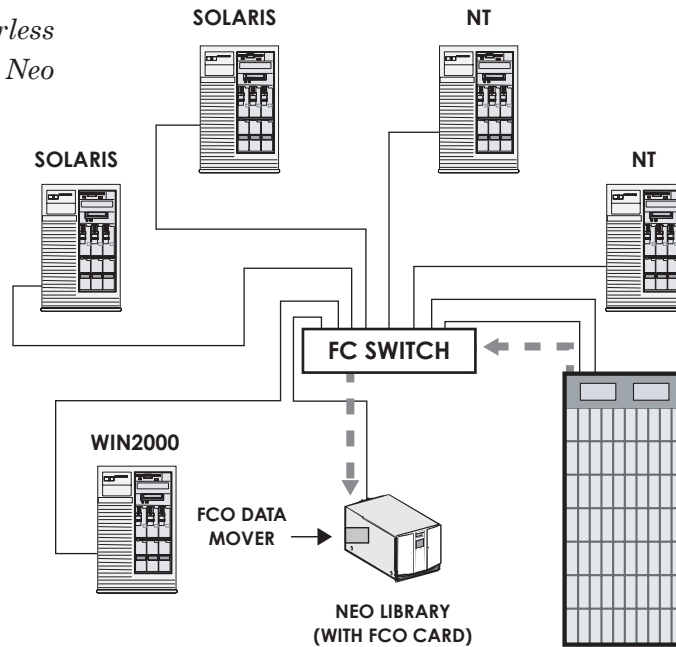


In using serverless backup, data flows directly between the primary and secondary storage devices only and it flows only once. The server and its operating system which own the LUN being backed up are no longer in the data path, allowing for high speed device to device copying.

Latencies are minimized when running serverless backup. Some backup software packages copy the logical disk as an image greatly reducing OS interaction and dramatically improve throughput performance.

*The assignment of drives and magazines to each partition is accomplished through the operator's panel on the Neo Series tape library or through the WebTLC remote library management software.*

### Single Move Serverless Backup with Neo



Serverless backup is here to stay for SAN providing a high-speed transfer mode. Multi threaded severless backup operations are the means by which MIS departments can cope with the ever-increasing data storage and backup demands.

Serverless backup is made possible by data mover code resident on the FCO card. In the severless mode of operation, the backup server tells the FCO card's data mover which blocks to copy. The FCO card reads the requested blocks of data from primary storage and writes those blocks to tape. Backup data never flows through the server.

Currently two major backup software vendors support serverless backup. This backup methodology will find its way into iSCSI Ethernet SANs also.

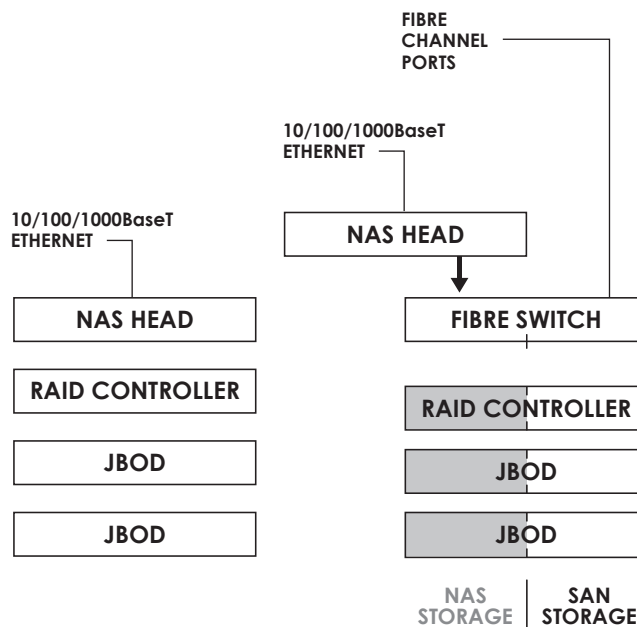
# HOST SYSTEM ATTACHMENT METHODS (cont.)

## SAN NAS CONVERGENCE

Much has been said recently about the SAN / NAS convergence that is supposed to occur over the next 18-24 months. "Convergence" of these two markets is occurring on two fronts.

The first front is the use of block level protocol like SCSI over Ethernet. With iSCSI customers will be able to use Ethernet technology to create SAN functionality. NICs (Network Interface Card) will serve a dual role in the future providing both the conventional file level access to data as well as block level access to data. Over the next few years we anticipate that most HW/protocol interfaces of today, including Fibre Channel will become a TCP/IP protocol on 10 GigE.

The second "convergence" of SAN and NAS is nothing more than a few high-end NAS vendors giving block level access to part of their disk farm. The NAS vendors that have chosen to offer this as a "feature" have an architecture as shown on the left.



What is sold as SAN NAS convergence is on the right. As one can see the vendor has opened up his private backend SAN disk farm by adding a Fibre Channel switch in front of his RAID controllers. Now host systems that attach to the switch will be able to use a portion of the RAIDed JBOD for block level SAN storage. Virtual disks can not be shared between the SAN and NAS front ends.

# APPENDIX A – HAVING FUN WITH EXCLUSIVE OR'ING

Below are the logic symbol and truth table for the lowly exclusive Or gate. The second illustration are three-ganged Xor's with the associated Boolean equation. With this ganged Xor array we can emulate RAID systems parity sector generator and sector re-creator functionality.

*See how the "mystery" of RAID can recover a lost sector or failed drive just as easily as your lost bit.*

To start with you will need to choose four single bit values, lets say 0,1,1, and 1, such that A=0, B=1, C=1 and D=1. We now percolate the data bits through the parity generator to create our parity bit. Let's see

A Xor B = 1 (Z1), Z1 Xor C = 0 (Z2), Z2 Xor D = 1 (Z final) our Parity bit.

Now drop one of the original four bits, say the C bit. You have lost or forgotten the C bit, but you have A=0, B=1, D=1 and Parity = 1. We can now use the same flow to recreate or "recover" our missing data. We will merrily substitute the parity bit where the C value went; the resultant (Z Final) will be the recovered bit. Let's give it a try.

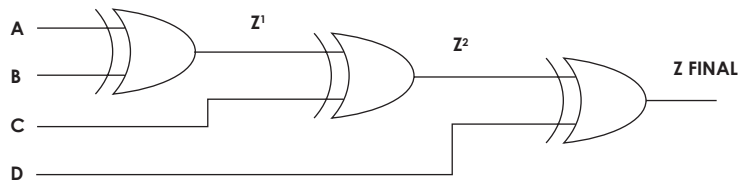
A Xor B = 1 (Z1),                      1(Z1) Xor 1(1st Z final) = 0 (Z2),                      0(Z2) Xor D = 1 our missing bit.



EXCLUSIVE OR SYMBOL

Xor Truth Table

A	B	Z
0	0	0
1	0	1
0	1	1
1	1	0



$$Z \text{ FINAL} = A \oplus B \oplus C \oplus D$$

Now try four bit values of your own choosing and see how the "mystery" of RAID can recover a lost sector or failed drive just as easily as your lost bit.



---

**WORLDWIDE HEADQUARTERS**

4820 Overland Avenue  
San Diego, CA 92123 USA  
TEL 1-800-729-8725  
1-858-571-5555  
FAX 1-858-571-3664  
EMAIL [sales@overlandstorage.com](mailto:sales@overlandstorage.com)

---

**UNITED KINGDOM (EMEA OFFICE)**

Overland House  
Ashville Way  
Wokingham, Berkshire  
RG41 2PL England  
TEL +44 (0) 118-9898000  
FAX +44 (0) 118-9891897  
EMAIL [europe@overlandstorage.com](mailto:europe@overlandstorage.com)

---

**FRANCE OFFICE**

Overland Storage  
126 rue Gallieni  
92643 Boulogne Cedex France  
TEL +33 (0) 1 55 19 23 93  
FAX +33 (0) 1 55 19 25 02  
EMAIL [europe@overlandstorage.com](mailto:europe@overlandstorage.com)

---

**GERMANY OFFICE**

Humboldtstr. 12  
85609 Dornach Germany  
TEL +49-89-94490-214  
FAX +49-89-94490-414  
EMAIL [europe@overlandstorage.com](mailto:europe@overlandstorage.com)

---

**ASIA PACIFIC REP. OFFICE**

30 Robertson Quay, #02-10  
Singapore, 238251  
TEL 65-6839-3510  
FAX 65-6738-3008  
EMAIL [asia@overlandstorage.com](mailto:asia@overlandstorage.com)

---

**[WWW.OVERLANDSTORAGE.COM](http://WWW.OVERLANDSTORAGE.COM)**