StoreVault™
A NetApp Division

an introduction to networked storage

How networked storage can simplify your data management

The key differences between SAN, DAS, and NAS

The business benefits of networked storage

introduction

This paper provides an introduction to the basics of networked storage and its relevance to smaller business operations where storage specialists are typically not found. It discusses the origins and development of the latest advances in storage technology, while illustrating the real-world advantages they bring.

Further reading and more advanced topics may be found at www.netapp.com

historical perspective

Mainframes

The development of networked storage has its roots in early computing models and has been enabled by technology developments of the past ten years. Early in the mainframe era, computers controlled their own disks and allocated space to applications based on priority. Disk storage was very expensive, so it was a prized computing resource and very carefully managed by the operating system and system administrators. Desktop computers were nothing more than dumb terminals, competing for the same resources on the central computer. This model became more difficult to manage as bottlenecks formed on the central computer.

Client-server computing

As computing and storage costs fell, we moved into the networked client-server model: smaller but powerful computers called servers managed specific applications like e-mail, databases, printing, and file sharing, while the emergence of desktop PCs enabled local processing and local storage. These computing resources were interconnected to form a network so that the servers and the PCs (or clients) were able to communicate with each other and exchange data. This network was known as a Local Area Network (LAN) and was based on Ethernet technology. As computing and storage costs fell further, servers proliferated and PCs got more powerful and gained more storage. This decentralization of processing power and storage led to the realization that management of the storage had become significantly more difficult. As more servers were added, backups became more challenging; adding storage often meant adding another server, which just made management worse.

While there had been great advances in the availability of processing power, storage resources were poorly used, management of the various machines was time consuming, and assessing what data was valuable became nearly impossible.

Today's environment

Today's computing networks typically have very powerful client computers with significant amounts of captive storage attached to each one. Processing power is very easily available and the Internet interconnects a huge number of private networks and individual computers. Captive storage for clients, servers, and dedicated applications has exploded. New regulations are affecting some industries, requiring more IT spending and planning around storage than ever before. In addition, increased dependence on computer applications for daily business operations has heightened the focus on data availability across all industries. Successfully executed backups and (more importantly) successful restores may mean the difference between gracefully recovering from data loss and starting the countdown to filing for bankruptcy.

The Challenges of Decentralized Storage

As processing power proliferates, and dedicated captive storage follows along with each new machine, storage "islands" are gradually formed. Islands are dedicated storage arrays or volumes that are only used by a single host or a single application. One example of a common island is an Exchange server with an external SCSI array; another could be a SQL server with some dedicated RAID-protected drives. Captive locally attached storage for a database cannot be shared out to other computers; nor can it be easily grown, moved, or duplicated without placing a significant burden on or even causing downtime for the host.

These islands represent wasted hardware investment dollars, since many times some are underutilized while others are overflowing. Also, as these islands propagate over time they increasingly burden the systems administrators with management of the new growth. Storage islands often cause backup problems, requiring dedicated tape drives and software to manage multiple individual backup tasks – not to mention the time and burden on the host when a restore is necessary.

Other examples of storage islands can be found with even simple applications. General-purpose file and print servers that are dedicated to sharing files out to individual computers require backup management, and antivirus and client software licensing, and are subject to similar capacity upgrade and downtime concerns as more specific application servers.

The Benefits of Centralized Storage

Application data can be combined into a single storage device or pool of devices called a SAN. SANs provide very high performance data storage that can be scaled as needed. File sharing services, usually provided by partially or wholly dedicated servers, can be

consolidated into a NAS server. In the simplest cases, a single application server can share files as well as host dedicated application data, and also run software for backup and restore tasks. But as servers grow in capacity while narrowing their specialization, the advantages of consolidating become clearer.

With increasing pressures of security, regulatory compliance, and corporate governance, management of business data is becoming ever more complex and important. Technology is allowing businesses to create and store exponentially more data; the key to being successful is managing that data explosion.

What would happen if storage were combined? Consolidation means reducing the amount of time spent managing tasks, jobs, applications, and data growth. Moving all types of data to a single, central, redundant and high-performing storage server could mean less time spent handling repeated storage-related tasks and worrying about backup failures.

Let us now consider some of the key technologies involved in networked storage.

networked storage technologies

Direct-attached Storage

The most common form of server storage today is still direct-attached storage. The disks may be internal to the server or they may be in an array that is connected directly to the server. Either way, the storage can be accessed only through that server. An application server will have its own storage; the next application server will have its own storage; and the file and print servers will each have their own storage. Backups must either be performed on each individual server with a dedicated tape drive or across the LAN to a shared tape device consuming a significant amount of bandwidth. Storage can only be added by taking down the application server, adding physical disks, and rebuilding the storage array. When a server is upgraded, its data must be migrated to the new server.

In small installations this setup can work well, but it gets very much more difficult to manage as the number of servers increases. Backups become more challenging, and because storage is not shared anywhere, storage utilization is typically very low in some servers and overflowing in others. Disk storage is physically added to a server based on the predicted needs of the application. If that application is underutilized, then capital cost has been unnecessarily tied up. If the application runs out of storage, it must be taken down and rebuilt after more disk storage has been added. In the case of file services, the typical response to a full server is to add another file server. This adds more storage, but all the clients must now be reset to point to the new network location, adding complexity to the client side. Additional cost in the form of Client Access Licenses (CAL) must also be taken into account.

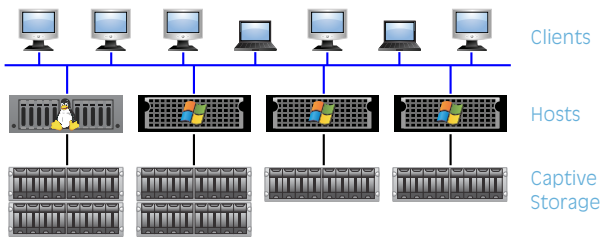

Figure 1: DAS Architecture

In a direct-attached storage environment, storage is a dedicated to the server to which it is attached. There is no sharing of storage resources.

Storage Area Networks

A SAN allows more than one application server to share storage. Data is stored at a block level and can therefore be accessed by an application, not directly by clients. The physical elements of the SAN (servers, switches, storage arrays, etc.) are typically connected with Fibre-Channel – an interconnect technology that permits high-performance resource sharing. Backups can be performed centrally and can be more easily managed to avoid interrupting the applications. The primary advantages of a SAN are its scalability and flexibility. Storage can be added without disrupting the applications, and different types of storage can be added to the pool.

With the advent of storage area networks, adding storage capacity has become simplified for systems administrators, because it's no longer necessary to bring down the application server. Additional storage is simply added to the SAN, and the new storage can then be configured and made immediately available to those applications that need it. Upgrading the application server is also simplified; the data can remain on the disk arrays, and the new server just needs to point to the appropriate data set. Backups can be centralized, reducing workload and providing greater assurance that the backups are complete. The time taken for backups is dramatically reduced because the backup is performed

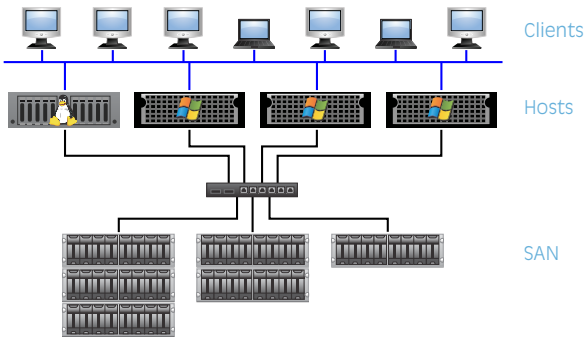

Figure 2: SAN Architecture

In a storage area network, data is consolidated behind the servers. Storage can be shared by more than one server. Clients access their data through the application or file server. Storage resources can be reallocated from one application to another.

over the high-speed SAN, and no backup traffic ever impacts users on the LAN. However, the actual implementation of a SAN can be quite daunting, given the cost and complexity of Fibre Channel infrastructure components. For this reason, SAN installations have primarily been confined to large organizations with dedicated storage management resources.

The last few years have seen the emergence of iSCSI (which means SCSI over IP or Internet Protocol) as a new interconnect for a SAN. iSCSI is a lower cost alternative to Fibre Channel SAN infrastructure and is an ideal solution for many small and medium sized businesses. Essentially all of the same capability of FC-SAN is provided, but the interconnect is Ethernet cable and the switches are Gigabit Ethernet – the same low-cost technology you use today for your LAN. The tradeoff is slightly lower performance, but most businesses simply will not notice. iSCSI also provides a simple migration path for businesses to use more comprehensive data storage management technologies without the need for a "fork-lift" upgrade.

Network Attached Storage

A NAS appliance is a simplified form of file server; it is optimized for file sharing in an organization. Authorized clients can see folders and files on the NAS device just as they can on their local hard drive. NAS appliances are so called because they have all of the required software preloaded and they are easy to install and simple to use. Installation consists of rack mounting, connecting power and Ethernet, and configuring via a simple browser-based tool. Installation is typically achieved in less than half an hour.

NAS devices are frequently used to consolidate file services. To prevent the proliferation of file servers, a single NAS appliance can replace many regular file servers, simplifying management and reducing cost and workload for the systems administrator. NAS appliances are also multiprotocol, which means that they can share files among clients using Windows® and UNIX®-based operating systems.

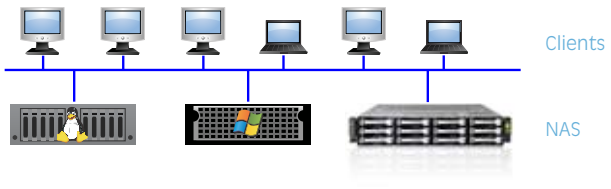

Figure 3: NAS Architecture

Network Attached Storage allows clients to directly access their files across the LAN as if those files were on their local hard drive. The files may be shared by UNIX or Windows clients.

Administrators manage the NAS device from anywhere they have network access. They can assign shares, security settings, and so on, just as they would for a regular file server, except that there are no CALs to manage or pay for. NAS fits right in with existing security and network management tools. As the business grows and needs more capacity, more storage can be added to the NAS device without disrupting users.

One of the most common requests to a systems administrator is to restore a single file or group of files. With traditional backup and restore processes, it is necessary to sort through the backup tapes, restore the correct directory, and look for the required file. With NAS, a feature called Snapshot™ is available that provides an almost instantaneous way for the systems administrator (or even an authorized user) to recover lost, deleted, or corrupted files.

all-in-one networked storage

An all-in-one network storage device combines the capabilities of SAN and NAS providing flexibility to businesses. By combining block and file-level storage, clients on the LAN have direct access to shared files on the all-in-one device and application servers share storage. Instead of having to predict ahead of time how much storage will be required for file sharing versus application data, storage can be allocated from a central pool as required. This approach eases management, reduces costs and improves storage utilization.

The StoreVault S500 from NetApp allows businesses to simplify their storage by consolidating their storage needs into one simple-to-use, all-in-one network storage device. The S500 supports iSCSI connectivity concurrently with CIFS and NFS file sharing, providing great flexibility. The StoreVault S500 is so flexible, it can even be used as direct attached storage until you are ready for NAS and or SAN. If, at a later stage, you want to move up to the performance of a Fibre Channel SAN, the S500 is ready with the addition of a Fibre Channel host bus adapter, protecting your investment.

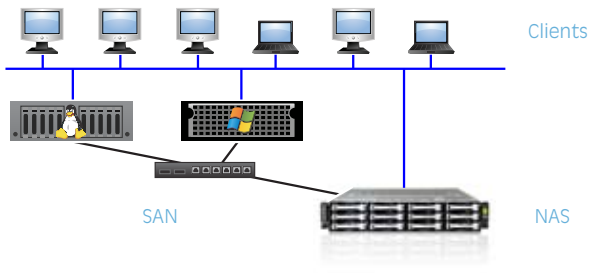

Figure 4: All-in-one Architecture

With an all-in-one network storage device, SAN and NAS storage can be consolidated. Block-level data is available to application servers, and file-level data is available to be shared by clients on the LAN. Storage resources can be allocated on the fly without disruption to users or applications.

The following table describes some of the key differences among these technologies.

	DAS	SAN	NAS
Applications	Storage is unique to one server	Provides shared storage for application servers	Provides shared file storage for heterogeneous clients
Adding storage	Take down application, add drives, rebuild, or add another server	Add storage, configure, and allocate centrally – no user disruption	Add storage, configure, and allocate centrally – no user disruption
Backup	By server or over LAN	Centralized	Centralized
Snapshots	No	Possible	Yes
Typical storage interconnect	SCSI	Fibre Channel or iSCSI	Ethernet
Multiprotocol file sharing	No	Application specific	Yes

summary

We have seen how storage networks have developed to help businesses manage their data. The evolution of SAN and NAS have led to a realization that the best features of both can be combined into a single powerful device that is capable of storing application data (as a SAN) and file data (as a NAS) at the same time, while also providing immediate backup and restore options using Snapshot.

The new StoreVault from Network Appliance is an all-in-one scalable network storage solution that allows first-time NAS and SAN users to simplify not only the installation process, but also the day-to-day protection and management of their data. StoreVault provides access to the powerful data storage management tools and capabilities that have previously been available only to large IT shops. With Snapshot built in, data recovery is simplified, and an in-box backup option allows unprecedented simplicity and reliability of backup and restore.

To find out how NetApp StoreVault can help you focus on growing your business instead of worrying about your growing storage, please visit www.netapp.com

glossary of networked storage terms

backup window
The time it takes to complete a backup of the data. Often done at night and on weekends when the systems are not in use.

client
A computer that accesses a host.

DAS
Direct Attached Storage: A traditional, less efficient storage model where the storage is directly connected to one single host. All client requests must pass through this host to access the storage.

Ethernet
The most common protocol for communicating over Local Area Networks.

Fibre Channel
A high-performance interconnect protocol used in Storage Area Networks.

host
A computer that has storage connected to it.

iSCSI
SCSI over IP or Internet Protocol: a lower cost alternative interconnect protocol to Fibre Channel for Storage Area Networks.

LAN
Local Area Network: the connection between servers and clients or desktop PCs.

NAS
Network-attached storage: a simple means of providing shared file storage.

RAID
Redundant Array of Independent Disks (originally Redundant Array of Inexpensive Disks) is a series of methodologies for providing data protection.

RAID-DP
A new RAID methodology developed by NetApp that provides protection from two simultaneous drive failures in a group without mirroring an entire RAID group. More detail can be found at www.storevault.com.

RAID levels 0, 1, 4, 5
RAID 0: data is striped across at least two drives to improve performance. This is not strictly RAID, because there is no redundancy, and the loss of just one drive causes all data to be lost. All of the disk capacity is available for data.

RAID 1: often called mirroring, the same data is written to two drives at the same time. If one drive fails, the data is still available on the other drive. Only half of the total disk capacity is available for data.

RAID 4: blocks of data are written to a group of drives and parity is written to an extra drive. If any drive fails, the data can be reconstructed from the surviving drives. In a group of **n** drives, **n**-1 drives of capacity are available for data, because one drive is dedicated to parity. More flexible than RAID 5 as additional drives can be added to the array without disruption to users or applications.

RAID 5: Similar to RAID 4, except that the parity is distributed among all the drives. The same capacity compromise also applies.

SAN
Storage Area Network: a high-performance network that is dedicated to sharing storage between application servers.

StoreVault
A new all-in-one network storage device from Network Appliance that provides scalable storage for application servers and for heterogeneous file sharing.

Snapshot
An instantaneous, point-in-time image of the file system that allows individual files, selected directories, or an entire file system to be immediately recovered.

NetApp®

Network Appliance, Inc.
495 East Java Drive
Sunnyvale, CA 94089
877.2STRVLT

About Network Appliance

Network Appliance is a world leader in network storage solutions for today's data-intensive world. Since its inception in 1992, Network Appliance has delivered technology, product, and partner firsts that simplify data management. Information about Network Appliance solutions and services is available at **www.netapp.com.**

For more information on StoreVault, A NetApp Division, go to **www.storevault.com.**

© 2006 Network Appliance, Inc. All rights reserved. Specifications subject to change without notice. NetApp, the Network Appliance logo, and WAFL are registered trademarks and Network Appliance, Snapshot, StoreVault, and the StoreVault logo are trademarks of Network Appliance, Inc. in the U.S. and other countries. All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such. Windows is a registered trademark of Microsoft Corporation.

20-195-101 08/17/2006