

DISC > White Papers

DEVELOPING A STORAGE STRATEGY

Businesses today are relying more and more on information systems for successful operation in the corporate world. Both employees and customers build these information systems around network data that is created on a daily basis. Data created in the course of running your business exists throughout the network and is one of your company's most valuable corporate assets. As you begin to look at data as a corporate asset, a strategy must be developed that will guarantee the security and availability of your data, while still allowing it to grow.

According to META Group:

- Most organizations are facing exponential data growth
- Business are demanding 24x7 access to their data
- IT managers need to adopt a storage strategy for managing, protecting and growing their network storage assets

How do you grow your available network storage and keep it under control? For most IT staffs, this is a difficult task because your network evolved as the company grew. You added workstations when needed, and you added storage to your existing server as needed. Eventually you ended up with a hodge-podge of technology as your network. In many networking environments adding more storage means hanging another server off of your network or taking down an existing server and installing additional hard drives. This type of storage increase just adds to the complexity of your existing storage environment. As your storage requirements grow to support data collection from broadband Internet access, creation of large graphic files, archival of e-commerce transactions and the exponential growth of email databases, you must develop a consistent, concise strategy for corporate network storage.

Developing a storage strategy for your network environment begins with answering three basic questions:

- How do you want to connect additional storage to your network?
- How will you manage all of the storage devices and the data on your network?
- How will you ensure the security and availability of the data on your network?

A good strategy for network storage addresses each of these questions and allows for the expansion of your network storage pool without affecting either the availability or security of the existing data on the network.

Storage Architecture -- Connecting Storage to Your Network

When you see the message "Out of disk space on Drive X," the network demand for storage is exceeding your plan or ability to add capacity. This is a sure sign that you need to develop and implement a more practical strategy for network storage. For the first part of your strategy, you must decide which storage architecture best suits your environment. There are 3 basic storage architectures available. Direct Attach Storage (DAS), Storage Area Network (SAN) and Network Attach Storage (NAS). The following definitions for each type of storage architecture can help you decide which one will work best for your application.

Direct Attach Storage (DAs)

Direct attach storage is the old way of doing things. DAs is storage connected directly to a file server via SCSI. With a direct attach storage system storage is local to a specific file server. This single server controls all information. DAs can also be referred to as captive storage or server attached storage. Adding storage to your network using this model requires installing another network server with additional storage capacity. Another method might also entail bringing down an existing network server and installing additional storage devices into it, or connecting new storage devices to it via external cabling. Again, this is the old way of doing things and the demands on today's networks do not allow for the downtime required for this type of implementation.

Performance: Direct Attach Storage is the slowest of all of the storage architectures. Attaching directly to storage servers means that processors in the server need to manage application requests and move data across the bus and monitor traffic on the network simultaneously. This creates too many demands on the network to provide quick access to attached storage.

Scalability: Storage volume is tied to server capacity in DAs model. Adding storage requires server downtime, physical space in the server and in some cases new servers. Scalability is severely limited.

Availability: Storage in the DAs model is also tied to server availability. If an individual server goes down, all of the attached storage becomes unavailable, leaving you without access to your data. The complexity of network server hardware and operating systems adds unnecessary failure points in your storage strategy.

Cost of ownership: Adding the cost of general purpose network servers to the cost of storage makes DAs one of the most expensive ways to add storage to your network.

Storage Area Network (SAN)

SANs are high-speed networks that enable the interconnection of heterogeneous systems and storage elements. SAN gateways deliver transparent performance that attaches devices across multiple interfaces while permitting each to deliver its full performance capability across the SAN network. By putting storage devices on a separate high-speed network via a SAN, data can be directly accessed by multiple servers, workstations and PCs and be managed as a centralized storage pool. Sans need the bandwidth of interconnects such as Fibre Channel for optimum performance, availability and scalability. But because TCP/IP does not run over Fibre Channel yet, some interim SAN implementations may use legacy networks such as Ethernet or FDDI. When IP on Fibre Channel becomes available, all control and data traffic for server backup will be offloaded from the LAN to the SAN. SAN implementation is very complex because of all of the interoperability issues that can arise. Some vendors are trying to make it easier by supplying a "SAN-in-a-box." Unlike NAS, installing a SAN is not a do-it-yourself project.

One of the greatest challenges for Sans is interoperability. The goal is for UNIX, Windows NT and Netware servers to have access to the same storage and share the same data. Today users cannot freely mix and match devices from different vendors because there are different device-level formats for each operating system. Once operating systems adopt a common structure at the device level sharing devices will become easier. Despite its current drawbacks, Sans promise relief for enterprise storage-level issues.

Performance: Sans improve performance by relieving congested LANs of high volume data traffic generated by backups, large data migrations, business intelligence systems and digital video and audio applications. Storage response time is faster because Fibre Channel links can transfer data at 100MBps. The potential problems with interoperability Sans can be difficult to manage because all of its components are designed for maximum throughput.

Scalability: Multi-channel SCSI controllers can only support a maximum of 30 devices, while a Fibre Channel fabric of interconnected switches can address thousands of ports. Bandwidth can be allocated on demand and network reconfigurations are relatively simple. Sans allow users to increase storage capacity or re-map department needs without bringing down the system and disrupting data access, as all the disks are centrally managed from one location

Availability: Sans allow distributed servers to access large, consolidated storage resources for data-intensive applications. Shared storage pools can be accessed by multiple systems. In SAN architecture, all servers can have direct access to all storage devices, allowing one server to provide fail-over protection for dozens of other servers.

Cost of Ownership: By creating a central storage pool for the entire user community, Sans can lower total cost of ownership. Fewer administrators are required to manage the storage, management is centralized from a single management interface and storage can be purchased separate from servers. The cost of storage can be amortized over more servers and the storage can be dynamically allocated and reallocated for maximum capacity usage. Also, Fibre Channel's high-speed and low latency shortens backup and restore times, freeing LANs and WANs for business applications that improve productivity and enhance revenue.

Currently, Sans are more expensive to implement than NAS because of the investment required in Fibre Channel hubs, switches and Fibre Channel-to-SCSI bridges. However, the price gap between Fibre Channel and SCSI is narrowing and the larger the enterprise, the higher the return on the investment. Given the right set of management tools, enterprises can see a return on investment in approximately two to three years.

Network Attach Storage (NAS)

The concept of Network Attach Storage is quite simple. NAS attaches special-purpose storage appliances to the LAN, which can be shared by application servers, workstations and PCs on the network. These appliances have only one job -- file serving. NAS devices can be distributed across a large network and managed centrally to provide a common pool of storage that can be shared by multiple servers and clients, regardless of their file or operating system. This enables efficient allocation of storage, alleviating the problem of one server running out of storage while another may have more than needed.

Unlike Storage Area Networks (Sans) implementation is simple and straightforward with Plug and Play compatibility. NAS appliances use standard file-system protocols, such as Network File System (NFS) and Common Internet File System (CIFS) for data sharing across multiple operating systems.

Performance: NAS is slower than SAN, but faster than DAs NAS offers users dedicated file server appliances that provide fast access and high-availability storage to UNIX and Windows NT clients on a network. Data access time is fast because the system only serves files. Files are offloaded from the host to free CPU cycles for other cycles. Separating the storage from the server also increases network reliability.

Scalability: NAS allows you to separate your storage capacity from your server capacity and with NAS companies add storage as needed. NAS products scale to multiple terabytes, and by offloading file serving to these devices, servers can support more users. But be sure to consider your future storage needs. As your requirements increase, large numbers of NAS devices can be difficult to manage. While small NAS devices are great for projects or workgroups, larger systems may be necessary for mainstream data storage.

Availability: The simplicity of NAS makes it more reliable than traditional LAN file servers and eliminates many failures induced by complex hardware and operating systems. Because the NAS device communicates directly with the client, files remain available, in the event of network server downtime, thus increasing data availability.

Cost of ownership: Specialized for high-speed file serving, NAS devices are significantly less expensive than general-purpose network file servers. Servers across different operating systems can share access to NAS devices so that enterprises can save money on hardware, maintenance and administration by consolidating data on fewer devices in a central location.

Managing Data and Storage Devices

Even though businesses are increasingly dependent upon information systems to sustain day-to-day operations, storage management has only recently become a hot topic for IT departments. Most companies have all their data reside on RAID with little thought on how to address future data growth. The common practice, up until now, of simply adding more RAID to the network and backing it up is detrimental to business operations because of the time and cost associated with it.

Another routine the IT departments used to consolidate network data had users manually remove all older files that were dormant or aged past a certain number of days. This practice worked in part because users weren't dealing with the exponential data growth. Today's electronic marketplace emphasizes accessibility of both current and past email, financial and healthcare records. So in environments where protocol to remove data from the network are not only inefficient, records for e-commerce transactions and accounting information, by law, cannot be altered or moved. So what is the solution? Many network administrators are still improperly addressing the issue by adding more hard drive space to the network via DAs or additional RAID subsystems without understanding the long-term ramifications. Adding storage to a network by increasing the RAID pool affects access speed and adds to backup window length and space requirements.

As you begin to develop a storage strategy for your network, many issues and potential problems should come to mind. How can your network operate with the increased downtime associated with backup? Where are you going to find the budget for an additional RAID system that may only last you another 6-8 months before maxing out? How are you going to finance and manage the increased IT staff needed to control your growing storage pool? The solution to your problems involves moving your network data to a cost-effective storage system while still providing users with a quick and seamless method to retrieve their data.

To manage data effectively you must understand how your users create information on the network. One of the best ways to do this is to look at your data allocation by age. Categorize your data into 0-30, 31-60 and 60-plus day old windows. Every allocation will be a little different based upon your business activities and data types. On average, most of the data on your network will fall into the "over 30" day category. Typical environments have somewhere between 20% and 30% of their data created in the last 30 days. After you complete this analysis, you should have enough information about your storage environment to forecast what potential problems are in your future.

A complete storage system could include any or all of the following types of storage devices:

Online storage devices

Online storage offers the highest performance available on the market today. As a result, hard disk systems have the highest price-point for entry or space. Active data best resides on an online storage device because of RAID's high write transfer and load speeds. Demand for primary storage is being driven by the digital convergence of automating commonplace business functions and the role of the Internet. More video, audio and image data makes its way through computer networks, driving the demand for 24/7 availability for mission-critical applications.

Industry studies have shown that only 30% of all data residing on a typical network is regularly accessed, leaving a major portion of expensive RAID systems to ineffectively store unused data. As a result, the practice of migrating data from online to NearLine storage devices has begun to grow in popularity. In applications where high volumes of data reside, a combination of online and NearLine storage devices co-existing on the same network behave as two components of a complete storage solution.

NearLine storage devices

NearLine storage devices are becoming increasingly popular in the storage market for offering multi-terabyte capacity and high performance at mid-range cost.

Data housed in NearLine storage is typically not needed on a regular basis, but when called upon, needs to be accessed quickly and automatically.

Although NearLine storage systems access data in terms of seconds, rather than the millisecond speeds that RAID offers, NearLine is not intended as an alternative to magnetic disk, but rather as a more efficient solution for growing storage requirements.

Although NearLine storage systems need software applications to operate, the need for second tier storage is increasing because current practices do not adequately protect critical data on any level. Tape backup requires excessive server downtime and provides minimal support for seamless, on-demand access to data resources. Available backup windows are shrinking as the average storage capacity grows and the organizational demand for 24/7 availability increases. By employing NearLine storage, administrators can migrate data back and forth between online and NearLine status automatically to effectively delay the next RAID purchase, lower the backup window and add the functionality of long-term archive.

NearLine storage maintains data permanently and securely, while still making it easy to retrieve, manage and control. Built around the most robust storage technology on the market today, NearLine Storage Devices provide random access capabilities, portability and a fifty-plus year shelf life, all at a fraction of the total cost of RAID.

Backup storage devices

In terms of pure functionality, and cost of ownership, tape is the best choice for network backup. With data transfer speeds of 15MB/second; no other removable technology streams data as fast.

Magnetic tape's popularity in the arena of backup storage is primarily due to its cost-per megabyte. Combined with high portability and improving random access capabilities, magnetic tape will continue to play a very important part in the complete storage strategy via backup and disaster recovery.

For large archival projects that don't need to re-visit old data, the price point and functionality of tape is unbeatable. Tape exists, and will continue to be used for network backup and disaster recovery. However in environments where archived data is retrieved and used on a regular basis, tape should not be regarded as the best choice for stable, long-term storage.

Long Term Archive devices

The final piece of the complete storage solution is long-term archive. Again, magnetic tape has traditionally been the favored as the medium of choice in this arena, however the storage industry is increasingly aware of the benefits of NearLine storage.

While primarily operating as a means to store dormant or less-requested data, the design characteristics of NearLine Storage Systems (NLSS) also lend themselves to operate as long-term archive devices. NearLine storage systems record data via laser, avoiding any physical contact with the media, eliminating any wear and tear issues. NearLine media is also not as sensitive to surrounding elements as tape, boasting fifty-plus year shelf life while offering unprecedented random access capabilities beyond the competition.

With the ability to store multiple terabytes of data, the need to store information offline is eliminated. The days of physically searching for tape cartridges stored in vaults and loading them into a drive have instead been replaced by a simple file search and execute command via computer to automatically retrieve and deliver the data quickly and seamlessly.

Summary

In order to formulate a comprehensive storage strategy, understanding your current and future storage needs will help you get a better idea of what a complete solution should look like and cost.

Up until now, system administrators and IT managers have been heavily dependent upon the old storage model of RAID and tape being their primary means to store and backup data. As a company grows, its data needs a place to live as email, applications, platforms and the advent of e-business operations emphasize consistent availability and accessibility.

IT departments know that without sufficient data storage, the company goes out of business. Only up until recently IT staffs have been resorting to what they have known all their lives by throwing more expensive RAID at their problem, hoping that their piecemeal solution will halt the progress of data growth. But in doing so, they add fuel to the fire by not only draining their annual budgets in a matter of months, but affect the overall efficiency of the network by increasing the backup window and hiring additional IT staff to manage the entire chaos.

Tape continues to serve as the primary method of backup, but without random access capabilities and a long shelf life, there is no real archive solution present. Does this sound like an efficient 21st century, e-business?

With the storage industry warming up to nearline storage, the storage model is re-written bringing not only a means to address growing data requirements, but doing so at a fraction of the total cost of RAID.

DISC Incorporated
372 Turquoise Street
Milpitas, CA 95035
P: 800.944.DISC
F: 408.934.7007

DISC GmbH
Im-Tiergarten 20-30/D-55411
Bingen, Rhein
P: +49 (0) 6721 964 430
F: +49 (0) 6721 964 414



www.disc-storage.com