Technology drives storage demand increases.

Looking forward: storage technologies on the horizon

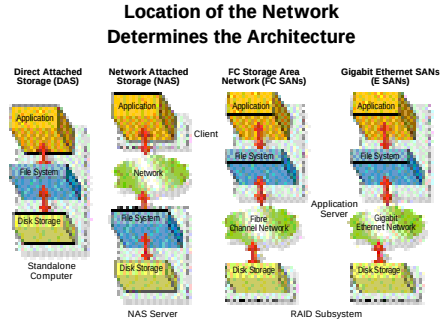
Lawrence Didsbury, Auspex Systems, Inc.

From an end user perspective, the end user has been presented with a continuing increase in computing technology over the past five years. Nearly every aspect of computing has made improvements in either speed or capacity, or both. Some examples include the increase in processor speeds from barely 33MHz to over 1GHz, the increase in local area network speeds from under 10MB/s to 10GB/s Ethernet, and hard drive speeds from sub-5000 rpm to as much as 15,000. Of course the internal bus speeds and widths on the system boards have been increased to handle the increased processing speeds. These various increases have coincided with a growth in storage capacity on hard drives from a couple hundred MB to as much as 72GB on a single drive. Other technologies such as SCSI, Ultra-Wide SCSI or Fibre Channel have been developed as well to reduce the possibility of bottlenecks in the delivery of data from the disk to the network. Many companies and alliances continue to provide the next generations of speed and capacity and are constantly carrying out further development work in this area.

While all of the hardware advances in computing are quite interesting, it is the software advances which provide the impetus for keeping an eye on the storage market. With each new advance in processing or other technology, some software is produced to take advantage of this improvement. While years ago users were content to create text files, and view very small files through DOS editors and the like, today's user expects a much more graphical and appealing interface. In addition, the user today rarely disposes of any files, and often expects to keep them for some time. It is not unusual for a corporate user to keep all of their pertinent e-mail, and expect it to be backed up and available at any given time. New software that creates movies, three-dimensional images, complex models, music and more is the order of the day. This new, compelling, content is rapidly consuming disk space, in other words, it requires storage, and lots of it. One analyst recently stated that due to the improvements in technology, more content would be created in the next 10 years than had been created in the history of mankind!

Storage, online and on demand

One very important improvement in technology that was not mentioned above was the development of the Internet and the subsequent increases in bandwidth available to both businesses and consumers. This one technology, more than the others, will have the most profound impact on the storage market. Now, users can not only create this content, but now they can share, access, and more importantly, SAVE this content for their own use. Much data that would have previously existed in only one or a few places (if on a network or on external media), will now exist in many places due to this sharing and saving practice. This activity presents new opportunities (and challenges) for IT departments. As high bandwidth connectivity to the Internet becomes more pervasive, more



One analyst recently stated that due to the improvements in technology, more content would be created in the next 10 years than had been created in the history of mankind!

large digital content will be created and shared. ISPs can provide services to their customers to enable this data storage and sharing. Eventually all users will have a significant amount of personal data, and will need to store it somewhere. As it is not cost efficient for every user to host their own file serving hardware (with as much as a terabyte of personal storage), the opportunity will present itself to the ISPs to offer online storage services.

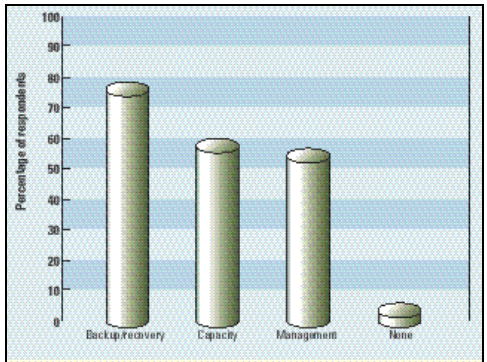
The online storage market, also called the storage on demand market, will likely start off slowly, with advances being made in time with the common availability and affordability of high-speed Internet access. The pricing of storage hardware and software will also affect the speed at which the market grows. ISPs will need to adopt cost models that enable the user to comfortably afford online storage to speed its adoption. The other hurdle that will need to be crossed by this market will be the perception by the end users that Internet storage is not secure. Security technologies like SSL, PKI, public and private key encryption will make this objection less valid. Soon all of these hurdles will be crossed and it will be common to purchase and maintain storage online. The following are some meaningful reasons why customers will consider purchasing online storage, or storage on demand:

- Purchases in accordance with need: The customer purchases storage 'on-demand,' meaning that they purchase only what they need at the time.
- Budget advantages: Customers will be able to purchase storage at guaranteed prices allowing the customer to budget accordingly for future storage needs.
- Guaranteed service levels: Customers purchasing online storage will be given service-level agreements guaranteeing certain availability of their storage. If the provider does not meet the level of service, the customer is compensated accordingly. The customer previously bore all of the cost (and headache) associated with downtime with no compensation.
- Cost reductions: 20-40% cost reductions over internal implementation. The reduction in cost is due to many factors including the training, maintenance, and overhead of storage, which will no longer be borne by the customer.
- Financial considerations: The customer will now be able to treat the storage expenditure as an operating expense rather than a capital expenditure, assigning the cost to the appropriate period of its utilisation.

Market makers like Enron Broadband Services are beginning to offer storage on demand through their online markets. Companies like Enron will create a more liquid market, and will standardize the expected prices for storage based on market demand and activity.

The initial ISP product offerings to achieve success in this space will be online backup, disaster recovery and data replication services. The reason that these services will likely be the most successful in the early stages of this market is the value that customers place on this service. Few companies have a successful backup strategy, and many backup methods, while secure, do not provide ease of operation and the speed of restore to make them entirely adequate. Online backup of files to remote storage at ISPs provides a secure, easy to access, online data vault where users can store files at their discretion. The additional value of an online backup through an ISP is the ability to access the files through any Internet connection, making the files accessible to the user virtually anywhere in the world. Disaster recovery and data replication are also services that can be offered by ISPs using existing network infrastructures and multiple geographic locations. Many companies do not have the infrastructure or expertise to effectively provide such services internally without the aid of an ISP. Market research shows that the demand for managed storage will increase significantly over the next few years.

Storage services likely to be outsourced



Backup and recovery are the services most likely to be outsourced. Capacity and managed services also are being considered for third-party management.

Source: Gartner Dataquest

Comparing DAS, NAS and SAN TCO

	DAS	NAS	FC-SAN	E-SAN
Acquisition Costs				
Hardware	Low	Moderate	High	Moderate
Software	High	Low	Low	Low
Network equipment	N/A	Low	High	Low
Operating Costs				
Management	High	Low	High	Moderate
Facilities/Maintenance	High	Low	High	Low
Opportunity Costs	High	Low	Moderate	High

NAS	FC-SAN
Network Centric	Data Centric
Network connection to data (TCP/IP – NFS, CIFS, HTTP)	Direct connection to data (Fibre Channel)
Uses existing infrastructure	Needs new infrastructure
Shared data	Co-located Storage
File system on storage server – single file system	File system on application server – multiple file systems
Mature product	Infant product
Network Security	Component Security
Plug & Play	Lack of interoperability
Highly Manageable	Lack of management standards

Source IDC

US Spending on Managed Storage Services ($M)

1998	1999	2000	2001	2002	2003
$0	**$11**	**$140**	**$643**	**$2213**	**$8000**

(Source IDC)

In addition, the overall worldwide demand for data storage will continue to rise rapidly.

Looking forward

Worldwide Data Demand (Terabytes)

1999	2000	2001	2002	2003
170,000	225,000	440,000	800,000	1,270,000

(Source IDC)

Market research also confirms that backup and recovery will be the most likely product offerings to succeed in the storage on demand market.

Which storage technology should you choose?

ISPs handle significantly more storage than typical businesses, even more than many large enterprise businesses. An ISP can have hundreds of thousands or even millions of users, each with a certain space requirement. Not all of their users will have storage space online, at least not yet…but most will have e-mail, a basic web site, and many will take advantage of their news servers. One ISP customer recently told me that they have over 200GB per month of news that gets stored on their news server. Many news groups serve large numbers of images, movies and other multimedia files, all of which are stored and served to the ISP's customers regularly. E-mail is not usually stored for any length of time on the ISP's servers, but it must be stored along with its attachments until the user retrieves it.

While the amount of content served by the ISP may require significant amounts of storage, it is not the most challenging part. Getting the data off of the disk and out to the end user in an acceptable amount of time can be a daunting task. Throughput, throughput, and faster throughput…ISPs implement high-speed networks, and pursue the most efficient storage technology in an effort to feed the insatiable performance demands of its customers. Internet customers are notoriously demanding. They want it now, and are not willing to wait. Figuring out which storage technology will give the best overall performance and value is a continuing search for ISPs.

Compounding the challenge of explosive content and performance demands, there are two major technology shifts ISPs must consider when developing an enterprise storage strategy: first, the impact of networking technology on storage architecture and content management; second, the impact of parallel processing on the design of storage products.

These two technology shifts have produced three mutually coexistent methods for connecting storage to computing platforms: Direct Attached Storage (DAS), Network Attached Storage (NAS), and Storage Area Networks (SAN). While investigating products and services to offer their customers, ISPs will need to consider the pros and cons of these storage technologies to determine which offers the best storage solution for the needs of their customers.

The following paragraphs describe each of these three storage models in detail:

(Source: Storage Architecture Guide, Auspex Systems, Inc.)

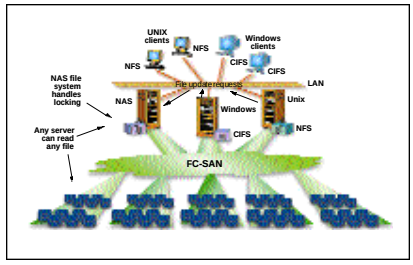
Direct Attached Storage (DAS):

Today, greater than 95% of all computer storage devices such as disk drives, disk arrays and RAID systems are directly attached to a client computer through various adapters with standardized software protocols such as SCSI, Fibre Channel and others. This type of storage is alternatively called captive storage, server attached storage or direct attached storage (DAS).

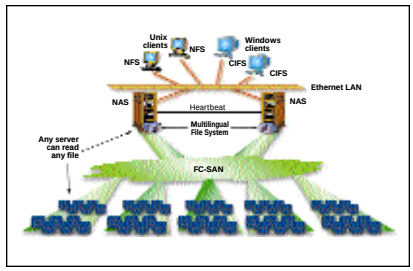
The committees that established these standards, however, allowed such wide flexibility in interoperability that there are many variations of SCSI and Fibre Channel (FC) for the many available UNIX and Windows NT systems. For example, there are seven variations of SCSI, and most UNIX vendors implement FC differently. This is because storage was local to a specific server when these standards were defined and server vendors implemented variations that were not compatible. Storage standards therefore are weak standards and driven by component considerations. In other words, the problem with storage standards is that there seems to be so many of them.

As a result of weak storage standards, third-party DAS vendors need to re-qualify their products with each revision of a server's operating system software. This can often lead to long lists of supported operating systems for SCSI or FC interconnects to different hosts. Each interconnect often requires special host software, special firmware and complicated installation procedures.

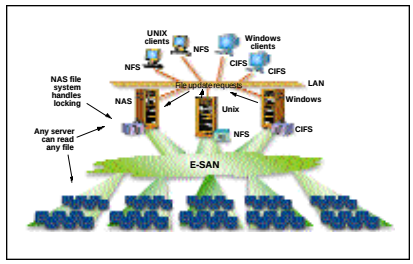
Horizontally Unified NAS and FC-SAN



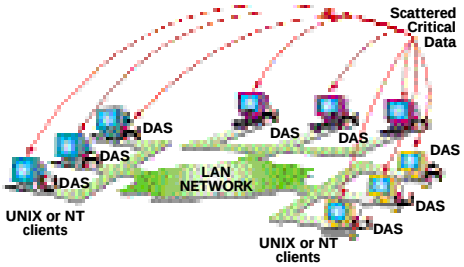
Vertically Unified NAS and FC-SAN



Horizontally Unified NAS and E-SAN



DAS Scatters Information



The Direct Attached Storage (DAS) model can be thought of as the way computer systems worked before networks. The DAS model contains three basic software layers: application software, file system software (which is part of the UNIX or NT operating system software) and disk controller software. The elements are usually located close together physically and operate as a single entity. In the DAS model, the UNIX or NT application software makes an I/O request to the file system, which organizes files and directories on each individual disk partition into a single hierarchy. The file system also manages buffer cache in UNIX. When database applications are installed, the database software sometimes bypasses the UNIX buffer cache and provides its own cache as with Oracle's System Global Area (SGA). The file system or database software determines the location of the I/O requested by the application and manages all caching activity. If the data is not in cache, the file system then makes a request to the disk controller software that retrieves the data from its disks or RAID array and returns the data to the file system to complete the I/O process.

Network Attached Storage (NAS):

In contrast to the standards inherent in DAS, network standards are strong standards that are driven by system considerations. There are two true network standards for accessing remote data that have been broadly implemented by virtually all UNIX and Windows NT system vendors. Developed and put into the public domain by Sun Microsystems, Network File System (NFS) is the de facto standard for UNIX. Developed by IBM and Microsoft, Common Internet File System (CIFS) is the standard for all flavors of the Windows operating system. As a result of these broadly accepted standards for network data access, storage devices that serve data directly over a network (called Network Attached Storage or NAS devices) are far easier to connect and manage than DAS devices. Also, NAS devices support true file sharing between NFS and CIFS computers, which together account for the vast majority of all computers sold.

The NAS model was made possible because NFS for UNIX or CIFS for Windows allows a file system to be located or mounted remotely and accessed over a network, instead of residing on the application server. In the NAS model the application software makes a network request for I/O to the remote file system mounted on a NAS server. The file system on the NAS server determines the location of the data requested by the client application and manages all caching activity. If the data is not in cache, the NAS file system then makes a request to the disk controller software, which retrieves the data from its disks or RAID array and returns the data to the NAS file system, which returns the data to the client across the network. Compared to DAS, NAS servers off-load all of the functions of organizing and accessing all directories and data on disk and managing cache. This frees the server's CPU to do additional work, thereby reducing potential CPU bottlenecks.

Storage Area Networks (SAN):

In the SAN model, the file system continues to reside on the application server. As in the case of the DAS model, the server performs its normal file system functions of organizing and accessing all files and directories on each individual disk partition and managing all caching activity. Reads and writes to the disk controller software, however, must be sent over the Storage Area Network , thereby adding latency to the I/O process and reducing performance. Unlike NAS there is no reduced workload for the server processor to offset added network latencies. However, a SAN does offer the benefits of storage resource pooling and LAN-free backup.

Comparing the Storage Technologies' Pros and Cons:

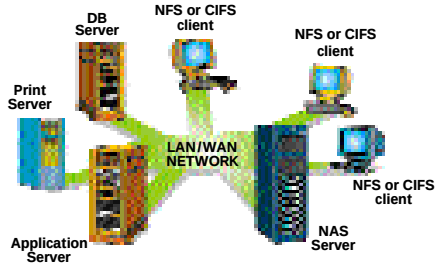
DAS:

While DAS may be easy to implement, it has many drawbacks, the largest of which is the scattering of data. With data dispersed among servers, management of enterprise data can be a nightmare. Locating, backup, and restore of data can be particularly challenging with this model. In addition to these challenges, security and asset management can be daunting as well with resources in many locations.

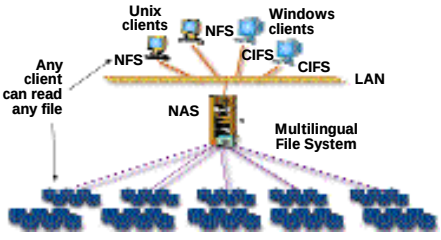
For most applications it is clear that DAS is not the best choice, particularly for ISPs, who benefit greatly from centralizing their data storage and being able to access the data over their high-speed networks.

NAS pros:

NAS Shares & Centralises Information



NAS Information can be Shared



- Easy to deploy as it uses established standards (Ethernet, NFS and CIFS)
- File level sharing among users of heterogeneous operating systems
- Fast, secure access to file level data
- Centralisation of data for access and LAN-free backup

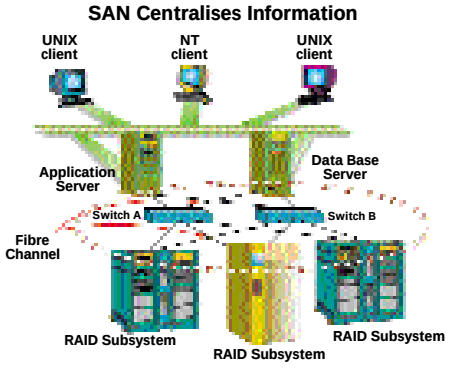
Looking forward

- Centralised asset management and security
- Low total cost of ownership
- Scalable storage independent of application servers
- Standards-based technology

NAS cons:

- A few applications require block level data
- UNIX / Windows file sharing applications are still being discovered
- Lack of benchmarks for mixed file sharing
- Lack of integration with SANs
- Initially higher server procurement costs
- Proprietary multi-lingual file systems

SAN benefits:



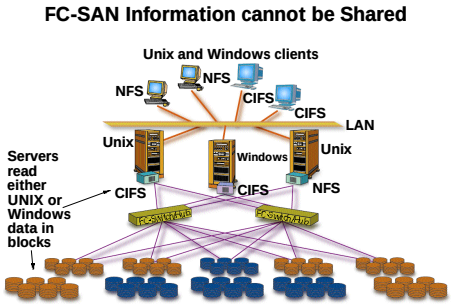
- Sharing of block-level data
- Centralisation of data and storage assets
- LAN-free backup over Fibre Channel
- Storage can be added independently of Servers
- High-speed flow of data over Fibre Channel Network.
- As much as 8 km distance between storage and servers with Fibre Channel.

SAN cons:

- High implementation costs
- Difficult to deploy and maintain
- Requires additional Fibre channel network
- Block level data requires additional server to share files
- Lack of trained resources
- Lack of reliable industry standards
- Lack of product interoperability.

What's new in the storage market?

Through continuing research in the area of storage technologies, and careful analysis of the



storage market, Auspex anticipates a convergence of NAS and SAN technologies taking place in the storage market. Auspex is helping drive this convergence and working through our alliances to drive industry standards that will enable some very compelling, enterprise storage solutions. In fact, our newest product offerings will soon incorporate the best of both worlds, providing the benefits of both NAS and SAN technologies.

ISPs should keep an eye on a few key technologies that will enable the convergence of NAS and SAN. Some of these technologies include:

- iSCSI – a proposed protocol standard to do block data calls over IP-based networks, specifically Ethernet. This proposed standard could remove the need to use additional Fibre Channel networks in the deployment of a Storage Area Network. The biggest challenge with this technology is the current size of the TCP stack. The instruction set is currently over 7000 lines, reducing the potential speed of this protocol. Hardware-assisted TCP and reduced instruction set versions of TCP are being investigated as a remedy for this problem.

- Storage over IP (SoIP) – SoIP is another proposed protocol by Nishan Systems. Nishan is trying to get this spec attached to the iSCSI standard. This proposed addition also requires interface cards, switches, and the fibre channel fabric, all of which have not been completely sorted out. Interoperability between suppliers could also be a challenge.

- Virtual Interface (VI) – Virtual interface allows for the connection of servers and NAS devices via existing IP networks. This interface is compelling because of the reduction in complexity and the fact that the file level transfers in VI can be faster than block level data transfers in a Fibre Channel SAN.

- Software Virtualisation – also called 'SAN in a Box,' software virtualisation allows disparate storage resources to be viewed as a single volume of storage to the end user.

- Fibre Channel additions to NAS products – When Fibre Channel Host Bus Adapters (HBA) are added to NAS products, a number of interesting possibilities exist. For example, a NAS server could provide a gateway to a SAN, providing the file sharing and locking capabilities that a traditional SAN lacks. Another scenario would be the ability of two NAS servers to share a common pool of disk resources, providing a High Availability (HA) solution. There are a number of configurations in this scenario such as an active-passive configuration where one system takes over in the event of a failure of the other system. Again, the end result will be a solution that achieves the best of both the NAS and SAN technologies.

- 10 Gigabit Ethernet and beyond. With Gigabit Ethernet, 10 Gigabit, and even 100 Gigabit or more, the reasons for using or continuing to develop Fibre Channel become less compelling. Fibre Channel speeds are currently reaching 4GB/s, while 10GB/s Ethernet is already stable and being deployed. Given the choice of connecting storage resources with standard Ethernet or implementing more complex Fibre Channel if the bandwidth speeds are the same or better, most ISPs would likely choose Ethernet.

Thinking outside of the box – with the convergence of NAS and SAN as described by some of the technologies above, the most interesting possibilities come into view when you begin to think outside of the traditional boundaries regarding NAS and SAN. For example, if you connect a number of NAS devices to a network switch using 10GB/s Ethernet, then connect that switch to another routing device, you have then isolated and consolidated your storage resources…Is this still NAS, or is it SAN? n

www.auspex.com