NAS - SAN Convergence Today: New Trends in Enterprise Storage

A Solution White Paper

Auspex Systems, Inc.

2002

Copyright

Auspex Systems, Inc., 2800 Scott Blvd., Santa Clara, CA. 95050 • USA

Authored by Auspex Systems, Inc.

1 Table of Contents

2 Executive Summary

The enterprise storage landscape is changing. New data additions to enterprise storage are continuing to increase at a rate that doubles required storage in less than two years. The result is a new crisis. Businesses may lack the ability to manage both the storage and the data since it is economically infeasible to increase administrative staff at the same rate.

In order to keep up with the immense amounts of data that are filling every available storage device, companies are always looking for a more efficient and cost-effective storage implementation that can be managed effectively.

Many companies have purchased a combination of storage technologies over the years. Recent trends and technology innovations provide the opportunity to consolidate all of these assets and continue to improve return on investment (ROI) and lower the total cost of ownership (TCO) for these items.

Direct Attached Storage (DAS), Network Attached Storage (NAS), and Storage Area Networks (SAN) have dominated the storage mindset for the past few years, with certain benefits derived from each. Each of these storage models has advantages, but none alone delivers the precise combination of functionality that businesses truly need. Therefore, storage vendors have been working to converge the storage technologies. However, most vendors have been focused only on their own offerings and not enabling customers to converge their existing multi-vendor storage environments.

Auspex Systems has a unique vision for storage technology convergence: the Auspex NSc3000 Network Storage Controller. Designed for flexibility, the NSc3000 will allow companies to maximize their ROI and lower TCO for their existing SAN and network storage investments.

3 Background – More, Better, and Faster Storage Solutions are Needed

Storage Demand Continues to Increase

The demand for devices to store enterprise data, storage networking devices and storage management software is increasing at a compound growth rate of 40% as estimated by organizations such as Gartner and IDC. At that rate, demand doubles in less than two years. The two major factors driving the need for more storage devices are:

- Data generated by new and existing software applications. According to Gartner, "New and existing applications are generating increasing amounts of digital data that will result in almost an annual doubling of the data stored on storage devices."
- High availability, archiving and disaster recovery requirements. The increased use of data replication for these purposes results in increased disk requirements.

Many existing and emerging applications place significant, often overwhelming demands on the existing storage infrastructures. Some of these applications include:

- **Digital Movies and Digital Imaging:** Increased use of digital imaging by individuals and business will raise storage demand substantially. Gartner/Dataquest expects worldwide digital still camera production to grow from 6.3 million units in 1999 to 16.1 million by 2004. Medical digital imaging will increase substantially. Digital imaging will also be used increasingly in security applications.
- **E-Mail:** Worldwide mailbox count was expected to reach 1 billion by the end of 2001, with more than 40 percent of the U.S. population connected at home or at work in that year. Not only has e-mail become an essential method for people to communicate at work and at home, it ranks with other corporate mission-critical applications as a major consumer of disk space.
- **E-Commerce and Other Database Applications:** Growth in enterprise data generated by database applications, including enterprise resource planning (ERP), supply chain, e-procurement, and customer relationship management (sales, marketing and service), results in the need for more storage.
- **Scientific Data Applications:** Biotechnical, geophysical, and pharmaceutical industries are making use of the increased computing capabilities in existence today, and are generating significant amounts of digital data in the process. Biotech research companies project their requirements for data storage will soon be in the range of petabytes instead of terabytes.
- **Movement of Industries to Digital Technology:** Industries such as printing, law enforcement and security are shifting to digital imaging.
- **Applications that Sample the Real World:** Applications that capture satellite data, seismic data, medical imaging, video, and audio data are increasing the demand for storage.

New storage topologies and management software are designed to use the storage better and to allow for more effective management with fewer people resources. The increased use of SANs, NAS and storage management software will help control growth in disk, tape, and optical devices in the near term. Current estimates put disk usage rates at 30 percent to 60 percent, with most companies targeting 85 percent to 90 percent.

Aberdeen Group notes that fundamentally a storage infrastructure need only do three things: store data, move data to where it's needed, and make the previous two manageable. Aberdeen also points out that the storage infrastructure must do all three things well; a failure in any one cripples the entire infrastructure. Storage and network hardware are fast becoming commodities with falling prices while storage infrastructure is becoming increasingly complex and demanding of relatively scarce, highly skilled personnel. Aberdeen notes that enterprises can typically expect storage infrastructure management costs to exceed the purchase price by more than double.

In a series of papers, Aberdeen has explained the benefits of storage virtualization, IP-based storage networking, and NAS/SAN convergence. Aberdeen has also shown that the three technologies, while independent, can provide the greatest return when combined. This is the vision of the Auspex NSc3000 Network Storage Controller.

4 The Drawbacks of Today's Storage Environment

DAS is still the most common storage used, but NAS and SANs have been gaining increasing market share. Recently the move has clearly been toward convergence of these three technologies. A convergence of the technologies makes sense because, while either NAS or SAN is an improvement over DAS, NAS and SAN tend to have exactly opposite strengths and weaknesses, as Table 1 from an Aberdeen Group presentation clearly illustrates.

Network Storage Services	DAS (Peripheral Device)	NAS (File-Oriented)	SAN (Storage-Oriented)
Storage Volume-Sharing	No	Almost never	Yes (same OS)
File-sharing	Does not apply	Yes	Almost never
Robustness	Very low	Lower	Higher
Manageability	Does not apply	Higher	Lower
Back-up/Recovery	Very poor solution	Poor solution	Good solution
Performance	Good	Lower	Higher
Purchase Cost	Very low	Low	Higher
Scalability	Very low	Lower	Higher
Network Topology	Does not apply	IP (Common)	Fibre channel (Uncommon)

Table 1: Storage Architecture Strengths and Weaknesses

The first two rows of the table say a lot. NAS is a server that provides file intelligence and, as a result, facilitates file sharing. SAN storage is actually switched DAS principally designed to provide volume sharing among servers and efficient backup.

A problem inherent with SANs is inability to provide file sharing without complicated and cumbersome additional emulation software and servers. This is because, without a layer of file intelligence, block-level data cannot be shared.

How a Storage Controller Can Help

There is clearly a difference between shared storage resources and shared files. SANs share the storage resources through a common fibre network while NAS shares files through a common IP network. NAS has the capability of delivering file sharing, with native OS permissions and file locking maintained. When SAN and NAS converge, taking the best of both worlds, you end up with Fibre Attached Storage (FAS) with a Storage Controller providing multi-OS, high performance file sharing. In addition, software features of a Storage Controller can provide additional benefits like high availability implementations, load balancing, LAN free backup, server free backup, and data replication for disaster recovery. The Storage Controller:

- Provides optimized file sharing across heterogeneous platforms;
- Supports existing SAN infrastructure;

- Uses a single vendor for shared data;
- Reduces administration costs;
- Optimizes use of existing storage;
- Offers continuous availability; and,
- Increases end-user productivity.

The Storage Controller will let users and applications deal with files without having to own the file system. It will also let system administrators manage files through a common platform. The SAN will feature storage virtualization, which provides online allocatable volume sizing and sharing among attached servers. Thus, if the files under management require more storage space, the space is created, configured, and assigned without interruption of data delivery.

The Storage Controller reduces networked storage to a single topology for shared files off the SAN as suggested in the Aberdeen table. The SAN topology still remains fibre for storage pools allocated to the Storage Controller as well as separate application servers, producing maximum ROI from the SAN.

Finally, the Storage Controller plays an important role in reducing total costs of ownership. While it does not alleviate the acquisition cost of the SAN, it reduces overall storage acquisition costs by utilizing the SAN for data storage rather than requiring the purchase of a separate NAS island. The Storage Controller also significantly reduces data managements costs. Since analysts like Aberdeen have stated that management costs far outstrip purchase costs over the lifetime of storage (as reflected in the Aberdeen table), these savings are significant over the long term.

5 Networked Storage is a Reality

A major objective of networked storage is to address the needs of corporate computing environments for storage systems that offer the scalability, flexibility, availability, and other attributes that are typically required by enterprise storage systems.

The industry is changing to a paradigm that recognizes storage as an individually managed resource – one that is best thought of separately from the computer systems (hosts) that are its consumers and beneficiaries. Storage is increasingly shared by multiple hosts and is acquired and managed independently from them. This is in contrast to the historical view (host-attached storage) that storage is an intrinsic part of a computer system, i.e. a "peripheral."

A SNIA report points out that this trend towards shared storage recognizes the critical value of the information entrusted to the storage system. It also shows that storage represents a significant portion of the investment in a typical computing environment. In turn, the trend offers the IT community several benefits, which can be summarized as:

- Increased flexibility in deploying and managing storage;
- Improved quality of service; and,
- Increased operational efficiency.

Each of these benefits is highly valued in today's businesses, which must react quickly to changing opportunities and meet rapidly rising expectations for service quality while being constrained by the need to accomplish more with limited resources.

The key enabling technology for shared storage is networking technology that can provide high bandwidth, large scale, reliable connectivity, and long-distance connectivity at a cost that makes shared storage an attractive alternative to the historical host-attached storage.

The challenge facing an information technology manager is how to protect his organization's investment in storage while meeting user demands. This requires:

- More data storage at a lower TCO and/or better ROI (through more efficient utilization) on existing storage investments.
- Better reliability and availability of data storage.
- Faster deployment and access to meet changing demands.

6 The Challenge of Data Sharing

While storage demand continues to increase, the number of users or clients of storage is also growing. With multiple users and multiple mirrored and backup copies of files, storage demand and proliferated file copies abound. In an ideal world, all users would share a single copy of each data element. Removing the redundancy in data stored would improve the efficiency and ROI of the storage investment, and lower its TCO.

An orchestra is a good analogy. The orchestra members are the users of information that support an organization's objective. In the orchestra's case the objective is to play a symphony. Not any symphony, but the same arrangement of the same one, at the same time. The sheet music (data to the orchestra) is issued once and does not change while the symphony is being played. All members are following the same, shared data, under the direction of one conductor.

Suppose the orchestra had thousands of members and the music could change for any one of them hundreds of times a minute. Alternatively, imagine if there was more than one conductor. There is a high probability that no two members of the orchestra would be playing from the same sheet of music. This is the case with data in many organizations today using host-attached storage.

The technology to insure that every member of an organization is playing from the same, shared sheet of music is available today.

- Mainframes capable of storing and distributing information from a central system have been in place for over 30 years.
- High capacity servers have been used to store and distribute information for the past fifteen years.
- Networked storage solutions have been commonly used for the past ten years.
- Database software systems that support data synchronization across multiple servers are available and used in many companies.
- Hardware and software technologies for redundant, highly available data protection ensure continued delivery of shared data to the organization's users.

The fact is that most organizations have not yet done an adequate job of making sure that all of the players have the same view of reality as depicted by the data. If IT managers could start with a clean slate — no computers, no data, no systems and no users or customers who cause changes to the data — what sort of system would they deploy? If they could make changes to the existing infrastructure to add improved file sharing capability and performance, what would provide such functionality? Systems are available today with technology that will enable effective, data sharing across an organization.

The reasons for the failure of organizations to create an effective data sharing system include:

- **Legacy technology:** Organizations are often running on multiple data sharing platforms that have been installed over the company's history. Dependence on older, inefficient, yet critical data systems often prevents successful data sharing.

- **Long implementation times:** New hardware systems take months to implement. New software systems often take years to implement. Often new systems increase the technical knowledge requirements of existing staff.

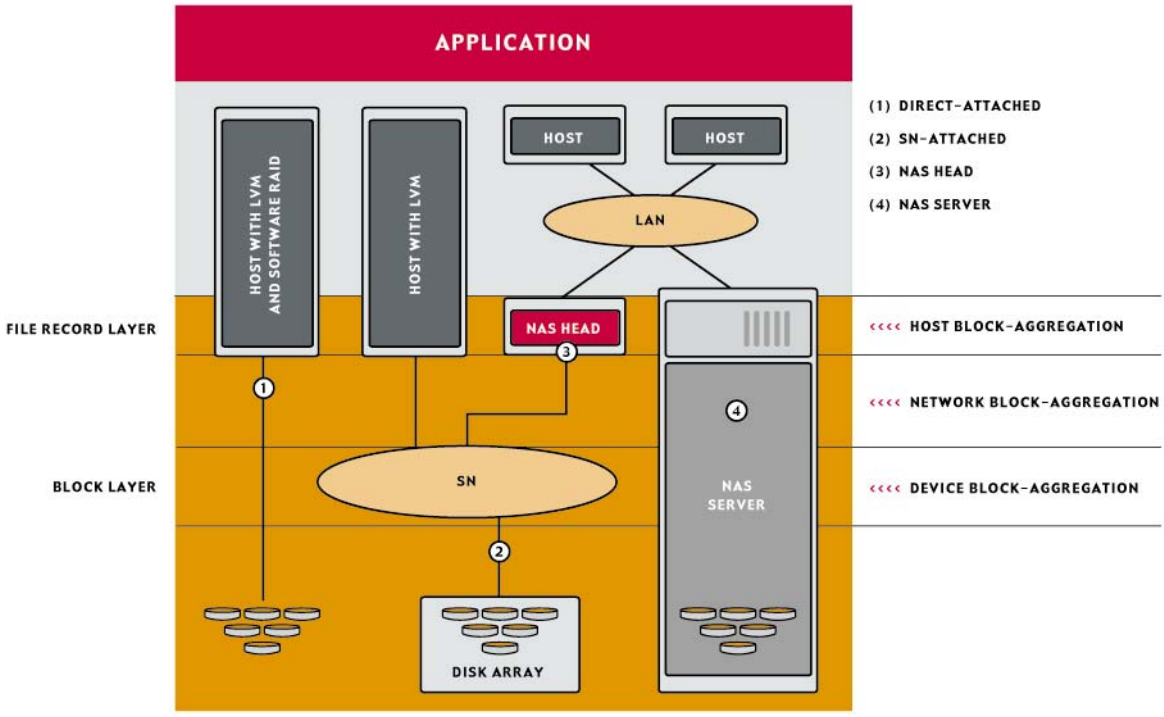
- **Changing user requirements:** Users have significant amounts of data and yet starve for information. Networking technology and the Internet have increased the amount of data available and the frequency with which it changes. IT organizations are hard pressed to build systems that create useful information from the pool of rapidly changing data delivered to the user.

The IT manager is faced with change management problems. IT systems must be architected and organized to allow for seamlessly adopting changing storage technologies while using the technology that is currently installed. This is like changing the wings on an aircraft while it is flying along.

7 Shared Storage Architecture Options

Storage Architecture Choices: DAS, SAN, or NAS

SNIA has defined a model[1] that describes the various shared storage architecture options. Figure 1 below describes that model.



Storage Architecture Choices (Figure 1)

The first two storage models, DAS and SN, are block-oriented. The remaining two are file-oriented.

Direct-attach storage (DAS): The leftmost column shows local, private storage directly connected to a single host. The host can manage the storage using Logical Volume Manager (LVM).

Storage-network (SN) attach: The second column shows a representative host connected to a (shared) disk array through a storage network. It describes a SAN. Note that a SAN is not storage. A SAN is a network for storage.

Storage Controller: The third column illustrates a dedicated-function "Storage Controller" (file server controller) interposed between the lower level, block-based storage network and its clients, connected to it through a second network. The second storage network is shown as a LAN, as that is the most common

[1] SNIA Shared Storage Model –01-06-05.doc Section 4.5 © 2001 Storage Network Industry Association

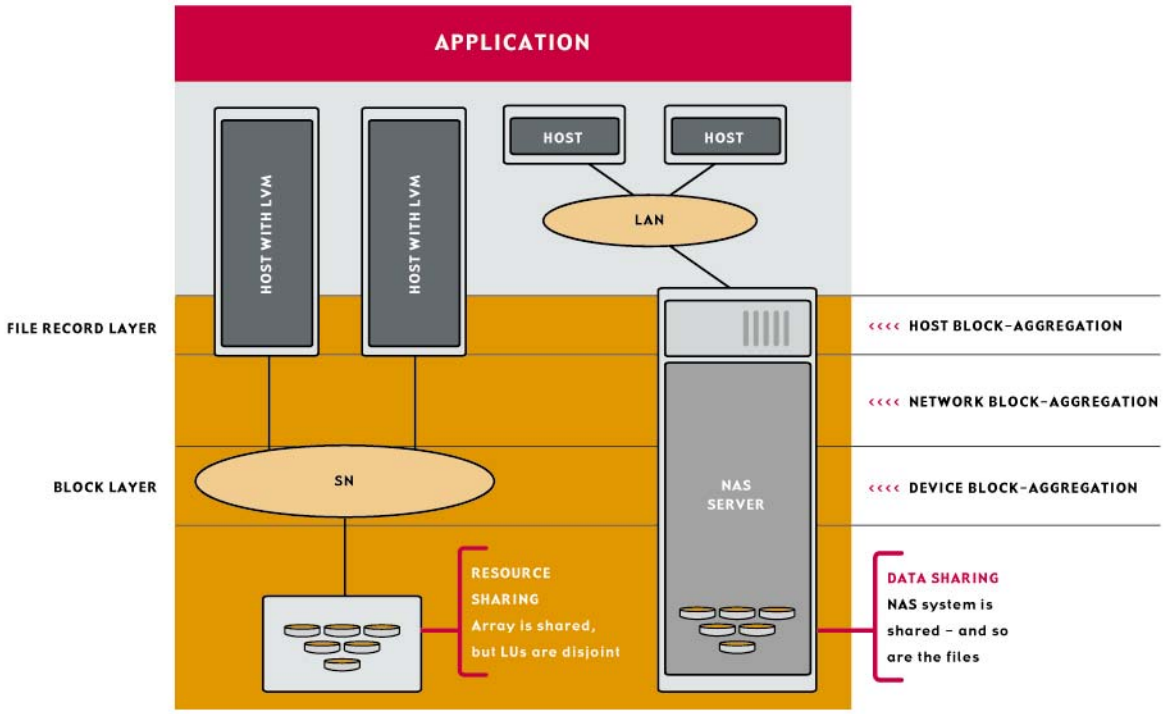
form, but it could be a SAN. Note that block aggregation functions can be used to support several Storage Controllers, as well as regular block-level hosts.

NAS server: Shown in the rightmost (fourth) column, this logically consists of a combined Storage Controller and its own local, private storage. Storage on a simple NAS is DAS to that NAS.

Sharing Data is Different from (and More Than) Sharing Storage

Storage sharing is the shared use of just the storage system resources – portions of the physical storage systems that provide its containers – not the containers (volumes) themselves, nor their contents (the files).

Shared resources commonly include the network and physical storage resources such as storage controllers, disk arrays, tape libraries, disk drives, and tape drives attached to it. A simple example is an array controller that provides a logical unit (LU) to one host and a different LU to a second host. The network, the array controller – and possibly even the array's disk drives – are shared resources, but the LUs are not shared in any way. Figure 2 below from SNIA's Shared Storage Model Report shows the difference between sharing data and sharing storage from an architectural point of view.



Sharing Data Versus Sharing Storage Resources (Figure 2)

Data sharing is the sharing of contents: two or more client entities (hosts) access and interpret the same data. (It is sometimes called "logical sharing" or "content sharing.") Data sharing requires attention to difficult issues such as coherence (maintaining cache copies "in synch" – think back to the orchestra analogy) and data format translation. Therefore, it is relatively uncommon at the block level, but much more prevalent at the file/record level.

In the most primitive form of data sharing, multiple hosts may be given read-only access to the same logical unit (or logical volume). Some logical volume systems can handle a single writer with multiple readers – but the real difficulties occur in the file/record level, where caching is done. Maintaining coherent caches can be problematic when multiple users can simultaneously alter a file. Thus, most often, such block-based data sharing is supported only within the context of cluster file systems or distributed databases on a set of clustered host systems.

File/record-based data sharing is common for the case of a single, shared server providing a client/server file access interface such as NFS or CIFS. Here, the protocols required to allow multiple hosts to access the same target file system are reasonably well embedded in the client side of distributed file system implementations. There is still room for improvement in cases where there are multiple writers for the same file, especially when it is shared among clients with different operating systems that have different rules for handling such situations. One set of rules, sometimes called "behavior," is enforced in such cases.

Successful data sharing implies that a user of an application running on an operating system such as Windows can seamlessly share data that was created by an application running on an operating system such as UNIX. Features required for successful data sharing include:

- Unified data management to reduce total cost of ownership;
- No emulation software, so that access is transparent, at high performance, and with high availability;
- The ability to use UNIX and Windows native protocols and management tools;
- The lack of "rules" (e.g. avoiding different, artificial sub-volumes to support UNIX and Windows environments), eliminating potential for file access errors and reducing administration costs;
- Common file locking that enables providing a central pool of secure, accessible storage from both Windows and UNIX, sharable without synchronization problems; and,
- Maintaining file permissions and metadata when a file is copied or restored.

The Auspex NSc3000 Storage Controller is dedicated at the file-level, can be clustered for "aggregated Storage Controller scalability," and can use virtualized SAN storage. In short, it can effectively provide the desirable functionalities of NAS while utilizing an enterprise SAN, thereby building on the strengths ofeach to create value that is greater than the sum of the parts.

8 How the Auspex Storage Controller Adds Value to Data

With nearly a decade and a half of expertise in the enterprise storage industry, Auspex Systems has the background to deliver exceptional value in the NSc3000 Network Storage Controller. Combining existing strengths in the area of high performance, high availability, file sharing and content delivery in heterogeneous OS environments, with new cutting edge technologies, the NSc3000 delivers business value with sound technical implementation.

The Auspex Storage Controller will have a unique position in the enterprise by storing all the shared files in the enterprise behind it. It is able to see what storage is behind it (from multiple vendors), what files are requested by specific users and applications, how often files are needed, and what the content of those files are. Being in this position, the NSc3000 can also add some key values to the data itself. These include::

- Storage Awareness;
- Demand Awareness; and
- Content Awareness.

Storage aware means understanding the storage pools available. Storage awareness includes cost, performance, and availability. As new storage becomes available for file sharing, there is a need to automate storage provisioning for file systems and specific files based on demand and service level agreements. In this way a storage controller can appropriately allocate files on storage with less expense, less performance and less availability based on the service level agreements.

Demand aware means understanding the usage patterns of files. Demand aware includes understanding when a file is needed and where. If a file is only used once a month, it does not need to be stored on highly available, high performance, and usually high cost storage for the remainder of the month. Instead, the file can be stored on less expensive storage during the majority of the time and then pre-staged to the storage that best fits the application just prior to its use.

Content aware file services enable information flow control. Content awareness has become prevalent on the network edges assisting with quality of service. Auspex believes that content awareness is also becoming important at the core of applications and even storage. By providing content awareness at the storage level, new capabilities become possible. Some of these include eliminating the need to run search engine spiders to capture content on changed files; the ability to establish storage and backup policies based on content; and the ability to control quality of service based on content including prioritization of file delivery as well as assisting quality of service through pre-staging data to the network edge.

The Auspex Storage Controller can provide significant value for storage customers today and is positioned to increase the value of the storage on which the data is kept as well as the value of the data itself.

9 Combining Storage Controller and Storage Virtualization Technology

Storage virtualization creates logical storage volumes that can be pooled and managed as entities while masking the complexities of physical storage devices. It can dramatically simplify the management of complex heterogeneous storage configurations by using logical representations of physical resources.

In its purest form, storage virtualization allows application or end users on any server to use, request and change available storage on the basis solely of its required attributes and without any regard for vendor, location, physical organization or media type. Ideally, this should work in a completely heterogeneous environment, which includes multiple vendors and platforms for servers, software, network elements, and storage devices. True storage virtualization requires new technologies and architectures in storage management.

The simplified management pays off to a business through lower TCO. As the costs of deploying and managing storage continue to grow faster than the actual cost of storage procurement, the idea of virtualizing enterprise storage is gaining popularity. The use of highly centralized Intranet and Internet architectures, requiring increasing amounts of storage capacity and availability of these digital assets, continuously reinforces this movement.

Storage virtualization can also improve ROI for storage devices such as redundant array of independent disk (RAID) and just a bunch of disk (JBOD) units. Without storage virtualization, such devices could not be used as part of a SAN, which for practical purposes had to come from one vendor. With virtualization, they can be.

Virtualization can also eliminate lost opportunity cost to businesses. Aberdeen Group has said that its research indicates that as much as 30 percent of perceived server downtime – during which impatient potential customers may move on to a competitor – are the result of planned downtime for storage reconfiguration. Virtual storage can be reconfigured dynamically, eliminating the downtime. Of course, eliminating downtime also raises worker productivity, so storage virtualization is worthwhile even for storage used in non customer-facing applications.

When combined with NAS and SAN architectures and server/storage consolidation, storage virtualization may be a panacea to gain control of growing operating costs for managing storage.

Combined with virtual storage, the ***STORAGE CONTROLLER*** can be to an enterprise storage manager what plug and play technology is on a modern workstation.

10 Summary

Meeting the increased demand for Enterprise Storage is a challenge that can be met with several possible approaches. Table 2 below summarizes those approaches and the potential benefits of each approach.

Approach	Benefits	Concerns
DAS	Familiar technology	Expensive to manage and wastes storage
NAS "Filer"	Easy to install and administer.	Single point of failure vulnerability, not very scalable, poor backup attributes.
SAN	Improved reliability, high availability, performance, and low storage cost.	Difficult to administer, high entry cost.
Storage Controller	Combines best of breed NAS and SAN solutions.	The concept is new.

Table 2: Enterprise Storage Approaches Summarized

Storage management issues plague organizations today. DAS has always worked in the past, but it cannot provide the data sharing organizations require, nor can it be managed or backed up effectively. NAS is a great way to enable file sharing across a user community, but is not very robust or scalable, and backup can still pose problems. Moreover, each NAS unit must be managed separately. SANs provide a good way to share storage resources across a network, and are great for mirroring and backup. A SAN can be robust and scalable but difficult to manage, and it does not offer file-sharing capability.

Business organizations are not enamored with technology, and do not necessarily care whether a storage solution is DAS, NAS or SAN. Enterprise managers care about results, and therefore demand a "storage utility" that combines the best attributes of all the varied storage infrastructure approaches.

Enter Storage Virtualization and the Storage Controller. The Storage Controller will make all of the technology in the enterprise storage pools transparent to the user. This will allow vendor-independent (heterogeneous) storage pools, which will lower the cost of physical storage and preserve existing storage investments. Disaster recovery will move from a discrete application to a standard by-product storage deployment. And, most importantly, both block and file level data can be delivered, maximizing application performance.

11 Conclusion

Enterprise data assumes its greatest meaningfulness and takes on its maximum value when it can be used efficiently and effectively throughout the entire enterprise. The Auspex NSc3000 Network Storage Controller makes that possible.

For more information, including configuration/application notes on the Network Storage Controller, or a quote for the Network Storage Controller, please contact Auspex Systems:

Auspex Systems, Inc.
2800 Scott Blvd.
Santa Clara, CA 95050
(408) 566-2000 phone
(800) 735-3177 toll-free phone

Visit us on the web at www.auspex.com

For a comprehensive review of the current storage technologies, download our **Free Storage Architecture Guides** (versions 1 and 2) at http://www.auspex.com/download/download.html