Auspex NS3000 Series

The most advanced Network Attached Storage Solutions available today

Product Guide

Auspex Technical Report

AUSPEX

Auspex NS3000 Series

The most advanced Network Attached Storage Solutions available today

Product Guide

NS3060HA Maximum Configuration
68 Terabytes of Usable Storage
1008 Disk Drives

Auspex Technical Report

AUSPEX

Table of Contents

Table of Figures

Tables

Introduction 1

The Auspex NS3000 is the World's Most Advanced Information Sharing Solution

The Auspex NS3000 series offers fault tolerance and high performance.

The Auspex Systems NS3000 is the world's most advanced solution available when large scale UNIX and Windows file serving and file sharing are required up to capcities of 68 TB. The Auspex NS3000 series offers virtually every conceivable option of fault tolerance and high performance for a secure, and consistently fast, user experience for both remote and local deployments. The NS3000 **universal file sharing solutions** directly benefit a company's business objectives, and the Auspex storage solutions directly benefit a company's IT bottom line. All Auspex products are based on a philosophy of hardware, software and service excellence that addresses every detail required by the most demanding business requirements.

Analysts predict major growth for Network Attached Storage (NAS) products such as the NS3000 over the next five years due to its major benefits in the areas of consistent availability, performance, scalability, and manageability. A thorough discussion of NAS, its benefits compared to Direct Attached Storage (DAS) and Storage Area Networks (SAN) can be found in a companion Auspex report titled ***A Storage Architecture Guide – Second Edition,*** which can be downloaded from http://www.auspex.com. This report contains specific decision criteria for an enterprise to use in making appropriate deployment choices for a particular application. It also discusses Internet Protocol (IP) based Storage Area Networks using iSCSI and other protocols.

The Six Major Business and Infrastructure Benefits of the Auspex NS3000

The Auspex NS3000 is the most advanced NAS product design available.

The Auspex NS3000 architecture is considered the most advanced NAS product design available for the specific task of serving network files with market leading performance, consistent data availability and robust security. This is because the NS3000 architecture is the only NAS product that allows any user connected to any network to share any information with full read and write privileges anywhere anytime with capacity up to 68 TB. This design improves system performance and provides for consistently high performance to users with 99.99+% data availability. Importantly, extra storage capacity can be added seamlessly without on-line system interruption or any decline in performance or availability of data. There are six major business and infrastructure benefits of the Auspex NS3000, which are shown in **Figure 1** and discussed further in this Chapter.

1. All Information can be shared
2. Performance is always fast
3. Information is easy to manage
4. Information is centralized and scaleable
5. Information is always available
6. Information is always protected from loss

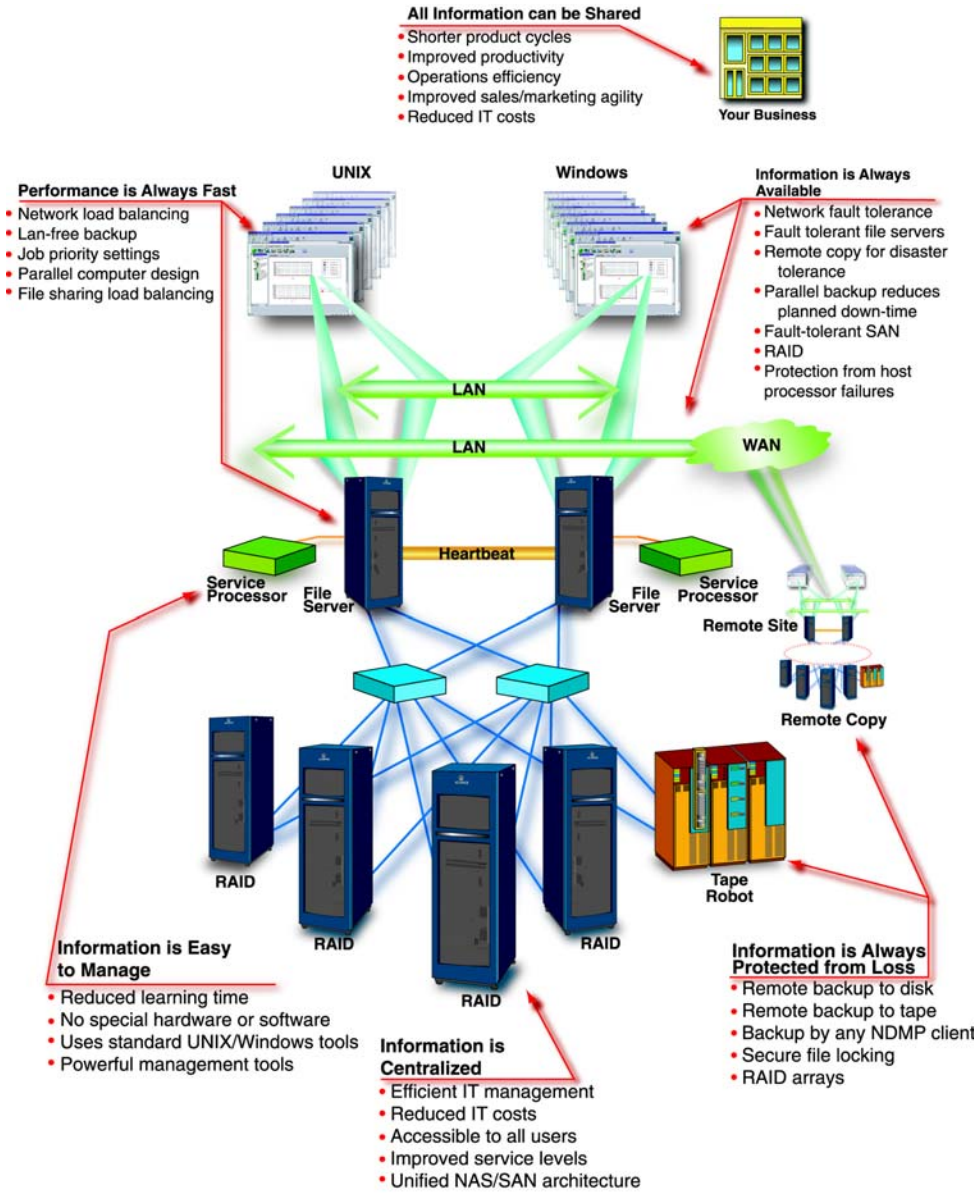

Figure 1 – The business and infrastructure benefits of the Auspex NS3000 Series.

All information can be shared

Excellence in information sharing is Auspex's commitment to its customers. The Auspex expertise in supporting file sharing applications and implementing file sharing infrastructure is unsurpassed in the market today. Unlike the NS3000, alternative solutions partition storage or do not allow full read write access to all files. Regardless of whether your employees and business partners have UNIX or Windows computers, they can't help but become more productive with an Auspex file server.

Performance is always fast

The Auspex NS3000 file server allows multiple nodes to serve files in parallel.

The commitment of Auspex is to look at the entire network, storage and computing performance environment to optimize the fast delivery of files to users regardless of the computers they are using. The parallel design of the Auspex NS3000 file server allows

multiple nodes to serve files in parallel. This provides for linear performance increases without having to add additional servers. In addition, the Auspex file server design prevents backup traffic from using network bandwidth except in the case of remote operations. This provides users with consistently fast file performance. Auspex File Servers also provide load balancing within each server for file sharing accesses. File locking and file system processing are distributed to different processors to further improve parallel performance, especially for mixed UNIX and Windows environments.

Information is easy to manage

Because Auspex file servers allow the use of all native UNIX and Windows management tools, there is virtually no learning time on the part of Network and System Administrators. In addition Auspex file servers allow for integration into existing enterprise system management frameworks through the support of industry standard protocols. No special hardware or software is required.

Information is centralized

It is well accepted in the industry that centralized files can be more efficiently and cost effectively managed. The Auspex NS3000 file server design is ideal for the centralization of files because additional storage can be easily added without degrading performance. In addition, these centralized files can be managed with existing UNIX and Windows tools that are well known to IT administrators. This feature improves IT service levels for availability and performance since fewer administrators using familiar tools have a reduced risk of human error.

Information is always available

To achieve high availability computing, it important to avoid not only *unplanned* interruptions to file and network availability, but also *planned* interruptions such as scheduled downtime for servicing and backup. Whether planned or unplanned, downtime means lack of information availability to your employees. Many network file serving and file sharing applications require files to be "always available" day and night all the time. Whether engineers are designing a new product using network file sharing or a manufacturing team is working three shifts to meet demand, availability of the latest information can be critical to effective business operations in all departments. For many companies if files are not available, employees cease to be productive. The cost of downtime can be very high!

Whether planned or unplanned, downtime means lack of information availability to your employees.

Information is always protected from loss

Whereas all of the availability features discussed in the prior section contribute to the protection of files against the possibility of loss or corruption, the ultimate protection from loss of files is multiple backup copies in a remote location.

The highest level of expertise available among NAS vendors

Being the originator of NAS, Auspex is widely considered by customers and analysts alike to be the authority in both storage and networking. Since the topic of NAS is new to many customers, Auspex is committed to provide the best public information available on optimizing the flow of accurate information and support on both a pre- and post-sales basis. Auspex sales and system engineering teams will often recruit additional technical support from Auspex resident specialists, who are experts in each of the areas mentioned in this

Auspex built the first NAS server in 1987.

report. As with any IT architecture decision, probably the most important issue is the selection of a vendor/partner with the best "total" solution. This means not only choosing a vendor who remains at the forefront of technology with the most advanced parallel architecture, but also making sure the vendor can supply the most knowledgeable professional services, and support personnel.

An NS3000 System Architecture Overview

2

The Auspex NS3000 architecture is the only NAS product that is based on a parallel processing architecture that distributes the software and processing workload to many processors including a separate service processor. The NS3000 is also unique among NAS devices with a built-in private Fibre Channel backend that is offered as a first step towards converging NAS and SAN. This design provides both NAS and SAN benefits to the enterprise. The Auspex architecture is known as the *Vertically Integrated Convergence Model* and is shown in **Figure 2**. The ultimate total integration of NAS and SAN will occur when Vertical Designs such as the NS3000 are expanded horizontally *(Horizontally Integrated Convergence Model)* to allow full file sharing with data contained on Fibre Channel SANs (FC-SANs) and SANs based on IP (Ethernet SANs or E-SANs).[1]

The NS3000 provides a unique backend Fibre Channel design as a first step toward converging NAS and SAN.

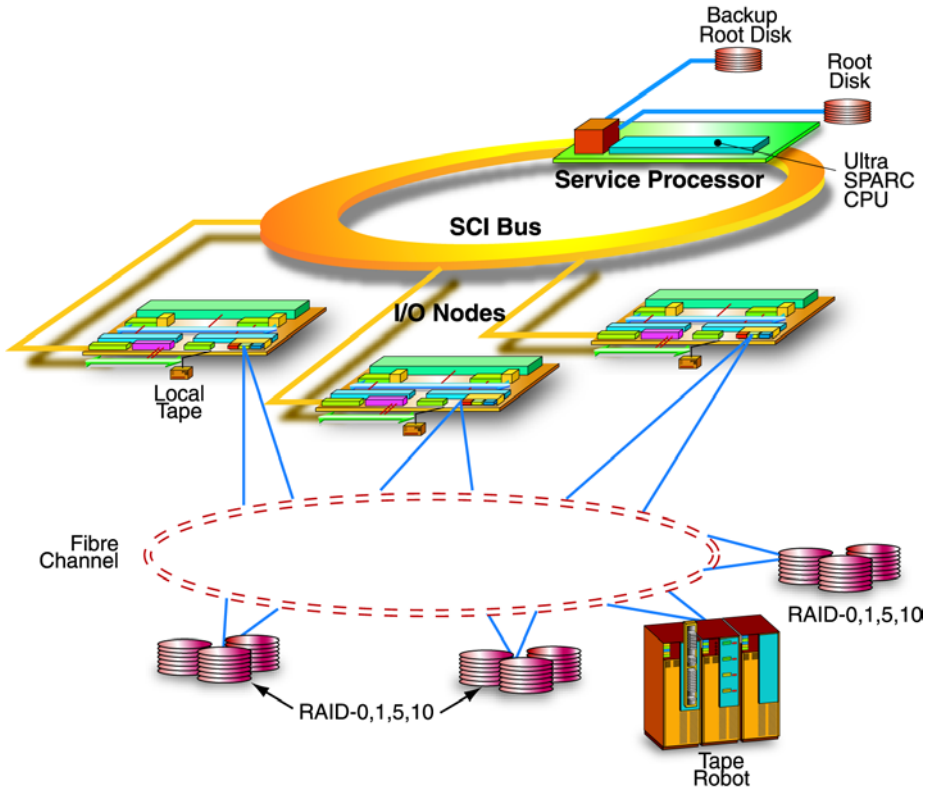

Figure 2 - System block diagram of the Auspex NS3010ER system.

[1]For a complete discussion of the converging architectures of NAS and SAN, see the companion Auspex report titled ***A Storage Architecture Guide – Second Edition,*** which can be downloaded from http://www.auspex.com or provided in hardcopy by your Auspex representative.

The NS3000 parallel architecture is known as the Functional Multi Processing or FMP architecture and has been patented by Auspex. FMP is based on a "building block" concept that is highly scaleable and easy to expand. This design integrates network storage, and file system processing thereby significantly increasing both system performance and throughput.

About the NS3000 product family

The NS3000 is available in three basic configurations, NS3000 Base Systems, NS3000ER (External RAID) Systems and NS3000HA (High Availability) Systems.

NS3000 Base Systems, Backend SAN and High Availability Configurations.

1. **NS3000 Base Systems –** The NS3000 Base System is available in a one node configuration designated the NS3010 (1-6 Terabytes of raw storage). The Base System features **direct attached SCSI disk drives** and can be configured with many fault tolerant and high performance features. It is illustrated in **Figure 3**.

2. **NS3000ER (External RAID) Systems –** The NS3000ER Systems provide multi-node support for capacities up to 18 Terabytes (TB) of storage (6TB per node). External RAID NS3000 systems differ from the NS3000 Base Systems in that they provide **for integrated backend Fibre Channel connections**. External RAID Systems are designated as the NS3010ER (1-6 TB of raw storage), the NS3020ER (6-12 TB of raw storage) and the NS3030ER (12-18TB of raw storage).

3. **NS3000HA (High Availability) Systems** – The NS3000HA High Availability Systems provide for full file server fault tolerance and are based on the Auspex **ServerGuard V software (Chapter 7)**. HA systems area available in System Pairs of **two, four or six redundant systems**. NS3000HA High Availability Systems are designated as the NS3020HA (2 redundant nodes – 4-12TB[2] of raw storage – 2-6TB of mirrored storage), the NS3040HA (4 redundant nodes - 24 TB of raw storage – 12TB of mirrored storage) and the NS3060HA (6 redundant nodes - 36 TB of raw storage – 18TB of mirrored storage).[3]

[2] The NS3020HA capacity of 12TB is available in 4Q 2001.
[3] The NS3040 and NS3060 are available in 4Q 2001.

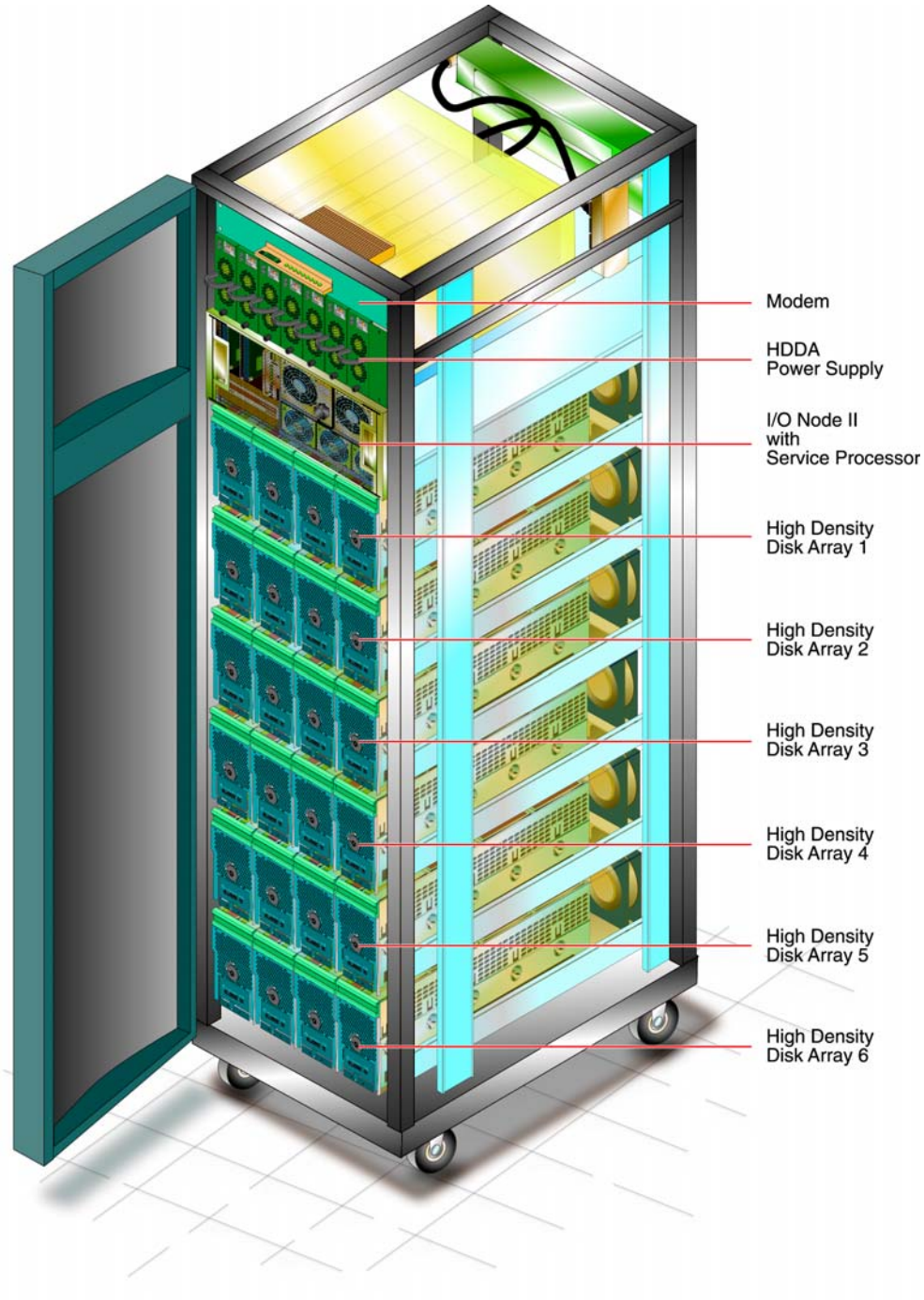

Figure 3 - A single node, single cabinet NS3010 system.

All products in the NS3000 Series family allow for efficient centralized and consolidated storage resulting in reduced management costs, enhanced network performance, increased data availability, and lower overall total cost of ownership.

Hot-swap and N+1 system power

Hot-swap and N+1 means non-stop file services.

The NS3000 power subsystem consists of one power shelf for High Density Disk Arrays (HDDAs), and redundant power cords to the I/O Node II and Service Processor module. The Power Distribution Unit (PDU) can contain from three to seven bulk 48-volt power supply modules that are hot-pluggable and hot swappable. This means they can be removed or installed during system operation. PDU's are 48v power modules that are N+1 redundant. Power supplies are added as HDDAs are added to maintain redundancy. This means that you have an extra power supply module in the shelf (i.e., the other power supplies are capable of handling the load), and a power supply can be physically removed without creating a power fault. The seven power supplies of a PDU are shown in **Figure 4** and the redundant power to the I/O Node II and Host Node is shown in **Figure 5**.

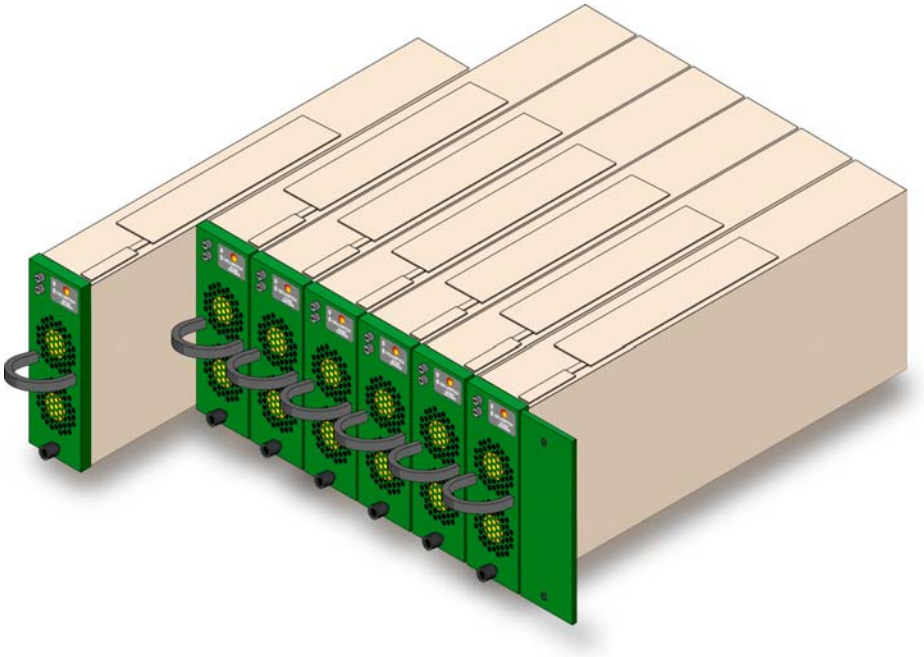

Figure 4 - The power supplies of a cabinet model NS3000 HDDA power shelf or Power Distribution Unit (PDU).

Cables carry 48-volt main system power from the back of the PDU to each HDDA in the cabinet or rack mount models where the power is converted to the required voltages as shown in **Figure 5**.

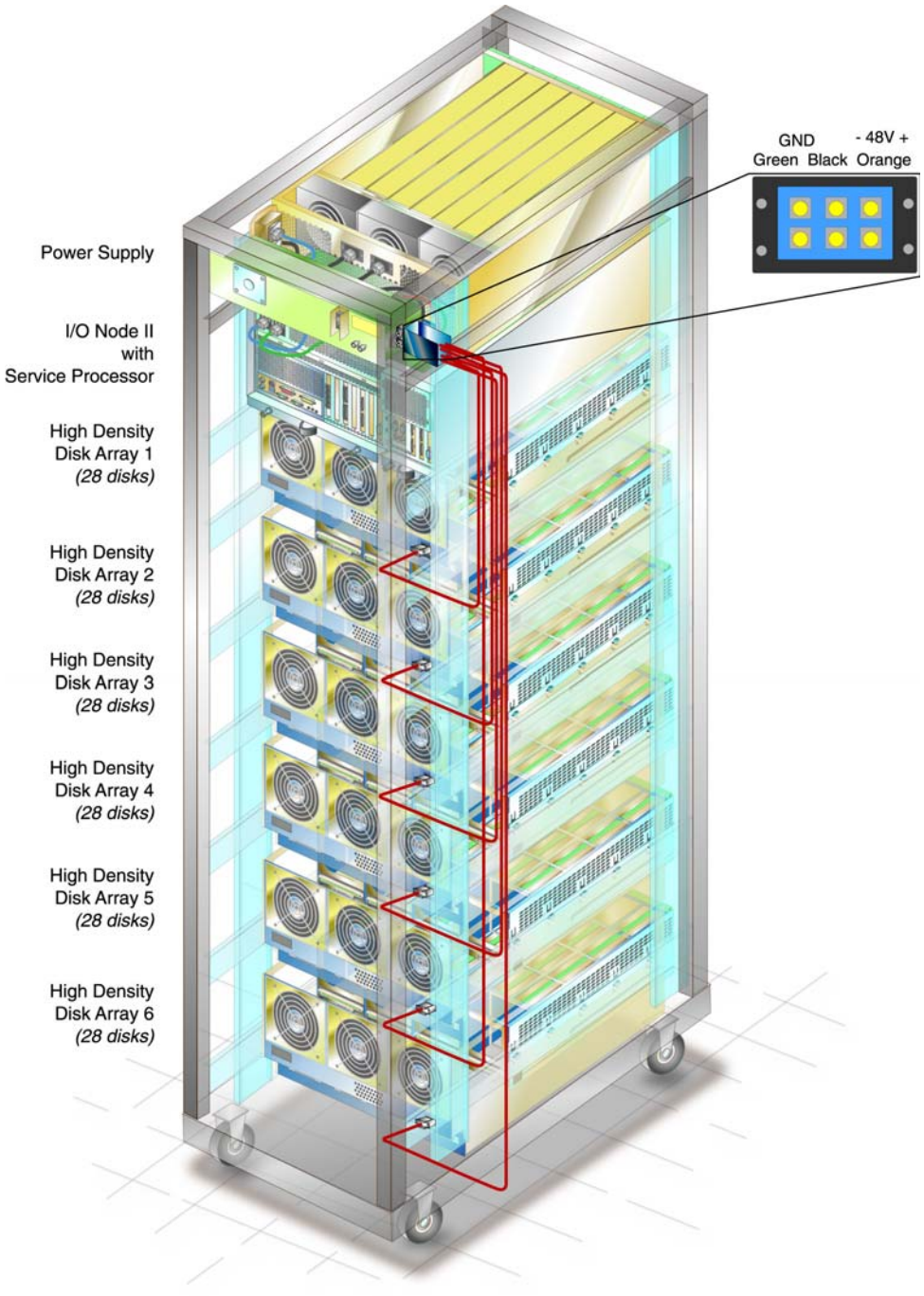

Figure 5 - Main system (48volt) power from the PDU to each chassis in a single cabinet model.

Different AC power cord options allow all NS3000 models to connect with the different electrical outlet configurations used around the world.

The Environmental Monitoring Network (EM-Net) for system status

EM-Net provides for effective preventative maintenance.

The NS3000 is designed with an environmental monitoring network (EM-Net) which is cabled in a logical ring as shown in **Figure 6**. The EM-Net is connected to all chassis and cabinets in an NS3000 system and monitors a variety of system parameters (e.g. fan speeds, power readings, voltages, and temperatures) and insures trouble free operation. The failure alerts in EM-Net can expedite system repair after a component has failed. EM-Net information is passed to the Service Processor where alerts can be viewed remotely on a UNIX console attached to the network or directly on the Service Processor console.

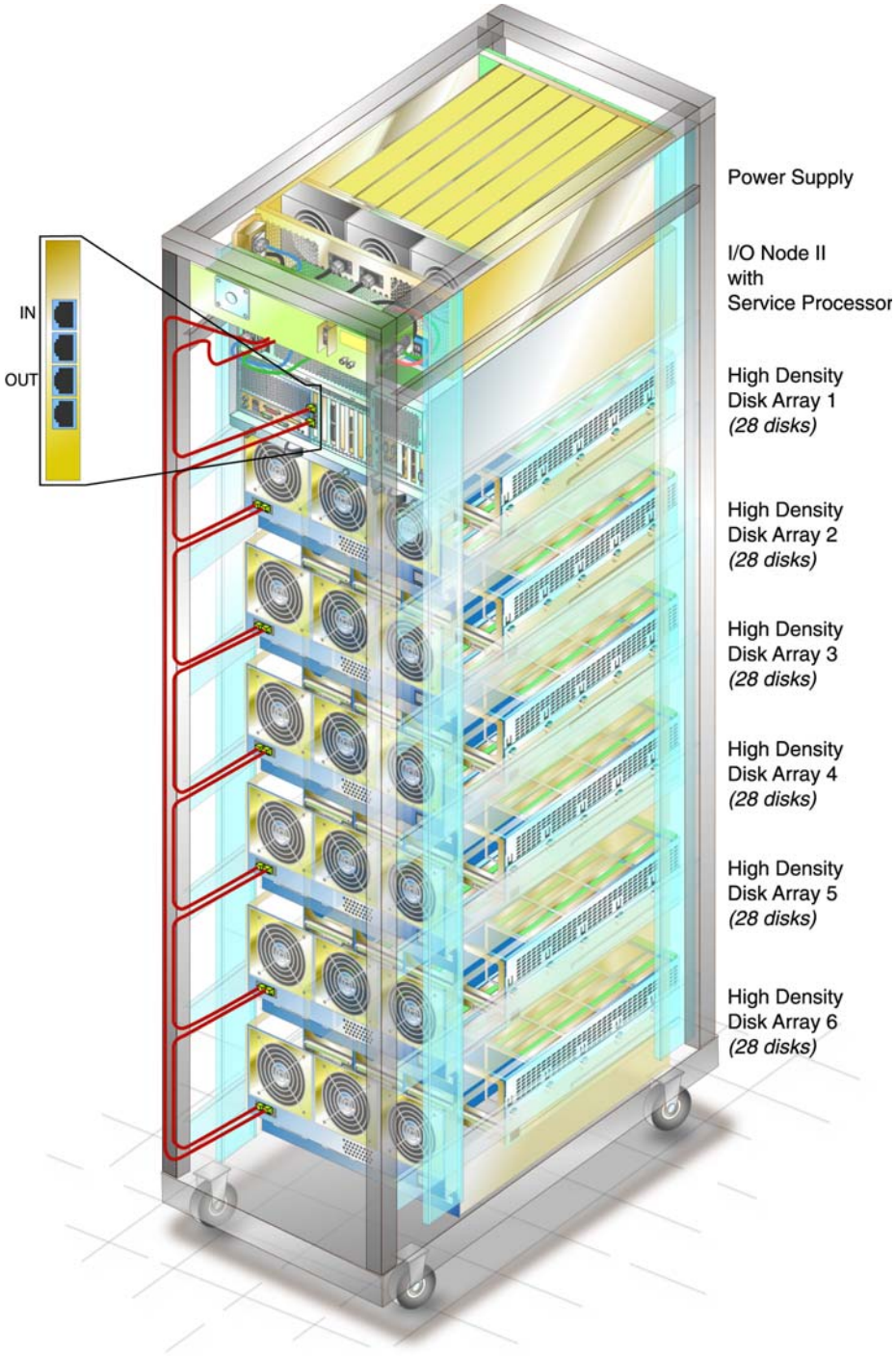

Figure 6 - Environmental monitoring cables for a single node, single cabinet NS3000 system.

The SCI bus - an efficient internal messaging network

A highly efficient 400 MB/sec Scaleable Coherent Interface (SCI) interconnect allows a single host to manage up to three I/O Node IIs. This is accomplished by the SCI bus that is a high-speed network for message exchange between the Service Processor and I/O Node IIs of the Auspex system. The NS3000 implements a unique and proprietary message passing protocol between I/O Node IIs and the Service Processor. It enables cross-mounts between I/O Node IIs so that data on any I/O Node II is available to any client on any network connection from any cabinet or any NS3000 Series system. The SCI bus is an industry standard and was specifically designed for parallel processing computer architectures.

Work is divided among seven processors.

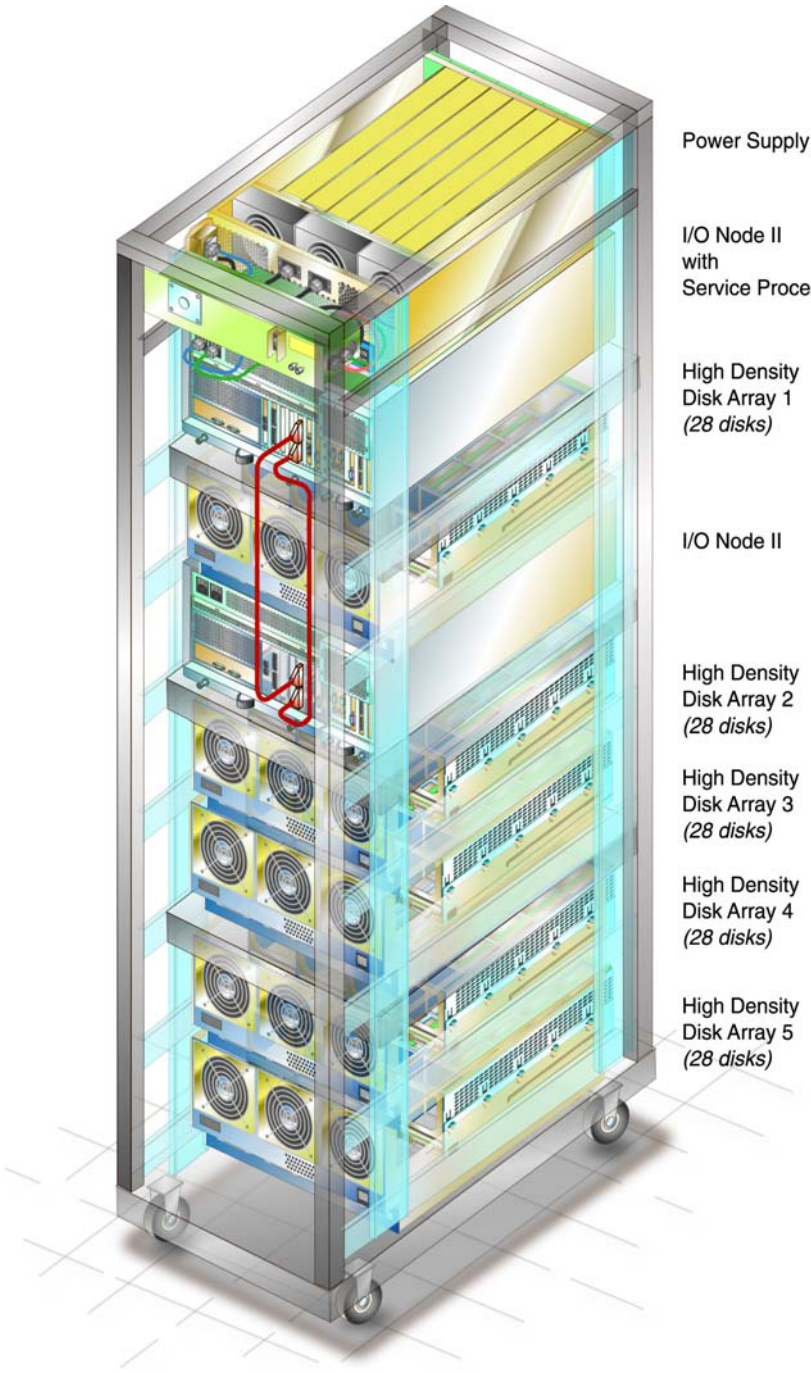

Figure 7 - 400 MB/sec SCI cable connections for a two node system.

Disk, tape, network connectivity and scalability

Each I/O Node II on the NS3000 has PCI slots for disk, tape, or network interfaces. One slot is reserved for tape (FibreChannel or SCSI), one for disk drives, and one for a high speed NIC (network interface card). The three remaining slots can be optionally used for Network Interface cards, FiberChannel or SCSI connections.

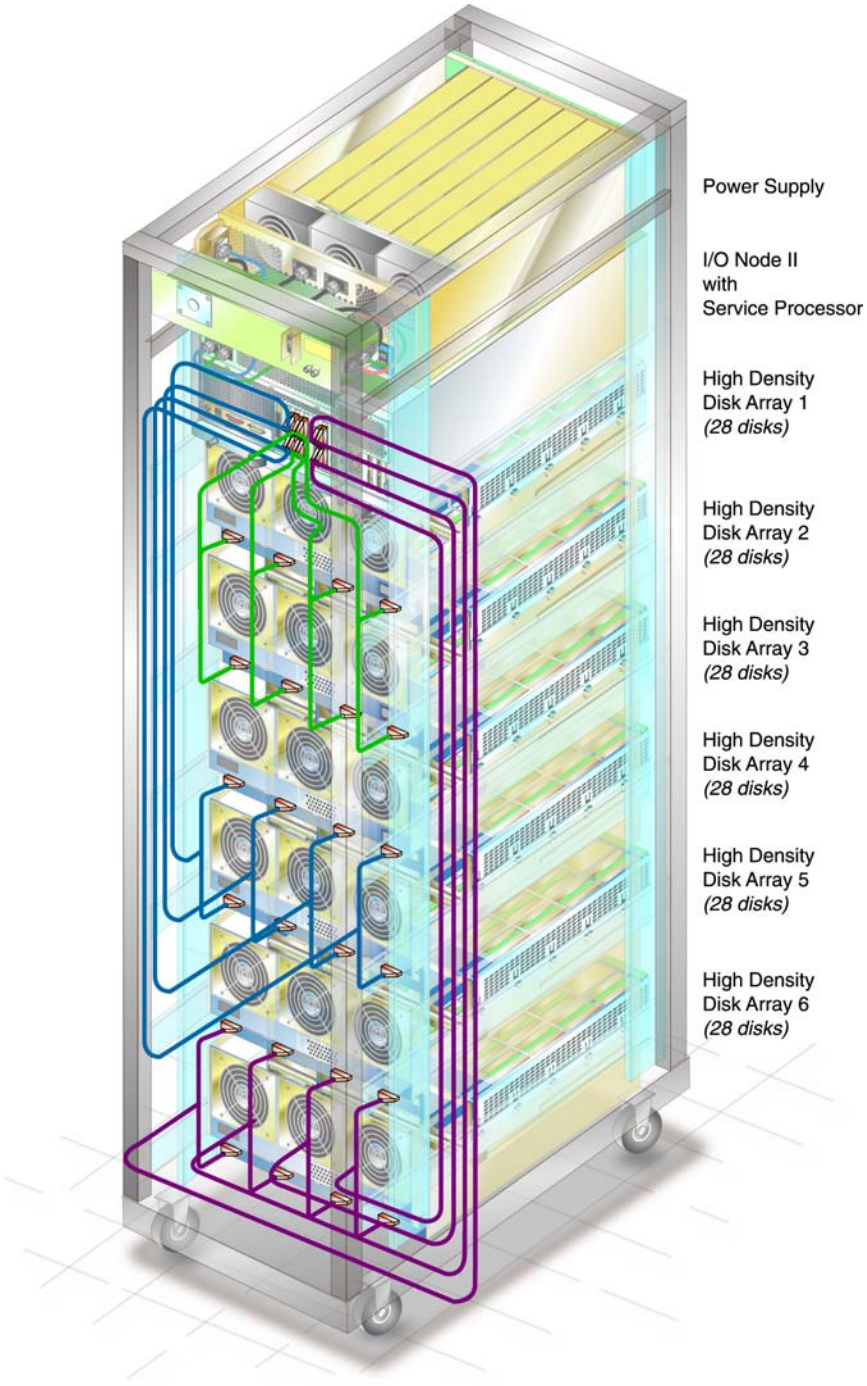

Figure 8 - SCSI cable connections from an NS3010 I/O Node II to disk array shelves.

About the I/O Node II and HDDA building blocks

Each I/O Node II, as shown in **Figure 9**, has two 933 MHz Intel Pentium III processors. Combined with the processing power of the RAID controller chips there is over 2 GHz of processing power per node.

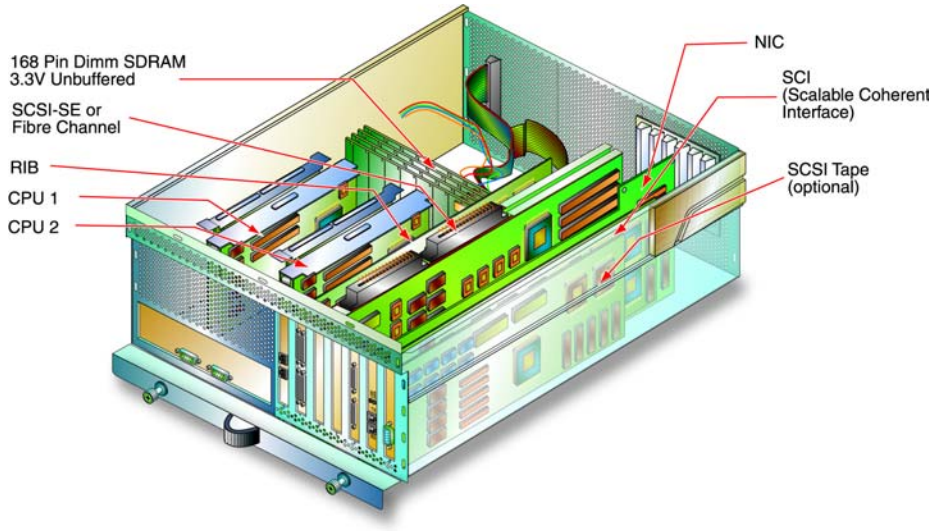

Figure 9 - An NS3000 I/O Node II and its major hardware components.

This means that a fully configured (six node) NS3060HA system has a total of fourteen processors, six for File System and Storage Processing, six for Network Processing and two for system management in the Service Processor. Each I/O Node II can be scaled to six High-density Disk Array (HDDA) chassis of 1TB capacity with four drawers of seven disk drives each. A complete six I/O Node II system can be scaled to 36 HDDAs. An HDDA contains 28 disk drives (4x7). Up to six HDDAs (or 6x28=168) drives are supported per I/O Node II. A complete six I/O Node II NS3060HA system supports 36 HDDAs and 1008 disk drives (36x28=1008). A High-density Disk Array (HDDA) shelf with 28 disk drives is shown in **Figure 10**.

Figure 10 - An NS3000 High-Density Disk Array (HDDA) shelf with 28 drives (4 drawers of 7 drives each).

In the case of 36 GB drives, this equals 6 TB per node and for 73GB drives, 12 TB per node. It is convenient to think of an HDDA as 1TB of "raw storage" for 36 GB drives and 2 TB for 73 GB drives[4].

The internal hardware architecture of each I/O Node II is based on dual Pentium III processors running on 66 MHz (528 Mbytes/sec) 64 bit system bus. Both processors reside on the bus along with a standard capacity of 2GB of DRAM memory that is used for read cache and program files. About 95% or more of the DRAM memory (1.9 Gbytes) is available for read cache with a PCI bridge to two high speed PCI busses, which act as I/O busses in the Intel-based architecture. The two PCI busses are known as the primary and secondary PCI busses. These busses support the Pentium Network Processor (NP) and the Pentium File and Storage Processor (FSP). Up to six Xtreme RAID controllers can reside on the (FSP) processor's PCI bus. Each controller contains 128 MB of Non-Volatile Memory (NVRAM) or 768 MB total per node or 2304 MB per three node system. The NVRAM write cache is battery backed up SRAM that serves as a cache for "fast writes." Writes therefore execute at memory speed for "write back" cache applications. An option exists for "write through" caching where the write goes to disk instead of cache, as required by some Oracle and other applications. A three node system therefore contains 3 x 2.0 = 6 GB of memory with 95% or 5.7GB available for read cache, and 2.36 GB of write cache.

About the Auspex NS3000 Functional Multiprocessing (FMP) parallel architecture

All hardware is based on the cost-effective industry-standard Intel architecture. Since upgrades can easily be done as faster processors become available, this is known as "riding the Intel wave". However, unlike most Intel SMP designs, each processor in the Auspex Functional Multi-Processing (FMP) design, is assigned a specific function instead of operating symmetrically. Each I/O Node II consists of industry-standard dual Intel processors, dual PCI busses, associated PCI cards and ECC memory as shown in **Figure 11**.

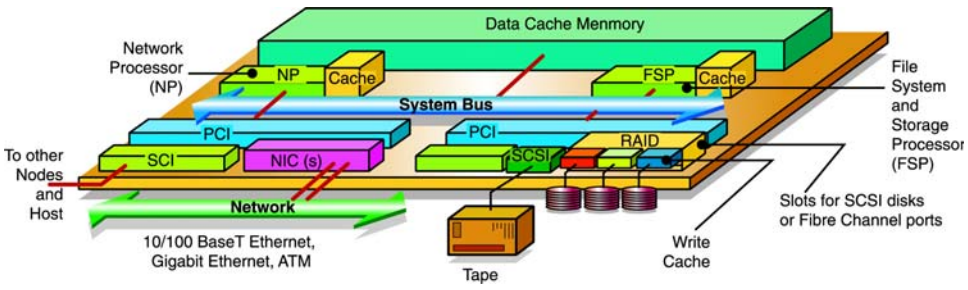

Figure 11 - I/O Node block diagram illustrating the Auspex patented Functional Multiprocessing (FMP) design.

One I/O Node II processor is known as the Network Processor (NP) and manages highly reliable proprietary software that controls all network protocol and caching functions. The other I/O Node II processor is known as the File and Storage Processor, or FSP, which also executes highly reliable proprietary software that handles file system processing and storage processing through the SCSI (NS3010) or FibreChannel (NS3000ER and NS3000HA) connectors. The proprietary software on both processors is known as the "DataXpress" kernel.

[4] 73 GB drives are available late in 2001.

FMP system software consists of a unique proprietary messaging system that enables efficient network and storage processing on the I/O Node IIs and efficient communication between the I/O Node IIs and the Service Processor. The Service Processor primarily supports system management functions. In the NS3000 architecture, each node performs its assigned function in an efficient manner. The NS3000 architecture improves data availability compared to other approaches by isolating the I/O Node IIs from unplanned outages of the general purpose Service Processor OS (Solaris). This Premier Software option is called Data Guard™ and permits I/O processing to continue even in the event of a Service Processor failure.

FMP provides functional specificity.

About the NS3000 Host Node and system management

The NS3000 Service Processor runs standard Solaris, which allows all management and control functions typically expected in a data center UNIX environment. This is in addition to Auspex Control Point™ proprietary management software that provides NS3000 specific features such as system configuration, monitoring, backup and system control. Control Point™ is a Java-based GUI program that runs in standard web browsers and allows simple and effective remote control of the Auspex NS3000 from either Windows or UNIX platforms. An NS3000 Service Processor is shown in **Figure 12**.

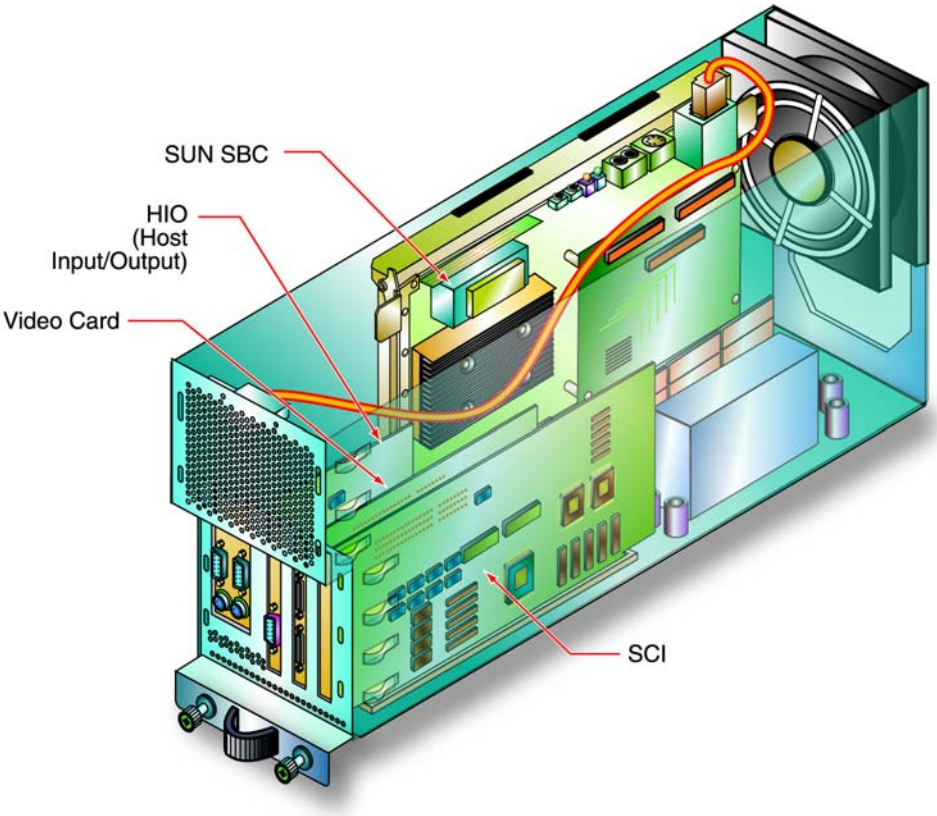

Figure 12 - An NS3000 Service Processor and it's major hardware components.

An excellent discussion of the Auspex Control Point™ management control software is found in Chapter 9 of this report.

FMP is a much better choice than SMP.

Why Auspex chose to create an FMP over an SMP architecture

Symmetric Multiprocessing (SMP) is a type of computer architecture that provides fast compute performance by making multiple processors available to simultaneously execute multiple software programs. SMP systems are suited for compute-intensive applications. In SMP, any processor executes any program or process. A variety of specialized operating systems are available to support SMP architectures.

In Auspex's patented FMP (Functional Multiprocessing) architecture, each processor executes a predefined set of programs or processes. In a highly predictable environment, like I/O and network processing, this architecture can provide superior performance and scalability characteristics.

Auspex NetServer 3000 links multiple I/O node IIs with a Service Processor for data and system management. This further distributes the work to many processors working in parallel. The efficient *Scaleable Coherent Interface* (SCI) interconnect allows the multiple nodes of an Auspex NetServer 3000 system to act as one. This provides a superfast network for message exchange between computer nodes of the Auspex system.

Multiprocessing systems are much more complicated than single-process operating systems because the operating system must allocate resources to competing programs, or processes, in a reasonable manner. The more processes an operating system must support, the more complex the scheduling algorithms necessary to accomplish this—and the more time the processors spend task switching and running scheduling programs to determine what to do next. Also, SMP machine performance degrades more quickly in performance than an FMP design. This is because as an SMP system gets busier, the processors must spend more time in scheduling work and less time performing work as in the FMP design of Auspex NetServer 3000.

Not only does the NS3000 distribute different functions to the multiple CPUs in each I/O node II, but additional functional specificity is achieved by distributing different functions, such as management control to the host node. In addition, the NS3000 design allows multiple I/O node IIs to function in parallel. This efficient parallel distribution of work illustrates the advantages of the NS3000 parallel FMP architecture compared to an SMP architecture (See **Table 1**).

This NS3000 distribution of work, both within nodes and between nodes, results in higher performance compared to SMP machines—especially at higher I/O work-loads. At the I/O node II level, this is due to reduced task switching and reduced time spent running scheduling routines. By taking advantage of parallel processing, the Auspex NetServer 3000 avoids the bottlenecks that result from scheduling complexities with heavy work-loads in an SMP environment (see **Table 1**). The Auspex design also provides greater predictability and consistency in file service performance compared to an SMP design. This arises from the greater predictability of the time to complete each program on each node and each processor because of the greatly reduced task switching and scheduling overhead compared to SMP.

Table 1 - The NS3030ER distributes processing functions not only among processors within nodes but also among processors between nodes. SMP computers perform all functions in one node.

Architecture	NS3030ER (FMP)							SMP
Processor Type	Service Processor	Network Processors			File System and Storage Processors			Host
# Processors	One	NP1	NP2	NP3	FSP1	FSP2	FSP3	One
Network Processing	No	Yes	Yes	Yes	No	No	No	Yes
File System Processing	No	No	No	No	Yes	Yes	Yes	Yes
Storage Processing	No	No	No	No	Yes	Yes	Yes	Yes
Management software	Yes	No	No	No	No	No	No	Yes
Peripheral management	Yes	No	No	No	No	No	No	Yes
Complex scheduling	No	No	No	No	No	No	No	Yes

Network Processors (NPs) of the NS3000 I/O Node IIs 3

Types of Network Interfaces Supported

The Auspex NS3000 supports 10/100baseT Ethernet, ATM OC12 and Gigabit Ethernet. The maximum transmission rates of these interfaces are shown in **Table 2**.

Supported Interface	Maximum Transmission Rate
10BaseT Ethernet	10 Mbits/sec
100BaseT Ethernet	100 Mbits/sec
ATM OC12	622 Mbits/sec
Gigabit Ethernet	1000 Mbits/sec

Table 2 - Transmission rates of supported network interfaces.

Number of Network Interfaces Supported

Complete network support.

Table 3 shows the number of each type of network interfaces supported by the NS3000.

Network Interface Adapters	Auspex NS3000 Support
10BaseT Ethernet PCI Adapters	Quad Port (4 ports per NIC)
Max 10BaseT ports per system	36 = (9 adapters x 4 ports)
100BaseT Ethernet = Fast Ethernet	Quad Port (4 ports per NIC)
Max 100BaseT ports per system	36 = (9 adapters x 4 ports)
Gigabit Ethernet	Yes (1 port per NIC)
Max Gigabit Ethernet ports per system	6 = 2/node
ATM OC12	Yes (1 port per NIC)
Max ATM ports per system	3 = 1/node

Table 3 - Number of network interface cards (NICs) supported and Network protocols supported.

Network protocols supported

The Auspex NS3000 supports the Network File System protocol or NFS version 2 (v.2) and version 3 (v.3) over UDP (Universal Data Protocol), TCP (Transmission Control Protocol) or FTP (File Transfer Protocol) with Internet Protocol (IP) routing. The NS3000 network protocol software architecture is shown in **Figure 13.**

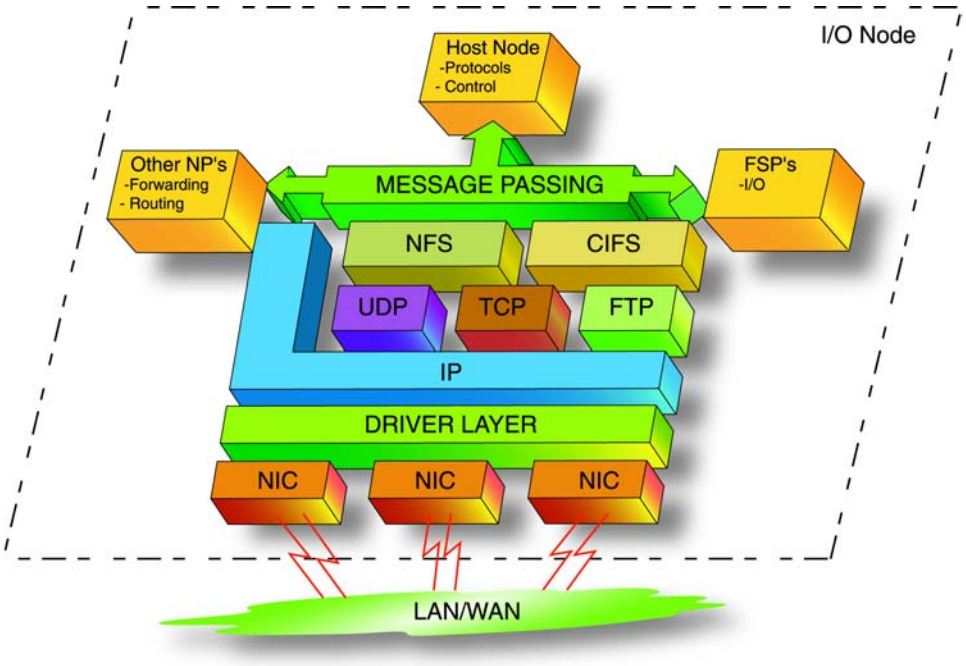

Figure 13 - NS3000 Network Processing (NP) software architecture.

NFS is the standard UNIX protocol for accessing files and printers remotely. In addition, the NS3000 supports the Common Internet File System or CIFS which is Microsoft's latest implementation of the Server Message Block (SMB) protocol. These protocols are fully integrated and provide for true UNIX and Windows file sharing which is discussed more thoroughly in Chapter 4 on File System Processing and Chapter 5 on File sharing.

The NS3000 FastFLO™ Journaling File System

4

The NS3000 FastFLO™ file system

The FastFLO file system provides local file operations similar to NFS remote operations.

At the heart of the Auspex Software Architecture is the FastFLO™ file system. The FastFLO™ file system is a file system type developed by Auspex and used in the NS3000 for file-system communication between the network processors and the file processor and between the service processor and the file processor. The FastFLO™ file system provides local file operations similar to NFS remote operations, but without the protocol processing overhead. It provides journaling, large file management (up to 512 GB per file and 1 TB per file system), flexibility, high performance and fast recovery. The Auspex software architecture is shown in **Figure 14**.

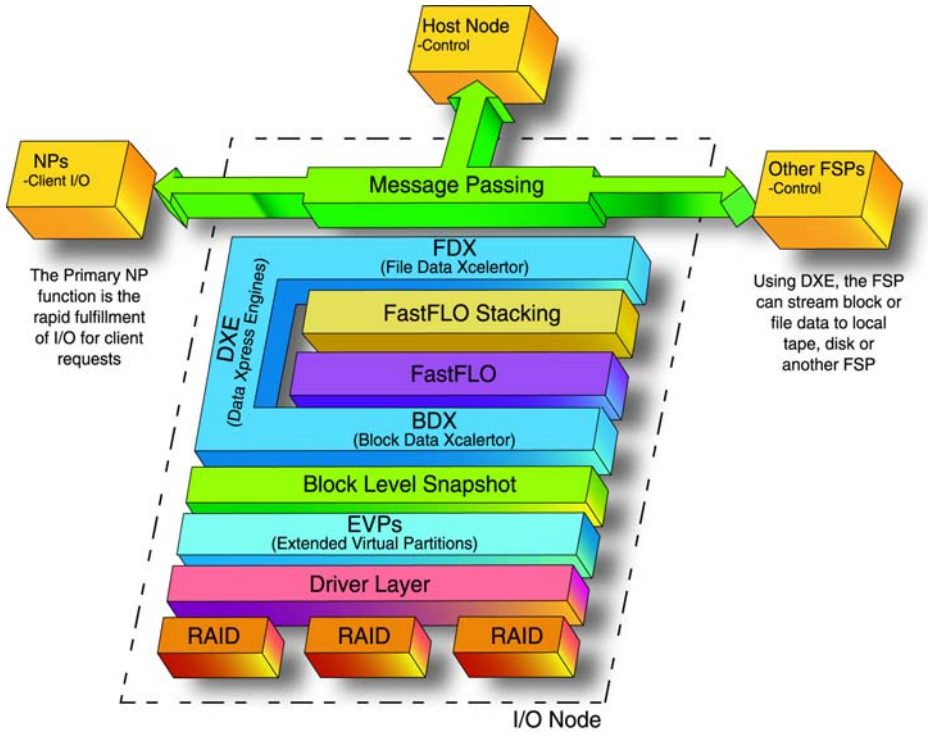

Figure 14 - NS3000 File System and Storage Processing (FSP) software architecture.

Auspex File System Innovations

File semantics and attributes of both UNIX and NT.

Although it is often overlooked, the file system plays a critical role in data management. As the hunger for disk storage continues to grow, the file system must be able to scale to handle large file systems, very large files, large numbers of files, single directories with thousands of entries, and the file semantics and attributes of both the UNIX and Windows environments. At the same time, the file system must be able to maintain its internal accounting and recover quickly in the face of an unexpected power outage or other failure.

The NS3000 architecture includes the innovative FastFLO™ File system, which is an integral part of the FSP kernel. FastFLO™ provides much higher throughput than the standard UNIX File System (UFS) and includes the following features:

- Journaling for reliability and fast recovery.
- Dramatically fast system reboot performance.
- Checkpointing increases reliability.
- Aggressive write-clustering for high performance.
- Contiguous block allocation for high speed sequential access.
- Read-ahead for high performance sequential reads.
- Support for very large files and large directories.
- Full support for both UNIX and NT semantics and metadata.
- Multiple access method support.
- Online File system expansion.
- Dynamic inode allocation.

Journaling for reliability and fast recovery

FastFLO™ uses journaling technology to ensure that file system structural integrity is guaranteed at all times. Traditional file systems always maintain a certain amount of important internal file system metadata (information about the files within the file system and about the file system internal organization) in memory to enhance performance. These data structures are periodically flushed to disk to ensure that the file system on disk is consistent. However, if a disruption occurs before the data is flushed, the file system on disk is left in an inconsistent state. This is why checking programs like fsck are required.

Unfortunately, as the administrators of most large systems have discovered, checking programs like fsck can take a prohibitively long time to run on systems with many large file systems. It is not unheard of for a system to take many hours to complete checking all file systems after a system crash. This amount of downtime is obviously unacceptable.

In addition, some of the internal data within the file system must always be written synchronously to ensure that the file system stays consistent. These synchronous operations decrease overall file system performance, since pending operations must wait while the disk (or disk volume) completes the I/O operation.

Journaling file systems like FastFLO write a record of each disk transaction to a separate log before any data is written to disk.

Journaling file systems like FastFLO™ avoid these problems by writing a record of each disk transaction to a separate log before any data is written to disk. If a failure occurs, the file system need only examine the contents of the log and verify that transactions in progress have been completed. Incomplete transactions are either completed from information in the log or backed out to return the file system to a consistent state. This typically takes a few seconds versus the hours that checking programs may spend checking all file system data structures for consistency. Within FastFLO™, a transaction daemon processes outstanding disk transactions, groups them together to minimize the number of I/O operations, and commits them to disk. By grouping the transactions together, multiple metadata operations that affect the same disk block can be written together, thus significantly reducing the total amount of I/O that must be performed. The transaction daemon is responsible for recording the data in the log before the FastFLO™ transactions are committed to disk. On the NS3000, FastFLO™ uses non-volatile RAM to store its transaction log.This further accelerates file system performance since NVRAM can be written much more rapidly than disk.

Dramatically fast system reboot performance

The FastFLO™ file system has a dramatic effect on overall system reboot performance, especially after a system crash or other unexpected outage. Journaling allows the entire system, regardless of configuration, to recover and come back online in a few minutes,

just as it would from a clean system boot. It is important to note that FastFLO™, like almost all other file systems whether they use journaling or not, will lose transactions that are in memory but have not yet been logged or committed to disk. For network clients using NFS this does not present a problem since NFS writes are either required to be synchronous (NFSv2), or are performed using a safe asynchronous write protocol (NFSv3). Both mechanisms ensure that data is committed to disk before the client recognizes the write as successful. CIFS, however, does not require this guarantee, and thus, some client data could be lost if a system failure occurred at the wrong time.

Checkpointing increases reliability

FastFlo™ checkpointing aggressively flushes data to disk.

FastFLO™ provides an optional checkpointing mechanism that increases the frequency with which user file data is flushed to disk. This mechanism aggressively flushes data to disk when there are no outstanding transactions or when demand on the file system is low. Checkpointing improves the reliability of the file system without adversely affecting performance. and is primarily applicable to file system metadata, CIFS, NFSv3 and data written from the host. As mentioned, NFSv2 requires that all file writes be performed synchronously. (These writes are actually cached in NVRAM on the RAID controllers and don't have to wait to be written to disk.)

Aggressive write-clustering for high performance

FastFLO™ uses a read and write cluster size of 128 Kilobytes. Clustering for both reads and writes is performed using a 128KB window. For write operations, when "dirty pages" (pages with updated data) are ready to be written to disk, the file system will cluster all the pages, construct one logical request to the disk driver and issue a write.

Contiguous block allocation for high speed sequential access

Disk blocks in FastFLO™ are allocated only when the file system is ready to write data to disk.

FastFLO™ uses a delayed block allocation mechanism for allocating disk blocks to files. Unlike UFS where blocks are allocated early, disk blocks in FastFLO™ are allocated only when the file system is ready to write data to disk. The file system attempts to allocate contiguous disk blocks for all the dirty pages that make up the cluster. This helps to maximize contiguous block allocation. Contiguous block allocation is highly advantageous for applications like Mechanical Design (MCAD) where a relatively small number of large files are read and written by a few clients. Under such conditions, the I/O request stream seen by the server is often highly sequential. By storing file data contiguously read performance can be substantially improved.

The block allocation mechanism is further tuned for the underlying RAID 5 implementation such that it generates full stripe writes whenever possible. Full stripe writes allow a full stripe of data plus parity information to be written to the RAID array without requiring any data to be read from the stripe. By comparison, partial stripe writes require data to be read from the stripe to generate correct parity.

Read-ahead for high performance sequential reads

Read clustering in FastFLO™ is enabled using built-in heuristics to read ahead for every file. The heuristics are applied to determine if the file is being accessed sequentially. If so, then the file system reads ahead pages corresponding to that file. Read-ahead helps to ensure that when a client read request is received the requested data will already be stored in the data cache, so the request can be satisfied immediately.

FastFLO allows file systems and individual files to scale up dramatically.

Support for large files and large directories

FastFLO™ features full support for large file systems, large files and large directories. All file offsets are stored internally as full 64-bit numbers, allowing file systems and individual files to scale up dramatically. The standard method is a linear search similar to the method used by UFS and works well in most situations, but a different method can be selected (either at mount time or using administrative tools) to accelerate directory access for directories with a very large number of files.

Multiple access methods support

FastFLO™ supports multiple access methods for data. An access method is an attribute of a file. Access methods allow an administrator to select the method of storing file data that is optimal for the particular data type. Example access methods include, standard byte-stream I/O, Windows file system compatibility, and record oriented I/O. As the need arises, new access methods can be designed to meet particular needs without necessitating a re-design of the file system. For instance, applications like those used in the Oil and Gas industry perform sequential access to very large files. When handling files that approach or even exceed the size of the data cache, the normal caching algorithms are no longer advantageous.

Dynamic inode allocation

A direct I/O access method would allow such applications to access files directly, bypassing standard caching mechanisms. FastFLO™ dynamically allocates inodes in the file system, as they are needed. There is no need for the system administrator to manually tune the number of inodes. This allows the file system to easily support a large number of small files such as are typically created by news feeds or a small number of large files, efficiently.

Online file system expansion

Should a file system need to increase in size, expansion can be accomplished online. FastFLO™ is expandable to 1 TB through RAID VP Expansion.

NeTservices: UNIX and NT File Sharing on the NS3000

5

The Importance of Universal File Sharing

The NS3000 supports both NFS and CIFS network file sharing protocols to allow ***universal file sharing*** between UNIX and Windows NT hosts. This is important because universal file sharing can often result in a business advantage for the enterprise. For this reason, about 75% of Auspex users share data between UNIX and Windows NT. The NS3000 file sharing feature offers competitively unique universal file sharing where any user can have full "read-write" file priviledges to any file on an NS3000 Series system. This is a major advantage compared to competitive products where emulation software is installed or files need to be partitioned where certain users have "read only" access. In a software development environment or CAD environment, some engineers may use UNIX workstations whereas others use NT to access and update the same data.

There are four basic approaches for supporting mixed UNIX/NT environments.

- Separate UNIX and NT servers accessed by different clients.
- Client-based emulation.
- Server-based emulation.
- Bilingual Network Attached Storage (NAS) with universal file sharing such as the Auspex NS3000.

These approaches are discussed in detail in Appendix A.

With NetServices all Information can be shared

Excellence in information sharing is the Auspex's commitment to its customers. The company's expertise in supporting information sharing applications and implementing information sharing infrastructure is unsurpassed. Regardless of whether users have UNIX or Windows computers, they can't help but become more productive with an Auspex file server. Heterogeneous information sharing between UNIX and Windows hosts from a single image of data is perhaps the most important feature to evaluate when selecting a file server for file sharing to enable business benefits. The Auspex NeTservices™ software solution for UNIX and Windows data sharing represents the *best of breed* available in the industry. This is due to NeTservices™ concurrent native file locking routines for both UNIX and Windows universal information sharing protocols **and** an Auspex unique performance load-balancing feature between UNIX and Windows users.

Universal File Sharing can be critical to productivity in all functional areas of business

Information sharing between employees and business partners has been shown to dramatically increase productivity across all functional departments of a business. If an enterprise's *Engineering* department designs products using computer software for UNIX and Windows platforms, then NetServices **universal file sharing** can be of critical importance to

productivity. Universal file sharing also helps a business increase the efficiency of *Manufacturing* operations in the areas of Supply Chain Management (SCM) and Business to Business Integration (B2Bi) systems. File sharing solutions can increase *Marketing* and *Sales* agility relative to competitors to generate more business through effective collaboration. The business benefit of the Auspex infrastructure solutions is in streamlining *IT* operations and reducing costs.

UNIX and Windows clients can share any file with full "read-write" privileges

Unparalleled performance.

NeTservices™ is an Auspex software solution that allows UNIX and Windows clients to share any file securely stored and managed on Auspex file servers with full "read-write" privileges. The functionality of this software is shown in **Figure 15**. Since only one physical copy of data is stored on the Auspex file server, costs are reduced and true high performance multi-lingual file sharing is enabled. Since no client emulation software is required, access is transparent to users and there is no performance penalty as in implementations with Samba or other intermediate software solutions. Administration is facilitated through native UNIX or Windows tools so all existing IT skills are preserved. Through Auspex's "One-World" view of UNIX and Windows files, and the distribution of file sharing operations to multiple processors in the file server, unparalleled performance and data security are possible. This file sharing load balancing is also industry unique and enables unparalleled performance and data security are possible. In addition, NeTservices™ requires only one physical copy of data which increases data reliability and ease of administration dramatically. Since no client software is required, access is transparent to users. Through an optimized implementation, the NeTservices™ file sharing, load-balancing design between UNIX and Windows provides even performance, reliability, and scalability to all users.

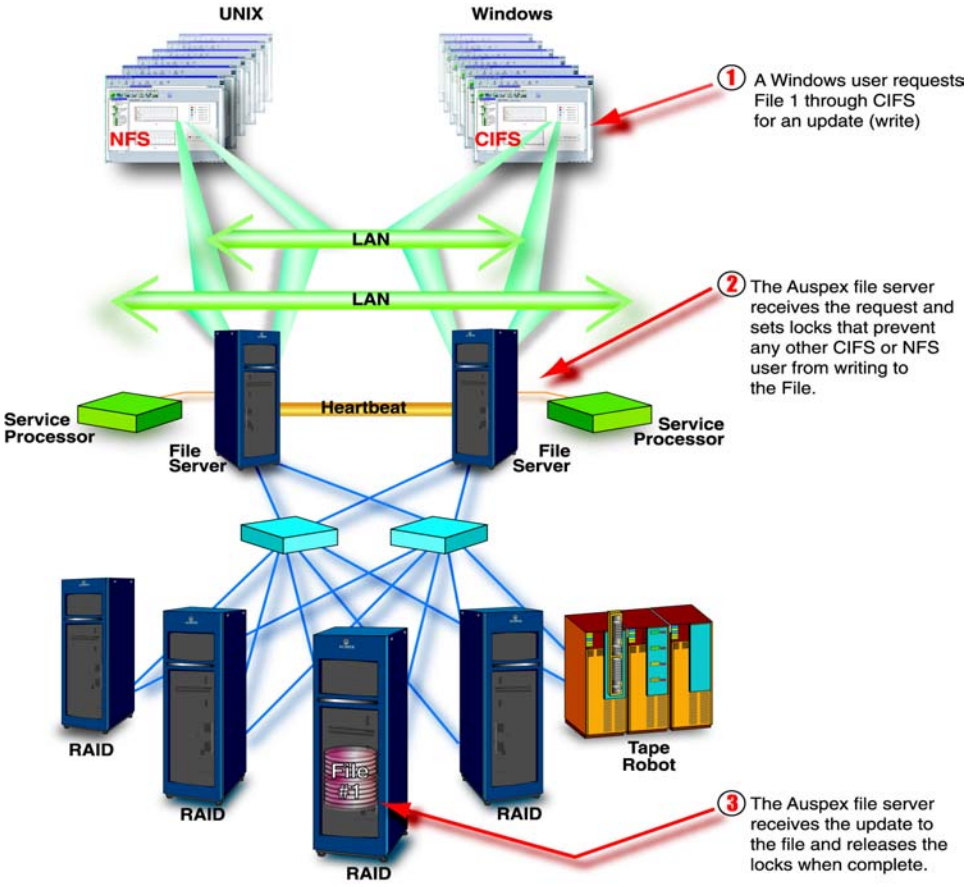

Figure 15 - NeTservices™ provides secure universal UNIX and Windows file sharing.

NeTservices supports both Windows NT and Windows 2000

NeTservices supports an implementation of Microsoft native CIFS file sharing protocol that leverages the proven FMP architecture.

NeTservices is the premier solution for deploying enterprise-level, shared file services for mixed UNIX and Windows environments. It allows Auspex NS3000 systems to provide data consolidation and file-sharing for such environments without compromising support of native file access performance, Windows NT or Windows 2000 integration and NT remote administration capabilities. NeTservices supports an implementation of Microsoft native CIFS file sharing protocol that leverages the proven FMP architecture to provide industry-leading performance and scalability. Furthermore, it delivers the NT 4.0 and Windows 2000 networking environment including directory services[5], file security, and remote administration, that is essential for deploying Windows NT or Windows 2000 in corporate environments. Finally, NeTservices delivers a best-of-breed administrative environment that fully supports remote Windows NT or Windows 2000 administrative tools for user/group account and file server properties management as well as enterprise-level administration tools for managing disk storage, RAID, and backup/restore. NeTservices was developed to deliver enterprise-class file services allowing data consolidation and file sharing in a mixed UNIX and Windows environment. It provides the following benefits:

Enterprise benefits of UNIX and Windows file consolidation

With NeTservices, both UNIX and Windows data can be managed on the same Auspex server, reducing costs, and simplifying management. Furthermore, customers can now obtain very high levels of data availability for both UNIX and Windows data by consolidating server attached storage to the NS3000.

Secure, "single image" file sharing among UNIX and Windows users

NeTservices allows customers to manage only one physical copy of the shared data. It provides features allowing transparent, yet secure sharing of individual files by UNIX and Windows clients. Further, it provides support for mechanisms allowing data protection in situations with concurrent file access by UNIX and Windows clients.

Enterprise-level Support

NeTservices is an Auspex-developed product. It is sold and supported by Auspex. As such, enterprises can benefit from Auspex's pre- and post-sales support organization's experience in supporting enterprise-level file services deployment.

NeTservices Performance

A key requirement for support of enterprise-level shared file services is to concurrently deliver NFS and CIFS file services.

In order to support the overall goal of enterprise-level consolidation and sharing of UNIX and Windows data, NeTservices delivers an optimized implementation of the CIFS file services protocol providing very high Windows file services performance. A single I/O Node II provides significantly greater performance than a 4-way SMP Intel-based NT file server does. The Auspex CIFS implementation delivers near-linear CIFS performance scalability as additional I/O Node IIs are added to the server system. A system with multiple I/O Node IIs can be expected to provide performance equivalent to multiple NT file servers, all in an easy to manage single system image.

[5] Directory services support will be available Q4, 2001. NeTservices has been certified by both Veritest and ZD Labs to be compatible with Windows 2000. In addition, the NS3000 was also certified compatible for use with Microsoft's SQL Server 2000, SQL Server 7.0, Exchange Server 5.5 and Internet Information Server (IIS) 4.0 and 5.0..

A key requirement for support of enterprise-level shared file services is for the data consolidation and sharing platform to concurrently deliver NFS and CIFS file services with undiminished performance and scalability. The Auspex CIFS implementation allows each I/O Node II to simultaneously deliver NFS and CIFS protocols with sustained performance and scalability.

Windows Domain Security

The Windows domain security model provides single log-on capability for NT networks. A Windows Domain is defined as a group of servers running Windows Servers that share common security policies and user group account databases. Therefore, the Windows NT Domain is the basic unit of security and centralized administration for Windows clients and servers in the domain, which in some ways, can be viewed as a single system.

One server running Windows Server acts as the Primary Domain Controller (PDC), that maintains the centralized security databases for the domain. Other computers running Windows Server in the domain function as Backup Domain Controllers (BDC) and can authenticate logon requests. The PDC or BDC authenticates users of a Windows Domain. Changes in security policies are implemented on the PDC and transparently replicated to BDCs. An alternative configuration is to setup a Windows server as a stand-alone server that can participate in a domain and share its resources with other nodes on the network. This is typically referred to as a Member Server.

Another key concept in Windows Domains is the Trust Relationship. A Trust Relationship is a link between two domains that enables a user with an account in one domain to have access to resources, such as files and directories, in another domain.

NeTservices provides full support for the Domain security model.

NeTservices provides full support for the Windows Domain security model including support for PDC, BDC, and Member Server, mode of operation. NeTservices support of the Domain security model also includes the capability to respond to validation requests from users/groups in trusted domains and support of authentication of local and global groups.

Windows NT Security

Windows uses a set of standard Access Control Lists (ACLs) for granting access to shares, directories, and files. The ACLs offer useful combinations of specific types of access, which are called individual permissions. Individual permissions are somewhat analogous to UNIX permissions. They consist of read (R), Write (W), Execute (X), Delete (D), Change Permissions (P), and Take Ownership (O). UNIX supports three sets of file and directory permissions: owner, group, and world. This is the familiar –rwx-rwx-rwx (read-write-execute) that shows up in the output from the 'ls' command.

NeTservices includes full support for both share-level and file/directory-level permissions.

With Windows Servers, permissions can be granted to either individual users or to groups. The major difference between Windows and UNIX is that in Windows each user or group can be granted its own set of permissions for each file and/or directory. This allows a finer degree of access control and therefore greater flexibility. In UNIX assignment of access control is effectively limited to three entities, the owner of the file/directory, the primary group, and the rest of the world.

NeTservices includes full support for both share-level and file/directory-level permissions. Authorized administrators using Windows Explorer, Server Manager or File Manager GUI tools can accomplish management of such permissions.

NeTservices Administration

NeTservices supports "best-of-breed" administration environment for Windows services and data. Administrators can autonomously administer all Windows services, such as file sharing, user account, and file security, running on NS3000 systems. In addition, enterprise-

class solutions running on Auspex servers that deliver high-performance and robust storage management for UNIX data can be leveraged for Windows data.

Windows administration is based on a client/server model utilizing Remote Procedure Call (RPC) technology. This is a decentralized model, which dramatically simplifies many of the tasks usually associated with system administration. The Windows administrative environment, which is fully integrated with its Windows-based, graphical user interface (GUI), includes the following tools:

- ***User Manager*** for Domains provides the same function as the UNIX method of manually editing the /etc/passwd and /etc/groups files such as adding, modifying, renaming users/groups, and managing security policies.
- With auditing enabled, ***Event Viewer*** can be used to monitor system events such as when a particular user last logged on to the domain.
- ***Server Manager*** is the Windows GUI tool for monitoring and managing server properties, such as who is connected to a server, how long they have been connected, and what resources they have open. It can also be used for management functions such as closing open resources, and disconnecting users connected to a share.

NeTservices provides full support for Windows tools for remote administration, including ***User Manager, Event Viewer,*** and ***Server Manager.*** This support is enabled by the inclusion of a full implementation of Windows RPC in NeTservices. Finally, centralized UNIX-based data management, for tasks such as managing disk storage, RAID, and backup, is fully supported for Windows data.

NeTservices provides full support for Windows tools for remote administration.

File Sharing Using NeTservices

Support of transparent sharing of individual files among UNIX and Windows users is a key goal behind the delivery of the NeTservices product. Each file system on a NetServer running NeTservices supports file systems that can be used by NFS, CIFS or both without special configurations. A number of value-added facilities in the locking and file-sharing areas are provided that enable secure and robust high-performance sharing of files among UNIX and Windows users.

File Locking

NeTservices includes support for PC-style mandatory file/record locking that is fully compatible with CIFS file/record locking. The goal behind such mandatory locking functionality is to allow robust access by multiple authorized Windows clients to the same file or record. CIFS locking is the locking mechanism used by Windows clients when accessing file systems. NFS Lock Manager (lockd) will continue to be supported as the advisory locking mechanism used by clients when accessing file systems.

NeTservices provides robust access by multiple authorized Windows clients to the same file or record.

Coordinated Locking allows "safe" access

The need for sharing individual files by UNIX and Windows users is primarily driven by the emergence of UNIX and Windows versions of specific applications supporting a common, interchangeable file format. File systems that can be shared across UNIX and Windows clients enable file sharing in a robust fashion. NeTservices has been designed to support two alternative means for coordinated, "safe access"[6] by UNIX and Windows clients to individual files.

[6] File data is fully protected against coruption.

1. NeTservices provides a facility that allows CIFS locking to be enforced with respect to NFS access so that Windows clients are fully protected against concurrent NFS accesses to the same file.

NeTservices is fully compatible with high-level application-based locking.

2. NeTservices is fully compatible with high-level application-based locking. This refers to applications using Relational Database Management Systems (RDBMS) or shadow files to support locking information. Examples include MCAD applications such as ProEngineer, EDS/Unigraphics and Catia that use Product Data Managers (PDMs), and Frame's use of shadow files to manage locked file access across UNIX and Windows Clients. NeTservices is designed to allow interoperation with these and other applications.

NeTservices also supports server-based coordination between NFS and CIFS locking to allow coordinated access across UNIX and Windows clients. This is to support safe, coordinated access to data by multi-platform applications using platform-specific locking, such as NFS Lock Manager and CIFS file/record locking (versus applications using built-in locking functionality to provide coordinated access as in alternative 2 above).

File Access Control

UNIX file permissions automatically work with NTFS permissions.

Automatic coordination between the UNIX and Windows file access control mechanisms is critical if file sharing across UNIX and Windows clients is to be supported in a flexible and secure fashion.

NeTservices provides mechanisms that allow UNIX file permissions to automatically work with NTFS permissions to allow secure file sharing among UNIX and Windows clients with a minimum of administrative overhead. Specifically, Windows ACLs (which are set by using NT Server Manager, File Manager, or Explorer tools) function as the first level of file security for access by Windows users to files. UNIX file/directory permissions are used as the second level of file security for file accesses by Windows users. This is accomplished through a mechanism that allows mapping of Windows domain-based user/group accounts to UNIX user accounts. (Note that logon validation for Windows users/groups continues to be provided by Windows PDC/BDC servers). UNIX permissions continue to be the means for managing file security for UNIX users.

NS3000 Storage Capacities and RAID Architecture

6

Storage Capacity of the NS3000 Series

Up to 68 TB of storage capacity.

Each of the three I/O Node IIs of the NS3000 can manage up to six shelves of four drawers, depending on configuration. Each drawer contains seven drives or 6x4x7=168 drives per I/O Node II. In the case of 36GB or 73GB drives[7], this equals 6TB or 12 TB per drawer or 68 TB for a fully configured system[8]. **Table 4** shows system capacities and number of disk drives for various NS3000 Series system configurations.

Table 4 - NS3000 system capacities and number of disk drives for various configurations.

System Capacities and Number of disk drives for various configurations.		
System Configuration	Capacity with 36 GB drives	Capacity with 72 GB drives
One Node NS3010	1-6 TB – 168 disks	2-12TB – 168 disks
One Node NS3010ER	1-6 TB – 168 disks	2-12TB – 168 disks
Two Node NS3020ER	6-12TB – 336 disks	12-24TB – 336 disks
Two Node NS3020HA	1-4TB – 112 disks	4-12TB – 336 disks
Three Node NS3030ER	12-18TB – 504 disks	24-36TB – 504 disks
Four Node NS3040HA	12-24TB – 672 disks	24-48TB – 672 disks
Six Node NS3060HA	24-36TB – 1008 disks	36-72TB – 1008 disks

Advantages of the NS3000 RAID Hardware

Hardware RAID is a significant advantage compared to software RAID.

Auspex best-of-breed intelligent PCI RAID controllers to interface the High Density Disk Arrays (HDDAs). These intelligent controllers provide disk interface and RAID management, offloading these tasks from the FSP CPU. This is a significant advantage compared to software based RAID subsystems that use the central CPU to manage RAID. Each NS3000 controller supports RAID 0 (striping), RAID 1 (mirroring), and RAID 5 (parity RAID). All disks in the system are defined as part of a RAID array of one type or another. Non-volatile RAM on each controller accelerates RAID functions, particularly disk writes.

RAID 0 – Striping

RAID 0 arrays consist of from 2 to 8 disks. RAID 0 accelerates disk access by striping data across the disk array, thereby spreading I/O evenly across multiple disk spindles. RAID 0 provides the best overall performance but provides no resilience to disk failures.

[7] 73GB drives will be supported in late 2001.

[8] 1-4 TB of the maximum capacity of large scale configuration are required for server failover and external RAID system information reducing the maximum capacity of a fully configured system from 72 TB to 68 TB for user information.

RAID 1 – Mirroring

RAID 1 arrays consist of pairs of disks in which each disk is maintained as an exact replica or mirror of the other. This method provides for data redundancy and resilience. It provides two paths to the data and does not incur the write penalty. It is also the most expensive RAID alternative since it requires disk space equivalent to twice the usable capacity.

RAID 5 – Distributed Parity

RAID 5 on the NS3000 Series consist of either four-member (3d + 1p) or eight-member (7d + 1p) arrays. Eight-member arrays protect against any drive failure while four-member arrays protect against any drive *or* channel failue. Data is striped across RAID 5 arrays in a fashion similar to RAID 0, but RAID 5 provides fault resilience by creating and maintaining parity information on each stripe of data. If a failure occurs, the contents of that block can be recreated by reading back the other blocks in the stripe along with the parity. Parity information is distributed throughout the array to minimize potential bottlenecks. The storage overhead of RAID 5 is equivalent to one disk drive regardless of the size of the array.

RAID Rebuild Capability

No CPU cycles are used to rebuild the array in the event of a drive failure.

In the event of a disk failure, RAID 1 or 5 arrays can be rapidly and automatically rebuilt using available "hot-spare" drives. The caching policy may also be specified for each array as either write-back or write-through. With write-back caching, once data is committed to Non-Volatile RAM (NVRAM), it is considered complete and written to disk at a later time. With the NS3000 "write-through" feature enabled, data is written to disk while a copy is preserved in controller NVRAM. The RAID controllers allow RAID arrays to be expanded online. Configurations for RAID 0, RAID 1, and RAID 5 are shown in **Figure 16**. In that the NS3000 uses hardware RAID controllers, no CPU cycles are used to rebuild the array in the event of a drive failure. A drive failure in competitive software based RAID systems will impact throughput during rebuild. This is not the case with the Auspex NS3000 design.

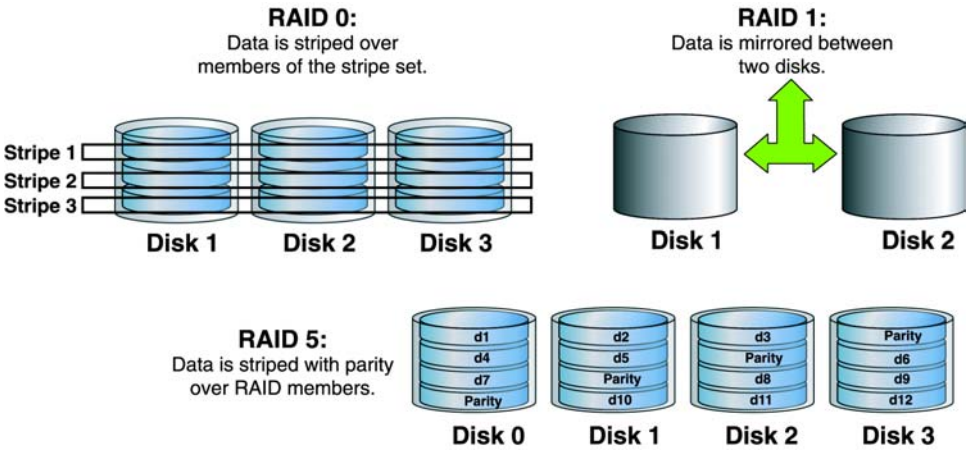

Figure 16 - RAID 0, RAID 1, and RAID 5 arrays.

High-density storage arrays

The Auspex RAID controllers connect directly to the High-density Disk Array (HDDA) with system capacities as shown in Table 4 above. The HDDA provides a very large amount of storage in a package that has the same form factor as the I/O Node II itself. All drives in the array are hot pluggable to ensure rapid, online replacement in the event of a problem.

Extended Virtual Partitions

The RAID arrays of the NS3000 are supported by the Auspex RAID controllers in each I/O Node II. The ability to define larger virtual partitions made up of multiple RAID arrays is known as the Extended Virtual Partition Feature or EVPs. EVPs can span multiple RAID controllers within an I/O Node II to create very large file systems. All EVPs take full advantage of the underlying redundancy of the individual RAID arrays that make up the EVP. Three types of EVPs are supported as shown in **Figure 17**.

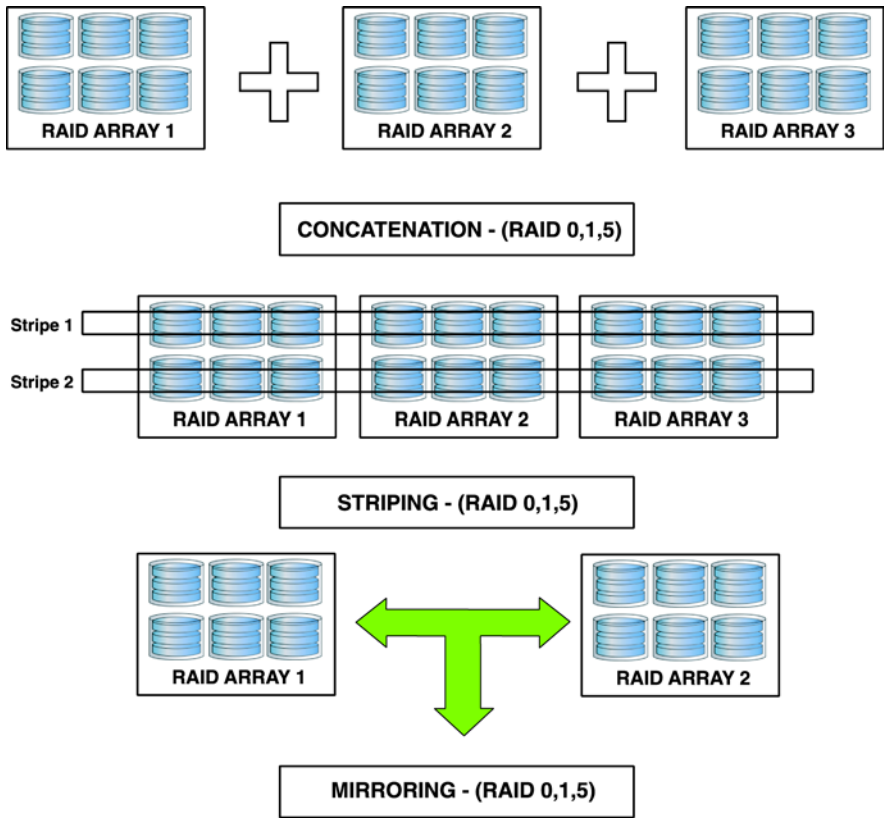

Figure 17 -Three types of Extended Virtual Partitions (EVPs) - Concatenation, Striping and Mirroring.

Extreme Virtual Partitions are accomplished through Software RAID 0,1 and 5 by the NS3000 NetOS operation system on each I/O Node II. EVPs allow mirroring of file systems across RAID controllers so that if one controller fails, the file systems remain available through the surviving controller.

Extreme Virtual Partitions provide the ultimate in protection against storage subsystem failures.

Extreme Virtual Partitions

Distinct from Extended Virtual Partitions, Extreme Virtual Partitions are used to protect against storage subsystem failures in Fibre Channel connected NS3000ER and NS3000HA configurations. Storage subsystems are protected against dual controller failures, multiplechannel failures and power failures. Fibre Channel connectivity assures access by separate I/O Node IIs in normal operation or fail-over operations when file systems are mirrored across dual redundant RAID pairs. This feature provides the ultimate in protection against storage subsystem failures. When combined with the fault tolerant server features of ServerGuard V (Chapter 7), Extreme Virtual Partitions allow the NS3000HA Series to have **no single point of failure** and protect fully against either server or array outages.

Concatenations

Concatenations join multiple RAID arrays (RAID 0, 1, or 5) together to create larger storage volumes.

Stripes

Stripes are similar to concatenations, but with stripes, the data is spread more evenly across the disks of the multiple RAID arrays, which can be RAID 0, 1, or 5.

Mirrors

Mirrors allow a duplicate copy of all data on one RAID to be duplicated on a second RAID array. When two RAID 0 arrays are mirrored it is known as RAID 0+1. This configuration of striping with mirroring creates outstanding performance and reliability.

High Availability and High Performance Considerations

7

Information is always available

Downtime means lost productivity for employees.

To achieve high availability computing, it important to avoid not only *unplanned* interruptions to file and network availability, but also *planned* interruptions such as scheduled downtime for servicing and backup. Whether planned or unplanned, downtime means lost productivity for employees. Many network file serving and file sharing applications require files to be "always available" day and night all the time. Whether engineers are designing a new product using network file sharing or a manufacturing team critical to effective business operations in all departments. For many companies if files are not available, employees cease to be productive. The cost of downtime can be very high!

From the network to the disk drive, the Auspex file servers offers a broad range of availability solutions from top to bottom of the availability pyramid to provide the highest level of file availability to customers, whether its minimizing *unplanned* downtime or reducing *planned* downtime for backup. This is illustrated in **Figure 18**.

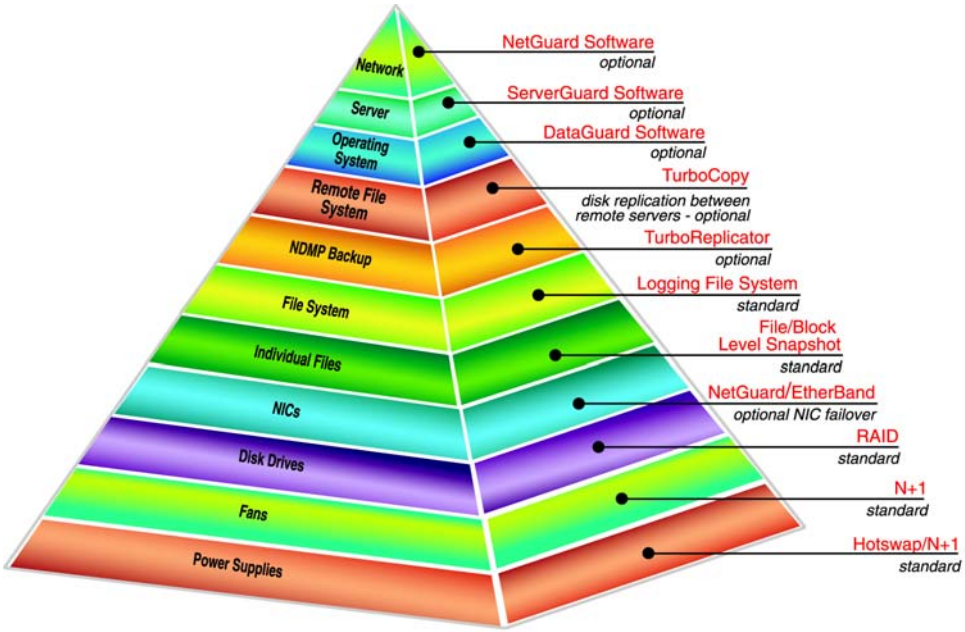

Figure 18 - From the disk to the network, the NS3000 offers a full range of availability options.

Superior availability compared to general-purpose servers

The availability features of the NS3000 results in much less unplanned downtime per year than for general-purpose computers that are used for network file serving. This difference is shown in **Table 5** below.

Table 5 - Availability of the NS3000 compared to general purpose computers used as file servers.

Measurement	General Purpose Computers used as File Servers	Auspex NS3000
Annual Availability of Data	99.86%	99.99+%
Unplanned downtime per year	12 hours = 720 minutes	<30 minutes

This dramatic difference relates to the fact that one of the major causes of system outages is the UNIX or Windows is the operating system. This is opposed to the specifically designed NS3000 that insulates the function of file service from failures in a general-purpose operating system. The NS3000 NetOS operating system (known as a "real time" operating system kernel) involves approximately 10MB of tightly integrated, and thoroughly tested, code that is designed to manage only those functions necessary for network file service. General-purpose file servers use the UNIX or NT operating system. These are known as "fat O/S" and can often exceed 3 GB of code. These large operating system software structures fail more often due to their complexity and the impossibility to test for all possible failure modes. The NS3000 kernel is approximately 10MB and has been more thoroughly tested for all possible failure modes. This results in significant advantages for the Auspex NS3000 in terms of performance compared to general-purpose file servers. This advantage is shown in **Table 5** above.

Standard hot swapped and N+1 power supplies

The NS3000 provides full system power in the event of a power supply failure.

As discussed in Chapter 2, the NS3000 provides an N+1 design for power supplies and fans to insure uninterrupted power and cooling to all storage chassis in the system. Each power distribution Unit (PDU) is independently redundant based on its N+1 power supply design. An N+1 power shelf design provides for one redundant power supply on the shelf to provide full system power in the event of a power supply failure. In the event of such a failure, the power supply that has failed can be "hot swapped" from the shelf without disruption to the power subsystem.

Standard N+1 fans for cooling

Each chassis in the NS3000 Series of products is equipped with standard N+1 cooling fans. The speed and operation of these fans are monitored by the EM-Net as discussed in Chapter 2. The fans utilize an N+1 design, which provides for one redundant fan in the event of a fan failure. This guarantees that a fan can fail without causing a cooling problem for the chassis. A failed cooling fan can therefore be quickly replaced with the remaining fans being sufficient to operate the NS3000 within factory cooling specifications.

Unmatched competitive advantage in RAID choices

The NS3000 offers the unique capability to intermix RAID levels on a file system by file system basis.

The flexibility in choice for NS3000 RAID protection for disk drives is unmatched. It is important to have choices in RAID protection since workloads have different read/write characteristics which influence performance throughput. RAID 1 (mirroring) offers the highest performance (read from either disk in the mirror pair) during normal operation. RAID 1 also offers the highest performance in the event of a failed disk since it is not necessary to read parity (and then data) in the event of a disk failure in other RAID configurations.

The NS3000 offers RAID 0, RAID 1, RAID 5 and RAID 10 (mixed RAID and RAID 1), with the unique capability to intermix RAID levels by file system within a controller.

EtherBand™ provides both high availability and consistent performance

Load balancing is the even distribution of file server traffic across a computer network.

EtherBand™ is an Auspex solution that provides load balancing for outbound network traffic and network interface failover. All of the interfaces in the EtherBand™ balance the load of outbound traffic among themselves, thereby improving network performance while also providing network fault tolerance in the case of network failure. Working in conjunction with EtherBand™ is an industry unique "throttling" feature in Auspex file servers that can reduce the performance impact of a remote copy to disk or tape operation.

Redundant components provide high availability and fault tolerance in computing. Network failover in computer systems is the ability for one network connection to seamlessly take over the workload of a failed connection of a similar type. The Auspex EtherBand™ design allows for multiple redundant network connections to provide protection in the event of a failed network adapter card or network link. If a channel fails, traffic is automatically routed to another channel, thereby providing network failover protection.

File server load balancing is the even distribution of file server traffic across a computer network so that no single device or network connection becomes bottlenecked. Load balancing is especially important for networks where it is difficult to predict the number of packets that will be issued over any one channel. If a particular network attachment experiences a disproportionate amount of traffic, load-balancing algorithms forward the traffic to another network channel. All of the interfaces in the EtherBand™ trunk balance the load of outbound traffic among themselves. This improves network performance and provides network fault tolerance in the case of network failure.

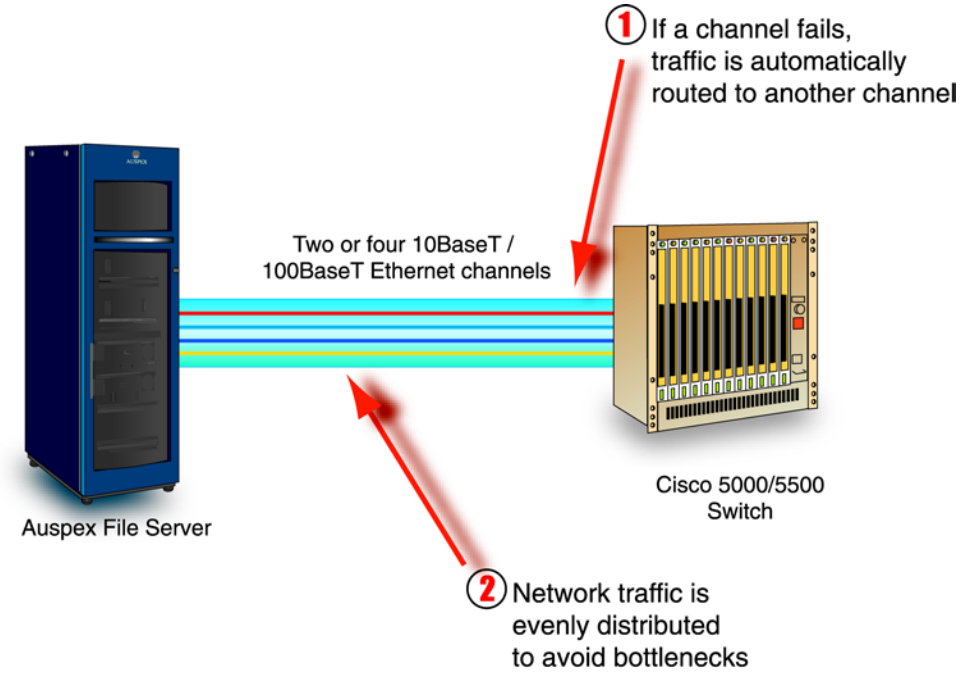

Figure 19 - EtherBand™ supports both load balancing and NIC failover.

EtherBand™ balances only outbound traffic from the Auspex file server. A network switch or customized load balancer controls and configures inbound traffic to the Auspex file server. EtherBand™ load balancing is based on a low level software algorithm that is activated when a packet is ready to go out on the network. Based on traffic conditions of the moment, EtherBand™ makes a decision at the driver level about which EtherBand™ port to use. With network interface failover, the network traffic is rerouted over the remaining active ports if any interface in the EtherBand™ fails. EtherBand™ failover is automatic and protects against cable, network or port failure.

EtherBand™ is based on the integration of Cisco Fast EtherChannel into the Auspex file serving and file sharing architecture. Without the EtherBand‘ feature, four separate IP and MAC addresses are required with no load balancing or failover capability in the event of an outage. With the EtherBand™ feature, there is only one IP and MAC address and load balancing and NIC failover are enabled. Up to 4 EtherChannels per I/O node are supported. This is also known as Ethernet port aggregation. Physically, the EtherBand™ trunk consists of two or four 10BaseT or 100BaseT Ethernet interfaces. Logically, the EtherBand™ trunk appears as one network connection to the Auspex file server.

Without the EtherBand™ or NetGuard™ featurs, four separate IP and MAC addresses are needed to balance the load and provide failover capability in the event of a NIC outage. **Figure 19** illustrates that with the EtherBand™ feature, only one IP and MAC address is required for load balancing and NIC failover. Up to 3 pipes with 4 ports each per I/O Node II per NS3000 are supported.

NS3000 logical volume level snapshot capability is standard

Snapshots are typically used to provide a consistent unchanging view of the file system for backup. Regularly performed, snapshots can also be used to provide a rudimentary form of file versioning, and also allow for rapid easy recovery of data that is accidentally overwritten or deleted due to user error. In this sense, snapshots are an important data availability characteristic of storage systems.

Device level snapshots occur at the level of the physical disk volume to create a snapshot.

Device level snapshots occur at the level of the physical disk volume to create a snapshot, all cached data is first flushed to disk and the volume is made momentarily quiescent to ensure that it is stable. Once the snapshot occurs, any time a data block is changed, a copy of the original is made and saved in a separate designated snapshot partition. These snapshots can be exported and mounted by network clients, thereby providing a view of the file system exactly as it was at the time the snapshot was taken. Snapshots can also be backed up using a variety of methods, which will be discussed shortly. Up to 16 simultaneous snapshots are supported for any given volume.

Figure 20 illustrates that after a snapshot has been taken, and as new data blocks are written to the source volume, the original data blocks are copied and saved to the snapshot volume. To maintain a static view of the snapshot data, the snapshot view accesses unmodified data blocks from the source volume and any modified data blocks from the snapshot volume. Two maps (bitmap and remap) are maintained in FSP memory and on the snapshot volume to provide the necessary bookkeeping information.

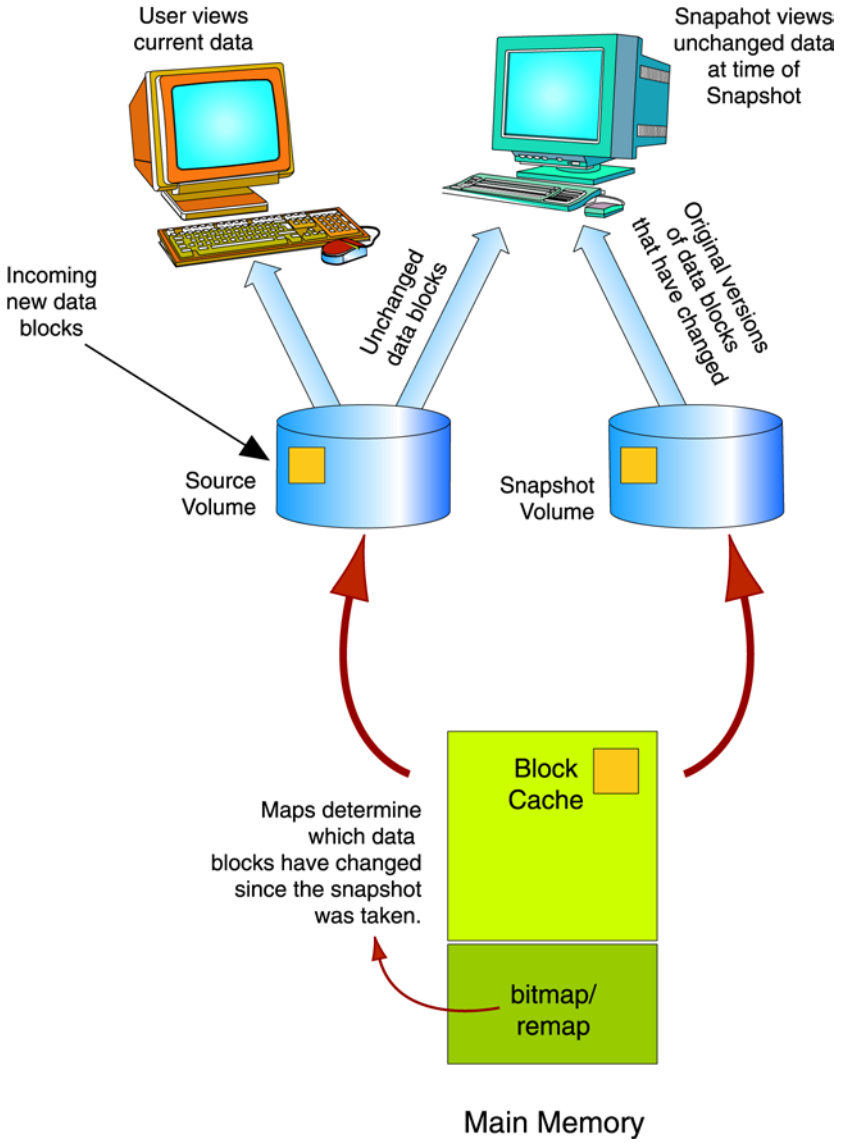

Figure 20 – Maintaining an unchanged snapshot view in the device level snapshot process of the NS3000.

Logging file system is standard

As discussed in Chapter 4 the NS3000 FastFLO™ file system is a logging file system, which is also very important to overall system availability. A logging file system keeps track of all changes that are made to data from a particular point in time and is therefore important to consistent high data availability. To restore a file system in the event of a system failure, the FastFLO™ file system restores only the data that has been changed instead of requiring the entire system to be restored. This feature saves time in rebuild performance and improves overall system availability.

With Turbo Copy™, information is always protected from loss

TurboCopy™ is an Auspex solution that provides high availability and disaster tolerance by allowing disk-to-disk Point in Time (PiT) remote copies of entire file systems, individual directories or files. TurboCopy™ accomplishes this by enabling high-speed transfer of files between two Auspex file servers through direct communication rather than through intermediate systems. Working in conjunction with TurboReplicator™, backup copies can be remotely made

after the files, directories or file systems have been copied to the off-site disk. TurboCopy™ meets a growing demand for online, off-site file copies that provides quicker recovery of recent data than tape, and affords the safety of a remote location in the event of a local disaster. This solution is ideal for enterprises with two Auspex file servers in remote geographies since the sites can reciprocally back each other up.

Disaster tolerance reduces the time to recover computing operations.

Disaster tolerance greatly reduces the time to recover computing operations following a disaster such as earthquakes, fires or floods. To achieve high availability computing, it important to avoid not only *unplanned* interruptions to file and network availability, but also *planned* interruptions such as scheduled downtime for servicing and backup. Whether planned or unplanned, downtime means lack of file availability to your employees. Many network file serving and file sharing applications require files to be "always available" day and night all the time. Since the cost of downtime can be very high, TurboCopy™ not only reduces downtime for remote copy but also provides protection against data loss following a disaster.

Historical remote disk-to-disk replication with UNIX or NT has been prohibitively slow because files were forced through an intermediate client or through host processors. TurboCopy™ solves this problem by offering a fast, efficient disk-to-disk replication by enabling high-speed transfer of data between two Auspex file servers through direct network processor-to-network processor communication rather than through the intermediate host processors. To improve remote copy performance and reduce performance impact to the primary system, TurboCopy™ embeds file metadata (data about data) and actual file data in the same packet stream. This process is known in the industry as Point in Time (PiT) copy.

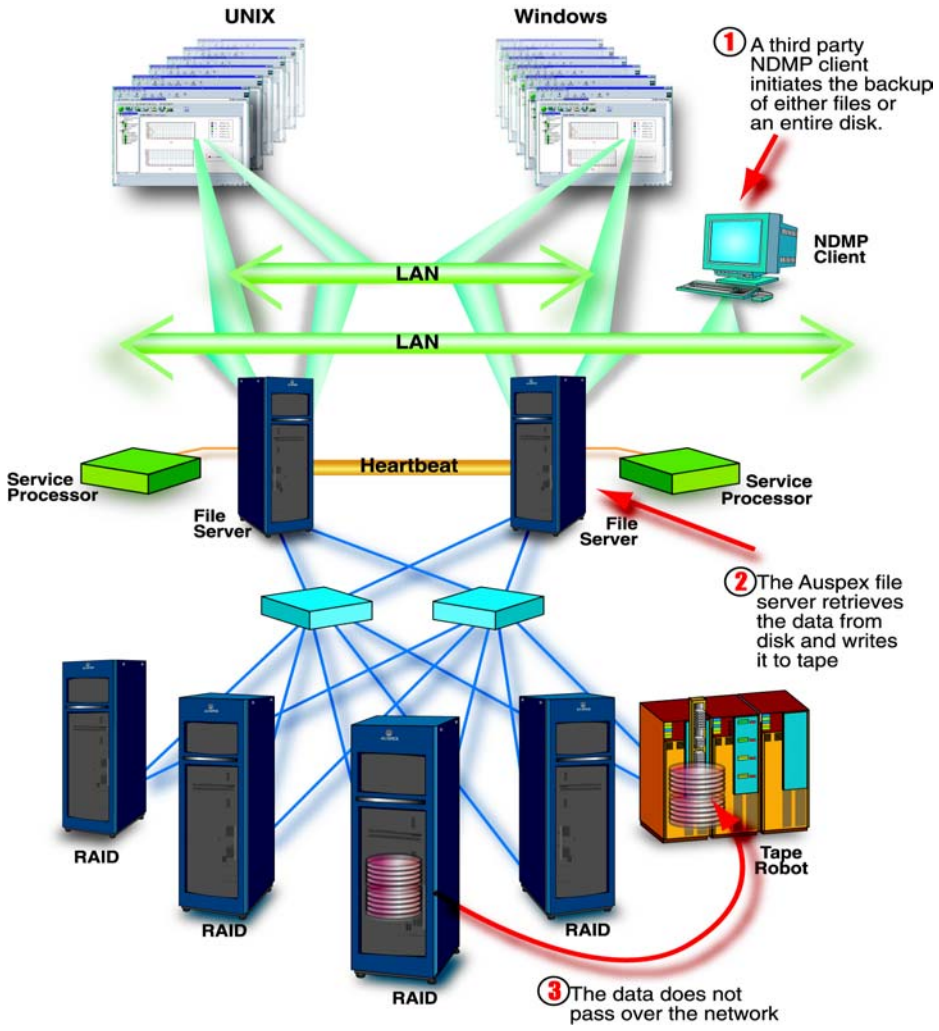

Figure 21 - TurboCopy software provides for disaster tolerance by remote file system replication.

As illustrated in **Figure 21,** TurboCopy™ remote data transfer occurs directly between two Auspex file servers, one acting as a data server and the other as a disk server. With TurboCopy™, the role of two Auspex file servers is interchangeable to enable utilization of the two file servers to copy files back and forth between the two. For system-to-system remote copy, TurboCopy™ must be installed on both the local and off-site Auspex file servers that serve as NDMP hosts. With file-based backup, the backup is on a file-by-file basis, so even when an entire file system is being backed up, individual files and directories are preserved. Additionally, TurboCopy™ supports **incremental remote copy** of only file data that has changed to reduce the amount of file data being transferred on the network. Since file server and network activity usually varies depending on time of day and week, Auspex provides an industry unique "task-throttling" feature among file server vendors to avoid consuming too much network bandwidth by scheduling replicity at non-peak network usage times.

DataGuard™ provides higher availability

DataGuard™ is an Auspex solution that provides high availability characteristics for the Auspex file serving architecture. DataGuard™ allows the Service Processor to reboot automatically without affecting UNIX file serving operations in the event of an operating system failure. During the reboot, the Auspex file server continues to deliver UNIX data and file services to users. However, NT file services and NT data sharing are temporarily interrupted until the Service Processor reboots. This industry unique feature greatly reduces downtime relative to competitors who do not implement this feature.

The Service Processor serves as the control center of the Auspex file server and manages intelligent file sharing between Auspex file server modules (know as I/O Nodes) which operate in parallel. The Auspex file serving and file sharing architecture is unique because the function of UNIX file serving is isolated from the Service Processor. Since all of the functions of the Auspex file server can be managed either locally from the Service Processor console or remotely from any compatible browser, monitoring of the Auspex file server can continue in the event of a Service Processor failure.

As shown in **Figure 22** DataGuard™ builds on the unique Auspex parallel processing architecture by protecting UNIX file operations from the effects of a Service Processor or application failure. Since UNIX file serving runs on the I/O Nodes, UNIX file services continue uninterrupted in the event of a Service Processor failure. Also since UNIX system administration applications run on either the Auspex Service Processor or on remote network clients, UNIX system administration can continue in the event of a Service Processor failure.

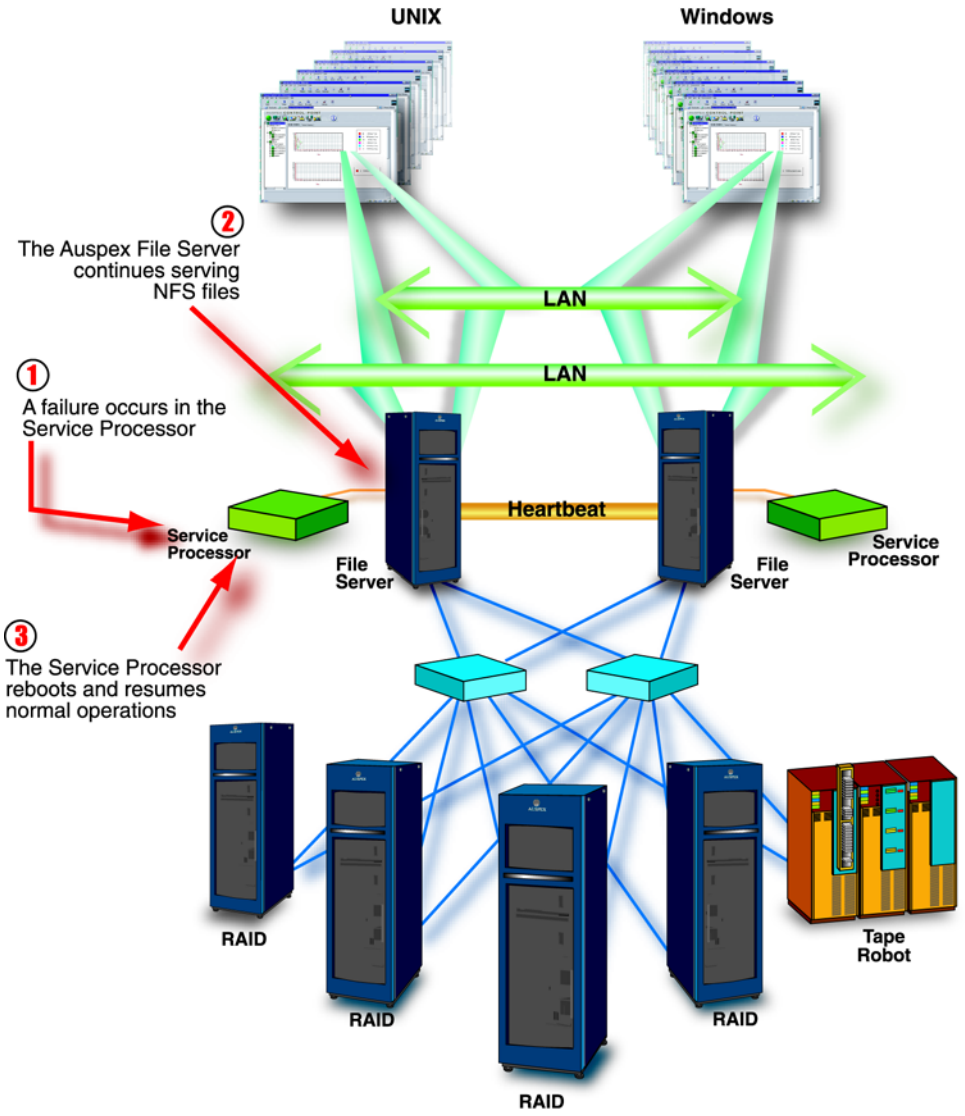

Figure 22 - Data Guard provides a firewall between the NS3000 I/O Node IIs and the Service Processor.

With DataGuard™, a majority of operating system failures are isolated from the host and services provided by the host. Service Processor failure isolation makes the Auspex file server more reliable than competitive products and independent of the Solaris operating system running on the host. This failure isolation improves reliability and maintainability.

ServerGuard V™ provides file server fault tolerance and very high data availability

ServerGuard V™ is a unique unified NAS and SAN software solution that provides file server high availability and fault tolerance against a file server or Fibre Channel failure for the NS3000HA Series of products. When combined with the features of Extreme Virtual Partitions (Chapter 6), ServerGuard V™ allows the NS3000HA Series to have **no single point of failure**. This provides full protection against either server or array outages ServerGuard V™ also provides Storage Area Network (SAN) high availability and fault tolerance in the event of a server or SAN element failure in the Fibre Channel Storage Area Network (SAN) architecture.

The Auspex ServerGuard V™ software solution provides both server and SAN failover protection. This feature provides redundancy and fault tolerance for the most important

elements in a file serving and file sharing architecture. File server fault tolerance is accomplished by the connection of a heartbeat between pairs of servers and so that one file server can determine if the other server in the pair has experienced a failure.

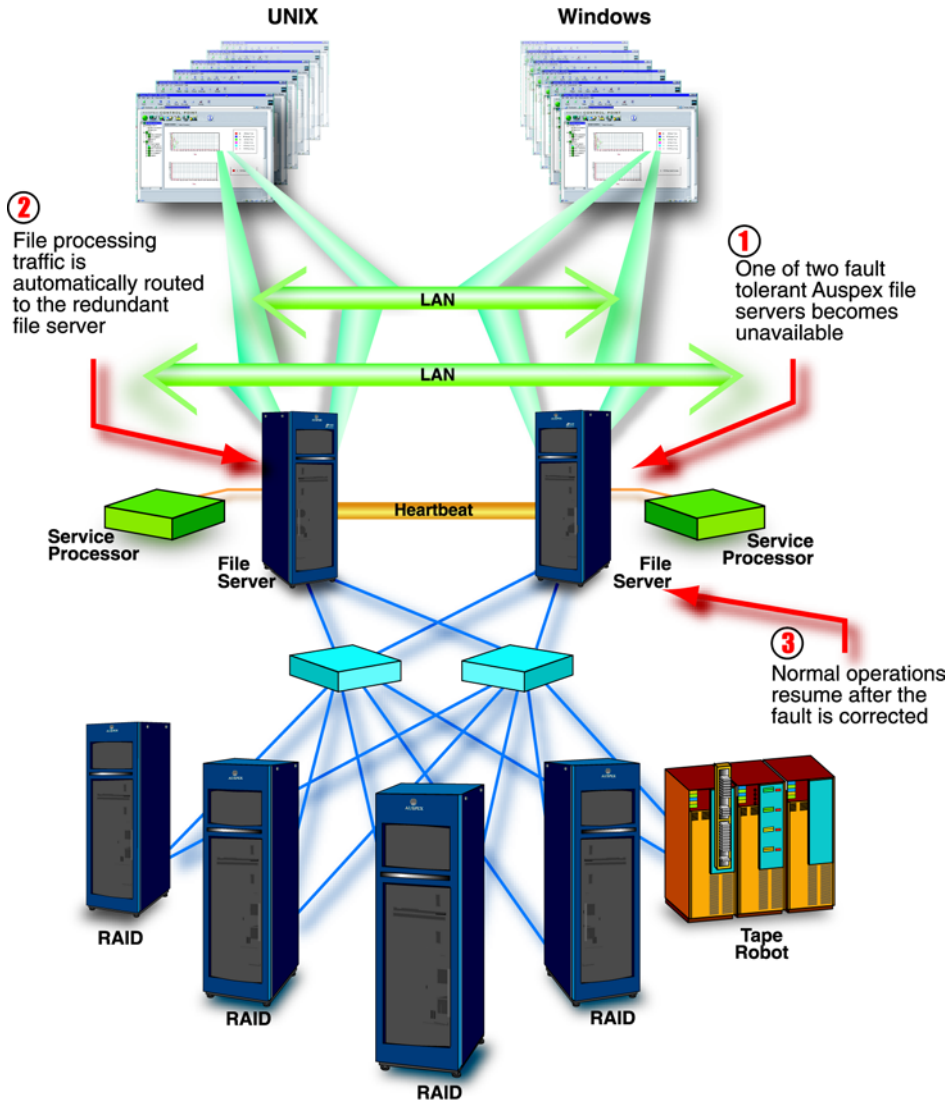

Figure 23 - ServerGuard V™ provides automatic failover and fail-back between servers in the event of a server failure.

As shown in **Figure 23**, ServerGuard V™ provides automatic failover and fail-back between servers in the event of a server failure. The affected work that was being done on the failed server is automatically switched over to the redundant server which will process the workload until the repair can be made and normal operations resume. This is accomplished through a software facility know as Virtual Server. Virtual Servers are logical identities that share physical resources but own logical resources running on two or more physical servers. When a physical server fails, the Virtual Servers are moved from the failed system to the surviving system. Network connections are restored with virtual IP addresses using a gratuitous ARP broadcast technique from the surviving server and either clients or a network device (router or switch) responds to the gratuitous ARP. The File Systems are then re-mounted on the surviving server with journal enabled fast fsck.

Virtual Servers are logical identities that share physical resources but own logical resources.

Fibre Channel fault tolerance is accomplished by redundant connections and failover software in the event of a Fibre Channel failure that connects storage arrays to the file server. In the event one Fibre Channel element or network connection in the SAN becomes

The Auspex NS3000HA Series is doubly fault tolerant and unlikely ever to fail.

unavailable, traffic on the network is automatically routed to a redundant path. In addition each SAN element is redundant within itself so the Auspex NS3000HA Series SAN is doubly fault tolerant and unlikely ever to fail.

NetGuard™ provides network fault tolerance

NetGuard™ is an Auspex software solution that that provides high availability characteristics and network fault tolerance for either single node or multi-node Auspex file servers. It provides automatic failover and fail-back between network server connections for network connection failures or lack of network availability. This high availability solution is transparent to UNIX and Windows users. NetGuard™ achieves network fault tolerance by providing network interface failover for both single node or multi-node Auspex file servers. This high availability solution is transparent to UNIX (NFS) and Windows NT (CIFS) file sharing protocols.

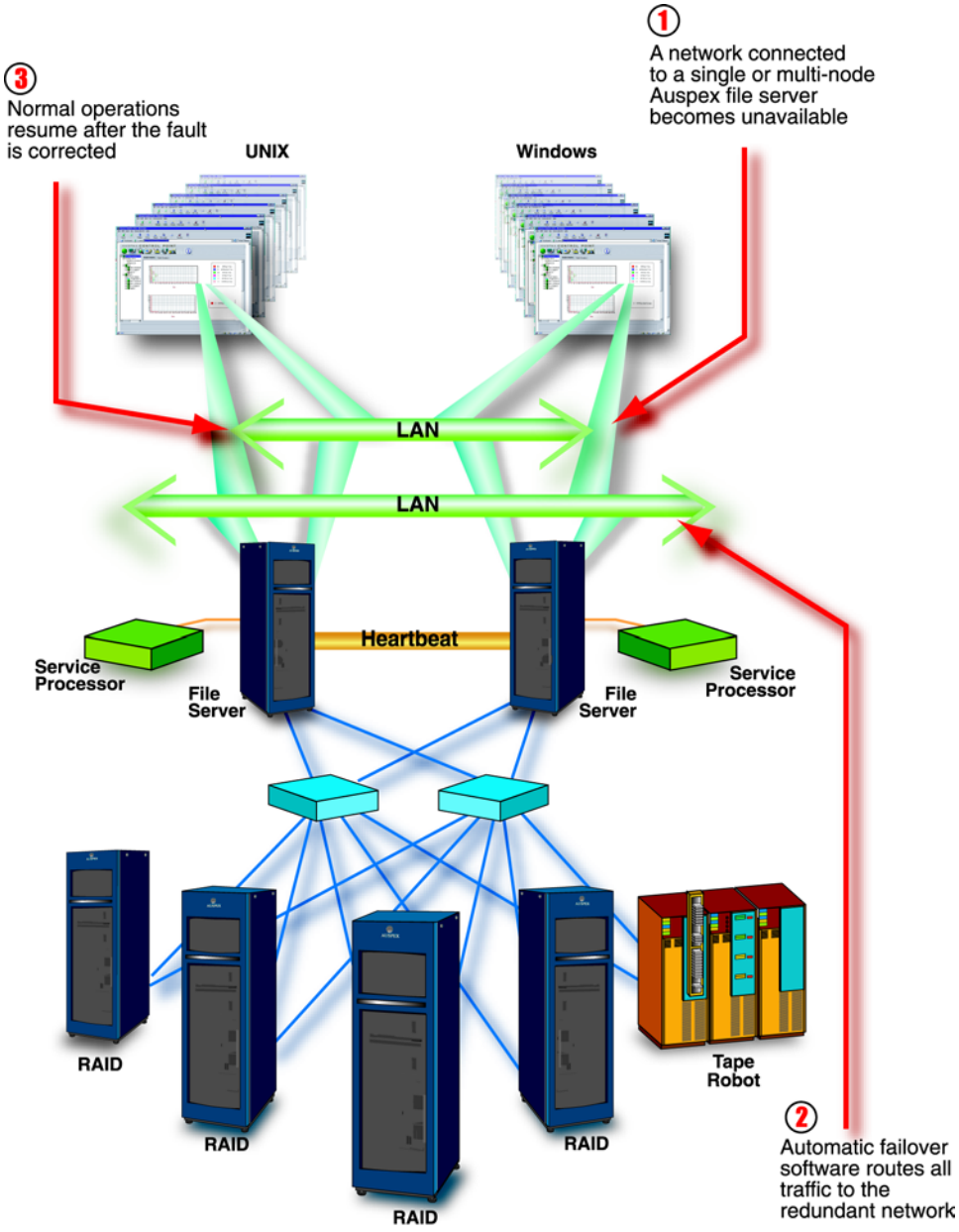

Figure 24 - NetGuard™ routes all traffic to the redundant network.

Distinct from the network failover provided by EtherBand™, network interface failover in computer systems is the ability for one network interface connection to seamlessly take over the workload of a failed connection of a similar type. The Auspex NetGuard™ provides port failover protection in the event of a failed interface adapter port. This is illustrated in **Figure 24**. If a port on a network interface card fails, traffic is automatically routed to another port on either the same or another network interface card thereby providing network interface failover protection.

NetGuard™ is based on the Virtual Internet Protocol address (VIP) or Internet Protocol (IP) aliasing that assigns an alternate logical IP address (the VIP) to a network connection. When a network failure occurs, traffic continues to access the VIP address, but the VIP transparently accesses another physical network connection. NetGuard™ determines the health of the network connection by pinging known good IP addresses. When configuring NetGuard™, the administrator must assign addresses to each VIP, then associate each VIP with a single Physical IP (PIP) address, called the home PIP. A home PIP is any PIP with which a selected VIP is associated under normal network operating conditions.

Remote tape backup and restore with TurboReplicator™

Backup and restore from any NDMP client.

TurboReplicator is a backup and restore software solution that works in harmony with the parallel backup solution of the Auspex file servers. Auspex offers not only industry unique fast parallel *local backup* to SCSI of Fibre Channel devices, but also the ability to *perform remote backup* to disk or tape in a geographically remote site. This feature provides fault tolerant protection against file loss. TurboReplicator™ allows backup and restore from any Network Data Management Protocol (NDMP) client. This allows Auspex file servers to easily fit into existing management practices.

Any NDMP client can initiate and manage backup and restore operations to a remote site. Auspex file servers support total storage capacities up to 68 terabytes. Traditionally, a backup or restore of such large amounts of data can significantly degrade system operations for hours. The Auspex TurboReplicator™ software solution is a workable, easy-to-use solution for large-system backup and minimizes downtime of the primary production system.

Faster backup is not possible with any other vendor.

To achieve high availability and disaster tolerance in computing, the ultimate protection from loss of files is multiple backup copies in remote geographies with no interruption to performance of the production system at the primary data center. Many network file serving and file sharing applications require files to be "always available" day and night all the time and in addition be protected from potential disasters such as earthquakes, fires, flood or hackers. Whether engineers are designing a new product using network file sharing or a manufacturing team is working three shifts to meet demand, availability of the latest information can be critical to effective business operations in all departments. For many companies if files are not available on line or cannot be recovered in the event of a disaster, employees cease to be productive. The cost of downtime and or lost data can be very high!

TurboReplicator™ supports high availability computing through fast and efficient local parallel tape backup and restore of large amounts of data thereby minimizing planned downtime for backup operations. Since Auspex is industry unique in its parallel backup design, faster backup is not possible with any other vendor.

TurboReplicator™ supports two types of backups: file-based backup and image-based backup. During file-based backup, individual files or directories are stored on tape and cataloged in a database. With this type of backup, individual files or directories can be backed up to tape from the management console user interface and restored without having to restore the entire file system. Image-based backup saves the block image of an entire file system to tape. However incremental backup is not possible in this mode. Image-based backup performs better than file-based backup when the majority of the files on the file system are small or the file system is heavily fragmented (i.e., data blocks belonging to the same file are scattered on a disk).

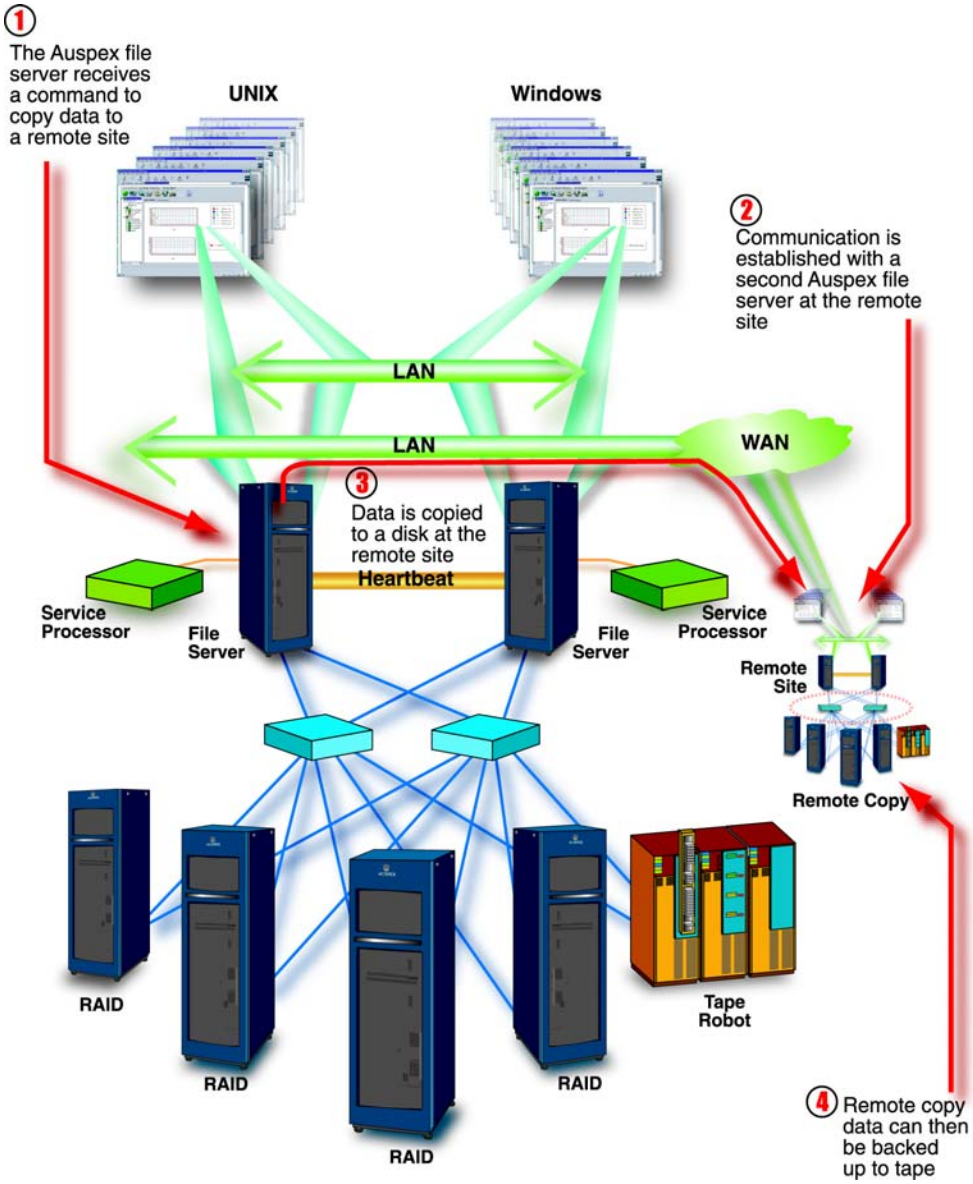

Figure 25 - A local and remote NS3000 Server remotely backup data to tape using TurboReplicator™.

As shown in **Figure 25,** the TurboReplicator™ remote backup of files is accomplished over a network between two Auspex file servers or between a Solaris workstation and an Auspex file server. In remote backup mode one Auspex file server or NDMP client workstation serves as a data server and another Auspex file server or Solaris workstation serves as a tape server. Three-way backup provides improved protection for networks by providing a central point for the backup of distributed Auspex file servers and Solaris networked systems.

NS3000 Backup and Restore Capabilities

8

Advantages of the NS3000 local parallel backup

Parallel backup provides major advantages in reducing backup windows.

Because of its modern parallel architecture, the Auspex NS3000 provides users with simplified scale-up of storage capacity, processors, network connections and performance. In addition, this design provides major advantages in reducing backup windows due to parallel backup of data on each I/O Node II. The importance of backup windows is illustrated by the fact that this was the single most important storage concern in a recent survey of 80 enterprises by ITcentrix.

Parallel local backup performance of the NS3000

Unsurpassed backup rates of 195 Gbytes/hr for BTE and 186 Gbytes/hr for FTE.

In evaluating backup performance in computer systems, the performance bottleneck is always the transfer rate of the tape drives themselves. Auspex has a major advantage over competitive designs such as Network Appliance and EMC in this respect since tape drives can be attached to each node. This allows backup to occur in parallel for an entire NS3000 system. With the NS3000, both block (BTE) and File (FTE) backup performance for a 9TB system can be accomplished at the rate of 195 Gbytes/hr for BTE and 186 Gbytes/hr for FTE with the maximum number of tape drives configured locally to each I/O Node II for parallel operation. Although backup speed depends on file size, compression, and the type of RAID applied to the file system, a 2:1 compression ratio is typical. Assuming this compression ratio a maximum configured NS3000 system can be backed up completely in 11.8 hours for RAID 1 and 20 hours for RAID 5 at the block level. Similarly, a complete system can be backed up completely in 12.4 hours for RAID 1 and 21 hours for RAID 5 at the file level. A more normal scenario is to do incremental file backup and only backup changed files. Assuming a 25% hit ratio for damaged data a maximum configured three I/O Node II NS3000 system can be incrementally backup up in 3.1 hours for RAID 1 and 5.2 hours for RAID 5. This is shown in **Table 6**.

Remote tape backups and restore with TurboReplicator™

Configuration for Backup	RAID 1 4.5TB useable	RAID 5 7.7TB useable
Backup BTE archive = 195 GB/hr	11.8 hrs	20.0 hrs
Level 0 (FTE) = 186 GB/hr	12.4 hrs	21.0 hrs
Incremental FTE assuming 25% hit ratio	3.1 hrs	5.2 hrs

Table 6 - Backup performance to tape drives directly attached to I/O Node IIs.

Third party UNIX-based backup tools

The Auspex NS3000 system supports a range of third party, UNIX-based backup tools, including Legato Networker™ and Verita Net Backup™. NeTservices allows Windows data to be backed up using these UNIX-based, enterprise-level backup products.

Device and file level acceleration

Auspex NS3000 system software provides two data moving engines and interfaces that are designed to dramatically accelerate system backup. NDMP is a standard protocol that can be implemented on any server or backup device.

The NS3000 uses snapshot along with the ***Block DataXceleration Engine (BDX)*** which provides the capability to stream blocks of data from disk to tape, and provides the foundation for extremely rapid image backup. NDMP can also use the ***File DataXceleration Engine (FDX)*** which provides a similar interface that acts at the file system level, and provides the basis for rapid file-by-file or directory backup. Data passes directly from disk to tape if the tape is attached to the same I/O Node II or from disk on one FSP across the SCI network to tape on another FSP if the tape is attached to another I/O Node II. These engines in conjunction with NDMP are robust data management tools, which enable high-performance backup and restore of file systems, directories, and individual files along with UNIX and Windows security information, including ACLs, SID, etc.

NDMP compliant products such as Veritas Net Backup™ and Legato Networker™, can be used for high-performance backups of data on Auspex servers. Auspex NS3000 system software also provides UNIX commands that are capable of backup and restore of UNIX and Windows data along with associated security information.

Device level and file level snapshots are supported

To create a device level snapshot all data is first flushed to disk and the volume is made momentarily quiescent to ensure it is stable. Once a snapshot occurs, any time a data block is changed a copy of the original is made and saved in a separate designated partition. Both types of snapshots preserve both UNIX and Windows data along with associated security information. Users can specify the interval at which these snapshot copies are made. A list of currently certified products is available from Auspex upon request.

Tape Robot support over Fiber Channel

An optional Fibre Channel interface card can be attached to the NS3000 I/O Node II to provide backup support to tape robots and other backup tape drives supporting Fiber Channel.

Network and System Management

9

All data processing managers realize the importance of configuration, maintenance, and monitoring of all the equipment under their control. Initial configuration, as well as any required re-configuration and maintenance of the NS3000 is accomplished through Auspex Control Point™ softwareas seen in **Figure 26**. This Java applet allows any Web browser, either on the server console or a remote network-attached client, to configure, maintain, and monitor single or multiple NS3000 servers once authenticated through root privleges. In addition to configuration and reconfiguration capacities, Control Point's Performance Monitor can display server performance in real-time as well as capture server performance metrics on an on-going basis, giving system managers the historical information they need to calibrate baseline server performance and plan for server growth.

Monitoring the NS3000 is also accomplished through Control Point's Event Manager. Critical events, such as high levels of network traffic on particular interfaces, which could be a sign of changing network requirements, can cause an alert to be set which will send an E-mail or other notification to appropriate system administrators.

The NS3000 also includes SNMP MIBs for integration with Enterprise Management Systems. This allows for monitoring of the NS3000 to be included with monitoring of other network devices on platforms such as HP OpenView, CA Unicenter, Tivoli TME and Microsft's Systems Management Server (SMS).

The Enterprise Management Platform vendors are signing cooperative agreements with systems vendors to provide greater levels of integration such as custom icons, expanded component views, and integration of traps and alerts with their more traditional "changed color" icon system views.

System Status

This selection provides a summary of NS3000 system status and system configuration options. Information includes server uptime, number of reboots, system date and time, number of active users, and information about system and network errors. In addition to this information, system configuration information is also available, detailing specific information about the system's inventory, including information about all the network interfaces and storage options contained in the server, down to serial number and firmware revision level.

Network Interface(s)

This selection reports on the specific configuration information about all the network interfaces contained in the system. Network interface type, IP address, subnet mask, speed, and EtherBand™ configuration information is all reported. In addition to being the reporting interface for this detailed information, this selection is also used to configure the server's network interfaces. Configuration can be done in real time, such that changes which are made to server network interface configuration are immediately reflected in server operation.

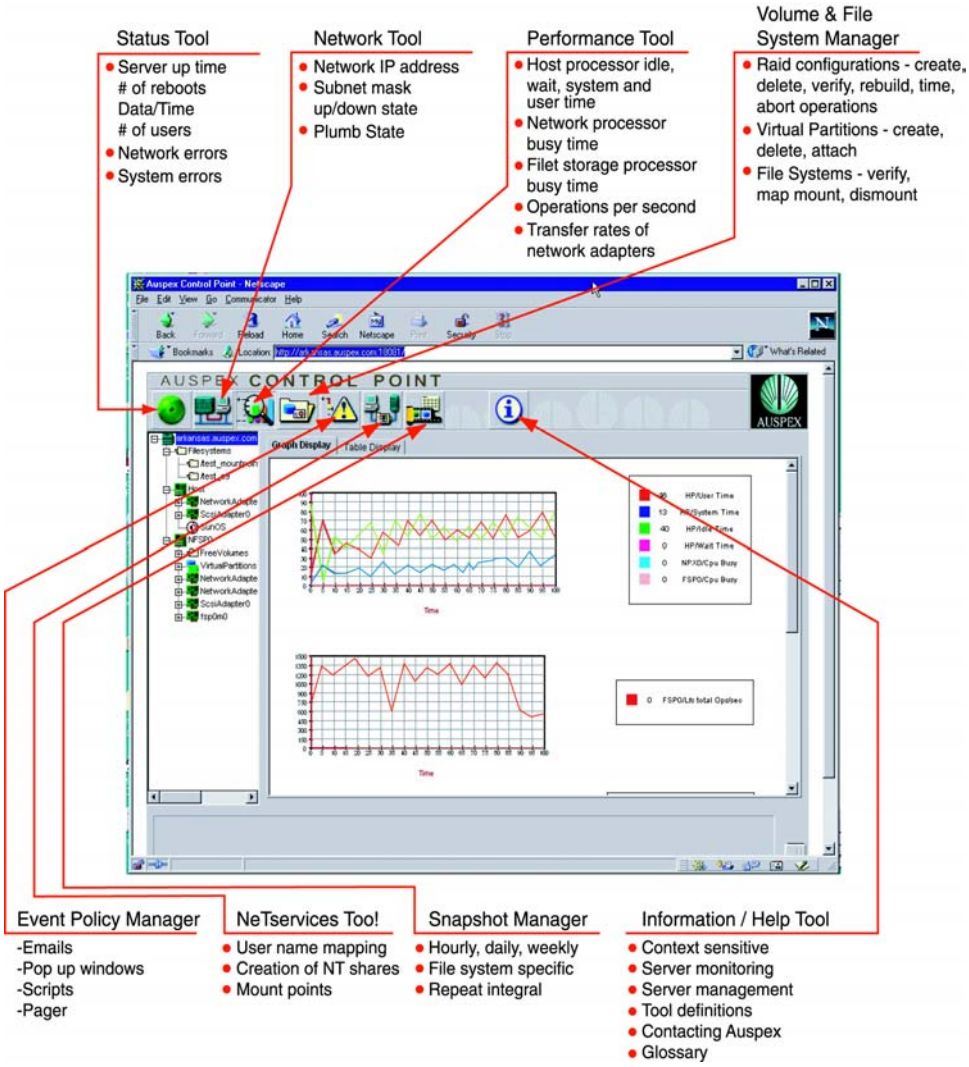

Figure 26 - Auspex Control Point provides a view of logical and physical server components (left panel) as well as rich configuration, maintenance, and monitoring functions.

Performance Monitor

Control Point offers robust systems management tools.

Control Point's Performance Monitor features two valuable functions for the system administrator. All server performance metrics, such as network interface utilization, host processor utilization, operations per second of file systems, and overall system cache utilization can be displayed in real time in either tabular or graphical format. The Performance Monitor also features a data logging utility which allows a system manager to select metrics to be monitored, frequency of that monitoring, and time to start and stop metric monitoring. These collected statistics are saved on the server in .CSV (Comma-Separated Variable) files for easy import into standard data analysis packages. This ability allows system administrators to gain additional visibility and perspective into the server's operation over time and use this information to plan for future requirements and server upgrades.

Volume and File System Manager

The Control Point™ Volume and File System Manager is where all details of the server's storage configuration is managed. All RAID configuration options, such as creation, deletion, verification, rebuild, and tuning are supported, as are all Virtual Partition

configuration options, which include creation, deletion, and member attachment. File System operations, including mount, dismount, map, and verfication are also supported through the Volume and File System Manager.

Event Policy Manager

One of the major requirements of any system management package is alerting the system administrator when certain critical event thresholds are reached. The Control Point™ Event Policy Manager allows for multiple thresholds (2 low thresholds and 2 high thresholds) to be set for all server performance metrics. When thresholds are crossed multiple event notification actions can be triggered. These actions include an E-mail or a pop-up browser notification, logging of the event to a file, or execution of a UNIX script.

NeTservices Manager

The Control Point™ NeTservices Manager allows for configuration of the Auspex NeTservices optional software. NeTservices allows an Auspex NetServer to serve Windows NT and Windows 2000 files as well as Unix files and enables true bilingual file sharing in a centralized storage environment. The NeTservices Manager allows a system administrator to map UNIX user account names to Windows user account names as well as creating Windows NT and Windows 2000 NT file share names and mountpoints. For further information about NeTservices, see **Chapter 5** and **Figure 15.**

Snapshot Manager

The Control Point™ Snapshot Manager provides for the time scheduling and administration of file system snapshots. Snapshots can be scheduled for any time of the day and can be automatically repeated hourly, daily, or weekly on a per file system basis. For further information about Snapshot, see **Chapter 8** and **Figure 20.**

Information/Online help

Control Point™ features context sensitive on-line help. For example, if a system administrator were configuring a network interface and wasn't comfortable with some of the terminology used, clicking the Information button will lauch the administrator into the specific portion of on-line help associated with configuration of network interfaces.

Auspex Service and Support 10

The Auspex commitment to *Solutions*

Service is often the key difference between a product and a solution. A company's ability to provide the best possible information and support from pre-sale through post-sale creates true customer partnerships and long term relationships. Auspex is committed to delivering the most responsive, comprehensive and cost-effective support and value-added services available. The Auspex service team will assist prospects and customers in realizing the highest return possible on their IT investments and maximizing the level of availability and productivity from NS3000 equipment. Being the primary innovator of NAS, Auspex is widely considered by customers and industry analysts alike to have the highest level of expertise in both storage and networking technology.

Being the primary innovator of NAS, Auspex is widely considered by customers and industry analysts alike to have the highest level of expertise in both storage and networking.

Auspex Professional Services

Auspex offers a full range of professional services designed to install, bring up, relocate, re-certify, recover from, and avoid any problems involved with a customer's Auspex technology investment. The Auspex professional services offering also includes a comprehensive Business Protection Services (BPS) that evaluates and provides a variety of client-specific environment services such as power conditioning and cabling for UPS.

Additionally the company can provide both technical and administrative staff with expertise in a variety of IT operational areas on an outsourced basis. Such an outsourcing program allows clients to save significant time and money compared to continually developing and training full time staff.

The Auspex professional services offering also includes a comprehensive Business Protection Services (BPS).

Basic NS3000 Warranty

The Basic NS3000 Warranty includes 24x7 telephone NetOS support for 90 days and telephone, parts and labor support for hardware from 8AM to 5PM Monday through Friday, with a next-business-day, on-site response for 1 year. In addition to a solid basic warranty, Auspex offers additional levels of warranty and support services.

Premier NS3000 Warranty

The Premier NS3000 Warranty extends NetOS telephone support to 1 year and enhances hardware support to 24x7 and shortens the on-site response time to 4 hours from time of problem diagnosis.

Auspex Premier Software Option

The Auspex Premier Software Option delivers trusted large scale support 24 hours a day, seven days a week, in mission-critical environments that require continuous access to crucial data. It is available as a 90-day warranty, a 9 month warranty or a 1 year warranty.

The Auspex Competitive Advantage

Robust on-board diagnostics combines the UNIX command set and extensive NetOS capabilities.

Auspex offers robust factory dial-in capability that provides world class ability to experts to diagnose problems if they occur. This significantly shortens the time for problem diagnosis and resolution. In addition the NS3000 comes with robust on-board diagnostics. This combined suite of tools use the UNIX command set and extensive NetOS capabilities. Remote capability allows the SE to run and view the performance monitor "perfmon," assess and adjust network configurations, and assess and adjust file system configurations. This provides a "Virtual SE" with each machine, with strict customer controlled access.

Most significantly Auspex has many years of experience in solving network problems. Auspex Customer Engineers (ACE) handle both pre- and post-sales work thereby providing for continuity in the installation, training and management of NS3000 systems.

Summary 11

World's best UNIX and Windows information sharing and file serving solution

The Auspex NS3000 series of file servers offers the most advanced set of solutions available in the market today.

We have discussed the Auspex Systems NS3000 Series of file servers and demonstrated their world class capabilities. The Auspex NS3000 series of file servers offers the most advanced set of solutions available in the market today when file serving and file sharing are required. The Auspex file sharing solutions directly benefits a company's business objectives. The unique Auspex universal file sharing solution allows any user connected to any network to access any file with full read and write privileges anywhere and at any time. File sharing is critical to achieving business benefits for companies in mechanical computer aided development, software development, oil and gas exploration, internet applications like email, FTP and News (NNTP), and streaming media. These applications are discussed in greater detail in the companion report titled ***The Auspex 3000 Storage Architecture Guide - Second Edition*** which is available from www.auspex.com.

Auspex infrastructure solutions directly benefits a company's IT infrastructure and bottom line. The full range of NS3000 infrastructure solutions consists of a complete set of availability, performance, file protection, and management features for robust network file serving and file sharing with consistently high performance. All solutions are based on the Auspex philosophy of hardware, software and service excellence that addresses every detail required by the most demanding business requirements. Auspex file servers are field proven by customers such as Motorola, Fujitsu UK, Nuevo Energy, LTX and the PeMEX Mexican Petroleum Company.

Auspex is focused on the entire network, storage and computing infrastructure.

While most storage companies are focused only on the ability of their storage products to provide secure, fast and available information storage, Auspex is instead focused on the entire network, storage and computing infrastructure. Auspex believes that little good comes from a fast storage device or a fancy SAN architecture, if files take forever to get to users due to network problems or inefficient file sharing software. Auspex believes that the characteristics of the networks are equally important to the characteristics of the storage system. The commitment of Auspex is to optimize the entire network, storage and computing infrastructure for fast, secure and "always-available" delivery of files to users AND the sharing of files among users regardless of the computers they are using. Finally, Auspex Service and Support is continually ranked as best among file server vendors.

Four Approaches to UNIX and Windows File Sharing

A

Separate servers make data sharing difficult

Sharing data is difficult with general purpose file systems.

Maintaining separate Windows and UNIX servers, with separate sets of clients, is the path of least resistance. However, this approach makes it difficult for UNIX and Windows systems to share files and system management facilities such as backup. Users moving from UNIX to NT are likely to be dissatisfied with performance, availability and reliability.

The biggest disadvantage of the approach is that sharing files between different types of clients may require a difficult-to-manage backup scheme resulting in additional administration cost for administrators to learn. In addition, UNIX system management facilities like backup are available only for UNIX users and Windows backup facilities are available only for Windows users. If Windows runs on typical PC servers, users moving from UNIX to Windows may become less productive and less satisfied due to the lower availability and performance of PC servers. This approach avoids the complexities involved in evaluation and implementing the other three approaches for administrative personnel. However, the cost of fragmenting the work environment and lost user productivity is often far greater than the initial investment in a superior architecture, such as the NS3000.

Client-based emulation uses processing power inefficiently

Client-based emulation can be an administrator's nightmare.

Client-based emulation implements an NFS protocol stack on a PC client or a CIFS protocol stack on a UNIX client. For example, SunSoft's PC-NFS implements an NFS protocol stack on a PC client. This approach has the advantage of requiring no changes at the server by system administration personnel and any problems with the product affect only clients using the emulation software. Technically savvy users can often manage the solution themselves. For this situation or when users only access files on a foreign system on an occasional basis, client-based emulation can be an appropriate solution to mixed UNIX and Windows environments.

However, these products tend to be relatively slow because of the extra work that the client processor must do to emulate the foreign protocol. They can also create stability problems on the client. To allow two-way file sharing, two different products are needed: one to allow Windows clients to access UNIX systems and another to allow UNIX clients to access Windows systems. There is also a major cost associated with client-based emulation that is the administration overhead of keeping all the clients current with new releases or getting the software installed on all new clients. Furthermore, since it is quite difficult to mask fundamental differences in file systems client-emulation is usually imperfect. Perhaps the biggest drawback to client-based emulation occurs when the user cannot solve all installation, configuration and administration problems placing a burden on system administrators. Finally, if there are large numbers of clients, total costs to the organization can be high.

Restrictions of server-based emulation

Server-based emulation implements foreign protocol conversion software on a server, for instance, CIFS on a UNIX server or NFS on an Windows server. An example this is the

Server-based emulation suboptimizes performance.

Samba suite of freeware components that implements a CIFS protocol stack on a UNIX server. TotalNET Advanced Server (TAS) from Syntax is bundled with Sun's Netra 150 server product and is an example of a commercial server-based emulation product.

Server-based emulation is generally better than client-based emulation in terms of availability, performance and manageability since no special software is required on the client machine. Since servers are usually more powerful than clients, performance tends to be better. Availability tends to be better than client-based emulation since servers tend to be more tightly controlled, monitored and configured. In addition management difficulties that do occur are confined to servers and not spread over an entire client population. Finally, server-based emulation is likely to be better because the product's central location is more strategic.

Server-based emulation products however execute as user-level processes as opposed to running in the UNIX kernel. This is not the most efficient way to tune a protocol on a server and performance is less than kernel based software since it takes far more instructions to accomplish the same amount of work. Like client-based emulation, server-based emulation is usually imperfect due to the difficulties of emulating facilities like file locking and security on a system that has different features and is a fundamentally different machine. It is particularly difficult to support Windows users on UNIX, since Windows is more flexible and offers more options. In essence clients that use emulation are still "second-class citizens" when compared to native clients. Although server-based emulation is a step up from client-based emulation and is appropriate for more users and provides more intensive file access of the foreign system. It's performance limitations make it less than ideal for large numbers of users or even moderate numbers of users with high-intensity application.

Advantages of bilingual Network Attached Storage

Bilingual file servers are the best choice.

Bilingual file servers such as the NS3000 are typically the most appropriate solution for high-intensity mixed UNIX and Windows environments. They become more attractive as the amount of data increases and as the number of clients requiring access to both UNIX and Windows files increases. Technically, the bilingual file server is greatly superior to both of the emulation-based approaches. Performance and reliability are likely to be much better, since protocol stacks are part of the kernel, not user-level add-ons. Furthermore there are no "second-class citizens" since a bilingual server treats CIFS and NFS as peers even under heavy concurrent loads on both protocols. Bilingual servers provide the best structure for integrated management of file locking and can make sure that NFS users cannot violate CIFS locks. The bilingual file server also provides "best of breed" administration and management. For example, UNIX backup tools can be used for all files while Windows administrators can manage the system using standard administrative tools. Because of the scalability, manageability, and reliability of this approach, the total cost of ownership over time is less than other approaches especially for high-intensity applications.

Although acquisition costs may be higher for a bilingual file server such as the NS3000, total cost of ownership (TCO) is lower over the life of the product due to:

- centralized backup
- higher productivity
- data integrity
- lower administration

Glossary of Terms

10BaseT
Ethernet with a data transfer rate of 10 Mbits/sec.

100BaseT
Also known as Fast Ethernet with a data transfer rate of 100 Mbits/sec.

ACL
Access Control List. Windows networking uses a set of standard Access Control Lists (ACLs) for granting access to shares directories, and files. The ACLs offer useful combinations of specific types of access, which are called individual permissions. Individual permissions are somewhat analogous to UNIX permissions. They consist of read (R), Write (W), Execute (X), Delete (D), Change Permissions (P), and Take Ownership (O). UNIX supports three sets of file and directory permissions: owner, group, and world. This is the familiar -rwx-rwx-rwx that shows up in the output from the UNIX ls-al command.

API
An Application Programmer's Interface or API is a standardized set of software commands (calls) that can be used to access a particular software program in a consistent and reliable way.

ATM
Asynchronous Transfer Mode. A suite of network protocols providing low-level services spanning local and wide-area networks. ATM is intended to provide the switching and multiplexing services necessary to carry voice, data and video and multimedia traffic using fixed 53-byte cells. Standards are defined to allow ATM to emulate traditional LANs (LANE).

b
Abbreviation for "bit" where 8 "bits" comprise a byte.

B
Abbreviation for byte or the equivalent of one character in text.

BDC
One server running Windows Server acts as the Primary Domain Controller (PDC), that maintains the centralized security databases for the domain. Other computers running Windows Server in the domain function as Backup Domain Controllers (BDC) and can authenticate logon requests. The PDC or BDC authenticates users of a Windows Domain. See also PDC.

BDX
The NS3000 Block DataXceleration Engine provides the capability to stream blocks of data from disk to tape, or node to node, and provides the foundation for extremely rapid backup.

CIFS
Common Internet File System. A connection-oriented, network file-sharing protocol developed by IBM and Microsoft as part of LAN Manager. CIFS is the native file sharing protocol for systems running Windows for Workgroups, Windows95 and Windows NT. Sometimes referred to as SMB.

Control Point
Auspex's proprietary management control software.

CPU
Central Processing Unit. Can refer to either a processor chip such as Sun's SPARC or Intel's Pentium, or to a processor chip or chips and support circuitry on a CPU board.

DataXpress
Communication among the NS3000's multiple hardware processors and software processes are handled by DataXpress, a low-overhead message-passing kernel executing on each processor.

ECAD
Electrical Computer Aided Design

EM-Net
The NS3000 Environmental Monitoring Network that connects to all chassis in an NS3000 system and reports a variety of control information to the Host Node.

Ethernet
A Local Area Network (LAN) protocol developed by Xerox in cooperation with Digital Equipment and Intel in 1976. Ethernet supports a star or bus topology and supports a data transfer rate of 10 megabits per second or 10 Mbps. The Ethernet specification formed the basis of the IEEE 802.3 standard, which specifies the physical and lower software layers. Ethernet uses the CSMA/CD access method for handling simultaneous demands and is one of the most widely implemented LAN standards.

Fast Ethernet
Fast Ethernet or 100BaseT, defined by the IEEE 802.3 committee, provides a 100 Mbps standard that is compatible with existing 10BaseT installations, preserving the CSMA/CD media access control (MAC) protocol.

FastFLO™
The NS3000 proprietary file system that is optimized for providing high performance and consistent file services.

FC
An acronym for Fibre Channel

FCP
FCP is an acronym for Fibre Channel Protocol, an ANSI standard covering Fibre Channel protocol for SCSI.

FDX
The NS3000 File DataXceleration Engine provides an interface that acts at the file system level to provide the basis for rapid file-by-file or directory backup.

Fibre Channel
ANSI standard designed to provide high-speed data transfers between workstations, servers, desktop computers and peripherals. Fibre channel makes use of a circuit packet switched topology capable of providing multiple simultaneous point-to-point connections between devices. The technology has gained interest as a channel for the

attachment of storage devices, but has limited popularity as high-speed networks interconnect. Fibre channel can be deployed in point-to-point, arbitrated loop (FC-AL), or switched topologies. Fibre channel nodes log in with each other and the switch to exchange operating information on node attributes and characteristics. This information includes port names and port IDs and is used to establish interoperability parameters.

Fibre Channel Protocol (FCP)
FCP is an ANSI standard covering Fibre Channel protocol for SCSI.

FMP
Functional Multiprocessing (FMP) is the term Auspex uses for its unique distributed parallel processing NS3000 architecture. Each NS3000 I/O Node II is based on an Asymmetric Multiprocessing design with two processors and a unique real time OS called the DataXpress™ kernel. Each processor simultaneously and efficiently executes different functions in the network file serving process. One processor handles network processing and the other processor handles File and Storage Processing. A Host Node is based on the traditional general-purpose single CPU computer running the general purpose Solaris OS, and is used primarily for system management activity. Up to three I/O Node IIs and one Host Node are connected by a Scalable Coherent Interface (SCI). System software consists of a unique custom messaging system that enables efficient network and storage processing on the I/O Node IIs and efficient system and data management on the Host Node. The FMP architecture improves system availability compared to other approaches by isolating the I/O Node IIs from unplanned outages of the general purpose OS (Solaris). I/O processing can therefore continue in the event that the Host Node is down. See also SMP, parallel processing, SCI.

FSP
The File System and Storage Processor refers to one of the two Intel Pentium processors on an I/O Node II of an NS3000 system. This processor runs highly optimized microcode that manages all file system and storage processing of the I/O Node II and communicates with other I/O Node IIs and the Host Node. See also Network Processor (NP).

Gigabit Ethernet
Standard of the IEEE 802.3 committee that provides a mechanism for conveying Ethernet format packets at GB/s speeds. Gigabit Ethernet preserves the CSMA/CD access method with support for 1 repeater, use the 802.3 frame format, provides simple forwarding between Ethernet, fast Ethernet and Gigabit Ethernet, support both fiber and copper, and accommodates the Ethernet standard for flow control.

Gigabyte
1024 Megabytes.

GUI
An acronym referring to a Graphical User Interface that is the screen presented to a user in any computer application.

HDDA
A term that refers to a High-Density Disk Array shelf in an NS3000 containing 28 drives arranged in four drawers of 7 drives each. A maximum three I/O Node II NS3000 system contains nine HDDAs or 9x28=252 disk drives.

Inode
In UNIX, an inode is an index to files.

I/O Node II

A module of the NS3000 architecture containing file server hardware and software components.

IP

The IP (Internet Protocol) is the underlying protocol for routing packets on the Internet and other TCP/IP-based networks. IP is an internetwork protocol that provides a communication standard that works across different types of linked networks for example Ethernet, FDDI or ATM. In an internetwork, the individual networks that are joined are called subnetworks or subnets. IP provides a universal way of packaging information for delivery across heterogeneous subnet boundaries. See also TCP Transmission Control Protocol.

Java

Developed by Sun Microsystems, Java is now a standard software language for developing plug-in applications.

Journaling

A journaling file system keeps track of all changes to files as transactions occur in real time. In the event of unexpected system problems, the file system can be restored to a consistent state by updating a prior copy of the file system for the changes made from the point in time that the copy was made.

LADDIS

An acronym formed by names of the group (Legato, Auspex, Data General, Digital Equipment Corporation, Interphase, and Sun) that developed and popularized SPEC's vendor-neutral NFS server benchmark of the same name. See also SPEC.

LAN

Local area networks or LANs are networks of computers that are geographically close together; this usually means within the same building.

MAC

Media Access Controls or MACs are the rules defined within a specific network type that determines how each station accesses the network cable. Using a token-passing method, a carrier sensing and collision detection method or a demand priority method prevents simultaneous access to the cable. The MAC used for 100BaseT Ethernet as implemented in the Auspex EtherBand™ feature is based on the "demand priority" access method in which the central hub scans all its ports in a round-robin fashion to detect stations that want to transmit a frame. Higher priorities can be requested by ports to transmit real-time information like video or audio.

MCAD

Mechanical Computer Aided Design

MIB

Management Information Base is a set of standards for detailed system information that is reported to a control console for SNMP compliance. Its intent is to provide common metrics for heterogeneous computer systems.

MTBF

Mean Time Between Failure. A key component of the availability equation, AVAILABILITY = (MTBF – MTTR) ÷ MTBF. Example: A server that on average fails once every 5,000 hours and on average takes 2 hours to diagnose, replace faulty components and reboot would have an availability rating of (5,000 – 2) ÷ 5,000 = 99.96%.

MTTR
Mean Time To Repair. Includes the time taken to diagnose the failure, replace or repair faulty component(s) and reboot the system. See MTBF.

N+1
An N+1 power supply design provides for one redundant power supply on the shelf to provide full system power in the event of a power supply failure.

NAS
Network Attached Storage.

NDMP
NDMP is a standard protocol for network-based backup of network-attached storage. NDMP hides the unique interfaces from third party backup software which allows this software to execute on any NDMP compliant system on the network (such as the NS3000 Host Node, and control backups on the NS3000 using standard commands.

NeTservices
The Auspex software product that provides for consistent high performance UNIX and NT file services and makes the NS3000 a true bilingual file server.

NIC
Network Interface Cards (or NICs) in the NS3000 support 10/100BaseT Ethernet, Gigabit Ethernet, FDDI or ATM. There are from one to three on each I/O Node II. See Table 3.

NIS/NIS+
Network Information Service. This is ONC's general name-binding and name-resolution protocol and service.

NFS
Network File System. NFS is an ONC application-layer protocol for peer-to-peer, distributed, file system communication. NFS allows a remote file system (often located on a file server) to be mounted transparently by client workstations. The client cannot perceive any functional difference in service between remote and local file systems (with trivial exceptions). NFS is the most popular ONC service, has been licensed to over 300 computer system vendors, and runs on an estimated 10 million nodes. It is a de facto UNIX standard. See also VFS, ONC, and NFSv3.

NFSv3
NFS version 3. References to NFS generally imply NFS version 2 protocol. NFS version 3 is an update to the NFS protocol. Significant among the many changes made for NFSv3 are the adoption of a safe asynchronous write protocol and the use of block sizes up to 64 KB. Other protocol changes are intended to improve the overall network and client efficiency and provide improved support for client-side caching.NFS ops/s NFS operations per second. Typical NFS operations include: lookup, read, write, getattr, readlink, readdir, create, remove, setattr, and statfs.

Node
See Fibre Channel.

NP
The Network Processor (NP) refers to one of the two Intel Pentium processors on an I/O Node II of an NS3000 system. This processor runs highly optimized microcode that manages all network processing of the NS3000 I/O Node II and communicates with other I/O Node IIs and the Host Node. See also File System and Storage Processor (FSP).

NS3000
A term that refers to the Auspex Net Server 3000. Series of products consisting of NS3000 Base Systems, NS3000ER (Extended RAID) Systems, and NS3000HA (High Availability) System.

NTFS
A term that refers to Windows NT file system.

NVRAM
Non-volatile random access memory such as static RAM that will not lose data in the event that power is lost to the chip.

ONC
Open Network Computing. The trade name for the suite of standard IP-based network services—including RPC, XDR and NFS—promulgated by Sun Microsystems.

Operating System
The operating system is the most important software program that runs on a computer. The Operating System (OS) performs basic tasks such as recognizing input from a keyboard, sending output to the display screen, keeping track of files and directories on the disk and controlling peripheral devices such as disk drive and printers or a mouse. The OS acts as a traffic cop and schedules the various programs that the computer executes. The OS is also responsible for security, ensuring that unauthorized users do not access the system. Operating systems can be classified as follows:
1) Multi-user – allows two or more users to run programs at the same time. 2) Multi-processing – supports running a program on more than one CPU. 3) Multi-tasking – allows more than one program to run concurrently. 4) Multi-threading – allows different parts of a single program to run concurrently. 5) Real Time – Usually a stripped down OS that responds to input instantly.

Parallel processing
Parallel processing refers to when a single computer simultaneously uses more than one CPU to execute a program. Ideally parallel processing makes a program run faster because there are more CPUs running it. In practice, it is often difficult to divide a program so that separate CPUs can execute different portions without interfering with each other. Among NAS vendors, only the Auspex NS3000 effectively overcomes this problem by designing each I/O Node II with two processors each performing separate portions of the network file-serving task. In addition the NS3000 links multiple I/O Node IIs together by a highly efficient Scaleable Coherent Interface (SCI) interconnect that allows the multiple nodes to act as one system. See also Functional Multiprocessing (FMP).

PCI
The Peripheral Channel Interconnect is an ANSI standard for an I/O bus used predominantly in PC design.

PDC
One server running Windows Server acts as the Primary Domain Controller (PDC), that maintains the centralized security databases for the domain. Other computers running Windows NT Server in the domain function as Backup Domain Controllers (BDC) and can authenticate logon requests. The PDC or BDC authenticates users of a Windows NT Domain.

PDU
Power Distribution Unit or Power Shelf in the NS3000. A cabinet model contains from three to seven power supplies and is N+1 redundant. See also N+1.

PIP
A Physical IP address.

Port / Port ID
See Fibre Channel.

RPC
Remote Procedure Call. An RPC is an (almost) transparent subroutine call between two computers in a distributed system. ONC RPC is a Sun-defined session-layer protocol for peer-to-peer RPC communication between ONC hosts.
ONC RPC underlies NFS.

RAID
Redundant Array of Independent Disks. RAID is used to increase the reliability of disk arrays by providing redundancy either through complete duplication of the data (RAID 1, i.e., mirroring) or through construction of parity data for each data stripe in the array (RAID 3, 4, 5). RAID 5, which distributes parity information across all disks in an array, is among the most popular means of providing parity RAID since it avoids the bottlenecks of a single parity disk.

RAID Controllers
The NS3000 RAID controllers provide a highly optimized scheme for securely managing RAID configurations on NS3000 systems. The Auspex RAID controllers allow RAID arrays to be expanded online, and support conversion of an array from one RAID level to another.

SCI
Scalable Coherent Interface is an ANSI standard (#1596-1992) that is the modern equivalent of a processor-memory-I/O bus and a Local Area Network combined and made parallel to support distributed multiprocessing. The SCI interconnect has very high bandwidth, very low latency and a scaleable architecture. This allows building large high performance systems and is used by Convex/HP supercomputers, Sun Clusters, Sequent, Auspex and others. Network latency has been measured at 150 times less than previous network connections for efficient and fast communication between computer nodes.

SCSI
Small Computer System Interface. An intelligent bus-level interface that defines a standard I/O bus and a set of high-level I/O commands. The SCSI busses in the NS3000 are used to connect multiple peripheral devices such as disk drives tape drives. Each SCSI device has an intelligent SCSI controller built into it. There are currently many flavors of SCSI defined by different bus widths and clock speeds. The seven major variations of SCSI are SCSI 1, SCSI 2 (Fast / Narrow), SCSI 2 (Fast / Wide), Ultra SCSI (Fast / Narrow), Ultra SCSI (Fast / Wide) – also called SCSI 3, Ultra 2 SCSI (Narrow), Ultra 2 SCSI Wide. Single ended SCSI is used when the peripheral device is close to the point of attachment as in the NS3000 method of attaching disk drives. Differential SCSI provides for reliable operation over greater distances and is used in the NS3000 for tape drive connections.

SE
System Engineer(s) perform a variety of technical pre and post sales services for customers and prospects.

SID
A Windows networking term meaning System Identification (SID).

SMB
Server Message Block protocol. See CIFS.

Snapshot
A term that refers to a copy of a file system at a certain point in time. Snapshots are used for backup and recovery.

SMP
Symmetric Multi-Processing. A computer architecture in which processing tasks are executed in parallel on multiple, identical, general-purpose CPUs that share a common memory. SMP computer systems usually have modified operating systems that can themselves execute concurrently. The SMP architecture offers high computational throughput, but not necessarily high I/O throughput. See FMP.

SNMP
Simple Network Management Protocol. SNMP is a protocol used for communication between simple, server-resident SNMP agents that respond to network administration requests from simple-to-sophisticated SNMP manager tools running on remote workstations.

Solaris 2.x
Sun's UNIX operating system.

SPARC
Scalable Processor Architecture. SPARC International's specification for the Reduced-Instruction-Set-Computer (RISC) CPUs found in systems sold by Sun Microsystems, Auspex, etc.

SPEC
Standard Performance Evaluation Corporation. A nonprofit corporation of vendors' technical representatives that develops and certifies accurate, vendor-neutral, computer-system benchmarks. As an example, popular SPEC CPU benchmark metrics include SPECint, SPECfp and the now obsolete SPECmarks. See LADDIS.

SPECnfs
SPECnfs measures ops/s as a measure of NFS performance standardized by SPEC. This unit of measure is often used interchangeably with SPECNFS-A93 ops/s. The A93 suffix indicates the first of what may evolve into a series of workloads, each corresponding to different LADDIS variations simulating the loads and traffic patterns of application environments like ECAD, MCAD, imaging, etc. The current version is SFS97 and incorporates NFSv3 testing.

Stack
In software protocols the CPU processes one instruction after another in a serial fashion. The exact sequence of these instructions is referred to as a stack.

Stripe
In RAID terminology, a stripe is when data is read or written in parallel to or from multiple disks instead of reading or writing all data to one disk. Striping provides much higher performance through its parallel design.

TB
A Terabyte (TB) equals 1024 Gigabytes.

TCP

Transmission Control Protocol or TCP is a transport layer component of the Internet's TCP/IP protocol suite. It sits above IP in the protocol stack and provides reliable data delivery services over connection-oriented links. TCP uses IP to deliver information across a network and makes up for the deficiency of IP providing a guarantee of reliable delivery services that IP does not. TCP messages and data are encapsulated into IP datagrams and IP delivers them across the network.

UFS

UNIX File System. UFS is the standard file system type in the BSD 4.3 kernel.

WAN

Wide Area Networks or WANs are networks of computers that are geographically dispersed and connected by radio waves, telephone lines or satellites.

Zoning

In a SAN environment this is a workaround to security problems with the Fibre Channel specification whereby data pools are assigned to a specific server. This defeats the basic premise of SAN whereby "any application" can have access to "any data."

Corporate Headquarters
2800 Scott Boulevard
Santa Clara, CA 95050
USA
408.566.2000 **phone**
800.735.3177 **fax**

http://www.auspex.com **web**

Japan
Auspex K.K.
ATT Shinkan, 6K
2-11-7 Akasaka
Minato-ku
Tokyo 107-0052
Japan

+81 3 3586 1751 **phone**
+81 3 3586 1778 **fax**

United Kingdom
Auspex Systems, Ltd
Portland House
Aldermaston Park
Church Road
ALDERMASTON
Reading RG7 4HR
United Kingdom

+44 1189690900 **phone**
+44 1189692266 **fax**

France
Auspex Systems, SA
40, Boulevard Henri Cedex
France

+33 1 42 04 9150 **phone**
+33 1 42 04 9151 **fax**

Germany
Auspex Systems, GmbH
Ingolstaedter Stasse 22
80807 Muenchen
Germany

+49 89 350151-0 **phone**
+49 89 350151-50 **fax**

Singapore
Auspex Asia Pacific
#06-02 International Building
360 Orchard Road
Singapore
238869

+65 732 9770 **phone**
+65 732 3660 **fax**

Auspex Worldwide

www.auspex.com

AUSPEX