

**APPLICATION NOTE**

March 2006

## Home directory storage consolidation with the StorageTek™ 5000 Family of NAS Appliances

**ABSTRACT**

In the early days of distributed computing, home directories were commonly assigned to individual desktops, making it difficult for administrators to protect and preserve data availability during backup, restore, and disaster recovery operations. This paper describes these problems and outlines steps for implementing and consolidating home directories on a product in the StorageTek™ 5000 Family of NAS Appliances.

<b>1.0</b>	<b>Executive summary</b>	<b>2</b>
<b>2.0</b>	<b>Introduction</b>	<b>2</b>
<b>3.0</b>	<b>The distributed desktop workstation and server architecture</b>	<b>2</b>
<b>4.0</b>	<b>Network-attached storage consolidation</b>	<b>3</b>
<b>5.0</b>	<b>NAS-based home directory storage consolidation benefits</b>	<b>3</b>
5.1	5000 NAS Family autohome capability	3
5.2	Logical volume management	3
5.3	General file sharing	4
5.4	Data protection	4
5.5	Data integrity	5
5.6	Information technology management	5
5.7	Backup/restore and disaster recovery	6
<b>6.0</b>	<b>Return on investment and reduced total cost of ownership</b>	<b>6</b>
<b>7.0</b>	<b>Conclusion</b>	<b>7</b>

## 1.0 Executive summary

In the quest to manage ever-growing amounts of data, distributed techniques used in the past are often no longer viable. Specifically, home directories located on the machines of individual users create severe headaches for administrators and impact the business bottom line through increased management overhead and total cost of ownership on hardware.

One answer is to consolidate data on network-attached storage (NAS) devices like the Sun StorageTek™ 5000 Family of NAS Appliances. This family of hardware and software solutions allows you to centralize data, thereby reducing management costs and maximizing your investment in existing hardware.

Home directories are a prime candidate for data consolidation, as hosting on a NAS appliance allows for better data security and file sharing. This paper describes the common problems with distributed data management and the benefits of home directory storage consolidation with a 5000 NAS appliance.

## 2.0 Introduction

IT environments require home directories so that users can store their data for fast and easy access. During the first wave of distributed computing, it was standard practice to assign home directories to each individual desktop and/or server on their associated direct-access storage (DAS) devices. The distributed nature of such a storage implementation created almost insurmountable problems when it came to protecting and preserving data availability during backup, restore, and disaster recovery.

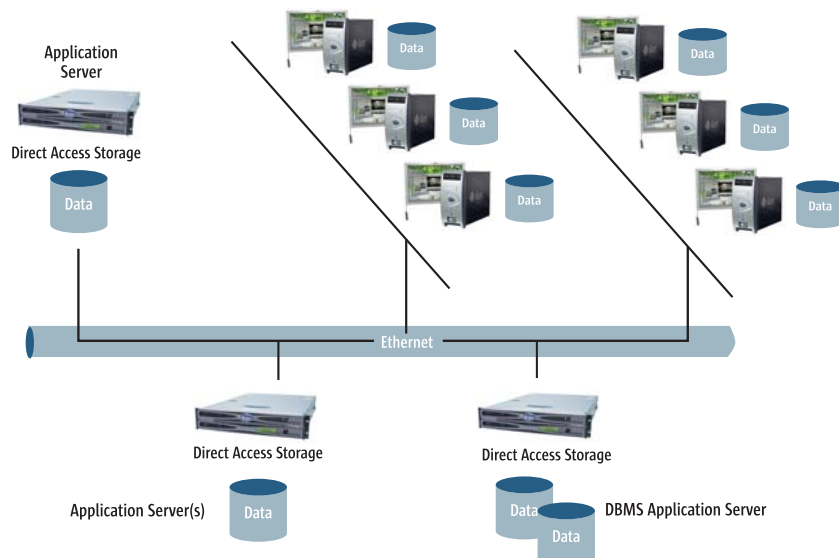
This document provides a framework for implementing and consolidating home directories on a 5000 NAS appliance — a simple, fast, reliable, and affordable implementation of NAS.

This document is intended for system and/or storage administrators with a basic working knowledge of UNIX®, NFS, Microsoft Windows, CIFS/SMB, RAID, and the 5000 NAS appliance.

## 3.0 The distributed desktop workstation and server architecture

Before looking at the benefits of home directory consolidation, we will examine the common problems with storage in distributed environments. Distributed data storage, as shown in Figure 1, has several serious shortcomings.

*Distributed home directories are difficult to back up and restore. The answer is centralized directories and management.*



*Figure 1. Distributed computing and storage architecture prior to NAS storage consolidation.*

*Desktop workstations and associated DAS tend to have no native data protection capabilities.*

Desktop workstations normally do not have data protection capabilities — such as RAID or tape backup — that are present in typical storage environments. This is because it is cost-prohibitive to implement these types of data protection capabilities in a distributed workstation and server environment. Labor, software licensing, and other hidden system administration costs tend to increase exponentially when implementing these types of capabilities in a distributed environment.

Nevertheless, preservation of data in case a particular desktop client fails or loses data still requires a nightly backup of all pertinent data on every workstation in the computing and workstation infrastructure. In a typical organization, 50–60 percent of all relevant mission-related data is stored on the desktop. This data is at great risk and is expensive to manage.

*Fast-growing storage requirements can obsolete desktop hardware prematurely. Centralization can preserve these investments.*

Finally, a typical desktop workstation or storage device will, at some point during its life cycle, no longer be able to meet user storage requirements. When this occurs, a replacement workstation or desktop device is usually ordered to provide more storage capacity. Thus, although the device may still meet user needs in every other way, its effective life is reduced unnecessarily because it cannot meet basic storage requirements.

Taken together, these problems increase the total cost of ownership of distributed storage architecture such that an alternative storage architecture is called for.

#### 4.0 Network-attached storage consolidation

The alternative is to implement an architecture in which all desktop and server components can utilize a common consolidated pool of storage. This is accomplished easily and simply on the 5000 NAS appliance utilizing the home directory feature, `autohome`. This feature leverages the existing capabilities of the Microsoft Windows desktop environment often called `Home Directory(s)` as well as the typical feature utilized in the Solaris™ Operating System and other UNIX environments called `automount`. `Autohome` consolidates all distributed storage from the desktop and server infrastructure to a centralized data content repository that resides on the 5000 NAS appliance.

#### 5.0 NAS-based home directory storage consolidation benefits

From an IT perspective, benefits of home directory consolidation range from increased management efficiency to increased data integrity and faster disaster recovery. The following sections provide more detail on each area.

##### 5.1 5000 NAS Family `autohome` capability

After data and content are copied to the 5000 NAS appliance, Microsoft-Windows-based “shares” are allocated that relate all desktop users to their corresponding data and content. Within the UNIX environment(s), `automount` allocates NFS mounts to connect all users automatically to their data and content when they log in to the computing domain. The Java™ technology-based Web graphical user interface (GUI) on the 5000 NAS appliance provides for easy allocation and administration of these home directories.

##### 5.2 Logical volume management

The Sun StorageTek NAS OS (operating system) includes a logical volume manager (LVM). Administrators use the LVM to create logical volumes with file systems and subdirectories that are, as mentioned earlier, exported intelligently to associated desktop and server elements in the infrastructure.

`Autohome` associates a logical volume with its corresponding file system and home directory account subdirectories. This file system essentially contains a root-level directory that, in turn, provides a subdirectory that is the corresponding home directory for all desktop and server elements in the storage architecture.

*The 5000 NAS appliance OS provides GUI tools to quickly associate centralized home directories with specific users.*

### 5.3 General file sharing

The 5000 NAS appliance (also often referred to descriptively as a “Filer”) can also be used to provide a general file sharing capability. Typical distributed desktop workstation and server architectures often require a centralized storage medium that can be used to share data and content among users, projects, and certain mission-critical applications. The LVM included with the NAS OS can allocate a volume for purposes of file sharing and general-purpose storage. This file system, in turn, can be sub-allocated to each group of users, projects, or applications and secured with industry-standard security tools included with the NAS application. The NAS OS supports Active Directory, Kerberos, and typical UNIX-based security paradigms while allowing for sharing of data among different OS architectures.

### 5.4 Data protection

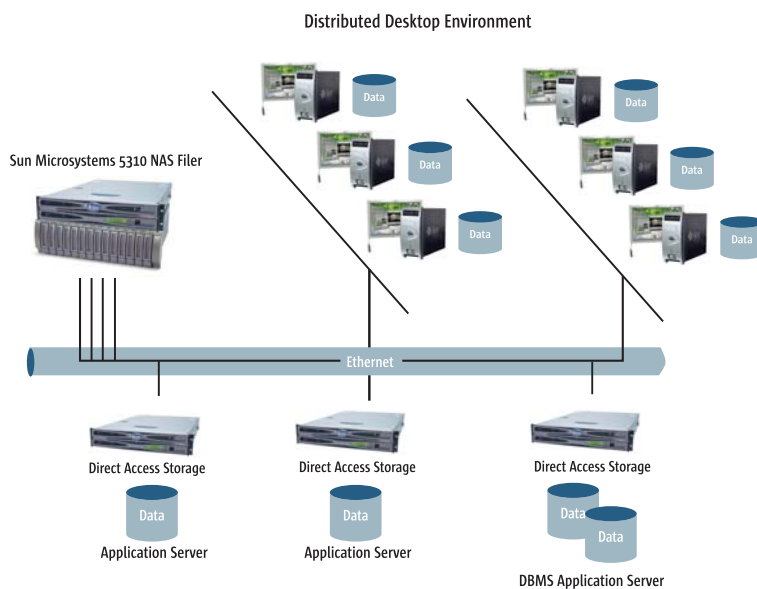
As mentioned earlier, typical DAS solutions implemented in a distributed desktop environment have little or no data protection capabilities. Also, system administration costs associated with providing these capabilities is too expensive and labor-intensive to accomplish practically.

With NAS-based home directory storage consolidation, all data and content is protected on the 5000 NAS appliance. The NAS appliance is equipped with hardware controllers that protect data using a RAID-5 data storage architecture. This RAID architecture provides two copies of the data: one is a primary copy and the other is a logical copy that can be utilized immediately in the event of a disk failure.

In addition, the 5000 NAS appliance has a “checkpoint” feature (often used synonymously with “snapshot”). This feature can provide a logical point-in-time copy of data and content either manually or on a scheduled basis. Currently, the 5000 NAS appliance provides up to 64 checkpoints per logical volume. The number of checkpoints per NAS Filer is constrained only by the available amount of storage.

Typical NAS implementations require storage to be reserved for snapshot/checkpoint data. When these NAS implementations exhaust this reserved capacity, a loss of data availability occurs. Uniquely, the NAS Filer does not require storage to be reserved for checkpoints, but rather uses storage from the file system from which it is providing the point-in-time copy. Thus, the administrator is relieved of concerns regarding a shortage of reserved allocated snapshot/checkpoint capacity and subsequent loss of data availability.

*Because it can manage file systems, the 5000 NAS appliance OS is often referred to as a “Filer.”*



*Figure 2. Distributed computing and storage architecture after NAS storage consolidation.*

When data or content is accidentally lost, erased, or corrupted, either the administrator or the user (depending on organizational policy) can restore the data or content instantaneously. All of these capabilities combine to provide a level of data protection not typically found in a distributed desktop workstation environment.

### 5.5 Data integrity

Along with data protection, the 5000 NAS appliance provides for exceptional data integrity by implementing a high-performance journaling file system (JFS).

*To provide data integrity, the 5000 NAS appliance OS includes a uniquely implemented journaling file system.*

Uniquely, the JFS for the NAS OS is implemented in an internal architecture where it is protected from data loss. Typical JFS implementations store journal data and metadata on RAM, NVRAM, or DAS — all of which provide only marginally for power loss and subsequent possible loss of data due to data corruption.

The NAS JFS deployment on the 5000 NAS appliance employs exceptional data protection measures to provide for data integrity and preserve performance. The file system journal is allocated on a RAID-protected file system, with its associated mechanisms for data protection as well as RAID controllers that include one gigabyte of mirrored cache each. If elements of the I/O infrastructure fail, data integrity is preserved via the RAID-based parity and mirrored cache. This architecture avoids loss of data availability while providing maximum protection for production data.

### 5.6 Information technology management

Hidden costs associated with maintaining mission-critical computing applications and storage infrastructure can be challenging in a distributed desktop workstation environment. Hidden costs can include excessive system and storage administration labor costs as well as costs associated with evolving mission-critical application requirements.

*Centralizing home directories can eliminate many of the hidden costs of managing distributed computing environments.*

System administration labor necessary to support a distributed desktop and server architecture can be time-consuming and very expensive. A large portion of administrative labor is associated with providing data protection, data availability, and quick recoverability. Consolidating all production data onto a single NAS 5000 appliance allows system administration to be accomplished quickly and easily, with dramatically less administrative overhead and superior data protection and data availability.

Over time, mission-critical applications usually increase in capability, which, in turn, increases computational and storage requirements. The desktop and server infrastructure must be able to quickly and seamlessly adapt to changing requirements. 5000 NAS appliances enable cost savings in the face of exponential storage growth by facilitating “on-the-fly” storage growth. It does this in two ways:

- Through a feature that expands a production logical volume without incurring outages.
- Through the very nature of storage consolidation in that it separates data from the mission-critical application platforms.

## 5.7 Backup/restore and disaster recovery

Daily, weekly, and monthly backup procedures are chores that challenge every data center today. The exponential growth of data and the associated storage infrastructure that has accommodated this growth requires special consideration to facilitate the everyday task of data protection.

As mentioned earlier, checkpoints provide a logical backup of data and content by providing a logical point-in-time copy of a logical volume and its associated file system. It is important to note that this is not a second physical copy of the data and thus should not be considered a permanent backup of data and content. Backup and restore of data, though, *is facilitated* using the 5000 NAS appliance checkpoint feature. This is because in order to provide a coherent backup of data, the file system must be static at the time of the backup. Checkpoints provide this critical element to the backup procedure.

Backup tools typically require some sort of system “snapshot” — a need met by NAS checkpoints. The 5000 NAS appliance provides backup/restore capabilities in concert with commercially available industry-standard tools. The 5000 NAS appliance embeds an NDMP backup/restore agent in the NAS OS at no charge. This agent is certified to function with these commercially available backup/restore tools.

Disaster recovery has become extremely visible and important in all modern data centers. The 5000 NAS appliance provides replication as well as near real-time logical volume/file system mirroring. This means that the 5000 NAS appliance can replicate and mirror a production file system to another 5000 NAS appliance. Supported replication and mirroring include:

- One-to-one
- Many-to-one
- Bi-directional

Easy volume mirror allocation, promotion, and recovery are provided via a Web-based GUI. Note that mirror coherency and performance are highly dependent on network connectivity.

## 6.0 Return on investment and reduced total cost of ownership

Typically, return on investment (ROI) and total cost of ownership (TCO) are very difficult to measure because associated metrics are hard to identify and assess.

However, when implementing NAS-based storage consolidation, real ROI and TCO can be realized and measured. As mentioned earlier, expensive desktop computing is often retired early only because the DAS associated with the asset is not capable of supporting the requirements of the application over time. In a centralized configuration, the DAS on each of the desktops and servers is retained only to warehouse the operating system and any other local required software tools. The centralization of home directories and generalized file sharing remove the application storage requirement from the desktop, thus extending the life of the desktop workstation and postponing replacement costs.

In addition, costly application servers often serve double duty on a daily basis —providing generalized file sharing storage while supporting production applications. By removing this storage and computing overhead from the production application server, the purchase of additional computing and storage resources is eliminated, and the effective life of the server is extended.

*The 5000 NAS appliance works in conjunction with industry-standard backup and recovery tools.*

*Centralizing storage offloads storage-related cycles from application servers, thus improving performance and increasing the effective life of hardware.*

Server-based DAS or SAN resources are often very expensive to manage. Placing home directory and general file server data on this resource is counterproductive and costly over time. Often, mission-critical applications require working storage to execute on a daily basis. When they cannot find this critical resource, wasted computing cycles and time result. Moving home directory data and general-purpose file sharing to the 5000 NAS appliance frees resources for critical application execution. This postpones the purchase of additional storage for the consolidated server infrastructure, providing real savings to the organization.

### **7.0 Conclusion**

NAS devices like the StorageTek 5000 Family of NAS Appliances allow you to centralize data, thereby reducing management costs and maximizing your investment in existing hardware.

Specifically, consolidating home directories allows for better data security and file sharing. Such centralization overcomes the problems inherent with distributed data management and introduces a range of sophisticated data management features to get the most out of your networking dollar.

**Sun Microsystems, Inc.** 4150 Network Circle, Santa Clara, CA 95054 USA **Phone** 1-650-960-1300 or 1-800-555-9SUN **Web** sun.com

© 2006 Sun Microsystems, Inc. All rights reserved. Sun, Sun Microsystems, the Sun logo, StorageTek, the StorageTek logo, Java, and Solaris are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries. UNIX is a registered trademark in the United States and other countries, exclusively licensed through X/Open Company, Ltd.

NT 0005 A 03/06

