

# Backup strategies for NAS and SAN

A distributed backup approach may provide advantages in network-attached storage (NAS) and storage area network (SAN) environments.

■ BY IRA GOODMAN and SOUBIR ACHARYA

**E**xciting breakthroughs in network storage technology are making it possible to keep pace with exploding data growth. Large tape libraries are now used with sophisticated strategies like multi-hosting. Powerful new architectures are also being implemented, such as network-attached storage (NAS) and storage area network (SAN). All of these innovations affect backup systems, leaving backup software

vendors working hard to adapt to changing technology and to make backup systems work as efficiently as possible in these new environments.

Years ago, the smart way to cope with storage requirements—which were comparatively small—was to use centralized backup software. Centralized backup organized all of a site’s low-capacity devices neatly in one place by attaching them to a single server. This allowed easy tape handling for operators and centralized administrative control.

However, as the amount of data on networks increased, restrictions imposed by centralized backup

	Centralized	Distributed
<b>Bandwidth</b>	Data must travel to a centralized location during backup, flooding the network with data.	Data can be backed up locally without traveling across the network.
<b>Performance</b>	Multiple storage devices centralized on a single SCSI device degrade in performance as they are added to the SCSI chain.	A single storage device can be directly attached to each server with an individual SCSI device for optimal performance.
<b>Mixed OS</b>	Different operating systems on a network require different versions of the backup product and multiple catalogs.	One version of the backup software and a single catalog handle multiple operating systems.
<b>Scheduling</b>	Streaming is constrained by the necessity of having data travel to a centralized location.	Multiple backups to directly connected distributed devices can take place simultaneously.

**FIGURE 1: Comparison of Centralized and Distributed Backup**

devices to be connected wherever needed on a network, control of these devices continued to be centralized.

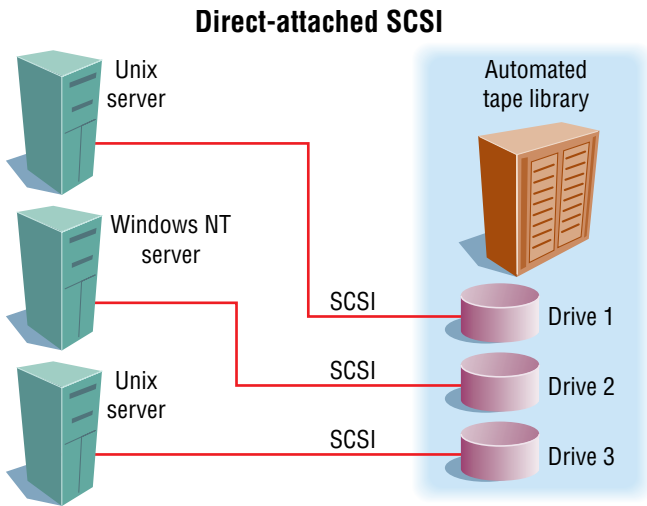
### Multi-hosting tape libraries

Because of its architectural flexibility, distributed backup software permits multi-hosting, a technique that can significantly improve tape library performance and reduce network bandwidth usage. And unlike more complex and costly strategies, such as SANs, multi-hosting uses technology that can be implemented now.

## Multi-hosting can significantly improve the performance of a tape library and reduce network bandwidth usage.

became serious obstacles. (See Figure 1.) As a result, a number of vendors developed backup and restore products designed with a distributed architecture. While this architecture solved bandwidth, performance, and scheduling problems by allowing storage

In a multi-hosting configuration, specific drives in a tape library use individual SCSI adapters to connect to specific servers (see Figure 2). The result of multi-hosting is the direct attachment of several servers to a single tape library. All



**FIGURE 2: In a multi-hosting configuration, specific drives in a tape library use individual SCSI adapters to connect to specific servers.**

of these servers can then stream data directly to the library, allowing faster backups while reducing network traffic. If a drive fails, the distributed backup software finds an alternative path.

However, with network growth rates of 50% to 100% per year and escalating demand for 7x24 access, automated libraries and multi-hosting are not enough. Fortunately, storage architectures are keeping pace, and distributed backup systems can adapt to the new architectures.

### NAS and NDMP

NAS devices are dedicated file server “appliances,” or “filers,” that provide fast access and high-availability storage to Unix and Windows/NT clients on a network. Because these devices are not connected in the traditional way to the network (e.g., through a server), the appliances need their own operating systems. And because a NAS appliance is neither a dumb device nor a server with complete operating system functionality, a special protocol called Network Data Management Protocol (NDMP) has been developed to allow backup software to communicate with these appliances.

When working with a variety of operating systems and storage devices, a distributed backup system often uses a universal protocol like TCP/IP for basic communications, adding new protocols like NDMP to make the backup system more versatile. During a backup operation, the software

users, is under discussion. System Independent Data Format (SIDF) is a likely candidate and is already used by a variety of backup software suites.

NDMP benefits both end users and vendors. Users can choose the best mix of hardware devices for their environment. Backup software vendors can concentrate more resources on enhancing their products, instead of incorporating additional protocols for the newest hardware.

### SAN backup

In retrospect, a logical progression seems obvious from distributed backup to “subnetting” to SAN technology. Distributed processing introduced the idea of unlimited flexibility into network backup. Different types of storage devices could easily be connected to a variety of servers running different operating systems. The kind of configuration that suited a site best could be chosen, while configuration could be centralized, distributed, or mixed. Of course, all backup operations were centrally controlled, regardless of configuration.

Subnetting added

sends NDMP commands to the filer, which in turn uses SCSI commands to move data to a backup device. If the storage device is in a tape library with multiple drives, a backup server—not the filer—controls the robotic arm. In a distributed backup system, any server can be set up to send the SCSI commands needed to move the robotic arm (see Figure 3).

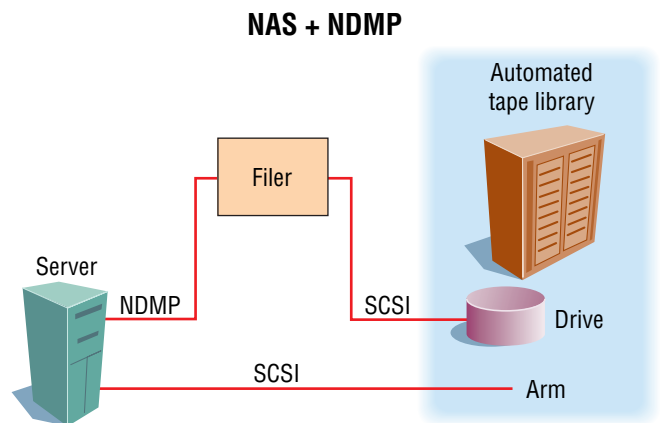
NDMP has been accepted as a standard, and Version 4, which deals with a common tape format for NDMP

the notion of physically segregating servers and storage devices through special hardware connections, allowing data to move within the subnet for backup without increasing traffic on the primary network. SAN technology works in a similar way, and has all of the advantages of subnetting and then some.

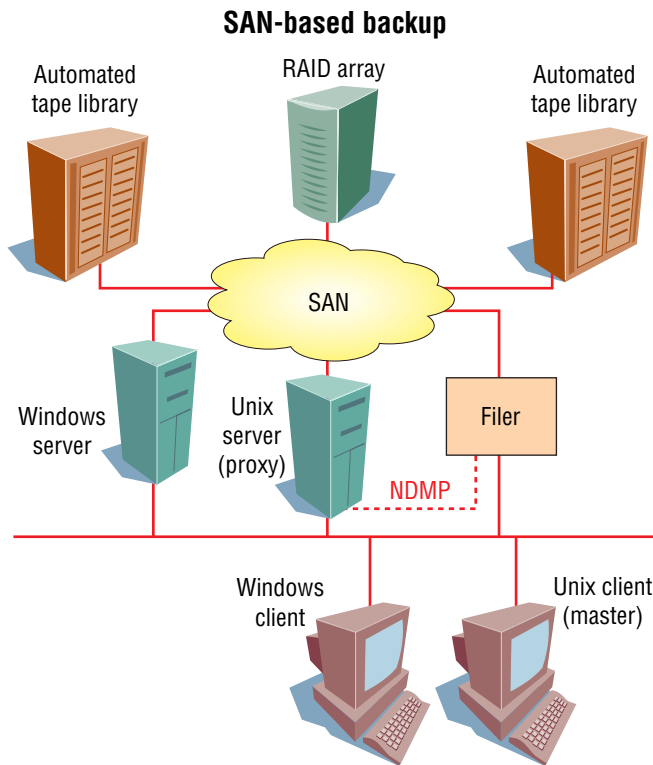
A SAN is a separate high-speed network to which shared storage is attached. SAN technology promises to allow network administrators to purchase multiple devices, and instead of attaching them to individual servers, connect them to all the servers on a SAN simultaneously and directly (see Figure 4).

In this example, two client machines, one of which is the backup “master” server holding the backup schedule and catalog, are connected to a LAN. Two servers and a filer are attached both to the LAN and to the SAN. The SAN has two tape libraries to which data from the clients, the servers, and the filer are backed up. Other types of devices for shared storage, such as RAID arrays, can also be attached to the SAN.

A Fibre Channel storage router would allow extended SCSI copy commands to be issued to back up data from disk drives (in Figure 4, the RAID array) to an automated library directly in what is often called (slightly inaccurately) “server-less backup.” (For more information, see “Server-less backup to take center stage.” *InfoStor*, March 2000, pp. 24-28.) The slight inaccuracy comes from the fact that messages to update the backup catalog must always be sent to the main backup server. However, this backup strategy works only on the partition level, not on the file level. Additional



**FIGURE 3: In a distributed backup system, any server can be set up to send the SCSI commands needed to move the robotic arm.**



**FIGURE 4:** In this example of a LAN-SAN-NAS setup with distributed backup, the Unix client system functions as the master backup server. It sends messages to the NAS filer through the proxy server, which then communicates via NDMP with the filer.

intelligence, most likely within a server, would be needed for file-level backup and restore in a server-less backup scenario.

The storage router has several functions. It sends the appropriate copy commands and translates the protocol for non-Fibre Channel (e.g., SCSI) devices, such as most tape libraries. Disk devices, such as RAID arrays, are often Fibre Channel ready.

### SAN and NAS

A filer must be handled in a special way during backup on a SAN (as illustrated in Figure 4). The filer cannot be addressed directly, but only through an NDMP proxy. This special proxy is usually an agent running on a backup device server, which controls the filer by issuing NDMP commands

and is in turn controlled by the master backup server.

In Figure 4, the Unix client machine is the master backup server. It sends messages to the filer through the proxy server, which then communicates via NDMP with the filer.

### Advantages of SAN

Once fully developed, SAN technology will provide many advantages, including redundancy and continuous access. For system administrators, it means a wealth of scheduling options for backup jobs. Need to do maintenance on Device A, where you usually store your backup images? Simply switch the

backup to Device B. Has Device C gone down? No problem. Do your backups to Device D today. Each server will have easy access to all the devices on the SAN as a common pool of storage, yet all storage-related activity will be managed centrally.

In addition, the use of SAN technology can shrink the time needed for backup, because the data transfer rate on a SAN is about 2.5x that of most SCSI devices (100MBps vs. 40MBps). And all of this high-speed data transfer from file servers to backup devices in a SAN takes place off the primary network, which is freed up for other applications.

However, SAN technology poses many new challenges for backup systems:

- **Device persistence**, or the relationship between devices and their addresses over time. SANs are designed for

dynamic storage configuration and for dynamic re-configuration when devices are added or taken offline for maintenance. Because backup software has traditionally been designed for static environments, it must be adapted to work under dynamic conditions. A backup system for SANs must be able to discover and configure devices intelligently, and automatically re-discover and re-configure them whenever conditions change.

- **Automatically routing backups** to the best mix of devices for optimal performance so that if a device fails or a path is unavailable, a backup system makes decisions on the fly about the best alternative routes. Dynamic routing has always been available in distributed backup.
- **Mechanism for lockout in a backup system.** Allocation must now be done in an environment of dynamic sharing. How will a backup system handle reserve and release arbitration during backup operations? Since it is designed to dynamically re-route, a distributed backup system may have advantages in this area.
- **Error handling.** Although errors are comparatively rare in a SAN environment that uses Fibre Channel technology, errors do happen. Which retry strategy will a backup system use to handle transient errors during a backup operation? Intelligent retry (how long and how often to retry) in a dynamic SAN environment is extremely important.

With the advent of SAN technology, the future of network (and backup) processing seems virtually unlimited. An architecture that allows direct, high-speed access to storage resources that are available, but not restricted, to any server on a network can satisfy the need for limitless growth and constant access. □

**Ira Goodman** is software services manager and **Soubir Acharya** is senior technical architecture at Syncsort Inc. in Woodcliff Lake, NJ. [www.syncsort.com](http://www.syncsort.com).



50 Tice Boulevard  
Woodcliff Lake, NJ 07677  
(201) 930-8200  
<http://www.syncsort.com>