



NAS Device Backup Solutions
Network Data Management Protocol (NDMP)
White Paper

Paul Mayer
Product Manager

Revision 2
April 2002



Table of Contents

Introduction	1
Network data management protocol (NDMP)	3
NAS backup configurations	6
NAS device-to-server NFS or CIFS configuration	7
Server-to-NAS device two-way NDMP configuration	8
NAS device-to-NAS device three way NDMP configuration	9
NAS device-to-server NDMP configuration	10
Replicated NAS device two-way backup configuration	11
Server-to-NAS device NDMP configuration	12
NAS device-to-network-attached tape library NDMP configuration	13
NAS device-to-SAN attached tape library configuration	14
NDMP limitations	15
Summary	16
Who supports NDMP?	17
Terms and definitions	18
Additional information	20



Introduction

Data generation, storage and new technology

The explosion of data generation and storage in recent years has led to the introduction of new technologies for simplifying the storage and management of this data. One of the on-line storage architectures that has emerged is Network Attached Storage (NAS), which essentially separates application servers and data, and stores the data on storage devices that perform dedicated file serving tasks. A NAS device is typically a dedicated, high-performance, high-speed communicating single-purpose machine or component. The benefits offered by NAS technology include:

- Improved performance based upon dedicated file service processing
- Ease of administration—the ability to serve a common set of files to a heterogeneous client population
- Ease of capacity expansion
- Ease of installation

Since the introduction of the NAS device concept to the marketplace in 1992, the technology has gained widespread acceptance, and many major storage manufacturers have added NAS devices to their product offerings.

The backup challenge

At the time of its introduction, the NAS architecture challenged the then existing data backup and restoration practices, which utilized a software agent running on each server that hosted data. In NAS environments, the NAS device hosts an operating system that is dedicated to the task of serving files, and to maintain its operational efficiency it does not readily accommodate the hosting of third party applications such as backup software agents and clients. This specialization forced backup administrators to mount each NAS volume via Common Interface File System (CIFS) or Network File System (NFS) from another server across the network, which hosted a backup software agent. This approach leads to performance degradation for backup operations due to overhead.

Continued on next page



Introduction, continued

Meeting the NAS challenge

In response to this challenge, Network Appliance and Legato (*then Intelliguard*) collaborated to develop the Network Data Management Protocol (NDMP). This protocol is intended to provide localized file movement capabilities on a NAS device, which can be leveraged for data backups and disaster recovery. The NDMP standard is now in its fourth generation, and is currently being maintained by a working group within the Storage Network Industry Association (SNIA). Figure 1 shows the NDMP development process timeline.

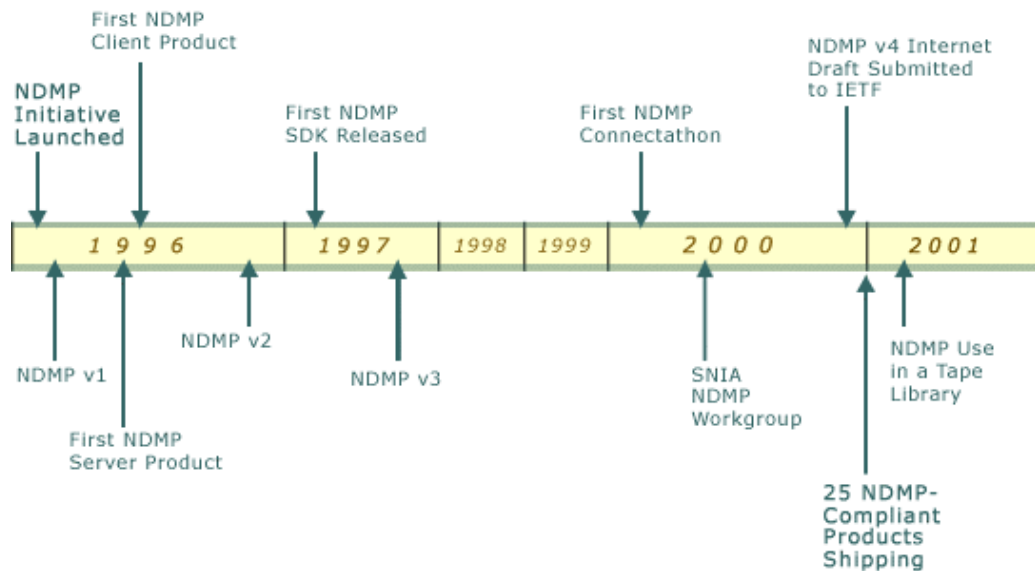


Figure 1: NDMP Development Timeline

Document purpose

This document presents an overview of NDMP, its function, features, benefits, and limitations. At the end of this document you will find the following:

- Who supports NDMP?
- Terms and definitions
- Where to find additional information about NDMP



Network data management protocol (NDMP)

What is NDMP NDMP is an open-standard protocol for conducting data backups and restores of heterogeneous network-attached storage devices. In this way file-system data is copied from the file system to the backup device using a common interface, regardless of the platform or device. Likewise, to control file metadata, common commands pass to and from the backup software, regardless of the software application being used.

NDMP design goal The NDMP design addresses the complexities of heterogeneous networks by creating a “common agent” used for centralized backup software and file servers. NDMP provides a common interface between any backup software application and any network-attached storage device. This enables backup software vendors to support a wide variety of network-attached storage devices without having to redesign their backup software applications. It also enables network-attached storage vendors to work seamlessly with any other NDMP-compliant application. This plug-and-play approach lets the administrator backup data throughout the enterprise, using a combination of NDMP-compliant network-attached servers, backup devices, and backup software applications.

NDMP and vendor proprietary device-specific code NDMP provides backup software vendors the ability to offer backup and restore capabilities on a NAS device without writing proprietary, device-specific code for NAS devices. The NDMP protocol runs natively on the NAS device, and includes a set of commands that can be called from any third party application running on another server.

Continued on next page



Network data management protocol (NDMP), continued

How NDMP functions

The NDMP protocol can be thought of as a dedicated application that runs on the NAS device, waiting for requests from backup software products on different file servers, running different platforms and platform versions; thus, creating a thin layer across a network to make easy data backup and restore operations. It provides several lower level services that can be grouped into two categories:

1. As a data server, it reads from a disk and produces an NDMP data stream (*in a specified format*) or reads an NDMP data stream and writes to disk—when the operation is either a backup or restore.
2. As a media server, NDMP reads the data stream and writes it to media or reads from the media and writes a NDMP data stream—when the operation is either a backup or restore. All media handling functions, such as split-image issues, and robotics controls, are handled by this service category.

The backup software uses NDMP to send requests to the NAS device, which initiates a backup or restore process. During a backup operation, information about the backed up files is sent from the NAS device to the backup software running on the backup server. The actual files are sent from the NAS device to one of three possible locations:

- A tape device directly attached to a NAS device (*direct-attached*)
 - A tape device directly attached to another NAS device of the same brand (*NAS device-to-NAS device*)
 - A tape device attached to a backup server (*NAS device-to-server*)
-

DAR, index backups

One of the shortcomings in earlier versions of NDMP was its inability to efficiently restore small sets of files upon request. It was designed, rather, to accommodate the restoration of entire backup sessions. Individual or small sets of files could be restored, but the operation required the real-time reading of the tape content at the time of restoration, which could take an inordinate amount of time. In the worst-case scenario, the file would be located at the end of the last tape in a backup session. In this case, the entire set of tapes would have to be read in order to access and restore the searched for file.

Direct Access Recovery (DAR) was incorporated into NDMP to resolve this issue. DAR provides the ability to index backup sessions at a file level, recording the tape volume and location of each file in a given backup session. During a single file recovery operation, this allows the appropriate tape to be loaded, and quickly spun directly to the precise location of the file to be recovered. DAR is part of the NDMP standard, and can be utilized by any software vendor that chooses to leverage this capability.

Continued on next page



Network data management protocol (NDMP), continued

NDMP backup process

The process of backing up a NAS device, using NDMP involves the following:

Stage	Description
1.	The backup server, which schedules the backup and determines when the backup will take place, signals the NDMP data service.
2.	The NDMP data service responds that it is ready to do the backup and provides the backup server with the location on the network where the data will be backed up (<i>IP address</i>).
3.	The backup server then scans for the first available tape drive(s) and makes note of their IP address and port. It also requests an available media cartridge and loads it in the drive.

Note: NDMP has a tape interface that allows an NDMP client to perform tasks such as positioning the tape, and reading and writing tape labels.

4.	The backup server then contacts the data service and gives it the IP address and port of the tape drive (<i>NDMP tape service</i>). After the host and server are connected, the backup server is released from the backup session, and then the NDMP host starts transferring the data.
----	--

Note: The NDMP data server produces an NDMP data stream that the NDMP tape server writes directly to tape. Controlling this stream of data and control characters manages the format on the tape—no backup software proprietary code is required.

5.	After the backup operation is completed, the session between the NDMP data service and tape service ends; the backup server is provided metadata detailing the results of the backup.
----	---



NAS backup configurations

Choosing a NAS backup configuration

This document highlights eight general configurations used to back up data stored on NAS devices, each one having advantages and disadvantages. The eight backup configurations are:

1. NAS device-to-server NFS or CIFS
2. Server-to-NAS device two-way NDMP
3. NAS device-to-NAS device three-way NDMP
4. NAS device-to-server NDMP
5. Replicated NAS device two-way NDMP
6. Server-to- NAS device NDMP
7. NAS device-to-network attached tape library three-way NDMP
8. NAS device-to-SAN attached tape library

It is the role of a storage architect to determine which configuration would provide the optimal solution in a given environment. The configuration of choice should be dictated by the customer's:

- Volume of NAS-hosted data
 - Volume of server-hosted data
 - Pattern and frequency of backups and restores
 - Backup window
 - Restore window
 - Network infrastructure
 - SAN infrastructure
 - Budget
-

NAS device-to-server NFS or CIFS configuration

NAS device-to-server NFS or CIFS

In this configuration, standard network drive-mounting technologies are used to capture the backed up data. Prior to the introduction of NDMP, this methodology was the predominant approach to NAS backup. Data streams are sent across the network to the server where they are transferred to the locally attached tape library. The data set is prepared for backup by the backup server rather than by the NDMP protocol running on the NAS device. See Figure 2.

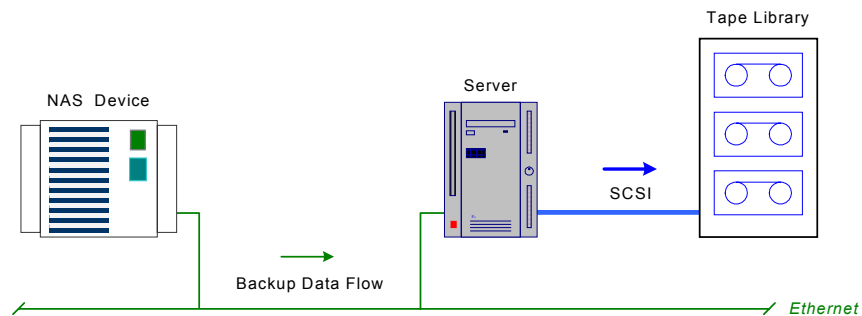


Figure 2 NAS Device-to-Server NFS or CIFS Configuration

Configuration advantages and disadvantages

NAS device-to-server NFS or CIFS configuration:

ADVANTAGES

- Provides rudimentary backup capability without requiring additional NDMP software modules
- Typically lowest cost NAS backup solution

DISADVANTAGES

- Generally not recommended by most backup software vendors due to performance and reliability issues
- Less than optimal performance due to CIFS and NFS overheads
- Inability to consistently retain file permissions upon backup
- Generates network traffic on production LAN unless backups are isolated on a dedicated network

Server-to-NAS device two-way NDMP configuration

Server-to-NAS device two-way NDMP

In this configuration, a tape library is attached directly to each NAS device, and the data that resides on the NAS device is backed up directly to the tape library. See Figure 3. The request for backing up the data set comes from the backup server, which is acting as an NDMP client. This was the original method of deployment for NDMP backups. Its primary purpose was to eliminate the transfer of data across the Ethernet medium by providing support for direct-attached tape devices.

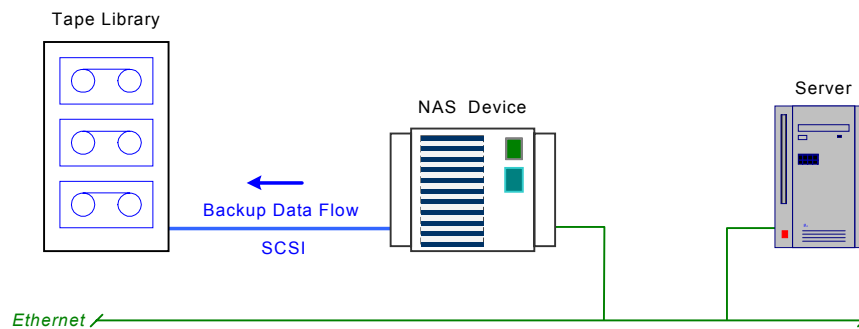


Figure 3: Server-to-NAS Device Two-Way NDMP Configuration

Configuration advantages and disadvantages

Server-to-NAS device two-way (*direct-attached*) NDMP configuration:

ADVANTAGES

- Minimizes network traffic by isolating data movement from the NAS device to the locally attached tape library
- Only control information and metadata is transferred over the network to a backup server
- Backup and restore speeds are optimized due to the high-speed SCSI connection between the tape library and NAS device
- Allows preservation of file permissions on both NFS and CIFS volumes

DISADVANTAGES

- Decentralized tape resources increase the complexity of tape administration
- Requires configuring a dedicated tape device for each NAS device
- Data transfer speeds are inferior to traditional server backups due to the current inability of NDMP to generate parallel data streams

NAS device-to-NAS device three way NDMP configuration

NAS device-to-NAS device three way NDMP

In this configuration, the data from one NAS device is backed up to the tape library directly attached to another NAS device. See Figure 4. The backup server maintains the backup schedule for all NAS devices; when it is time to begin a backup session it coordinates the activity between the NAS devices. Once the backup session is initiated, the backup server is released from the backup session, and the data transfer occurs between the two NAS devices—ultimately directed to the tape library. Upon completing the backup session, the metadata is sent to the backup server, which updates the index with information pertaining to the backup session.

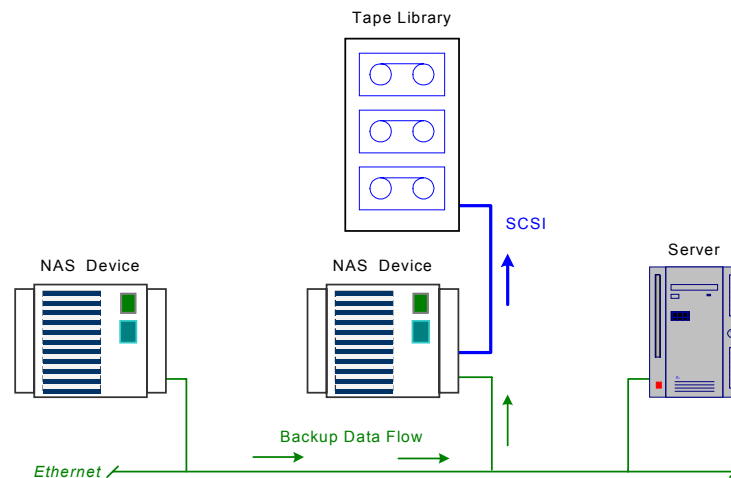


Figure 4: NAS Device-to-NAS Device Three Way NDMP Configuration

Configuration advantages and disadvantages

NAS device-to-NAS device three-way NDMP configuration:

ADVANTAGES

- Leverages investment in tape resources by sharing among multiple NAS devices
- Allows preservation of file permissions on both NFS and CIFS volumes
- Most NAS devices allow for a dedicated network interface card, which can isolate the backup traffic to a dedicated network segment, preventing saturation of the production network during backups

DISADVANTAGES

- Can generate network traffic on production LAN unless backups are isolated on a dedicated network
- Consumes CPU resources on both NAS devices during backup
- Data transfer speeds are inferior to traditional server backups due to the current inability of NDMP to generate parallel data streams

NAS device-to-server NDMP configuration

NAS device-to-server NDMP

In this configuration, the backup server sends a request to the NAS device for a data stream to send to the tape library. See Figure 5. The backup server acts as an NDMP client when it requests the backup, and then as an NDMP tape server as it receives the data stream from the NAS device. This configuration is suitable in environments where servers manage the majority of data with a relatively small amount remaining on the NAS device, and a sizeable investment has been made in a server-managed tape library. All vendors do not support this approach.

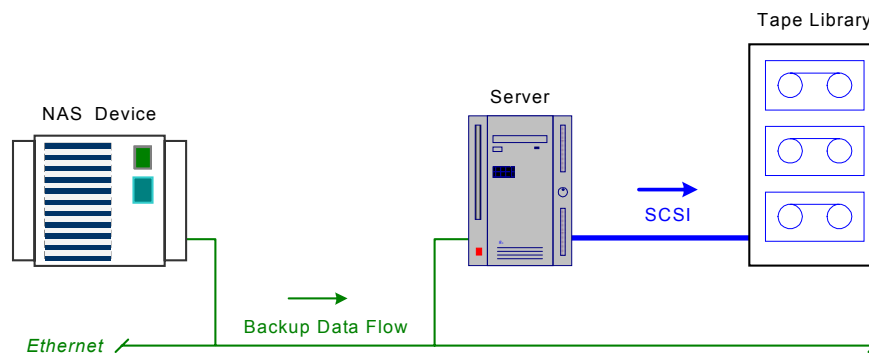


Figure 5: NAS Device-to-Server–NDMP Configuration

Configuration advantages and disadvantages

NAS device-to-server NDMP configuration:

ADVANTAGES

- Provides centralized backup capability
- Provides centralized tape administration

DISADVANTAGES

- Requires additional software on the backup server to permit the server to act as an NDMP tape service
- Can generate network traffic on production LAN unless backups are isolated on a dedicated network
- Data transfer speeds are inferior to traditional server backups due to the current inability of NDMP to generate parallel data streams

Replicated NAS device two-way backup configuration

Replicated NAS device two-way NDMP

In this configuration, the mirror relationship between NAS devices is broken temporarily, after which a backup takes place on the secondary NAS device. Meanwhile, the primary NAS device continues its production operation. Upon completion of a successful backup, the mirror is reestablished and the data is returned to a synchronous state. See Figure 6. Many NAS devices provide the ability to make a mirrored copy of data on another NAS device. The data is then available for several different functions, one of which is an off-host backup.

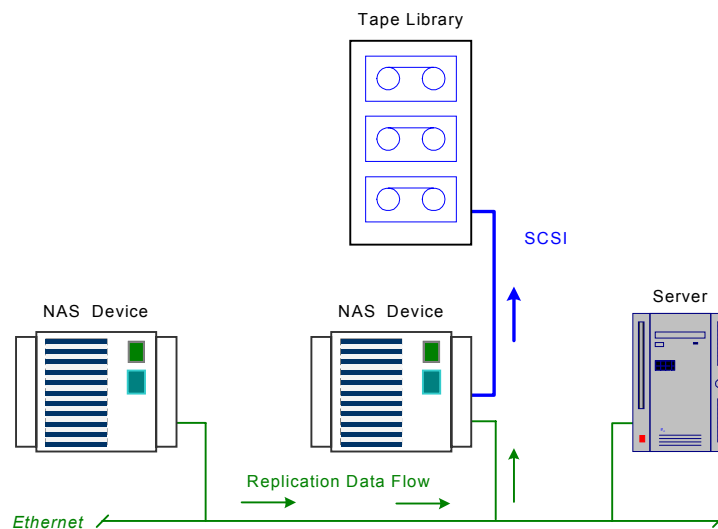


Figure 6: Replicated NAS Device Backup Configuration

Configuration advantages and disadvantages

Replicated NAS device backup configuration:

ADVANTAGES

- Provides backup with minimal impact on production performance
- Mirrored data provides fast recovery from primary NAS device physical failures
- Leverages investment in the disaster recovery infrastructure
- Allows preservation for file permissions on both NFS and CIFS volumes

DISADVANTAGES

- Requires additional replication software on the NAS device in many cases
- Requires additional expense for duplicate NAS storage hardware
- Replication data flow generates network traffic on the production LAN unless replication activity is isolated on a dedicated network

Server-to-NAS device NDMP configuration

Server-to-NAS device NDMP

In this configuration, the tape library is directly attached to the NAS device; data streams are transferred from the server to the NAS device and backed up to the direct-attached tape library. See Figure 7. This is not a method for backing up a NAS device, but rather a way to leverage an investment in NAS backup technology. Under this scenario, the server acts as an NDMP data server, and the NAS device hosts the NDMP tape server. Although this configuration is not common, it might prove to be a good alternative if the majority of data is hosted on a NAS device, while a relatively small amount of server-hosted data requires backup. In this case, the tape resource attached to the NAS device is leveraged, eliminating the need to add additional tape storage to the server. All vendors do not support this approach.

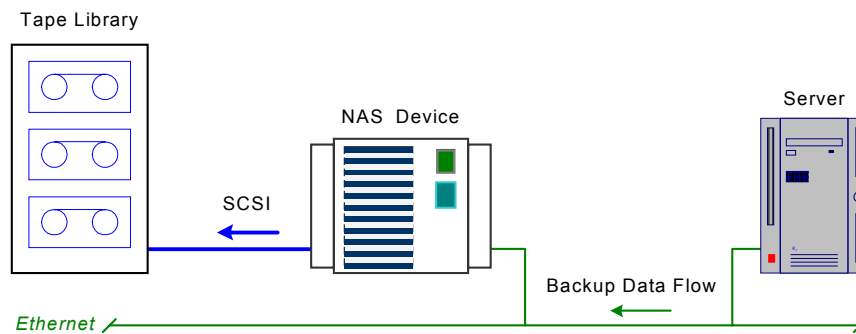


Figure 7: Server-to NAS Device NDMP Configuration

Configuration advantages and disadvantages

Server-to-NAS device NDMP configuration:

ADVANTAGES

- Leverages investment in the NAS device attached library by using it for backing up servers
- Centralizes tape administration
- Allows preservation of file permissions on both NFS and CIFS volumes

DISADVANTAGES

- NDMP does not currently support data multistreaming, so performance is not as robust as with standard server-attached or SAN backups
- Generates some network traffic when data is transferred from server(s) to the NAS device
- Requires additional software on the data server to allow the server to act as an NDMP data service

NAS device-to-network-attached tape library NDMP configuration

NAS device-to-network-attached tape library three-way NDMP

In this configuration, the tape library is directly attached to the network, rather than to a second NAS device. See Figure 8. The tape library houses an on-board NDMP tape server, which is capable of receiving an NDMP data stream from an NDMP data server. This approach allows multiple NAS devices to share a single tape library, without burdening a single NAS device with the backup I/O from other NAS devices.

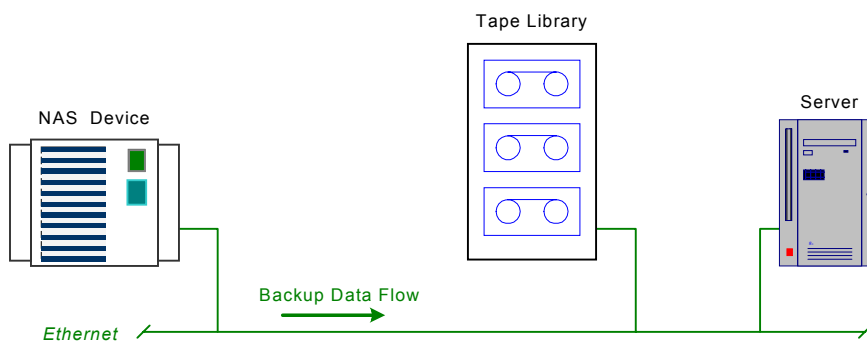


Figure 8: NAS Device-to-Network-Attached Tape Library NDMP Configuration

Configuration advantages and disadvantages

NAS device-to-network-attached tape library, three-way NDMP configuration:

ADVANTAGES

- Leverages investment in tape resources by sharing them among multiple NAS devices
- Does not bottleneck a second NAS device with backup traffic
- Centralizes tape administration
- Eliminates the need for a second NAS device as a point of failure for backup operations
- Allows preservation of file permissions on both NFS and CIFS volumes

DISADVANTAGES

- Generates network traffic unless backups are isolated on a dedicated network
- Can incur additional costs for network interface and NDMP support

NAS device-to-SAN attached tape library configuration

NAS device-to-SAN attached tape library

In this configuration, a Fibre channel HBA is added to the NAS device and is connected to a SAN for backup purposes. See Figure 9. The NAS device has the ability to utilize shared tape resources in the SAN, without having to transfer its backup data stream across the LAN. This configuration still does not provide multi-streaming capabilities, which is common with standard open-architecture servers. This is due to the current limitations of the NDMP version 3 standard, but look for advances in this capability in the future.

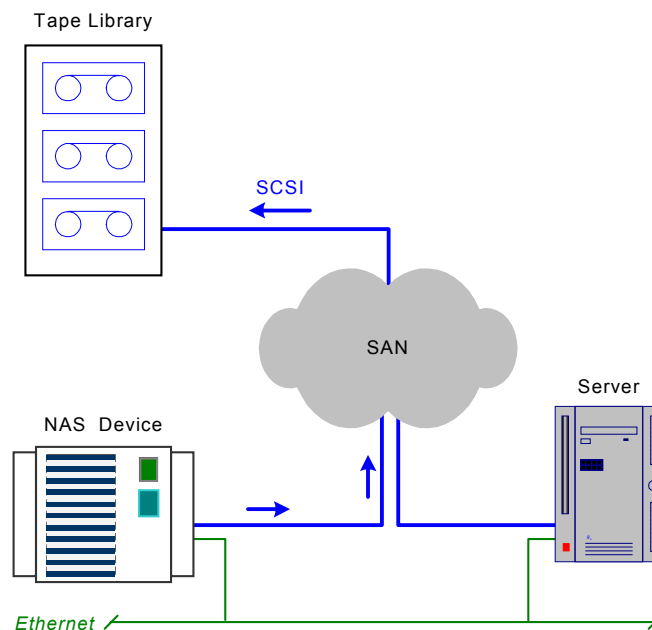


Figure 9: NAS Device-to-SAN Attached Tape Library configuration

Configuration advantages and disadvantages

NAS device-to-SAN attached tape device configuration:

ADVANTAGES

- Leverages investment in tape resources by sharing among multiple NAS devices
- Does not bottleneck a second NAS device or server with backup traffic
- Provides high-speed connectivity to secondary storage
- Eliminates the need for a second NAS device as a point of failure for backup operations
- Allows preservation of file permissions on both NFS and CIFS volumes

DISADVANTAGE

- Additional costs incurred for HBAs and switch ports



NDMP limitations

What is missing from NDMP?

NDMP has certainly filled a critical void in NAS environments by providing a vendor-independent standard method for backing up and restoring data on NAS devices. It is not without its limitations, however. Relative to traditional server-based backups in non-NDMP environments, NDMP currently imposes the following limitations:

- No multiplexing capability – In a traditional server environment, performance scalability can be achieved by sending multiple data streams simultaneously to a single tape device, or sending a single data stream to multiple tape devices. The current NDMP standard does not provide this capability, which can be a considerable hindrance as NAS device capacity continues to grow.
- No checkpointing capability – Session management in traditional backup schemes allow the placement of checkpoints in a backup session that facilitate quick recovery of a failed backup session by resuming operation at the checkpoint rather than starting from the beginning. This feature is not available in the current NDMP standard, making recovery from a backup failure a much slower process.
- No compression capability – The current NDMP standard does not allow data compression during a backup session. This can cause inefficient tape utilization and inferior performance compared to traditional backups.
- No generalized interface-to-vendor-specific snapshot capabilities – Many NAS vendors provide proprietary utilities for making a point-in-time copy of a resident file system. This capability can be a valuable component of an effective backup system, but the current NDMP specification does not provide a common interface for these utilities.

Future releases of the NDMP standard are expected to address these areas.



Summary

Network attached storage

A key element of implementing a complete NAS solution is configuring an appropriate combination of data protection technologies that complement the simplicity and performance of NAS. The NAS backup solutions discussed in this document presented solutions of varying degrees of performance, production impact, restoration capabilities, and preservation of file system integrity. It is the role of the storage system architect to examine the organization's business demands, which should ultimately dictate its data protection requirements. Then those requirements should be configured with the appropriate technologies to architect the optimal solution for the organization's environment.



Who supports NDMP?

Companies supporting NDMP

There is a large and growing number of diverse companies that support the NDMP standard, such as backup software providers, NAS hardware providers, tape hardware providers, messaging device providers, and NAS OS providers. The following is a partial list of NDMP supporting companies and their product offerings.

Backup Software Providers

Company	Product
Atempo	TimeNavigator for NDMP
Backbone Software	NetVault
Computer Associates	BrightStor Enterprise Backup
CommVault	Galaxy
Legato Systems	Celestra Power NetWorker SnapImage Module
SyncSort, Inc.	Backup Express
Veritas	NetBackup
Workstation Solutions	QuickRestore

NAS Hardware Providers

Advance Media Services	Dakota RAID Pro
Aspex Systems	NS2000
BlueArc	Si7500
Comex	N3100
ECCS	Synchronecton II
EMC	Celerra File Server, IP4700
Network Appliance	F85, F7xx, and F8xx Filers
Network Engines	StorageEngine
Network Storage Solutions	Thunderbolt, NASengine, SPANStor GT
Procom Technology	Netforce
R.A.R.E. Systems	Razor NAS
StorLogic	StorCom 4000
VA Linux	9205 NAS

NAS OS Providers

Land-5	IceNAS Software
Network Appliance	Data ONTAP

Tape hardware Providers

Quantum/ATL	Tape Libraries
StorageTek	Tape Libraries
Spectra Logic	Tape Libraries
ADIC	Tape Libraries

Messaging Device Providers

Mirapoint	Internet Message Server
-----------	-------------------------



Terms and definitions

Terms	Definitions
Backup Data	The resulting data from a backup operation.
Backup or Backup Operation	Copying selected data from primary storage to secondary storage.
Checkpoint	A placeholder positioned in a backup session to mark a recovery point in the event of a failure during the backup process. In the event of a failure, the backup can resume at the checkpoint rather than restarting the entire session. This capability is currently not supported in NDMP version 3.
Direct Access Recovery (DAR)	An optional capability of NDMP data and tape services whereby only relevant portions of secondary media are accessed during recovery operations. This capability greatly enhances restore performance for individual or small file sets.
Data Service	A NDMP Service, which transfers data between primary storage and the data connection. A primary storage system hosts an NDMP data service.
Data stream	A unidirectional byte stream of data flowing over a data connection between two-peer NDMP services in an NDMP session. During a backup, the data stream is generated by the data service and consumed by the tape service. The data stream can be backup data, recovered data, data to be replicated, etc.
Data Management Application (DMA)	The software application that controls the NDMP session. In NDMP there is a master/slave relationship, the DMA is the session master and the NDMP services are the slaves. In NDMP versions 1, 2 and 3 the term "NDMP client" was used instead of the DMA.
Mover	An aspect of the tape service that transfers data between the secondary storage and the data connection.
Multiplexing	The process of simultaneously writing multiple data streams to a single tape device, or the process of writing one data stream over multiple tape devices for the purpose of performance enhancement.
Multistreaming	The process of routing multiple sets of data to a single backup server simultaneously for the purpose of achieving performance enhancement.
NDMP Client	An entity that requests an NDMP service. The backup server generally plays to role of the NDMP client because it makes a request of the NDMP tape service to get a data stream from the NDMP data service.
NDMP Host	The host computer system that executes the NDMP server application. Data is backed up from the NDMP host to either a local tape drive or to a backup device on a remote NDMP host.

Continued on next page


Terms and definitions, continued

Terms	Definitions
NDMP Server	An instance of one or more distinct NDMP services controlled by a single NDMP control connection. Thus a data/tape/SCSI server is an NDMP server providing a data/tape/SCSI service.
NDMP Service	An operating utility controlled by NDMP that can be used by an NDMP client. There are three types of NDMP Services: Data Service, Tape Service, and SCSI Service.
NDMP Session	The configuration of one DMA and two NDMP services to perform a data management operation, such as a backup or a recovery.
Primary Storage System	A storage system that stores live or in production data. Examples are direct or SAN attached storage in application servers, or dedicated storage devices such as NAS devices.
Recovery data	The resulting data from a recovery operation.
Recovery or Recovery Operation	Copying selected data from secondary storage to primary storage.
Replication	The copying of data between two services of the same type. Examples are data-to-data service replication and tape-to-tape service replication.
Replication data	The resulting data from a replication operation.
SCSI Service	An NDMP service which passes low level SCSI commands to a SCSI device typically used by the DMA to manipulate a SCSI media changer.
Secondary Storage System	A storage system used for archiving or data protection. Examples are application servers with direct attached tape drives, libraries or robots, and dedicated network attached archiving/data protection devices. A secondary storage system hosts an NDMP tape service and often a SCSI service.
Tape Service	A NDMP Service transfers data between secondary storage and the data connection and allows the DMA to manipulate and access secondary storage.



Additional information

For more information

If you would like to know more about NDMP, a wealth of additional information is available from the following sources:

NDMP Standards Organization

<http://www.ndmp.org>

NDMP Support in Networker 6

<http://www.legato.com/resources/whitepapers/W054.pdf>

INFOSTOR: Tape Backup Strategies for NAS

http://is.pennnet.com/Articles/Article_Display.cfm?&Section=Articles&SubSection=Display&ARTICLE_ID=88071&PUBLICATION_ID=23&VERSION_NUM=1

Data Protection Strategies for Network Appliance Filers

http://www.netapp.com/tech_library/3066.html

Backup Solutions for the Clarion IP4700

http://www.emc.com/pdf/products/clarion/backup_solutions_wp.pdf

Using a Network-Attached Enterprise Tape Library with NDMP

<http://www.quantumatl.com/Products/tech/WP00005.pdf>

Datalink Corporation
8170 Upland Circle
Chanhassen, MN, 55317
www.datalink.com

Datalink Corporation is an information storage architect. The Company analyzes, designs, implements, and supports information storage infrastructures that store, protect, and provide continuous access to information. Datalink's specialized capabilities and solutions span Storage Area Networks, Network-Attached Storage, Direct-Attached Storage, and IP-based Storage, using industry-leading hardware, software, and technical services.