EMC²
where information lives®

White Paper

EMC Data Domain Storage System with EMC NetWorker

Best Practices Planning

Abstract

This white paper contains a compilation of specific configuration and best practices information for the EMC® Data Domain® Storage System with EMC NetWorker®.

March 2010

Part Number h6984, 300-010-619 A02

Table of Contents

Executive summary

The EMC® Data Domain® Storage System appliances provide simple and reliable disk-based backup and recovery systems with the added feature of deduplication. The Data Domain products with deduplication integrate seamlessly into your current backup environment to provide leading-edge backup and restore operations. The Data Domain Storage System was designed and implemented for optimal performance and ease of use, and easily scales to meet your storage needs.

Introduction

This white paper provides specific configuration and best practices recommendations for the Data Domain Storage System appliances when used with EMC NetWorker® in NAS and SAN environments.

Audience

This white paper is intended for EMC customers, system engineers, and members of the EMC and partners professional services community who are interested in configuration and best practices information when using the Data Domain appliances with NetWorker.

Terminology

- **Backup-to-disk (B2D)** – A backup solution where data is written to hard disk instead of tape.
- **CIFS** – Protocol for a Windows-based network fileshare.
- **Cloned storage node** – A computer with NetWorker software installed whose primary purpose is to make copies of save sets.
- **Data Domain Storage System** – A target for storing backups in deduplicated form.
- **Data Domain Replicator** – A Data Domain software option that provides network-efficient, automated, replication for disaster recovery.
- **Data Protection Advisor (DPA)** – An EMC data protection management solution that provides automatic and continuous data collection, conditional analysis that triggers alerts, and a single, consistent interface for reporting.
- **Deduplication** – Process of detecting and identifying the redundant variable-length blocks (or data segments) within a given set of data to eliminate redundancy.
- **Failback operation** - The replicated NAS share or VTL is transferred from the target system to the source system.
- **Logical Unit Number (LUN)** – Identifying number of a SCSI object that processes SCSI commands.
- **Multiplexing** – The interleaving of data on a target, required to make concurrency/parallelism technically realizable on anything but an Advanced File Type Device (AFTD). See *parallelism*.
- **Metadata description** – Sequence of unique blocks and pointers that represent the original data. The receipt of this metadata description by the target system (and the data that the metadata describes) results in the generation of a recoverable image.
- **NAS** – Network-attached storage.
- **NFS** – Protocol for cross-platform network fileshares.
- **Parallelism** – The generic term for concurrency in NetWorker, from the name of the client, server, and group settings controlling their respective concurrency. See *multiplexing*.
- **Partial replication –** The recovery point representing that a VTL or NAS share or readable image representing a cartridge or directory/file on the target system is missing data present in its source equivalent.
- **Recovery operation** – When the replicated NAS share or VTL (recovery point) is made available for use on the target Data Domain Storage System.

- **Recovery point** – A replicated NAS share or VTL on the target system whose metatdata description has also been replicated. A replicated virtual tape or directory/file on the target system.
- **Remote replication** – When backup data residing on a Data Domain Storage System is copied over a LAN or WAN to another Data Domain Storage System for disaster recovery protection.
- **Save set** – A group of files or a file system that has been backed up on storage media.
- **Shoeshining** – An interruption in the data stream to a physical tape drive that requires the tape to be repositioned on the drive's head in a manner that causes a stop-start or shoeshine motion of the tape device. This behavior is due to an inadequate data flow to the tape drive and slows overall performance. See http://en.wikipedia.org/wiki/Tape_drive.
- **Source-based deduplication** – Deduplication of backup data occurs at the client before it is sent to the primary storage system.
- **Staging** – The movement (or migration) of a save set versus. the copying of a save set. Staging is typically associated with the more generic backup-to-disk approach using Advanced File Type Devices in NetWorker – where once a save set is staged, the original copy is deleted.
- **Storage node** – A computer with NetWorker software installed whose primary purpose is to receive a backup stream from one or more backup clients, and then write to save sets on storage media.
- **Target-based deduplication** – When deduplication of backup data occurs at the target storage system.
- **Virtual tape library (VTL)** – Software emulation of a physical tape library system.

Data Domain Storage System overview

The Data Domain Storage Systems are stand-alone backup-to-disk (B2D) appliances offering a network-attached storage (NAS) front end (over IP) and an optional virtual tape library (VTL) front end (over FC). Both configurations of Data Domain Storage System include data deduplication to efficiently store data written to the appliance. These appliances can be configured to present an NFS share or CIFS share, in NAS personality, and/or a virtual tape library, in VTL personality, to the backup storage node. When enabled, both NAS backup to disk (B2D) and VTL functionality can be used simultaneously. The Data Domain Storage Systems are 100% inline deduplication appliances, which means that all data being sent is immediately deduplicated.

The Data Domain Storage System appliance with its data deduplication capability does the following:

- Eliminates redundant data from backups to reduce storage, enabling longer onsite retention and reduced replication costs.
- Performs sub-file comparisons on variable length data blocks to capture small block inserts and overstrikes in unstructured and structured data.
- Provides very high-performing inline deduplication.
- Includes built-in data compression that is additive to deduplication in the data reduction process

The Data Domain replication option leverages its deduplication capability, thus substantially reducing the amount of backup data that needs to be migrated to a remote site. Replication provides rapid local and remote restores with the following benefits:

- Supports replication between any combination of Data Domain Storage System appliances.
- Permits bi-directional replication between appliances.
- Permits cascaded replication (A->B->C) between appliances.
- Replicates deduplicated NAS shares and/or virtual tape pools to reduce bandwidth.
- Checks for the presence of each data block before replication so that only unique data blocks are sent to the target Data Domain Storage System to further reduce network traffic.
- Supports up to 180 source Data Domain Storage Systems to one target.
- Maintains a common data deduplication repository at the target for maximum storage savings.

- Replication of tape pools within VTLs or NAS shares/directories.
- Detailed replication reporting.
- Replicated data is immediately recoverable from the replica system.

The deduplication process

The deduplication process divides the data into variable length blocks and calculates a signature for each block. For each signature, the deduplication process determines if any other identical signature already exists, and if not, hardware-compresses the unique block and writes it to its data pool. If it finds any matches to these signatures, it replaces that redundant block with a pointer to the unique block so that only one copy of that data block exists on disk.

As additional data is written to the Data Domain Storage System, its data is also broken down into variable length blocks whose signatures are compared against those of existing unique blocks. Only new unique blocks are stored along with pointers to existing unique blocks in the appropriate sequence that can be used to reconstitute the data.

Key factors affecting the deduplication ratio

The ratio of the amount of storage space required to store the total number of save sets in a conventional disk storage system compared to the storage capacity used in a deduplication system is referred to as the deduplication ratio. There are many factors that affect deduplication ratios. Some key factors are discussed in this section.

Backup policies

Retaining data for longer periods of time improves the chance that common data will already exist in storage, resulting in greater storage savings and better deduplication ratios.

Whether performing backups of similar data according to a daily full or weekly full/incremental model, the amount of data storage required is essentially the same as only unique data is stored using either model. Since the amount of data sent to the Data Domain Storage System is substantially greater if doing daily full backups, the deduplication ratio will be considerably higher for a policy of daily fulls rather than a policy of performing a weekly full/daily incremental model. Of course, the amount of data moved by the client would be much, much higher if doing daily fulls and so daily full is not recommended. But if switching from a daily full backup to a weekly full, expect deduplication rate to fall approximately 7x.

Change rate

When the first data stream is deduplicated, the number of unique blocks within it can vary widely depending on both the data type and backup application. The deduplication process may find little or no data redundancy to 50 percent or more data redundancy. Because the Data Domain Storage System compresses unique data blocks, there is still considerable data reduction on the first full backup. In most environments, typically a 2-3x data reduction (50-75%) is achieved on the first full backup.

When multiple backups of the same policy are written to the Data Domain Storage System, however, storage savings often increase significantly with each new save set as only those data blocks that are unique to each backup need to be written to disk. In conventional business operations, this unique data may represent only 1-2 percent of the data present in each additional backup set. The remainder of the backup consists of pointers to data already present.

Data types

Backups of user data such as text documents, PowerPoint presentations, spreadsheets, most database types, source code, and Exchange are known to contain redundant data and are good deduplication candidates.

Other data sources such as audio, video, and scanned images consist of precompressed data. The first full backup typically yields no data reduction at all, but consecutive backups will typically deduplicate well if the data has low change rate.

Compression

Compression does not affect deduplication of files that do not change. However, compression can cause even a small change in a file to ripple throughout the compressed file, greatly reducing the commonality that can be found between current and prior versions, or similar but not identical documents on different machines. As a result, software client-side compression prevents optimal deduplication from taking place since data that is already compressed cannot be efficiently deduplicated once it reaches the target storage.

Data that changes frequently should be uncompressed when transmitted to the Data Domain Storage System to increase the matches that the deduplication engine can find in the data stream. Compressing frequently changing data prior to sending will most likely result in negligible redundancy and poor deduplication ratios. The Data Domain Storage System, in addition to deduplication, compresses the deduplicated data when writing to storage.

EMC recommends avoiding the use of standard NetWorker directives “Standard Windows with compression” or “UNIX with compression” when writing to a Data Domain Storage System.

Encryption

Encryption with the same key does not affect deduplication of identical files that do not change between backups; only space for one copy of the file will be consumed. However, encryption will cause even a small change in a file to ripple throughout a file when it is encrypted, eliminating the commonality that can be found between current and prior versions, or similar but not identical documents. Changing the data zone encryption passphrase will also change the encrypted form of every document, preventing deduplication from finding commonality between current and prior versions. As a result, software client-side encryption prevents optimal deduplication from taking place since data that is already encrypted cannot be efficiently deduplicated once it reaches target storage.

Moreover, encryption prevents compression. The Data Domain Storage System typically provides 2:1 compression in addition to the benefits of deduplication.

For best use of the Data Domain Storage System, do not send encrypted data to it.

However, encryption is an important part of securing a company's data from theft or accidental loss. EMC NetWorker encrypts at the client to distribute the encryption workload and ensure all copies of the backup data are encrypted. NetWorker does not have a mechanism to encrypt just a clone to physical tape sent offsite.

To encrypt backup data, consider the use of an encryption appliance deployed between the Data Domain Storage System and the physical tape library.

If encryption is required without a physical encryption appliance, avoid changing the data zone passphrase frequently, so that the current version of an unchanged file is more likely to be identical to the prior version. If poor deduplication ratios are achieved, it may be appropriate to send those backups to a different backup target.

Multiplexing and parallelism

EMC NetWorker customers typically use parallelism between 4x and 12x, for many reasons:

- To avoid shoeshining on a physical tape drive.
- To increase concurrency for a given number of devices.
- To reduce complexity and user management effort for a given level of concurrency.
- To reduce the load on the storage nodes and server.

While shoeshining does not apply to a VTL, the other reasons do.

When parallelism is applied to a tape or tape-emulating device like a VTL, parallelism often means multiplexing. Multiplexing interleaves backup streams, writing a little of save set 1, then a little of save set 2, and so on, so that none of the clients sending save sets need to wait for the other clients to finish. Unfortunately, this interleaving of save sets has a significant impact on deduplication efficiency when the Data Domain Storage System is used as a virtual tape library (VTL). Multiplexed streams hinder the deduplication process from efficiently identifying blocks of common data because of the additional header information added to the data with parallelism. In order to realize the full benefit of deduplication, EMC

recommends multiplexing be turned off when using the Data Domain Storage System as a VTL with NetWorker by setting both target sessions and max sessions on each virtual tape drive to a value of 1.

Note that the use of parallelism is not an issue if the Data Domain Storage System is used as a NAS backup-to-disk target by creating an Advanced File Type Device (AFTD) on CIFS or NFS shares. In this case each save set is stored as a separate file regardless of parallelism settings. The older File Type Device (FTD) stores savesets as multiplexed however, and is affected.

IMPORTANT: Backup environments designed for backup to NAS are architected very differently from backup environments designed for physical or virtual tape. Engage EMC Solution Architects in a rearchitecture if you are interested in moving from an environment designed for physical or virtual tape to NAS.

More recommendations on dealing with multiplexing and the Data Domain Storage System appear in "Coping with previously multiplexed data" on page 17.

Sizing

Storage capacity needs to be sized to adequately handle the amount of data anticipated to be retained. Backups that are larger than expected or contain data that deduplicates poorly can require much more storage space.

EMC strongly recommends performing a sizing assessment when including replication in the backup environment. The sizing assessment can help determine if replication can occur in the desired timeframe based on the replication network bandwidth and estimated amount of data to replicate on each day.

Contact your EMC representative for information about performing a sizing assessment

Space management

EMC recommends the following practices for space management:

- Expired media is not subject to space reclamation until the volume is also relabeled. Relabeling the expired volume places it in a state that allows the space reclamation process to dereference and subsequently delete the unique blocks associated with the backups on that volume.
- Either NetWorker's Auto Media Management feature needs to be configured on the jukebox, or a backup script using NetWorker CLI commands needs to be created to force relabeling of volumes when they are expired, or the backup administrator must properly size the number of volumes required so they are relabeled as they are expired. NetWorker will always use a blank volume in preference to one with customer data and if there are a lot of unnecessary volumes, space reclamation will be inefficient.

Deduplicated data

After NetWorker relabels a volume for reuse, the metadata index that described that volume is no longer valid. The unique blocks and pointers associated with that volume need to be dereferenced. When unique blocks are no longer referenced by any volumes, the storage space they occupy will be recovered. The Data Domain Storage System's space reclamation feature performs this dereferencing and unique block removal. The space reclamation feature is set to run once per week by default. Unlike other competing products, this feature is not very resource-intensive and typically does not negatively impact other processes such as inline deduplication, and replication.

Best practices

EMC recommends running space reclamation weekly as per the default. This feature can be scheduled or run manually.

If possible, schedule space reclamation to occur outside peak ingestion windows. This will reduce the contention for resources and minimize any impact on ingestion, deduplication, or replication.

Phasing deduplication into the environment

EMC recommends that backup clients are phased in over time to a new Data Domain Storage System, starting with an initial full backup followed by the other backup types.

For example, if the Data Domain Storage System is installed during a normal incremental schedule, the initial backups are tied to the existing technology's last full backup. If a restore is needed, data would be required both from the previous technology as well as the Data Domain Storage System. If the customer does not want to perform a full backup outside the normal policy schedule, they need to wait until the full backup is about to occur to begin sending data to the Data Domain Storage System.

General NetWorker settings and considerations

The following describes general EMC NetWorker settings and best practices for optimizing the backup environment. Specific recommendations for NAS and VTL environments appear later in this paper:

- Avoid running disk-intensive applications such as virus scanning on the backup client when it is backing up or restoring files.
- If the source data for the backup is located on a single, non-RAID, physical disk and multiple streams are running in parallel, the source physical disk could become a performance bottleneck because of parallel reading (parallelism). Therefore, only run a single backup stream on NetWorker clients when the data to be backed up is on a single physical disk. On the other hand, use parallelism on the client when backing up multiple drives.
- Balance backup start times rather than schedule hundreds of backups to begin at 8 P.M. For example, schedule 50 to start at 8 P.M., another 50 to start at 8:30 P.M., and so on. Look at savegroup and client completion times, or drive activity, to balance the load.
- To ensure steady-state load, examine the drive target sessions, and try to keep 10 percent more sessions running throughout the backup window — fewer than this risks stalling target devices, more than this places unnecessary load on the infrastructure.
- On large systems with more than several hundred gigabytes to protect, eliminate data travel through the network by configuring the client as a storage node or dedicated storage node. If the client doubles as a storage node, data is transmitted directly to the Data Domain Storage System, thus providing a more efficient use of network resources.
- Depending on the characteristics of the data, a small number of large files could render better performance than a large number of small files.
- Increase the number of storage nodes and devices, within reasonable limits. for better performance.
- Every device requires memory and CPU resources on the storage node and on the NetWorker server. Carefully build any environment with more than 100 devices. After reaching 100 devices, add devices gradually, perhaps 20 percent of the desired number at a time, to be assured that the infrastructure can cope with the load.
- If considering a switch from a SAN environment to a NAS environment or vice versa, carefully check the LAN infrastructure. A dedicated backup LAN may be appropriate.

NetWorker performance tuning tips are outlined in the *NetWorker 7.6 Performance Optimization Planning Guide*, which replaces the 7.5 and earlier Performance Tuning Guide.

SAN vs. NAS environments

The Data Domain Storage System can operate either as a VTL or as a NAS device.

In existing SAN environments, NetWorker can use the Data Domain Storage System as a VTL. It is possible to reduce or eliminate multiplexing while achieving the benefits of deduplication with adequate performance in this environment through careful planning and management.

In NAS environments, EMC NetWorker's Advanced File Type Device accepts concurrent streams, writing them into separate files in the directory structure of the AFTD. This makes it possible for Data Domain Storage Systems to find and remove common data, thus achieving optimal deduplication performance.

Careful consideration to the type of backup environment to use must be given prior to Data Domain Storage System integration. Consult the EMC Solution Architect team to assist in determining the best environment to meet objectives and involve them in rearchitecting if moving from SAN to NAS or NAS to SAN.

Note: Using Advanced File Type Devices is very different from using VTLs and requires an understanding of the benefits and limitations before deployment. Briefly, AFTD can simultaneously receive many save sessions. They can simultaneously handle many recover sessions. However, only one clone can be performed at a time, so ensure that a single stream can deliver the performance needed to clone data before growing an AFTD too large.

Some considerations to keep in mind:

- Existing infrastructure can be a key consideration. An environment with an investment in tape, either physical or virtual like the EMC Disk Library line, is often architected around SAN, not LAN and Ethernet. Administrators and operators know how to manage it, monitor it, and configure it. AFTDs on NAS require different planning, different infrastructure, different configuration, and different monitoring.
- Customers already using AFTDs can adopt NAS on Data Domain Storage System without much of an infrastructure or mindset change, but those well versed in tape libraries will see significant differences.

Using the Data Domain Storage System with NetWorker in a SAN environment

The Data Domain Storage System, when used with its optional virtual tape library (VTL) feature, can be configured to look like one or more tape libraries with associated tape drives to the storage nodes on the SAN.

The VTL feature is a licensed option of the Data Domain Storage System and is used only in SAN environments.

This section covers Data Domain Storage System behaviors to expect as well as configuration recommendations on to how to achieve optimal performance when using it with NetWorker in a Fibre Channel SAN environment.

For the NetWorker versions and supported operating systems, refer to the *EMC Support Matrix.*

Refer to "Key factors affecting the deduplication ratio" on page 7 for additional recommendations and settings.

Data Domain Storage System SAN attach (VTL)

The Data Domain Storage System has two or four 4-gigabit Fibre Channel ports for target mode FC attach. All connections to these ports are made via a Fibre Channel switch. Direct attachment of a device to these ports is also supported. The following recommendations apply when connecting the Data Domain Storage System to a backup server via a Fibre Channel switch:

- When implementing a Data Domain VTL, use a Fibre Channel switch listed on the Data Domain FC_SwitchCompatibilityList for the software release that is applicable to the specific Data Domain Storage System.
- Because the Data Domain VTL provides LUN Masking capabilities, consider using port zoning on the SAN switches.
- Switch encryption solutions are not supported by the Data Domain Storage System.
- Limit FC extended fabric (ISL link) configurations to three hops between the backup server/storage node and the Data Domain Storage System.

Device discovery (persistent binding)

When an operating system scans its storage buses (Fibre Channel or SCSI) for devices, it may assign a particular device to a different device name from the one it used during a previous scan. Since backup software saves, in its database, information about the device names of the various tape drives it sees during the configuration process, device reassignment during a scan will impact the availability of backup devices to the application.

Persistent binding is the ability to fix a particular device to a particular device name. In general, this is done by associating a unique device characteristic (such as a WWPN or serial number) with a device name to maintain persistence of a particular device to a specific device name. The method for doing this varies for different operating systems and Fibre Channel disk and tape device drivers.

If persistent binding is improperly configured, the operating system may give a device multiple names — giving a device one name after a boot, but giving it a different name on a subsequent boot.

IMPORTANT: EMC recommends the use of persistent binding at the operating system level. This will prevent interruption of backup operations to the Data Domain Storage System and difficult-to-diagnose problems after a system reboot.

Device drivers

Please see the latest copy of the "Configuring Tape Devices for EMC NetWorker" technical note on Powerlink for current information on device drivers.

Scanning devices

The operating system can scan for new devices. If all devices are not detected, a gap may exist in the LUN numbering. It is recommended that you use the NMC GUI to scan for devices.

If scanning from the NMC GUI does not work, use the inquire and jbconfig commands to find new devices.
Run **inquire –l** to force LUN detection across gaps in LUN numbering.
If there is a gap in the LUN numbering, run **jbconfig –l** to configure the environment.

Configuring the backup server

NetWorker operates in a client/server model consisting of NetWorker servers, clients, storage nodes, and storage devices. When backing up to a virtual tape library on the Data Domain Storage System using a storage node, a backup client sends the data over the LAN to the storage node and the storage node sends data across the SAN. If the backup client is also a storage node, the client sends the data directly across the SAN. The backup metadata associated with the backup goes via IP over LAN/WAN from the backup client to the backup server where it is stored in the client file index and media database. The metadata is instrumental to NetWorker's ability to quickly locate and restore backup files.

1. Create a virtual tape library on the Data Domain Storage System, specifying the emulation type (currently only STK L180), number of slots, number of tape drives, and their emulation type (LTO1,2 or 3). This VTL will be a backup device on the NetWorker storage node. If desired, multiple VTLs can be created to service different backup servers/storage nodes.

2. Create virtual tape cartridges on the Data Domain Storage System for the VTL, specifying the cartridge type, quantity, capacity, and barcodes. If save sets will be cloned to a physical library, use the same cartridge type and blocksize as the physical library for simplicity — however, capacities may be different between physical and virtual cartridges.

Note: NetWorker inserts multiple tape markers into the data stream to the VTL that can affect deduplication ratios. DDOS versions up to and including 4.7 do not fully compensate for these tape markers. To ensure good deduplication rates, it is strongly recommended to set the block size on all virtual tape drives to 256k or 512k.

3. Assign the VTL to a NetWorker storage node on the SAN to which the Data Domain Storage System is attached. This is done using Access Groups (the equivalent of LUN masking) on the Data Domain VTL.

4. Configure persistent binding on the NetWorker storage node.

5. Scan for devices at the NetWorker storage node. The VTL and its tape drives will appear.

6. Apply NetWorker autochanger license(s) to enable use of the Data Domain device. A licensing bundle for Data Domain is available with the necessary enablers to support a Data Domain Storage System as a VTL and a NAS device. It provides entitlement for 10 TB of physical disk capacity within the system. Additional capacity can be licensed as needed.

7. Perform the normal NetWorker setup for backups. See the *EMC NetWorker Administration Guide*.

A single client with multiple disk spindles, or multiple clients, can concurrently back up multiple save sets as parallel backup streams. In general, multiple parallel backup streams will increase the aggregate throughput of the backup process and will shorten the backup window of the client, for a given number of devices.

EMC strongly recommends turning off multiplexing in NetWorker so that backup data from multiple sources is not interleaved on the virtual tapes because this will significantly impact deduplication ratios. Keep in mind, however, that turning off multiplexing will adversely affect performance unless the environment is adjusted to compensate as described next.

1. Edit the properties of each virtual device in NetWorker to set both the **target sessions** and **max sessions** to 1. Note that in some NetWorker versions, setting them both to 1 at the same moment does not actually change the value of both settings. It is therefore necessary to set the target sessions to 1 first, then come back and change the max sessions to 1.

2. In the NSR resource, increase the **nsrmmd polling interval** to at least 15 and the **restart interval** to at least 10.

3. Re-examine schedules so that the number of active sessions in NetWorker is always approximately 10 percent over the aggregate target sessions.

4. Add more virtual devices gradually to the existing environment. Do not assume that the storage node and server infrastructure can support significantly more devices. Monitor your environment and confirm that the storage nodes can handle the load: CPU, memory and IO.

Simultaneous backups

The Data Domain Storage System emulates the STK L180 industry-standard library and LTO tape drive types (LTO1, 2, and 3). These appliances support multiple virtual libraries and tape drives simultaneously.

To address backup bottlenecks commonly seen with shared devices, two solutions can be deployed:

- Create a library and tape drives for each backup host's exclusive use to ensure the best possible performance.
- Create a single library with many tape drives, but only assign a few tape drives to each storage node. In this configuration, each storage node has its own dedicated resources. This also simplifies the initial installation and avoids any driver-related issues you could run into when sharing a tape device between unlike OS platforms.

With either solution, it is a good practice to limit the number of virtual tape drives so when all are servicing I/O at the same time each drive can sustain at least 20 MB/s. You may need to stagger backups to support fast and slow data streams to obtain maximum utilization of the available bandwidth.

Note: A NetWorker storage node license enables either 16 devices (NetWorker Edition) or 32 devices (Power Edition). The devices can be logical, physical or a combination of logical/physical. Note that you may saturate your storage node's CPU, system bus, or interfaces before using that many devices; monitor your storage nodes.

Single library – multiple hosts

For multiple hosts to use the same devices, the Data Domain Storage System requires you to create different access groups for each host. A group consists of exactly one host (initiator), one or more target FC ports on the Data Domain Storage System, and one or more devices. The Data Domain Storage System does not permit multiple hosts to access the same group.

Virtual tape libraries

Emulations

The VTL permits you to create combinations of any emulated tape drive. NetWorker is generally insensitive to these combinations.

Other backup software applications are sensitive to combinations that do not exist in the real world, and will not work properly when unsupported combinations of library and tape drive types are specified. When they fail, it may not be obvious that it is an invalid combination.

EMC recommends creating virtual library and drive combinations that are formally supported by the actual physical libraries/drives.

Number of slots

The number of slots for a particular virtual tape library is not limited to the amount supported by the physical tape library it is emulating, but the number defined must be licensed appropriately in NetWorker.

Note: The NetWorker licensing bundle for Data Domain supports a Data Domain Storage System configuration with up to three (3) virtual libraries on the same physical system with an unlimited number of virtual cartridge slots. When using more than one physical system, additional licensing is required.

Assigning and unassigning virtual libraries

Until a virtual library is assigned to a SAN client, it is not enabled for use by the backup servers. To ensure correct library behavior when assigning or unassigning virtual libraries, ensure that device numbering on the bus starts with LUN 0 and that there are no LUN numbering “holes.” LUN numbering holes can occur when devices are unassigned.

Virtual tape cartridges

When you create virtual tape cartridges, they are spread across the storage in a round-robin fashion. This ensures maximizes performance by spreading the I/O load across many disk spindles.

The cartridge size can be customized and can exceed the capacity of its physical equivalent. This ensures that the compressed data stored on a virtual tape cartridge will fit on a physical tape once it has been decompressed and then recompressed by the physical tape drive. Although tape cartridge sizes of up to 800 GB are supported, it is recommended to use a size of 100 GB per tape when the tapes are being replicated to another Data Domain Storage System. This is to ensure the fastest possible time to DR since the Data Domain VTL will replicate a tape as soon as NetWorker stops writing to it for several minutes. By keeping the cartridge size relatively small, the replication will be granular and very fast.

Reclaiming space

Much like physical tape, the bytes from expired save sets still reside on the virtual cartridge until the cartridge has been relabeled. When a virtual tape becomes recyclable in NetWorker, relabel it to free up space in the Data Domain Storage System.

By default, NetWorker will not automatically label or relabel tapes. But automatic tape labeling will be done if the Auto Media Management feature is set on the jukebox resource:

The following is from the *EMC NetWorker Administration Guide*:

> Auto Media Management gives the NetWorker server automatic control over media loaded in the storage device. When Auto Media Management is enabled during device configuration, the NetWorker server automatically:
>
> - Labels the volume
> - Mounts the volume.
> - Overwrites volumes it considers to be unlabeled.
> - Recycles volumes eligible for reuse that are loaded into the device.
>
> The NetWorker server considers a volume to be unlabeled under the following conditions:
>
> - Has no internal label.
> - Is labeled with information other than a NetWorker label.
> - Is labeled with a NetWorker label, but the density indicated on the internal label differs from that of the device where the volume is mounted.
>
> **Note:** The NetWorker server considers volumes that were labeled by a different application to be valid relabel candidates if Auto Media Management is enabled. Once the NetWorker server relabels the volume, the previously stored data is lost.

If NetWorker automatically recycles [=reuses] a volume after all of the save sets on the volume have passed the customer-set useful lifetime, the two 32 KB labels at the beginning of the tape are *not* overwritten; the tape is written beginning at the tape position just after the second tape label. In other words, the tape usage size does not return to zero.

Note: When the NetWorker server recycles a volume, the volume label does not change if the volume remains in the same media pool. That is, if a volume labeled Dev.006 is recycled, it will retain the volume label Dev.006 and will not receive a new label with the next sequence number. The original data on the volume, however, will be overwritten by the new data

Coping with previously multiplexed data

If a Data Domain Storage System has too much space already consumed by existing multiplexed data, and that data needs to be retained, the data can be un-multiplexed by careful cloning. The cloned tapes will deduplicate when rewritten. Follow this by relabeling the original multiplexed tapes to free the non-deduplicated space. (Be sure everything is off of the originals before relabeling!)

Note: This process will require 1) scripting; 2) extra storage capacity; and 3) processing through a storage node.

There is an additional benefit from this extra clone step beyond deduplication. The data is now in an un-multiplexed format for faster recovery.

Look to the hardware and time requirements for existing cloning to judge what is needed for an extra clone step. Be conservative when estimating as single save set cloning will need at least N reads of the tape, where N is the amount of multiplexing. (Normal cloning of multiplexed data only requires one read of the tape.)

Demultiplexing through cloning requires sufficient space to accommodate the multiplexed and demultiplexed cloned copies of the same data. This will not be practical if the Data Domain Storage System is already full.

However, if sufficient capacity exists, cloning can occur. The cloning operation is performed from the command line and is on a scale that will likely require scripting. *Save sets must be cloned individually*.

For example:

```
# pseudocode: EMC does not support scripts
for vol in $volume_names
do
    ssids=`mminfo -q volume="$vol" -r ssid`
    for ss in $ssids
    do
        nsrclone -S $ssid
    done
done
```

The script must not try to clone all the save sets on a volume with a single nsrclone *command. A performance optimization in NetWorker will leave data from those save sets multiplexed on the clone volume, which defeats the purpose here.*

Using the Data Domain Storage System with NetWorker in a NAS environment

This section covers Data Domain Storage System behaviors to expect as well as configuration recommendations on how to achieve optimal performance when using it with NetWorker in a NAS environment.

General recommendations

Refer to "Key factors affecting the deduplication ratio" on page 7 for additional recommendations and settings.

Network connectivity

Aggregation of multiple gigabit Ethernet links is supported on the Data Domain Storage System. Currently no dynamic link aggregation protocol like LACP is supported. When interconnecting with Cisco switches, the Port Channel therefore needs to be set in On mode.

In general:

- Use a dedicated network by configuring a separate network or use QoS features that guarantee network bandwidth.
- Alternatively, use virtual networks (VLANs) to segregate backup from production network traffic.
- Alternatively, direct connect Storage Nodes to the Data Domain Storage System using Gigabit or 10Gigabit Ethernet connections.
- Set network interface cards (NICs) for servers and clients to full duplex. Set all routers to full duplex.
- Use proper cables. Gigabit networks require CAT 5e or CAT 6 cables. Ten gigabit networks require CAT 7 cables.
- If using a DNS server, verify the DNS server configuration settings are correct. It is extremely important that the Data Domain Storage System can reverse lookup each IP address that sends data to it. The use of host entries is recommended.
- Use multiple Data Domain Storage System ports when connecting to the network. The more ports used, the better the load balancing will be across the ports. Note that link aggregation load balancing for backups will be dictated by the link aggregation load balancing that the Ethernet switch(es) perform. The link aggregation load balancing scheme only plays a role when restoring data.
- For redundancy, connect at least two ports to an Ethernet switch.
- Set each switch port used to auto-negotiate/auto-sensing. The network interface cards are preset to auto/auto but can be changed.
- When connecting multiple Ethernet ports to the network using link aggregation make sure the proper settings on the switch are configured. Consult the switch vendor's documentation for the configuration steps.

Keep the following in mind:

- Concurrent operations on the Data Domain Storage System *will* impact performance
- The Data Domain Storage System does support jumbo frames. With many servers, optimum performance can only be obtained when using jumbo frames between Storage Node and Data Domain Storage System.
- Use OS vendor recommendations to perform network optimizations appropriate for your OS.

Backup-to-disk configuration

NetWorker operates in a three-tier client/storage node/server model consisting of NetWorker servers, clients, storage nodes, and storage devices. When backing up to disk using a storage node, a backup client sends the data over the LAN to the storage node and the storage node sends the data to the Data Domain Storage System. If the backup client is also a storage node, the client sends the data directly to the Data Domain Storage System. The backup metadata associated with the backup goes from the backup client to the backup server where it is stored in the client file index and in the media database. The metadata is instrumental to NetWorker's ability to quickly locate and restore backup files.

Best practice for AFTDs is to create one per pool on a storage node and not to place more than one on a file system. The AFTD should be the only thing on the file system.

In the simplest LAN backup-to-disk configuration, there is a single NetWorker client, a backup server, and Data Domain Storage System as the storage device all connected via an Ethernet LAN.

Setting up the backup-to-disk environment involves the following steps:

1. Create an NFS or CIFS share on the Data Domain Storage System. This share will be mounted (or mapped) as a backup device on the NetWorker server/storage node. If desired, multiple shares can be created to store different collections of save sets.

2. Mount (or map) the share on the NetWorker server/storage node.

   a. Create a mount point for the NFS share to be used by the backup application and mount it using the commands for your particular operating system as described in "Mounting NFS shares" on page 21.

   b. Map the CIFS share to a drive in Windows Explorer and create a directory to be used by NetWorker as the location for the backup device as described in "Configuring CIFS shares" on page 23.

3. Apply the NetWorker DiskBackup Option license provided with the Data Domain licensing bundle. This enables the Data Domain Storage System to be used as a disk target for backup with NetWorker.

4. Create a NetWorker Advanced File Type Device that points to the NFS share or UNC path to the CIFS share.

   a. Open the NetWorker Management Console and then invoke Administration.

   b. Create a new device as follows:

      Select Devices and under the Devices tree, create a new device:

      "Name" = to full path to the CIFS share (example: \\10.40.166.100\pep_cifs1\nw)

      "Media Type" = adv_file

      Also specify "netadmin" and the password for "username" and "password" on the second tab.

      Important: If using a NetWorker AFTD as CIFS from a windows server, you must change the account that the NetWorker Remote Exec Service runs as. By default, NetWorker services run as the local system account, which does not have permissions to use CIFS shares. Create a local windows administrator with the same name and password as the DD user, and give the new user Local Admin Privileges so that it can back up all files on the system. Then use the **Control Panel > Services** to change the logon of the service, and restart the service.

5. Perform the normal NetWorker setup for backups. See the *EMC NetWorker Administration Guide*.

During a backup operation, the NFS or CIFS share designated as the backup device receives the save set directly from the client/storage node or backup server.

In large environments it is not likely that every client will have storage node software installed, or storage node licenses available. When there are a large number of clients, it is much more likely that a small number of storage nodes will serve as an aggregation point for a large number of clients. The clients will send their data to the storage node with which they are associated, and the storage node will send the data to the share.

Mounting NFS shares

The Data Domain Storage System can serve as a NAS appliance for backup purposes. If it is an NFS environment, access to the shares is restricted by hostname, IP address, domain name or IP subnet range.

Data Domain Storage System only supports NFSv3.

NFS shares are mounted on the NFS client in the following environments as described next:

AIX 5.2 and later

Enter the following command:

```
nfso -o nfs_use_reserved_ports=1
```

This mount command does not persist across AIX reboots. For AIX 5.2 or later, use the –p option to mount the share permanently. For releases of AIX prior to 5.2, the mount command must be included in a boot up script.

If you are using NFSv3, mount the NFS share using this command:

```
mount –o timeo=600 {nfs_server}:/{export path} /{mountpoint}
mount -V nfs -o llock,intr,hard,rsize=32768,wsize=32768,
proto=tcp,combehind,timeo=600,retrans=2 {nfs_server}:/{export path} /{mountpoint}
```

For example: mount -V nfs -o llock,intr,hard,rsize=32768,wsize=32768, proto=tcp,combehind,timeo=600,retrans=2 ddr:/backup /mnt/dd

To show the list of file systems exported by the Data Domain Storage System:

```
nfs show clients
```

In addition, to optimize TCP/IP performance on the AIX host, apply the following parameters:

```
Set large_send to no for each NIC interface

Other changes that will likely increase throughput:

# no -p -o sack=1
# no -p -o tcp_newreno=0
# chdev -l en10 -a rfc1323=1
# chdev -l en10 -a tcp_nodelay=1
# chdev -l en10 -a tcp_recvspace=262144
# chdev -l en10 -a tcp_sendspace=262144
(change en10 with the respective Ethernet device(s) on the Storage Node(s) you
are working with)

# nfso -o nfs_rfc1323=1

The setting above needs to be re-applied with -p, to become permanent, as the
following:
# nfso -p -o nfs_rfc1323=1
```

HP-UX 11i

If you are using NFSv3, mount the NFS share using this command:

```
mount –F nfs –o rsize=32768,wsize=32768,hard {nfs_server}:/{export path}
/{mountpoint}
```

For example: `mount –F nfs –o rsize=32768,wsize=32768,hard,vers=3 ddr:/backup /mnt/dd`

To show the list of file systems exported by the Data Domain Storage System:

```
nfs show clients
```

```
HPUX NFS Additional Tuning Parameters:
Add the following line to the file /etc/rc.config.d/nfsconf:
NFS_CLIENT="1"
NUM_NFSIOD=24
Stop and restart the NFS daemons with the commands:
# /sbin/init.d/nfs.client stop
# /sbin/init.d/nfs.client start
```

In addition, to optimize TCP/IP performance on the HP-UX host, apply the following parameters:

```
Set the TCP send and receive sizes for HP-UX 11.0 and 11i backup servers.
• To set the sizes immediately, enter the following two commands on the HP-UX
server and then remount the restorer NFS share to enable the values:
ndd -set /dev/tcp tcp_recv_hiwater_def 262144
ndd -set /dev/tcp tcp_xmit_hiwater_def 262144
• To make the changes persistent over system reboots, create a startup script
that runs before the NFS automount. The numbering in the script name and
location depends on how startup scripts are set up on your system, but as an
example:
/sbin/rc3.d/S99dd. Enter the following two lines in the script:
ndd -set /dev/tcp tcp_recv_hiwater_def 262144
ndd -set /dev/tcp tcp_xmit_hiwater_def 262144
```

Linux

Mount the NFS share using this command:

mount -T nfs -o hard,intr,nfsvers=3,tcp,rsize=32768,wsize=32768,bg {nfs_server}:/{export path} /{mountpoint}

For example: `mount -t nfs -o hard,intr,nfsvers=3,tcp,rsize=32768,wsize=32768,bg` ddr:/backup /mnt/dd

To show the list of file systems exported by the Data Domain Storage System:

```
nfs show clients
```

For TCP/IP tuning, see the Linux tech note on the EMC Data Domain support site https://my.datadomain.com

Solaris

Mount the NFS share using this command:

```
mount -F nfs -o
hard,intr,vers=3,proto=tcp,rsize=32768,wsize=32768{nfs_server}:/{export path}
/{mountpoint}
```

For example: `mount` -F nfs -o hard,intr,vers=3,proto=tcp,rsize=32768,wsize=32768 `ddr:/backup /mnt/dd`

To show the list of file systems exported by the Data Domain Storage System:

```
nfs show clients
```

Solaris system settings to improve TCP/IP NFS performance

```
On the Solaris backup server:
•  Create a file /etc/rc3.d/S90ddr. Enter the following two lines in the file:
ndd -set /dev/tcp tcp_recv_hiwat 131072
ndd -set /dev/tcp tcp_xmit_hiwat 131072
•  In the file /etc/system, add the following lines:
set nfs:nfs3_max_threads=16
set nfs:nfs3_async_clusters=4
set nfs:nfs3_nra=16
set rpcmod:clnt_max_conns=1
set fastscan=131072
set handspreadpages=131072
set maxpgio=65536

Note that SUN T-processor (aka coolthreads) servers have notoriously bad NFS
performance. The only adequate resolution for this is to use Jumbo Frames.
```

Configuring CIFS shares

If you are operating in a Windows environment, the Data Domain Storage System can either operate in Workgroup mode or it can join an Active Directory Domain. When in AD mode, the Data Domain Storage System can be setup so that CIFS shares can be configured from the Microsoft Management Console (MMC). CIFS shares can also be administered from the command line interface.

1. If operating in Workgroup mode, add a user named nwadmin and a password on the Data Domain Storage System:

   ```
   admin@ddsystem# user add nwadmin password xxxx
   ```

2. Open up CIFS access on the Data Domain Storage System to the appropriate NetWorker Storage Node(s) and NetWorker server:

   ```
   admin@ddsystem# cifs add /backup <ip address or name of server>
   ```

3. Add a Windows user (local user when in workgroup mode, AD user when in AD mode) named nwadmin and make the user a member of the Windows Administrators and Backup Operators group.

4. If using Active Directory, ensure that the Data Domain Storage System is added to the active directory that is used by the backup server that is targeting the Data Domain Storage System.

   ```
   admin@ddsystem# cifs add /backup <clien-list>
   ```

5. Login to the NetWorker Server as the nwadmin user and open a UNC path to the Data Domain Storage System (Start>Run>\\ddsystemname\backup). Create a directory (structure) for the NetWorker AFTDs.

6. On the NetWorker Server(s) under Control Panel>Administrative Tools>Services, edit the **NetWorker Backup and Recover Server** and **NetWorker Remote Exec Service**.

   a. Under **Log on as**, choose **This accoun**t and enter the nwadmin username and password.

   b. Restart these services.

7. With respect to CIFS performance, on Windows versions prior to 2008, it is necessary to tune the TCP/IP parameters on each Storage Node. Add the following settings to the registry of each storage node and reboot the servers for these settings to take effect:

```
Windows Registry Editor Version 5.00

[HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\lanmanworkstation\paramet
ers]
"SESSTIMEOUT"=dword:00000e10

[HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\AFD\Parameters]
"DefaultSendWindow"=dword:00040000
"DefaultReceiveWindow"=dword:00040000

[HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters]
"GlobalMaxTcpWindowSize"=dword:00040000
"TcpWindowSize"=dword:00040000
"Tcp1323Opts"=dword:00000003
```

   In addition, check the version of NIC drivers. Many older NIC drivers have severe performance issues, so it is strongly recommended to upgrade NIC drivers to a recent version.

   EMC has issued patches for NetWorker 7.4 and 7.5 that specifically improve the AFTD performance over CIFS. Because these patches provide a 200% or better improvement over prior versions, it is critical that they are installed.

Staging limitation

When used with Data Domain Storage System, NetWorker AFTDs support automatic staging to tape based on data age.

NetWorker AFTD used with Data Domain Storage System does not support staging triggered if or when a Data Domain Storage System filesystem fills up.

When savesets are staged off of the Data Domain Storage System, space is not available for subsequent backups until the next time the cleaning process runs. Best practice is to run cleaning weekly.

Monitor the size of your Data Domain Storage System NAS shares and avoid relying on high and low watermarks.

General: Moving data from disk to physical tape

There are many options for copying backups on virtual tape to physical tape for long-term data retention, which apply to both SAN and NAS environments. Physical tapes can be created from backups residing on AFTD or VTL with the standard method of *cloning* with backup software. The NetWorker Storage Node will read the data off the original target and copy it to physical tapes.

NetWorker cloning

When data backed up to a Data Domain Storage System in VTL mode needs to be written onto physical tape, NetWorker can clone save sets written to emulated tape in the Data Domain Storage System through a production storage node to a SAN-attached tape library. This operation is a standard NetWorker cloning procedure. This procedure uses emulated tape drives in the Data Domain Storage System in conjunction with a tape device in another tape library to perform the cloning process. Cloning from a production storage node to a second storage node can also be performed over IP between the storage nodes.

Similarly, cloning will move data from a AFTD to a physical tape library.

Important: Only one backup can be cloned off of an AFTD at a time.

The advantages and disadvantages of this strategy are as follows:

Advantages

- Enables cloning from the Data Domain Storage System under NetWorker control, with standard NetWorker policy support; that is, multiple retention policies for different cloned copies of data.
- Enables cloning at the saveset level.
- Enables copying from one emulated tape type to another tape type (physical or virtual). This is sometimes referred to as tape conversion.
- Enables copying from multiple emulated tape cartridges to a single physical or virtual tape (tape stacking).

Disadvantages

- Requires storage node licenses.
- Must maintain additional front-end SAN infrastructure for connections to a separate tape library as well as the emulated tape library.
- Consumes SAN bandwidth – data must be copied from a virtual tape cartridge in the Data Domain Storage System over the SAN to another physical device on the SAN.
- Consumes storage node CPU, memory, and I/O bandwidth.

Replicating deduplicated data

Replication is an optional licensed feature. The Data Domain Storage System can replicate tape pools (collections of virtual tapes) and/or directories on the Data Domain Storage System. Replication is independent of NetWorker.

Important: Manual tracking of the replicated copies is required when using replication, since Data Domain Storage System replication occurs outside of the backup application.

Replication considerations

Replication transmits data over Ethernet using TCP. By default, the Data Domain Storage System uses TCP port 2051 for replication but this port can be changed by the user. There is no physical Ethernet port reserved for the replication feature. This can however be controlled by IP routing. See "Network connectivity" on page 19 for recommendations when connecting to the network.

When adding Data Domain Storage System replication to the backup environment, keep the following in mind:

- Always assess the network between the source and target systems in *both* directions as part of the test and acceptance process and prior to putting the systems into production. This will ensure that sufficient bandwidth is available to support the maximum amount of data desired to be replicated within the replication window. This can also be done by performing a formal assessment if necessary.
  - Use "ping" or similar analyzer software to get an idea of latency between the two systems as well as determine if any network bottlenecks are occurring. Run this test multiple times on different days at different times to confirm the numbers.
  - Run iperf between the Data Domain Storage Systems to assess the available bandwidth.
- Seeding of the target system is almost always required. See "Seeding (pre-populating the target system)" for recommended options.
- Do not place replication configurations into production until required replication windows are met. When replication falls behind (exceeds the replication window), it is hard to catch up.

Replication performance depends on daily change rates between new backup data and data residing on the target system.

- The higher the change rate, the more data that needs to be moved from the source to the target system.
- Lower than expected deduplication rates will result in higher bandwidth utilization (if more bandwidth is available) or longer replication times and may exceed the required replication window.
- Network latency has very little impact on Data Domain Storage System replication performance.

Since the Data Domain Storage System replicates data within minutes after it has been written to the source, and since the metadata travels concurrently with the data blocks, each file or virtual tape that has finished replicating is immediately available for recovery at the replication site.

See Data Domain whitepapers on replication for more detail.

Seeding (pre-populating the target system)

Initial replication (first data copy) is a time-consuming operation and will usually take significantly more time than later replications because the data is typically new and unknown to the target Data Domain Storage System.

For replication, the goal is to facilitate the transfer of the large amount of unique data produced by initial backups to the source shares/VTLs to the target Data Domain Storage System. This is done by seeding the target with the same data as on the source so that only the minimum amount of data needs to be transferred between units on subsequent replications to maintain synchronization. The source and target systems should not be put into production until this seeding process is complete. Seeding is considered complete when the required replication window has been demonstrated to have been met.

There are a number of options that can be employed to accomplish this:

- **OPTION 1**: Locate the source and target so that they are on a dedicated Ethernet network and replicate locally. This will allow replication to proceed at the fastest possible network rate. EMC recommends replicating at least two full backups in this local configuration to demonstrate replication savings. If savings are not seen, further local replications may be needed to ensure the target system has sufficient data. (That is, the data transferred by subsequent replications is more representative of the quantity of data that will typically replicate within the window.) When the seeding completes, the target can be deployed to its intended location.
- **OPTION 2**: Use tape to seed the data in the target. Use tapes from at least two full backups present on the source, and write the data to temporary NAS shares or a VTL on the target. Then setup and initialize replication from the source to the target. When the replication has finished initialization, delete the temporary NAS share or VTL on the target.

If replication is not enabled for a NAS share or tape pool until multiple backup streams destined for that particular virtual device or NAS share are ingested, there will be a large number of unique blocks on the source device that will need to be transferred over the network when the replication is setup.

If there are disruptions during the replication process for any reason (for example, a network infrastructure problem), reseeding may be necessary to eliminate a large replication backlog that can occur as a result of the disruption. Large backlogs, when they exist, will resolve if sufficient excess network bandwidth is available.

Recovery/Failback

VTL:

Should a VTL become unavailable on a source, two methods are available to access the data either at the target system (recovery) or by replicating the VTL to the source (failback).

In order for the replicated tapes to be accessed from the target by a secondary storage node, a Virtual Tape Library has to be configured on the target, after which the replicated tapes can be imported into the library and be presented to the secondary storage node.

When failing back to the source, export the tapes from the VTL on the replica, have the secondary storage node reinventory that library, then have NetWorker inventory the source and finally resume operations as normal. Exporting the tapes from the replica VTL is important since otherwise NetWorker may end up seeing the same barcode twice.

If the VTL and volume names are the same on the replicated side, disable the original library in NetWorker, create a new library, and issue reset and inventory commands once it is up. That will result in NetWorker updating the location field for all volumes so it will be able to use them automatically from that point on.

This can be done proactively with both VTLs (source and replicated one) configured in NetWorker all the time. Just make sure that only a single library is enabled in NetWorker at any given point in time – if both libraries are enabled, conflicts will occur.

If such a configuration is done beforehand, failover from the NetWorker perspective can be performed in less than one minute, but this is a manual process. It cannot be automated.

NAS:

Recovering from a failure of a NAS device is similar. Before the disaster, it is helpful to define a AFTD device to be used to mount the replicated NAS. This AFTD should use the path that will be available on the replicated Data Domain Storage System / remote storage node. This AFTD MUST BE UNMOUNTED AT ALL TIMES WHILE THE PRIMARY IS MOUNTED. (The AFTD contains a unique ID that will be replicated; NetWorker cannot see the same unique ID in two places at once.)

- Unmount the failed device from NetWorker.
- Mount the replicated NAS onto the remote storage node.
- Recover files.
- When failing back, insure that the volume is only shown to NetWorker from one Data Domain Storage System at a time.

Conclusion

The processes and configuration recommendations described in this white paper are intended to assist you in maximizing the performance and usage of the Data Domain Storage System in the majority of NetWorker backup environments. Since no two backup environments are exactly the same, it may be necessary to vary individual settings to meet each specific environment's requirements and goals.

References

The following EMC documents and white papers may provide further information. They can be found on EMC Powerlink.

- *Configuring Tape Devices for EMC NetWorker*
- *Configuring TCP Networks and Network Firewalls for EMC NetWorker*