**Preface**

Chapter 1: Introduction

Overview

System requirements

Restrictions and limitations

User interface choices

Related information

Chapter 2: Concepts

Overview

Deduplication states

Duplicate file detection

Space-reduced backups

CIFS compressed file attribute

Externally triggered deduplication

Cloning capabilities

Planning application integration

Backup and restore

VNX File System Migration

VNX file-level retention

VNX FileMover archiving

File system space usage

Multi-Path File System

Point-in-time views of the file system

Quotas

Replication
As part of an effort to improve and enhance the performance and capabilities of its product lines, EMC periodically releases revisions of its hardware and software. Therefore, some functions described in this document may not be supported by all versions of the software or hardware currently in use. For the most up-to-date information on product features, refer to your product release notes.

If a product does not function properly or does not function as described in this document, please contact your EMC representative.
**Special notice conventions**

EMC uses the following conventions for special notices:

---

**CAUTION:** A caution contains information essential to avoid data loss or damage to the system or equipment.

---

**Important:** An important note contains information essential to operation of the software.

---

**Note:** A note presents information that is important, but not hazard-related.

---

**Hint:** A note that provides suggested advice to users, often involving follow-on activity for a particular action.

---

**Where to get help**

EMC support, product, and licensing information can be obtained as follows:

- **Product information** — For documentation, release notes, software updates, or for information about EMC products, licensing, and service, go to the EMC Online Support website (registration required) at [http://Support.EMC.com](http://Support.EMC.com).
- **Troubleshooting** — Go to the EMC Online Support website. After logging in, locate the applicable Support by Product page.
- **Technical support** — For technical support and service requests, go to EMC Customer Service on the EMC Online Support website. After logging in, locate the applicable Support by Product page, and choose either Live Chat or Create a service request. To open a service request through EMC Online Support, you must have a valid support agreement. Contact your EMC sales representative for details about obtaining a valid support agreement or with questions about your account.

---

**Note:** Do not request a specific support representative unless one has already been assigned to your particular system problem.

---

**Your comments**

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications.

Please send your opinion of this document to:
Introduction

Topics included are:

* Overview on page 10
* System requirements on page 10
* Restrictions and limitations on page 11
* User interface choices on page 13
* Related information on page 13
Overview

The EMC® VNX™ File Deduplication and Compression feature increases file storage efficiency by eliminating redundant data from the files stored in the file system, thereby saving storage space and money. For each file system, file-level deduplication gives the Data Mover the ability to process files in order to compress them, and the ability to share the same instance of the data only if they happen to be identical. Deduplication functionality operates on whole files and is applicable to files that are static or nearly static with a last-access time greater than 15 days (default value in systems that use version 5.6.47 or later) or greater than 30 days (default value in systems that use version 5.6.43 through 5.6.46).

For example, if there are 50 unique files in a file system that can be deduplicated, 50 unique files will still exist but the data will be compressed, yielding space savings of up to 50 percent. If there are 70 identical copies of a presentation document in a file system, 70 files will still exist but they will all share the same file data. In the second example, the data usage will decrease by a factor of almost 70. In addition, the one instance of the file data shared by the 70 files will also be compressed, providing further space savings.

Note: Prior to systems that used version 5.6.47, deduplication options were set by using the VNX CLI server_param command. Starting with systems that use version 5.6.47, deduplication options are set by using either the VNX CLI fs_dedupe command or the file system Deduplication Settings tabs in the EMC Unisphere™ software.

VNX File Deduplication and Compression processes file data, not metadata. If multiple files contain the exact same data but have different names, the files are deduplicated. Duplicate files can also have different permissions and timestamps.

This document is part of the EMC VNX documentation set and is intended for use by the system administrators responsible for creating and managing a deduplication-enabled file system.

System requirements

Table 1 on page 10 describes the EMC VNX series software, hardware, network, and storage configurations.

Table 1. System requirements

<table>
<thead>
<tr>
<th>Software</th>
<th>VNX series version 7.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware</td>
<td>VNX Data Mover</td>
</tr>
<tr>
<td></td>
<td>• NS series (except for NS600)</td>
</tr>
<tr>
<td>Network</td>
<td>No specific network requirements</td>
</tr>
<tr>
<td>Storage</td>
<td>No specific storage requirements</td>
</tr>
</tbody>
</table>
Restrictions and limitations

If any of this information is unclear, contact your EMC Customer Support Representative for assistance:

◆ The file system must have at least 1 MB of free space before deduplication can be enabled. If there is not enough free space, an error message is generated and the server log is updated.

◆ By default, deduplication does not process files smaller than 24 KB.

◆ During the deduplication process, the file system must have enough free space available that is equivalent to the size of the original file to be deduplicated, plus the size of the compressed version of the file to be stored:
  ◆ An additional 1 percent of the file system must be free.
  ◆ If auto-extension is enabled, an additional 1 percent below the auto-extension threshold must be free.

◆ File systems enabled for processing by VNX File Deduplication and Compression may be replicated by using VNX Replicator. All destination file systems are required to support VNX File Deduplication and Compression.
  ◆ If the source system is running version 5.6.43 through 5.6.46, the destination system must be running version 5.6.43 or later.
  ◆ If the source system is running version 5.6.47, the destination system must be running version 5.6.47 or later.

◆ For systems that use versions earlier than 5.6.47, and to avoid CIFS client timeouts when modifying deduplicated files, the system did not deduplicate any file over 200 MB in size. Starting with version 5.6.47, there is no limit to the maximum size of a file that can be deduplicated. For NFS-only environments, this value can be set higher to potentially achieve greater space savings.

◆ You might not want to use Data Movers that have heavy usage for deduplication. The deduplication scan and ingest process adaptively throttles itself when the Data Mover is very busy. Therefore, a Data Mover that maintains a high level of usage cannot scan or deduplicate files as quickly as a less busy system. Accessing deduplicated files also uses more system resources than accessing normal files. This additional load for deduplication access may negatively impact other client access if the system is already very busy.

◆ VNX deduplication does not deduplicate data across or between file systems.

◆ Multi-Path File System (MPFS) may be used to access file systems on which deduplication is enabled. However, the MPFS client transparently falls back to standard CIFS or NFS when accessing deduplicated files.

◆ Deduplication-enabled file systems can be backed up by using VNX NDMP Volume Backup (NVB) and restored in full by using the full destructive restore (FDR) method.
However, a single file restore or a file-by-file restore of deduplicated files from NVB backups is not supported and will be rejected by the system. Perform a single file restore or a file-by-file restore by using local or remotely replicated EMC SnapSure™ checkpoints instead of by using NVB backups.

- Systems that use version 6.0 or later support the NDMPcopy utility. NDMPcopy, which is enabled by default, allows an administrator to copy files and directories from one file system to another while preserving all files and directory meta-data (for example, CIFS and NFS security attributes), FileMover stub files, and deduplicated files. The NDMPcopy data flow is more efficient than a traditional client-based copy operation. Data is moved directly from the source to the destination Data Mover instead of from the source to a client machine and then to the destination. You cannot use NDMPcopy to copy deduplicated files from a system that is running version 5.6.47 or later to a system that is running version 5.6.46 or earlier.

- Systems that use version 5.6.47 or later support the space-reduced backup feature for NDMP. This feature, which is enabled by default, allows for faster backups and uses less space on tape by transferring the files to tape in their compressed format. You cannot restore deduplicated files from an NDMP backup performed on a system that is running version 5.6.47 or later to a system that is running version 5.6.46 or earlier, unless you disable this feature. When the space-reduced backup feature is disabled, deduplicated files are decompressed in memory and written to the tape as a normal online file. The files can be restored to earlier versions of the system.

- Restoring space-reduced files from a Portable Archive Interchange (PAX)-based NDMP backup into a file system that does not currently have deduplication enabled causes the deduplication state to be set to suspended for that file system.

- If deduplication is enabled on a file system that contains iSCSI LUNs, the iSCSI LUNs will not be deduplicated.

- VNX deduplication does not process or affect alternate data streams (also known as named attributes) associated with files and directories in the file system.

- A file system cannot contain both iSCSI LUN snapshots and file clones. The system prevents the creation of the following:
  - iSCSI LUNs in a file system that contains fast clones of files.
  - Fast clones of files in file systems that contain iSCSI LUNs.

- Restrictions on fast clones and the base image:
  - For an NDMP/PAX backup, each fast clone is backed up in its entirety. On restore, the files are not fast clones.
  - Space-reduced backups do not apply to fast clones because they are not compressed.
  - Exclude virtual machine (VM) disk files from system backups by using the NDMP exclude list feature. Back up these images by the computer that hosts the VM virtual disk files. Using the NDMP exclude list feature is not enforced. You can back up these images in their entirety with NDMP/PAX.
User interface choices

The system offers flexibility in managing networked storage based on your support environment and interface preferences. This document describes how to configure deduplicated files and a deduplication-enabled file system by using the VNX command line interface (CLI). You can also perform all of these tasks by using the Unisphere software.

The Unisphere software online help contains additional information about managing your system.

*Installing Management Applications on VNX for File* includes instructions on launching the Unisphere software.

The *VNX Release Notes* contain additional, late-breaking information about VNX management applications.

Related information

Specific information related to the features and functionality described in this document are included in:

- *EMC VNX Command Line Interface Reference for File*
- Online VNX man pages
- *Parameters Guide for VNX for File*

**EMC VNX documentation on the EMC Online Support website**

The complete set of EMC VNX series customer publications is available on the EMC Online Support website. To search for technical documentation, go to [http://Support.EMC.com](http://Support.EMC.com). After logging in to the website, click the VNX Support by Product page to locate information for the specific feature required.

**VNX wizards**

Unisphere software provides wizards for performing setup and configuration tasks. The Unisphere online help provides more details on the wizards.
The concepts and planning considerations to understand deduplication are:

- Overview on page 16
- Deduplication states on page 17
- Duplicate file detection on page 17
- Space-reduced backups on page 18
- CIFS compressed file attribute on page 18
- Externally triggered deduplication on page 19
- Cloning capabilities on page 19
- Planning application integration on page 19
- Automatic resource management on page 24
Overview

For each file system, file-level deduplication gives the Data Mover the ability to:

- Process files in order to compress them.
- Share the same instance of the data only if they happen to be identical.

Deduplication functionality operates on whole files and is applicable to files that are static or nearly static and that use one of the following last-access times:

- Greater than 30 days (default value in systems that use version 5.6.43 through 5.6.46)
- Greater than 15 days (default value in systems that use version 5.6.47 or later).

During deduplication, each deduplication-enabled file system on the Data Mover is scanned for files that match specific criteria, such as last-access time and a modification time older than a certain date. After a file that matches the criteria is found, the file data is deduplicated, and compressed if appropriate. Different instances of the file can have different names, security attributes, and timestamps. None of the metadata is affected by deduplication. When a user reads a file that was deduplicated, the system retrieves the data so that the NAS clients are unaware that the data was deduplicated.

Read operations decompress the portion of the file requested on-the-fly, in memory. Read operations do not cause the file to be decompressed on disk.

When a user writes to a deduplicated file, the Data Mover stores the new or modified data in addition to the original data until the amount of new or modified data exceeds a threshold relative to the amount of original data or current logical file size. This effectively removes the limit on the maximum size of a file that can be deduplicated that existed in systems that use versions earlier than 5.6.47.

The system can select files for deduplication by using the file selection policies defined for the file system and periodic file system scans. These files are decompressed and reduplicated on disk when the number of individual changed blocks plus the number of blocks in the corresponding deduplicated file is equal to or larger than the original amount of data or current logical file size, whichever is smaller. This accounts for when the file is truncated to be smaller than it was when deduplicated. This means that a modified deduplicated file will never be larger on disk that in its original, not deduplicated, form. Having been decompressed and reduplicated, the file may be eligible to be deduplicated again during the next scan operation.

For files that were explicitly selected for deduplication by using the FileMover API through the VMware® vCenter™ plug-in or by using the CIFS compressed file attribute, the file is reprocessed (compressed and deduplicated) when the number of written blocks for a file exceeds 20 percent of the original amount of data or current logical file size. Reduplication is also scheduled when a truncate-down operation for a file results in its size becoming 3 percent or more below the compressed file size. In effect, the system interprets an instruction to deduplicate a file by using the CIFS compressed file attribute or vCenter plug-in as an instruction to keep that file deduplicated regardless of any changes made to it.
Deduplicating a file, and then reduplicating that file, may result in a file that has fewer bytes than the original file. When reduplicating files, the system does not allocate disk space for blocks that contain all zeros, therefore allowing space savings. The byte size reported for these files is less than their file size.

**Deduplication states**

There are three possible states for deduplication: on, off, or suspended:

- **On** — Enable deduplication processing on a file system. Deduplication is the process used to compress redundant data, which allows space to be saved on a file system. Setting the state to on schedules the file system to be the next file system that is scanned. If there are no active scans, it starts immediately.

- **Off** — Undo all deduplication processing. Do not perform any new space reduction. Any data that was deduplicated is now reduplicated, which is the process used to restore a file that was deduplicated to its original condition. This process may take some time. If reduplication fails, such as when the system detects that the file system is too small to contain the reduplicated data, then the state transitions to the suspended state. The file system is left in a consistent, usable state, and an error message is sent to the server’s event log. If reduplication succeeds, then it remains in the off state.

- **Suspended** — Suspend deduplication processing on a file system. The state transitions to the suspended state. No new space reduction is performed by the scanning process. However, CIFS compression is still allowed and existing space-reduced files remain as is. The system still periodically scans the file system to update the deduplication statistics for the file system.

**Duplicate file detection**

To detect identical files, file-level deduplication uses either:

- SHA-1 hash calculations (default).
- SHA-1 hash calculations followed by byte-by-byte data comparisons if the hash calculations match.

The byte-by-byte data comparison method is used as an additional check to compare each byte of files to determine whether they are duplicates. It is slower than SHA-1 hashing alone but may be preferred if you are concerned about using a SHA-1 hash-based duplicate data detection method.

However, if you do not want to use SHA-1 calculations or byte-by-byte comparisons, the duplicate detection method can be disabled, which still allows files to be compressed for space savings. Set duplicate data detection method for deduplication on page 43 contains instructions on modifying the duplicate detection method setting.

If the system detects over 65,535 identical copies of a specific deduplicated file on the same file system, it will no longer deduplicate that specific file. However, it will still compress the data of any additional files that match this working text.
Space-reduced backups

VNX File Deduplication and Compression propagates space-savings to PAX-based NDMP backups that are similar to file system space-savings:

◆ NDMP is able to back up the compressed version of the file which allows less space to be used on tape than if the file had been backed up in its normal format. However, file-level deduplication is not performed.

◆ Backup-to-tape of compressed files may take less time overall, as the amount of physical data which needs to be read from disk and transferred to tape may be smaller than if the files were in their normal state.

◆ NDMP also restores the compressed file in its compressed state.

◆ Configure NDMP backups to set the option for creating a checkpoint in a file system. This ensures that all of the data that is present at the start of the operation is backed up.

A full file system restore of compressed files can use a smaller size file system than that of the source file system as its recover target. Restoring compressed files can free up space achieved by deduplication when all the data is transferred from its original file system to a file system of a smaller size. The destination file system overhead, restore operation overhead, and the amount of physical data that was backed up to tape determines how small the target file system can be. The physical data that was backed up to tape is reported by the backup application.

CIFS compressed file attribute

By using the CIFS compressed file attribute, Windows CIFS clients can deduplicate and reduplicate individual files on the system. By default, the system marks deduplicated files with the CIFS compressed file attribute. Windows Explorer, by default, shows the names of files marked with the CIFS compressed file attribute in a different color (the default is blue). The filename coloring functionality in Windows Explorer can be disabled if necessary. Microsoft KnowledgeBase article ID 307987 describes how to disable the filename coloring functionality in Windows Explorer. The marking of deduplicated files with the CIFS compressed file attribute can be disabled on the system. Set CIFS compression for deduplication on page 46 describes how to disable this attribute.

Directories can also be marked as deduplicated or reduplicated by using the CIFS compressed file attribute.

If a file is deduplicated, the system tries to keep it in the deduplicated state, even when it is written to. You can write to a file during the deduplication operation. As soon as deduplication is turned off, CIFS compression is no longer allowed until the administrator turns on deduplication. CIFS compression is also allowed if deduplication has been suspended.
Externally triggered deduplication

If a file stored in the VNX for file NFS data store is marked to be deduplicated either manually by using the CIFS compression file attribute or by using the FileMover API, if the file is modified, then the system tries to keep it deduplicated. However, if the file is modified so that the size of the changed blocks plus the number of blocks in the corresponding deduplicated file is larger than the logical file size, then the file is deduplicated again to save as much space as possible. This is in contrast to files that are automatically selected to be deduplicated by the system itself. These files are reduplicated after the threshold is passed.

Cloning capabilities

Full cloning creates full copies of files. Fast cloning is the ability to create writeable snapshots of files. A fast clone, also known as a thin clone, shares the original blocks with a source file and saves only the changed blocks.

In VMware environments, full clones are used to rapidly create copies of a VM, and fast clones are used to rapidly create snapshot copies of a VM. Fast cloning is also useful in Virtual Desktop Image (VDI) implementations where the VDIs are essentially identical. Use of fast clones to create virtual desktops is similar to deduplicating the desktop images and dramatically reduces the storage required for VDI implementations.

A vCenter plug-in that allows a VMware administrator to provision NFS datastores from the system and make full or fast clones of VMs is available. To access the vCenter plug-in, go to the EMC Online Support website. After logging in, click Search or Support by Product to locate information for the vCenter plug-in.

Planning application integration

The applications used to ensure a successful integration with VNX File Deduplication and Compression are:

- Backup and restore on page 20
- VNX File System Migration on page 21
- VNX file-level retention on page 21
- VNX FileMover archiving on page 21
- File system space usage on page 21
- Multi-Path File System on page 22
- Point-in-time views of the file system on page 22
- Quotas on page 23
- Replication on page 23
Backup and restore

This section discusses VNX NDMP Volume Backup (NVB) and local area network (LAN) and VNX NDMP (non-NVB) backups.

**VNX NDMP Volume Backup**

VNX File Deduplication and Compression-enabled file systems can be backed up by using NVB (also known as Volume-Based Backup or VBB) and restored in full by using the full destructive restore (FDR) method. Because NVB operates at the block level (while preserving the history of which files it backs up), backing up a deduplicated file system does not cause any data reduplication. The data in the file system is backed up in its reduced form. The benefits of the space-saving storage efficiency realized in the production file system, from both compression and file-level deduplication, flow through to backups.

However, a single file restore or a file-by-file restore of deduplicated files from NVB backups is not supported and will be rejected by the system. NVB backups of deduplicated file systems must be used as part of a strategy where a single file restore or file-by-file restore is done from locally or remotely replicated SnapSure checkpoints, and not from “tape.”

The VBB facility’s skipDedupFiles parameter setting controls whether a restore from a VBB or NVB backup of a deduplicated file system fails or whether deduplicated files are skipped:

- When set to 0 (the default setting), the VBB facility’s skipDedupFiles parameter causes a restore from a VBB or NVB backup of a deduplicated file system to fail while attempting to restore deduplicated files. This also fails to restore any non-deduplicated files during the restore process.

- When set to 1, the VBB facility’s skipDedupFiles parameter instructs the restore process to skip all deduplicated files, therefore restoring any subsequent non-deduplicated files during the restore process.

The *Parameters Guide for VNX* provides further information.

**Network (LAN) and VNX NDMP (non-NVB) backups**

When backed up over the network by using either CIFS or NFS, space-reduced files are reduplicated to their original size for transfer to the backup application, although the data is not reduplicated on disk. The benefits of the space-saving storage efficiency realized in the production file system will not flow through to backups when using network-based backups of deduplicated file systems.

With PAX-based NDMP, however, the compressed version of the file is backed up, which allows less space to be used on tape than if the file was backed up in its normal format. A file is also restored from a PAX-based NDMP backup in its compressed state. A restore operation does not reduplicate the file, and is less likely to cause “file system full” errors.
VNX File System Migration

Deduplication cannot be used with VNX File System Migration (also known as CDMS).

You cannot enable the deduplication functionality on a migration file system (MGFS).

Deduplication cannot be used to migrate data out of a file system while you are using VNX File System Migration to migrate data into it. You can only enable deduplication on a file system after you have finished migrating data and converted the file system to a Universal Extended File System (UxFS).

After completing a migration and prior to enabling the deduplication functionality for the first time, reboot all Data Movers.

VNX file-level retention

VNX File Deduplication and Compression can be enabled on both enterprise and compliance styles of VNX file-level retention (FLR) file systems without compromising the protection offered to the data that the file systems contain. Files, including locked files, contained on both styles of FLR file systems can be deduplicated.

VNX FileMover archiving

VNX File Deduplication and Compression is transparent to VNX FileMover archiving. The two features can be used together to maximize the storage efficiency of the file storage solution. Any files archived from a file system on which VNX File Deduplication and Compression is enabled are written to and read from the archive system in their un-deduplicated form. However, the archive system may deduplicate the archived data. File systems used as repositories for archived data are good candidates for VNX File Deduplication and Compression.

File system space usage

When the file system is configured to auto-extend, the deduplication process aborts if the space usage is greater than the configured auto-extension threshold minus 1 percent (default value in systems that use version 5.6.47 or later) or minus 5 percent (default value in systems that use versions earlier than 5.6.47). Aborting deduplication prevents the file system from unexpectedly auto-extending. For example, if the auto-extend high water mark (HWM) value is set to 90 percent, the deduplication process aborts after the file system reaches 89 percent full (systems that use version 5.6.47 or later) or 85 percent full (systems that use versions earlier than 5.6.47).

If the file system is not configured to auto-extend, the deduplication process aborts if the file system usage is equal to or greater than 99 percent (default value in systems that use
version 5.6.47 or later) or equal to or greater than 95 percent (default value in systems that use versions earlier than 5.6.47).

The behavior described above ensures that the deduplication process has enough free space in which to work, and thereby reduces the need to automatically extend the file system. Known problems on page 53 describes workarounds.

Multi-Path File System

When accessing deduplicated files, Multi-Path File System (MPFS) clients use either standard CIFS or NFS, not MPFS accelerated I/O.

Note: An application on the MPFS client is not affected, except that it will likely experience some performance degradation if it accesses a migrated file.

Point-in-time views of the file system

The deduplication process releases space in the production file system immediately. However, it may cause blocks to be copied to the SnapSure save volume (SavVol) in the process. Deduplicating data associated with a file involves copying the data within the file system so it can be compressed and deduplicated. Because SnapSure checkpoints copy changed blocks to the SavVol on first write, the blocks that are deduplicated may need to be copied to the SavVol in order to preserve a previous checkpoint point-in-time view of the file system. These blocks are freed when the corresponding checkpoint gets deleted or refreshed and are then available for re-use by other checkpoints. How many blocks will need to be copied to the SavVol during the deduplication process is a function of how full the file system is, the rate of change in it, and so on, and is difficult to predict. By default, the system is configured to abort deduplication operations on a file system before it causes the SavVol to extend. This avoids the SavVol expanding due to deduplication activity. If the deduplication process is aborted in this way, an alert is generated that explains what happened. The system administrator can choose to extend the SavVol or simply let the deduplication process execute again on its next scheduled run.

The SavVol high water mark (threshold) is the percentage of the configured SavVol auto-extension threshold that can be used during deduplication. After the specified amount of SavVol is used, deduplication stops on the specified file system. By default, this value is 90 percent and the SavVol auto-extension is also 90 percent. This option will apply when the SavVol is 81 percent full (90 * 90).

Note: When VNX File Deduplication and Compression is first enabled on a file system with a SavVol that is relatively full, the system may not be able to achieve the maximum space-savings the first time the deduplication process is run. You may need to let the deduplication process run multiple times before achieving the maximum space-savings. Known problems on page 53 describes workarounds.
Quotas

When you set file system quotas, those quotas monitor and control the usage of the file systems on that system.

The system can track user, group, and tree quotas by using either of two quota policies: blocks or file size. By default, the quota policy is set to blocks. When configured to use the blocks quota policy, the system calculates quota usage by counting the number of file system blocks that each file occupies on disk. For example, a 1 KB file counts as 8 KB in the quota because the file consumes one 8 KB block on disk.

When the system deduplicates a file, the block count is subtracted from the applicable quotas from the file. Similarly, when a file is reduplicated, the usage of a block-based quota will increase. When viewing the bytes used for a deduplicated file (by using the du command from NFS or by viewing the Size on disk field from Windows Explorer), the compressed size of the file is reported. However, this compressed size is not charged to a user's block quota because the data may be shared. Therefore, a deduplicated file counts only as one 8 KB block toward any applicable quotas when the blocks quota policy is used.

When configured to use the filesize quota policy, the system calculates quota usage by counting the logical size of files. Because deduplication activity does not effect the logical size of files, it has no effect with the filesize quota policy. Therefore, even though a file is compressed or shared, each instance of it counts as the full logical size towards any applicable quotas.

When calculating the storage required in the file system, remember that each deduplicated file consumes a minimum of one inode, regardless of the quota policy setting.

In addition, the VNX system will allocate an additional inode in the file system while it:

- Deduplicates a file in the file system.
- Reduplicates the content of a file.

These temporary inodes count toward any inode-based quotas in use. If inode quotas prevent reduplication and write operations to deduplicated files, the server log displays a message to indicate either that a hard quota was reached or exceeded, or that the quota was exceeded and the reduplication of the file failed. Do not allow the system to reach hard-inode quota limits to avoid disrupting deduplication-related activity.

The server_df command always displays the inodes and blocks in use in the file system.

Using Quotas on VNX provides additional information about setting quotas.

Replication

Deduplicating the contents of a file system before it is replicated by using VNX Replicator can greatly reduce the amount of data that has to be sent over the network as part of the initial baseline copy process. After replication and deduplication are running together, the impact of deduplication on the amount of data transferred over the network will depend...
on the relative timing of replication updates and deduplication runs. In all but the most extreme circumstances, replication updates will be more frequent than deduplication scans of a file system. New and changed data in the file system will almost always be replicated in its non-deduplicated form first. Any subsequent deduplication of that data will prompt additional replication traffic due to the changes within the file system. This effect will be true of any deduplication solution that post-processes data and updates remote replicas of the data set more frequently than the deduplication process is run. The space-savings realized by the production file system will be reflected on the destination file system.

Deduplication can add a maximum of 25 MB/s of write activity during processing. If you use a lower bandwidth network, ensure that your replication bandwidth can accommodate the additional load, and that your service level agreements (SLAs) are being met. If this impacts your VNX Replicator SLA in your environment, you can lower the deduplication rate to about 2 MB/s by changing the deduplication CPU usage low water mark and CPU usage high water mark settings to 0.

*Using VNX Replicator* provides additional information about using replication.

### Automatic resource management

File systems on each Data Mover are scanned one at a time by a single threaded process to select files for deduplication processing. Different Data Movers can be scanned simultaneously. While the scanning operation is in progress, the system monitors the CPU usage of the Data Mover.

- If the Data Mover CPU usage exceeds the specified CPU % high water mark threshold, the system will throttle the deduplication activity to avoid impacting client activity. The CPU % high water mark threshold is the average percent of CPU usage that can be used during the deduplication process before triggering a slow throttle mode.

- Deduplication activity remains throttled until the CPU usage falls below the specified low water mark threshold. The CPU % low water mark threshold is the average percent of CPU usage that can be used during the deduplication process at which full throttle mode is re-entered. This allows the deduplication process to use the CPU that is free.
The task to configure a deduplication-enabled file system is:

- Enable file system deduplication on page 26
Enable file system deduplication

**Note:** If you want to exclude file extensions or pathnames from the deduplication process, set up the file extension exclusion list or pathname exclusion list before running deduplication for the first time. If you add exclusions after deduplication has been run on a file system, the new exclusions take effect from that point forward and do not affect what has already been deduplicated. **Exclude file extensions from deduplication** on page 36 and **Exclude pathnames from deduplication** on page 37 contain instructions on excluding file extensions and pathnames.

<table>
<thead>
<tr>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>To enable deduplication space-reduction processing, use this command syntax:</td>
</tr>
<tr>
<td>$ fs_dedupe -modify {&lt;fs_name&gt;</td>
</tr>
<tr>
<td>where:</td>
</tr>
<tr>
<td>&lt;fs_name&gt; = name of the file system</td>
</tr>
<tr>
<td>&lt;fs_id&gt; = identifier of the file system</td>
</tr>
<tr>
<td><strong>Example:</strong></td>
</tr>
<tr>
<td>To enable deduplication space-reduction processing on file system fs_ufs1, type:</td>
</tr>
<tr>
<td>$ fs_dedupe -modify fs_ufs1 -state on</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Done</td>
</tr>
</tbody>
</table>

**Note**

This feature is disabled by default. Setting the deduplication state to on prompts an immediate scan of a file system, even if the state was already set to on. If another scan is in progress on the Data Mover, it is aborted. The file system must have at least 1 MB of free space before deduplication can be enabled. If there is not enough free space, an error message is generated and the server log is updated.
Note: Deduplication settings can be modified for both a Data Mover and a file system, unless otherwise noted. Modifying deduplication settings for a specific file system overrides the global Data Mover deduplication settings. If you clear a file system setting, it is reset to the global Data Mover setting.

The tasks to manage a deduplication-enabled file system are:

- Manually test deduplication on page 28
- Display deduplication-enabled file systems on page 30
- List all deduplication-enabled file systems on page 31
- Query settings on deduplication-enabled file systems on page 32
- Suspend file system deduplication on page 34
- Reduplicate files in a file system on page 35
- Exclude file extensions from deduplication on page 36
- Exclude pathnames from deduplication on page 37
- Set access time for deduplication on page 38
- Set maximum file size for deduplication on page 39
- Set minimum file size for deduplication on page 40
- Set minimum scan interval for deduplication on page 41
- Set modification time for deduplication on page 42
- Set duplicate data detection method for deduplication on page 43
- Set backup data threshold for deduplication on page 45
- Set CIFS compression for deduplication on page 46
- Reset deduplication default settings on page 48
Manually test deduplication

1. Create a small, test file system.
2. Copy identical files to more than one directory on the file system.
3. To review the amount of used and available disk space for the file system, use this command syntax:

   \[
   \text{
   $ server\_df \ <movername> \ <fs\_name>
   
   where:
   
   \<movername> = name of the Data Mover
   \<fs\_name> = name of the file system
   
   Example:
   
   To review the amount of used and available disk space for file system fs\_ufs1, type:
   
   $ server\_df \ server\_2 \ fs\_ufs1
   
   Output:
   
   server\_2 :
   Filesystem  kbytes used  avail  capacity Mounted on
   fs\_ufs1  1007984  519792  488192  52\%  /fs\_ufs1
   \]
4. To disable the access time, use this command syntax:

   \[
   $ \text{fs\_dedupe -modify \ <fs\_name> -access\_time 0}
   
   Example:
   
   To disable the access time for file system fs\_ufs1, type:
   
   $ \text{fs\_dedupe -modify fs\_ufs1 -access\_time 0}
   
   5. To disable the modification time, use this command syntax:

   \[
   $ \text{fs\_dedupe -modify \ <fs\_name> -modification\_time 0}
   
   Example:
   
   To disable the modification time for file system fs\_ufs1, type:
   
   $ \text{fs\_dedupe -modify fs\_ufs1 -modification\_time 0}
   
   6. To verify the settings, use this command syntax:

   \[
   $ \text{fs\_dedupe -info \ <fs\_name>}
   
   Example:
   
   To verify the settings on file system fs\_ufs1, type:
   
   $ \text{fs\_dedupe -info fs\_ufs1}
   
   Output:
7. To enable deduplication on the file system, use this command syntax:

```bash
$ fs_dedupe -modify <fs_name> -state on
```

Example:

To enable deduplication on file system fs_ufs1, type:

```bash
$ fs_dedupe -modify fs_ufs1 -state on
```

8. To review the amount of used and available disk space for the file system after deduplication is completed, use this command syntax:

```bash
$ server_df <movername> <fs_name>
```

Example:

To review the amount of used and available disk space for file system fs_ufs1, type:

```bash
$ server_df server_2 fs_ufs1
```

Output:

```
server_2:
Filesystem  kbytes  used  avail  capacity Mounted on
fs_ufs1    1007984  145176  862808   14%  /fs_ufs1
```

9. Reset the access time and modification time to the appropriate values.
Display deduplication-enabled file systems

Action

To display deduplication-enabled file systems, use this command syntax:

```
$ nas_fs -info [-size] {all|<fs_name>|id=<fs_id>}
```

where:

- `<fs_name>` = name of the file system
- `<fs_id>` = identifier of the file system

To display the deduplication-enabled file system `fs_ufs1`, type:

```
$ nas_fs -info fs_ufs1
```

Output

```
id  = 401
name = fs_ufs1
acl  = 0
in_use = True
type = uxf
worm = off
volume = v1232
pool = symm_std_rdf src
member_of = root_avm fs_group_8
rw_servers = server_2
ro_servers =

auto_ext = no,thin=no
deduplication = On
stor_devs = 000190100563-0277
disks = d61
disk=d61 stor_dev=000190100563-0277 addr=c0t4l6-00-0 server=server_2
disk=d61 stor_dev=000190100563-0277 addr=c48t4l6-00-0 server=server_2
```
List all deduplication-enabled file systems

**Action**

To list all deduplication-enabled file systems, type:

```
$ fs_dedupe -list
```

**Output**

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>state</th>
<th>status</th>
<th>time_of_last_scan</th>
<th>original_data_size</th>
<th>usage</th>
<th>space_saved</th>
</tr>
</thead>
<tbody>
<tr>
<td>401</td>
<td>fs_ufs1</td>
<td>On</td>
<td>Idle</td>
<td>Wed Jun 24 10:31:16 EST 2009</td>
<td>352 MB</td>
<td>119%</td>
<td>93 MB (26%)</td>
</tr>
</tbody>
</table>
Query settings on deduplication-enabled file systems

Action

To query Data Mover settings for deduplication-enabled file systems, use this command syntax:

```
$ fs_dedupe -default -info {<movername>|-all}
```

where:

- `<movername>` = name of the Data Mover

Example:

To query the settings on server_2, type:

```
$ fs_dedupe -default -info server_2
```

Output

Server parameters:

```
server_2
Case Sensitive = yes
Duplicate Detection Method = sha1
Access Time = 30
Modification Time = 30
Minimum Size = 20
Maximum Size = 200
File Extension Exclude List = .jpg:.db:.pst
Minimum Scan Interval = 1
SavVol Threshold = 90
Backup Data Threshold = 90
CPU % Usage Low Water Mark = 25
CPU % Usage High Water Mark = 90
Cifs Compression Enabled = yes
```

Action

To query the file system settings for deduplication-enabled file systems, use this command syntax:

```
$ fs_dedupe -info {<fs_name>|id=<fs_id>|-all}
```

where:

- `<fs_name>` = name of the file system
- `<fs_id>` = identifier of the file system

Example:

To query the settings on fs_ufs1, type:

```
$ fs_dedupe -info fs_ufs1
```
### Output

<table>
<thead>
<tr>
<th>Id</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>fs_ufs1</td>
</tr>
<tr>
<td>Deduplication</td>
<td>On</td>
</tr>
<tr>
<td>Status</td>
<td>Idle</td>
</tr>
</tbody>
</table>

As of the last file system scan (Wed Jun 24 10:31:16 EST 2009):

- Files scanned = 1986265
- Files deduped = 606472 (30% of total files)
- File system capacity = 1032575 MB
- Original data size = 875459 MB (84% of current file system capacity)
- Space saved = 341622 MB (39% of original data size)

File system parameters:

- Case Sensitive = yes
- Duplicate Detection Method = sha1
- Access Time = 30
- Modification Time = 30
- Minimum Size = 20
- Maximum Size = 200
- File Extension Exclude List = .jpg:.db:.pst
- Minimum Scan Interval = 1
- SavVol Threshold = 90
- Backup Data Threshold = 90
- Cifs Compression Enabled = yes
- Pathname Exclude List = root;etc

### Note

The Status option can be Idle, Scanning, or Reduplicating.
Suspend file system deduplication

Setting the deduplication state to suspended stops deduplication processing on a file system. No new space reduction is performed. Existing space-reduced and compressed files remain unchanged.

Action

To suspend deduplication space-reduction operation but keep existing space-reduced and compressed files, use this command syntax:

```
$ fs_dedupe -modify {<fs_name>|id=<fs_id>} -state suspended
```

where:

- `<fs_name>` = name of the file system
- `<fs_id>` = identifier of the file system

Example:

To suspend deduplication processing on file system `fs_ufs1`, type:

```
$ fs_dedupe -modify fs_ufs1 -state suspended
```

Output

Done
Reduplicate files in a file system

Reduplicating, or undoing deduplication processing, does not perform any new space reduction. Any data that was deduplicated is now reduplicated, which is the process used to restore a file that was deduplicated to its original condition. This process may take some time.

- If reduplication fails, such as when the system detects that the file system is too small to contain the reduplicated data, then the following occurs:
  - State transitions to the suspended state.
  - File system remains in a consistent, usable state.
  - Error message appears in the server’s event log.

- If reduplication succeeds, then the file system remains in the off state.

**Action**

To reduplicate files and remove all space-reduced data, use this command syntax:

```
$ fs_dedupe -modify {<fs_name>|id=<fs_id>} -state off
```

where:

- `<fs_name>` = name of the file system
- `<fs_id>` = identifier of the file system

**Example:**

To remove all space-reduced data from file system `fs_ufs1`, type:

```
$ fs_dedupe -modify fs_ufs1 -state off
```

**Output**

Done
Exclude file extensions from deduplication

**Note:** If you want to exclude certain file extensions from the deduplication process, set up the file extension exclusion list before running deduplication for the first time. If you add exclusions after deduplication has been run on a file system, the new exclusions take effect from that point forward and do not affect files that were deduplicated in previous scans.

<table>
<thead>
<tr>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>To exclude file extensions from the deduplication process on a Data Mover, use this command syntax:</td>
</tr>
<tr>
<td>$ fs_dedupe -default -set {&lt;movername&gt;</td>
</tr>
<tr>
<td>where:</td>
</tr>
<tr>
<td>&lt;movername&gt; = name of the Data Mover.</td>
</tr>
<tr>
<td>&lt;ext_list&gt; = semicolon-delimited or colon-delimited list of filename extensions to be excluded from deduplication. Filename extensions can be any valid alphanumeric character. Each extension must include the leading dot. The list of filename extensions must be 64 bytes or less. The default value is ‘’ (empty).</td>
</tr>
<tr>
<td>Example:</td>
</tr>
<tr>
<td>To exclude MP3 and Zip files from the deduplication process on server_2, type:</td>
</tr>
<tr>
<td>$ fs_dedupe -default -set server_2 -file_ext_exclude_list .mp3;.zip</td>
</tr>
<tr>
<td>Output</td>
</tr>
<tr>
<td>Done</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>To exclude file extensions from the deduplication process on a file system, use this command syntax:</td>
</tr>
<tr>
<td>$ fs_dedupe -modify {&lt;fs_name&gt;</td>
</tr>
<tr>
<td>where:</td>
</tr>
<tr>
<td>&lt;fs_name&gt; = name of the file system.</td>
</tr>
<tr>
<td>&lt;fs_id&gt; = identifier of the file system.</td>
</tr>
<tr>
<td>&lt;ext_list&gt; = semicolon-delimited or colon-delimited list of filename extensions to be excluded from deduplication. Filename extensions can be any valid alphanumeric character. Each extension must include the leading dot. The list of filename extensions must be 1,024 bytes or less. The default value is ‘’ (empty).</td>
</tr>
<tr>
<td>Example:</td>
</tr>
<tr>
<td>To exclude MP3 and Zip files from the deduplication process on fs_ufs1, type:</td>
</tr>
<tr>
<td>$ fs_dedupe -modify fs_ufs1 -file_ext_exclude_list .mp3;.zip</td>
</tr>
<tr>
<td>Output</td>
</tr>
<tr>
<td>Done</td>
</tr>
</tbody>
</table>
Exclude pathnames from deduplication

You can exclude pathnames only on a file system.

Note: If you want to exclude certain pathnames from the deduplication process, set up the pathname exclusion list before running deduplication for the first time. If you add exclusions after deduplication has been run on a file system, the new exclusions take effect from that point forward and do not affect files that were deduplicated in previous scans.

Action

To exclude pathnames for VNX File Deduplication and Compression on a file system, use this command syntax:

$ fs_dedupe -modify {<fs_name>|id=<fs_id>} -pathname_exclude_list <path_list>

where:

<fs_name> = name of the file system.
<fs_id> = identifier of the file system.
<path_list> = semicolon-delimited list of relative pathnames, in UTF-8 format, to be excluded from deduplication. Any directory below a specified pathname is excluded from deduplication. You can specify a maximum of 10 pathnames and each one can be up to 1024 bytes. This option does not support regular expressions. The default value is ‘’ (empty).

Example:

To exclude the user1/data pathname on fs_ufs1, type:

$ fs_dedupe -modify fs_ufs1 -pathname_exclude_list user1/data

Output

Done

Note

- The semicolon (;) is used as a delimiter when multiple pathnames are included in a list. It cannot be used as part of a pathname because escape characters are not supported.
- If a semicolon is used as a delimiter, the pathname list must be enclosed in either single quotes or double quotes.
- The forward slash (/) is always a valid directory delimiter within a single pathname.
- The backward slash (\) is not a valid directory delimiter within a single pathname.
## Set access time for deduplication

### Action

To set the access time for VNX File Deduplication and Compression on a Data Mover, use this command syntax:

```bash
$ fs_dedupe -default -set {<movername>|-all} -access_time <days>
```

where:

- `<movername>` = name of the Data Mover.
- `<days>` = minimum required file age in days based on read access time. Files that have been read within the specified number of days are not deduplicated. This setting does not apply to files with an FLR locked state. Setting this value to 0 disables it. The range of values is 0 to 365, and the default value is 15 days.

Example:

To set the access time to 20 on all Data Movers, type:

```bash
$ fs_dedupe -default -set -all -access_time 20
```

### Output

Done

### Action

To set the access time for VNX File Deduplication and Compression on a file system, use this command syntax:

```bash
$ fs_dedupe -modify {<fs_name>|id=<fs_id>} -access_time <days>
```

where:

- `<fs_name>` = name of the file system.
- `<fs_id>` = identifier of the file system.
- `<days>` = minimum required file age in days based on read access time. Files that have been read within the specified number of days are not deduplicated. This setting does not apply to files with an FLR locked state. Setting this value to 0 disables it. The range of values is 0 to 365, and the default value is 15 days.

Example:

To set the access time to 20 on fs_ufs1, type:

```bash
$ fs_dedupe -modify fs_ufs1 -access_time 20
```

### Output

Done
Set maximum file size for deduplication

**Action**

To set the maximum file size for VNX File Deduplication and Compression on a Data Mover, use this command syntax:

```
$ fs_dedupe -default -set {<movername>|all} -maximum_size <MB>
```

where:

- `<movername>` = name of the Data Mover.
- `<MB>` = file size in megabytes of the largest file to be processed for deduplication. Files larger than this size in megabytes are not deduplicated. Setting this value too high may affect system write performance as the first write operation reduplicates the file in its entirety. Setting this value to 0 disables it. The range of values is 0 to 8388608, and the default value is 8388608 MB.

**Example:**

To set the maximum file size to 1,000,000 MB on server_2, type:

```
$ fs_dedupe -default -set server_2 -maximum_size 1000000
```

**Output**

Done

---

**Action**

To set the maximum file size for VNX File Deduplication and Compression on a file system, use this command syntax:

```
$ fs_dedupe -modify {<fs_name>|id=<fs_id>} -maximum_size <MB>
```

where:

- `<fs_name>` = name of the file system.
- `<fs_id>` = identifier of the file system.
- `<MB>` = file size in megabytes of the largest file to be processed for deduplication. Files larger than this size in megabytes are not deduplicated. Setting this value too high may affect system write performance as the first write operation reduplicates the file in its entirety. Setting this value to 0 disables it. The range of values is 0 to 8388608, and the default value is 8388608 MB.

**Example:**

To set the maximum file size to 1,000,000 MB on fs_ufs1, type:

```
$ fs_dedupe -modify fs_ufs1 -maximum_size 1000000
```

**Output**

Done
Set minimum file size for deduplication

**Action**

To set the minimum file size for VNX File Deduplication and Compression on a Data Mover, use this command syntax:

$ `fs_dedupe -default -set {<movername>|-all} -minimum_size <KB>`

where:

- `<movername>` = name of the Data Mover.
- `<KB>` = file size in kilobytes that limits deduplication. File sizes equal to this value or smaller are not deduplicated. File sizes greater than this value are candidates for deduplication. Setting this value to 0 disables it. This value should not be set lower than 24 KB. The range of values is 0 to 1000, and the default value is 24 KB.

Example:

To set the minimum file size to 30 KB on server_2, type:

$ `fs_dedupe -default -set server_2 -minimum_size 30`

**Output**

Done

---

**Action**

To set the minimum file size for VNX File Deduplication and Compression on a file system, use this command syntax:

$ `fs_dedupe -modify {<fs_name>|id=<fs_id>} -minimum_size <KB>`

where:

- `<fs_name>` = name of the file system.
- `<fs_id>` = identifier of the file system.
- `<KB>` = file size in kilobytes that limits deduplication. File sizes equal to this value or smaller are not deduplicated. File sizes greater than this value are candidates for deduplication. Setting this value to 0 disables it. This value should not be set lower than 24 KB. The range of values is 0 to 1000, and the default value is 24 KB.

Example:

To set the minimum file size to 30 KB on fs_ufs1, type:

$ `fs_dedupe -modify fs_ufs1 -minimum_size 30`

**Output**

Done
Set minimum scan interval for deduplication

**Action**

To set the minimum scan interval for VNX File Deduplication and Compression on a Data Mover, use this command syntax:

```bash
$ fs_dedupe -default -set {<movername>|-all} -minimum_scan_interval <days>
```

where:

- `<movername>` = name of the Data Mover.
- `<days>` = minimum number of days between completing one scan of a file system and before scanning the same file system again. The range of values is 1 to 365, and the default value is 7 days.

**Example:**

To set the minimum scan interval to 14 days on server_2, type:

```bash
$ fs_dedupe -default -set server_2 -minimum_scan_interval 14
```

**Output**

Done

**Action**

To set the minimum scan interval for VNX File Deduplication and Compression on a file system, use this command syntax:

```bash
$ fs_dedupe -modify {<fs_name>|id=<fs_id>} -minimum_scan_interval <days>
```

where:

- `<fs_name>` = name of the file system.
- `<fs_id>` = identifier of the file system.
- `<days>` = minimum number of days between completing one scan of a file system and before scanning the same file system again. The range of values is 1 to 365, and the default value is 7 days.

**Example:**

To set the minimum scan interval to 14 days on fs_ufs1, type:

```bash
$ fs_dedupe -modify fs_ufs1 -minimum_scan_interval 14
```

**Output**

Done
### Set modification time for deduplication

**Action**

To set the modification time for VNX File Deduplication and Compression on a Data Mover, use this command syntax:

```
$ fs_dedupe -default -set {<movername>|-all} -modification_time <days>
```

where:

- `<movername>` = name of the Data Mover.
- `<days>` = minimum required file age in days based on modification time. Files updated within the specified number of days are not deduplicated. Setting this value to 0 disables it. The range of values is 0 to 365, and the default value is 15 days.

**Example:**

To set the modification time to 30 days on server_2, type:

```
$ fs_dedupe -default -set server_2 -modification_time 30
```

**Output**

Done

**Action**

To set the modification time for VNX File Deduplication and Compression on a file system, use this command syntax:

```
$ fs_dedupe -modify {<fs_name>|id=<fs_id}> -modification_time <days>
```

where:

- `<fs_name>` = name of the file system.
- `<fs_id>` = identifier of the file system.
- `<days>` = minimum required file age in days based on modification time. Files updated within the specified number of days are not deduplicated. Setting this value to 0 disables it. The range of values is 0 to 365, and the default value is 15 days.

**Example:**

To set the modification time to 30 days on fs_ufs1, type:

```
$ fs_dedupe -modify fs_ufs1 -modification_time 30
```

**Output**

Done
Set duplicate data detection method for deduplication

**Note:** The duplicate data detection method is set to SHA-1 by default. If you want to modify the duplicate data detection method capability, modify this setting before running deduplication for the first time. If you modify this setting after deduplication has been run on a file system, the modification takes effect from that point forward and does not affect files that were deduplicated in previous scans. Compression remains enabled regardless of which option is set.

### Action

To set the duplicate data detection method capability on a Data Mover, use this command syntax:

```bash
$ fs_dedupe -default -set {<movername>|all} -duplicate_detection_method {sha1|byte|off}
```

**Example:**

To disable duplicate data detection method on server_2, type:

```bash
$ fs_dedupe -default -set server_2 -duplicate_detection_method off
```

### Output

*Done*

### Note

- The sha1 option identifies that a SHA-1 hash is used to detect duplicate data. This is the default method.
- The byte option identifies that to detect duplicate data, a SHA-1 hash calculation is used, followed by a byte-to-byte comparison if the hash calculation matches. This byte option adds considerable overhead, especially for large files.
- The off option identifies that the duplicate detection method is disabled. With this setting, every deduplicated file is considered unique and the only space savings made are accomplished with compression.

### Action

To set the duplicate data detection method capability on a file system, use this command syntax:

```bash
$ fs_dedupe -modify {<fs_name>|id=<fs_id>} -duplicate_detection_method {sha1|byte|off}
```

**Example:**

To disable duplicate data detection method on fs_ufs1, type:

```bash
$ fs_dedupe -default -set fs_ufs1 -duplicate_detection_method off
```
The sha1 option identifies that a SHA-1 hash is used to detect duplicate data. This is the default method.

The byte option identifies that to detect duplicate data, a SHA-1 hash calculation is used, followed by a byte-to-byte comparison if the hash calculation matches. This byte option adds considerable overhead, especially for large files.

The off option identifies that the duplicate detection method is disabled. With this setting, every deduplicated file is considered unique and the only space savings made are accomplished with compression.
Set backup data threshold for deduplication

Action

To set the backup data threshold for VNX File Deduplication and Compression on a Data Mover, use this command syntax:

```
$ fs_dedupe -default -set {<movername>|-all} -backup_data_threshold <percent>
```

where:

- `<movername>` = name of the Data Mover.
- `<percent>` = full percentage that a deduplicated file has to be below in order to trigger space-reduced backups for NDMP. For example, when set to 90, any deduplicated file whose physical size (compressed file plus changed blocks) is greater than 90% of the logical size of the file will have the entire file data backed up without attempting to back it up in a space-reduced format. Any deduplicated file whose physical size is less than 90% of the logical file size will be backed up in a space-reduced format. Setting this value to 0 disables a space-reduced backup. The range of values is 0 to 200, and the default value is 90%.

Example:

To set the backup data threshold to 85% on server_2, type:

```
$ fs_dedupe -default -set server_2 -backup_data_threshold 85
```

Output

Done

Action

To set the backup data threshold for VNX File Deduplication and Compression on a file system, use this command syntax:

```
$ fs_dedupe -modify {<fs_name>|id=<fs_id>} -backup_data_threshold <percent>
```

where:

- `<fs_name>` = name of the file system.
- `<fs_id>` = identifier of the file system.
- `<percent>` = full percentage that a deduplicated file has to be below in order to trigger space-reduced backups for NDMP. For example, when set to 90, any deduplicated file whose physical size (compressed file plus changed blocks) is greater than 90% of the logical size of the file will have the entire file data backed up without attempting to back it up in a space-reduced format. Any deduplicated file whose physical size is less than 90% of the logical file size will be backed up in a space-reduced format. Setting this value to 0 disables a space-reduced backup. The range of values is 0 to 200, and the default value is 90%.

Example:

To set the backup data threshold to 85% on fs_ufs1, type:

```
$ fs_dedupe -modify fs_ufs1 -backup_data_threshold 85
```

Output

Done
Set CIFS compression for deduplication

Action

To set CIFS compression and display for VNX File Deduplication and Compression on a Data Mover, use this command syntax:

$ fs_dedupe -default -set {<movername>|~all} -cifs_compression_enabled {yes|no}

where:

<movername> = name of the Data Mover

Example:

To enable CIFS compression on server_2, type:

$ fs_dedupe -default -set server_2 -cifs_compression_enabled yes

Output

Done

Note

This CIFS compression option allows CIFS clients to manage the deduplication state of files by setting or removing the CIFS compressed file attribute. The default is yes, enable CIFS compression. When set to yes and the deduplication state of the file system is either on or suspended, then CIFS compression is allowed. If the deduplication state is either off or in the process of being turned off, then CIFS compression is not allowed, regardless of whether this option is set to yes.

Action

To set CIFS compression and display for VNX File Deduplication and Compression on a file system, use this command syntax:

$ fs_dedupe -modify {<fs_name>|id=<fs_id>} -cifs_compression_enabled {yes|no}

where:

<fs_name> = name of the file system

<fs_id> = identifier of the file system

Example:

To enable CIFS compression on fs_ufs1, type:

$ fs_dedupe -modify fs_ufs1 -cifs_compression_enabled yes

Output

Done
<table>
<thead>
<tr>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>This CIFS compression option allows CIFS clients to manage the deduplication state of files by setting or removing the CIFS compressed file attribute. The default is yes, enable CIFS compression. When set to yes and the deduplication state of the file system is either on or suspended, then CIFS compression is allowed. If the deduplication state is either off or in the process of being turned off, then CIFS compression is not allowed, regardless of whether this option is set to yes.</td>
</tr>
</tbody>
</table>
Reset deduplication default settings

**Action**

To reset a deduplication default setting on a Data Mover, use this command syntax:

```
$ fs_dedupe -default -clear {<movername>|-all} <param_name>
```

where:

- `<movername>` = name of the Data Mover.
- `<param_name>` = deduplication setting to be reset to the default value. Valid settings are: -access_time; -backup_data_threshold; -case_sensitive; -cifs_compression_enabled; -cpu_usage_high_watermark; -cpu_usage_low_watermark; -duplicate_detection_method; -file_ext_exclude_list; -maximum_size; -minimum_scan_interval; -minimum_size; -modification_time; and -savvol_threshold.

Examples:

To reset the CIFS compression option on server_2, type:

```
$ fs_dedupe -default -clear server_2 -cifs_compression_enabled
```

To reset all options on server_2, type:

```
$ fs_dedupe -default -clear server_2
```

**Note**

The -all option resets the deduplication settings to default values on all Data Movers.

**Action**

To reset a deduplication default setting on a file system, use this command syntax:

```
$ fs_dedupe -clear {<fs_name>|id=<fs_id>} <param_name>
```

where:

- `<fs_name>` = name of the file system.
- `<fs_id>` = identifier of the file system.
- `<param_name>` = deduplication setting to be reset to the default value. Valid settings are: -access_time; -backup_data_threshold; -case_sensitive; -cifs_compression_enabled; -duplicate_detection_method; -file_ext_exclude_list; -maximum_size; -minimum_scan_interval; -minimum_size; -modification_time; and -savvol_threshold.

Examples:

To reset the CIFS compression option on fs_ufs1, type:

```
$ fs_dedupe -clear fs_ufs1 -cifs_compression_enabled
```
### Action

To reset all options on fs_ufs1, type:

```bash
$ fs_dedupe -clear fs_ufs1
```

### Output

Done

### Note

The clearing of a file system setting resets it to the global Data Mover setting.
As part of an effort to continuously improve and enhance the performance and capabilities of its product lines, EMC periodically releases new versions of its hardware and software. Therefore, some functions described in this document may not be supported by all versions of the software or hardware currently in use. For the most up-to-date information on product features, refer to your product release notes.

If a product does not function properly or does not function as described in this document, contact your EMC Customer Support Representative.

*Problem Resolution Roadmap for VNX* contains additional information about using the *EMC Online Support* website and resolving problems.

Topics included are:

- **EMC E-Lab Interoperability Navigator on page 52**
- **Error messages on page 52**
- **Known problems on page 53**
- **EMC Training and Professional Services on page 54**
EMC E-Lab Interoperability Navigator

The EMC E-Lab™ Interoperability Navigator is a searchable, web-based application that provides access to EMC interoperability support matrices. It is available at http://Support.EMC.com. After logging in to the EMC Online Support website, locate the applicable Support by Product page, find Tools, and click E-Lab Interoperability Navigator.

Error messages

All event, alert, and status messages provide detailed information and recommended actions to help you troubleshoot the situation.

To view message details, use any of these methods:

◆ Unisphere software:
  • Right-click an event, alert, or status message and select to view Event Details, Alert Details, or Status Details.

◆ CLI:
  • Type nas_message -info <MessageID>, where <MessageID> is the message identification number.

◆ Celerra Error Messages Guide:
  • Use this guide to locate information about messages that are in the earlier-release message format.

◆ EMC Online Support:
  • Use the text from the error message's brief description or the message's ID to search the Knowledgebase on the EMC Online Support website. After logging in to EMC Online Support, locate the applicable Support by Product page, and search for the error message.
Known problems

Table 2 on page 53 describes known problems that might occur when using deduplication, and presents workarounds.

Table 2. Known problems and workarounds

<table>
<thead>
<tr>
<th>Known problems</th>
<th>Symptom</th>
<th>Workaround</th>
</tr>
</thead>
<tbody>
<tr>
<td>The deduplication process aborts on any file system where the SavVol threshold is reached.</td>
<td>Deduplication process (scanning) fails due to lack of space.</td>
<td>If the deduplication process stops because of a SavVol restriction, you can use one of the following alternatives:</td>
</tr>
<tr>
<td><strong>Note:</strong> Deduplication and compression are still enabled for the file system. However, scanning cannot continue until the SavVol contains enough space to allow the deduplication process to run.</td>
<td></td>
<td>• Let deduplication rerun in a week (or as scheduled) because data may have been purged from the SavVol, which allows more deduplication to occur. This method has the advantage that the SavVol is conserved. However, it may greatly slow down the rate of deduplication.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reduce the minimum scan interval setting to increase the frequency that deduplication runs. This method more quickly uses space in the SavVol as it becomes available.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Delete the oldest checkpoints and force deduplication to rerun by re-enabling deduplication on the file system. This method frees space in the SavVol and allow deduplication to process more data.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Manually extend the SavVol and then immediately restart the deduplication process. This method uses more of the SavVol but allows the user to control the expansion.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Use the fs_dedupe command to change the SavVolThreshold setting that allows the SavVol to extend. This command allows deduplication to complete, but the SavVol will grow up to the size of the compressed version of the data. This allows deduplication to achieve its maximum potential. However, it has the greatest likelihood of expanding the SavVol, which is non-reclaimable.</td>
</tr>
</tbody>
</table>
Table 2. Known problems and workarounds (continued)

<table>
<thead>
<tr>
<th>Known problems</th>
<th>Symptom</th>
<th>Workaround</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not enough space available on the file system to run deduplication.</td>
<td>If the file system is configured to auto-extend, the deduplication process aborts if the space usage is greater than the configured auto-extension threshold minus 5% (default value in systems that use versions earlier than 5.6.47) or minus 1% (default value in systems that use version 5.6.47 or later). If the file system is not configured to auto-extend, the deduplication process aborts if the file system usage is equal to or greater than 95% (default value in systems that use versions earlier than 5.6.47) or equal to or greater than 99% (default value in systems that use version 5.6.47 or later).</td>
<td>Either extend the file system manually, or move some files out of the file system temporarily to let the deduplication process run. After space is available, move the files back again. You can also archive files on a secondary storage to clear space on the primary file system.</td>
</tr>
</tbody>
</table>

Note: Deduplication and compression are still enabled for the file system. However, scanning cannot continue until the file system contains enough space to allow the deduplication process to run.

EMC Training and Professional Services

EMC Customer Education courses help you learn how EMC storage products work together within your environment to maximize your entire infrastructure investment. EMC Customer Education features online and hands-on training in state-of-the-art labs conveniently located throughout the world. EMC customer training courses are developed and delivered by EMC experts. Go to the EMC Online Support website at http://Support.EMC.com for course and registration information.

EMC Professional Services can help you implement your VNX series efficiently. Consultants evaluate your business, IT processes, and technology, and recommend ways that you can leverage your information for the most benefit. From business plan to implementation, you get the experience and expertise that you need without straining your IT staff or hiring and training new personnel. Contact your EMC Customer Support Representative for more information.
C

Common Internet File System (CIFS)
File-sharing protocol based on the Microsoft Server Message Block (SMB). It allows users to share file systems over the Internet and intranets.

compression
Process of encoding data to reduce its size by representing repeating patterns of data by using a smaller number of bits than the original.

D

Data Mover
In VNX for file, a cabinet component that is running its own operating system that retrieves data from a storage device and makes it available to a network client. This is also referred to as a blade.

deduplication
Process used to compress redundant data, which allows space to be saved on a file system. When multiple files have identical data, the file system stores only one copy of the data and shares that data between the multiple files. Different instances of the file can have different names, security attributes, and timestamps. None of the metadata is affected by deduplication.

F

fast clone
A writeable virtual, or thin, point-in-time copy of a file. A fast clone of a file appears as a separate file in the file system, but shares the original blocks with the source file from which it was cloned. Any changes to either the source file or the clone file result in creating additional blocks for the fast clone.

full clone
A complete, block independent, copy of a source file.
I

Internal Policy Engine
Internal process that performs periodic scans to deduplicate files, or reduplicates files on request.

Internet SCSI (iSCSI)
Protocol for sending SCSI packets over TCP/IP networks.

iSCSI LUN
VNX for file iSCSI software feature that processes SCSI commands such as reading from and writing to storage media.

L

link aggregation
High-availability feature based on the IEEE 802.3ad Link Aggregation Control Protocol (LACP) standard which allows Ethernet ports with characteristics similar to the same switch to combine into a single virtual device, or link with a single MAC address and potentially multiple IP addresses.

M

Multi-Path File System (MPFS)
VNX for file feature that allows heterogeneous servers with MPFS software to concurrently access, directly over Fibre Channel or iSCSI channels, shared data stored on a EMC Symmetrix or VNX for block storage system. MPFS adds a lightweight protocol called File Mapping Protocol (FMP) that controls metadata operations.

N

NDMP Volume Backup (NVB)
EMC-specific type of NDMP backup mechanism that backs up data blocks at a volume level, rather than at a file level. NVB reads a set of disk data blocks in an efficient manner compared to the method used for traditional, file-based backups. NVB works only with EMC-qualified vendor backup software. NVB is commonly known as Volume Based Backup (VBB).

Network Data Management Protocol (NDMP)
Open standard network protocol designed for enterprise-wide backup and recovery of heterogeneous network-attached storage.

network file system (NFS)
Network file system (NFS) is a network file system protocol that allows a user on a client computer to access files over a network as easily as if the network devices were attached to its local disks.

P

Portable Archive Interchange (PAX)
VNX for file archive protocol that works with standard UNIX tape formats and provides file-level backup and recovery operations.
primary storage
VNX for file that provides clients access to normal files and archived files through the stub files that represent them. The VNX for file contains all the stub files.

Q

quota
Limit on the amount of allocated disk space and the number of files (inodes) that a user or group of users can create in a Production File System. Quotas control the amount of disk space or the number of files that a user or group of users can consume or both.

R

reduplication
Process to undo the effect of deduplication on a file. If the file was compressed, it will be decompressed. If there are multiple instances of the file data, then a copy of the file data is made so that blocks are not shared between instances of the file. This process consumes additional space in the file system. Therefore, there must be sufficient free space in the file system to hold an additional copy of the original file for this process to complete.

S

service level agreement (SLA)
Contract or agreement that formally defines the level of expected availability, serviceability, performance, operation, or other attributes of service and even penalties in the case of violation of the SLA. An SLA may include acceptable downtime or disaster recovery time. In theory, an SLA is a formal agreement. However, in practice the agreement is often informal in which case it may be called a service level expectation (SLE).

SnapSure
On the VNX for file, a feature that provides read-only, point-in-time copies, also known as checkpoints, of a file system.
A
access time, set 38
alternate data streams limitations 12
application integration 19

B
backup data threshold, set 45
backups, PAX-based NDMP 18
byte-by-byte calculation 17

C
CIFS compressed file attribute 18
CIFS compression, set 46
cloning capabilities 19
concepts overview 16
configure deduplication 25

deduplication (continued)
set duplicate data detection method 43
set maximum file size 39
set minimum file size 40
set minimum scan interval 41
set modification time 42
states 17
suspend 34
test manually 28
undo 35
default settings, reset 48
display deduplicated file systems 30
duplicate data detection method, set 43

E
EMC E-Lab Navigator 52
enable deduplication 26
error messages 52
exclude file extensions 26, 36
exclude pathnames 26, 37
external deduplication 19

F
fast clone 19
file system deduplication restrictions 11
file system migration 21
file system space usage integration 21
file systems
display deduplicated 30
list deduplicated 31
suspend deduplication 34
undo deduplication 35
file-level retention integration 21
FileMover API 19
FileMover integration 21
free space restrictions 11
full clone 19

H
heavy usage restrictions 11

I
integrate applications 19
iSCSI LUN restrictions 12

K
known problems 53

L
LAN integration 20
limitations 11
list deduplicated file systems 31

M
manage deduplication 27
manually test deduplication 28
maximum file size restrictions 11
maximum file size, set 39
messages, error 52
minimum file size restrictions 11
minimum file size, set 40
minimum scan interval, set 41
modification time, set 42
modify data comparison capability 17
MPFS integration 22
MPFS restrictions 11

N
NDMP non-NVB integration 20
NDMP NVB integration 20
NDMP NVB restrictions 12
NDMP restore restrictions 12
NDMPcopy restrictions 12

P
PAX-based NDMP backups 18
point in time file system views integration 22
problems, known 53

Q
query deduplication settings 32

R
read operations 16
related information 13
replication integration 24
replication restrictions 11
requirements, system 10
restrictions 11

S
set duplicate data detection method 43
SHA-1 hash calculation 17
space-reduced backups 18
space-reduced files restrictions 12
states, deduplication 17
suspend deduplication 34
system requirements 10

test deduplication, manually 28
troubleshooting 51

U
undo deduplication 35
user interface choices 13

W
write operations 16