



EMC<sup>®</sup> NetWorker<sup>®</sup>

Avamar<sup>®</sup>

Release 8.0

## Integration Guide

P/N 300-013-563

REV A02

EMC<sup>2</sup>

Copyright © 2011-2012 EMC Corporation. All rights reserved. Published in the USA.

Published July, 2012

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided as is. EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose. Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC<sup>2</sup>, EMC, and the EMC logo are registered trademarks or trademarks of EMC Corporation in the United States and other countries. All other trademarks used herein are the property of their respective owners.

For the most up-to-date regulatory document for your product line, go to the technical documentation and advisories section on the EMC online support website.

# CONTENTS

## Contents

### Preface

### Revision History

### Chapter 1

#### Introduction

NetWorker Avamar Integration features .....	12
About NetWorker and Avamar .....	12
Source-based deduplication .....	12
NetWorker management of backups .....	12
Configuration and monitoring .....	12
Replication of backups .....	13
Deduplication plus tape backup .....	13
Support for NetWorker module applications .....	13
Unsupported NetWorker directives .....	13
Version requirements .....	14
License requirements .....	14
Hosts and software .....	14
Avamar server .....	14
Avamar Administrator .....	14
NetWorker server .....	15
NetWorker storage node .....	15
NetWorker Management Console server .....	15
NMC portal .....	15
NetWorker client .....	15

### Chapter 2

#### Planning and Practices

The NetWorker with Avamar environment .....	18
Choosing deduplication clients .....	18
Include clients that can provide high deduplication ratios .....	19
Exclude clients with low deduplication benefits .....	20
Faster backups versus faster recoveries .....	21
Maintaining best practices .....	21
Metadata storage requirements .....	23
Metadata device .....	23
Reclaiming expired storage space .....	24

### Chapter 3

#### Configuring the Software

Configuring Avamar for NetWorker .....	26
Configure the Avamar server for NetWorker .....	26
Configuring NetWorker for Avamar .....	27
Methods .....	27
Creating an Avamar deduplication node with the wizard .....	28
Configuring an Avamar deduplication node with the property windows .....	28
Create a metadata storage device .....	28

	Create an Avamar deduplication node .....	30
	Configure a client for deduplication backup .....	31
<b>Chapter 4</b>	<b>Replicating and Copying Backups</b>	
	Copying methods .....	34
	Replicating backup data .....	34
	Clone reports .....	35
	Replication and clone requirements .....	35
	Creating a replication node .....	36
	Creating rehydrated copies of backups .....	38
	Rehydrate options .....	38
	More information .....	39
<b>Chapter 5</b>	<b>Recovering Data</b>	
	Recovering deduplicated data .....	42
	Recovery requirements .....	42
	Supported NetWorker recovery procedures .....	42
	Auto and manual recovery from an Avamar replication node .....	42
	Failover autorecovery from a replication node .....	43
	Manual recovery from a replication node .....	43
	Disaster recovery .....	44
	Causes of disaster .....	44
	Potential losses .....	44
	Disaster recovery requirements .....	44
	Disaster recovery scenarios .....	45
<b>Chapter 6</b>	<b>Monitoring and Troubleshooting</b>	
	Viewing Avamar deduplication events in NMC .....	48
	View deduplication save sets with the mminfo command .....	48
	Cleanup of deleted deduplication save sets .....	49
	Refreshing Avamar deduplication node information .....	49
	Troubleshooting .....	49
	Restore operation fails .....	49
	Avamar server fills with expired save sets that were not deleted .....	50
	<b>Glossary</b>	

# PREFACE

*As part of an effort to improve its product lines, EMC periodically releases revisions of its software and hardware. Therefore, some functions described in this document might not be supported by all versions of the software or hardware currently in use. The product release notes provide the most up-to-date information on product features.*

*Contact your EMC representative if a product does not function properly or does not function as described in this document.*

---

**Note:** This document was accurate at publication time. New versions of this document might be released on the EMC online support website. Check the EMC online support website to ensure that you are using the latest version of this document.

---

## Purpose

This document provides information to configure, maintain, and monitor Avamar data deduplication solutions within the EMC NetWorker backup and storage management environment.

## Audience

This document is intended for use by system administrators and database administrators (DBAs). Readers of this document should be familiar with the following tasks:

- ◆ Identify the different hardware and software components that make up the NetWorker datazone.
- ◆ Follow procedures to configure storage management operations.
- ◆ Follow guidelines to locate problems and implement solutions.

## Related documentation

This document is part of the NetWorker documentation set. The following EMC publications provide additional information:

- ◆ *EMC Information Protection Software Compatibility Guide* provides a list of client, server, and storage node operating systems supported by the EMC information protection software versions.
- ◆ *EMC NetWorker Release 8.0 Installation Guide* provides instructions for installing or updating the NetWorker software for clients, console, and server on all supported platforms.
- ◆ *EMC NetWorker Release 8.0 Cluster Installation Guide* contains information related to installation of the NetWorker software on cluster servers and clients.
- ◆ *EMC NetWorker Release 8.0 Administration Guide* describes how to configure and maintain the NetWorker software.
- ◆ *EMC NetWorker Release 8.0 Release Notes* contain information on new features and changes, fixed problems, known limitations, environment, and system requirements for the latest NetWorker software release.

- ◆ *EMC NetWorker Licensing Guide* provides information about licensing NetWorker products and features.
- ◆ *EMC NetWorker License Manager 9th Edition Installation and Administration Guide* provides information on installation, setup, and configuration for the NetWorker License Manager product.
- ◆ *EMC NetWorker 8.0 Error Message Guide* provides information on common NetWorker error messages.
- ◆ *EMC NetWorker 8.0 Command Reference Guide* provides reference information for NetWorker commands and options.
- ◆ *EMC NetWorker Management Console Online Help* describes how to perform the day-to-day administration tasks in the NetWorker Management Console and the NetWorker Administration window.
- ◆ *EMC NetWorker User Online Help* describes how to use the NetWorker User program, which is the Microsoft Windows client interface for the NetWorker server, to back up, recover, archive, and retrieve files over a network.

## Conventions used in this document

EMC uses the following conventions for special notices:

**NOTICE**

NOTICE presents information related to hazards.

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

## Typographical conventions

EMC uses the following type style conventions in this document:

<b>Normal</b>	Used in running (nonprocedural) text for: <ul style="list-style-type: none"> <li>• Names of interface elements, such as names of windows, dialog boxes, buttons, fields, and menus</li> <li>• Names of resources, attributes, pools, Boolean expressions, buttons, DQL statements, keywords, clauses, environment variables, functions, and utilities</li> <li>• URLs, pathnames, filenames, directory names, computer names, links, groups, service keys, file systems, and notifications</li> </ul>
<b>Bold</b>	Used in running (nonprocedural) text for names of commands, daemons, options, programs, processes, services, applications, utilities, kernels, notifications, system calls, and man pages  Used in procedures for: <ul style="list-style-type: none"> <li>• Names of interface elements, such as names of windows, dialog boxes, buttons, fields, and menus</li> <li>• What the user specifically selects, clicks, presses, or types</li> </ul>
<i>Italic</i>	Used in all text (including procedures) for: <ul style="list-style-type: none"> <li>• Full titles of publications referenced in text</li> <li>• Emphasis, for example, a new term</li> <li>• Variables</li> </ul>

<code>Courier</code>	Used for: <ul style="list-style-type: none"> <li>• System output, such as an error message or script</li> <li>• URLs, complete paths, filenames, prompts, and syntax when shown outside of running text</li> </ul>
<b>Courier bold</b>	Used for specific user input, such as commands
<i>Courier italic</i>	Used in procedures for: <ul style="list-style-type: none"> <li>• Variables on the command line</li> <li>• User input variables</li> </ul>
< >	Angle brackets enclose parameter or variable values supplied by the user
[ ]	Square brackets enclose optional values
	Vertical bar indicates alternate selections — the bar means “or”
{ }	Braces enclose content that the user must specify, such as x or y or z
...	Ellipses indicate nonessential information omitted from the example

## Where to get help

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

**Product information** — For documentation, release notes, software updates, or information about EMC products, licensing, and service, go to the EMC online Support by Product website (registration required) at:

[EMC Online Support Site](#)

**Technical support** — For technical support, go to EMC online support and select Support. On the Support page, you will see several options, including one to create a service request. Note that to open a service request, 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.

## Your comments

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications. Send your opinions of this document to:

BSGdocumentation@emc.com





# REVISION HISTORY

Email your clarifications or suggestions for this document to:

[BSGdocumentation@emc.com](mailto:BSGdocumentation@emc.com)

The following table lists the revision history of this document.

Revision	Date	Description of added or changed sections
A02	July 12, 2012	Updated the following: <ul style="list-style-type: none"><li>• <a href="#">“Creating rehydrated copies of backups” on page 38</a> - updated nsrconsolidate command syntax.</li><li>• Removed several incorrect references to the use of FQDNs in the local host file.</li><li>• <a href="#">“Name resolution of Avamar server, nodes, and clients” on page 21</a> - Clarified</li></ul>
A01	June 21, 2012	First release of this document for Restricted Availability (RA) of <i>EMC NetWorker 8.0 Avamar Integration</i> .



# CHAPTER 1

## Introduction

This chapter contains the following topics:

- ◆ NetWorker Avamar Integration features ..... 12
- ◆ Version requirements ..... 14
- ◆ License requirements ..... 14
- ◆ Hosts and software ..... 14

## NetWorker Avamar Integration features

The EMC® NetWorker® integration with EMC Avamar® software offers the following features that you should become familiar with before you attempt to administer the system.

### About NetWorker and Avamar

NetWorker is a high-performance backup and recovery software solution that provides protection of business-critical data in mixed client environments. The client-server architecture of NetWorker provides high scalability and high performance from small LANs to enterprise-wide networks. NetWorker supports disk and tape storage devices as well as other data protection appliances.

Avamar is a backup and recovery solution that uses data deduplication technology. Deduplication at the client source reduces backup data by up to 500 times before it is transferred across the network to an Avamar server.

### Source-based deduplication

The NetWorker and Avamar products, which are deployed on separate hardware platforms, are integrated by the use of the Avamar deduplication node. The deduplication node is available as a NetWorker client configuration attribute.

After an initial full backup is performed to an Avamar deduplication node, the Avamar software identifies redundant sub-file data segments at the client source during subsequent backups. Only unique file segments that indicate changes, instead of entire files, are transferred to the Avamar server for storage. This feature reduces the network bandwidth, the storage space used for backups, and the time required to perform backups.

### NetWorker management of backups

Deduplication save sets are handled by the integration in the same way as regular NetWorker save sets are, except deduplication save sets consist of two parts that are stored separately. Both of these parts are tracked by the media index on the NetWorker server:

- ◆ The Avamar deduplication node stores the client backup data on the Avamar server.
- ◆ The backup metadata consists of a hash ID per file. This information is used to identify which sub-file segments in a backup must be deduplicated or have already been deduplicated. The metadata is essential to recover the stored data back to its original non-deduplicated format.

[“Metadata storage requirements” on page 23](#) provides details.

### Configuration and monitoring

The NetWorker Management Console (NMC) portal provides centralized configuration, monitoring, and reporting of backup and restore operations on Avamar deduplication nodes. Deduplication clients are supported for add, delete, and drag-and-drop operations, similar to non-deduplication clients. The NMC server is accessible from any supported remote Internet browser.

## Replication of backups

Backups that are stored on an Avamar deduplication node may be copied by Avamar procedures to a separate Avamar replication node for efficient disaster recovery or automatic failover of backups.

NMC does *not* manage or track Avamar replication operations. Replication configurations and operations must be configured, launched, and monitored from the Avamar Console. Replication nodes must be created and configured with the assistance of EMC Customer Support.

## Deduplication plus tape backup

For added protection, or if there is a requirement to archive backups to tape for offsite storage, clients may be backed up to both a deduplication node and tape. Two instances of the NetWorker Client resource can be created for the same client. One is assigned to a group that is backed up to a deduplication node and the other is backed up to on conventional tape. “[Creating rehydrated copies of backups](#)” on page 38 provides an optional method.

## Support for NetWorker module applications

The Avamar deduplication node supports clients that use the following NetWorker application modules:

- ◆ NetWorker Module for Microsoft Applications (NMM)
- ◆ NetWorker Module for MEDITECH (NMMEDI)
- ◆ NetWorker Module for Databases and Applications (NMDA)
- ◆ NetWorker Module for SAP with Oracle (NMSAP)

Other NetWorker application modules may *not* be supported, even though they may be supported by NetWorker and Avamar as separate products. Contact EMC Customer Support for details.

## Unsupported NetWorker directives

Some NetWorker directives that are not useful for the Avamar integration are *not* supported:

- ◆ Directives that filter data are supported. For example, the skip filter directive is supported.
- ◆ Directives that modify data are *not* supported. For example, the compressasm, encryptasm, and rawasm directives are ignored.
- ◆ The ocfsasm directive that is used to back up database files on OCFS file systems is unnecessary for this purpose and *not* supported.

## Version requirements

The hosts in the NetWorker with Avamar backup environment should meet the following minimum version requirements:

- ◆ EMC NetWorker 7.5 Service Pack 1 (SP1) or later
- ◆ EMC Avamar server 4.1 or later

The *EMC Information Protection Software Compatibility Guide* and the *EMC NetWorker Release Notes* provide version support details.

## License requirements

A fresh installation of a NetWorker release (not an upgrade) enables evaluation of all features, including the Avamar features, with no enabler (license key) required for 30 days. You can extend this evaluation period by 15 additional days by typing “grace” as an authorization code before the 30-day expiry period ends. After expiration, backups cannot be performed unless permanent enablers are applied.

Contact your EMC BRS sales representative for details on how to obtain temporary enablers to configure the Avamar deduplication node and Avamar features.

Two types of licenses must be enabled for the Avamar deduplication node to interact with the Avamar server:

- ◆ A deduplication option enabler for each Avamar server
- ◆ Disk Backup Option (DBO) for an AFTD or other supported storage device (supplied free for the first 1 TB of backup data)

## Hosts and software

This section describes hosts and software used in the Avamar deduplication node backup and recovery environment.

### Avamar server

The Avamar server may be installed with a system of rack-mounted storage computers that use Avamar software. An Avamar Datastore System (ADS) uses Avamar Datastore computer hardware although other hardware is also supported by the Avamar software.

The Avamar system, sometimes referred to as a “grid” may consist of a utility node and one or more storage nodes that all communicate with each other over their own dedicated network switch. A replication node would be located on a separate system.

### Avamar Administrator

Avamar Administrator is also known as the Avamar Console or the Management Console Server (MCS). MCS is a graphical user interface software application that is used to remotely administer an Avamar system from a supported web browser that has Java Runtime Environment (JRE) software installed. Multiple users can access the Avamar system concurrently from different Avamar Administrator consoles.

## NetWorker server

NetWorker 7.5 SP1 or later software must be installed on the NetWorker server that stores data to Avamar deduplication nodes.

The NetWorker server software provides services to back up and recover data for client computers within a NetWorker datazone.

## NetWorker storage node

A release of NetWorker software that is supported by the Avamar version must be installed on each storage node that is used to store the metadata of deduplicated backups. The storage nodes within the same NetWorker datazone must be of the same release and patch level.

## NetWorker Management Console server

The NMC server provides a graphical user interface for NetWorker configuration and monitoring operations. The NMC server must run NetWorker 7.4 or later software and have network access to the NetWorker server and storage nodes.

## NMC portal

The NMC server may be accessed and operated from a portal on a remote computer with a supported Internet web browser and JRE software installed. Multiple users can access the NMC server concurrently from different portals.

The *EMC NetWorker Installation Guide* provides information on supported web browsers and versions of the JRE.

## NetWorker client

A NetWorker client is a supported host whose data requires backup and restore services. The NMC server, NetWorker server, and NetWorker storage nodes are all NetWorker clients.

NetWorker client 7.4 SP1 or later software must be installed and running before the Avamar deduplication nodes are configured.

The *EMC Information Protection Software Compatibility Guide* provides information on supported releases.





# CHAPTER 2

## Planning and Practices

This chapter describes concepts that will help you plan and configure your deduplication backup environment for optimum performance. It contains the following topics:

- ◆ The NetWorker with Avamar environment ..... 18
- ◆ Choosing deduplication clients ..... 18
- ◆ Faster backups versus faster recoveries ..... 21
- ◆ Maintaining best practices ..... 21
- ◆ Metadata storage requirements ..... 23
- ◆ Reclaiming expired storage space ..... 24

## The NetWorker with Avamar environment

The NetWorker with Avamar environment stores not only deduplicated client data but also the deduplication metadata and information that is necessary to recover the client data.

Figure 1 on page 18 shows the paths of the backup data and metadata that are processed as follows:

1. The NetWorker server initiates the backup of several client groups within its datazone.
2. The NetWorker storage node for the clients communicates with the clients to identify which data needs to be deduplicated and stored.
3. The backup data is deduplicated at each client and stored on an Avamar deduplication node located on the Avamar server.
4. The metadata that identifies the deduplicated information is stored on a supported device, such as an AFTD, on the NetWorker storage node.
5. The NetWorker server maintains tracking information for the backup operation in its media index, which is typically stored on a locally-attached disk or tape device.

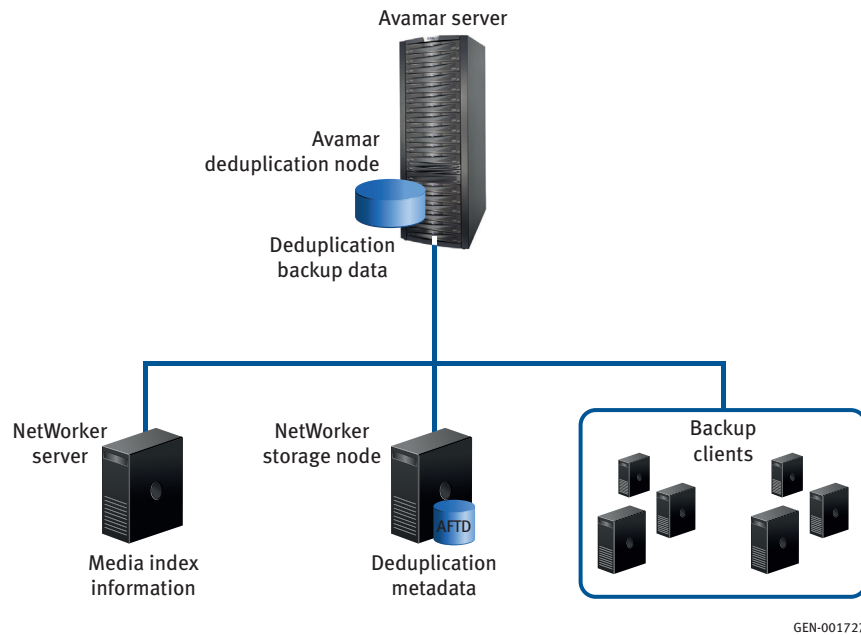


Figure 1 Integration environment showing storage of data and metadata

## Choosing deduplication clients

You can decide which clients are good candidates for deduplication backups with the help of the following sets of factors:

- ◆ “Include clients that can provide high deduplication ratios” on page 19
- ◆ “Exclude clients with low deduplication benefits” on page 20

## Include clients that can provide high deduplication ratios

To get the most efficient use of storage and backup schedules, choose clients that can provide high deduplication ratios.

Deduplication ratio is the reduction in storage space that results from the deduplication/compression technology that is applied to save sets. Some backups may achieve ratios up to 500:1, but reductions of even 5:1 are extremely valuable.

Factors that impact deduplication ratio are as follows:

- ◆ Change rate
- ◆ Retention period
- ◆ Data type
- ◆ Frequency of full backups

---

**Note:** Earlier releases used a different method to calculate deduplication ratio and the reported ratio may be slightly different for the same data previously backed up.

---

### Change rate

Client data with few changes between backups (low change rate) produce high deduplication ratios and are good candidates for deduplication. Data that has already been stored will not be stored again in subsequent backups.

In conventional business operations, the change rate for a client is typically about 1-2 percent in each subsequent backup after the initial full backup. A client with a daily change rate of 20 percent or more would not be a good choice for deduplication.

### Retention period

Client data with long retention periods tend to produce high deduplication ratios and are generally good candidates for deduplication. The longer the save sets are retained, the greater are the chances that the data already stored can be used to deduplicate subsequent backups. However, a long retention period will also reduce the available capacity for metadata on the storage node and backup data on the Avamar server.

In most cases, it is best to avoid retention periods of less than 7 days or greater than 12 months.

### Data type

Some types of data, such as text documents, slide presentations, spreadsheets, most database types, source code, and e-mail, contain redundant data and low change rates and are good deduplication candidates.

Some other types of data, such as audio, video, and scanned images, already consist of compressed data. Typically, the first full deduplication backup of these data types yields low reductions, but subsequent backups will typically produce high deduplication ratios if the data has a low change rate.

---

**Note:** Ensure that databases are not in compressed or encrypted format, and are not configured for multiplexed data streams, such as the Oracle RMAN filesperset parameter setting.

---

## Frequency of full backups

In general, use the same type of schedule as for non-deduplication backups, which would be a mix of full, incremental, and optional level backups.

Frequent full backups produce high deduplication ratios and often this can indicate good candidates for deduplication. However, frequent full backups can also quickly increase the amount of metadata that NetWorker must store and this can be a limiting factor. In some cases, the best backup plan may be to perform only incremental backups after an initial full backup.

## Exclude clients with low deduplication benefits

Certain performance factors can exclude clients from the benefits of deduplication backups.

These factors include the following:

- ◆ Number of client files
- ◆ Client application support
- ◆ Network bandwidth
- ◆ Client priority
- ◆ Client host performance

## Number of client files

The number of client files can also be a factor. A very large number of files (for example, over one million files per client) can cause client performance issues and such clients might not be good candidates for deduplication.

## Client application support

The application that generates the client data can also be a factor. Some applications may provide only limited support for deduplication, for example, by special flat file export. Another consideration might be whether there is an ability to quiesce the application or restore the data to a consistent state.

## Network bandwidth

The network speed or bandwidth available to the client to transmit its data is a factor for deduplication backups. Source-based deduplication normally benefits clients with low bandwidth network connections. Clients that require high bandwidths might not be suitable.

## Client priority

The priority of the client within a business continuity plan can be a factor for deduplication backups. A high-priority client will have a low recovery time objective (RTO) or recovery point objective (RPO). Low RTO requires restore time to be fast and repeatable, and this might be difficult to guarantee with deduplication.

## Client host performance

Client performance can be a consideration. The client will require sufficient memory, disk space, and processing speed to comfortably perform the deduplication backups. Although these resource demands are typically low and short in duration, they can become a factor.

If a client is old and already resource constrained, then the performance of the deduplication process and any application running on the client will be affected.

## Faster backups versus faster recoveries

Your choice of backup schedule will depend on whether you give more importance to faster backups or faster recoveries.

[Table 1 on page 21](#) compares estimated save set backup and recovery times relative to a non-deduplication full backup ( $x$  hours).

**Table 1** Backup schedule comparison

Backup type	Full initial backup	Subsequent daily full backups	Subsequent daily incremental backups (after a weekly full)	Data recovery time after 1 week
Non-deduplication	$x$ hours	$x$ hours	$0.1 x$ hours	$1.6 - 2.0 x$ , if the full backup and all differential backups are to be restored.
Deduplication, daily full	$1.5 x$ hours	$0.2 x$ hours	n/a	Less than or equal to $x$ (only one backup need be restored).
Deduplication, daily differential	$1.5 x$ hours	n/a	$0.05 x$ hours	$1.5 - 2.0 x$ , if the full backup and all differential backups are to be restored.

The *EMC NetWorker Administration Guide* provides details.

## Maintaining best practices

The performance of Avamar deduplication nodes depends on good scheduling practices and attention to the size and configuration of the Avamar server.

### Name resolution of Avamar server, nodes, and clients

The network environment has an impact on hostname resolution methods and you should follow the recommendations of the operating system vendor:

- ◆ The deduplication client must be able to resolve the names of the NetWorker server and the configured Avamar deduplication node.
- ◆ Avoid renaming the Avamar server, deduplication nodes, or deduplication clients. These operations should be done with assistance from EMC Technical Support:

---

**Note:** Renaming the Avamar server, or renaming deduplication or replication nodes that are hosting unexpired backups, will have an adverse effect on the recovery of the existing save sets.

---



---

**Note:** Renaming clients that have unexpired deduplication backups will have an adverse effect on data recovery.

---

[“Recover from a renamed node” on page 43](#) provides more details.

[“Avamar server fills with expired save sets that were not deleted” on page 50](#) provides more details.

## Backup and replication schedules

Schedule backups and replications so they do not extend into the Avamar maintenance blackout window or the NetWorker maintenance window:

- ◆ At the start of the blackout window, running backups have a 15 minute grace period before they are cancelled. If this cancellation produces a partial backup, it will be stored on the Avamar server for seven days to facilitate the next backup of the client.
- ◆ The NetWorker software will retry interrupted backups up to the value set for client backup retries.
- ◆ Schedule and perform replication operations only at times when backup operations will not be impaired, typically at the end of the backup window.
- ◆ A recovery from a replication node may be started only after all of the replicated data is stored and on the node.

[“Reclaiming expired storage space” on page 24](#) provides more details on maintenance operations.

## Extending your retention policies

If you extend a client’s retention policy, carefully monitor the Avamar server’s capacity usage. The Avamar server can fill up much faster than expected.

## Rollbacks to checkpoints

The NetWorker software is unaware of changes that result from rollbacks to checkpoints on the Avamar system.

### NOTICE

Rollbacks to checkpoints can result in lost backups, lost clients, and lost deletions.

The following situations can occur:

- ◆ Avamar deduplication save sets that were created between the checkpoint and the start of the rollback become unrecoverable.
- ◆ Deduplication clients created in the Avamar domain during the rollback period become lost.
- ◆ Save sets that were deleted during the rollback period become undeleted. Provide EMC Customer Support with the `/nsr/logs/nsravamar.raw` log file if you need assistance to list the deletions on the Avamar server.

Protect deduplicated save sets by the procedures described in [Chapter 4, “Replicating and Copying Backups.”](#)

## Avamar server workloads

Other operations performed with the Avamar server can impact performance:

- ◆ The Avamar server may be shared among different NetWorker datazones as well as the native Avamar backup software. However, added complexity impacts server performance, scaling, and storage space usage. Use EMC Data Protection Advisor (DPA) to manage complex environments.
- ◆ The use of the Avamar server for other operations that are not available through the NetWorker Management Console can impair performance and should be done only with the assistance of EMC Customer Support.

## Metadata storage requirements

The backup metadata consists of a hash ID per file backed up to the Avamar server. This information is used to identify which sub-file segments in a backup must be deduplicated or have already been deduplicated. The metadata is essential to recover the stored data back to its original non-deduplicated format.

The metadata also enables all backup, clone copy, restore, disaster recovery, monitoring, and reporting operations to be managed and tracked through the NMC display.

### Metadata device

A supported storage device, such as an AFTD, is required to store deduplication metadata on the NetWorker storage node:

#### NOTICE

Clone the metadata device volume or save sets to protect this essential information and ensure disaster recovery.

- ◆ The metadata device should be dedicated for deduplication clients and not used by other types of clients, pools, or data.
- ◆ The volume for the metadata device should be labeled and belong to a deduplication pool.
- ◆ Metadata device names should be simple:
  - The device path and name should be unique for each NetWorker datazone.
  - The name of the device volume should not exceed 60 characters.
  - Locate each device within its own directory even if there is a dedicated file system or partition.
  - Use a clear and consistent naming convention if other similar devices are used.
- ◆ The metadata device should have enough space to accommodate all the full and differential backups for the retention periods involved.

Depending on the file characteristics, each metadata file could generate between 160 bytes to 1 KB of metadata. Thus 200 million files could require as much as 200 GB of metadata on the metadata device.

## Reclaiming expired storage space

When save sets on a deduplication node exceed their NetWorker retention policies, the NetWorker server begins the process of deleting the expired data and recycling the storage memory. Either on startup or every 6 hours, NetWorker invokes a temporary maintenance process to start the deletion of expired data on the Avamar deduplication node and the associated metadata on the NetWorker storage node.

If you delete the NetWorker server's media index entry or overwrite the volume label, this action will only free up space that the deduplicated save set used.

If an Avamar server is removed from the NetWorker datazone, the internal Avamar retention policy reverts to "no expiration" and saveset data will be retained on the Avamar server. The deduplication node cannot be deleted in NetWorker as long as there are expired save sets in the NetWorker queue to be deleted on the Avamar server.

The deletion of expired data from the Avamar server should be monitored with NMC or the Avamar Console to ensure that this process runs successfully and that the server does not fill up with redundant data, especially if Avamar replication is being used. [Chapter 4, "Replicating and Copying Backups,"](#) provides details.

["Cleanup of deleted deduplication save sets" on page 49](#) provides instructions on how to delete Avamar deduplication save sets.

["Avamar server fills with expired save sets that were not deleted" on page 50](#) provides more details.



# CHAPTER 3

## Configuring the Software

This chapter contains the following topics:

- ◆ [Configuring Avamar for NetWorker.....](#) 26
- ◆ [Configuring NetWorker for Avamar.....](#) 27
- ◆ [Creating an Avamar deduplication node with the wizard.....](#) 28
- ◆ [Configuring an Avamar deduplication node with the property windows .....](#) 28

## Configuring Avamar for NetWorker

Configure the Avamar server for the NetWorker integrated backup environment as described in the following sections.

### Configure the Avamar server for NetWorker

To enable an Avamar server for storage and recovery operations with the NetWorker server:

1. Ensure the required versions and licenses are available and ready to use. “[Version requirements](#)” on page 14 provides details.
2. Ensure that an Avamar 4.x or later server is installed, running, and fully operational on an appropriate platform for the Avamar server.

The *EMC Avamar System Installation Manual* provides details.

3. Install the NetWorker 7.5 SP1 or later client software on the Avamar server:
  - a. On the Avamar server utility, open a command shell and log in as the root user.
  - b. Change to the directory containing the NetWorker software.
  - c. Install the NetWorker client software in the default location. For example:

```
rpm -ivh --nodeps lgtocInt-7.x-y.i686.rpm
```

- d. Start the NetWorker daemons. For example:

```
/etc/init.d/networker start
```

- e. Verify that the NetWorker **nsrexecd** daemon has started. For example:

```
ps -ef | grep nsr
```

4. Change Avamar server preferences to allow duplicate client names:
  - a. On the Avamar server, log out as root and log in as the admin user.

---

**Note:** You must *not* be logged in as root.

---

- b. Set the default ssh session shell and load the admin user’s OpenSSH key:

```
ssh-agent bash
ssh-add ~admin/.ssh/admin_key
```

- c. Stop the Avamar Management Console Server (MCS) service:

```
dpnctl stop mcs
```

- d. Open the mcsserver.xml preferences file in a UNIX text editor. For example:

```
/usr/local/avamar/var/mc/server_data/prefs/mcsserver.xml
```

- e. Set “allow\_duplicate\_client\_names” to **true**.

- f. Save your changes and restart MCS:

```
dpnctl start mcs
```

5. Add a NetWorker domain to the Avamar server:
  - a. Launch the **Avamar Administrator Graphical Management Console**.
  - b. Select **Navigation > Administration** and click the **Account Management** tab.
  - c. In the navigation tree, select the **server icon** (root domain).
  - d. Select **Actions > Account Management > New Domain**.
  - e. In the **New Domain** window, type the following information described in [Table 2 on page 27](#).

**Table 2** New Domain fields

Field	Description
New Domain Name	NetWorker
Contact	Optional contact name
Phone	Optional contact telephone number
Email	Optional contact e-mail address
Location	Optional contact location

To avoid ambiguity in name resolution, specify the fully qualified domain name (FQDN) to identify the deduplication node and the replication node, if used.

If configuring Avamar replication, verify that the replication setup on the original (replicated) deduplication node uses the same destination directory (domain) as NetWorker uses. If a difference exists, change the Avamar replication setup.

- f. Click **OK** to confirm the configuration.

The new NetWorker domain appears in the **Account Management** tree.

## Configuring NetWorker for Avamar

After an Avamar server has been configured for the integrated environment, you may configure the NetWorker datazone with an AFTD or other supported storage device, a deduplication node, deduplication clients, and optional replication nodes if required.

### Methods

The NMC software provides the Device Configuration Wizard that is used to create and configure the Avamar deduplication nodes.

After a deduplication node has been created with the wizard, its configuration may be modified, if required by using the individual NMC property windows:

- ◆ [“Creating an Avamar deduplication node with the wizard” on page 28](#)
- ◆ [“Configuring an Avamar deduplication node with the property windows” on page 28](#)

## Creating an Avamar deduplication node with the wizard

The Device Configuration Wizard is designed to reduce user errors during the creation of Avamar deduplication nodes and is the preferred method.

To create a deduplication node with the wizard:

1. Ensure the required NetWorker licenses are enabled. [“License requirements” on page 14](#) provides details.
2. In the NMC **Enterprise** view, select the NetWorker server name, and double-click the NetWorker application to launch it.
3. In the NetWorker Administration application, click **Devices**.
4. In the left panel, right-click **Avamar Deduplication Nodes** and select **New Device Wizard**.
5. On each wizard page that appears, specify the options and values required for the backup configuration and click **Next**.
6. On the **Select the Device Type** wizard page, select the **Avamar Deduplication Node** type.
7. Proceed through the wizard steps to configure the Avamar deduplication node, review settings, and confirm results.

[“Configuring an Avamar deduplication node with the property windows” on page 28](#) provides additional configuration details.

## Configuring an Avamar deduplication node with the property windows

The Device Configuration Wizard is the recommended method to create Avamar deduplication nodes. You may modify the nodes and perform other tasks by following the procedures in this section.

The following sections provide details:

- ◆ [“Create a metadata storage device” on page 28](#)
- ◆ [“Create an Avamar deduplication node” on page 30](#)
- ◆ [“Configure a client for deduplication backup” on page 31](#)

### Create a metadata storage device

Avamar deduplication backups require a device on the NetWorker storage node to store the metadata of the deduplication save sets. [“Metadata storage requirements” on page 23](#) provides details.

The AFTD or other supported storage device may be associated with either:

- ◆ A single deduplication client
- ◆ A pool that includes numerous deduplication clients

The *EMC NetWorker Administration Guide* provides more information.

To configure a NetWorker Device resource for a device that will store the metadata:

1. Ensure that the requirements for the metadata storage device are met.  
[“Metadata storage requirements” on page 23](#) provides details.
2. On the NetWorker storage node host or a NetWorker server host, create a single dedicated directory for the metadata device within its own disk or partition.

#### NOTICE

Do *not* use the temp or root directories. Data in the temp directory can be overwritten and data in the root directory can cause performance problems.

3. In the NMC NetWorker **Administration** window, click the **Devices** view.
4. In the navigation tree, right-click **Devices** and select **New**.
5. In the **Create Device** window, **General** tab, replace the default **Name** value with the complete pathname of the directory that you created:
  - If the device is configured on the NetWorker server host (local storage node), the name is the simple device path, such as `/backup/d0`.
  - If the device is configured on a NetWorker storage node host (remote storage node), then prefix the name with `rd=` and separate the hostname and device pathname with a colon (`:`).

For example, for a remote device on the storage node named “jupiter,” the device path might be:

```
rd=jupiter:/backup/d0
```

6. In the **Media Type** attribute, select the type of metadata storage device to be used, for example **adv\_file** (for an AFTD).
7. Ensure that **Auto Media Management** is *not* selected. This attribute pertains to multiple tape volumes.
8. Ensure the **Cleaning** options are disabled (their default settings), so that automatic cleaning is *not* invoked.
9. On the **Configuration** tab, set the number of concurrent save sessions (streams) and the number of nsrmmmd (data mover) processes the device may handle:
  - **Target Sessions** is the number of sessions that a nsrmmmd process on the device will handle before another device on the host will take the additional sessions. Use this setting to balance the sessions among nsrmmmd processes.  
  
If another device is not available, then another nsrmmmd process on the same device will take the additional sessions.  
  
Typically, set this attribute to a low value. The default value for an AFTD is 4. It may not be set to a value greater than 60.
  - **Max Sessions** is the maximum number sessions the device may handle. If no additional devices are available on the host, then another available storage host takes the additional sessions, or retries are attempted until sessions become available.

The default value for an AFTD is 32, which typically provides best performance. It cannot be set to a value greater than 60.

- **Max nsrmmmd count** limits the number of nsrmmmd processes that can run on the device. Use this setting to balance the nsrmmmd load among devices.

If **Dynamic nsrmmmds**, instead of the default setting for static nsrmmmds (NetWorker continuously runs all available nsrmmmd processes), is enabled on the storage node, this value is automatically adjusted by the formula  $\text{max}/\text{target} + 4$ , with the default value for AFTDs being 12. Otherwise the default value is 4. The *NetWorker Administration Guide* provides details on Dynamic nsrmmmds.

To modify this value, first adjust the sessions attributes, apply and monitor the effects, then update max nsrmmmd count.

---

**Note:** At least one nsrmmmd process will always be reserved for restore or clone operations.

---

10. In the **Local Backup** area, leave **Dedicated Storage Node** at **No**.
11. In the **Remote Host** area, if a CIFS (Microsoft Windows) or NFS (UNIX) path is specified in the **Device Access Information**, then type a **Remote User** name and **Password**. An NFS server must allow access by root.
12. Click **OK** when the configuration is complete.
 

The NetWorker **Administration** window displays details of the new device resource.
13. Label and mount the NetWorker DiskBackup Device in the same manner as other offline media used for NetWorker backups.
  - Limit the volume name to 60 characters when labeling the volume.
  - Because an AFTD is never marked full, there is no need to add another device to the pool.
14. If a new password for a CIFS AFTD is provided, it might be necessary to unmount and re-mount the device to ensure that the change is in effect.

## Create an Avamar deduplication node

To create an Avamar deduplication Node resource:

1. In the NetWorker **Administration** window, click the **Devices** view.
2. In the navigation tree, right-click **Avamar Deduplication Nodes**, and select **New**.
3. In the **Create Deduplication Node** window, **General** tab, **Name** attribute, type the fully qualified domain name (FQDN) of the deduplication node, which must already exist on the Avamar server.
4. In the **Access** attribute, type a the username and password to access the deduplication node.

---

**Note:** Any subsequent changes to the remote password on the Avamar server must be updated in this Avamar Deduplication Node resource.

---

5. (Optional) If the **General** tab shows a value for **Replication node**, then a replication node that this deduplication node can use has already been created on this deduplication node server. [“Creating a replication node” on page 36](#) provides details.

The Avamar documentation provides details on creating a replication node.

6. Click **OK**.

The **Avamar Deduplication Nodes** table lists the new node.

## Configure a client for deduplication backup

After an Avamar deduplication node has been created, you may configure Avamar deduplication clients either from scratch or by reconfiguring existing clients.

To configure a deduplication client:

1. Ensure that the Avamar deduplication node already exists on the Avamar server.
2. In the NetWorker **Administration** window, click the **Configuration** view.
3. In the navigation tree, right-click **Clients** and select **New**, or you may convert the **Properties** of an existing non-deduplication client.
4. Unless you are creating a new Client resource, skip this step. The following attributes do not need to be modified for clients whose backups will be redirected to a deduplication node:
  - a. In the **Name** attribute, type the hostname of the deduplication client. The host must be running an operating system that supports deduplication.
  - b. Select values for the **Browse Policy** and **Retention Policy**.
  - c. Select the **Scheduled Backups** checkbox.
  - d. In the **Save Sets** attribute, click **Browse** and in the **Choose Directory** window, select all the directories or individual files to be backed up:
    - Each item should appear on a separate line.
    - To back up *all* client data, type All in the **Save Sets** attribute.
  - e. When finished, click **OK**.

The *EMC NetWorker Administration Guide* provides details for this step.

5. In the **Group** attribute, select a backup group. A backup group consists of clients that are backed up according to the schedule for the group.

---

**Note:** This attribute might show pre-selected groups based on the pool selected. Do *not* separate groups from their associated pools at the client level.

---

For deduplication backups:

- The group must be exclusively devoted to deduplication clients. Do *not* mix regular and deduplication clients within a group.
- Select a group whose schedule avoids the deduplication node’s read-only periods, when cron jobs, such as checkpoint creation or garbage collection are run. The Avamar documentation provides details.

- Ensure that the group's schedule includes periodic full backups. All backup levels are supported by deduplication.
  - Enable cloning for the group, preferably on a different storage node. This creates a second copy of the deduplication metadata and provides necessary protection if the first copy is lost.
6. (Optional) In the **Schedule** attribute, select a backup schedule. The schedule selected in this attribute overrides any other schedules that might be configured for the client or its save sets.
  7. In the **Pool** attribute, select a pool that targets the Avamar deduplication node that you want to use. The pool selected in this attribute overrides any other pool that might be configured for the client or its save sets.

For deduplication backups:

- The pool must be exclusively devoted to deduplication clients.
- The pool must target devices on the NetWorker server or storage node that store the metadata of the deduplication save sets.
- The NetWorker server's bootstrap and index data are not deduplicated. The bootstrap and index files should be configured in a separate pool and is typically backed up to a traditional disk or tape device that is locally attached to the NetWorker server.

The *EMC NetWorker Administration Guide* provides details.

8. On the **Apps & Modules** tab:
  - a. Select **Avamar deduplication backup**.
  - b. Select the name of the Avamar deduplication node that will store the client's backup data. Do *not* select the name of a replication node.

---

**Note:** If you change the client to a different deduplication node, then the next backup *must* be a full (Level 0) backup.

---

9. On the **Globals (1 of 2)** tab, ensure that **Parallelism** is set to a value of 4 or less for optimal performance. The parallelism value configured in this attribute overrides any other parallelism value that may also be set in the Group resource or in the Schedule resource.
10. Complete the remaining settings the same as for any regular client and click **OK**.

The *EMC NetWorker Administration Guide* provides details.



# CHAPTER 4

## Replicating and Copying Backups

This chapter contains the following sections:

- ◆ Copying methods ..... 34
- ◆ Replicating backup data ..... 34
- ◆ Creating rehydrated copies of backups ..... 38

## Copying methods

There are basically two ways to copy backup data stored on an Avamar server. Each is suited to specific purposes, as described in the following sections:

- ◆ [“Replicating backup data” on page 34](#)
- ◆ [“Creating rehydrated copies of backups” on page 38](#)

## Replicating backup data

To ensure added data protection of backup data, for example for disaster recovery or failover scenarios, save sets and backup metadata stored at the primary location can be replicated to a secondary location. This process involves two separate procedures:

1. Avamar replication of backup data

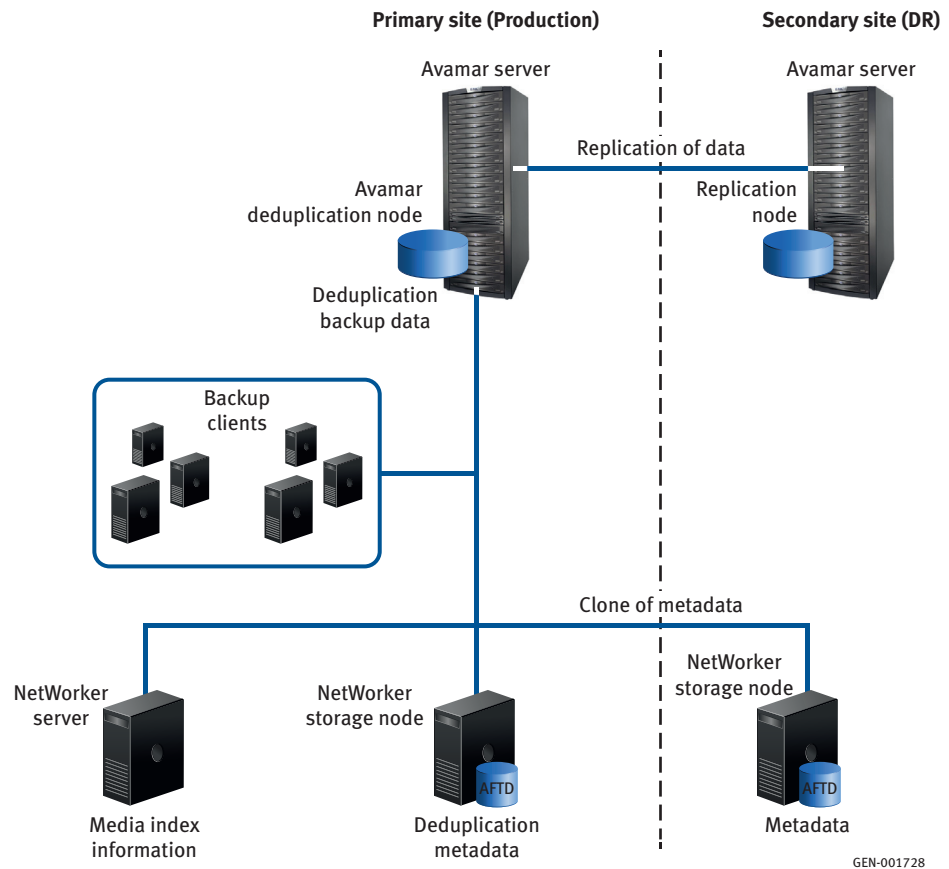
The backup data, stored on a deduplication node on the primary Avamar server is replicated to a replication node on an Avamar server at a secondary location. Single save sets or the entire volume of the deduplication node may be replicated.

2. NetWorker clone of backup metadata

The backup metadata, stored on the primary NetWorker storage node is cloned to a NetWorker storage node at a secondary location. By default, the cloned metadata retains the original NetWorker browse and retention policies, but these can be changed to give the replicated data different policies. Staging (moving) of the metadata is *not* recommended.

The *EMC NetWorker Administration Guide* provides details on cloning.

[Figure 2 on page 35](#) illustrates a basic disaster recovery environment. In this scenario, a data loss at the primary site can be recovered by restoring data from the secondary site. A failover scenario, in which the entire primary site fails and the secondary site continues backup and recovery operations for the clients would require a NetWorker server (not shown) at the secondary site with media index information cloned from the primary NetWorker server.



**Figure 2** Basic disaster recovery environment

## Clone reports

The NetWorker software provides reports on clone operations of Avamar metadata that can be viewed in the NMC Reports view.

[“Viewing Avamar deduplication events in NMC” on page 48](#) provides details.

## Replication and clone requirements

The following requirements must be met before replication of Avamar deduplicated backup data and clone of the corresponding metadata can be performed:

- ◆ Replication nodes should be created and already exist before integration with the NetWorker server, preferably when the Avamar server is installed.
- ◆ Data replication must be configured, scheduled, and monitored on the Avamar server. The NetWorker software cannot perform these operations.
- ◆ An Avamar replication node must also be an Avamar deduplication node and must be part of the NetWorker server’s datazone, even if it is at a geographically distant location.

- ◆ The NetWorker storage node, which will store a clone copy of the metadata for the replicated backups, must also be part of the NetWorker server's datazone.
- ◆ A replication node does not require a dedicated host. The host may also serve a deduplication node for different clients and backup groups.
- ◆ Schedule or manually start replication operations at times that will ensure backup operations are not impaired, typically at the end of the backup window. A replication started immediately after a backup finishes will experience a short delay.
- ◆ All of the replicated data must be available on the replication node before a recovery operation can be started from that node.

The *EMC Avamar System Administration Guide* provides further details.

## Creating a replication node

An Avamar replication node must also be an Avamar deduplication node that enables the NetWorker server to track the replicated backups.

---

**Note:** Avamar replication nodes must be created and configured with the assistance of EMC Customer Support.

---

To create a NetWorker resource for a replication node:

1. Define the primary Avamar deduplication node that the replication node will copy:
  - a. In the NetWorker **Administration** window, click the **Devices** view.
  - b. In the navigation tree, right-click the primary node and select **Properties**.
  - c. Click the **General** tab and record the following information:
    - The **Name** value, which is the fully qualified domain name (FQDN) of the Avamar deduplication node.
    - The **Replication node** value, if it already exists, which is the FQDN of the Avamar replication node that this Avamar deduplication node can use.
  - d. Close the **Properties** window.
2. Configure the Avamar replication node. This is a secondary Avamar deduplication node, located on a different Avamar server, where the replication data will be stored:
  - a. In the navigation tree, right-click **Deduplication Nodes**, and select **New**.
  - b. In the **Create Deduplication Node** window, **General** tab, **Name** attribute, type the FQDN of the replication node from [step 1](#).

### NOTICE

Ensure that this name is the same as the Avamar */destination\_directory/REPLICATE/node\_name*. The use of a different name prevents the deletions of expired save sets on the replication node and results in the node becoming filled to capacity — a situation that requires assistance from EMC Customer Support to resolve.

---

- c. In the **Access** attribute, type the username and password to access the replication node.

---

**Note:** Any subsequent changes to the password on the Avamar server must be updated in this Replication Node resource.

---

- d. (Optional) If the **General** tab shows a value for **Replication node**, then a replication node that this replication node can use has already been created on this Avamar server. This value could be used for advanced configurations such as for cross-replications between the two production Avamar servers, or for a “cascading” replication from the secondary server to a third server.

The Avamar documentation provides details on creating a replication node.

- e. Click **OK**.

The **Deduplication Nodes** table lists the new replication node.

3. Give the replication node access to the Avamar deduplication node:
  - a. In the Avamar Management Console Server (MCS), open the **Administration** interface for the primary deduplication node.
  - b. On the **Services Administration** tab, right-click the **Replication Cron job** attribute, and select **View Properties**.
  - c. Ensure that the **Destination Directory:/REPLICATE/** attribute matches the exact value of the deduplication node. If necessary, make changes and click **Apply**.

---

**Note:** For supported Avamar server releases earlier than 4.1, the **Destination Directory:/REPLICATE/** attribute was labeled **System Name**.

---

- d. Ensure that the **Destination** attribute matches the exact value of the replication node. If necessary, make changes and then click **Apply**.
- e. Click **OK**.

---

**Note:** Changes to the replication node must be performed with assistance from EMC Customer Support and updated in the corresponding Avamar deduplication node.

---

The *EMC Avamar System Administration Guide* provides further details.

## Creating rehydrated copies of backups

In some cases, for example to meet long-term data retention requirements, it is necessary or desirable to revert (“rehydrate”) deduplicated backups to their original non-deduplicated state and store the rehydrated backups on traditional tape or disk media.

This method provides the following features:

- ◆ A rehydrated traditional tape or disk copy of a backup can be created at any time, unlike the backup-to-tape method described in “[Deduplication plus tape backup](#)” on [page 13](#), which must be performed at the time of backup.
- ◆ The optional compression and encryption directives, and NetWorker retention policies can be applied to the rehydrated copies.

---

**Note:** Rehydrating multiple save sets concurrently can impact Avamar performance.

---

## Rehydrate options

The **nsrconsolidate** command, which uses NetWorker synthetic full backup operations, can be used to rehydrate one or more Avamar deduplication save sets and store the resultant data as one or more traditional save sets.

---

**Note:** To rehydrate save sets from a replication node, set the `RECOVER_FROM_REP_HOST` environment variable before invoking the **nsrconsolidate** command.

---

**Example 1** To rehydrate all deduplication save sets for client “mars” and create the new save sets in the destination *rehydrate\_pool*:

```
nsrconsolidate -b rehydrate_pool -f dirfile -R -S 'mminfo -C mars
-q dedupe -r ssid'
```

**Example 2** To identify specific deduplication save sets and rehydrate them in the destination *rehydrate\_pool*:

- a. List the save sets for the client “mars” in ssid format, redirected to a file (*input\_file*).

```
mminfo -C mars -q dedupe -r ssid >input_file
```

- b. Rehydrate the save sets, either from the input file or individually, by typing one of the following commands:

```
nsrconsolidate -b rehydrate_pool -f dirfile -R -S -I input_file
nsrconsolidate -b rehydrate_pool -f dirfile -R -S ssid1 ssid2 ...
```

Where:

*rehydrate\_pool* is the destination media pool.

*dirfile* is the file that contains compression and encryption directives.

A directive file may contain only compressasm (compression) and aes (encryption) options and should be used only when storing data in a non-deduplicated format, for example:

```
<<"H:\TechDocs\pdfs2">>
+compressasm : *
<<"H:\TechDocs\important">>
+aes : *
```

Table 3 on page 39 lists command line options that support **nsrconsolidate** rehydration of Avamar deduplicated save sets.

**Table 3** Typical **nsrconsolidate** options for rehydration of Avamar deduplicated save sets

Option	Description
-b <i>rehydrate_pool</i>	Specifies the destination media pool to store the new rehydrated save set.
-f <i>dirfile</i>	Apply the directives that are listed in the directive file. Supported directives are "compressasm" and "aes" (encryption).
-i on off	When used with the -R option, specifies whether to generate the client file index when rehydrating a save set. The default setting (-i off) does not create the index and typically would be used for disaster recovery operations. As an alternative, the <b>scanner</b> command can create indexes after rehydration.
-l <i>input_file</i>	Specifies the input filename used by the <b>nsrconsolidate</b> program to read the "client:saveset-name" pairs or save set identifiers (ssid) for the rehydrating operation. Each entry must be separated by a new line.
-R	(Valid only in combination with the -S option) Rehydrate each save set (identified by the -S <i>ssid</i> option) individually instead of consolidating them together (synthetic full operation) as one new save set. A new save set is generated for each ssid.
-S	Specifies the list of full and incremental save sets for rehydration in ssid or cloneid format, specified on the command line or through the -l <i>input_file</i> .

## More information

The *EMC NetWorker Command Reference Guide* or man pages provide details of the **nsrconsolidate** command.

The *EMC NetWorker Administration Guide* provides details on synthetic full operations.





# CHAPTER 5

## Recovering Data

This chapter contains the following topics:

- ◆ [Recovering deduplicated data .....](#) 42
- ◆ [Auto and manual recovery from an Avamar replication node.....](#) 42
- ◆ [Disaster recovery.....](#) 44

## Recovering deduplicated data

Provided that the NetWorker and Avamar software have been properly configured with a deduplication node or replication node, the data recovery from either of the nodes is basically the same as for a NetWorker storage node, except that the data is stored in two components in different places:

- ◆ Deduplicated client backup data is stored on an Avamar deduplication node, which is on an Avamar server.
- ◆ Backup metadata, which is the hash information used to deduplicate and track the stored backup data, is stored on a NetWorker storage node. This metadata is required to revert the deduplicated data to its original non-deduplicated format.

If the backup-to-tape method described in [“Deduplication plus tape backup” on page 13](#) is used regularly, it provides a regular non-deduplicated tape backup of the client data up to a particular point-in-time.

### Recovery requirements

The requirements for recovery from a deduplication node or replication node are as follows:

- ◆ All of the deduplicated or replicated data must be available on the deduplication or replication node before the data can be recovered.
- ◆ Both the Avamar deduplication node and the NetWorker storage node must be online during the recovery of deduplicated data.

### Supported NetWorker recovery procedures

Most data recovery procedures are the same as for non-deduplicated NetWorker backups:

- ◆ NetWorker software can be used to browse the media index and recover a file or save set.
- ◆ The NetWorker recovery commands may be used.
- ◆ As many as four recovery sessions may be run concurrently, the same as for backups.
- ◆ Directed recoveries are supported for NetWorker 7.4 SP1 or later clients and storage nodes.
- ◆ Attempts to recover expired backup data can use NetWorker **scanner** program to rebuild a media index. However, this index provides access to only the surviving backup metadata for deduplication backups.

The *EMC Administration Guide* provides procedures for data recovery.

## Auto and manual recovery from an Avamar replication node

Deduplicated backup data may be recovered from a secondary replication Avamar node by autorecovery or by manual methods.

## Failover autorecovery from a replication node

If recovery from the primary deduplication node fails, Avamar will use a failover autorecovery operation to transparently enable the NetWorker software to recover the data from a secondary replication node, if such a node exists.

The requirements for autorecovery are as follows:

- ◆ Both the primary and the replication node must be configured on the NetWorker server.
- ◆ A replication node must be configured on an Avamar server by EMC Customer Support. [“Creating a replication node” on page 36](#) provides details.
- ◆ The replication node must meet all its requirements. A node may be visible in NMC, but not be qualified for use. [“Replication and clone requirements” on page 35](#) provided details.

## Manual recovery from a replication node

If the autorecovery from a correctly configured and qualified replication node fails, data may still be recovered from a replication node by using either of the following manual recovery methods, depending on the circumstances.

### Recover from the replication node first

If an autorecovery operation fails, you can cause the NetWorker software to attempt to recover the backup data from the replication node first before it attempts to recover the data from the primary deduplication node:

1. From a command prompt, set the RECOVER\_FROM\_REP\_HOST environment variable to **Yes**.
2. Export the variable value.
3. Recovery from the command prompt by using the NetWorker **recover** command.

The *EMC NetWorker Administration Guide* provides details.

### Recover from a renamed node

If the name of the replication node or deduplication node was changed after the backup was performed, use the **recover -X** command.

For example:

```
recover -c scip2b081 -d /tmp -X "Replication Node=replication_node" /usr/sbin/nsrd
```

where *replication\_node* is the name of the new replication node.

To manually recover from a deduplication node whose name has changed, substitute the **-X** value with **-X “Deduplication Node=deduplication\_node”**.

The *EMC NetWorker Administration Guide* provides details.

## Disaster recovery

For the purpose of this guide, a disaster is any loss of data in which the computing environment required to restore that data is not available. Disaster recovery (DR) is necessary when ordinary data recovery procedures are not sufficient to recover the computing environment and its data to normal day-to-day operations.

### Causes of disaster

A disaster can result from any of the following situations:

- ◆ Hardware or debilitating software failures
- ◆ Computer viruses that corrupt the computing system
- ◆ Infrastructure interruptions, inconsistencies, or loss of services, such as communications or network connections that result in damage to the computing environment

The *EMC NetWorker Disaster Recovery Guide* provides information on how to recover the NetWorker server, and a NetWorker client or storage node, from various types of disasters.

### Potential losses

Disaster recovery of the primary site must cover the potential loss of any or all of the following systems at the primary site:

- ◆ Avamar server where the deduplicated client backups are stored
- ◆ NetWorker storage node where the deduplication metadata for the backups is stored
- ◆ NetWorker server where the storage information (media indexes) for the backups is stored

### Disaster recovery requirements

A complete DR environment provides a secondary site with systems that copy all of the information involved in each completed backup at the primary site.

The two sites may be configured to provide disaster recovery for each other, with each serving as both a primary and secondary site with different datazones for different clients.

DR requires the maintenance of the following systems:

- ◆ DR Avamar server with deduplicated client data replicated from the primary Avamar server
- ◆ DR NetWorker storage node with deduplication metadata cloned from the primary NetWorker storage node
- ◆ DR NetWorker server with media indexes cloned from the primary NetWorker server

[Figure 2 on page 35](#) shows an example of a simple disaster recovery environment.

## Disaster recovery scenarios

The DR procedures used to recover from disaster will vary depending on the circumstances, which could include:

- ◆ The deployment of the disaster recovery environment
- ◆ Which systems are affected by the disaster
- ◆ Time required to successfully recover from the disaster

The following sections provide typical DR procedures.

### Primary Avamar server is lost, secondary Avamar server becomes primary

In this scenario, the primary NetWorker server and storage node survive the disaster, but the primary Avamar server is lost. A DR Avamar server at the secondary site contains backup data that was replicated over a WAN or IP connection from the primary site.

To recovery from this scenario:

1. Reconfigure the NetWorker server to promote the secondary Avamar server to become the primary deduplication node.

[“Create an Avamar deduplication node” on page 30](#) provides details.

2. Manually configure the NetWorker Client resources for the new deduplication node.

The new primary Avamar server at the DR site will store client backups over the WAN or IP connection.

3. After the primary site is fully restored and enabled for backup operations, demote and resynchronize the secondary site.

### Entire primary site is lost and service is restored at the secondary (DR) site

In this scenario the primary site has entirely lost its NetWorker server, NetWorker storage node, and the primary Avamar node. The secondary Avamar server is available at the disaster recover (DR) site, which will become the new primary site.

To make the DR site the new primary site:

1. Install the NetWorker software on the system at the new site.
2. Configure the NetWorker server with the storage devices and the surviving (secondary) Avamar node.

[“Create an Avamar deduplication node” on page 30](#) provides details.

3. Enable the devices with the No Mount option.
4. Run the **scanner -B device\_path** command and note the bootstrap SSID information.
5. Run the **mmrecov** utility with the bootstrap information to recover the NetWorker server:
6. After the recovery is complete, shut down the NetWorker services.
7. Rename the res folder to res.old and res.R to res.
8. Restart the NetWorker services.



# CHAPTER 6

## Monitoring and Troubleshooting

This chapter contains the following topics:

- ◆ [Viewing Avamar deduplication events in NMC](#) ..... 48
- ◆ [Refreshing Avamar deduplication node information](#) ..... 49
- ◆ [Troubleshooting](#) ..... 49

## Viewing Avamar deduplication events in NMC

NMC can be enabled to view the following types of Avamar events:

- ◆ **hfscheck** — Validation of the data stored on the Avamar server.
- ◆ **Disk health** — Status of “bad” or “failed” disks on the Avamar server.
- ◆ **Capacity** — Monitor or report on the available free space on the deduplication node. Issue an alert when the threshold is met.
- ◆ **Replication events** that occur on the Avamar server.

To view these events in NMC:

1. Log in to the root domain of the Avamar server in the role of Administrator.
2. Start the **Avamar Administrator** program.
3. Click **Administration** or under **Navigation**, select **Administration**.
4. In the **Administration** window, **Event Management** tab, select the **Unacknowledged Events** tab.
5. Select the event types you wish to monitor.
6. Under **Actions** > **Event Management**, select **Acknowledge Unacknowledged Events**.  
The events are removed from the **Unacknowledged Events** list.
7. On NMC, open the **Console** window and click **Events** to view the selected Avamar events.

## View deduplication save sets with the mminfo command

Deduplicated save set and clients can be queried the same as for standard NetWorker save sets and clients.

Use the **mminfo** program to view the group completion report and error status messages. The **dedupe** flag identifies which save sets were created with Avamar deduplication.

To limit the output of **mminfo** to only deduplicated save sets, use the **-q dedupe** option. All empty or “skipped” save sets are treated as non-deduplicated save sets and will not be listed.

For example, to query deduplication save sets for the local host, use the following command:

```
mminfo -S -q dedupe
ssid=4246722279 savetime=10/27/07 04:35:07 PM (1193258967)
scrip2b081.legato.com:/
  *Data set size: 81,035,784,884.00;
  *De-Dup session id_ 1;
  *De-Dup snapup time: 2007-10-24;
  *Deduplication: Yes;
  *De-Duplication host: scrip2d089.legato.com;
    *Domain: /NetWorker/scrip2b081.legato.com;
  *New data on Dedupe Node: 1,819,753,069.00;
  *New files: 103426;
  *Replication host: scrip2d088.legato.com;
  *Size on Dedupe Node: 4,748,760,053.00;
  *group: Test
```



---

**Note:** The NMC and the optional EMC Data Protection Advisor (DPA) software also provide detailed reports on deduplicated clients and save sets.

---

## Cleanup of deleted deduplication save sets

When deduplication save sets expire or are manually deleted, space on the Avamar deduplication node is not freed up until a periodic cleanup operation is performed. The NetWorker software performs the following procedure:

1. Marks the save sets to be deleted in the deduplication node resource.

---

**Note:** If the resource no longer exists, the save sets will not be removed.

---

2. Deletes the media database records for the save sets.
3. Upon periodic cleanup, instructs the Avamar server to delete the marked save sets.

## Refreshing Avamar deduplication node information

In NMC, the Deduplication Nodes table reads Name and Replication Node values directly from the NetWorker server. Other information may need to be manually updated from the Avamar server.

To update NMC for the deduplication node information:

1. In NMC, run the **NetWorker Administration** program and open the **Devices** view.
2. Right-click **Avamar Deduplication Nodes** in the navigation tree, or right-click one or more nodes in the **Avamar Deduplication Nodes** table, and select **Refresh**.

---

**Note:** A refresh performed from the navigation tree refreshes all deduplication nodes, which can take considerable time if there are many nodes.

---

## Troubleshooting

The following sections will help you to identify and resolve common configuration and operation issues.

### Restore operation fails

A restore fails. For example a disaster recovery with Windows OSSR may fail to recover “critical volumes.”

To resolve, ensure that the client has access to the Avamar server.

## Avamar server fills with expired save sets that were not deleted

Expired save sets might not be removed from the Avamar server. In some cases the server fills to capacity and requires assistance from EMC Customer Support to resolve. The following situations can contribute to this problem.

### Replication node is not properly configured in NetWorker

The Avamar replication node configured in NetWorker must use the FQDN of the Avamar */destination\_directory:/REPLICATE/node\_name*.

[“Creating a replication node” on page 36](#) provides details.

### Requested deletions become out of sync

Save sets on the NetWorker deduplication node can become out of sync with the save sets on the replication node.

For example, NetWorker sends a request to delete a save set and while the deletion is completed on the replication node, a blackout window prevents the deletion on the NetWorker deduplication node. When the next replication is performed, the save set is recopied from the NetWorker deduplication node to the replication node. This save set will never be deleted from the replication node.

Until this synchronization problem is resolved, a workaround can be used. Install a crontab script that replays the deletions on the replication node three days back in time.

For example:

```
cd /nsr/logs
rm todelete.sh
grep "bin/nsravamar -u" nsravamar.raw|grep `date -d now +%m/%d/%Y` | grep -v
  deleted|grep -v "\-D"| cut -d " " -f 4- >> todelete.sh
grep "bin/nsravamar -u" nsravamar.raw|grep `date -d now-1days +%m/%d/%Y` | grep -v
  deleted|grep -v "\-D"| cut -d " " -f 4- >> todelete.sh
grep "bin/nsravamar -u" nsravamar.raw|grep `date -d now-2days +%m/%d/%Y` | grep -v
  deleted|grep -v "\-D"| cut -d " " -f 4- >> todelete.sh
ex todelete.sh << 'EOF'
%s/xxxxxxx/EBRub{3 -D 1/
wq
EOF
chmod +x todelete.sh
./todelete.sh
```

# GLOSSARY

## A

<b>administrator</b>	Person who normally installs, configures, and maintains software on network computers, and who adds users and defines user privileges.
<b>advanced file type device (AFTD)</b>	Disk storage device that uses a volume manager to enable multiple concurrent backup and recovery operations and dynamically extend available disk space.
<b>attribute</b>	Feature of a NetWorker resource. It is a setting or information that the resource provides.
<b>authorization code</b>	Unique code that in combination with an associated enabler code unlocks the software for permanent use on a specific host computer. <a href="#">See also "license key."</a>
<b>Avamar Administrator</b>	Graphical management console software application that is used to remotely administer an Avamar system from a supported Windows or client computer.
<b>Avamar Datastore System (ADS)</b>	System of rack-mounted storage computers that use Avamar software and Datastore hardware. The ADS may consist of a utility node (Avamar server), one or more storage nodes, and other optional nodes such as replication nodes, which all communicate with each other over their own dedicated network switch.

## B

<b>backup</b>	<ol style="list-style-type: none"><li>1. Duplicate of database or application data, or entire computer system, stored separately from the original, which can be used to recover the original if it is destroyed or damaged.</li><li>2. Operation that saves data to a volume for use as a backup.</li></ol>
<b>backup group</b>	<a href="#">See "group."</a>
<b>backup volume</b>	<a href="#">See "volume."</a>
<b>bootstrap</b>	Save set that is essential for NetWorker disaster recovery procedures with the NetWorker server. The bootstrap consists of three components that reside on the NetWorker server: the media database, the resource database, and a server index.
<b>browse policy</b>	NetWorker policy that specifies the period of time during which backup entries are retained in the client file index. The index makes the associated backed-up data readily accessible for recovery by desktop users. <a href="#">See also "retention policy."</a>

## C

<b>client</b>	Computer, workstation, or application server whose data can be backed up and restored with the NetWorker or NMDA software.
---------------	--

<b>client file index</b>	Database maintained by the NetWorker server that tracks every database object, file, or file system backed up. The NetWorker server maintains a single index file for each client computer. The tracking information is purged from the index after the browse time of each backup expires.
<b>Client resource</b>	NetWorker server resource that identifies the save sets to be backed up on a client. The Client resource also specifies information about the backup, such as the schedule, browse policy, and retention policy for the save sets.
<b>clone</b>	Duplicate copy of backed-up data, which is indexed and tracked by the NetWorker server. Single save sets or entire volumes can be cloned.
<b>clone volume</b>	Exact duplicate of a backup or archive volume. NetWorker software can track four types of volumes (backup, archive, backup clone, and archive clone). Save sets of these different types may not be intermixed on one volume. Clone volumes may be used in exactly the same way as the original backup or archive volume.
<b>Console server</b>	<a href="#">See “NetWorker Management Console (NMC).”</a>

## D

<b>daemon</b>	Process on UNIX systems that runs in the background and performs a specified operation at predefined times or in response to certain events.
<b>database</b>	<ol style="list-style-type: none"> <li>1. A collection of data arranged for ease and speed of update, search, and retrieval by computer software.</li> <li>2. An instance of a database management system (DBMS), which in a simple case might be a single file containing many records, each of which contains the same set of fields.</li> </ol>
<b>dataset</b>	Policy that defines a set of files, directories, and file systems that will be included or excluded in backups across a group of clients. A dataset is a persistent and reusable Avamar policy that can be named and attached to multiple groups.
<b>datazone</b>	Group of clients, storage devices, and storage nodes that are administered by a NetWorker server.
<b>deduplication</b>	Process used to compress redundant data.
<b>deduplication backup</b>	Type of backup in which redundant data blocks replaced by metadata pointers and only unique blocks of data are stored. When the deduplicated data is restored, the data is returned to its original native format.
<b>deduplication node</b>	EMC Avamar server that stores deduplication backup data.
<b>deduplication ratio</b>	Reduction in storage space required to store data as a result of deduplication technology, usually combined with data compression, for example, a 20:1 space reduction.
<b>device</b>	<ol style="list-style-type: none"> <li>1. Storage folder or storage unit that can contain a backup volume. A device can be a tape device, optical drive, autochanger, or disk connected to the server or storage node.</li> <li>2. Access path to the physical drive, when dynamic drive sharing (DDS) is enabled.</li> </ol>

3. General term that refers to storage hardware.

**disaster recovery** Recovery from any disruptive situation, such as hardware failure or software corruption, in which ordinary data recovery procedures are not sufficient to restore a system and its data to normal day-to-day operations.

**domain** Feature in Avamar Administrator that is used to organize large numbers of clients into named areas of control and management.

**domain name server (DNS)** Dynamic and distributed directory service for assigning domain names to specific IP addresses.

## E

**enabler code** Unique code that activates the software:

- ◆ Evaluation enablers or temporary enablers, such as a Powerlink entitlement for NetWorker, expire after a fixed period of time.
- ◆ Base enablers unlock the basic features for software.
- ◆ Add-on enablers unlock additional features or products, for example, library support.

[See also "license key."](#)

## F

**failover** Safeguard capability in a cluster network that automatically switches activity from a failed or abnormally terminated computer server, disk drive, or network to a redundant standby server, drive, or network, with little or no disruption of service. Failover is a feature of systems that require high reliability and continuous availability.

## G

**group**

1. Client computer or group of clients that are configured to back up files during a NetWorker scheduled backup, according to a single designated schedule or set of conditions.
2. Level of organization in Avamar Administrator for one or more Avamar clients. All clients in an Avamar group use the same group policies (dataset, backup schedule, and retention policy).

## H

**hash file system (HFS)** The content addressed storage area inside the Avamar server that is used to store client backups.

**host** Computer on a network.

**hostname** Name or address of a physical or virtual host computer that is connected to a network.

## I

**index** [See "media index."](#)

**L**

- label** Electronic header on a volume used for identification by NetWorker or other data mover application.
- license key** Combination of an enabler code and authorization code for a specific product release to permanently enable its use. Also called an activation key or license enabler.

**M**

- managed application** Program that can be monitored or administered, or both from the Console server.
- media** Physical storage, such as magnetic tape, optical disk, or file system, to which backup data is written. [See also "volume."](#)
- media index** Database that contains indexed entries of storage volume location and the lifecycle status of all data and volumes managed by the NetWorker server. Also known as media database.
- media pool** [See "pool."](#)
- metadata** Hash information that identifies stored sub-file information for deduplication, and is required to revert deduplicated client backup data to the regular non-deduplicated format.

**N**

- NetWorker Management Console (NMC)** Software product that is used to manage NetWorker servers and clients. The NMC server also provides reporting and monitoring capabilities for all NetWorker processes.
- NetWorker server** Computer on a network that runs the NetWorker server software, contains the online indexes, and provides backup and restore services to the clients and storage nodes on the same network.
- node** Networked storage subsystem with both processing power and disk storage, running Avamar software.
- notification** Message sent to the NetWorker administrator about important NetWorker events.

**O**

- online indexes** Databases located on the NetWorker server that contain all the information pertaining to the client backups (client file index) and backup volumes (media database).

**P**

- pathname** Set of instructions to the operating system for accessing a file. An absolute pathname indicates how to find a file starting from the root directory and working down the directory tree. A relative pathname indicates how to find a file starting from the current location.
- policy** Set of defined rules for client backups that can be named and applied to multiple groups. Groups have dataset, schedule, and retention policies.

- pool**
1. NetWorker feature that assigns specific backup data to be stored on selected media volumes.
  2. Collection of NetWorker backup volumes to which specific data has been backed up.

## R

- recover** To restore data files from a backup volume to a client disk and apply transactional (redo) logs to the data to make it consistent with a given point in time.
- remote device**
1. Storage device that is attached to a storage node that is separate from the NetWorker server.
  2. Storage device located at an offsite location that stores a copy of data from a primary storage device for disaster recovery.
- replication** Optional feature that allows one Avamar server to store a read-only copy of its data on another Avamar server in order to support future disaster recovery of that server.
- replication node** EMC Avamar server that stores replication copies of Avamar utility node and storage node data for the purpose of disaster recovery of those nodes.
- resource** Software component that describes details of the NetWorker server or its clients. Clients, devices, schedules, groups, and policies are all NetWorker resources. Each resource has configurable attributes that define its properties.
- resource database** NetWorker database of information about each configured resource.
- restore** To retrieve individual data files from backup media and copy the files to disk, without applying transaction logs. [See also "recover."](#)
- retention policy** NetWorker setting that determines the minimum period of time that backup data is retained on a volume available for recovery. After this time is exceeded, the data is eligible to be overwritten. [See also "browse policy."](#)
- retrieve** To locate and recover archived files and directories.

## S

- save** NetWorker command that backs up client files to back up media volumes and makes data entries in the online index.
- save set** Group of files or a file system that has been backed up on storage media by using the NetWorker software.
- save set ID (ssid)** Internal identification number assigned to a save set.
- save stream** Data and save set information that is written to a storage volume during a backup. A save stream originates from a single save set.
- scheduled backup** Type of backup that is configured to start automatically at a specified time for a group of one or more NetWorker clients. A scheduled backup generates a bootstrap save set.
- ssid** [See "save set ID \(ssid\)."](#)

**storage device** See “[device.](#)”

**storage node** Storage device physically attached to a computer other than the NetWorker server, whose backup operations are administered from the controlling NetWorker server.

## T

**trap** Setting of an SNMP event management program to report an error or status message.

## U

**utility node** EMC Avamar server that stores metadata from Avamar backup and replication operations and provides processes and services such as management console, scheduler, name services, external authentication, time protocol, and web access.

## V

- volume**
1. Unit of physical storage medium, such as a magnetic tape, optical disk, or file system used to store data. Backup data must be stored on a backup volume and cannot be stored on an archive volume or a clone volume.
  2. Identifiable unit of data storage that may reside on one or more computer disks.

**volume ID (volid)** Internal identification that NetWorker software assigns to a backup volume.

**volume name** Name that you assign to a backup volume when it is labeled. See also “[label.](#)”

**VTL** Virtual tape library. Software emulation of a physical tape library storage system.