Symantec NetBackup™ Deduplication Guide

Release 7.0

20654102

Symantec NetBackup™ Deduplication Guide

Documentation version 7.0

PN: 20654102

Legal Notice

Symantec Corporation
350 Ellis Street
Mountain View, CA 94043

http://www.symantec.com

Technical Support

Symantec Technical Support maintains support centers globally. Technical Support's primary role is to respond to specific queries about product features and functionality. The Technical Support group also creates content for our online Knowledge Base. The Technical Support group works collaboratively with the other functional areas within Symantec to answer your questions in a timely fashion. For example, the Technical Support group works with Product Engineering and Symantec Security Response to provide alerting services and virus definition updates.

Symantec's maintenance offerings include the following:

- A range of support options that give you the flexibility to select the right amount of service for any size organization
- Telephone and Web-based support that provides rapid response and up-to-the-minute information
- Upgrade assurance that delivers automatic software upgrade protection
- Global support that is available 24 hours a day, 7 days a week
- Advanced features, including Account Management Services

For information about Symantec's Maintenance Programs, you can visit our Web site at the following URL:

www.symantec.com/techsupp/

Contacting Technical Support

Customers with a current maintenance agreement may access Technical Support information at the following URL:

www.symantec.com/techsupp/

Before contacting Technical Support, make sure you have satisfied the system requirements that are listed in your product documentation. Also, you should be at the computer on which the problem occurred, in case it is necessary to replicate the problem.

When you contact Technical Support, please have the following information available:

- Product release level
- Hardware information
- Available memory, disk space, and NIC information
- Operating system

- Version and patch level
- Network topology
- Router, gateway, and IP address information
- Problem description:
  - Error messages and log files
  - Troubleshooting that was performed before contacting Symantec
  - Recent software configuration changes and network changes

Licensing and registration

If your Symantec product requires registration or a license key, access our technical support Web page at the following URL:

www.symantec.com/techsupp/

Customer service

Customer service information is available at the following URL:

www.symantec.com/techsupp/

Customer Service is available to assist with the following types of issues:

- Questions regarding product licensing or serialization
- Product registration updates, such as address or name changes
- General product information (features, language availability, local dealers)
- Latest information about product updates and upgrades
- Information about upgrade assurance and maintenance contracts
- Information about the Symantec Buying Programs
- Advice about Symantec's technical support options
- Nontechnical presales questions
- Issues that are related to CD-ROMs or manuals

Maintenance agreement resources

If you want to contact Symantec regarding an existing maintenance agreement, please contact the maintenance agreement administration team for your region as follows:

Asia-Pacific and Japan	customercare_apac@symantec.com
Europe, Middle-East, and Africa	semea@symantec.com
North America and Latin America	supportsolutions@symantec.com

Additional enterprise services

Symantec offers a comprehensive set of services that allow you to maximize your investment in Symantec products and to develop your knowledge, expertise, and global insight, which enable you to manage your business risks proactively.

Enterprise services that are available include the following:

Symantec Early Warning Solutions	These solutions provide early warning of cyber attacks, comprehensive threat analysis, and countermeasures to prevent attacks before they occur.
Managed Security Services	These services remove the burden of managing and monitoring security devices and events, ensuring rapid response to real threats.
Consulting Services	Symantec Consulting Services provide on-site technical expertise from Symantec and its trusted partners. Symantec Consulting Services offer a variety of prepackaged and customizable options that include assessment, design, implementation, monitoring, and management capabilities. Each is focused on establishing and maintaining the integrity and availability of your IT resources.
Educational Services	Educational Services provide a full array of technical training, security education, security certification, and awareness communication programs.

To access more information about Enterprise services, please visit our Web site at the following URL:

www.symantec.com

Select your country or language from the site index.

Contents

Chapter 1

Introducing NetBackup deduplication

This chapter includes the following topics:

About NetBackup deduplication

The proprietary Symantec PureDisk deduplication technology powers NetBackup integrated deduplication. Symantec packaged PureDisk into modular components. The components plug-in to NetBackup through the NetBackup OpenStorage framework.

With these components, Symantec NetBackup provides the deduplication options that let you deduplicate data everywhere, as close to the source of data as you require.

Deduplication everywhere provides significant return on investment, as follows.

- Reduce the amount of data that is stored.
- Reduce backup bandwidth.
  Reduced bandwidth can be especially important when you want to limit the amount of data that a client sends over the network. Over the network can be to a backup server or for image duplication between remote locations.
- Reduce backup windows.
- Reduce infrastructure.

About NetBackup deduplication options

Deduplication everywhere lets you choose at which point in the backup process to perform deduplication. NetBackup can manage your deduplication wherever you implement it in the backup stream.

Figure 1-1 shows the options for deduplication.

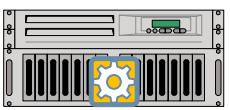
Table 1-1 describes the options for deduplication.

Figure 1-1 NetBackup deduplication

Deduplicate on NetBackup clients

Deduplicate on NetBackup media servers

Deduplicate on disk appliances by using the OpenStorage option



Deduplicate using NetBackup PureDisk, including at remote offices

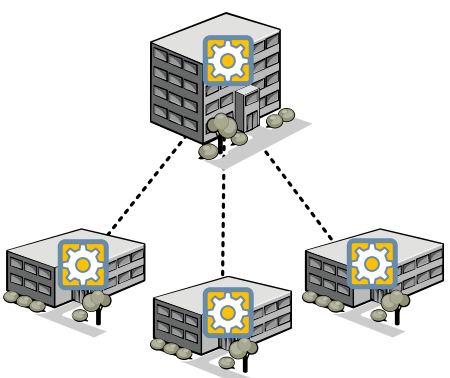


Table 1-1 NetBackup deduplication options

Type	Description
NetBackup Client Deduplication Option	With NetBackup client-side deduplication, clients deduplicate their backup data and then send it directly to the storage destination. A media server does not deduplicate the data. NetBackup Client Deduplication is a useful deduplication solution if a client host has unused CPU cycles or if the load balancing servers are overloaded. See "About NetBackup Client Deduplication" on page 24.
NetBackup Media Server Deduplication Option	NetBackup clients send their backups to a NetBackup media server, which deduplicates the backup data. A NetBackup media server hosts the NetBackup Deduplication Engine, which writes the data to the storage and manages the deduplicated data. NetBackup Media Server Deduplication is a useful deduplication solution if a client does not have enough CPU cycles to deduplicate its own data. See "About the NetBackup Media Server Deduplication Option" on page 19.
Appliance deduplication	The NetBackup OpenStorage option lets third-party vendor appliances function as disk storage for NetBackup. The disk appliance provides the storage and it manages the storage. A disk appliance may provide deduplication functionality. NetBackup backs up and restores client data and manages the life cycles of the data. Appliance deduplication is a storage optimization or reduction strategy. It reduces the storage that you may require. See "How deduplication works" on page 14. Conversely, NetBackup integrated deduplication reduces storage requirements and provides other benefits that a disk appliance deduplication solution cannot. See "About NetBackup deduplication" on page 11.

Table 1-1 NetBackup deduplication options *(continued)*

Type	Description
PureDisk deduplication	NetBackup PureDisk is a deduplication solution for bandwidth-optimized backups of data in remote offices. You can use PureDisk to reduce the amount of backup data that is stored in a datacenter by NetBackup. You use PureDisk interfaces to install, configure, and manage the PureDisk servers, storage pools, and client backups. You do not use NetBackup to configure or manage the storage or backups. PureDisk has its own documentation set. See the *NetBackup PureDisk Getting Started Guide*. A PureDisk storage pool can be a storage destination for both the NetBackup Client Deduplication Option and the NetBackup Media Server Deduplication Option.

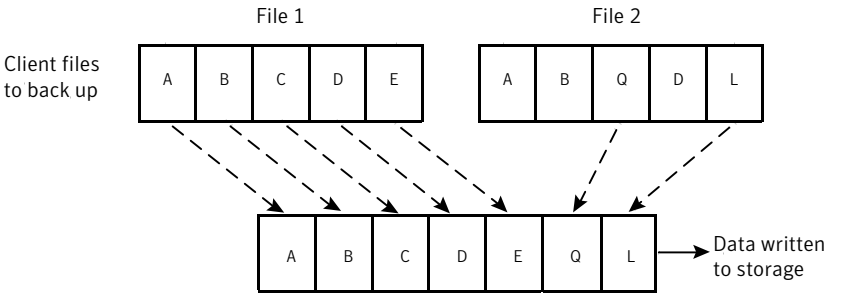
How deduplication works

Deduplication is a method of retaining only one unique instance of backup data on storage media. Redundant data is replaced with a pointer to the unique data copy. Deduplication occurs on both a file level and a file segment level. When two or more files are identical, deduplication stores only one copy of the file. When two or more files share identical content, deduplication breaks the files into segments and stores only one copy of each unique file segment.

Deduplication significantly reduces the amount of storage space that is required for the NetBackup backup images.

Figure 1-2 is a diagram of file segments that are deduplicated.

Figure 1-2 File deduplication



The following list describes how NetBackup derives unique segments to store:

- The deduplication engine breaks file 1 into segments A, B, C, D, and E.
- The deduplication engine breaks file 2 into segments A, B, Q, D, and L.
- The deduplication engine stores file segments A, B, C, D, and E from file 1 and file segments Q, and L from file 2. The deduplication engine does not store file segments A, B, and D from file 2. Instead, it points to the unique data copies of file segments A, B, and D that were already written from file 1.

More detailed information is available.

See "Media server deduplication process" on page 111.

Chapter 2

Planning your deployment

This chapter includes the following topics:

Planning your deduplication deployment

Table 2-1 provides an overview of planning your deployment of NetBackup deduplication.

Table 2-1 Deployment overview

Deployment task	Where to find the information
Determine the storage type	See "About the deduplication storage type" on page 19.
Determine which type of deduplication to use	See "About the NetBackup Media Server Deduplication Option" on page 19. See "About NetBackup Client Deduplication" on page 24.
Determine the requirements for deduplication hosts	See "About deduplication servers" on page 21. See "About deduplication server requirements" on page 23. See "About client deduplication host requirements" on page 25. See "About the network interface for deduplication" on page 27. See "About firewalls and the deduplication hosts" on page 27. See "About scaling deduplication" on page 27. See "About deduplication performance" on page 32.
Determine the credentials for deduplication	See "About NetBackup Deduplication Engine credentials" on page 26.
Read the compression and encryption recommendation	See "About compression and encryption" on page 28.
Determine the requirements for optimized duplication	See "About optimized duplication of deduplicated data" on page 28.
Determine the storage requirements and provision the storage	See "About provisioning the storage" on page 37. See "About deduplication storage requirements" on page 37. See "About deduplication storage capacity" on page 38. See "About the deduplication storage paths" on page 38.

Table 2-1 Deployment overview *(continued)*

Deployment task	Where to find the information
Replace a PDDO host or migrate from PDDO to NetBackup deduplication	See "Replacing the PureDisk Deduplication Option with Media Server Deduplication on the same host" on page 33. See "Migrating from PureDisk to the NetBackup Media Server Deduplication option" on page 34.
Migrate from other storage to NetBackup deduplication	See "Migrating from another storage type to deduplication" on page 35.

About the deduplication storage type

The deduplication storage type depends on the destination for the deduplicated data, as follows:

- The disk storage that is attached to a NetBackup media server.
  If you use this destination, use this guide to plan, implement, configure, and manage deduplication and the storage. When you configure the storage server, select **Media Server Deduplication Pool** as the storage type.
- A PureDisk storage pool.
  If you use a PureDisk storage pool, use the PureDisk documentation to plan, implement, configure, and manage the storage.
  NetBackup deduplication requires that PureDisk be at release 6.6 or later.
  See the *NetBackup PureDisk Getting Started Guide*.
  After you configure the storage, use this guide to configure backups and deduplication in NetBackup. When you configure the storage server, select **PureDisk Deduplication Pool** as the storage type.

You can use one or both of the destinations for NetBackup deduplication.

About the NetBackup Media Server Deduplication Option

NetBackup Media Server Deduplication Option exists in the Symantec OpenStorage framework. A storage server writes data to the storage and reads data from the storage; the storage server must be a NetBackup media server. The storage server hosts the core components of deduplication. The storage server also deduplicates the backup data. It is known as a deduplication storage server.

For a backup, the NetBackup client software creates the image of backed up files as for a normal backup. The client sends the backup image to the deduplication storage server, which deduplicates the data. The deduplication storage server writes the data to disk.

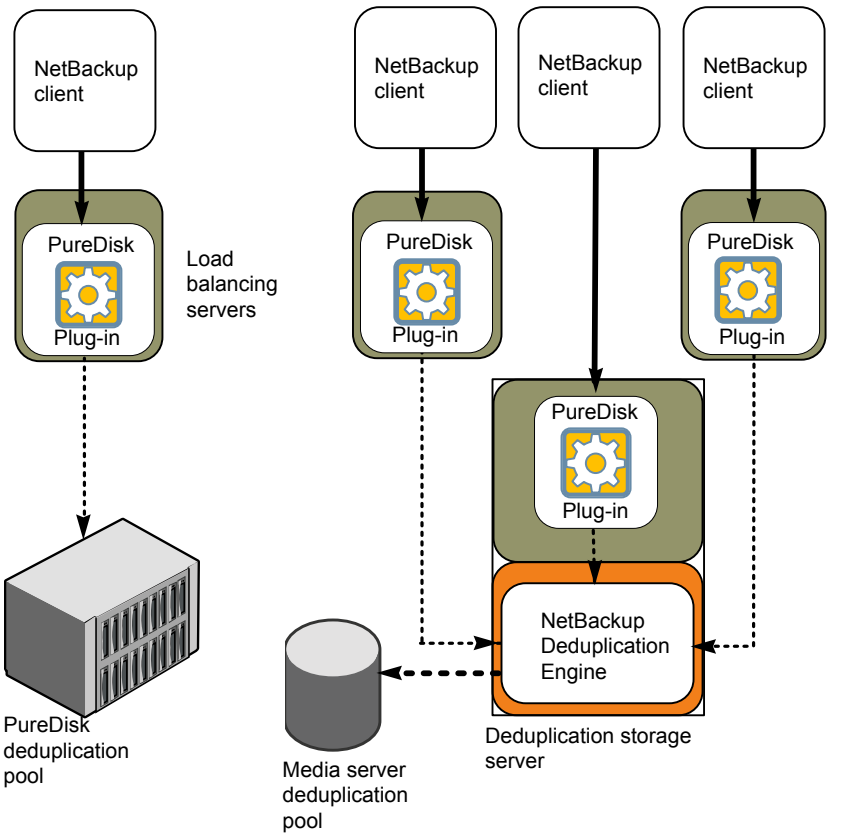
See "About deduplication servers" on page 21.

The NetBackup Media Server Deduplication Option is integrated into NetBackup. It uses the NetBackup administration interfaces, commands, and processes for configuring and executing backups and for configuring and managing the storage. Deduplication occurs when NetBackup backs up a client to a deduplication storage destination. You do not have to use the separate PureDisk interfaces to configure and use deduplication.

The NetBackup Media Server Deduplication Option integrates with NetBackup application agents that are optimized for the client stream format. Agents include but are not limited to Microsoft Exchange and Microsoft SharePoint Agents.

Figure 2-1 shows NetBackup media server deduplication. The deduplication storage server is a media server on which the deduplication core components are enabled.

Figure 2-1 NetBackup media server deduplication



More detailed information is available.

See "Deduplication server components" on page 109.

See "Media server deduplication process" on page 111.

About deduplication servers

Table 2-2 describes the servers that are used for NetBackup deduplication.

Table 2-2 NetBackup deduplication servers

Host	Description
Deduplication storage server	One host functions as the storage server for a deduplication node; that host must be a NetBackup media server. The storage server does the following: ■ Writes the data to and reads data from the disk storage. ■ Manages that storage. The storage server also deduplicates data. Therefore, one host both deduplicates the data and manages the storage. Only one storage server exists for each NetBackup deduplication node. You can use NetBackup deduplication with one media server host only: the media server that is configured as the deduplication storage server.
Load balancing server	You can configure other NetBackup media servers to help deduplicate data. They perform file fingerprint calculations for deduplication, and they send the unique results to the storage server. These helper media servers are called load balancing servers. See "About deduplication fingerprinting" on page 117. You configure load balancing servers when you configure the deduplication storage server. Also, you can add a deduplication server later to a deduplication node. Load balancing servers also perform restore and duplication jobs. Symantec recommends that you add load balancing servers only after the storage server reaches maximum CPU utilization. For more information about how to use load balancing servers, see the following Symantec tech note: http://entsupport.symantec.com/docs/338123

About deduplication nodes

A media server deduplication node is a deduplication storage server, load balancing servers (if any), the clients that are backed up, and the storage. Each node manages its own storage. Deduplication within each node is supported; deduplication between nodes is not supported.

Multiple media server deduplication nodes can exist. Nodes cannot share servers, storage, or clients.

About deduplication server requirements

All hosts that are used for deduplication must be NetBackup 7.0 or later. Hosts include the master server, the media servers, and the clients.

The computer's CPU and memory constrain how many jobs can run concurrently.

Table 2-3 Deduplication server minimum requirements

Hardware	Requirement
CPU	CPU speed is the most important factor for performance. Minimum CPU speed should be 2.2 GHz. The deduplication storage server should have a minimum of 4 CPU cores. Symantec recommends eight cores. Symantec recommends Intel, AMD, and Sun SPARC processors (in order of effectiveness).
RAM	Symantec recommends 4 GBs of memory minimum.
Operating system	The operating system must be a supported 64-bit operating system. For supported systems, see the *NetBackup Release Notes*.

Note: Symantec recommends that you do not use the master server as a deduplication storage server. Master server activity and media server deduplication activity on the same host degrades performance.

Note: Symantec recommends that you do not use an existing media server for deduplication. Similarly, Symantec recommends that you do not repurpose older host hardware for deduplication.

About media server deduplication limitations

NetBackup media server deduplication and Symantec Backup Exec deduplication cannot reside on the same host. If you use both NetBackup and Backup Exec deduplication, each product must reside on a separate host.

NetBackup deduplication components cannot reside on the same host as a PureDisk Deduplication Option (PDDO) agent. Therefore, you cannot use the same media server for both NetBackup deduplication and as a PDDO host.

You cannot upgrade to NetBackup 7.0 or later a NetBackup media server that hosts a PDDO agent. If the NetBackup 7.0 installation detects the PDDO agent,

the installation fails. To upgrade a NetBackup media server that hosts a PDDO agent, you must first remove the PDDO agent.

See the *NetBackup PureDisk Deduplication Option (PDDO) Guide*.

Deduplication within each media server deduplication node is supported; global deduplication between nodes is not supported.

About NetBackup Client Deduplication

With normal deduplication, the client sends the full backup data stream to the media server. The deduplication engine on the media server processes the stream, saving only the unique segments.

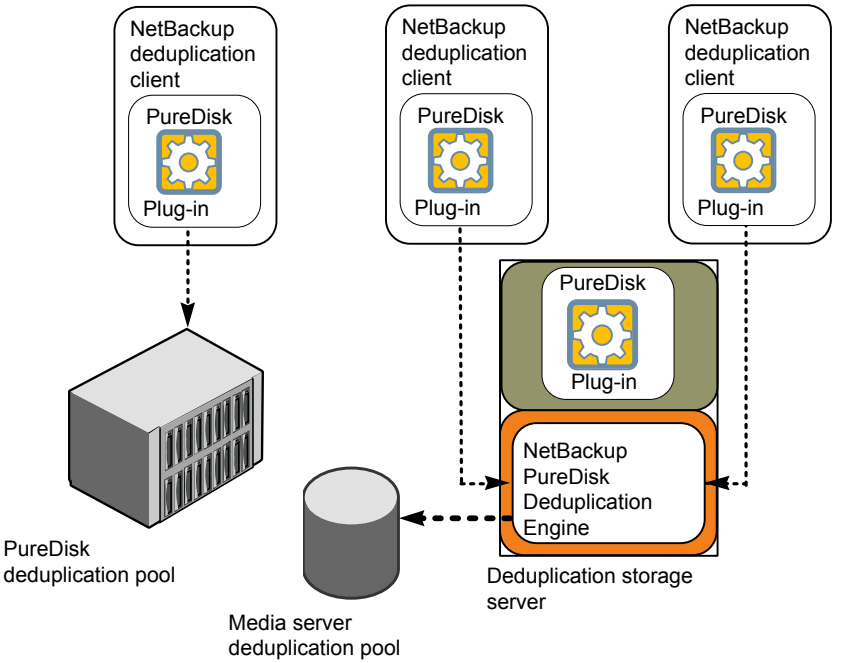
With NetBackup Client Deduplication, the client hosts the PureDisk plug-in that duplicates the backup data. The NetBackup client software creates the image of backed up files as for a normal backup. Next, the PureDisk plug-in breaks the backup image into segments and compares them to all of the segments that are stored in that deduplication node. The plug-in then sends only the unique segments to the NetBackup Deduplication Engine on the storage server. The engine writes the data to a media server deduplication pool.

Client deduplication does the following:

- Reduces network traffic. The client sends only unique file segments to the storage server. Duplicate data is not sent over the network.
- Distributes some deduplication processing load from the storage server to clients. (NetBackup does not balance load between clients; each client deduplicates its own data.)

Figure 2-2 shows client deduplication. The deduplication storage server is a media server on which the deduplication core components are enabled.

Figure 2-2 NetBackup client deduplication



More detailed information is available.

See "Deduplication client components" on page 114.

See "Deduplication client backup process" on page 114.

About client deduplication host requirements

The operating system must be a supported 64-bit operating system.

For supported systems, see the *NetBackup Release Notes*.

About client deduplication requirements

All hosts that are used for client deduplication must be NetBackup 7.0 or later.

A media server deduplication pool or a PureDisk deduplication pool must be configured. Storage units must be configured for the deduplication pool.

About client deduplication limitations

Client deduplication does not support multiple copies per job. For the jobs that specify multiple copies, the backup images are sent to the storage server and may be deduplicated there.

Client deduplication does not support encryption.

Client deduplication is not tolerant of high latency network connections. Therefore, Symantec recommends that you use NetBackup PureDisk for remote office backups.

About NetBackup Deduplication Engine credentials

The NetBackup Deduplication Engine requires credentials. The deduplication components use the credentials when they communicate with the NetBackup Deduplication Engine. The credentials are for the engine, not for the host on which it runs.

You enter the NetBackup Deduplication Engine credentials when you configure the storage server.

The following are the rules for the credentials:

- For user names and passwords, you can use characters in the printable ASCII range (0x20-0x7E) except for the following characters:
  - Asterisk (*)
  - Backward slash (\) and forward slash (/)
  - Double quote (")
  - Left parenthesis [(] and right parenthesis [)]
- The user name can be up to 127 characters in length. The password can be up to 100 characters in length.
- Leading and trailing spaces and quotes are ignored.
- The user name and password cannot be empty or all spaces.

Record and save the credentials in case you need them in the future.

Caution: You cannot change the NetBackup Deduplication Engine credentials after you enter them. Therefore, carefully choose and enter your credentials. If you must change the credentials, contact your Symantec support representative.

About the network interface for deduplication

If the server host has more than one network interface, by default the host operating system determines which network interface to use. However, you can specify which interface NetBackup should use for the deduplication traffic.

To use a specific interface, enter that interface name when you configure the deduplication storage server.

Caution: You cannot change the network interface after NetBackup configures the deduplication storage server. Therefore, carefully enter the properties.

About firewalls and the deduplication hosts

If firewalls exist between the various deduplication hosts, open ports 10082 and 10102 between those hosts. Deduplication hosts are the deduplication storage server, the load balancing servers, and the clients that deduplicate their own data.

About scaling deduplication

You can scale deduplication processing to improve performance by using load balancing servers or client deduplication or both.

If you configure load balancing servers, those servers also perform deduplication. The deduplication storage server still functions as both a deduplication server and as a storage server. NetBackup uses standard load balancing criteria to select a load balancing server for each job. However, deduplication fingerprint calculations are not part of the load balancing criteria.

To completely remove the deduplication storage server from deduplication duties, do the following for every storage unit that uses the deduplication disk pool:

- Select **Only use the following media servers.**
- Select all of the load balancing servers but do not select the deduplication storage server.

The deduplication storage server performs storage server tasks only: storing and managing the deduplicated data, file deletion, and optimized duplication.

If you configure client deduplication, the clients deduplicate their own data. Some of the deduplication load is removed from the deduplication storage server and loading balancing servers.

Symantec recommends the following strategies to scale deduplication:

- For the initial full backups of your clients, use the deduplication storage server. For subsequent backups, use load balancing servers.
  Do not expect the deduplication storage server to be the media server that is used for restores to that client. If a media server deduplicates a client backup, your restore settings may require that the media server also be used for restores to that client.
- Enable client-side deduplication gradually.
  If a client cannot tolerate the deduplication processing workload, be prepared to move the deduplication processing back to a server.

See "About deduplication performance" on page 32.

About compression and encryption

For compression or encryption, Symantec recommends that you enable them so they occur during the NetBackup deduplication process. If you compress or encrypt the data before it is deduplicated, deduplication rates are low.

See "About the deduplication configuration file" on page 60.

See "Editing the deduplication configuration file" on page 60.

See "pd.conf file settings" on page 61.

About optimized duplication of deduplicated data

Optimized duplication of deduplicated data reduces the amount of data that is transmitted over your network. Therefore, you can use optimized duplication for off-site storage of data for disaster recovery. It can improve recovery times and minimize the use of off-site tape storage.

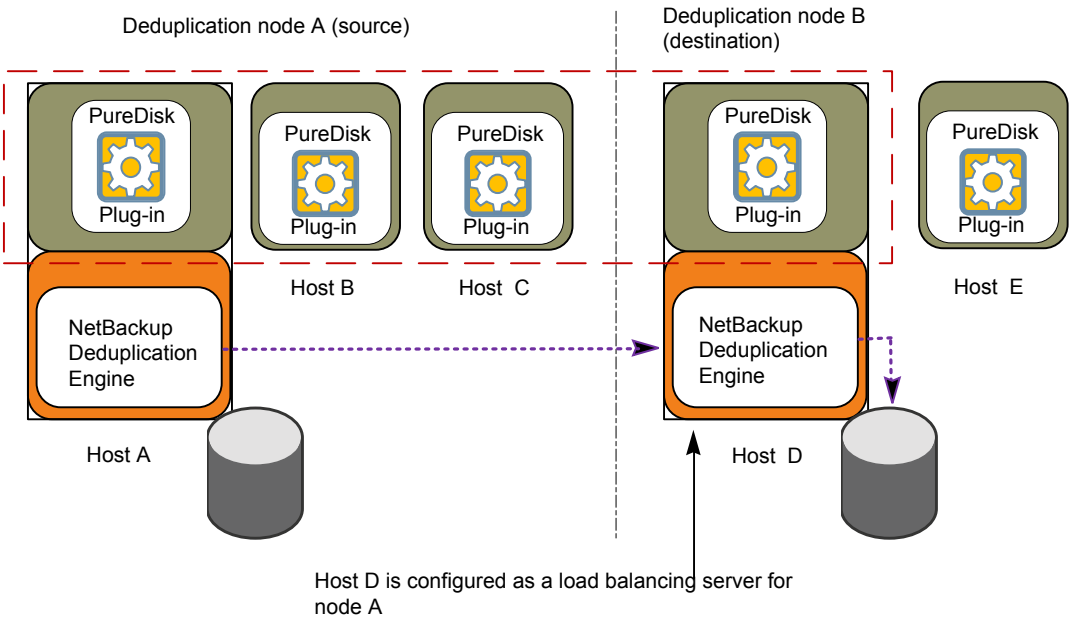
Only the unique data segments are transferred.

See "Configuring optimized deduplication copy" on page 57.

Optimized deduplication copy requirements

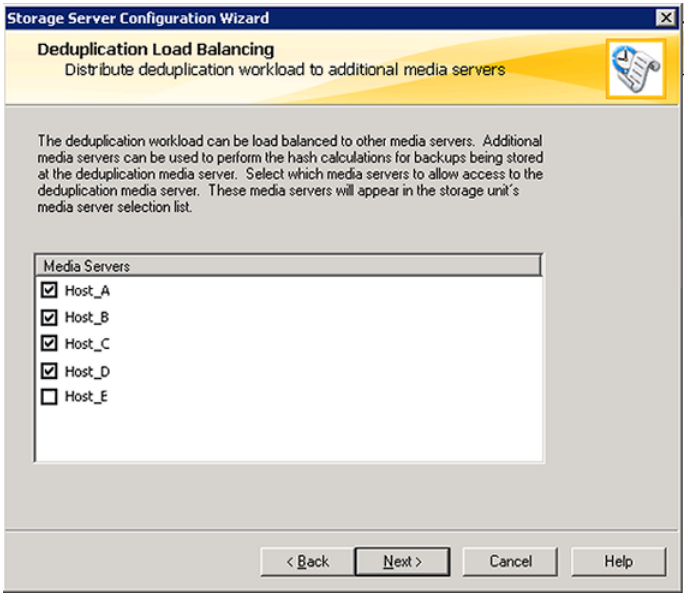
Figure 2-3 shows a source deduplication node and a destination deduplication node for optimized deduplication copy. The requirements description follows the figure.

Figure 2-3 Optimized duplication copy example



The following are the requirements for optimized duplication:

- The source images must be on a NetBackup media server deduplication pool.
- The destination disk storage can be another **Media Server Deduplication Pool** or a **PureDisk Deduplication Pool**. The destination storage unit cannot be the same as the source storage unit.
  If the destination is a **PureDisk Deduplication Pool**, the PureDisk environment must be at release level 6.6 or later.
- At least one media server must be common between the source deduplication node and the destination, as follows:
  - If the destination is another **Media Server Deduplication Pool**: Configure a server in the destination deduplication node as a load balancing server for the source storage server.
    For example, Figure 2-3 shows two deduplication nodes. Host D from the destination node is configured as a load balancing server for the source node. It is the common host. The following **Storage Server Configuration Wizard** screen shows the load balancing servers that are configured for deduplication node A:



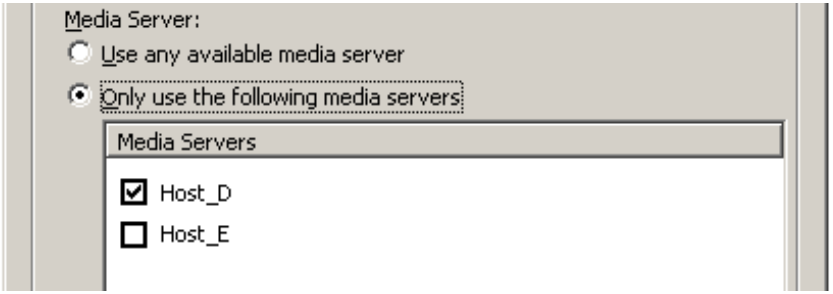
- If the destination is a **PureDisk Deduplication Pool**: Configure a media server that accesses the PureDisk Storage Pool Authority host as a load balancing server for the source storage server.

To use more than one media server for the optimized copy operation, each additional one must be common between them .If you select more than one, NetBackup balances the optimized copy job load among them.

- All of the media servers that are selected in the destination storage unit must be common with the source storage server.

  In the storage unit for the destination disk pool, select **Only use the following media servers**. Then, select the media server or media servers that are common to both the source storage server and the destination storage server.

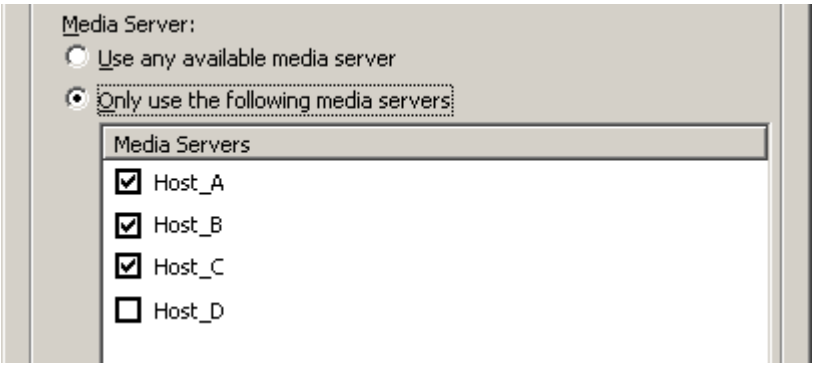
  For example, the following figure shows the destination storage unit media server selection for the optimized duplication that is show inFigure 2-3. Host D is the only common host, so it is selected in the destination storage unit.



If you use your destination storage unit to back up clients, you can create a different storage unit for those jobs. In that storage unit, select all of the hosts in that node that you want to use for deduplication.

If you select the common server from the destination node in the source node storage unit, NetBackup uses i5 for deduplication. Therefore, do not select it in the storage unit for source node (unless you want to use it for the source node).

For example, for the storage unit for the backup jobs for node A in Figure 2-3, do not select Host D (shown in the following figure):



Optimized deduplication copy limitations

The following are limitations for optimized deduplication copy:

- NetBackup does not support a PureDisk storage pool as the source for optimized duplication. Therefore, you cannot use optimized duplication from a PureDisk storage pool to a media server deduplication pool or to another PureDisk storage pool.
- If an optimized duplication job fails, NetBackup does not run the job again. See "Configuring optimized deduplication copy behavior" on page 57.

About deduplication performance

Many factors affect performance, especially the server hardware and the network capacity.

Table 2-4 provides information about performance during backup jobs for a deduplication storage server. The deduplication storage server conforms to the minimum host requirements. Client deduplication or load balancing servers are not used.

See "About deduplication server requirements" on page 23.

Table 2-4 Deduplication job load performance for a deduplication storage server

When	Description
Initial seeding	Initial seeding is when all clients are first backed up. Approximately 15 to 20 jobs can run concurrently under the following conditions: ■ The hardware meets minimum requirements. (More capable hardware improves performance.) ■ No compression. If data is compressed, the CPU usage increases quickly, which reduces the number of concurrent jobs that can be handled. ■ The deduplication rate is between 50% to 100%. The deduplication rate is the percentage of data already stored so it is not stored again. ■ The amount of data that is stored is less than 30% of the capacity of the storage.
Normal operation	Normal operation is when all clients have been backed up once. Approximately 15 to 20 jobs can run concurrently and with high performance under the following conditions: ■ The hardware meets minimum requirements. (More capable hardware improves performance.) ■ No compression. If data is compressed, the CPU usage increases quickly, which reduces the number of concurrent jobs that can be handled. ■ The deduplication rate is between 10% and 50%. The deduplication rate is the percentage of data already stored so it is not stored again. ■ The amount of data that is stored is between 30% to 90% of the capacity of the storage.

Table 2-4 Deduplication job load performance for a deduplication storage server *(continued)*

When	Description
Clean up periods	Clean up is when the NetBackup Deduplication Engine performs maintenance such as deleting expired backup image data segments. NetBackup maintains the same number of concurrent backup jobs as during normal operation. However, the average time to complete the jobs increases significantly.
Storage approaches full capacity	NetBackup maintains the same number of concurrent backup jobs as during normal operation under the following conditions: ■ The hardware meets minimum requirements. (More capable hardware improves performance.) ■ The amount of data that is stored is between 85% to 90% of the capacity of the storage. However, the average time to complete the jobs increases significantly.

How file size may affect the deduplication rate

The small file sizes that are combined with large file segment sizes may result in low initial deduplication rates. However, after the deduplication engine performs file fingerprint processing, deduplication rates improve. For example, a second backup of a client shortly after the first does not show high deduplication rates. But the deduplication rate improves if the second backup occurs after the file fingerprint processing.

How long it takes the NetBackup Deduplication Engine to process the file fingerprints varies.

Replacing the PureDisk Deduplication Option with Media Server Deduplication on the same host

You can replace a PureDisk Deduplication Option agent from its media server host with a NetBackup PureDisk plug-in on the same host. The storage remains the PureDisk storage pool, and NetBackup maintains access to all of the valid backup images in the PureDisk storage pool.

Note: The PureDisk storage pool must be part of a PureDisk 6.6 or later environment.

Table 2-5 Replacing a PDDO host with a media server deduplication host

Task	Procedure
Ensure that no activity occurs on the host	Deactivate all backup policies that use the host. See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.
Remove the PDDO plug-in	NetBackup deduplication components cannot reside on the same host as a PureDisk Deduplication Option (PDDO) agent. Therefore, remove the PDDO plug-in from the host. See the *NetBackup PureDisk Deduplication Option Guide.*
Upgrade the media server to 7.0 or later	If the media server runs a version of NetBackup earlier than 7.0, upgrade that server to NetBackup 7.0 or later. See the *NetBackup Installation Guide for UNIX and Linux.* See the *NetBackup Installation Guide for Windows.*
Configure the host	In the **Storage Server Configuration Wizard**, select **PureDisk Deduplication Pool** and enter the name of the Storage Pool Authority. See "Configuring a deduplication storage server" on page 48.
Activate your backup policies	See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.

Migrating from PureDisk to the NetBackup Media Server Deduplication option

NetBackup cannot use the storage hardware while PureDisk uses it for storage. The structure of the PureDisk storage is different than the structure of the storage for integrated NetBackup deduplication. The disk systems cannot be used simultaneously by both NetBackup and PureDisk. The PureDisk images on the storage cannot be transferred to the deduplication storage server storage.

Therefore, to migrate from NetBackup PureDisk to the NetBackup Media Server Deduplication Option, Symantec recommends that you age the PureDisk storage pool backups until they expire.

Table 2-6 To migrate from PureDisk to NetBackup deduplication

Task	Procedure
Install and configure NetBackup	See the *NetBackup Installation Guide for UNIX and Linux*. See the *NetBackup Installation Guide for Windows*.
Configure NetBackup deduplication	See "Configuring deduplication" on page 47.
Redirect your backup jobs	Redirect your backup jobs to the NetBackup media server deduplication pool. See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I*. See the *NetBackup Administrator's Guide for Windows, Volume I*.
Uninstall PureDisk	After the PureDisk backup images expire, uninstall PureDisk. See your NetBackup PureDisk documentation.

Migrating from another storage type to deduplication

To migrate from another NetBackup storage type to deduplication storage, Symantec recommends that you age the backup images on the other storage until they expire. Symantec recommends that you age the backup images if you migrate from disk storage or tape storage.

You should not use the same disk storage for NetBackup deduplication while you use it for other storage such as AdvancedDisk, BasicDisk, or SharedDisk. Each type manages the storage differently and each requires exclusive use of the storage. Also, the NetBackup Deduplication Engine cannot read the backup images that another NetBackup storage type created. Therefore, you should age the data so it expires before you repurpose the storage hardware. Until that data expires, two storage destinations exist: the media server deduplication pool and the other storage. After the images on the other storage expire and are deleted, you can repurpose it for other storage needs.

Table 2-7 Migrating to NetBackup deduplication

Task	Procedure
Configure NetBackup deduplication	See "Configuring deduplication" on page 47.

Table 2-7 Migrating to NetBackup deduplication *(continued)*

Task	Procedure
Redirect your backup jobs	Redirect your backup jobs to the media server deduplication pool storage unit. To do so, change the backup policy storage destination to the storage unit for the deduplication pool. See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I*. See the *NetBackup Administrator's Guide for Windows, Volume I*.
Repurpose the storage	After all of the backup images that are associated with the storage expire, repurpose that storage. If it is disk storage, you cannot add it to an existing media server deduplication pool. You can use it as storage for another, new deduplication node.

Chapter 3

Provisioning the storage

This chapter includes the following topics:

About provisioning the storage

How to provision the storage is beyond the scope of the NetBackup documentation. For help, consult the storage vendor's documentation.

What you choose as your storage destination affects how you provision the storage. NetBackup requirements also may affect how you provision the storage.

See "About the deduplication storage type" on page 19.

About deduplication storage requirements

The storage for the NetBackup Media Server Deduplication Option is disk storage. The storage must be attached to the NetBackup media server that functions as the storage server. Attached means a file system mount on the storage. You must provision the storage such that a file system is mounted on the media server.

The storage can be SAN storage or network attached storage. The minimum requirement is for the disk storage that is capable of read and write performance of 130 MB/sec.

The storage must be configured and operational before you can configure deduplication in NetBackup.

NetBackup requires exclusive use of the disk resources. If the storage is used for purposes other than backups, NetBackup cannot manage disk pool capacity or manage storage lifecycle policies correctly. Therefore, NetBackup must be the only entity that uses the storage.

See "About the deduplication storage paths" on page 38.

About deduplication storage capacity

Storage capacity for a deduplication node (deduplication storage server and storage) is 32TB.

The deduplication database consumes approximately 10 percent of the storage capacity. Therefore, approximately 90 percent of the storage capacity is usable space for unique backup data. The actual percentages vary depending on the data.

For performance optimization, Symantec recommends that you use a separate disk, volume, partition, or spindle for the catalog database.

If your storage requirements exceed the capacity of a media server deduplication node, do one of the following:

- Use more than one media server deduplication node.
- Use a PureDisk deduplication pool as the deduplication destination. A PureDisk deduplication pool provides larger storage capacity. It also provides global deduplication.

About the deduplication storage paths

When you configure the deduplication storage server, you must enter the path name to the storage. The storage path is the directory in which NetBackup stores the raw backup data.

Because the storage requires a directory path, do not use only a root node (/) or drive letter (`G:\`) as the storage path.

You also can specify a different location for the deduplication database. The database path is the directory in which NetBackup stores and maintains the structure of the stored deduplicated data.

For performance optimization, Symantec recommends that you use a separate disk, volume, partition, or spindle for the deduplication database.

If the directory or directories do not exist, NetBackup creates them and populates them with the necessary subdirectory structure. If the directory or directories exist, NetBackup populates them with the necessary subdirectory structure.

The path names must use ASCII characters only.

The NetBackup Media Server Deduplication Option does not support NFS mounted file systems.

Caution: You cannot change the paths after NetBackup configures the deduplication storage server. Therefore, carefully decide during the planning phase where and how you want the deduplicated backup data stored.

Chapter 4

Installing deduplication

This chapter includes the following topics:

About installing deduplication

The NetBackup deduplication components are installed by default on the supported host systems. However, you must enter a license key to enable deduplication.

Before you try to install or upgrade to a NetBackup version that supports deduplication, be aware of the following:

- NetBackup supports deduplication on specific 64-bit host operating systems. If you intend to upgrade an existing media server and use it for deduplication, that host must be supported.
  For the supported systems, see the *NetBackup Release Notes*.
- NetBackup deduplication components cannot reside on the same host as a PureDisk Deduplication Option agent.
  To use a PDDO agent host for NetBackup deduplication, first remove the PDDO agent from that host.
  See the *NetBackup PureDisk Deduplication Option (PDDO) Guide*.
  Then, upgrade that host to NetBackup 7.0 or later.
  Finally, configure that host as a deduplication storage server or as a load balancing server.

About the deduplication license key

NetBackup deduplication is licensed separately from base NetBackup.

The NetBackup Deduplication Option license key enables both NetBackup Media Server Deduplication and NetBackup Client Deduplication. The license is a front-end capacity license. It is based on the size of the data to be backed up, not on the size of the deduplicated data.

You may have a single license key that activates both NetBackup and optional features. Alternatively, you may have one license key that activates NetBackup and another key that activates deduplication.

If you remove the NetBackup Deduplication Option license key or if it expires, you cannot create new deduplication disk pools. you also cannot create the storage units that reference NetBackup deduplication pools.

NetBackup does not delete the disk pools or the storage units that reference the disk pools. You can use them again if you enter a valid license key.

Licensing NetBackup deduplication

If you installed the license key when you installed or upgraded NetBackup, you do not need to perform this procedure.

Enter the license key on the NetBackup master server. The following procedure describes how to use the NetBackup Administration Console to enter the license key.

To license NetBackup deduplication

1. To add a license to a specific server, on the **File** menu select **Change Server** and then select the server.
2. In the **NetBackup License Keys** dialog box, click **New**.
3. In the **Add a New License Key** dialog box, enter the license key and click **Add** or **OK**.
4. Click **Close**.
5. Restart all the NetBackup services and daemons.

Replacing the deduplication storage server host computer

If you replace the deduplication storage server host computer, use these instructions to install NetBackup and reconfigure the deduplication storage server.

For the new host, you must use the same host name. The new host cannot host a deduplication storage server already.

Reasons to replace the host include a lease swap or perhaps the current deduplication storage server host does not meet your performance requirements.

Warning: The new host must use the same byte order as the old host. If it does not, you cannot access the deduplicated data.

In computing, endianness describes the byte order that represents data: big endian and little endian. For example, Sun SPARC processors and Intel processors use different byte orders. Therefore, you cannot replace a Solaris SPARC host with a host that has an Intel processor.

Table 4-1 How to replace the deduplication storage server host

Task	Procedure
Change the disk volume state and disk pool state to DOWN	See "Changing the deduplication disk volume state" on page 77. See "Changing the deduplication pool state" on page 77.
Configure the new host so it meets deduplication requirements	See "About deduplication servers" on page 21. See "About deduplication server requirements" on page 23.
Move the storage to the new host.	See the storage vendor's documentation.
Install the NetBackup media server software on the new host	See the *NetBackup Installation Guide for UNIX and Linux.* See the *NetBackup Installation Guide for Windows.*
Delete the NetBackup Deduplication Engine credentials	If you have load balancing servers, delete the NetBackup Deduplication Engine credentials on those media servers. On each load balancing server, run the following command: See "Deleting credentials from a load balancing server" on page 75.
Add the credentials to the storage server	Add the NetBackup Deduplication Engine credentials to the storage server. See "Adding NetBackup Deduplication Engine credentials" on page 74.

Table 4-1 How to replace the deduplication storage server host *(continued)*

Task	Procedure
Get a configuration file template	If you did not save a storage server configuration file before the failure, get a template configuration file. See "Getting the storage server configuration" on page 68.
Edit the configuration file	See "Editing a storage server configuration file" on page 68.
Configure the storage server	Configure the storage server by uploading the configuration from the file you edited. If you saved a configuration file before the storage server failure, use that file. See "Setting the storage server configuration" on page 70.
Configure the load balancing servers	If you have load balancing servers, add them to the configuration. See "Adding a load balancing server" on page 58.
Change configuration settings	If you edited the deduplication configuration file, make the same changes to that file. See "About the deduplication configuration file" on page 60. See "Editing the deduplication configuration file" on page 60.
Change the disk volume state and disk pool state to UP	See "Changing the deduplication disk volume state" on page 77. See "Changing the deduplication pool state" on page 77.
Change the disk volume state and disk pool state to UP	See "Changing the deduplication disk volume state" on page 77. See "Changing the deduplication pool state" on page 77.
Restart the backup jobs	If any backup jobs failed, restart those jobs. Alternatively, wait until the next scheduled backup, at which time the backup jobs should succeed.

Uninstalling media server deduplication

The NetBackup deduplication components are uninstalled when you uninstall NetBackup software.

However, you can disable media server deduplication and remove the configuration files and storage files from the media server. The following procedure disables

NetBackup media server deduplication components and the deduplication storage. The host remains a NetBackup media server.

This process assumes that all backup images that reside on the deduplication disk storage have expired.

Caution: If you uninstall deduplication and valid NetBackup images reside on the deduplication storage, data loss may occur.

Table 4-2 Disable media server deduplication

Disable client deduplication	Remove the clients that deduplicate their own data from the client deduplication list. See "Disabling deduplication for a client" on page 90.
Delete the storage units that use the disk pool	See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.
Delete the disk pool	See "Deleting a deduplication pool" on page 78.
Stop the services on the storage server	See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.
Delete the storage directories	Delete the storage directory and database directory (if you configured a database directory). See the operating system documentation.
On Windows, delete accounts and files	On Windows, delete the following: ■ The `puredisksdbuser` account. The account is for the deduplication database administration. ■ The `puredisksdbuser` folder. See the operating system documentation.

Table 4-2 Disable media server deduplication *(continued)*

On UNIX and Linux, remove files	On UNIX and Linux systems, remove the following files: ■ `etc/pdregistry.cfg` ■ `opt/pdag` ■ `opt/pdshared` ■ The *`hostname`*`.cfg` file The file resides in the `/usr/openv/lib/ost-plugins` directory. The *servername* is the name of the configured deduplication storage server. If you entered a fully-qualified domain name for the server, that is the name used for *servername*.
Delete the deduplication storage server	See "Deleting a deduplication storage server" on page 67.
Remove the NetBackup Deduplication license key.	See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.
Start the NetBackup services on the media server	See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.

Chapter 5

Configuring deduplication

This chapter includes the following topics:

Configuring deduplication

This guide describes how to configure deduplication in NetBackup.

Table 5-1 describes the configuration tasks.

The NetBackup administrator's guides describe how to configure a base NetBackup environment.

See the *NetBackup Administrator's Guide for Windows, Volume I.*

See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I.*

Table 5-1 Deduplication configuration tasks

Task	Procedure
Configure a deduplication storage server	See "Configuring a deduplication storage server" on page 48.
Configure a disk pool	See "About deduplication pools" on page 49. See "Configuring a deduplication pool" on page 50.
Configure a storage unit	See "Configuring a deduplication storage unit" on page 52.
Enable client-side deduplication	See "Enabling client deduplication" on page 56.
Configure a backup policy	Use the deduplication storage unit as the destination for the backup policy. See the *NetBackup Administrator's Guide for Windows, Volume I.* See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I.*
Configure optimized duplication	See "Configuring optimized deduplication copy" on page 57. See "Configuring optimized deduplication copy behavior" on page 57.
Optionally, specify advanced deduplication settings	See "About the deduplication configuration file" on page 60. See "Editing the deduplication configuration file" on page 60. See "pd.conf file settings" on page 61.

Configuring a deduplication storage server

Configure in this context means to configure a NetBackup media server as a storage server for deduplication.

When you configure a storage server for deduplication, you specify the following:

- The type of storage server.
  For NetBackup media server deduplication, select **Media Server Deduplication Pool** for the type of disk storage.
  For a PureDisk deduplication pool, select **PureDisk Deduplication Pool** for the type of disk storage.
- The credentials for the deduplication engine.
  See "About NetBackup Deduplication Engine credentials" on page 26.
- The storage paths.
  See "About the deduplication storage paths" on page 38.
- The network interface.
  See "About the network interface for deduplication" on page 27.
- The load balancing servers, if any.
  See "About deduplication servers" on page 21.

When you create the storage server, the wizard lets you create a disk pool and storage unit also.

To configure a deduplication storage server in NetBackup

1. In the NetBackup Administration Console, expand **Media and Device Management > Configure Disk Storage Servers**.
2. Follow the wizard screens to configure a deduplication storage server.
3. After NetBackup creates the deduplication storage server, you can click **Next** to continue to the Disk Pool Configuration Wizard.

About deduplication pools

Deduplication pools are the disk pools that are the storage destination for deduplicated backup data. NetBackup media servers or NetBackup clients deduplicate the backup data that is stored in a deduplication pool.

NetBackup deduplication disk pools are of type PureDisk.

NetBackup requires exclusive ownership of the disk resources that comprise the disk pool. If you share those resources with other users, NetBackup cannot manage disk pool capacity or storage lifecycle policies correctly.

Configuring a deduplication pool

When you create a deduplication storage server, you can launch the Disk Pool Configuration Wizard after NetBackup creates the deduplication storage server. If you created the disk pool already, you do not have to follow these instructions.

When you configure a disk pool for deduplication, you specify the following:

- The type of disk pool (**PureDisk**).
- The NetBackup deduplication storage server to query for the disk storage to use for the pool.
- The disk volume to include in the pool.
  NetBackup exposes the storage as a single volume.
- The disk pool properties.
  See "Media server deduplication pool properties" on page 50.

Symantec recommends that disk pool names be unique across your enterprise.

To create a NetBackup disk pool

1. In the **NetBackup Administration Console**, select the **Media and Device Management** node.
2. From the list of wizards in the **Details** pane, click **Configure Disk Pool** and follow the wizard instructions.

   For help, see the wizard help.
3. After NetBackup creates the deduplication pool, you have the option to create a storage unit that uses the pool.

Media server deduplication pool properties

Table 5-2 describes the disk pool properties.

Table 5-2 Media server deduplication pool properties

Property	Description
Name	The disk pool name.
Storage server	The storage server name. The storage server is the same as the NetBackup media server to which the storage is attached.

Table 5-2 Media server deduplication pool properties *(continued)*

Property	Description
Disk volume	For a media server deduplication pool, all disk storage is exposed as a single volume. **PureDiskVolume** is a virtual name for the storage that is contained within the directories you specified for the storage path and the database path.
Available space	The amount of space available in the disk pool.
Raw size	The total raw size of the storage in the disk pool.
Comment	A comment that is associated with the disk pool.
High water mark	The high water mark is a threshold that indicates the **PureDiskVolume** is full. When the **PureDiskVolume** is at the high water mark, NetBackup fails any backup jobs that are assigned to the disk pool storage unit. NetBackup also fails backup jobs if the **PureDiskVolume** does not contain enough storage for its estimated space requirement. NetBackup again assigns jobs to the storage unit when the capacity of the **PureDiskVolume** drops below the high water mark. Capacity is regained as backup images expire. NetBackupdoes not assign backup jobs to the disk pool if used space in the **PureDiskVolume** is greater than the high water mark. The default is 98%.
Low water mark	The low water mark has no affect on the **PureDiskVolume**. The low water mark setting cannot be greater than or equal to the high water mark setting. The default is 80%

Table 5-2 Media server deduplication pool properties *(continued)*

Property	Description
Limit I/O streams	Select to limit the number of read and write streams (that is, jobs) for each volume in the disk pool. A job may read backup images or write backup images. By default, there is no limit. If you select this property, also configure the number of streams to allow per volume. When the limit is reached, NetBackup chooses another volume for write operations, if available. If not available, NetBackup queues jobs until a volume is available. Too many streams may degrade performance because of disk thrashing. Disk thrashing is excessive swapping of data between RAM and a hard disk drive. Fewer streams can improve throughput, which may increase the number of jobs that complete in a specific time period.
per volume	Select or enter the number of read and write streams to allow per volume. Many factors affect the optimal number of streams. Factors include but are not limited to disk speed, CPU speed, and the amount of memory.

Configuring a deduplication storage unit

Create one or more storage units that reference the disk pool.

The **Disk Pool Configuration Wizard** lets you create a storage unit; therefore, you may have created a storage unit when you created a disk pool. To determine if storage units exist for the disk pool, see the **NetBackup Management > Storage > Storage Units** window of the Administration Console.

To configure a storage unit from the Actions menu

1. In the **NetBackup Administration Console**, expand **NetBackup Management > Storage > Storage Units**.
2. On the **Actions** menu, select **New > Storage Unit**.
3. Complete the fields in the **New Storage Unit** dialog box.

   See "Deduplication storage unit properties" on page 53.

   See "Deduplication storage unit recommendations" on page 54.

Deduplication storage unit properties

The following are the configuration options for a PureDisk disk pool storage unit.

Table 5-3 Deduplication storage unit properties

Property	Description
Storage unit name	A unique name for the new storage unit. The name can describe the type of storage. The storage unit name is the name used to specify a storage unit for policies and schedules. The storage unit name cannot be changed after creation.
Storage unit type	Select **Disk** as the storage unit type.
Disk type	Select **PureDisk** for the disk type for a media server deduplication pool, a PureDisk deduplication pool, or a PureDisk Deduplication Option storage pool.
Disk pool	Select the disk pool that contains the storage for this storage unit. All disk pools of the specified **Disk type** appear in the **Disk pool** list. If no disk pools are configured, no disk pools appear in the list.
Media server	The **Media server** setting specifies the NetBackup media servers that can move data to and from the disk pool for this storage unit. Only the load balancing servers appear in the media server list. Specify the media server or servers as follows: ■ To allow any server in the media server list to access the disk storage (default), select **Use any available media server**. ■ To restrict the media servers that can access the disk storage, select **Only use the following media servers**. Then, select the media servers to allow. NetBackup selects the media server to use when the policy runs.
Maximum fragment size	For normal backups, NetBackup breaks each backup image into fragments so it does not exceed the maximum file size that the file system allows. You can enter a value from 20 MBs to 51200 MBs.

Table 5-3 Deduplication storage unit properties *(continued)*

Property	Description
Maximum concurrent jobs	The **Maximum concurrent jobs** setting specifies the maximum number of jobs that NetBackup can send to a disk storage unit at one time. (Default: one job. The job count can range from 0 to 256.) This setting corresponds to the Maximum concurrent write drives setting for a Media Manager storage unit. NetBackup queues jobs until the storage unit is available. If three backup jobs are scheduled and **Maximum concurrent jobs** is set to two, NetBackup starts the first two jobs and queues the third job. If a job contains multiple copies, each copy applies toward the **Maximum concurrent jobs** count. **Maximum concurrent jobs** controls the traffic for backup and duplication jobs but not restore jobs. The count applies to all servers in the storage unit, not per server. If you select multiple media servers in the storage unit and 1 for **Maximum concurrent jobs**, only one job runs at a time. The number to enter depends on the available disk space and the server's ability to run multiple backup processes. **Warning:** A Maximum concurrent jobs setting of 0 disables the storage unit.

Deduplication storage unit recommendations

You can use storage unit properties to control how NetBackup performs.

Increase the Maximum concurrent jobs gradually

Symantec recommends that you increase the **Maximum concurrent jobs** value gradually. The initial backup jobs (also known as initial seeding) require more CPU and memory than successive jobs. After initial seeding, the storage server can process more jobs concurrently. Gradually increase the jobs value over time.

Testing shows that the upper limit for a storage server with 8GB of memory and 4GB of swap space is 50 concurrent jobs.

Configure a client-to-server ratio

For a favorable client-to-server ratio, you can use one disk pool and configure multiple storage units to separate your backup traffic. Because all storage units use the same disk pool, you do not have to partition the storage.

For example, assume that you have 100 important clients, 500 regular clients, and four media servers. You can use two media servers to back up your most important clients and two media servers to back up you regular clients.

The following example describes how to configure a favorable client-to-server ratio:

- Configure the media servers for NetBackup deduplication and configure the storage.
- Configure a disk pool.
- Configure a storage unit for your most important clients (such as STU-GOLD). Select the disk pool. Select **Only use the following media servers**. Select two media servers to use for your important backups.
- Create a backup policy for the 100 important clients and select the STU-GOLD storage unit. The media servers that are specified in the storage unit move the client data to the deduplication storage server.
- Configure another storage unit (such as STU-SILVER). Select the same disk pool. Select **Only use the following media servers**. Select the other two media servers.
- Configure a backup policy for the 500 regular clients and select the STU-SILVER storage unit. The media servers that are specified in the storage unit move the client data to the deduplication storage server.

Backup traffic is routed to the wanted data movers by the storage unit settings.

Note: NetBackup uses storage units for media server selection for write activity (backups and duplications) only. For restores, NetBackup chooses among all media servers that can access the disk pool.

Throttle traffic to the media servers

You can use the **Maximum concurrent jobs** settings on disk pool storage units to throttle the traffic to the media servers. Effectively, this setting also directs higher loads to specific media servers when you use multiple storage units for the same disk pool. A higher number of concurrent jobs means that the disk can be busier than if the number is lower.

For example, two storage units use the same set of media servers. One of the storage units (STU-GOLD) has a higher **Maximum concurrent jobs** setting than the other (STU-SILVER). More client backups occur for the storage unit with the higher **Maximum concurrent jobs** setting.

Enabling client deduplication

To enable Client Deduplication, set an attribute in the NetBackup master server Client Attributes host properties.

To specify the clients that deduplicate backups

1. In the **NetBackup Administration Console**, expand **NetBackup Management > Host Properties > Master Servers**.
2. In the details pane, select the master server.
3. On the **Actions** menu, select **Properties**.
4. On the **Host Properties General** tab, add the clients that use client direct to the **Clients** list.
5. Select one of the following **Deduplication Location** options:
   - **Always use the media server** disables client deduplication. By default, all clients are configured with the **Always use the media server** option.
   - **Prefer to use client-side deduplication** uses client deduplication if the PureDisk plug-in is active on the client. If it is not active, a normal backup occurs; client deduplication does not occur.
   - **Always use client-side deduplication** uses client deduplication. If the deduplication backup job fails, NetBackup retries the job.

   You can override the **Prefer to use client-side deduplication** or **Always use client-side deduplication** host property in the backup policies.

   See **Disable client-side deduplication** in the *NetBackup Administrator's Guide for UNIX and Linux, Volume I*.

   See **Disable client-side deduplication** in the *NetBackup Administrator's Guide for Windows, Volume I*.

Configuring backups

When you configure a backup policy, for the **Policy storage** select a storage unit that uses a deduplication pool.

For a storage lifecycle policy, for the **Storage unit** select a storage unit that uses a deduplication pool.

For VMware backups, select the **Mapped full VM backup** option when you configure a VMware backup policy. The **Mapped full VM backup** option provides the best deduplication rates.

NetBackup deduplicates the client data that it sends to a deduplication storage unit.

Configuring optimized deduplication copy

You can configure optimized copy of deduplicated backups.

See "About optimized duplication of deduplicated data" on page 28.

To configure optimized duplication of deduplicated data

1 Ensure that all requirements are met.

See "Optimized deduplication copy requirements" on page 28.

2 Use one of the following methods to duplicate backup images:

- A storage lifecycle policy to copy images automatically.
  For the **Backup** destination **Storage unit**, select the storage unit for the source media server deduplication pool. For the **Duplication** destination **Storage unit**, select the storage unit for the destination disk pool. That disk pool may be a media server deduplication pool or a PureDisk deduplication pool.
  See the *NetBackup Administrator's Guide for UNIX and Linux* or the *NetBackup Administrator's Guide for Windows*.
- A Vault policy to copy images automatically.
  On the **Profile** dialog box **Choose Backups** tab, choose the backup images in the source media server deduplication pool. For the **Destination Storage unit** on the **Duplication** tab, select the storage unit for the destination disk pool. That disk pool may be a media server deduplication pool or a PureDisk deduplication pool.
  See the *NetBackup Vault Administrator's Guide*.
- The NetBackup `bpduplicate` command to copy images manually.
  Duplicate from the source media server deduplication pool to another media server deduplication pool or to a PureDisk deduplication pool.
  See *NetBackup Commands*.

You can apply separate retention periods to each copy. For example, you can retain the primary copy for three weeks and the destination copy for a longer period of time. If you delete the source image, the copy is not deleted.

Configuring optimized deduplication copy behavior

You can configure several optimized deduplication copy behaviors, as follows:

- Optimized duplication failover. By default, if an optimized duplication job fails, NetBackup does not run the job again.
- Number of optimized duplication attempts. You can specify the number of times NetBackup retries an optimized deduplication job before it fails the jobs.
- Storage lifecycle policy retries. If the optimized deduplication job is configured in a storage lifecycle policy, NetBackup retries the job three times.

Caution: These settings affect all optimized duplication jobs, not just optimized deduplication copy jobs.

To configure NetBackup to revert to normal duplication if an optimized job fails

- Add the following entry to the `bp.conf` file on the NetBackup master server:

  ```
  RESUME_ORIG_DUP_ON_OPT_DUP_FAIL = TRUE
  ```

  On Windows systems, NetBackup configuration options are in the Windows registry.

To configure the number of duplication attempts

- Add an `OPT_DUP_BUSY_RETRY_LIMIT` entry in the NetBackup `behavior` file. For example, the following entry configures NetBackup to retry the job four times before NetBackup fails the job:

  ```
  OPT_DUP_BUSY_RETRY_LIMIT 4
  ```

  The `behavior` file resides in the following directories:

  - UNIX: `/usr/openv/netbackup/db/config/behavior`
  - Windows: *`install_path`*`\NetBackup\db\config\behavior`.

To configure the number of storage lifecycle policy retries

- Delay more retries by adding an `IMAGE_EXTENDED_RETRY_PERIOD_IN_HOURS` entry in the NetBackup `behavior` file. The default for this value is three hours.

Adding a load balancing server

You can add a load balancing server to an existing media server deduplication node.

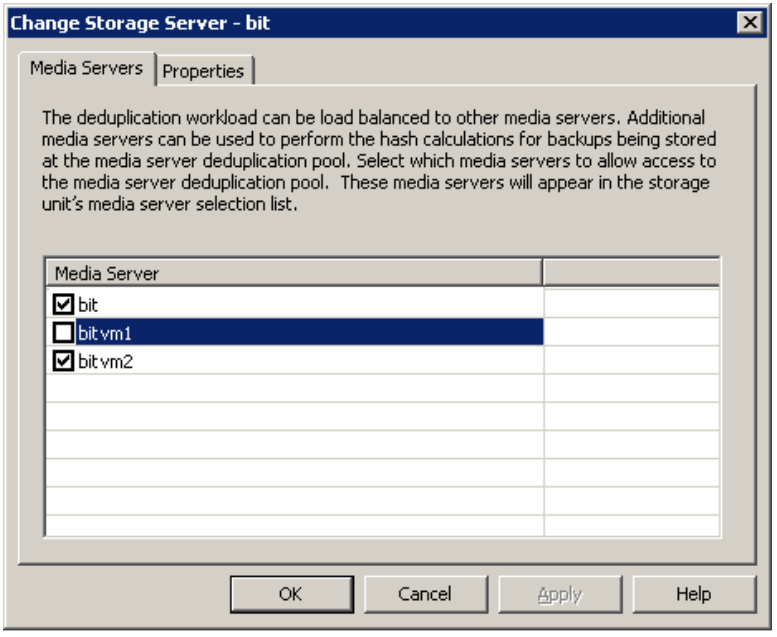
See "About deduplication servers" on page 21.

To add a load balancing server

1 In the NetBackup Administration Console, expand **Media and Device Management > Credentials > Storage Server**

2 Select the deduplication storage server.

3 On the **Edit**, select **Change**.



4 In the **Change Storage Server** dialog box, select the **Media Servers** tab (Windows) or the **Login Credentials** tab (UNIX).

5 Select the media server or servers that you want to use as a load balancing server. It must be a supported host.

The media servers that are checked are configured as load balancing servers.

6 Click **OK**.

7 For all storage units in which **Only use the following media servers** is configured, ensure that the new load balancing server is selected.

About the deduplication configuration file

On each host that deduplicates data, a `pd.conf` file contains configuration settings for the deduplication. You can edit the file to configure advanced settings for that host.

If you change the `pd.conf` file on a host, it changes the settings for that host only. If you want the same settings for all of the hosts that deduplicate data, you must change the `pd.conf` file on all of the hosts.

The `pd.conf` file resides in the following directories:

- (UNIX) `/usr/openv/lib/ost-plugins/`
- (Windows) *`install_path`*`\Veritas\NetBackup\bin\ost-plugins`

Editing the deduplication configuration file

The `pd.conf` file contains the configuration settings that control the operation of deduplication on each host.

If you change the `pd.conf` file on a host, it changes the settings for that host only. If you want the same settings for all of the hosts that deduplicate data, you must change the `pd.conf` file on all of the hosts.

See "About the deduplication configuration file" on page 60.

To edit the pd.conf file

1. Use a text editor to open the `pd.conf` file.

   The `pd.conf` file resides in the following directories:

   - (UNIX) `/usr/openv/lib/ost-plugins/`
   - (Windows) *`install_path`*`\Veritas\NetBackup\bin\ost-plugins`

   See "About the deduplication configuration file" on page 60.

2. To activate a setting, remove the pound character (`#`) in column 1 from each line that you want to edit.
3. To change a setting, specify a new value.

   Note: The spaces to the left and right of the equal sign (=) in the file are significant. Ensure that the space characters appear in the file after you edit the file.

   See "pd.conf file settings" on page 61.

4 Save and close the file.

5 Restart the NetBackup Remote Manager and Monitor Service (`nbrmms`) on the host.

pd.conf file settings

Table 5-4 describes the deduplication settings that you can configure.

Table 5-4 pd.conf file values

Setting	Default value	Possible values	Action
`DEBUGLOG`	`C:\pdplugin.log` (Windows) `tmp/pdplugin.log` (UNIX)	Any path	Writes the log information to the specified file. Uncomment the `DEBUGLOG` line that corresponds to your operating system and then specify the log file path. You can specify a different location and log file name.
`LOGLEVEL`	`0`	`0` through `10`	Specifies the amount of information that is written to the log file. The range is from 0 to 10, with 10 being the most logging. **Note:** Do not change this setting unless directed to by a Symantec Technical Support representative.
`COMPRESSION`	`0`	`0` (off) or `1` (on)	Specifies whether you want compression. By default, files are not compressed. If you want compression, change the value to `1`. See "About compression and encryption" on page 28.
`ENCRYPTION`	`0`	`0` (off) or `1` (on)	Specifies whether you want encryption. By default files are not encrypted. If you want encryption, change the value to `1`. See "About compression and encryption" on page 28.
`PREFETCH`			This keyword is reserved for internal use. **Note:** Do not change this setting unless directed to by a Symantec Technical Support representative.

Table 5-4 pd.conf file values *(continued)*

Setting	Default value	Possible values	Action
PDALIGN			This keyword is reserved for internal use. **Note:** Do not change this setting unless directed to by a Symantec Technical Support representative.
OPTDUP_TIMEOUT	N/A	The value expressed in minutes	Specifies the number of minutes before the optimized duplication times out. Indicated in minutes.
SEGKSIZE	N/A	N/A	This keyword is reserved for internal use. **Warning:** Changing this value can reduce capacity and decrease performance. Modify this setting only if directed to do so by Symantec support representative.
MINFILE_KSIZE	16	From 1 to the practical system limit	Determines the smallest size file (in KBs) that NetBackup segments. Files smaller than the threshold are combined into a single large segment to reduce the overhead of managing many small segments. However, such segments have less chance of being deduplicated. **Caution:** Large numbers of files smaller than the threshold may adversely affect backup performance.
MATCH_PDRO	1	1 (on) or 0 (off)	Specifies that NetBackup should use the PureDisk Remote Office Agent dynamic segmentation deduplication algorithm. Enabling it means that the data that is backed up is globally deduplicated with each other. If disabled (set to 0), PDDO uses a static 128KB segmentation algorithm for deduplication. **Caution:** Changing the default segmentation algorithm most likely causes the next set of backups to not deduplicate with the existing storage pool data. Subsequent backups then deduplicate using the PDRO algorithm.
OPTDUP_BANDWITH	0	0 (no limit) to the practical system limit	Determines the maximum bandwidth that is allowed for optimized duplication. The value is specified in KBytes/second.

Table 5-4 pd.conf file values *(continued)*

Setting	Default value	Possible values	Action
OPTDUP_ENCRYPTION	1	1 (on) or 0 (off)	Determines if the data to replicate is encrypted before it is sent out over the network.
DONT_SEGMENT_TYPES	N/A	Any file extension	Allows a list of file name extensions to be specified. Files in the backup stream that have these extensions are given a single segment if smaller than 16MB. Larger files are deduplicated using the maximum 16MB segment size. Example: DONT_SEGMENT_TYPES = mp3,avi This setting prevents NetBackup from analyzing and managing segments within the file types that do not deduplicate globally.
BANDWIDTH_LIMIT	0	0 (default - no limit) to the practical system limit	Determines the maximum bandwidth that is allowed when backing up or restoring data between the media server and the deduplication pool. The value is specified in KBytes/second. The default is no limit.
MAX_IMG_MBSIZE	50,000	0 to 50,000	This keyword is reserved for internal use. **Note:** Do not change this setting unless directed to by a Symantec representative.

Reconfiguring the deduplication storage server and storage paths

Only one deduplication storage path can exist on a media server. Therefore, you cannot rerun the **Storage Server Configuration Wizard**. An activated NetBackup Deduplication Engine rejects all configuration attempts that abandon its previously defined storage and backup images. You must manually deactivate the engine and physically delete the storage directory.

Two aspects to the configuration exist: the record of the deduplication storage in the EMM database and the physical presence of the storage on disk (the populated storage directory). Deleting the deduplication storage server does not alter the contents of the storage on physical disk. To protect against inadvertent data loss, NetBackup does not automatically delete the storage when you delete the storage server.

Warning: Deleting valid backup images may cause data loss.

Table 5-5 Disable media server deduplication

Task	Procedure
Ensure that no deduplication activity occurs	Deactivate all backup policies that use deduplication storage. See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.
Expire backup images	Expire all backup images that reside on the deduplication disk storage. See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.
Delete the storage units that use the disk pool	See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.
Delete the disk pool	See "Deleting a deduplication pool" on page 78.
Stop the services on the media server	See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.
Delete the storage directories	Delete the storage directory and database directory (if you configured a database directory).
Reset the registry	See "Resetting the deduplication registry" on page 73.
Delete the deduplication storage server	See "Deleting a deduplication storage server" on page 67.
Start the NetBackup services on the media server	See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* See the *NetBackup Administrator's Guide for Windows, Volume I.*.
Reconfigure	See "Configuring deduplication" on page 47.

Chapter 6

Managing deduplication

This chapter includes the following topics:

Managing deduplication servers

After you configure deduplication, you can perform various tasks to manage deduplication servers.

See "Changing deduplication storage server properties" on page 66.

See "Deleting a deduplication storage server" on page 67.

See "Determining the deduplication storage server state" on page 67.

See "Getting the storage server configuration" on page 68.

See "Editing a storage server configuration file" on page 68.

See "Setting the storage server configuration" on page 70.

See "Deleting a load balancing server configuration file" on page 70.

See "Removing a load balancing server" on page 70.

See "Viewing deduplication storage servers" on page 72.

See "Viewing deduplication storage server attributes" on page 72.

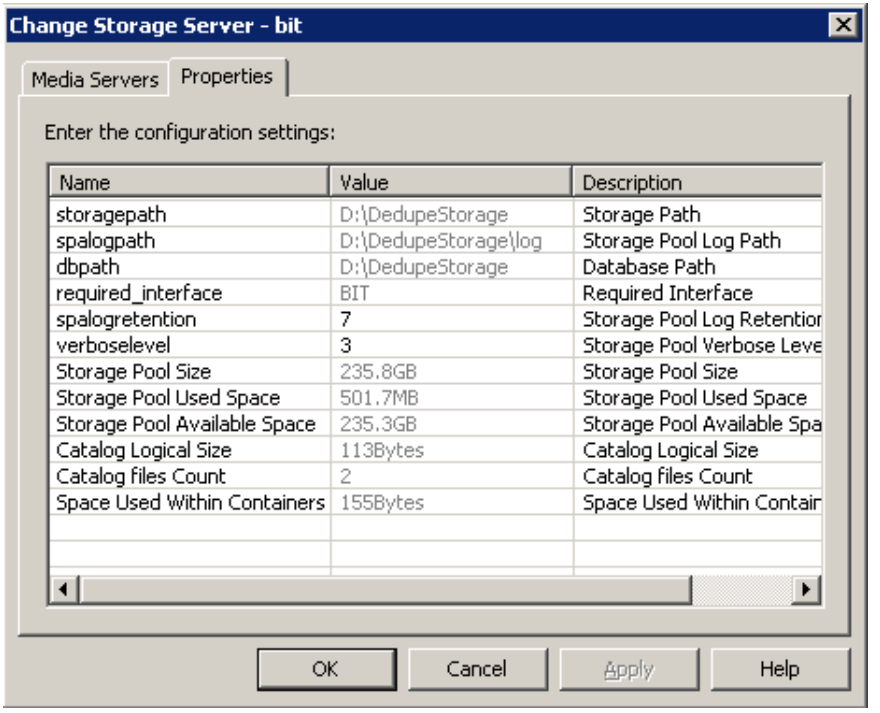
See "Resetting the deduplication registry" on page 73.

Changing deduplication storage server properties

You can change the retention period and logging level for the NetBackup Deduplication Manager.

To change deduplication storage server properties

1. In the NetBackup Administration Console, expand **Media and Device Management > Credentials > Storage Server**
2. Select the deduplication storage server.
3. On the **Edit** menu, select **Change**.
4. In the **Change Storage Server** dialog box, select the **Properties** tab.



5. For the property to change, select the value in the **Value column**.

6 Change the value.

7 Click **OK**.

Deleting a deduplication storage server

If you delete a deduplication storage server, NetBackup disables the deduplication functionality and the storage server functionality on that media server.

NetBackup does not delete the media server from your configuration. To delete the media server, use the NetBackup `nbemmcmd` command.

If a disk pool is configured from the disk volume that the deduplication storage server manages, you cannot delete the deduplication storage server.

Warning: Do not delete a deduplication storage server if its storage contains unexpired NetBackup images; if you do, data loss may occur.

To delete a deduplication storage server

1 In the **NetBackup Administration Console**, expand **Media and Device Management > Credentials > Storage Server**

2 On the **Edit** menu, select **Delete**.

3 Click **Yes** in the confirmation dialog box.

Determining the deduplication storage server state

Use the NetBackup `nbdevquery` command to determine the state of a deduplication storage server. The state is either UP or DOWN.

To determine deduplication storage server state

- Run the following command on the NetBackup master server or a deduplication storage server:

  UNIX: `/usr/openv/netbackup/bin/admincmd/nbdevquery -liststs -storage_server` *`server_name`* `-stype PureDisk –U`

  Windows: *`install_path`*`\NetBackup\bin\admincmd\nbdevquery -liststs -storage_server` *`server_name`* `-stype PureDisk –U`

  The following is example output:

```
Storage Server      : bit
Storage Server Type : PureDisk
Storage Type        : Formatted Disk, Network Attached
State               : UP
```

  This example output is shortened; more flags may appear in actual output.

Getting the storage server configuration

Symantec recommends that you get and save the storage server configuration. Getting and saving the configuration can help you with recovery of your environment.

If you get the configuration of a storage server that is unavailable because of a disaster, NetBackup returns a template configuration file.

To get the storage server configuration

- On the master server, enter the following command:

UNIX: `/usr/openv/netbackup/bin/admincmd/nbdevconfig -getconfig -storage_server` *`sshostname`* `-stype PureDisk -configlist` *`file.txt`*

Windows: *`install_path`*`\NetBackup\bin\admincmd\nbdevconfig -getconfig -storage_server` *`sshostname`* `-stype PureDisk -configlist` *`file.txt`*

For *sshostname*, use the name of the storage server. For *file.txt*, use a file name indicates its purpose.

Editing a storage server configuration file

In some very limited situations, you may need to create a storage server configuration file that includes the configuration settings for your environment.

For example, for disaster recovery you may need to edit a template file to create a configuration file for your environment.

To edit the storage server configuration

1 If you did not save a storage server configuration file, get a storage server configuration file.

See "Getting the storage server configuration" on page 68.

If you get the configuration of a storage server that is unavailable because of a disaster, NetBackup returns a template configuration file. The following is an example of a template configuration file:

```
V6.5.5 "storagepath" "none" string
V6.5.5 "spalogin" "n" string
V6.5.5 "spapasswd" " "  string
V6.5.5 "dbpath" "db_path" string
V6.5.5 "required_interface" "" string
V6.5.5 "spalogretention" "7" int
V6.5.5 "verboselevel" "3" int
```

V6.5.5 represents the version of the input and output format not the NetBackup release level. That version may differ on your system.

2 Use a text editor to enter or change values.

For a template configuration file, enter the appropriate information in the second set of quotation marks in each line, replacing the default values. The values should be the same as those you used when you configured the storage server initially.

The following are the values that are required:

- `storagepath`.
- `spalogin`.
- `spapasswd`.
- `dbpath`.

  If the database path is the same as the storage path, enter the same value for `storagepath` and `dbpath`.
- `required_interface`.

  The `required_interface` is required only if you configured one initially; if a specific interface is not required, leave it blank. The required interface defaults to the computer's hostname.

  Values for the other configuration parameters are optional and not required for a recovery situation.

Setting the storage server configuration

You can set the storage server configuration (that is, configure the storage server) by importing the configuration from a file. Setting the configuration can help you with recovery of your environment.

The file should be file of your configuration that you saved.

See "Preparing for disaster" on page 103.

Alternative, the file may be an edited configuration file.

See "Editing a storage server configuration file" on page 68.

To set the storage server configuration

- On the master server, run the following command:

  UNIX: `/usr/openv/netbackup/bin/admincmd/nbdevconfig -setconfig -storage_server sshostname -stype PureDisk -configlist file.txt`

  Windows: `install_path\NetBackup\bin\admincmd\nbdevconfig -setconfig -storage_server sshostname -stype PureDisk -configlist file.txt`

  For *sshostname*, use the name of the storage server. For *file.txt*, use the name of the file that contains the configuration.

Deleting a load balancing server configuration file

Each load balancing server for deduplication has a `sshostname.cfg` file. The *sshostname* is the name of the configured deduplication storage server. The *sshostname* is the fully qualified domain name if that was used to configure the storage server.

You may need to delete the storage server configuration file from a load balancing server. For example, disaster recovery may require that you delete the configuration file.

To delete the configuration file

- Delete the file; it's location depends on the operating system type, as follows:

  UNIX: `/usr/openv/lib/ost-plugins`

  Windows: `install_path\Veritas\NetBackup\bin\ost-plugins`

Removing a load balancing server

You can remove a load balancing server from a deduplication node. The media server no longer deduplicates client data.

See "About deduplication servers" on page 21.

After you remove the load balancing server, restart the NetBackup Enterprise Media Manager service. The NetBackup disk polling service may try to use the removed server to query for disk status. Because the server is no longer a load balancing server, it cannot query the disk storage. Consequently, NetBackup may mark the disk volume as DOWN. When the EMM service restarts, it chooses a different deduplication server to monitor the disk storage.

If the host failed and is unavailable, you can use the `tpconfig` device configuration utility in menu mode to delete the server. However, you must run the `tpconfig` utility on a UNIX or Linux NetBackup server.

For procedures, see the *NetBackup Administrator's Guide for UNIX and Linux, Volume II*.

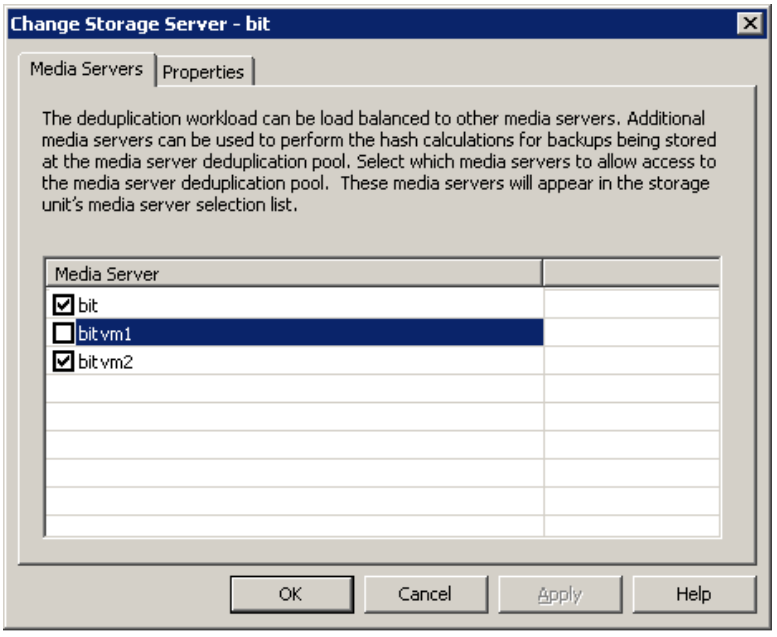
To remove a media server from a deduplication node

1. For every storage unit that specifies the media server in **Use one of the following media servers**, clear the check box that specifies the media server.

   This step is not required if the storage unit is configured to use any available media server.

2. In the NetBackup Administration Console, expand **Media and Device Management > Credentials > Storage Server**.

3. Select the deduplication storage server, then select **Edit > Change**.



4. In the **Change Storage Server** dialog box, select the **Media Servers** tab (Windows) or the **Login Credentials** tab (UNIX).
5. Clear the check box of the media server you want to remove.
6. Click **OK**.

Viewing deduplication storage servers

Use the NetBackup Administration Console to view a list of deduplication storage servers already configured.

To view deduplication storage servers

- In the NetBackup Administration Console, expand **Media and Device Management > Credentials > Storage Server**.

  The **All Storage Servers** pane shows all configured deduplication storage servers. deduplication storage servers show **PureDisk** in the **Disk Type** column.

Viewing deduplication storage server attributes

Use the NetBackup `nbdevquery` command to view the deduplication storage server attributes.

The *server_name* you use in the `nbdevquery` command must match the configured name of the storage server. If the storage server name is its fully-qualified domain name, you must use that for *server_name*.

To view deduplication storage server attributes

- Run the following command on the NetBackup master server or a deduplication storage server:

  UNIX: `/usr/openv/netbackup/bin/admincmd/nbdevquery -liststs -storage_server server_name -stype PureDisk –U`

  Windows: `install_path\NetBackup\bin\admincmd\nbdevquery -liststs -storage_server server_name -stype PureDisk –U`

  The following is example output:

```
Storage Server      : bit
Storage Server Type : PureDisk
Storage Type        : Formatted Disk, Network Attached
State               : UP
Flag                : OpenStorage
Flag                : CopyExtents
Flag                : AdminUp
Flag                : InternalUp
Flag                : LifeCycle
Flag                : CapacityMgmt
Flag                : FragmentImages
Flag                : Cpr
Flag                : FT-Transfer
```

  This example output is shortened; more flags may appear in actual output.

Resetting the deduplication registry

If you reconfigure your deduplication environment, one of the steps is to reset the deduplication registry.

See "Reconfiguring the deduplication storage server and storage paths" on page 63.

Warning: Only follow these procedures if you are reconfiguring your storage server and storage paths.

The procedure differs on UNIX and on Windows.

To reset the deduplication registry file on UNIX and Linux

- Enter the following command on the storage server to reset the deduplication registry file:

  ```
  cp -f /usr/openv/pdde/pdconfigure/cfg/userconfigs/pdregistry.cfg
  /etc/pdregistry.cfg
  ```

To reset the deduplication registry on Windows

1. Delete the contents of the following keys in the Windows registry:
   - `HKLM\SOFTWARE\Symantec\PureDisk\Agent\ConfigFilePath`
   - `HKLM\SOFTWARE\Symantec\PureDisk\Agent\EtcPath`

   Warning: Editing the Windows registry may cause unforeseen results.

2. Delete the storage path in the following key in the Windows key. That is, delete everything after `postgresql-8.3 -D` in the key.

   ```
   HKLM\SYSTEM\ControlSet001\Services\postgresql-8.3\ImagePath
   ```

   For example, in the following example registry key, you would delete the content of the key that is in italic type:

   ```
   "C:\Program Files\Veritas\pdde\pddb\bin\pg_ctl.exe" runservice
   -N postgresql-8.3 -D "D:\DedupeStorage\databases\pddb\data" -w
   ```

Managing NetBackup Deduplication Engine credentials

You can manage existing credentials in NetBackup.

See "Adding NetBackup Deduplication Engine credentials" on page 74.

See "Changing NetBackup Deduplication Engine credentials" on page 75.

See "Deleting credentials from a load balancing server" on page 75.

See "Determining which media servers have deduplication credentials" on page 76.

Adding NetBackup Deduplication Engine credentials

You may need to add the NetBackup Deduplication Engine credentials to an existing storage server or load balancing server. For example, disaster recovery may require that you add the credentials.

Add the same credentials that you already use in your environment.

Another procedure exists to add a load balancing server to your configuration.

See "Adding a load balancing server" on page 58.

To add NetBackup Deduplication Engine credentials by using the tpconfig command

- On the host to which you want to add credentials, run the following command:

  UNIX: `/usr/openv/volmgr/bin/tpconfig -add -storage_server sshostname -stype PureDisk -sts_user_id UserID -password PassWord`

  Windows: `install_path\Veritas\NetBackup\Volmgr\bin\tpconfig -add -storage_server sshostname -stype PureDisk -sts_user_id UserID -password PassWord`

  For *sshostname*, use the name of the storage server.

Changing NetBackup Deduplication Engine credentials

You cannot change the NetBackup Deduplication Engine credentials after you enter them. If you must change the credentials, contact your Symantec support representative.

See "About NetBackup Deduplication Engine credentials" on page 26.

Deleting credentials from a load balancing server

You may need to delete the NetBackup Deduplication Engine credentials from a load balancing server. For example, disaster recovery may require that you delete the credentials on a load balancing server.

Another procedure exists to remove a load balancing server from a deduplication node.

See "Removing a load balancing server" on page 70.

To delete credentials from a load balancing server

- On the load balancing server, run the following command:

  UNIX: `/usr/openv/volmgr/bin/tpconfig -delete -storage_server sshostname -stype PureDisk -sts_user_id UserID`

  Windows: `install_path\Veritas\NetBackup\Volmgr\bin\tpconfig -delete -storage_server sshostname -stype PureDisk -sts_user_id UserID`

  For *sshostname*, use the name of the storage server.

Determining which media servers have deduplication credentials

You can determine which media servers have credentials configured for the NetBackup Deduplication Engine. The servers with credentials are load balancing servers.

To determine if NetBackup Deduplication Engine credentials exist

1. In the **NetBackup Administration Console**, expand **Media and Device Management > Credentials > Storage Server**.
2. Select the storage server, then select **Edit > Change**.
3. In the **Change Storage Server** dialog box, select the **Media Servers** tab (Windows) or the **Login Credentials** tab (UNIX).

   The media servers for which credentials are configured are checked.

Managing deduplication disk pools

After you configure NetBackup deduplication, you can perform various tasks to manage your disk pools.

See "Changing deduplication disk pool properties" on page 76.

See "Changing the deduplication pool state" on page 77.

See "Changing the deduplication disk volume state" on page 77.

See "Deleting a deduplication pool" on page 78.

See "Determining the deduplication pool state" on page 78.

See "Determining the deduplication disk volume state" on page 79.

See "Viewing deduplication disk pools" on page 79.

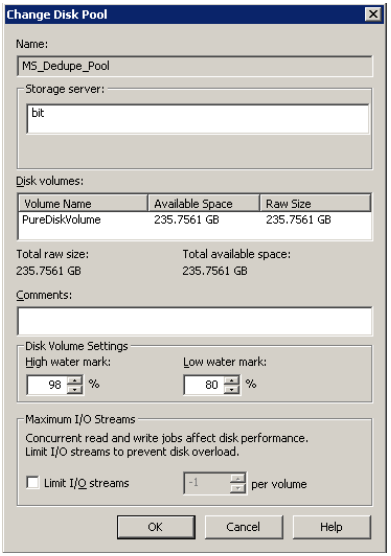
Changing deduplication disk pool properties

You can change the properties of a deduplication disk pool.

To change disk pool properties

1. In the NetBackup Administration Console, expand **Media and Device Management > Devices > Disk Pools**.
2. Select the disk pool you want to change in the details pane.

3 On the **Edit** menu, select **Change**.



4 In the **Change Disk Pool** dialog box, change properties.

See "Media server deduplication pool properties" on page 50.

Changing the deduplication pool state

Disk pool state is UP or DOWN.

To change the state to DOWN, the disk pool must not be busy. If backup jobs are assigned to the disk pool, the state change fails. Cancel the backup jobs or wait until the jobs complete.

To change deduplication pool state

1 In the NetBackup Administration Console, expand **Media and Device Management > Device Monitor**.

2 Select the **Disk Pools** tab.

3 Select the disk pool.

4 Select either **Actions > Up** or **Actions > Down**.

Changing the deduplication disk volume state

The disk volume state is UP or DOWN.

To change the state to DOWN, the disk pool in which the volume resides must not be busy. If backup jobs are assigned to the disk pool, the state change fails. Cancel the backup jobs or wait until the jobs complete.

To change the deduplication disk volume state

1. Determine the name of the disk volume. The following command lists all volumes in the specified disk pool:

   ```
   nbdevquery -listdv -stype PureDisk -dp disk_pool_name
   ```

   The `nbdevquery` and the `nbdevconfig` commands reside in the following directory:

   - UNIX: /usr/openv/NetBackup/bin/admincmd
   - Windows: *install_path*\NetBackup\bin\admincmd

   To display the disk volumes in all disk pools, omit the `-dp` option.

2. Change the disk volume state; the following is the command syntax:

   ```
   nbdevconfig -changestate -stype PureDisk -dp disk_pool_name -dv
   vol_name -state state
   ```

   The *state* is either UP or DOWN.

Deleting a deduplication pool

You can delete a disk pool if it does not contain valid NetBackup backup images or image fragments. If it does, you must first expire and delete those images or fragments.

If you delete a disk pool, NetBackup removes it from your configuration.

If a disk pool is the storage destination of a storage unit, you must first delete the storage unit.

To delete a deduplication disk pool

1. In the **NetBackup Administration Console**, expand **Media and Device Management > Devices > Disk Pools**.
2. Select a disk pool
3. On the **Edit** menu, select **Delete**.
4. In the **Delete Disk Pool** dialog box, verify that the disk pool is the one you want to delete and then click **OK**.

Determining the deduplication pool state

Disk pool state is UP or DOWN.

To determine disk pool state

1. In the **NetBackup Administration Console**, expand **Media and Device Management > Device Monitor**.
2. Select the **Disk Pools** tab.
3. The state is displayed in the **Status** column.

Determining the deduplication disk volume state

Use the NetBackup `nbdevquery` command to determine the state of the volume in a deduplication disk pool. The command shows the properties and attributes of the `PureDiskVolume`.

To determine deduplication disk volume state

- Display the volume state by using the following command:

  UNIX: `/usr/openv/netbackup/bin/admincmd/nbdevquery -listdv -stype PureDisk -U`

  Windows: *`install_path`*`\NetBackup\bin\admincmd\nbdevquery -listdv -stype PureDisk -U`

  The *state* is either UP or DOWN.

  The following is example output

```
Disk Pool Name      : PD_Disk_Pool
Disk Type           : PureDisk
Disk Volume Name    : PureDiskVolume
Disk Media ID       : @aaaab
Total Capacity (GB) : 49.98
Free Space (GB)     : 43.66
Use%                : 12
Status              : UP
Flag                : ReadOnWrite
Flag                : AdminUp
Flag                : InternalUp
Num Read Mounts     : 0
Num Write Mounts    : 1
Cur Read Streams    : 0
Cur Write Streams   : 0
```

Viewing deduplication disk pools

Use the NetBackup Administration Console to view configured disk pools.

To view disk pools

- In the NetBackup Administration Console, expand **Media and Device Management > Devices > Disk Pools**.

Monitoring deduplication activity

You can monitor deduplication activity.

See "Monitoring the deduplication rates" on page 80.

See "Viewing disk reports" on page 81.

See "Monitoring deduplication processes" on page 83.

See "Monitoring deduplication logs" on page 83.

Monitoring the deduplication rates

The deduplication rate is the percentage of data that was stored already. That data is not stored again.

NetBackup reports the rate of deduplication as follows:

- The **Deduplication Rate** column of the **Activity Monitor Jobs** tab.
- The **Job Details** dialog box.
  The **Detailed Status** tab shows detailed information, including the deduplication rate.

  The information depends on whether it is media server deduplication or client deduplication, as follows:

  - For media server deduplication, the **Detailed Status** tab shows the deduplication rate on the server that performed the deduplication. The following job details excerpt shows details for `client1`, for which `DedupeServer` deduplicated the data:

    ```
    10/6/2009 10:02:09 AM - Info DedupeServer(pid=30695)
      StorageServer=PureDisk:DedupeServer; Report=PDDO Stats
      for (DedupeServer):   scanned: 30126998 KB, stream rate:
      162.54 MB/sec, CR sent: 1720293 KB, dedup: 94.3%, cache
      hits: 214717 (94.0%)
    ```

  - For client deduplication jobs, the **Detailed Status** tab shows two deduplication rates. The first deduplication rate is always for the client data. The second duplication rate is for the load balancing server. The client sends the disk image header and **True Image Restore** information (if

applicable) to a load balancing server for deduplication. Typically, deduplication rates for that information are zero or very low. The following job details excerpt is an example of the two rates:

```
10/8/2009 11:54:21 PM - Info DedupeServer(pid=2220)
  Using OpenStorage client direct to backup from client
  DedupeClient to DedupeServer
10/8/2009 11:58:09 PM - Info DedupeServer(pid=2220)
  StorageServer=PureDisk:bitume; Report=PDDO Stats for
  (DedupeServer): scanned: 3423425 KB, stream rate: 200.77
  MB/sec, CR sent: 122280 KB, dedup: 96.4%, cache hits:
  49672 (98.2%)
10/8/2009 11:58:09 PM - Info DedupeServer(pid=2220) Using
  the media server to write NBU data for backup
  DedupeClient_1254987197 to DedupeServer
10/8/2009 11:58:19 PM - Info DedupeServer(pid=2220)
  StorageServer=PureDisk:DedupeServer; Report=PDDO Stats
  for (DedupeServer): scanned: 17161 KB, stream rate: 1047.42
  MB/sec, CR sent: 17170 KB, dedup: 0.0%, cache hits: 0 (0.0%)
the requested operation was successfully completed(0)
```

- The `bpdbjobs` command shows the deduplication rate if you configure a `COLDREFS` entry for `DEDUPRATIO` in the `bp.conf` file on the media server on which you run the command.
  See the *NetBackup Administrator's Guide for UNIX and Linux, Volume II*.

Many factors affect deduplication performance.

See "About deduplication performance" on page 32.

See "About deduplication server requirements" on page 23.

See "About client deduplication host requirements" on page 25.

Viewing disk reports

The NetBackup disk reports include information about the disk pools, disk storage units, disk logs, images that are stored on disk media, and storage capacity.

Table 6-1 describes the disk reports available.

Table 6-1 Disk reports

Report	Description
Images on Disk	The Images on Disk report generates the image list present on the disk storage units that are connected to the media server. The report is a subset of the Images on Media report; it shows only disk-specific columns. The report provides a summary of the storage unit contents. If a disk becomes bad or if a media server crashes, this report can let you know what data is lost.
Disk Logs	The Disk Logs report displays the media errors or the informational messages that are recorded in the NetBackup error catalog. The report is a subset of the Media Logs report; it shows only disk-specific columns. Either PureDisk or Symantec Deduplication Engine in the description identifies a deduplication message. (The identifiers are generic because the deduplication engine does not know which application consumes its resources. NetBackup, Symantec Backup Exec, and NetBackup PureDisk are Symantec applications that use deduplication.
Disk Storage Unit	The Disk Storage Unit Status report displays the state of disk storage units in the current NetBackup configuration. For disk pool capacity, see the disk pools window in **Media and Device Management > Devices > Disk Pools**. Multiple storage units can point to the same disk pool. When the report query is by storage unit, the report counts the capacity of disk pool storage multiple times.
Disk Pool Status	The **Disk Pool Status** report displays the state of disk pool and usage information.

To view disk reports

1. In the NetBackup Administration Console, expand **NetBackup Management > Reports > Disk Reports**.
2. Select the name of a disk report.
3. In the right pane, select the report settings.
4. Click **Run Report**.

Monitoring deduplication processes

The following are the deduplication processes about which NetBackup reports:

- NetBackup Deduplication Engine in the Administration **Activity Monitor Services** tab on Windows systems. On UNIX, the NetBackup Deduplication Engine appears as `spoold` on the **Daemons** tab.
- NetBackup Deduplication Manager in the **Activity Monitor Services** tab on Windows systems. On UNIX, the NetBackup Deduplication Manager appears as `spad` on the **Daemons** tab.
- The database processes (`postgres`) appear in the **Activity Monitor Processes** tab on Windows systems.
- The NetBackup `bpps` command shows the `spoold`, `spad`, and `postgres` processes.
  See "Deduplication server components" on page 109.

Monitoring deduplication logs

The NetBackup deduplication components write information to various log files. The following subsections describe the log files for each component.

Deduplication configuration script log

The log file is in the *storage_path*`/log/spoold` directory. The log file name is `pdde-config.log`.

NetBackup creates this log file during the configuration process. If your configuration through the wizard succeeded, you do not need to examine the log file. The only reason to look at the log file is if the configuration failed. If the configuration process failed after it created and populated the storage directory, this log file identifies when the configuration failed.

NetBackup Deduplication Engine log

The NetBackup Deduplication Engine writes several log files, as follows:

- VxUL log files for the events and errors that NetBackup receives from polling.
  See "About VxUL logs" on page 86.
- Log files in the *storage_path*`/log/spoold` directory, as follows:
  - The `spoold.log` file is the main log file
  - The `storaged.log` file is for queue processing.

- A log file for each connection to the engine is stored in a directory structure. The following describes the pathname to a log file for a connection:
  IP address/application/TaskName/FirstDigitofSessionID/sessionID-current_time_in_seconds.log
  For example, the following is an example of a `crcontrol` connection log pathname on a UNIX system:
  `/storage_path/log/spoold/127.0.0.1/crcontrol/Control/2/2916742631-1257956402.log`

Usually, the only reason to examine these connection log files is if a Symantec support representative asks you to.

NetBackup Deduplication Manager logs

The log files are in the `/storage_path/log/spad` directory.

- `spad.log`
- `sched_GarbageCollection.log`
- `sched_QueueProcess.log`
- `SchedClass.log`

You can set the log level and retention period in the **Change Storage Server** dialog box **Properties** tab.

See "Changing deduplication storage server properties" on page 66.

Deduplication database log

The deduplication database log file (`postgresql.log`) is in the `storage_path/databases/pddb` directory.

You can configure log parameters. For more information, see the following:

http://www.postgresql.org/docs/current/static/runtime-config-logging.html

The default configuration for the PostgreSQL database does not add timestamps to log entries on Windows systems. Therefore, Symantec recommends that you edit the configuration file on Windows hosts so timestamps are added to the log file.

To configure log file timestamps on Windows

1. Use a text edit to open the following file:

   `dbpath\databases\pddb\data\postgresql.conf`

   The database path may be the same as the configured storage path.

2. In the line that begins with `log_line_prefix`, change the value from `%%t` to `%t`. (That is, remove one of the percent signs (`%`).)

3 Save the file.

4 Run the following command:

install_path\Veritas\pdde\pddb\bin\pg_ctl reload -D *dbpath*\databases\pddb\data

If the command output does not include `server signaled`, use Windows Computer Management to restart the PostgreSQL Server 8.3 service.

PureDisk plug-in log

You can configure the location and name of the log file and the logging level. To do so, edit the `DEBUGLOG` entry and the `LOGLEVEL` in the `pd.conf` file.

See "About the deduplication configuration file" on page 60.

See "Editing the deduplication configuration file" on page 60.

Client deduplication proxy plug-in log

The client deduplication proxy plug-in on the media server runs under `bptm`, `bpstsinfo`, and `bpbrm` processes. Examine the log files for those processes for proxy plug-in activity. The strings `proxy` or `ProxyServer` embedded in the log messages identify proxy server activity.

They write log files to the following directories:

- For `bptm`:
  UNIX: `/usr/openv/netbackup/logs/bptm`
  Windows: *install_path*\Veritas\NetBackup\logs\bptm
- For `bpstsinfo`:
  Windows: `/usr/openv/netbackup/logs/admin`
  UNIX: `/usr/openv/netbackup/logs/bpstsinfo`
  Windows: *install_path*\Veritas\NetBackup\logs\admin
  Windows: *install_path*\Veritas\NetBackup\logs\stsinfo
- For `bpbrm`:
  UNIX: `/usr/openv/netbackup/logs/bpbrm`
  Windows: *install_path*\Veritas\NetBackup\logs\bpbrm

Client deduplication proxy server log

The deduplication proxy server `nbostpxy` on the client writes messages to files in an eponymous directory, as follows:

UNIX: `/usr/openv/netbackup/logs/nbostpxy`

Windows: *install_path*\Veritas\NetBackup\logs\nbostpxy.

About VxUL logs

Some NetBackup commands or processes write messages to their own log files. Other processes use Veritas unified log (VxUL) files. VxUL uses a standardized name and file format for log files. An originator ID (OID) identifies the process that writes the log messages.

Table 6-2 shows the NetBackup logs for disk-related activity.

The messages that begin with a sts_ prefix relate to the interaction with the storage vendor software plug-in. Most interaction occurs on the NetBackup media servers.

Table 6-2 NetBackup VxUL logs

Activity	VxUL OID	Processes that use the ID
Backups and restores	N/A	Messages appear in the log files for the following processes: ■ The bpbrm backup and restore manager ■ The bpdbm database manager ■ The bpdm disk manager ■ The bptm tape manager for I/O operations
Backups and restores	117	The nbjm Job Manager.
Deduplication Engine	364	The NetBackup Deduplication Engine that runs on the deduplication storage server.
Device configuration	111	The nbemm process.
Device configuration	178	The Disk Service Manager process that runs in the Enterprise Media Manager (EMM) process.
Device configuration	202	The storage server interface process that runs in the Remote Manager and Monitor Service. RMMS runs on media servers.
Device configuration	230	The Remote Disk Service Manager interface (RDSM) that runs in the Remote Manager and Monitor Service. RMMS runs on media servers.

To view and manage VxUL log files, you must use NetBackup log commands. For information about how to use and manage logs on NetBackup servers, see the *NetBackup Troubleshooting Guide*.

Monitoring deduplication storage capacity and usage

Several options exist to monitor your deduplication storage capacity and usage.

See "About deduplication capacity and usage reporting" on page 87.

See "About deduplication container files" on page 89.

See "Viewing capacity within deduplication container files" on page 89.

About deduplication capacity and usage reporting

Several factors may affect the expected usage results, as follows:

- Expired backups may not change the available size and the used size. An expired backup may have no unique data segments. Therefore, the segments remain valid for other backups.
- NetBackup Deduplication Manager clean-up may not have run yet. The Deduplication Manager performs clean up twice a day. Until it performs clean-up, deleted image fragments remain on disk.

If you use operating system tools to examine storage space usage, their results may not match the usage reported by NetBackup, as follows:

- The operating system tools cannot report usage accurately. The storage implementation uses container files. Deleted segments can leave free space in container files, but the container file sizes do not change.
- If other applications use the storage, NetBackup cannot report usage accurately. NetBackup requires exclusive use of the storage.

Table 6-3 describes the options for monitoring capacity and usage.

Table 6-3 Capacity and usage reporting

Option	Description
Disk Pools window	The **Disk Pools** window of the NetBackup Administration Console displays a value that was stored when NetBackup polled the disk pools. The value may not be as current as the value that is displayed in the **Storage Server** window. To display the window, expand **Media and Device Management > Devices > Disk Pools**.

Table 6-3 Capacity and usage reporting *(continued)*

Option	Description
Storage Server window	The **Storage Server** window of the NetBackup Administration Console displays real-time values. To display the window, expand **Media and Device Management > Credentials > Storage Servers**.
Change Storage Server dialog box	The **Change Storage Server** dialog box **Properties** tab displays storage capacity and usage. See "Changing deduplication storage server properties" on page 66.
Disk Pool Status report	The **Disk Pool Status** report displays the state of the disk pool and usage information. See "Viewing disk reports" on page 81.
View container command	A command that is installed with NetBackup provides a view of storage capacity and usage within the deduplication container files. See "About deduplication container files" on page 89. See "Viewing capacity within deduplication container files" on page 89.
The `nbdevquery` command	The `nbdevquery` command shows the state of the disk volume and its properties and attributes. It also shows capacity, usage, and percent used. See "Determining the deduplication disk volume state" on page 79.
License Keys dialog box	The summary of active capacity-based license features in the **NetBackup License Keys** dialog box. The summary displays the storage capacity for which you are licensed and the capacity used. It does not display the amount of physical storage space. One the **Help** menu in the **NetBackup Administration Console**, select **License Keys**.
NetBackup OpsCenter	The NetBackup OpsCenter also provides information about storage capacity and usage. See the *NetBackup OpsCenter Administrator's Guide*.

About deduplication container files

The deduplication storage implementation allocates container files to hold backup data. Deleted segments can leave free space in containers files, but the container file sizes do not change. Segments are deleted from containers when backup images expire and the NetBackup Deduplication Manager performs clean-up.

Viewing capacity within deduplication container files

A command reports on storage usage within containers, as follows:

- On UNIX and Linux systems, the path name of the command is `/usr/openv/pdde/pdcr/bin/crcontrol`.
- On Windows systems, the path name of the command is *`install_path`*`\Veritas\pdde\Crcontrol.exe`.

The following is an example of the command usage on a Windows deduplication storage server. The command shows the data store statistics (`--dsstat` option).

```
C:\Program Files\Veritas\pdde>Crcontrol.exe --dsstat 1

************ Data Store statistics ************
Data storage       Size   Used  Avail Use%
                  68.4G  46.4G  22.0G  68%
Number of containers             : 67
Average container size           : 187441541 bytes (178.76MB)
Space allocated for containers   : 12558583274 bytes (11.70GB)
Space used within containers     : 12551984871 bytes (11.69GB)
Space available within containers: 6598403 bytes (6.29MB)
Space needs compaction           : 508432 bytes (0.48MB)
Records marked for compaction    : 3
Active records                   : 95755
Total records                    : 95758
```

The NetBackup Deduplication Manager periodically compacts the space available inside the container files. Therefore, space within a container is not available as soon as it is free.. Various internal parameters control whether a container file is compacted. Although space may be available within a container file, the file may not be eligible for compaction. The NetBackup Deduplication Manager checks for space every 20 seconds.

For help with the command options, enter `crcontrol --help`.

Deleting backup images

Image deletion may be time consuming. Therefore, if you delete images manually, Symantec recommends the following approach.

See "Data removal process" on page 118.

To delete backup images manually

1. Expire all of the images by using the `bpexpdate` command and the `-notimmediate` option. The `-notimmediate` option prevents `bpexpdate` from calling the `nbdelete` command, which deletes the image.

   Without this option, `bpexpdate` calls `nbdelete` to delete images. Each call to `nbdelete` creates a job in the Activity Monitor, allocates resources, and launches processes on the media server.

2. After you expire the last image, delete all of the images by using the `nbdelete` command with the `-allvolumes` option.

   Only one job is created in the Activity Monitor, fewer resources are allocated, and fewer processes are started on the media servers. The entire process of expiring images and deleting images takes less time.

Disabling deduplication for a client

If you remove a client from the list of clients that deduplicate their own data, NetBackup backs up the client normally.

To disable client deduplication for a client

1. In the NetBackup Administration Console, expand **NetBackup Management > Host Properties > Master Servers**.
2. In the details pane, select the master server.
3. On the **Actions** menu, select **Properties**.
4. On the **Host Properties Client Attributes General** tab, select the client that deduplicates its own data.
5. In the **Deduplication Location** drop-down list, select **Always use the media server**.
6. Click **OK**.

About maintenance processing

The following are background maintenance processes:

- NetBackup Deduplication Engine queue processing.
  Operations that require database updates accumulate in a transaction queue. Twice a day, the NetBackup Deduplication Manager directs the deduplication engine to process the queue as one batch. The schedule is frequency-based. By default, queue processing occurs every 12 hours, 20 minutes past the hour.
  Queue processing is CPU-bound. The `postgres` database process usually consumes 100% of the CPU cycles during queue processing.
- NetBackup Deduplication Engine garbage collection.
  In a few rare scenarios, some data segments may become orphaned. Garbage collection cleans these segments up by removing them.
  Garbage collection is an unobtrusive process; once a week the NetBackup Deduplication Manager directs the deduplication engine to collect and remove garbage.

NetBackup cannot change the maintenance process schedules. Because these processes do not block any other deduplication process, rescheduling should not be necessary. However, it you must reschedule these processes, contact your Symantec support representative.

See "NetBackup Deduplication Manager logs" on page 84.

Chapter 7

Troubleshooting

This chapter includes the following topics:

Troubleshooting installation issues

The following sections may help you troubleshoot configuration issues.

See "Installation on SUSE Linux fails" on page 93.

Installation on SUSE Linux fails

The installation trace log shows an error when you install on SUSE:

```
....NetBackup and Media Manager are normally installed in /usr/openv.
Is it OK to install in /usr/openv? [y,n] (y)

Reading NetBackup files from /net/nbstore/vol/test_data/PDDE_packages/
suse/NB_FID2740_LinuxS_x86_20090713_6.6.0.27209/linuxS_x86/anb

/net/nbstore/vol/test_data/PDDE_packages/suse/NB_FID2740_LinuxS_x86_
20090713_6.6.0.27209/linuxS_x86/catalog/anb/NB.file_trans: symbol
lookup error: /net/nbstore/vol/test_data/PDDE_packages/suse/
NB_FID2740_LinuxS_x86_20090713_6.6.0.27209/linuxS_x86/catalog/anb/
NB.file_trans: undefined symbol: head /net/nbstore/vol/test_data/
```

```
PDDE_packages/suse/NB_FID2740_LinuxS_x86_20090713_6.6.0.27209/
linuxS_x86/catalog/anb/NB.file_trans failed. Aborting ...
```

Verify that your system is at patch level 2 or later, as follows:

```
cat /etc/SuSE-release
SUSE Linux Enterprise Server 10 (x86_64)
VERSION = 10
PATCHLEVEL = 2
```

Troubleshooting configuration issues

The following sections may help you troubleshoot configuration issues.

See "Cannot configure deduplication storage server" on page 94.

See "The disk pool wizard does not display a volume" on page 95.

Cannot configure deduplication storage server

If you cannot configure a deduplication storage server or load balancing servers, your network environment may not be configured for DNS reverse name lookup.

A configuration error message about license information failure indicates that the NetBackup servers cannot communicate with each other.

You can edit the hosts file on the media servers that you use for deduplication. Alternatively, you can configure NetBackup so it does not use reverse name lookup.

To prohibit reverse host name lookup by using the Administration Console

1. In the **NetBackup Administration Console**, expand **NetBackup Management > Host Properties > Master Servers**.
2. In the details pane, select the master server.
3. On the **Actions** menu, select **Properties**.
4. In the **Master Server Properties** dialog box, select the **Network Settings** properties.
5. Select one of the following options:
   - **Allowed**
   - **Restricted**
   - **Prohibited**

   For a description of these options, see the NetBackup online Help or the administrator's guide.

To prohibit reverse host name lookup by using the `bpsetconfig` command

- Enter the following command on each media server that you use for deduplication:

  ```
  echo REVERSE_NAME_LOOKUP = PROHIBITED | bpsetconfig -h host_name
  ```

  The `bpsetconfig` command resides in the following directories:

  UNIX: `/usr/openv/netbackup/bin/admincmd`

  Windows: *`install_path`*`\Veritas\NetBackup\bin\admincmd`

The disk pool wizard does not display a volume

The **Disk Pool Configuration Wizard** does not display a disk volume for the deduplication storage server.

First, restart all of the NetBackup daemons or services. The step ensures that the NetBackup Deduplication Engine is up and ready to respond to requests.

Second, restart the **NetBackup Administration Console**. This step clears cached information from the failed attempt to display the disk volume.

Troubleshooting operational issues

The following sections may help you troubleshoot operational issues.

See "Verify that the server has sufficient memory" on page 95.

See "Backup jobs fail" on page 96.

See "Volume state changes to DOWN when volume is unmounted" on page 96.

See "Errors, delayed response, hangs" on page 97.

Verify that the server has sufficient memory

Insufficient memory on the storage server can cause operation problems. If you have operation issues, you should verify that your storage server has sufficient memory.

See "About deduplication server requirements" on page 23.

If the NetBackup deduplication processes do no start on Red Hat Linux, configure shared memory to be at least 128 MB (`SHMMAX=128MB`).

Backup jobs fail

If backup jobs fail with an `Error 800: Disk Volume is Down` message, examine the disk error logs to determine why the volume was marked DOWN.

If the storage server is busy with jobs, it may not respond to master server disk polling requests in a timely manner. A busy load balancing server also may cause this error. Consequently, the query times out and the master server marks the volume DOWN.

If the error occurs for an optimized duplication job: verify that source storage server is configured as a load balancing server for the target storage server. Also verify that the target storage server is configured as a load balancing server for the source storage server.

See "Viewing disk errors and events" on page 97.

Volume state changes to DOWN when volume is unmounted

If a volume becomes unmounted, NetBackup changes the volume state to DOWN. NetBackup jobs that require that volume fail.

To determine the volume state

- Invoke the following command on the master server or the media server that functions as the deduplication storage server:

  The following example output shows that the `DiskPoolVolume` is UP:

```
Disk Pool Name      : PD_Disk_Pool
Disk Type           : PureDisk
Disk Volume Name    : PureDiskVolume
Disk Media ID       : @aaaab
Total Capacity (GB) : 49.98
Free Space (GB)     : 43.66
Use%                : 12
Status              : UP
Flag                : ReadOnWrite
Flag                : AdminUp
Flag                : InternalUp
Num Read Mounts     : 0
Num Write Mounts    : 1
Cur Read Streams    : 0
Cur Write Streams   : 0
```

To change the volume state to UP

- Mount the file system

  After a brief period of time, the volume state changes to UP. No further action is required.

Errors, delayed response, hangs

Insufficient memory or inadequate host capabilities may cause multiple errors, delayed response, and hangs.

See "About deduplication server requirements" on page 23.

For virtual machines, Symantec recommends that you do the following:

- Set the memory size of each virtual machine to double the physical memory of the host.
- Set the minimum and the maximum values of each virtual machine to the same value (double the physical memory of the host). These memory settings prevent the virtual memory from becoming fragmented on the disk because it does not grow or shrink.

These recommendations may not be the best configuration for every virtual machine. However, Symantec recommends that you try this solution first when troubleshooting performance issues.

Viewing disk errors and events

You can view disk errors and events in several ways, as follows:

- The Disk Logs report.
  See "Viewing disk reports" on page 81.
- The NetBackup `bperror` command with the `-disk` option reports on disk errors. The command resides in the following directories:

  UNIX: `/usr/openv/netbackup/bin/admincmd`

  Windows: *`install_path`*`\Veritas\NetBackup\bin\admincmd`

Deduplication event codes and messages

The following table shows the deduplication event codes and their messages. Event codes appear in the `bperror` command `-disk` output and in the disk reports in the NetBackup Administration Console.

Table 7-1 Deduplication event codes and messages

Event #	Event Severity	NetBackup Severity	Message example
1000	2	Error	`Operation configload/reload failed on server PureDisk:server1.symantecs.orgon host server1.symantecs.org.`
1001	2	Error	`Operation configload/reload failed on server PureDisk:server1.symantecs.orgon host server1.symantecs.org.`
1002	4	Warning	`The open file limit exceeded in server PureDisk:server1.symantecs.orgon host server1.symantecs.org. Will attempt to continue further.`
1003	2	Error	`A connection request was denied on the server PureDisk:server1.symantecs.orgon host server1.symantecs.org.`
1004	1	Critical	`Network failure occurred in server PureDisk:server1.symantecs.org on host server1.symantecs.org.`
1013	1	Critical	`Task session start request on server PureDisk:server1.symantecs.org on host server1.symantecs.org got an unexpected error.`
1008	2	Error	`Task Aborted; An unexpected error occurred during communication with remote system in server PureDisk:server1.symantecs.orgon host server1.symantecs.org.`
1009	8	Authorization	`Authorization request from <IP> for user <USER> denied (<REASON>).`
1010	2	Error	`Task initialization on server PureDisk:server1.symantecs.org on host server1.symantecs.org got an unexpected error.`
1011	2	Error	`Task ended on server PureDisk:server1.symantecs.org on host server1.symantecs.org.`

Table 7-1 Deduplication event codes and messages *(continued)*

Event #	Event Severity	NetBackup Severity	Message example
1012	2	Error	A request for agent task was denied on server PureDisk:server1.symantecs.orgon host server1.symantecs.org.
1014	1	Critical	Task session start request on server PureDisk:server1.symantecs.orgon host server1.symantecs.org got an unexpected error.
1015	1	Critical	Task creation failed, could not initialize task class on server PureDisk:server1.symantecs.orgon host server1.symantecs.org.
1017	1	Critical	Service Symantec DeduplicationEngine exit on server PureDisk:server1.symantecs.org on host server1.symantecs.org. Please check the server log for the probable cause of this error. The application has terminated.
1018	16	Info	Startup of Symantec DeduplicationEngine completed successfully on server1.symantecs.org.
1019	1	Critical	Service Symantec DeduplicationEngine restart on server PureDisk:server1.symantecs.orgon host server1.symantecs.org. Please check the server log for the probable cause of this error. The application has restarted.
1020	1	Critical	Service Symantec Deduplication Engine connection manager restart failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org. Please check the server log for the probable cause of this error.The application has failed to restart.

Table 7-1 Deduplication event codes and messages *(continued)*

Event #	Event Severity	NetBackup Severity	Message example
1028	1	Critical	`Service Symantec DeduplicationEngine abort on server PureDisk:server1.symantecs.orgon host server1.symantecs.org. Please check the server log for the probable cause of this error.The application has caught an unexpected signal.`
1029	1	Critical	`Double backend initialization failure; Could not initialize storage backend or cache failure detected on host PureDisk:server1.symantecs.org in server server1.symantecs.org.`
1030	1	Critical	`Operation Storage Database Initialization failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.`
1031	1	Critical	`Operation Content router context initialization failed on server PureDisk:server1.symantecs.orgon host server1.symantecs.org.`
1032	1	Critical	`Operation log path creation/print failed on server PureDisk:server1.symantecs.orgon host server1.symantecs.org.`
1036	4	Warning	`Operation a transaction failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.`
1037	4	Warning	`Transaction failed on server PureDisk:server1.symantecs.orgon host server1.symantecs.org. Transaction will be retried.`

Table 7-1 Deduplication event codes and messages *(continued)*

Event #	Event Severity	NetBackup Severity	Message example
1044	multiple	multiple	The usage of one or more system resources has exceeded a warning level. Operations will or could be suspended. Please take action immediately to remedy this situation.
1040	2	Error	Operation Database recovery failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.
1043	2	Error	Operation Storage recovery failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.
1047	2	Error	CRC mismatch detected; possible corruption in server PureDisk:server1.symantecs.org on host server1.symantecs.org.
2000		Error	Low space threshold exceeded on the partition containing the storage database on server PureDisk:server1.symantecs.org on host server1.symantecs.org.

Chapter 8

Disaster recovery

This chapter includes the following topics:

Preparing for disaster

Symantec recommends that you get and save the storage server configuration. Getting and saving the configuration can help you with recovery of your environment. For disaster recovery, you may need to set the storage server configuration by using a saved configuration file.

See "Getting the storage server configuration" on page 68.

Moving images off-site

Optimized duplication moves the primary backup data to other deduplication pools. It provides the easiest, most efficient method to move data off-site. You then can recover from a disaster that destroys the storage on which the primary copies reside by retrieving images from the other deduplication pool.

See "About optimized duplication of deduplicated data" on page 28.

See "Configuring optimized deduplication copy" on page 57.

Recovering from a deduplication storage server disk failure

If recovery mechanisms do not protect the disk on which the NetBackup software resides, the deduplication storage server configuration is lost if the disk fails. This topic describes how to recover from a system disk or program disk failure where the disk was not backed up.

Note: This procedure describes recovery of the disk on which the NetBackup media server software resides not the disk on which the deduplicated data resides. The disk may or may not be the system boot disk.

Symantec recommends that you use NetBackup to protect the deduplication storage server system or program disks. You then can use NetBackup to restore that media server if the disk on which NetBackup resides fails and you have to replace it.

Table 8-1 Process to recover from media server disk failure

Replace the disk.	If the disk is a system boot disk, also install the operating system. See the hardware vendor and operating system documentation.
Mount the storage.	Ensure that the storage and database are mounted at the same locations. See the storage vendor's documentation.
Install and license the NetBackup media server software.	See the *NetBackup Installation Guide for UNIX and Linux*. See the *NetBackup Installation Guide for Windows*. See "About the deduplication license key" on page 42.
Delete the configuration file on media servers	If you use load balancing servers in your environment, delete the storage server configuration files on those servers. See "Deleting a load balancing server configuration file" on page 70.

Table 8-1 Process to recover from media server disk failure *(continued)*

Delete the credentials on media servers	If you have load balancing servers, delete the NetBackup Deduplication Engine credentials on those media servers. See "Deleting credentials from a load balancing server" on page 75.
Add the credentials to the storage server	Add the NetBackup Deduplication Engine credentials to the storage server. See "Adding NetBackup Deduplication Engine credentials" on page 74.
Get a configuration file template	If you did not save a storage server configuration file before the disk failure, get a template configuration file. See "Getting the storage server configuration" on page 68.
Edit the configuration file	See "Editing a storage server configuration file" on page 68.
Configure the storage server	Configure the storage server by uploading the configuration from the file you edited. If you saved a configuration file before the disk failure, use that file. See "Setting the storage server configuration" on page 70.
Add load balancing servers	If you use load balancing servers in your environment, add them to your configuration. See "Adding a load balancing server" on page 58.

Recovering from a permanent deduplication storage server failure

To recover from a permanent media server failure, use the process that is described in the following table. For the new host, you must use the same host name.

Information about recovering the master server is available.

See the *NetBackup Troubleshooting Guide*.

Warning: The new host must use the same byte order as the old host. If it does not, you cannot access the deduplicated data.

(In computing, endianness describes the byte order that represents data: big endian and little endian. For example, Sun SPARC processors and Intel processors use different byte orders. Therefore, you cannot replace a Solaris SPARC host with a host that has an Intel processor.

Table 8-2 Process to recover from a permanent media server failure

Task	Procedure
Change the disk volume state and disk pool state to DOWN	See "Changing the deduplication disk volume state" on page 77. See "Changing the deduplication pool state" on page 77.
Configure the new host so it meets deduplication requirements	Use the same host name as the failed server. See "About deduplication servers" on page 21. See "About deduplication server requirements" on page 23.
Move the storage to the new host.	Ensure that the storage and database are mounted at the same locations. See the storage vendor's documentation.
Install the NetBackup media server software on the new host	See the *NetBackup Installation Guide for UNIX and Linux*. See the *NetBackup Installation Guide for Windows*. See "About the deduplication license key" on page 42.
Delete the credentials on media servers	If you have load balancing servers, delete the NetBackup Deduplication Engine credentials on those media servers. See "Deleting credentials from a load balancing server" on page 75.
Add the credentials to the storage server	Add the NetBackup Deduplication Engine credentials to the storage server. See "Adding NetBackup Deduplication Engine credentials" on page 74.
Get a configuration file template	If you did not save a storage server configuration file before the failure, get a template configuration file. See "Getting the storage server configuration" on page 68.
Edit the configuration file	See "Editing a storage server configuration file" on page 68.

Table 8-2 Process to recover from a permanent media server failure *(continued)*

Task	Procedure
Configure the storage server	Configure the storage server by uploading the configuration from the file you edited. If you saved a configuration file before the storage server failure, use that file. See "Setting the storage server configuration" on page 70.
Configure the load balancing servers	If you have load balancing servers, add them to the configuration. See "Adding a load balancing server" on page 58.
Change configuration settings	If you edited the deduplication configuration file, make the same changes to that file. See "About the deduplication configuration file" on page 60. See "Editing the deduplication configuration file" on page 60.
Change the disk volume state and disk pool state to UP	See "Changing the deduplication disk volume state" on page 77. See "Changing the deduplication pool state" on page 77.
Restart the backup jobs	If any backup jobs failed, restart those jobs. Alternatively, wait until the next scheduled backup, at which time the backup jobs should succeed.

Recovering the storage server after NetBackup catalog recovery

If a disaster requires a recovery of the NetBackup catalog, you must set the storage server configuration after the NetBackup catalog is recovered.

See "Setting the storage server configuration" on page 70.

Symantec recommends that you save your storage server configuration.

See "Preparing for disaster" on page 103.

Information about recovering the master server is available.

See the *NetBackup Troubleshooting Guide*.

Chapter 9

Deduplication architecture

This chapter includes the following topics:

Deduplication server components

Figure 9-1 is a diagram of the storage server components.

Figure 9-1 Storage server deduplication components

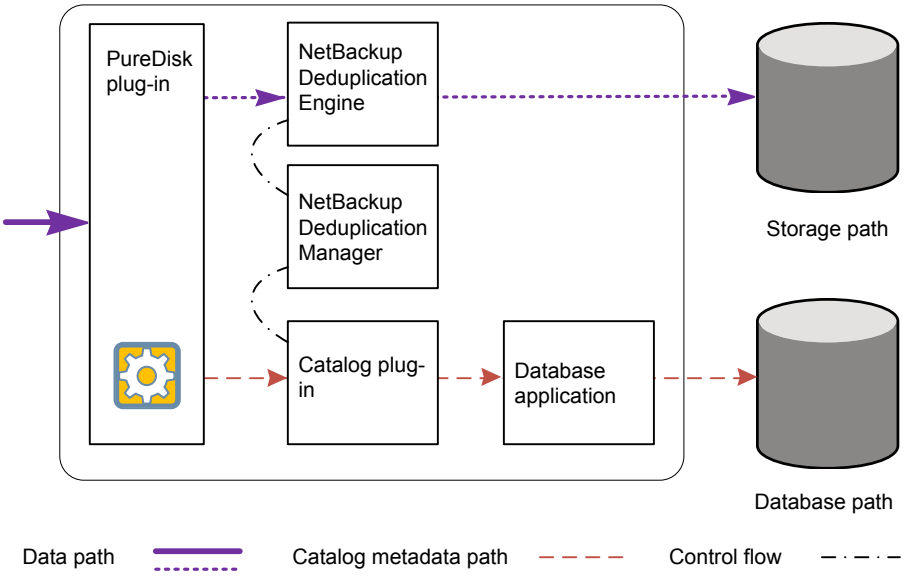


Table 9-1 describes the components.

Table 9-1 NetBackup deduplication components

Component	Description
PureDisk plug-in	The PureDisk plug-in is the data interface to the NetBackup Deduplication Engine on the storage server. The PureDisk plug-in does the following: ■ Separates the file's metadata from the file's content. ■ Deduplicates the content (separates files into segments). ■ Controls the data stream from NetBackup to the NetBackup Deduplication Engine and vice versa. The plug-in runs on the deduplication storage server. The plug-in also runs on load balancing servers and on the clients that deduplicate their own data.
NetBackup Deduplication Engine	The NetBackup Deduplication Engine is one of the storage server core components. It stores and manages deduplicated file data. The binary file name is `spoold`, which is short for storage pool daemon; do not confuse it with a print spooler daemon. The `spoold` process appears as the NetBackup Deduplication Engine in the NetBackup Administration Console.

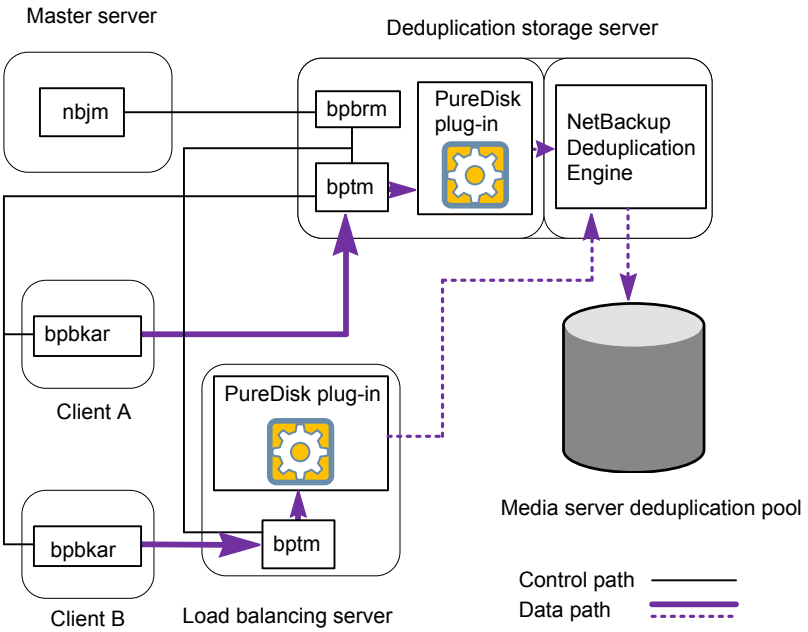
Table 9-1 NetBackup deduplication components *(continued)*

Component	Description
NetBackup Deduplication Manager	The deduplication manager is one of the storage server core components. The deduplication manager maintains the configuration and controls internal processes, optimized duplication, security, and event escalation. The deduplication manager binary file name is `spad`. The `spad` process appears as the NetBackup Deduplication Manager in the NetBackup Administration Console.
Catalog plug-in	The catalog plug-in implements a standardized catalog API, which lets the NetBackup Deduplication Engine communicate with the back-end database process. The catalog plug-in translates deduplication engine catalog calls into the calls that are native to the back-end database.
Deduplication database	The deduplication database stores and manages the metadata of deduplicated files. The metadata includes a unique fingerprint that identifies the file's content. The metadata also includes information about the file such as its owner, where it resides on a client, when it was created, and other information. NetBackup uses the PostgresSQL database for the deduplication database. You can use the NetBackup `bpps` command to view the database process (`postgres`). The deduplication database is separate from the NetBackup catalog. The NetBackup catalog maintains the usual NetBackup backup image information. On Windows systems, NetBackup creates a `purediskdbuser` account for database management.

Media server deduplication process

Figure 9-2 shows the backup process when a media server deduplicates the backups. The destination is a media server deduplication pool. A description follows.

Figure 9-2 Deduplication to a media server deduplication pool



The following list describes the backup process when a media server deduplicates the backups and the destination is a media server deduplication pool:

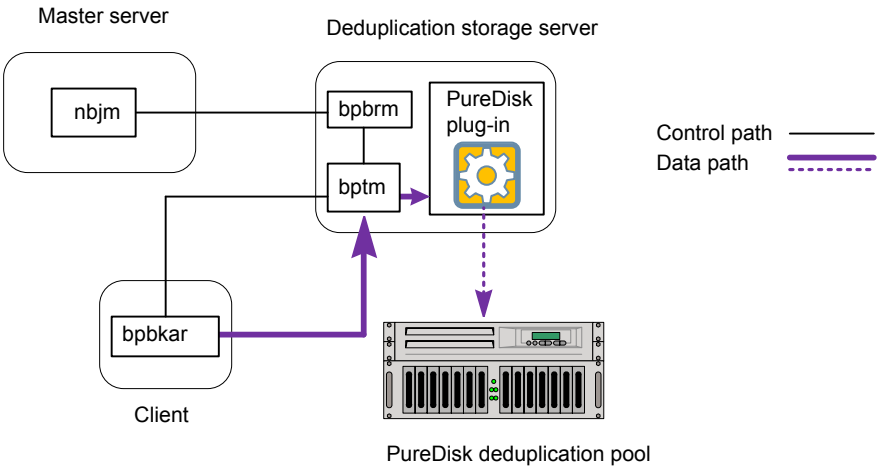
- The NetBackup Job Manager (`nbjm`) starts the Backup/Restore Manager (`bpbrm`) on a media server.
- The Backup/Restore Manager starts the `bptm` process on the media server and the `bpbkar` process on the client.
- The Backup/Archive Manager (`bpbkar`) generates the backup images and moves them to the media server `bptm` process.

  This example shows both a deduplication storage server and a load balancing server. The storage server both deduplicates data and manages the storage.
- The `bptm` process moves the data to the PureDisk plug-in.
- The PureDisk plug-in retrieves a list of fingerprints from the last full backup for the client from the NetBackup Deduplication Engine. The list is used as a cache so the plug-in does not have to request each fingerprint from the engine.
- The PureDisk plug-in performs file fingerprinting calculations.
- The PureDisk plug-in compares the file fingerprints and the segment fingerprints against the fingerprint list it its cache.

- The PureDisk plug-in sends only unique data segments to the NetBackup Deduplication Engine on the storage server. The NetBackup Deduplication Engine writes the data to the media server deduplication pool.

Figure 9-3 shows the backup process when a media server deduplicates the backups. The destination is a PureDisk storage pool. A description follows.

Figure 9-3 Deduplication to a PureDisk storage pool



The following list describes the backup process when a media server deduplicates the backups and the destination is a PureDisk storage pool:

- The NetBackup Job Manager (`nbjm`) starts the Backup/Restore Manager (`bpbrm`) on a media server.
- The Backup/Restore Manager starts the `bptm` process on the media server and the `bpbkar` process on the client).
- The Backup/Archive Manager (`bpbkar`) generates the backup images and moves them to the media server `bptm` process.

  This example shows two media servers.
- The `bptm` process moves the data to the PureDisk plug-in.
- The PureDisk plug-in performs file fingerprinting calculations.
- The PureDisk plug-in sends only unique data segments to the PureDisk storage pool.

Deduplication client components

Table 9-2 describes the client deduplication components.

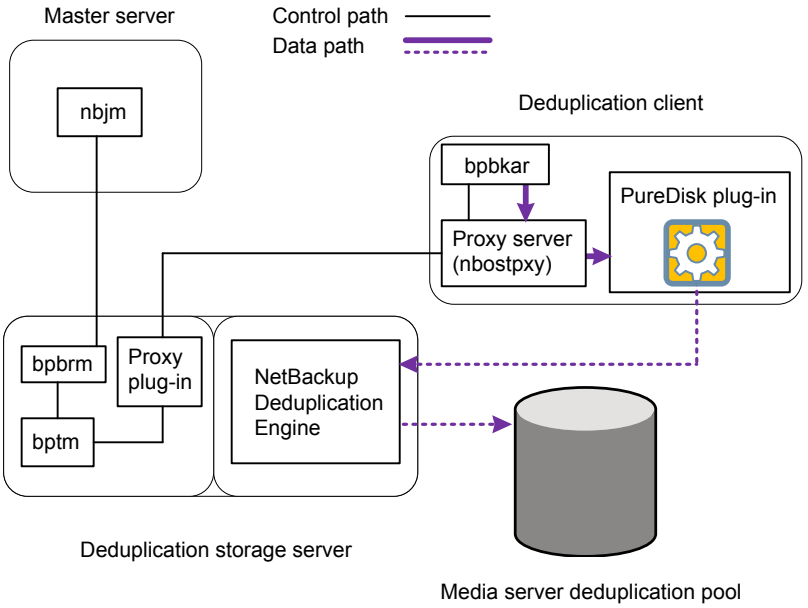
Table 9-2 Client deduplication components

Component	Host	Description
PureDisk plug-in	Client	The PureDisk plug-in is the data interface to the NetBackup Deduplication Engine on the deduplication storage server. The PureDisk plug-in does the following: ■ Separates the file's metadata from the file's content. ■ Deduplicates the content (separates files into segments). ■ Controls the data stream from NetBackup to the NetBackup Deduplication Engine and vice versa.
Proxy server	Client	The OpenStorage proxy server (`nbostpxy`) manages control communication with the media server.
Proxy plugin	Media server	The proxy plug-in manages control communication with the client.

Deduplication client backup process

Figure 9-4 shows the backup process of a client that deduplicates its own data. The destination is a media server deduplication pool. A description follows.

Figure 9-4 Deduplication client backup to a media server deduplication pool



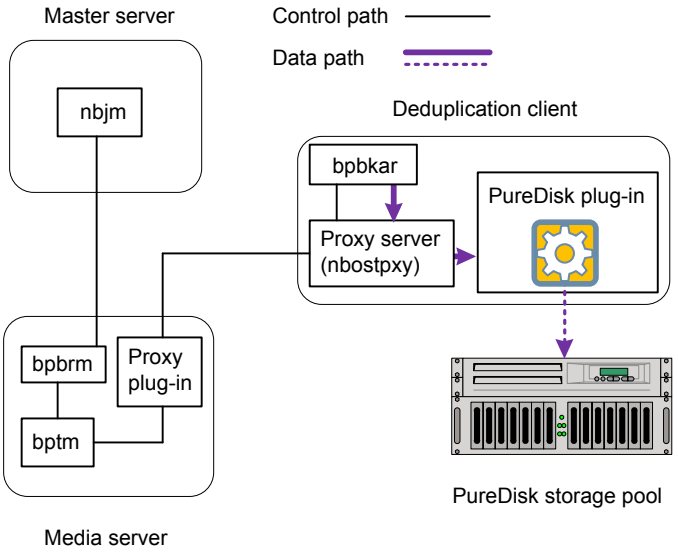
The following list describes the backup process for a deduplication client to a media server deduplication pool:

- The NetBackup Job Manager (`nbjm`) starts the Backup/Restore Manager (`bpbrm`) on a media server.
- The Backup / Restore Manager probes the client to determine if it is configured and ready for deduplication.
- If the client is ready, the Backup/Restore Manager starts the following processes: The OpenStorage proxy server (`nbostpxy`) on the client and the data moving processes (`bpbkar` on the client and `bptm` on the media server). NetBackup uses the proxy plug-in on the media server to route control information from `bptm` to `nbostpxy`.
- The Backup/Archive Manager (`bpbkar`) generates the backup images and moves them to the client `nbostpxy` process by shared memory.
- The client `nbostpxy` process moves the data to the PureDisk plug-in.
- The PureDisk plug-in retrieves a list of fingerprints from the last full backup for the client from the NetBackup Deduplication Engine. The list is used as a cache so the plug-in does not have to request each fingerprint from the engine.

- The PureDisk plug-in performs file fingerprinting calculations.
- The PureDisk plug-in sends only unique data segments to the storage server, which writes the data to the media server deduplication pool.

Figure 9-5 shows the backup process of a client that deduplicates its own data. The destination is a PureDisk storage pool. A description follows.

Figure 9-5 Deduplication client backup to a PureDisk storage pool



The following list describes the backup process for a deduplication client to a media server deduplication pool:

- The NetBackup Job Manager (`nbjm`) starts the Backup/Restore Manager (`bpbrm`) on a media server.
- The Backup / Restore Manager probes the client to determine if it is configured and ready for deduplication.
- If the client is ready, the Backup/Restore Manager starts the following processes: The OpenStorage proxy server (`nbostpxy`) on the client and the data moving processes (`bpbkar` on the client and `bptm` on the media server). NetBackup uses the proxy plug-in on the media server to route control information from `bptm` to `nbostpxy`.
- The Backup/Archive Manager (`bpbkar`) generates the backup images and moves them to the client `nbostpxy` process by shared memory.

- The client `nbostpxy` process moves the data to the PureDisk plug-in.
- The PureDisk plug-in retrieves a list of fingerprints from the last full backup for the client from the NetBackup Deduplication Engine. The list is used as a cache so the plug-in does not have to request each fingerprint from the engine.
- The PureDisk plug-in performs file fingerprinting calculations.
- The PureDisk plug-in sends only unique data segments to the PureDisk storage pool.

About deduplication fingerprinting

The NetBackup Deduplication Engine uses a unique identifier to identify each file and each file segment that is backed up. The engine identifies files inside the backup images and then processes the files.

The process is known as fingerprinting.

For the first deduplicated backup, the following is the process:

- The PureDisk plug-in reads the backup image and separates the image into files.
- The plug-in separates files into segments.
- For each segment, the plug-in calculates the hash key (or fingerprint) that identifies each data segment. To create a hash, every byte of data in the segment is read and added to the hash.
- The plug-in compares it calculated fingerprints to the fingerprints that the NetBackup Deduplication Engine stores on the media server. Two segments that have the same fingerprint are duplicates of each other.
- The plug-in sends unique segments to the deduplication engine to be stored. A unique segment is one for which a matching fingerprint does not exist in the engine already.

  The first backup may have a 0% deduplication rate; however, a 0% deduplication rate is unlikely. Zero percent means that all file segments in the backup data are unique.
- The NetBackup Deduplication Engine saves the fingerprint information for that backup.

For subsequent backups, the following is the process:

- The PureDisk plug-in retrieves a list of fingerprints from the last full backup for the client from the NetBackup Deduplication Engine. The list is used as a cache so the plug-in does not have to request each fingerprint from the engine.

- The PureDisk plug-in reads the backup image and separates the image into files.
- The PureDisk plug-in separates files into segments and calculates the fingerprint for each file and segment.
- The plug-in compares each fingerprint against the local fingerprint cache. If the fingerprint is not known in the cache, the plug-in requests that the engine verify if the fingerprint already exists.
- If the fingerprint does not exist, the segment is sent to the engine. If the fingerprint exists, the segment is not sent.

The fingerprint calculations are based on the MD5 algorithm. However, any segments that have different content but the same MD5 hash key get different fingerprints. So NetBackup prevents MD5 collisions.

Data removal process

The following list describes the data removal process for expired backup images:

- NetBackup removes the image record from the NetBackup catalog.
  NetBackup directs the NetBackup Deduplication Manager to remove the image.
- The deduplication manager immediately removes the image entry and adds a removal request for the image to the database transaction queue.
  From this point on, the image is no longer accessible.
- When the queue is next processed, the NetBackup Deduplication Engine executes the removal request. The engine also generates removal requests for underlying data segments
- At the successive queue processing, the NetBackup Deduplication Engine executes the removal requests for the segments.

Storage is reclaimed after two queue processing runs; that is, in one day. However, data segments of the removed image may still be in use by other images.

If you manually delete an image that has expired within the previous 24 hours, the data becomes garbage. It remains on disk until removed by the next garbage collection process.

See “About maintenance processing” on page 90.

Index

E

F

G

H

I

L

M

N

O

P

R

S

T

U