Symantec NetBackup™ Deduplication Guide

Release 7.0.1
The software described in this book is furnished under a license agreement and may be used only in accordance with the terms of the agreement.

Documentation version: 7.0.1

Legal Notice

Copyright © 2010 Symantec Corporation. All rights reserved.

Symantec, the Symantec Logo, and NetBackup are trademarks or registered trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

This Symantec product may contain third party software for which Symantec is required to provide attribution to the third party ("Third Party Programs"). Some of the Third Party Programs are available under open source or free software licenses. The License Agreement accompanying the Software does not alter any rights or obligations you may have under those open source or free software licenses. Please see the Third Party Legal Notice Appendix to this Documentation or TPIP ReadMe File accompanying this Symantec product for more information on the Third Party Programs.

The product described in this document is distributed under licenses restricting its use, copying, distribution, and decompilation/reverse engineering. No part of this document may be reproduced in any form by any means without prior written authorization of Symantec Corporation and its licensors, if any.

THE DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID. SYMANTEC CORPORATION SHALL NOT BE LIABLE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH THE FURNISHING, PERFORMANCE, OR USE OF THIS DOCUMENTATION. THE INFORMATION CONTAINED IN THIS DOCUMENTATION IS SUBJECT TO CHANGE WITHOUT NOTICE.

The Licensed Software and Documentation are deemed to be commercial computer software as defined in FAR 12.212 and subject to restricted rights as defined in FAR Section 52.227-19 "Commercial Computer Software - Restricted Rights" and DFARS 227.7202, "Rights in Commercial Computer Software or Commercial Computer Software Documentation", as applicable, and any successor regulations. Any use, modification, reproduction release, performance, display or disclosure of the Licensed Software and Documentation by the U.S. Government shall be solely in accordance with the terms of this Agreement.
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 support 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/or Web-based support that provides rapid response and up-to-the-minute information
- Upgrade assurance that delivers software upgrades
- Global support purchased on a regional business hours or 24 hours a day, 7 days a week basis
- Premium service offerings that include Account Management Services

For information about Symantec’s support offerings, you can visit our Web site at the following URL:

www.symantec.com/business/support/

All support services will be delivered in accordance with your support agreement and the then-current enterprise technical support policy.

Contacting Technical Support

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

www.symantec.com/business/support/

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/business/support/

Customer service

Customer service information is available at the following URL:

www.symantec.com/business/support/

Customer Service is available to assist with non-technical questions, such as 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 support 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
Support agreement resources

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

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

Chapter 1 Introducing NetBackup deduplication

- About NetBackup deduplication ................................................................. 13
- About NetBackup deduplication options ........................................... 14
- How deduplication works ........................................................................ 15

Chapter 2 Planning your deployment

- Planning your deduplication deployment .............................................. 18
- For more information ............................................................................. 19
- About the deduplication storage destination ........................................ 19
- About the NetBackup Media Server Deduplication Option ................. 21
  - About deduplication servers ................................................................. 22
  - About deduplication nodes ................................................................. 23
  - About deduplication server requirements ......................................... 24
  - About media server deduplication limitations ..................................... 25
- About NetBackup Client Deduplication .................................................. 26
  - About client deduplication host requirements ..................................... 27
  - About client deduplication requirements ........................................... 27
  - About client deduplication limitations ............................................... 28
- About NetBackup Deduplication Engine credentials ............................... 28
- About the network interface for deduplication ...................................... 28
- About deduplication port usage ............................................................. 29
- About compression and encryption ......................................................... 29
- About optimized synthetic backups and deduplication ......................... 30
- About optimized duplication of deduplicated data .................................. 31
  - Optimized duplication copy requirements ......................................... 31
  - Optimized duplication copy limitations ............................................. 32
  - About the media server in common .................................................... 32
- About deduplication and SAN Client ...................................................... 41
- About deduplication performance .......................................................... 42
  - How file size may affect the deduplication rate ................................... 43
- About deduplication stream handlers ..................................................... 43
- About iSCSI, CIFS, and NFS ................................................................. 44
- Deployment best practices ...................................................................... 44
Use fully qualified domain names ............................................. 44
About scaling deduplication ..................................................... 44
Send the initial full backups to the storage server .................... 45
Increase the number of jobs gradually ...................................... 45
Introduce load balancing servers gradually .............................. 46
Implement client deduplication gradually ............................... 46
Use deduplication compression or encryption .......................... 47
About the optimal number of backup streams .......................... 47
About storage unit groups for deduplication ............................ 47
About protecting the deduplicated data ................................. 48
Save the storage server configuration ...................................... 48
How deduplication restores work .......................................... 48
Replacing the PureDisk Deduplication Option with Media Server
  Deduplication on the same host ........................................... 49
Migrating from PureDisk to the NetBackup Media Server
  Deduplication option .......................................................... 50
Migrating from another storage type to deduplication .............. 51

Chapter 3  Provisioning the storage ................................ .......... 53
  About provisioning the storage .......................................... 53
  About deduplication storage requirements .......................... 53
  About deduplication storage capacity ................................ 54
  About the deduplication storage paths ............................... 54
  About adding additional storage ...................................... 55

Chapter 4  Licensing deduplication ......................................... 57
  About Licensing deduplication ......................................... 57
  About the deduplication license key .................................. 58
  Licensing NetBackup deduplication .................................. 58

Chapter 5  Configuring deduplication ...................................... 59
  Configuring deduplication ................................................. 60
  Configuring a deduplication storage server .......................... 61
  About deduplication pools ................................................. 62
  Configuring a deduplication pool ....................................... 62
    Media server deduplication pool properties ....................... 63
  Configuring a deduplication storage unit ............................ 64
    Deduplication storage unit properties ............................ 65
    Deduplication storage unit recommendations ..................... 66
  Enabling client deduplication ......................................... 68
  Enabling compression or encryption ............................... 68
Client deduplication fails .......................................................... 130
Volume state changes to DOWN when volume is
unmounted ..................................................................... 130
Errors, delayed response, hangs .......................................... 131
Deduplication fails after services are restarted or a domain
controller is restarted .................................................. 131
Cannot delete a disk pool ................................................... 132
Media open error (83) ....................................................... 133
Media write error (84) ..................................................... 134
Storage full conditions .................................................. 135
Viewing disk errors and events ......................................... 136
Deduplication event codes and messages ........................... 136

Chapter 9 Host replacement, recovery, and uninstallation .... 141
Replacing the deduplication storage server host computer .... 141
Recovering from a deduplication storage server disk failure .... 143
Recovering from a permanent deduplication storage server
failure ........................................................................ 145
Recovering the storage server after NetBackup catalog recovery .. 147
Uninstalling media server deduplication ............................ 147

Chapter 10 Deduplication architecture ................................. 151
Deduplication server components .................................... 151
Media server deduplication process ................................. 153
Deduplication client components .................................... 156
Deduplication client backup process ................................. 156
About deduplication fingerprinting .................................. 159
Data removal process ................................................... 160

Index .............................................................................. 163
Introducing NetBackup deduplication

This chapter includes the following topics:

- About NetBackup deduplication

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

NetBackup client-side deduplication, including clients at remote offices that send the backups to a central data center

NetBackup media server deduplication

Disk appliance deduplication by using the OpenStorage option

PureDisk deduplication without using NetBackup
### Table 1-1 NetBackup deduplication options

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetBackup Client Deduplication Option</td>
<td>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. See “About NetBackup Client Deduplication” on page 26.</td>
</tr>
<tr>
<td>NetBackup Media Server Deduplication Option</td>
<td>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. See “About the NetBackup Media Server Deduplication Option” on page 21.</td>
</tr>
<tr>
<td>Appliance deduplication</td>
<td>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.</td>
</tr>
<tr>
<td>PureDisk deduplication</td>
<td>NetBackup PureDisk is a deduplication solution that is not part of the NetBackup distribution. PureDisk provides bandwidth-optimized backups of data in remote offices. 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 <em>NetBackup PureDisk Getting Started Guide</em>. A PureDisk storage pool can be a storage destination for both the NetBackup Client Deduplication Option and the NetBackup Media Server Deduplication Option.</td>
</tr>
</tbody>
</table>

### 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.

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 153.
Planning your deployment

This chapter includes the following topics:

- Planning your deduplication deployment
- For more information
- About the deduplication storage destination
- About the NetBackup Media Server Deduplication Option
- About NetBackup Client Deduplication
- About NetBackup Deduplication Engine credentials
- About the network interface for deduplication
- About deduplication port usage
- About compression and encryption
- About optimized synthetic backups and deduplication
- About optimized duplication of deduplicated data
- About deduplication and SAN Client
- About deduplication performance
- About deduplication stream handlers
- About iSCSI, CIFS, and NFS
- Deployment best practices
- Replacing the PureDisk Deduplication Option with Media Server Deduplication on the same host
- Migrating from PureDisk to the NetBackup Media Server Deduplication option
Migrating from another storage type to deduplication

Planning your deduplication deployment

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

Table 2-1 Deployment overview

<table>
<thead>
<tr>
<th>Deployment task</th>
<th>Where to find the information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read the deduplication tech note</td>
<td>See “For more information” on page 19.</td>
</tr>
<tr>
<td>Determine the storage destination</td>
<td>See “About the deduplication storage destination” on page 19.</td>
</tr>
</tbody>
</table>
| Determine which type of deduplication to use         | See “About the NetBackup Media Server Deduplication Option” on page 21.  
| Determine the requirements for deduplication hosts   | See “About deduplication servers” on page 22.                      |
|                                                     | See “About deduplication server requirements” on page 24.          |
|                                                     | See “About client deduplication host requirements” on page 27.     |
|                                                     | See “About the network interface for deduplication” on page 28.    |
|                                                     | See “About deduplication port usage” on page 29.                   |
|                                                     | See “About scaling deduplication” on page 44.                      |
|                                                     | See “About deduplication performance” on page 42.                  |
| Determine the credentials for deduplication         | See “About NetBackup Deduplication Engine credentials” on page 28. |
| Read about compression and encryption               | See “About compression and encryption” on page 29.                 |
| Read about optimized synthetic backups              | See “About optimized synthetic backups and deduplication” on page 30. |
| Determine the requirements for optimized duplication| See “About optimized duplication of deduplicated data” on page 31.  |
### Table 2-1 Deployment overview (continued)

<table>
<thead>
<tr>
<th>Deployment task</th>
<th>Where to find the information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read about stream handlers</td>
<td>See “About deduplication stream handlers” on page 43.</td>
</tr>
<tr>
<td>Read about best practices for implementation</td>
<td>See “Deployment best practices” on page 44.</td>
</tr>
<tr>
<td>Determine the storage requirements and provision the storage</td>
<td>See “About provisioning the storage” on page 53.</td>
</tr>
<tr>
<td></td>
<td>See “About deduplication storage requirements” on page 53.</td>
</tr>
<tr>
<td></td>
<td>See “About deduplication storage capacity” on page 54.</td>
</tr>
<tr>
<td></td>
<td>See “About the deduplication storage paths” on page 54.</td>
</tr>
<tr>
<td>Replace a PDDO host or migrate from PDDO to NetBackup deduplication</td>
<td>See “Replacing the PureDisk Deduplication Option with Media Server Deduplication on the same host” on page 49.</td>
</tr>
<tr>
<td></td>
<td>See “Migrating from PureDisk to the NetBackup Media Server Deduplication option” on page 50.</td>
</tr>
<tr>
<td>Migrate from other storage to NetBackup deduplication</td>
<td>See “Migrating from another storage type to deduplication” on page 51.</td>
</tr>
</tbody>
</table>

### For more information

Symantec provides a tech note that includes the following:

- Currently supported systems
- Media server and client sizing information
- Configuration, operational, and troubleshooting updates
- And more

See the following link:

http://entsupport.symantec.com/docs/338123

### About the deduplication storage destination

In NetBackup 7.0 and later, several destinations exist for the deduplication storage, as follows:
A **Media Server Deduplication Pool** represents 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.

For a **Media Server Deduplication Pool** storage destination, all hosts that are used for the deduplication must be NetBackup 7.0 or later. Hosts include the master server, the media servers, and the clients that deduplicate their own data. Integrated deduplication means that the components installed with NetBackup perform deduplication.

A **PureDisk Deduplication Pool** represents a PureDisk storage pool. If you use a **PureDisk Deduplication Pool**, use the PureDisk documentation to plan, implement, configure, and manage the storage. A **PureDisk Deduplication Pool** destination 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.

For a **PureDisk Deduplication Pool** storage destination, all hosts that are used for the deduplication must be NetBackup 7.0 or later. Hosts include the master server, the media servers, and the clients that deduplicate their own data. Integrated deduplication means that the components installed with NetBackup perform deduplication.

A PureDisk storage pool destination that is not represented by a NetBackup construct. The required PureDisk Deduplication Option agent presents the PureDisk storage pool to NetBackup. In PDDO, the PureDisk Storage Pool Authority provides the PureDisk Deduplication Option agent that is installed on the NetBackup media servers. If you use a PureDisk storage pool, use the PureDisk documentation to plan, implement, configure, and manage the storage.

For a PureDisk storage pool destination, you can use NetBackup 6.5 or 7.0 NetBackup hosts. Hosts include the master server and the media servers. See the **NetBackup PureDisk Deduplication Option Guide**.

*Table 2-2* shows the storage destinations supported by the two NetBackup releases.

<table>
<thead>
<tr>
<th>NetBackup release</th>
<th>Media Server Deduplication Pool</th>
<th>PureDisk storage pool</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Table 2-2** NetBackup storage destination support
### About the NetBackup Media Server Deduplication Option

The 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 22.

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. The storage destination is a **Media Server Deduplication Pool**.

### Table 2-2: NetBackup storage destination support (continued)

<table>
<thead>
<tr>
<th>NetBackup release</th>
<th>Media Server Deduplication Pool</th>
<th>PureDisk storage pool</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Planning your deployment

About the NetBackup Media Server Deduplication Option
A PureDisk storage pool may also be the storage destination.

See “About the deduplication storage destination” on page 19.

More detailed information is available.

See “Deduplication server components” on page 151.

See “Media server deduplication process” on page 153.

About deduplication servers

Table 2-3 describes the servers that are used for NetBackup deduplication.
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

<table>
<thead>
<tr>
<th>Host</th>
<th>Description</th>
</tr>
</thead>
</table>
| 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.  
  See “About deduplication nodes” on page 23.                                                                                                                                                                                                                                           |
| 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 159.  
  A NetBackup media server becomes a load balancing server when two things occur:  
  ■ You enable the media server for deduplication load balancing duties.  
    You do so when you configure the storage server or later by modifying the storage server properties.  
  ■ You select it in the storage unit for the deduplication pool.  
  See “Introduce load balancing servers gradually” on page 46.  
  Load balancing servers also perform restore and duplication jobs.  
  Load balancing servers can be any supported server type for deduplication. They do not have to be the same type as the storage server.                                                                                                                                                                                          |
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

The host computer's CPU and memory constrain how many jobs can run concurrently. The storage server requires enough capability for deduplication and for storage management unless you offload some of the deduplication to load balancing servers.

Table 2-4 shows the minimum requirements for deduplication servers. NetBackup deduplication servers are always NetBackup media servers.

Processors for deduplication should have a high clock rate and high floating point performance. Furthermore, high throughput per core is desirable. Each backup stream uses a separate core.

Intel and AMD have similar performance and perform well on single core throughput.

Newer SPARC processors, such as the SPARC64 VII, provide the single core throughput that is similar to AMD and Intel. Alternatively, UltraSPARC T1 and T2 single core performance does not approach that of the AMD and Intel processors. Tests show that the UltraSPARC processors can achieve high aggregate throughput. However, they require eight times as many backup streams as AMD and Intel processors to do so.

Table 2-4  Deduplication server minimum requirements

<table>
<thead>
<tr>
<th>Component</th>
<th>Storage server</th>
<th>Load balancing server or PureDisk Deduplication Option host</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>Symantec recommends at least a 2.2-GHz clock rate. A 64-bit processor is required. At least four cores are required. Symantec recommends eight cores. See “About deduplication performance” on page 42. See “About maintenance processing” on page 114.</td>
<td>Symantec recommends at least a 2.2-GHz clock rate. A 64-bit processor is required. At least two cores are required. Depending on throughput requirements, more cores may be helpful.</td>
</tr>
</tbody>
</table>
Table 2-4  Deduplication server minimum requirements (continued)

<table>
<thead>
<tr>
<th>Component</th>
<th>Storage server</th>
<th>Load balancing server or PureDisk Deduplication Option host</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAM</td>
<td>4 GBs to 32 GBs. If your storage exceeds 4 TBs, Symantec recommends at least 1 GB more of memory for every terabyte of additional storage. For example, 10 TBs of back-end data require 10 GBs of RAM, 32 TBs require 32 GBs of RAM, and so on.</td>
<td>4 GBs.</td>
</tr>
<tr>
<td>Operating system</td>
<td>The operating system must be a supported 64-bit operating system.</td>
<td>The operating system must be a supported 64-bit operating system.</td>
</tr>
</tbody>
</table>

**Note:** In some environments, a single host can function as both a NetBackup master server and as a deduplication server. Such environments typically run fewer than 100 total backup jobs a day. (Total backup jobs means backups to any storage destination, including deduplication and nondeduplication storage.) If you perform more than 100 backups a day, deduplication operations may affect master server operations.

**Note:** If you use an existing media server for deduplication, performance may be inferior to the equipment that meets the deduplication server minimum requirement guidelines.

**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 the PureDisk Deduplication Option (PDDO) agent that is installed from the PureDisk distribution.

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. You then can use that host as a front end for your PureDisk Storage Pool Authority. (The host must be a host type that is supported for NetBackup deduplication.)
See “Replacing the PureDisk Deduplication Option with Media Server Deduplication on the same host” on page 49.

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.)

NetBackup Client Deduplication is a solution for the following cases:

- Remote office or branch office backups to the data center.
- LAN connected file server
- Virtual machine backups.

Client-side deduplication is also a useful solution if a client host has unused CPU cycles or if the storage server or load balancing servers are overloaded.

*Figure 2-2* shows client deduplication. The deduplication storage server is a media server on which the deduplication core components are enabled. The storage destination is a **Media Server Deduplication Pool**
A PureDisk storage pool may also be the storage destination.
See “About the deduplication storage destination” on page 19.
More detailed information is available.
See “Deduplication client components” on page 156.
See “Deduplication client backup process” on page 156.

About client deduplication host requirements
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.
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.

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.

The NetBackup `REQUIRED_INTERFACE` setting has no affect on deduplication processes.

### About deduplication port usage

The following table shows the ports that are used for NetBackup deduplication. If firewalls exist between the various deduplication hosts, open the indicated ports on the deduplication hosts. Deduplication hosts are the deduplication storage server, the load balancing servers, and the clients that deduplicate their own data.

If you have only a storage server and no load balancing servers or clients that deduplicate their own data, you do not have to open firewall ports.

<table>
<thead>
<tr>
<th>Port</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>10082</td>
<td>The NetBackup Deduplication Engine (<code>spoold</code>). Open this port between the hosts that deduplicate data.</td>
</tr>
<tr>
<td>10085</td>
<td>The deduplication database (<code>postgres</code>). The connection is internal to the storage server, from <code>spad</code> to <code>spoold</code>. You do not have to open this port.</td>
</tr>
<tr>
<td>10102</td>
<td>The NetBackup Deduplication Manager (<code>spad</code>). Open this port between the hosts that deduplicate data.</td>
</tr>
</tbody>
</table>

### About compression and encryption

NetBackup deduplication includes compression and encryption that are separate from and different than NetBackup policy-based compression and encryption.

Symantec recommends that you use compression and encryption during the deduplication process, not in the backup policy.

See “[Enabling compression or encryption](#)” on page 68.

(In a backup policy, compression is enabled when you select the **Compression** setting on the **Attributes** tab of the **Policy** dialog box. Similarly, encryption is
enabled when you select the **Encryption** setting. NetBackup compresses and encrypts the data before it reaches the PureDisk plug-in that deduplicates it. If you enable them in a backup policy, deduplication rates are low.)

The following is the behavior for the encryption that occurs during the deduplication process:

- If you enable encryption on the client, the data is encrypted before data transfer and remains encrypted on the storage. Data also is transferred from the client over a Secure Sockets Layer to the server regardless of whether or not the data is encrypted.
- If you enable encryption on a load balancing server, NetBackup encrypts the data at the load balancing server. It remains encrypted on storage.
- If you set encryption on the storage server, NetBackup encrypts the data at the storage server. It remains encrypted on storage.

Deduplication uses the Blowfish algorithm for encryption.

---

**About optimized synthetic backups and deduplication**

Optimized synthetic backups are a more efficient form of synthetic backup. A media server uses messages to instruct the storage server which full and incremental backup images to use to create the synthetic backup. The storage server constructs (or synthesizes) the backup image directly on the disk storage. Optimized synthetic backups require no data movement across the network.

Optimized synthetic backups are faster than a synthetic backup. Regular synthetic backups are constructed on the media server. They are moved across the network from the storage server to the media server and synthesized into one image. The synthetic image is then moved back to the storage server.

The target storage unit’s deduplication pool must be the same deduplication pool on which the source images reside.

In NetBackup, the **OptimizedImage** attribute enables optimized synthetic backup. It applies to both storage servers and deduplication pools. After you configure the storage server, you must set the **OptimizedImage** attribute on the storage server. If you have existing deduplication pools, you must set the attribute on them also.

See “Configuring optimized synthetic backups for deduplication” on page 70.

If NetBackup cannot produce the optimized synthetic backup, NetBackup creates the more data-movement intensive synthetic backup.
About optimized duplication of deduplicated data

Optimized duplication of deduplicated data copies the backup images from one pool to another pool more efficiently than normal duplication. Only the unique, deduplicated data segments are transferred. Optimized duplication reduces the amount of data that is transmitted over your network. The destination for optimized duplication must be part of the same NetBackup domain. That is, the source and the destination must use the same NetBackup master server.

Optimized duplication is a good method to copy your backup images off-site for disaster recovery.

See “About the deduplication storage destination” on page 19.

See “Configuring optimized duplication of deduplicated data” on page 71.

Optimized duplication copy requirements

The following are the requirements for optimized duplication:

- If the source images reside on a NetBackup Media Server Deduplication Pool, the destination can be another Media Server Deduplication Pool or a PureDisk Deduplication Pool. (In NetBackup, a PureDisk storage pool is configured as a PureDisk Deduplication Pool.)
  If the destination is a PureDisk storage pool, the PureDisk environment must be at release level 6.6 or later.

- If the source images reside on a PureDisk storage pool, the destination must be another PureDisk storage pool.
  Both PureDisk environments must be at release level 6.6 or later.

- The source storage and the destination storage must have at least one media server in common.
  See “About the media server in common” on page 32.

- In the storage unit you use for the destination for the optimized duplication, you must select only the common media server or media servers.
  If you select more than one, NetBackup assigns the duplication job to the least busy media server. If you select a media server or servers that are not common, the optimized duplication job fails.
  See “About the media server in common” on page 32.
  For more information about media server load balancing, see the NetBackup Administrator’s Guide for UNIX and Linux, Volume I or the NetBackup Administrator’s Guide for Windows, Volume I.

- The destination storage unit cannot be the same as the source storage unit.
Optimized duplication copy limitations

The following are limitations for optimized duplication copy:

- You cannot use optimized duplication from a PureDisk storage pool (a PureDisk Deduplication Pool) to a Media Server Deduplication Pool.

- If an optimized duplication job fails after the configured number of retries, NetBackup does not run the job again. By default, NetBackup retries an optimized duplication job three times. You can change the number of retries. See “Configuring optimized duplication copy behavior” on page 75.

- Optimized duplication does not work with storage unit groups. If you use a storage unit group as a destination for optimized duplication, NetBackup uses regular duplication.

- Optimized duplication does not support multiple copies during an optimized duplication job. If NetBackup is configured to make multiple new copies from the (source) copy of the backup image, the following occurs:
  - In a storage lifecycle policy, one duplication job creates one optimized duplication copy. If multiple optimized duplication destinations exist, a separate job exists for each destination. This behavior assumes that the device for the optimized duplication destination is compatible with the device on which the source image resides. If multiple remaining copies are configured to go to devices that are not optimized duplication capable, NetBackup uses normal duplication. One duplication job creates those multiple copies.
  - For other duplication methods, NetBackup uses normal duplication. One duplication job creates all of the copies simultaneously. The other duplication methods include the following: NetBackup Vault, the bpdeduplicate command line, and the duplication option of the Catalog utility in the NetBackup Administration Console.

About the media server in common

The source storage and the destination storage must have at least one media server in common. The common server initiates, monitors, and verifies the copy operation. The common server requires credentials for both the source storage and the destination storage. (For deduplication, the credentials are for the NetBackup Deduplication Engine, not for the host on which it runs.)

Usually, you would use one of the storage servers as the common server, but you can use another media server to control the optimized duplication. It must have the credentials and the connectivity for both the source storage and the destination storage.
storage. (You configure the credentials when you configure the storage server or you can add credentials to a media server later.)

Which server initiates the optimized duplication determines if the duplication is a push duplication or a pull duplication. Technically, no advantage exists with a push duplication or a pull duplication. However, the media server that initiates the duplication operation also becomes the write host for the new image copies.

**Push deduplication**

*Figure 2-3* shows a push configuration. Deduplication node A contains normal backups; deduplication node B is the destination for the optimized duplication copies. Host A has credentials for both nodes; it is the common server.

*Figure 2-3*  Push duplication

For *Figure 2-3*, the following screen shot shows the settings for the storage unit for the normal backups. The disk pool is the **Media Server Deduplication Pool** in the local environment. Because host A has credentials for both nodes, host A appears in the storage units for both of the nodes. For node A normal backups, you do not want a remote host deduplicating data, so only host A is selected.
For optimized duplication for Figure 2-3, the following screen shot shows the storage unit settings. The destination is the Media Server Deduplication Pool in the remote environment. You must select the common server, so only host A is selected.
If you use node B for backups also, select host B and not host A in the storage unit for the node B backups. If you select host A, it becomes a load balancing server for the node B deduplication pool.

Figure 2-4 shows a push deduplication from a Media Server Deduplication Pool to a PureDisk storage pool. (In NetBackup, a PureDisk storage pool is configured as a PureDisk Deduplication Pool.) The Media Server Deduplication Pool contains normal backups; the PureDisk Deduplication Pool is the destination for the optimized duplication copies. The local media server has credentials for both environments; it is the common server.
For Figure 2-4, the following screen shot shows the settings for the storage unit for the normal backups. The disk pool is the Media Server Deduplication Pool in the local environment. For normal backups, you do not want a remote host deduplicating data, so only the local host is selected.
For optimized duplication for Figure 2-4, the following screen shot shows the storage unit settings. The disk pool is the **PureDisk Deduplication Pool** in the remote environment. You must select the common server, so only the local media server is selected. If this configuration were a pull configuration, the remote host would be selected in the storage unit.
Figure 2-5 shows optimized duplication between two PureDisk storage pools. NetBackup media server A has credentials for both storage pools; it initiates, monitors, and verifies the optimized duplication. In the destination storage unit, the common server (media server A) is selected. This configuration is a push configuration. For a **PureDisk Deduplication Pool** (that is, a PureDisk storage pool), the PureDisk content router functions as the storage server.
You can use a load balancing server when you duplicate between two NetBackup deduplication pools. However, it is more common between two PureDisk storage pools.

**Pull deduplication**

Figure 2-6 shows a pull configuration. Deduplication node A contains normal backups; deduplication node B is the destination for the optimized duplication copies. Host B has credentials for both nodes; it is the common server.
For Figure 2-6, the storage unit settings for the normal backups are the same as for the push duplication example. For the duplication destination, the following screen shot shows the storage unit settings. They are similar to the push example except host B is selected. Host B is the common server, so it must be selected in the storage unit.
If you use node B for backups also, select host B and not host A in the storage unit for the node B backups. If you select host A, it becomes a load balancing server for the node B deduplication pool.

**About deduplication and SAN Client**

SAN Client is a NetBackup optional feature that provides high speed backups and restores of NetBackup clients. Fibre Transport is the name of the NetBackup high-speed data transport method that is part of the SAN Client feature. The backup and restore traffic occurs over a SAN.

SAN clients can be used with the deduplication option; however, the deduplication must occur on the media server, not the client. Configure the media server to be both a deduplication storage server (or load balancing server) and an FT media server. The SAN client backups are then sent over the SAN to the deduplication server/FT media server host. At that media server, the backup stream is deduplicated.

Do not enable client deduplication on SAN Clients. The data processing for deduplication is incompatible with the high-speed transport method of Fibre...
Transport. Client-side deduplication relies on two-way communication over the LAN with the media server. A SAN client streams the data to the FT media server at a high rate over the SAN.

### About deduplication performance

Many factors affect performance, especially the server hardware and the network capacity. Table 2-6 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 24.

<table>
<thead>
<tr>
<th>When</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial seeding</td>
<td>Initial seeding is when all clients are first backed up. Approximately 15 to 20 jobs can run concurrently under the following conditions:</td>
</tr>
<tr>
<td></td>
<td>■ The hardware meets minimum requirements. (More capable hardware improves performance.)</td>
</tr>
<tr>
<td></td>
<td>■ No compression. If data is compressed, the CPU usage increases quickly, which reduces the number of concurrent jobs that can be handled.</td>
</tr>
<tr>
<td></td>
<td>■ The deduplication rate is between 50% to 100%. The deduplication rate is the percentage of data already stored so it is not stored again.</td>
</tr>
<tr>
<td></td>
<td>■ The amount of data that is stored is less than 30% of the capacity of the storage.</td>
</tr>
<tr>
<td>Normal operation</td>
<td>Normal operation is when all clients have been backed up once. Approximately 15 to 20 jobs can run concurrently and with high performance</td>
</tr>
<tr>
<td></td>
<td>under the following conditions:</td>
</tr>
<tr>
<td></td>
<td>■ The hardware meets minimum requirements. (More capable hardware improves performance.)</td>
</tr>
<tr>
<td></td>
<td>■ No compression. If data is compressed, the CPU usage increases quickly, which reduces the number of concurrent jobs that can be handled.</td>
</tr>
<tr>
<td></td>
<td>■ The deduplication rate is between 10% and 50%. The deduplication rate is the percentage of data already stored so it is not stored again.</td>
</tr>
<tr>
<td></td>
<td>■ The amount of data that is stored is between 30% to 90% of the capacity of the storage.</td>
</tr>
</tbody>
</table>
Table 2-6  Deduplication job load performance for a deduplication storage server (continued)

<table>
<thead>
<tr>
<th>When</th>
<th>Description</th>
</tr>
</thead>
</table>
| 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.

About deduplication stream handlers

NetBackup provides the stream handlers that process various backup data stream types. Stream handlers improve backup deduplication rates by processing the underlying data stream.

The initial release of deduplication in NetBackup included stream handlers for NetBackup, Backup Exec, Hyper-V, VMWare vmdk, and vSphere backup types.

With the NetBackup 7.0.1 release, new stream handlers include one for NDMP and one for Windows System State.

Symantec continues to develop additional stream handlers for improved backup deduplication and performance.
About iSCSI, CIFS, and NFS

NetBackup supports storage area networks (SANs), direct-attached storage (DAS), or internal disks for the deduplication storage. NetBackup deduplication does not support iSCSI, CIFS, or NFS for the deduplication storage.

Deployment best practices

Symantec recommends that you consider the following practices when you implement NetBackup deduplication.

Use fully qualified domain names

Symantec recommends that you use fully qualified domain names for your NetBackup servers (and by extension, your deduplication servers). Fully qualified domain names can help to avoid host name resolution problems, especially if you use client-side deduplication.

Deduplication servers include the storage server and the load balancing servers (if any).

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.
- 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 42.

Send the initial full backups to the storage server

If you intend to use load balancing servers or client deduplication, use the storage server for the initial full backups of the clients. Then, send subsequent backups through the load balancing servers or use client deduplication for the backups.

Deduplication uses the same file fingerprint list regardless of which host performs the deduplication. So you can deduplicate data on the storage server first, and then subsequent backups by another host use the same fingerprint list. If the deduplication plug-in can identify the last full backup for the client and the policy combination, it retrieves the fingerprint list from the server. The list is placed in the fingerprint cache for the new backup.

See “About deduplication fingerprinting” on page 159.

Symantec also recommends that you implement load balancing servers and client deduplication gradually. Therefore, it may be beneficial to use the storage server for backups while you implement deduplication on other hosts.

See “Implement client deduplication gradually” on page 46.

See “Introduce load balancing servers gradually” on page 46.

Increase the number of jobs gradually

Symantec recommends that you increase the Maximum concurrent jobs value gradually. (The Maximum concurrent jobs is a storage unit setting.) 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 8 GB of memory and 4 GB of swap space is 50 concurrent jobs.

See “About deduplication performance” on page 42.
Introduce load balancing servers gradually

Symantec recommends that you add load balancing servers only after the storage server reaches maximum CPU utilization. Then, introduce load balancing servers one at a time. It may be easier to evaluate how your environment handles traffic and easier to troubleshoot any problems with fewer hosts added for deduplication.

See “About deduplication servers” on page 22.

Many factors affect deduplication server performance.

See “About deduplication performance” on page 42.

Because of the various factors, Symantec recommends that you maintain realistic expectations about using multiple servers for deduplication. If you add one media server as a load balancing server, overall throughput should be faster. However, adding one load balancing server may not double the overall throughput rate, adding two load balancing servers may not triple the throughput rate, and so on.

If all of the following apply to your deduplication environment, your environment may be a good candidate for load balancing servers:

■ The deduplication storage server is CPU limited on any core.
■ Memory resources are available on the storage server.
■ Network bandwidth is available on the storage server.
■ Back-end I/O bandwidth to the deduplication pool is available.
■ Other NetBackup media servers have CPU available for deduplication.

Gigabit Ethernet should provide sufficient performance in many environments. If your performance objective is the fastest throughput possible with load balancing servers, you should consider 10 Gigabit Ethernet.

Implement client deduplication gradually

If you configure clients to deduplicate their own data, do not enable all of those clients at the same time. Implement client deduplication gradually, as follows:

■ Use the storage server for the initial backup of the clients.
■ Enable deduplication on only a few clients at a time.
   It may be easier to evaluate how your environment handles traffic and easier to troubleshoot any problems with fewer hosts added for deduplication.

If a client cannot tolerate the deduplication processing workload, be prepared to move the deduplication processing back to the storage server.
Use deduplication compression or encryption

Do not use compression or encryption in a NetBackup policy; rather, use the compression or the encryption that is part of the deduplication process.

See “About compression and encryption” on page 29.

See “Enabling compression or encryption” on page 68.

In a backup policy, compression is enabled when you select the Compression setting on the Attributes tab of the Policy dialog box. Similarly, encryption is enabled when you select the Encryption setting. NetBackup compresses and encrypts the data before it reaches the PureDisk plug-in that deduplicates it. If you enable them in a backup policy, deduplication rates are low.

About the optimal number of backup streams

A backup stream appears as a separate job in the NetBackup Activity Monitor. Various methods exist to produce streams. In NetBackup, you can use backup policy settings to configure multiple streams. The NetBackup for Oracle agent lets you configure multiple streams; also for Oracle the RMAN utilities can provide multiple backup channels.

For client deduplication, the optimal number of backup streams is two.

Media server deduplication can process multiple streams on multiple cores simultaneously. For large datasets in applications such as Oracle, media server deduplication leverages multiple cores and multiple streams. Therefore, media server deduplication may be a better solution when the application can provide multiple streams or channels.

About storage unit groups for deduplication

You can use a storage unit group as a backup destination when the data is to be deduplicated. If you configure more than one deduplication destination, you can put the associated storage units in a storage unit group. Then, you can select the storage unit group as a destination for your backup jobs. This configuration avoids a single point of failure that can interrupt backup service.

The best storage savings occur when a backup policy stores its data in the same deduplication destination disk pool instead of across multiple disk pools. For this reason, the Failover method for the Storage unit selection provides the best deduplication performance. All of the other methods are designed to use different storage every time the backup executes.
Optimized duplication does not work with storage unit groups. If you use a storage unit group as a destination for optimized duplication, NetBackup uses regular duplication.

**About protecting the deduplicated data**

Symantec recommends the following methods to protect the deduplicated backup data and the deduplication database:

- Use NetBackup optimized duplication to copy the images to another deduplication node off-site location.
  
  Optimized duplication copies the primary backup data to another deduplication pool. It provides the easiest, most efficient method to copy 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 31.

  See “Configuring optimized duplication of deduplicated data” on page 71.

- For the primary deduplication storage, use a SAN volume with resilient storage methodologies to replicate the data to a remote site. If the deduplication database is on a different SAN volume, replicate that volume to the remote site also.

  Both of the preceding methods help to make your data highly-available.

  Also, you can use NetBackup to back up the deduplication storage server system or program disks. If the disk on which NetBackup resides fails and you have to replace it, you can use NetBackup to restore the media server.

**Save the storage server configuration**

Symantec recommends that you 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 99.

**How deduplication restores work**

The data is first reassembled on a media server before it is restored, even for the clients that deduplicate their own data. The media server that performs the restore always is a deduplication server (that is, hosts the PureDisk plug-in).

The backup server may not be the server that performs the restore. If another server has credentials for the NetBackup Deduplication Engine (that is, for the
storage server), NetBackup may use that server for the restore job. NetBackup chooses the least busy server for the restore.

The following other servers can have credentials for the NetBackup Deduplication Engine:

- A load balancing server in the same deduplication node.
- A deduplication server in a different deduplication node that is the target of optimized duplication.

  Optimized duplication requires a server in common between the two deduplication nodes.

  See “About the media server in common” on page 32.

You can specify the server to use for restores.

See “Specifying the restore server” on page 118.

---

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

The PureDisk Deduplication Option provides deduplication of NetBackup backups for NetBackup release 6.5. The destination storage for PDDO is a PureDisk storage pool. The PureDisk agent that performs the deduplication is installed from the PureDisk software distribution, not from the NetBackup distribution. PDDO is not the same as integrated NetBackup deduplication.

You can upgrade to 7.0 a NetBackup media server that hosts a PDDO agent and use that server for integrated NetBackup deduplication. The storage can remain the PureDisk storage pool, and NetBackup maintains access to all of the valid backup images in the PureDisk storage pool.

If you perform this procedure, the NetBackup PureDisk plug-in replaces the PureDisk agent on the media server. The NetBackup PureDisk plug-in can deduplicate data for either integrated NetBackup deduplication or for a PureDisk storage pool. The PDDO agent can deduplicate data only for a PureDisk storage pool.

---

**Note:** To use the NetBackup PureDisk plug-in with a PureDisk storage pool, the PureDisk storage pool must be part of a PureDisk 6.6 or later environment.
Table 2-7 Replacing a PDDO host with a media server deduplication host

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
</table>
| Deactivate all backup policies that use the host | Deactivate the policies to ensure that no activity occurs on 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 agent 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 61. |
| 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.
To migrate from PureDisk to NetBackup deduplication

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Install and configure NetBackup</td>
<td>See the <em>NetBackup Installation Guide for UNIX and Linux</em>. See the <em>NetBackup Installation Guide for Windows</em>.</td>
</tr>
<tr>
<td>Configure NetBackup deduplication</td>
<td>See “Configuring deduplication” on page 60.</td>
</tr>
<tr>
<td>Redirect your backup jobs</td>
<td>Redirect your backup jobs to the NetBackup media server deduplication pool.</td>
</tr>
<tr>
<td></td>
<td>See the <em>NetBackup Administrator's Guide for UNIX and Linux, Volume I</em>.*</td>
</tr>
<tr>
<td></td>
<td>See the <em>NetBackup Administrator's Guide for Windows, Volume I</em>.*</td>
</tr>
<tr>
<td>Uninstall PureDisk</td>
<td>After the PureDisk backup images expire, uninstall PureDisk.</td>
</tr>
<tr>
<td></td>
<td>See your NetBackup PureDisk documentation.</td>
</tr>
</tbody>
</table>

### 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>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configure NetBackup deduplication</td>
<td>See “Configuring deduplication” on page 60.</td>
</tr>
</tbody>
</table>
Table 2-9  Migrating to NetBackup deduplication (continued)

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redirect your backup jobs</td>
<td>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 <em>NetBackup Administrator's Guide for UNIX and Linux, Volume I</em>. See the <em>NetBackup Administrator's Guide for Windows, Volume I</em>.</td>
</tr>
<tr>
<td>Repurpose the storage</td>
<td>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.</td>
</tr>
</tbody>
</table>
Provisioning the storage

This chapter includes the following topics:

- About provisioning the storage
- About deduplication storage requirements
- About deduplication storage capacity
- About the deduplication storage paths
- About adding additional storage

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 destination” on page 19.

About deduplication storage requirements

The following defines the storage for the NetBackup Media Server Deduplication Option:

<table>
<thead>
<tr>
<th>Storage media</th>
<th>Disk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection</td>
<td>Storage area network (SAN), direct-attached storage (DAS), or internal disks</td>
</tr>
<tr>
<td>Minimum performance</td>
<td>130 MB/sec minimum read and write</td>
</tr>
</tbody>
</table>
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.

Local disk storage may leave you vulnerable in a disaster. SAN attached disk can be remounted at a newly provisioned server with the same name.

See “About the deduplication storage paths” on page 54.

About deduplication storage capacity

For NetBackup 7.0, the maximum deduplication capacity is 32 TBs. The deduplication database and the transaction logs consume some of that storage. NetBackup uses a 90 percent threshold for data to prevent storage overload when the deduplication database capacity is factored in. Although it is unlikely that 10 percent is required for database and transaction logs, it some situations 10 percent may be consumed.

For performance optimization, Symantec recommends that you use a separate disk, volume, partition, or spindle for the deduplication database. (If you use separate storage for the deduplication database, NetBackup still uses the 90 percent threshold to protect the data storage from any possible overload.)

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 storage pool as the storage destination.
  A PureDisk storage pool provides larger storage capacity; PureDisk 6.6 supports up to 100 TB of deduplicated data. It also provides global deduplication.
  See “About the deduplication storage destination” on page 19.

Only one deduplication storage path can exist on a media server. You cannot add another storage path to increase capacity beyond 32 TBs.

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.

---

**About adding additional storage**

The storage for a Media Server Deduplication Pool is exposed as a single disk volume. You cannot add another volume to an existing Media Server Deduplication Pool.

To increase the capacity of a Media Server Deduplication Pool, grow the existing volume.

See “About deduplication storage capacity” on page 54.

See “Resizing the storage partition” on page 117.
Provisioning the storage

About adding additional storage
Licensing deduplication

This chapter includes the following topics:

- About Licensing deduplication
- About the deduplication license key
- Licensing NetBackup deduplication

About Licensing 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 for deduplication 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. On the Help menu of the NetBackup Administration Console, select License Keys.
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.
Configuring deduplication

This chapter includes the following topics:

- Configuring deduplication
- Configuring a deduplication storage server
- About deduplication pools
- Configuring a deduplication pool
- Configuring a deduplication storage unit
- Enabling client deduplication
- Enabling compression or encryption
- Configuring backups
- Configuring optimized synthetic backups for deduplication
- Configuring optimized duplication of deduplicated data
- Configuring optimized duplication copy behavior
- Throttling optimized duplication traffic
- Adding a load balancing server
- About the deduplication pd.conf file
- Editing the deduplication pd.conf file
- Reconfiguring the deduplication storage server and storage paths
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

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Install the license key</td>
<td>See “Licensing NetBackup deduplication” on page 58.</td>
</tr>
<tr>
<td>Configure a deduplication storage server</td>
<td>See “About deduplication servers” on page 22.</td>
</tr>
<tr>
<td></td>
<td>See “Configuring a deduplication storage server” on page 61.</td>
</tr>
<tr>
<td>Configure a disk pool</td>
<td>See “About deduplication pools” on page 62.</td>
</tr>
<tr>
<td></td>
<td>See “Configuring a deduplication pool” on page 62.</td>
</tr>
<tr>
<td>Configure a storage unit</td>
<td>See “Configuring a deduplication storage unit” on page 64.</td>
</tr>
<tr>
<td>Optionally, enable client-side deduplication</td>
<td>See “Enabling client deduplication” on page 68.</td>
</tr>
<tr>
<td>Optionally, enable compression and encryption</td>
<td>See “Enabling compression or encryption” on page 68.</td>
</tr>
<tr>
<td>Configure a backup policy</td>
<td>Use the deduplication storage unit as the destination for the backup policy.</td>
</tr>
<tr>
<td></td>
<td>See the NetBackup Administrator's Guide for Windows, Volume I.</td>
</tr>
<tr>
<td></td>
<td>See the NetBackup Administrator's Guide for UNIX and Linux, Volume I.</td>
</tr>
<tr>
<td>Optionally, configure optimized synthetic backups</td>
<td>See “Configuring optimized synthetic backups for deduplication” on page 70.</td>
</tr>
</tbody>
</table>
Configuring a deduplication storage server

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

See “About deduplication servers” on page 22.

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 28.

- The storage paths.
  - See “About the deduplication storage paths” on page 54.

- The network interface.
  - See “About the network interface for deduplication” on page 28.

- The load balancing servers, if any.
  - See “About deduplication servers” on page 22.

When you create the storage server, the wizard lets you create a disk pool and storage unit also.
To configure a deduplication storage server by using the wizard

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 63.

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>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>The disk pool name.</td>
</tr>
<tr>
<td>Storage server</td>
<td>The storage server name. The storage server is the same as the NetBackup media server to which the storage is attached.</td>
</tr>
<tr>
<td>Disk volume</td>
<td>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.</td>
</tr>
<tr>
<td>Available space</td>
<td>The amount of space available in the disk pool.</td>
</tr>
<tr>
<td>Raw size</td>
<td>The total raw size of the storage in the disk pool.</td>
</tr>
<tr>
<td>Comment</td>
<td>A comment that is associated with the disk pool.</td>
</tr>
</tbody>
</table>
Table 5-2  Media server deduplication pool properties *(continued)*

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High water mark</strong></td>
<td>The high water mark is a threshold that indicates the <strong>PureDiskVolume</strong> is full. When the <strong>PureDiskVolume</strong> 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 <strong>PureDiskVolume</strong> does not contain enough storage for its estimated space requirement. NetBackup again assigns jobs to the storage unit when the capacity of the <strong>PureDiskVolume</strong> drops below the high water mark. Capacity is regained as backup images expire. NetBackup does not assign backup jobs to the disk pool if used space in the <strong>PureDiskVolume</strong> is greater than the high water mark. The default is 98%.</td>
</tr>
<tr>
<td><strong>Low water mark</strong></td>
<td>The low water mark has no affect on the <strong>PureDiskVolume</strong>.</td>
</tr>
<tr>
<td><strong>Limit I/O streams</strong></td>
<td>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. 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.</td>
</tr>
</tbody>
</table>

### 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.

   For a storage unit for optimized duplication destination, select **Only use the following media servers**. Then select the media servers that are common between the two deduplication nodes.

   See “**Deduplication storage unit properties**” on page 65.

   See “**Deduplication storage unit recommendations**” on page 66.

### Deduplication storage unit properties

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

<table>
<thead>
<tr>
<th><strong>Table 5-3</strong></th>
<th>Deduplication storage unit properties</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Property</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Storage unit name</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>Storage unit type</strong></td>
<td>Select <strong>Disk</strong> as the storage unit type.</td>
</tr>
<tr>
<td><strong>Disk type</strong></td>
<td>Select <strong>PureDisk</strong> for the disk type for a media server deduplication pool, a PureDisk deduplication pool, or a PureDisk Deduplication Option storage pool.</td>
</tr>
<tr>
<td><strong>Disk pool</strong></td>
<td>Select the disk pool that contains the storage for this storage unit. All disk pools of the specified <strong>Disk type</strong> appear in the <strong>Disk pool</strong> list. If no disk pools are configured, no disk pools appear in the list.</td>
</tr>
</tbody>
</table>
### Table 5-3

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media server</td>
<td>The Media server setting specifies the NetBackup media servers that can deduplicate data for this storage unit. Only the load balancing servers appear in the media server list.</td>
</tr>
<tr>
<td></td>
<td>Specify the media server or servers as follows:</td>
</tr>
<tr>
<td></td>
<td>■ To allow any server in the media server list to deduplicate data, select <strong>Use any available media server</strong>.</td>
</tr>
<tr>
<td></td>
<td>■ To use specific media servers to deduplicate the data, select <strong>Only use the following media servers</strong>. Then, select the media servers to allow.</td>
</tr>
<tr>
<td></td>
<td>NetBackup selects the media server to use when the policy runs.</td>
</tr>
<tr>
<td>Maximum fragment size</td>
<td>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.</td>
</tr>
<tr>
<td>Maximum concurrent jobs</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>NetBackup queues jobs until the storage unit is available. If three backup jobs are scheduled and <strong>Maximum concurrent jobs</strong> 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.</td>
</tr>
<tr>
<td></td>
<td><strong>Maximum concurrent jobs</strong> 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 <strong>Maximum concurrent jobs</strong>, only one job runs at a time.</td>
</tr>
<tr>
<td></td>
<td>The number to enter depends on the available disk space and the server’s ability to run multiple backup processes.</td>
</tr>
<tr>
<td></td>
<td><strong>Warning</strong>: A Maximum concurrent jobs setting of 0 disables the storage unit.</td>
</tr>
</tbody>
</table>

---

### Deduplication storage unit recommendations

You can use storage unit properties to control how NetBackup performs.
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 \textbf{Maximum concurrent jobs} setting than the other (STU-SILVER). More client backups occur for the storage unit with the higher \textbf{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 \texttt{NetBackup Administration Console}, expand \texttt{NetBackup Management > Host Properties > Master Servers}.
2. In the details pane, select the master server.
3. On the \texttt{Actions} menu, select \texttt{Properties}.
4. On the \texttt{Host Properties General} tab, add the clients that use client direct to the \texttt{Clients} list.
5. Select one of the following \textbf{Deduplication Location} options:
   - \textbf{Always use the media server} disables client deduplication. By default, all clients are configured with the \texttt{Always use the media server} option.
   - \textbf{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.
   - \textbf{Always use client-side deduplication} uses client deduplication. If the deduplication backup job fails, NetBackup retries the job.

You can override the \textbf{Prefer to use client-side deduplication} or \textbf{Always use client-side deduplication} host property in the backup policies.

See \texttt{Disable client-side deduplication} in the \textit{NetBackup Administrator's Guide for UNIX and Linux, Volume I}.

See \texttt{Disable client-side deduplication} in the \textit{NetBackup Administrator's Guide for Windows, Volume I}.

**Enabling compression or encryption**

Two procedures exist to enable compression and encryption during deduplication, as follows:

- You can enable compression or encryption on individual hosts.
Use this procedure to enable compression or encryption on the storage server, on load balancing servers, or on the clients that deduplicate their own data. See “To enable compression or encryption on a single host” on page 69.

- You can enable compression or encryption on all clients that deduplicate their own data without configuring them individually. Use this procedure if you want all of your clients that deduplicate their own data to compress and encrypt that data. See “To enable compression or encryption on all hosts” on page 70.

See “About compression and encryption” on page 29.

### To enable compression or encryption on a single host

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

   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 `pd.conf` file” on page 78.

2. For each setting you want to change, remove the pound character (#) in column 1 from that line.

   The file contains separate lines for `COMPRESSION` and `ENCRYPTION`.

3. For each setting you want to change, replace the 0 (zero) with a 1.

   **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 80.

4. Save and close the file.

5. If the host is the storage server or a load balancing server, restart the NetBackup Remote Manager and Monitor Service (`nbrmms`) on the host.
To enable compression or encryption on all hosts

1. On the storage server, open the `contentrouter.cfg` file in a text editor; it resides in the following directory:

   ```
   storage_path/etc/puredisk/contentrouter.cfg
   ```

2. To enable compression, add `compress` to the `ServerOptions` line of the file. To enable encryption, add `agent_crypt` to that line. For example, the following line enables both compression and encryption.

   ```
   ServerOptions=fast,verify_data_read,compress,agent_crypt
   ```

3. If you use load balancing servers, make the same edits to the `contentrouter.cfg` files on those hosts.

### 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 synthetic backups for deduplication

The following table shows the procedures that are required to configure optimized synthetic backups for a deduplication environment.

See “About optimized synthetic backups and deduplication” on page 30.

<table>
<thead>
<tr>
<th>Table 5-4</th>
<th>To configure optimized synthetic backups for deduplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>Procedure</td>
</tr>
<tr>
<td>Set the OptimizedImage attribute on the storage server.</td>
<td>See “Setting deduplication storage server attributes” on page 95.</td>
</tr>
</tbody>
</table>
To configure optimized synthetic backups for deduplication

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set the OptimizedImage attribute on your existing deduplication pools. (Any deduplication pools that you create after you set the storage server attribute inherit the new functionality.)</td>
<td>See “Setting deduplication pool attributes” on page 108.</td>
</tr>
</tbody>
</table>
| Configure a Standard or MS-Windows backup policy. Select the Synthetic backup attribute on the Schedule Attribute tab of the backup policy. | See the administrator's guide for your operating system:  

Configuring optimized duplication of deduplicated data

You can configure optimized duplication of deduplicated backups. Before you begin, review the requirements.

See “About optimized duplication of deduplicated data” on page 31.

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
</table>
| Configure the storage servers | One server must be common between the source storage and the destination storage. Which you choose depends on whether you want a push or a pull configuration.  
See “About the media server in common” on page 32.  
For a push configuration, configure the common server as a load balancing server for the storage server for your normal backups. For a pull configuration, configure the common server as a load balancing server for the storage server for the copies at your remote site. Alternatively, you can add a server later to either environment. (A server becomes a load balancing server when you select it in the storage unit for the deduplication pool.)  
See “Optimized duplication copy requirements” on page 31.  
See “Configuring a deduplication storage server” on page 61. |
### Table 5-5

To configure optimized duplication of deduplicated data *(continued)*

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
</table>
| Configure the deduplication pools | If you did not configure the deduplication pools when you configured the storage servers, use the Disk Pool Configuration Wizard to configure them.  
See “Configuring a deduplication pool” on page 62. |
| Configure the backup storage unit | In the storage unit for your backups, do the following:  
- For the Disk type, select **PureDisk**.  
- For the Disk pool, select one of the following:  
  - If you back up to integrated NetBackup deduplication, select your Media Server Deduplication Pool.  
  - If you back up to a PureDisk environment, select the PureDisk Deduplication Pool.  
If you use a pull configuration, do not select the common media server in the backup storage unit. If you do, NetBackup uses it to deduplicate backup data. (That is, unless you want to use it for a load balancing server for the source deduplication node.)  
See “Configuring a deduplication storage unit” on page 64. |
Symantec recommends that you configure a storage unit specifically to be the target for the optimized duplication. Configure the storage unit in the deduplication environment that performs your normal backups. Do not configure it in the environment that contains the copies.

In the storage unit that is the destination for your duplicated images, do the following:

- For the **Disk type**, select **PureDisk**.
- For the **Disk pool**, the destination can be a **Media Server Deduplication Pool** or a **PureDisk Deduplication Pool**.

  **Note:** If the backup destination is a **PureDisk Deduplication Pool**, the duplication destination also must be a **PureDisk Deduplication Pool**.

Also 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. If you select more than one, NetBackup assigns the duplication job to the least busy media server.

If you select only a media server (or servers) that is not common, the optimized duplication job fails.

See “Configuring a deduplication storage unit” on page 64.

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
</table>
| Configure the duplication storage unit       | Symantec recommends that you configure a storage unit specifically to be the target for the optimized duplication. Configure the storage unit in the deduplication environment that performs your normal backups. Do not configure it in the environment that contains the copies. In the storage unit that is the destination for your duplicated images, do the following:
- For the **Disk type**, select **PureDisk**.
- For the **Disk pool**, the destination can be a **Media Server Deduplication Pool** or a **PureDisk Deduplication Pool**.
  **Note:** If the backup destination is a **PureDisk Deduplication Pool**, the duplication destination also must be a **PureDisk Deduplication Pool**.
Also 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. If you select more than one, NetBackup assigns the duplication job to the least busy media server.
If you select only a media server (or servers) that is not common, the optimized duplication job fails. See “Configuring a deduplication storage unit” on page 64. |
| Configure optimized duplication behaviors     | See “Configuring optimized duplication copy behavior” on page 75.  
See “Throttling optimized duplication traffic” on page 77. |
To configure optimized duplication of deduplicated data  *(continued)*

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
</table>
| Configure a storage lifecycle policy for the duplication | Configure a storage lifecycle policy only if you want to use one to duplicate images. The storage lifecycle policy manages both the backup jobs and the duplication jobs. Configure the lifecycle policy in the deduplication environment that performs your normal backups. Do not configure it in the environment that contains the copies. When you configure the storage lifecycle policy, do the following:  
  ■ For the **Backup** destination, select the storage unit that is the target of your backups. That storage unit may use a **Media Server Deduplication Pool** or a **PureDisk Deduplication Pool**. These backups are the primary backup copies; they are the source images for the duplication operation.  
  ■ For the **Duplication** destination, select the storage unit for the destination deduplication pool. That pool may be a **Media Server Deduplication Pool** or a **PureDisk Deduplication Pool**. If the backup destination is a **PureDisk Deduplication Pool**, the duplication destination also must be a **PureDisk Deduplication Pool**. See the *NetBackup Administrator’s Guide for UNIX and Linux* or the *NetBackup Administrator’s Guide for Windows*. |
| Configure a backup policy | Configure a policy to back up your clients. Configure the backup policy in the deduplication environment that performs your normal backups. Do not configure it in the environment that contains the copies.  
  ■ If you use a storage lifecycle policy to manage the backup job and the duplication job: Select that storage lifecycle policy in the **Policy storage** field of the Policy **Attributes** tab.  
  ■ If you do not use a storage lifecycle policy to manage the backup job and the duplication job: Select the storage unit for the **Media Server Deduplication Pool** that contains your normal backups. These backups are the primary backup copies. |
To configure optimized duplication of deduplicated data (continued)

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configure NetBackup Vault for the duplication</td>
<td>Configure Vault duplication only if you use NetBackup Vault to duplicate the images. Configure Vault in the deduplication environment that performs your normal backups. Do not configure it in the environment that contains the copies. For Vault, you must configure a Vault profile and a Vault policy.</td>
</tr>
<tr>
<td>■ Configure a Vault profile.</td>
<td></td>
</tr>
<tr>
<td>■ On the Vault Profile dialog box Choose Backups tab, choose the backup images in the source Media Server Deduplication Pool.</td>
<td></td>
</tr>
<tr>
<td>■ On the Profile dialog box Duplication tab, select the destination storage unit in the Destination Storage unit field.</td>
<td></td>
</tr>
<tr>
<td>■ Configure a Vault policy to schedule the duplication jobs. A Vault policy is a NetBackup policy that is configured to run Vault jobs.</td>
<td>See the NetBackup Vault Administrator’s Guide.</td>
</tr>
<tr>
<td>Duplicate by using the bpduplicate command</td>
<td>Use the NetBackup bpduplicate command to copy images manually. If you use a storage lifecycle policy or NetBackup Vault for optimized duplication, you do not have to use the bpduplicate command. Duplicate from the source storage to the destination storage. The destination storage may be a Media Server Deduplication Pool or a PureDisk Deduplication Pool.</td>
</tr>
<tr>
<td></td>
<td>See NetBackup Commands.</td>
</tr>
</tbody>
</table>

Configuring optimized duplication copy behavior

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

■ Optimized duplication failover. The default NetBackup behavior is to use optimized duplication between the source and the target storage systems that support optimized duplication such as between two deduplication media servers. If an optimized duplication job fails, NetBackup does not run the job again. You can configure NetBackup to use normal duplication if optimized duplication fails. For example, NetBackup does not support optimized duplication from PureDisk Storage Pool Authority source to a Media Server Deduplication Pool. Both entities support optimized duplication, however not in this direction. Therefore, to pull or migrate data out of a PureDisk SPA into a Media Server Deduplication Pool, you must change the default NetBackup failover behavior for optimized duplication.
Caution: This setting affects all optimized duplication jobs; it is not limited to optimized duplication to a Media Server Deduplication Pool or a PureDisk Deduplication Pool.

- Number of optimized duplication attempts. The default is three attempts. You can specify the number of times NetBackup retries an optimized deduplication job before it fails the jobs.

Caution: This setting affects all optimized duplication jobs; it is not limited to optimized duplication to a Media Server Deduplication Pool or a PureDisk Deduplication Pool.

- Storage lifecycle policy retry wait period. If the optimized deduplication job is configured in a storage lifecycle policy and the job fails, NetBackup retries the job three times. If the job is unsuccessful after three tries, NetBackup waits two hours and then retries the job. You can change the wait period.

Caution: This setting affects all storage lifecycle policy jobs. It is not limited to optimized duplication to a Media Server Deduplication Pool or a PureDisk Deduplication Pool in a storage lifecycle policy.

To configure optimized duplication failover

- On the master server, add the following configuration option:

```plaintext
RESUME_ORIG_DUP_ON_OPT_DUP_FAIL = TRUE
```

See “Setting NetBackup configuration options by using bpsetconfig” on page 116.

Alternatively on UNIX systems, add the entry to the `bp.conf` file on the NetBackup master server.

To configure the number of duplication attempts

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

```plaintext
OPT_DUP_BUSY_RETRY_LIMIT 4
```

The `behavior` file resides in the following directories:

- UNIX: `/usr/openv/netbackup/db/config`
To configure the storage lifecycle policy wait period

- Change the wait period for retries by adding an `IMAGE_EXTENDED_RETRY_PERIOD_IN_HOURS` entry to the NetBackup `LIFECYCLE_PARAMETERS` file. The default for this value is two hours. For example, the following entry configures NetBackup to wait four hours before NetBackup tries the job again:

```
IMAGE_EXTENDED_RETRY_PERIOD_IN_HOURS 4
```

The `LIFECYCLE_PARAMETERS` file resides in the following directories:

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

### Throttling optimized duplication traffic

You can control how much network bandwidth that optimized duplication consumes. The `pd.conf` file `OPTDUP_BANDWIDTH` parameter specifies the amount of data to allow.

To specify the amount of bandwidth, edit the `pd.conf` file on the source storage server or the source PureDisk storage pool. Optimized duplication traffic to any deduplication destination is throttled.

See “About the deduplication `pd.conf` file” on page 78.

### 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 22.

**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.

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 pd.conf file

On each host that deduplicates data, a *pd.conf* file contains various configuration settings that control the operation of deduplication for the host. 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:
Editing the deduplication pd.conf file

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 pd.conf file” on page 78.

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 pd.conf file” on page 78.

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 80.

4. Save and close the file.

5. Restart the NetBackup Remote Manager and Monitor Service (nbrmms) on the host.
### pd.conf file settings

The following table describes the deduplication settings that you can configure. The parameters in this table are in alphabetical order; the parameters in a pd.conf file may not be in alphabetical order.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Default value</th>
<th>Possible values</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BANDWIDTH_LIMIT</strong></td>
<td>0</td>
<td>0 (default - no limit) to the practical system limit</td>
<td>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.</td>
</tr>
<tr>
<td><strong>COMPRESSION</strong></td>
<td>0</td>
<td>0 (off) or 1 (on)</td>
<td>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 29.</td>
</tr>
<tr>
<td><strong>DEBUGLOG</strong></td>
<td>C:\pdplugin.log (Windows)</td>
<td>Any path</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>tmp/pdplugin.log (UNIX)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **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. |
<p>| <strong>ENCRIPTION</strong>        | 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 29. |</p>
<table>
<thead>
<tr>
<th>Setting</th>
<th>Default value</th>
<th>Possible values</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOGLEVEL</td>
<td>0</td>
<td>0 through 10</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Note:</strong> Do not change this setting unless directed to by a Symantec Technical Support representative.</td>
</tr>
<tr>
<td>MATCH_PDRO</td>
<td>1</td>
<td>1 (on) or 0 (off)</td>
<td>Specifies that NetBackup should use the PureDisk Remote Office Agent dynamic segmentation deduplication algorithm. Enabling it means</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>that the data that is backed up is globally deduplicated with each other. If disabled (set to 0), PDDO uses a static 128KB segmentation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>algorithm for deduplication. <strong>Caution:</strong> Changing the default segmentation algorithm most likely causes the next set of backups to</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>not deduplicate with the existing storage pool data. Subsequent backups then deduplicate using the PDRO algorithm.</td>
</tr>
<tr>
<td>MAX_IMG_MBSIZE</td>
<td>50,000</td>
<td>0 to 50,000</td>
<td>This keyword is reserved for internal use.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Note:</strong> Do not change this setting unless directed to by a Symantec representative.</td>
</tr>
<tr>
<td>MINFILE_KSIZE</td>
<td>16</td>
<td>From 1 to the practical system limit</td>
<td>Determines the smallest size file (in KBs) that NetBackup segments. Files smaller than the threshold are combined into a single large</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>segment to reduce the overhead of managing many small segments. However, such segments have less chance of being deduplicated.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Caution:</strong> Large numbers of files smaller than the threshold may adversely affect backup performance.</td>
</tr>
<tr>
<td>OPTDUP_TIMEOUT</td>
<td>N/A</td>
<td>The value expressed in minutes</td>
<td>Specifies the number of minutes before the optimized duplication times out. Indicated in minutes.</td>
</tr>
<tr>
<td>Setting</td>
<td>Default value</td>
<td>Possible values</td>
<td>Action</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>PDALIGN</td>
<td></td>
<td></td>
<td>This keyword is reserved for internal use.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Note:</strong> Do not change this setting unless directed to by a Symantec Technical Support representative.</td>
</tr>
<tr>
<td>PREFETCH</td>
<td></td>
<td></td>
<td>This keyword is reserved for internal use.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Note:</strong> Do not change this setting unless directed to by a Symantec Technical Support representative.</td>
</tr>
<tr>
<td>OPTDUP_BANDWITH</td>
<td>0</td>
<td>0 (no limit) to the practical system limit</td>
<td>Determines the maximum bandwidth that is allowed for optimized duplication. The value is specified in KBytes/second.</td>
</tr>
<tr>
<td>OPTDUP_ENCRYPTION</td>
<td>1</td>
<td>1 (on) or 0 (off)</td>
<td>Determines if the data to replicate is encrypted before it is sent out over the network.</td>
</tr>
<tr>
<td>SEGKSIZE</td>
<td>N/A</td>
<td>N/A</td>
<td>This keyword is reserved for internal use.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Warning:</strong> Changing this value can reduce capacity and decrease performance. Modify this setting only if directed to do so by Symantec support representative.</td>
</tr>
</tbody>
</table>

### Reconfiguring the deduplication storage server and storage paths

An activated NetBackup Deduplication Engine rejects all configuration attempts that abandon its previously defined storage and backup images. Therefore, if you made a mistake during configuration, you must manually deactivate the engine and physically delete the storage directory before you can reconfigure deduplication. Similarly, if you want to change the storage path, you must deactivate the engine and delete the storage directory.

Common configuration mistakes include the following:

- Choosing a media server with an unsupported operating system. All NetBackup media servers appear in the **Storage Server Configuration Wizard**.

- Choosing a storage path that does not include a directory name. For example, `D:\DedupeData` is a valid storage path, but `D:` is not.
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 5-7  Reconfigure media server deduplication

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
</table>
| 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 111.                           |
| Stop the NetBackup 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 directory and the database directory | Delete the storage directory and database directory (if you configured a database directory). |
| Reset the deduplication registry           | See “Resetting the deduplication registry” on page 104.                   |
Table 5-7  Reconfigure media server deduplication (continued)

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start the NetBackup services on the media server</td>
<td>See the <em>NetBackup Administrator's Guide for UNIX and Linux, Volume I</em>.</td>
</tr>
<tr>
<td></td>
<td>See the <em>NetBackup Administrator's Guide for Windows, Volume I</em>.</td>
</tr>
<tr>
<td>Delete the deduplication storage server</td>
<td>See “Deleting a deduplication storage server” on page 97.</td>
</tr>
<tr>
<td>Delete the storage server configuration file</td>
<td>The storage server and every load balancing server contain a deduplication configuration file. Delete that file from every server that you use for deduplication.</td>
</tr>
<tr>
<td></td>
<td>See “Deleting a deduplication host configuration file” on page 101.</td>
</tr>
<tr>
<td>Reconfigure</td>
<td>See “Configuring deduplication” on page 60.</td>
</tr>
</tbody>
</table>
Monitoring deduplication activity

This chapter includes the following topics:

■ Monitoring the deduplication rate
■ Monitoring deduplication storage capacity and usage
■ Viewing disk reports
■ Monitoring deduplication processes

Monitoring the deduplication rate

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-side 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 a client for which Server_A deduplicated the data (the dedup field shows the deduplication rate):
10/6/2010 10:02:09 AM - Info Server_A(pid=30695)
StorageServer=PureDisk:Server_A; Report=PDDO Stats for (Server_A):  scanned: 30126998 KB, stream rate: 162.54 MB/sec, CR sent: 1720293 KB, dedup: 94.3%, cache hits: 214717 (94.0%)

The other fields that show deduplication information are highlighted in the example. For the field descriptions, see Table 6-1.

- For client-side 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 disk image header and True Image Restore information (if applicable). That information is always deduplicated on a server; typically, deduplication rates for that information are zero or very low. The following job details example excerpt shows the two rates:

10/8/2010 11:54:21 PM - Info Server_A(pid=2220)
Using OpenStorage client direct to backup from client
Client_B to Server_A

10/8/2009 11:58:09 PM - Info Server_A(pid=2220)
StorageServer=PureDisk:Server_A; Report=PDDO Stats for (Server_A): scanned: 3423425 KB, stream rate: 200.77 MB/sec, CR sent: 122280 KB, dedup: 96.4%, cache hits: 49672 (98.2%)

10/8/2010 11:58:09 PM - Info Server_A(pid=2220) Using
the media server to write NBU data for backup
Client_B_1254987197 to Server_A

10/8/2010 11:58:19 PM - Info Server_A(pid=2220)
StorageServer=PureDisk:Server_A; Report=PDDO Stats for (Server_A): 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 I.

Many factors affect deduplication performance.
See “About deduplication performance” on page 42.
See “About deduplication server requirements” on page 24.
See “About client deduplication host requirements” on page 27.
Table 6-1  Deduplication activity field descriptions

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cache hits</td>
<td>The percentage of time that the local fingerprint cache contained a record of the segment. The deduplication plug-in did not have to query the database about the segment.</td>
</tr>
</tbody>
</table>
| CR sent  | The amount of data that is sent from the deduplication plug-in to the component that stores the data. (In NetBackup, the NetBackup Deduplication Engine stores the data. In PureDisk, a content router stores the data.)  
If the storage server deduplicates the data, it does not travel over the network. The deduplicated data travels over the network when the deduplication plug-in runs on a computer other than the storage server, as follows:  
  - On a NetBackup client that deduplicates its own data (client-side deduplication).  
  - On a fingerprinting media server that deduplicates the data. The plug-in on the fingerprinting server sends the data to the storage server, which writes it to a **Media Server Deduplication Pool**.  
  - On a media server that then sends it to a PureDisk environment for storage. (In NetBackup, a **PureDisk Storage Pool** represents the storage of a PureDisk environment.) |
| dedup    | The percentage of data that was stored already. That data is not stored again.                                                                                                                                                                                                 |
| scanned  | The amount of data that the deduplication plug-in scanned.                                                                                                                                                                                                                       |
| stream rate | The speed of the scan: The kilobytes of data that are scanned divided by how long the scan takes.                                                                                                                                                                                  |

**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 90.

**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 6-2 describes the options for monitoring capacity and usage.

**Table 6-2** Capacity and usage reporting

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Change Storage Server</strong></td>
<td>The Change Storage Server dialog box Properties tab displays storage capacity and usage. This dialog box displays the most current capacity usage that is available in the NetBackup Administration Console. See “Changing deduplication storage server properties” on page 94.</td>
</tr>
<tr>
<td><strong>Disk Pools</strong></td>
<td>The Disk Pools window of the Administration Console displays a value that was stored when NetBackup polled the disk pools. NetBackup polls every 5 minutes; therefore, 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 &gt; Devices &gt; Disk Pools.</td>
</tr>
<tr>
<td><strong>Disk Pool Status</strong></td>
<td>The Disk Pool Status report displays the state of the disk pool and usage information. See “Viewing disk reports” on page 90.</td>
</tr>
</tbody>
</table>
### Table 6-2  Capacity and usage reporting (continued)

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Disk Logs report** | The **Disk Logs** report displays event and message information. A useful event for monitoring capacity is event 1044; the following is the description of the even in the **Disk Logs** report:  
The usage of one or more system resources has exceeded a warning level.  
The threshold for this message is at 90% capacity. No more data can be stored.  
See “Viewing disk reports” on page 90.  
See “Deduplication event codes and messages” on page 136. |
| **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.  
On the **Help** menu in the **NetBackup Administration Console**, select **License Keys**. |
| 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 90. |
| **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 111. |
| 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, use the `--help` option.

**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 6-3 describes the disk reports available.
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.

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.

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.

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.
Monitoring deduplication processes

The following table shows the deduplication processes about which NetBackup reports:

See “Deduplication server components” on page 151.

**Table 6-4** Where to monitor the main deduplication processes

<table>
<thead>
<tr>
<th>What</th>
<th>Where to monitor it</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetBackup Deduplication Engine</td>
<td>On Windows systems, in the NetBackup Administration Console <strong>Activity Monitor Services</strong> tab.</td>
</tr>
<tr>
<td></td>
<td>On UNIX, the NetBackup Deduplication Engine appears as <code>spoold</code> in the Administration Console <strong>Activity Monitor Daemons</strong> tab.</td>
</tr>
<tr>
<td></td>
<td>The NetBackup <code>bpps</code> command shows the <code>spoold</code> process.</td>
</tr>
<tr>
<td>NetBackup Deduplication Manager</td>
<td>On Windows systems, NetBackup Deduplication Manager in the <strong>Activity Monitor Services</strong> tab.</td>
</tr>
<tr>
<td></td>
<td>On UNIX, the NetBackup Deduplication Manager appears as <code>spad</code> in the Administration Console <strong>Activity Monitor Daemons</strong> tab.</td>
</tr>
<tr>
<td></td>
<td>The NetBackup <code>bpps</code> command shows the <code>spad</code> process.</td>
</tr>
<tr>
<td>The database processes (<code>postgres</code>)</td>
<td>On Windows systems, the <code>postgres</code> database processes appear in the <strong>Activity Monitor Processes</strong> tab.</td>
</tr>
<tr>
<td></td>
<td>The NetBackup <code>bpps</code> command shows the <code>postgres</code> processes.</td>
</tr>
</tbody>
</table>
This chapter includes the following topics:

- Managing deduplication servers
- Managing NetBackup Deduplication Engine credentials
- Managing deduplication pools
- Deleting backup images
- Disabling deduplication for a client
- About maintenance processing
- Performing maintenance manually
- Setting NetBackup configuration options by using bpsetconfig
- Resizing the storage partition
- Restoring files at a remote site
- Specifying the restore server

Managing deduplication servers

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

See “Changing deduplication storage server properties” on page 94.
See “Setting deduplication storage server attributes” on page 95.
See “Clearing deduplication storage server attributes” on page 96.
See “Deleting a deduplication storage server” on page 97.
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.
In the Change Storage Server dialog box, select the Properties tab.

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

Change the value.

Click OK.

Setting deduplication storage server attributes

You may have to set storage server attributes to enable new functionality. For example, to support optimized synthetic backups, you must set an attribute.

If you set an attribute on the storage server, you may have to set the same attribute on existing deduplication pools. The overview or configuration procedure for the new functionality describes the requirements.

See “Setting deduplication pool attributes” on page 108.

To set a storage server attribute

◆ The following is the command syntax to set a storage server attribute. Run the command on the master server or on the storage server.
nbdevconfig -changests -storage_server storage_server -stype PureDisk -setattribute attribute -media_server media_server

The following describe the options that require your input:

- **storage_server**
  The name of the storage server.

- **setattribute**
  The *attribute* is the name of the argument that represents the new functionality.
  For example, *OptimizedImage* specifies that the environment supports the optimized synthetic backup method.

- **media_server**
  A NetBackup media server that connects to the storage server. The media server queries the storage server for its capabilities. Because the storage server is also a media server, use the storage server name.

The following is the path to the `nbdevconfig` command:

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

### Clearing deduplication storage server attributes

Use the `nbdevconfig` command to remove storage server attributes.

**To clear deduplication storage server attributes**

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

```bash
nbdevconfig -changests -storage_server storage_server -stype PureDisk -clearattribute attribute
```

- **storage_server**
  The name of the storage server.

- **setattribute**
  The *attribute* is the name of the argument that represents the functionality.

The following is the path to the `nbdevconfig` command:

- UNIX: `/usr/openv/netbackup/bin/admincmd`
- Windows: `install_path\NetBackup\bin\admincmd`
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.
About the deduplication storage server configuration file

The storage server configuration file is a file that contains the configuration settings for your storage server. The file does not exist unless you create. The file may help you with recovery of your storage server. Therefore, Symantec recommends that you get and save the storage server configuration.

If you create the file after you configure your environment, it contains the configuration settings for your storage server. The following is an example of a populated configuration file:

```
V6.5.5 "storagepath" "F:\DedupeData" string
V6.5.5 "spalogpath" "F:\DedupeData\log" string
V6.5.5 "dbpath" "E:\DedupeDB" string
V6.5.5 "required_interface" "DEDUPESERVER" string
V6.5.5 "spalogretention" "7" int
V6.5.5 "verboselevel" "3" int
V6.5.5 "Storage Pool Size" "235.8GB" string
V6.5.5 "Storage Pool Used Space" "3.7GB" string
V6.5.5 "Storage Pool Available Space" "232.0GB" string
V6.5.5 "Catalog Logical Size" "129Bytes" string
V6.5.5 "Catalog files Count" "2" string
V6.5.5 "Space Used Within Containers" "171Bytes" string
```

V6.5.5 represents the version of the I/O format not the NetBackup release level. The version may differ on your system.

Normally, the file should be used as reference only.

If you create the file when a storage server is not configured or is down and unavailable, NetBackup creates a template file. For example, for disaster recovery you may need to edit a template file to create a configuration file to recover your storage server. 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
```

See “Save the storage server configuration” on page 48.

See “Getting the storage server configuration” on page 99.

See “Editing a storage server configuration file” on page 99.
See “Setting the storage server configuration” on page 100.

Getting the storage server configuration

Symantec recommends that you get and save the storage server configuration file.

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

See “About the deduplication storage server configuration file” on page 98.

See “Save the storage server configuration” on page 48.

See “Recovering from a deduplication storage server disk failure” on page 143.

See “Recovering from a permanent deduplication storage server failure” on page 145.

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 that indicates its purpose.

Editing a storage server configuration file

In some very limited situations, you may need to edit a storage server configuration file so that it 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 to recover the storage server.

See “About the deduplication storage server configuration file” on page 98.

See “Recovering from a deduplication storage server disk failure” on page 143.

See “Recovering from a permanent deduplication storage server failure” on page 145.
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 99.
   
   If you get the configuration of a storage server that is unavailable because of a disaster, NetBackup returns a template configuration file.

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`.
   - `spallogin`.
   - `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 a file of your configuration that you saved. Alternatively, the file may be an edited configuration file.

See “About the deduplication storage server configuration file” on page 98.

See “Save the storage server configuration” on page 48.

See “Recovering from a deduplication storage server disk failure” on page 143.
See “Recovering from a permanent deduplication storage server failure” on page 145.

See “Editing a storage server configuration file” on page 99.

---

**Note:** The only time you should use the `nbdevconfig` command with the `-setconfig` option is to recover from a storage server failure.

---

**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.

---

**About the deduplication host configuration file**

Each NetBackup host that is used for deduplication has a configuration file; the file name matches the name of the storage server, as follows:

```
storage_server_name.cfg
```

The `storage_server_name` is the fully qualified domain name if that was used to configure the storage server. For example, if the storage server name is `DedupeServer.symantecs.org`, the configuration file name is `DedupeServer.symantecs.org.cfg`.

The following is the location of the file:

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

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

---

**Deleting a deduplication host configuration file**

You may need to delete the configuration file from the deduplication hosts. For example, to reconfigure your deduplication environment or disaster recovery may require that you delete the configuration file on the servers on which it exists.

See “About the deduplication host configuration file” on page 101.

See “Reconfiguring the deduplication storage server and storage paths” on page 82.
See “Recovering from a deduplication storage server disk failure” on page 143.

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 22.

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.

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

- The following is the command syntax to set a storage server attribute. Run the command on the NetBackup master server or on the 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 : OptimizedImage
  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 82.

**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 commands on the storage server to reset the deduplication registry file:

  ```
  rm /etc/pdregistry.cfg
  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

   The result is as follows:

   "C:\Program Files\Veritas\pdde\pddb\bin\pg_ctl.exe" runservice -N postgresql-8.3 -D

**Managing NetBackup Deduplication Engine credentials**

You can manage existing credentials in NetBackup.

See “**Adding NetBackup Deduplication Engine credentials**” on page 106.

See “**Changing NetBackup Deduplication Engine credentials**” on page 106.

See “**Deleting credentials from a load balancing server**” on page 106.

See “**Determining which media servers have deduplication credentials**” on page 107.
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 77.

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 28.

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 102.
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.
   The media servers for which credentials are configured are checked.

Managing deduplication pools

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

See “Changing deduplication disk pool properties” on page 108.

See “Setting deduplication pool attributes” on page 108.

See “Clearing deduplication pool attributes” on page 109.

See “Changing the deduplication pool state” on page 110.

See “Changing the deduplication disk volume state” on page 110.

See “Deleting a deduplication pool” on page 111.

See “Determining the deduplication pool state” on page 111.

See “Determining the deduplication disk volume state” on page 111.

See "Viewing deduplication pools” on page 112.

See “Viewing deduplication pool attributes” on page 112.
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 63.

Setting deduplication pool attributes

You may have to set attributes on your existing media server deduplication pools. For example, if you configure the storage server to support optimized synthetic backups, you must then configure the existing deduplication pools for the same support.

See “About optimized synthetic backups and deduplication” on page 30.
To set a Media Server Deduplication Pool attribute

- The following is the command syntax to set a deduplication pool attribute. Run the command on the master server or on the storage server.

```
nbdevconfig -changedp -dp pool_name -stype PureDisk -setattribute attribute
```

The following describe the options that require your input:

- `pool_name` The name of the disk pool.

- `attribute` The `attribute` is the name of the argument that represents the new functionality.

  For example, `OptimizedImage` specifies that the environment supports the optimized synthetic backup method.

The following is the path to the `nbdevconfig` command:

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

---

Clearing deduplication pool attributes

You may have to clear attributes on your existing media server deduplication pools.

To clear a Media Server Deduplication Pool attribute

- The following is the command syntax to clear a deduplication pool attribute. Run the command on the master server or on the storage server.

```
nbdevconfig -changedp -dp pool_name -stype PureDisk -clearattribute attribute
```

The following describe the options that require your input:

- `pool_name` The name of the disk pool.

- `attribute` The `attribute` is the name of the argument that represents the new functionality.

The following is the path to the `nbdevconfig` command:

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

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

  The *state* is either UP or DOWN.

  The following is example output

<table>
<thead>
<tr>
<th>Disk Pool Name</th>
<th>PD_Disk_Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk Type</td>
<td>PureDisk</td>
</tr>
<tr>
<td>Disk Volume Name</td>
<td>PureDiskVolume</td>
</tr>
<tr>
<td>Disk Media ID</td>
<td>@aaaab</td>
</tr>
<tr>
<td>Total Capacity (GB)</td>
<td>49.98</td>
</tr>
<tr>
<td>Free Space (GB)</td>
<td>43.66</td>
</tr>
<tr>
<td>Use%</td>
<td>12</td>
</tr>
<tr>
<td>Status</td>
<td>UP</td>
</tr>
<tr>
<td>Flag</td>
<td>ReadOnWrite</td>
</tr>
<tr>
<td>Flag</td>
<td>AdminUp</td>
</tr>
<tr>
<td>Flag</td>
<td>InternalUp</td>
</tr>
<tr>
<td>Num Read Mounts</td>
<td>0</td>
</tr>
<tr>
<td>Num Write Mounts</td>
<td>1</td>
</tr>
<tr>
<td>Cur Read Streams</td>
<td>0</td>
</tr>
<tr>
<td>Cur Write Streams</td>
<td>0</td>
</tr>
</tbody>
</table>

Viewing deduplication 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.**

Viewing deduplication pool attributes

Use the NetBackup `nbdevquery` command to view deduplication pool attributes.
To view deduplication pool attributes

◆ The following is the command syntax to view the attributes of a deduplication pool. Run the command on the NetBackup master server or on the deduplication storage server:

UNIX: /usr/openv/netbackup/bin/admincmd/nbdevquery -listdp -dp pool_name -stype PureDisk -U

Windows: install_path\NetBackup\bin\admincmd\nbdevquery -listdp -dp pool_name -stype PureDisk -U

The following is example output:

<table>
<thead>
<tr>
<th>Disk Pool Name</th>
<th>MediaServerDeduplicationPool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk Pool Id</td>
<td>MediaServerDeduplicationPool</td>
</tr>
<tr>
<td>Disk Type</td>
<td>PureDisk</td>
</tr>
<tr>
<td>Status</td>
<td>UP</td>
</tr>
<tr>
<td>Flag</td>
<td>OpenStorage</td>
</tr>
<tr>
<td>Flag</td>
<td>AdminUp</td>
</tr>
<tr>
<td>Flag</td>
<td>InternalUp</td>
</tr>
<tr>
<td>Flag</td>
<td>LifeCycle</td>
</tr>
<tr>
<td>Flag</td>
<td>CapacityMgmt</td>
</tr>
<tr>
<td>Flag</td>
<td>OptimizedImage</td>
</tr>
<tr>
<td>Raw Size (GB)</td>
<td>235.76</td>
</tr>
<tr>
<td>Usable Size (GB)</td>
<td>235.76</td>
</tr>
<tr>
<td>Num Volumes</td>
<td>1</td>
</tr>
<tr>
<td>High Watermark</td>
<td>98</td>
</tr>
<tr>
<td>Low Watermark</td>
<td>80</td>
</tr>
<tr>
<td>Max IO Streams</td>
<td>-1</td>
</tr>
<tr>
<td>Storage Server</td>
<td>DedupeServer.symantecs.org (UP)</td>
</tr>
</tbody>
</table>

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

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 160.
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 consumes two CPU cores. Queue processing writes to the deduplication engine `storaged.log` file. See “NetBackup Deduplication Engine logs” on page 122.

- 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.

Because maintenance processing occurs automatically, you should not need to invoke those processes manually. However, you may do so.

See “Performing maintenance manually” on page 115.

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

### Performing maintenance manually

Usually, you should not need to run the maintenance processes manually. However, you can do so; the processes are transaction queue processing and garbage collection.

See “About maintenance processing” on page 114.

A control command launches the queue processing and garbage collection. The following is the path name of the command:

- On UNIX and Linux systems, `/usr/openv/pdde/pdcr/bin/crcontrol`.
- On Windows systems, `install_path\Veritas\pdde\Crcontrol.exe`.

**To process the transaction queue manually**

1. Run the control command with the `--processqueue` option. The following is an example on a Windows system:
   
   `install_path\Veritas\pdde\Crcontrol.exe --processqueue`

2. To examine the results, run the control command with the `--dsstat 1` option (number 1 not lowercase letter l). The command may run for a long time; if you omit the 1, results return more quickly but they are not as accurate.

   See “Viewing capacity within deduplication container files” on page 90.
To collect garbage manually

- Run the control command with the `-v -m +1,+2 --noreport` options. The following is an example on a UNIX system:

```
/usr/openv/pdde/pdcr/bin/crcollect -v -m +1,+2 --noreport
```

Setting NetBackup configuration options by using `bpsetconfig`

Symantec recommends that you use the **NetBackup Administration Console Host Properties** to configure NetBackup.

However, some configuration options cannot be set by using the **Administration Console**. You can set those configuration options by using the `bpsetconfig` command.

Alternatively, on UNIX systems you can set configuration options in the `bp.conf` file.

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

To set configuration options by using the `bpsetconfig` command

1. On the host on which you want to set configuration options, write the current configuration to a file by running the following command:

   **UNIX:**
   ```bash
   /usr/openv/netbackup/bin/admincmd/bpgetconfig -h hostname > file.txt
   ```

   **Windows:**
   ```bash
   install_path\NetBackup\bin\admincmd\bpgetconfig -h hostname > file.txt
   ```

2. Edit and save the file.

   You can change the values of the options that are in the file.
   
   You can add option and value pairs.
   
   Ensure that you understand the values that are allowed and the format of any new options that you add.

3. Upload the configuration by running the following command:

   **UNIX:**
   ```bash
   /usr/openv/netbackup/bin/admincmd/bpsetconfig -h hostname file.txt
   ```

   **Windows:**
   ```bash
   install_path\NetBackup\bin\admincmd\bpsetconfig -h hostname file.txt
   ```
Resizing the storage partition

If the volume that contains the deduplication storage is resized dynamically, restart the NetBackup services on the storage server. You must restart the services so that NetBackup can use the resized partition correctly. If you do not restart the services, NetBackup reports the capacity as full prematurely.

To resize the deduplication storage

1. Stop all NetBackup jobs on the storage on which you want to change the disk partition sizes and wait for the jobs to end.
2. Deactivate the media server that hosts the storage server.
   See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* or the *NetBackup Administrator's Guide for Windows, Volume I*.
3. Stop the NetBackup services on the storage server.
   Be sure to wait for all services to stop.
4. Use the operating system or disk manager tools to dynamically increase or decrease the deduplication storage area.
5. Restart the NetBackup services.
6. Activate the media server that hosts the storage server.
   See the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* or the *NetBackup Administrator's Guide for Windows, Volume I*.
7. Restart the deduplication jobs.

See “About adding additional storage” on page 55.

Restoring files at a remote site

If you use optimized duplication to copy images from a local site to a remote site, you can restore from the copies at the remote site to clients at the remote site. To do so, use a server-directed restore or a client-redirected restore, which restores files to a client other than the original client.

For information about how to redirect restores, see “Managing client restores” in the *NetBackup Administrator's Guide for UNIX and Linux, Volume I* or the *NetBackup Administrator's Guide for Windows, Volume I*.

You may have to configure which media server performs the restore. In optimized duplication, the media server that initiates the duplication operation becomes the write host for the new image copies. The write host restores from those image copies. If the write host is at the local site, it restores from those images at the remote site to the alternate client at the remote site. That host reads the image
across the WAN and then writes the image back across the WAN to the alternate client. In this case, you can specify that the media server at the remote site as the restore server.

See “About optimized duplication of deduplicated data” on page 31.

See “Specifying the restore server” on page 118.

Specifying the restore server

NetBackup may not use the backup server as the restore server for deduplicated data.

See “How deduplication restores work” on page 48.

You can specify the server to use for restores. The following are the methods that specify the restore server:

- Always use the backup server. Two methods exist, as follows:
  - Use NetBackup Host Properties to specify a Media host override server. All restore jobs for any storage unit on the original backup server use the media server you specify. Specify the same server for the Restore server as for the Original backup server.
    
    
    This procedure sets the FORCE_RESTORE_MEDIA_SERVER option. Configuration options are stored in the bp.conf file on UNIX systems and the registry on Windows systems.
  
  - Create the touch file USE_BACKUP_MEDIA_SERVER_FOR_RESTORE on the NetBackup master server in the following directory:
    
    UNIX: usr/openv/netbackup/db/config
    
    Windows: install_path\veritas\netbackup\db\config
    
    This global setting always forces restores to the server that did the backup. It applies to all NetBackup restore jobs, not just deduplication restore jobs. If this touch file exists, NetBackup ignores the FORCE_RESTORE_MEDIA_SERVER and FAILOVER_RESTORE_MEDIA_SERVER settings.

- Always use a different server.
  Use NetBackup Host Properties to specify a Media host override server.
  See the previous explanation about Media host override, except: Specify the different server for the Restore server.
A single restore instance. Use the `bprestore` command with the `-disk_media_server` option. Restore jobs for each instance of the command use the media server you specify. See the *NetBackup Commands* guide.
Managing deduplication

Specifying the restore server
Troubleshooting

This chapter includes the following topics:

■ About deduplication logs
■ Troubleshooting installation issues
■ Troubleshooting configuration issues
■ Troubleshooting operational issues
■ Viewing disk errors and events
■ Deduplication event codes and messages

About 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 following is the path name of the log file for the deduplication configuration script:

■ UNIX: storage_path/log/pdde-config.log
■ Windows: storage_path\log\pdde-config.log

NetBackup creates this log file during the configuration process. If your configuration 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 logs

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

- 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.

- A VxUL log file for the events and errors that NetBackup receives from polling. The originator ID for the deduplication engine is 364. See “About VxUL logs” on page 124.

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 94.

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 pd.conf file” on page 78.

See “Editing the deduplication pd.conf file” on page 79.

**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 8-1 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.

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*.

**Table 8-1** NetBackup VxUL logs

<table>
<thead>
<tr>
<th>Activity</th>
<th>VxUL OID</th>
<th>Processes that use the ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetBackup Deduplication Engine</td>
<td>364</td>
<td>The NetBackup Deduplication Engine that runs on the deduplication storage server.</td>
</tr>
</tbody>
</table>
Table 8-1  NetBackup VxUL logs (continued)

<table>
<thead>
<tr>
<th>Activity</th>
<th>VxUL OID</th>
<th>Processes that use the ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backups and restores</td>
<td>N/A</td>
<td>Messages appear in the log files for the following processes:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The bpbrm backup and restore manager</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The bpdbm database manager</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The bptm disk manager</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The bptm tape manager for I/O operations</td>
</tr>
<tr>
<td>Backups and restores</td>
<td>117</td>
<td>The nbjm Job Manager.</td>
</tr>
<tr>
<td>Device configuration and monitoring</td>
<td>111</td>
<td>The nbemm process.</td>
</tr>
<tr>
<td>Device configuration and monitoring</td>
<td>178</td>
<td>The Disk Service Manager process that runs in the Enterprise Media Manager (EMM) process.</td>
</tr>
<tr>
<td>Device configuration and monitoring</td>
<td>202</td>
<td>The storage server interface process that runs in the Remote Manager and Monitor Service.</td>
</tr>
<tr>
<td>Device configuration and monitoring</td>
<td>230</td>
<td>The Remote Disk Service Manager interface (RDSM) that runs in the Remote Manager and Monitor Service.</td>
</tr>
</tbody>
</table>

Troubleshooting installation issues

The following sections may help you troubleshoot installation issues.

See “Installation on SUSE Linux fails” on page 125.

Installation on SUSE Linux fails

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

```
....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
```
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 “Deduplication configuration script log” on page 121.

See “Database system error (220)” on page 126.

See “Server not found error” on page 127.

See “License information failure during configuration” on page 127.

See “The disk pool wizard does not display a volume” on page 128.

### Database system error (220)

A database system error indicates that an error occurred in the storage initialization.

**Error message**

```
ioctl() error, Database system error (220)
```

**Example**

```
RDSM has encountered an STS error:

Failed to update storage server ssname, database system error
```
The `PDDE_initConfig` script was invoked, but errors occurred during the storage initialization.

First, examine the deduplication configuration script log file for references to the server name. See “Deduplication configuration script log” on page 121.

Second, examine the `tpconfig` command log file errors about creating the credentials for the server name. The `tpconfig` command writes to the standard NetBackup administrative commands log directory.

**Server not found error**

The following information may help you resolve a server not found error message that may occur during configuration.

**Error message**

Server not found, invalid command parameter

**Example**

RDSM has encountered an issue with STS where the server was not found: `getStorageServerInfo` failed to create storage server `ssname`, invalid command parameter

**Diagnosis**

Possible root causes:

- When you configured the storage server, you selected a media server that runs an unsupported operating system. All media servers in your environment appear in the Storage Server Configuration Wizard; be sure to select only a media server that runs a supported operating system.
- If you used the `nbdevconfig` command to configure the storage server, you may have typed the host name incorrectly. Also, case matters for the storage server type, so ensure that you use `PureDisk` for the storage server type.

**License information failure during configuration**

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

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

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:

  ```bash
  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 129.
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 24.

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 136.
Client deduplication fails

NetBackup client-side agents (including client deduplication) depend on reverse host name look up of NetBackup server names. Conversely, regular backups depend on forward host name resolution. Therefore, the backup of a client that deduplicates it's own data may fail, while a normal backup of the client may succeed.

If a client-side deduplication backup fails, verify that your Domain Name Server includes all permutations of the storage server name.

Also, Symantec recommends that you use fully-qualified domain names for your NetBackup environment.

See “Use fully qualified domain names” on page 44.

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 24.

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.

Deduplication fails after services are restarted or a domain controller is restarted

Windows systems only.

After the NetBackup services are restarted or a domain controller is restarted, deduplication may fail on Microsoft Windows systems.

The deduplication postgresql-8.3 database service may fail during the restart if the purediskdbuser account does not have the "log on as a service" right. That right may be removed inadvertently; for example, a security policy on a domain controller may remove that right after the service is running. As a result, the postgresql-8.3 service cannot start during the next restart.

To work around the problem, do one of the following:

- Amend the security policy to allow the purediskdbuser account to have the "log on as a service" right.
- Change the service manually to a new domain account or to any other account that can run services.
Note: If you choose to switch the service to a different pre-existing account, that account must be granted full permissions to the <DBPATH>\databases\pddb\data directory.

■ Change the service to run as a Local System.

Note: If you change the service to run as a Local System, the account that runs the postgresql-8.3 service must be granted full permissions to the <DBPATH>\databases\pddb\data directory.

Cannot delete a disk pool

If you cannot delete a disk pool that you believe contains no valid backup images, the following information may help you troubleshoot the problem.

Expired fragments remain on disk

Under some circumstances, the fragments that compose an expired backup image may remain on disk even though the images have expired. For example, if the storage server crashes, normal clean-up processes may not run. In those circumstances, you cannot delete a disk pool because image fragment records still exist. The error message may be similar to the following:

DSM has found that one or more volumes in the disk pool diskpoolname has image fragments.

To delete the disk pool, you must first delete the image fragments. The nbdelete command deletes expired image fragments from disk volumes.

To delete the fragments of expired images

◆ Run the following command on the master server:

UNIX: /usr/openv/netbackup/bin/admincmd/nbdelete -allvolumes -force

Windows: install_path\NetBackup\bin\admincmd\nbdelete -allvolumes -force

The -allvolumes option deletes expired image fragments from all volumes that contain them.

The -force option removes the database entries of the image fragments even if fragment deletion fails.
Incomplete SLP duplication jobs

Incomplete storage lifecycle policy duplication jobs may prevent disk pool deletion. You can determine if incomplete jobs exist and then cancel them.

To cancel storage lifecycle policy duplication jobs

1. Determine if incomplete SLP duplication jobs exist by running the following command on the master server:
   - UNIX: `install_path\NetBackup\bin\admincmd\nbstlutil stlilist -incomplete`
   - Windows: `/usr/openv/netbackup/bin/admincmd/nbstlutil stlilist -incomplete`

2. Cancel the incomplete jobs by running the following command for each backup ID returned by the previous command (xxxxx represents the backup ID):
   - UNIX: `install_path\NetBackup\bin\admincmd\nbstlutil cancel -backupid xxxx`
   - Windows: `/usr/openv/netbackup/bin/admincmd/nbstlutil cancel -backupid xxxx`

Media open error (83)

To diagnose and resolve media open errors that may occur during backups, see the following possible causes and corrective actions:

Possible causes

- The NetBackup Deduplication Engine (spoold) was too busy to respond to the deduplication process in a timely manner.
- The NetBackup Deduplication Manager (spad) was too busy to respond to the deduplication process in a timely manner.

Diagnosis

- Examine what caused the core media server deduplication processes (spad and spoold) to be unresponsive. Were they temporarily busy (such as queue processing in progress)? Do too many jobs run concurrently?
- See “About deduplication performance” on page 42.

Status 83 is a generic error for the duplications. The NetBackup bpdm log provides additional information for determining the specific issue.
Media write error (84)

To diagnose and resolve media write errors that may occur during backups, see the following possible causes and corrective actions:

The NetBackup Deduplication Engine (spoold) was too busy to respond.
- Examine the Disk Logs report for PureDisk errors.
- Examine the disk monitoring services log files for details from the deduplication plug-in.
- See “Viewing disk reports” on page 90.

The NetBackup Deduplication Engine rebuilt the deduplication cache when the backup occurred.
- Restarting the NetBackup Deduplication Engine causes the cache to be rebuilt. During the cache rebuild no backups are accepted.

Data removal is running.
- Data cannot be backed up at the same time as it is removed.
- See “About maintenance processing” on page 114.

A user tampered with the storage.
- Users must not add files to, change files on, or delete files from the storage. If a file was added, remove it.

Storage capacity was increased.
- If you grew the storage, you must restart the NetBackup services on the storage server so the new capacity is recognized.

The storage is full.
- If possible, increase the storage capacity.
- See “About adding additional storage” on page 55.

The deduplication pool is down.
- Change the state to up.
- See “Changing the deduplication pool state” on page 110.

Firewall ports are not open.
- Ensure that ports 10082 and 10102 are open in any firewalls between the deduplication hosts.
Host name resolution problems. Client-side deduplication can fail if the client cannot resolve the host name of the server. More specifically, the error can occur if the storage server was configured with a short name and the client tries to resolve a fully qualified domain name.

To determine which name the client uses for the storage server, examine the deduplication host configuration file on the client.

See “About the deduplication host configuration file” on page 101.

To fix this problem, configure your network environment so that all permutations of the storage server name resolve.

Symantec recommends that you use fully qualified domain names.

See “Use fully qualified domain names” on page 44.

Storage full conditions

Operating system tools such as the UNIX df command do not report deduplication disk usage accurately. The operating system commands may report that the storage is full when it is not. NetBackup tools let you monitor storage capacity and usage more accurately.

The NetBackup Deduplication Guide includes information about how to monitor the storage.

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 90.

Examining the disk log reports for threshold warnings can give you an idea of when a storage full condition may occur.

How NetBackup performs maintenance can affect when storage is freed up for use.

See “About maintenance processing” on page 114.
See “Data removal process” on page 160.

While not advised, you can accelerate the process of reclaiming free space.

See “Performing maintenance manually” on page 115.
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 90.
- The NetBackup `berror` 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 `berror` command `-disk` output and in the disk reports in the NetBackup Administration Console.

<table>
<thead>
<tr>
<th>Event #</th>
<th>Event Severity</th>
<th>NetBackup Severity</th>
<th>Message example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>2</td>
<td>Error</td>
<td>Operation configload/reload failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1001</td>
<td>2</td>
<td>Error</td>
<td>Operation configload/reload failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1002</td>
<td>4</td>
<td>Warning</td>
<td>The open file limit exceeded in server PureDisk:server1.symantecs.org on host server1.symantecs.org. Will attempt to continue further.</td>
</tr>
<tr>
<td>1003</td>
<td>2</td>
<td>Error</td>
<td>A connection request was denied on the server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1004</td>
<td>1</td>
<td>Critical</td>
<td>Network failure occurred in server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>Event #</td>
<td>Event Severity</td>
<td>NetBackup Severity</td>
<td>Message example</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>1013</td>
<td>1</td>
<td>Critical</td>
<td>Task session start request on server PureDisk:server1.symantecs.org on host server1.symantecs.org got an unexpected error.</td>
</tr>
<tr>
<td>1008</td>
<td>2</td>
<td>Error</td>
<td>Task Aborted; An unexpected error occurred during communication with remote system in server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1009</td>
<td>8</td>
<td>Authorization</td>
<td>Authorization request from &lt;IP&gt; for user &lt;USER&gt; denied (&lt;REASON&gt;).</td>
</tr>
<tr>
<td>1010</td>
<td>2</td>
<td>Error</td>
<td>Task initialization on server PureDisk:server1.symantecs.org on host server1.symantecs.org got an unexpected error.</td>
</tr>
<tr>
<td>1011</td>
<td>2</td>
<td>Error</td>
<td>Task ended on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1012</td>
<td>2</td>
<td>Error</td>
<td>A request for agent task was denied on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1014</td>
<td>1</td>
<td>Critical</td>
<td>Task session start request on server PureDisk:server1.symantecs.org on host server1.symantecs.org got an unexpected error.</td>
</tr>
<tr>
<td>1015</td>
<td>1</td>
<td>Critical</td>
<td>Task creation failed, could not initialize task class on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>Event #</td>
<td>Event Severity</td>
<td>NetBackup Severity</td>
<td>Message example</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>--------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1017</td>
<td>1</td>
<td>Critical</td>
<td>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.</td>
</tr>
<tr>
<td>1018</td>
<td>16</td>
<td>Info</td>
<td>Startup of Symantec DeduplicationEngine completed successfully on server1.symantecs.org.</td>
</tr>
<tr>
<td>1019</td>
<td>1</td>
<td>Critical</td>
<td>Service Symantec DeduplicationEngine restart 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 restarted.</td>
</tr>
<tr>
<td>1020</td>
<td>1</td>
<td>Critical</td>
<td>Service Symantec DeduplicationEngine 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.</td>
</tr>
<tr>
<td>1028</td>
<td>1</td>
<td>Critical</td>
<td>Service Symantec DeduplicationEngine abort 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 caught an unexpected signal.</td>
</tr>
<tr>
<td>1029</td>
<td>1</td>
<td>Critical</td>
<td>Double backend initialization failure; Could not initialize storage backend or cache failure detected on host PureDisk:server1.symantecs.org in server server1.symantecs.org.</td>
</tr>
</tbody>
</table>
Table 8-2  Deduplication event codes and messages (continued)

<table>
<thead>
<tr>
<th>Event #</th>
<th>Event Severity</th>
<th>NetBackup Severity</th>
<th>Message example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1030</td>
<td>1</td>
<td>Critical</td>
<td>Operation Storage Database Initialization failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1031</td>
<td>1</td>
<td>Critical</td>
<td>Operation Content router context initialization failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1032</td>
<td>1</td>
<td>Critical</td>
<td>Operation log path creation/print failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1036</td>
<td>4</td>
<td>Warning</td>
<td>Operation a transaction failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1037</td>
<td>4</td>
<td>Warning</td>
<td>Transaction failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org. Transaction will be retried.</td>
</tr>
<tr>
<td>1044</td>
<td>multiple</td>
<td>multiple</td>
<td>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.</td>
</tr>
<tr>
<td>1040</td>
<td>2</td>
<td>Error</td>
<td>Operation Database recovery failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1043</td>
<td>2</td>
<td>Error</td>
<td>Operation Storage recovery failed on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
<tr>
<td>1047</td>
<td>2</td>
<td>Error</td>
<td>CRC mismatch detected; possible corruption in server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
</tbody>
</table>
Table 8-2  Deduplication event codes and messages (continued)

<table>
<thead>
<tr>
<th>Event #</th>
<th>Event Severity</th>
<th>NetBackup Severity</th>
<th>Message example</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>Error</td>
<td></td>
<td>Low space threshold exceeded on the partition containing the storage database on server PureDisk:server1.symantecs.org on host server1.symantecs.org.</td>
</tr>
</tbody>
</table>
Host replacement, recovery, and uninstallation

This chapter includes the following topics:

- Replacing the deduplication storage server host computer
- Recovering from a deduplication storage server disk failure
- Recovering from a permanent deduplication storage server failure
- Recovering the storage server after NetBackup catalog recovery
- Uninstalling media server deduplication

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 9-1 How to replace the deduplication storage server host

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change the disk volume state and disk pool state to DOWN</td>
<td>See “Changing the deduplication disk volume state” on page 110. See “Changing the deduplication pool state” on page 110.</td>
</tr>
<tr>
<td>Configure the new host so it meets deduplication requirements</td>
<td>See “About deduplication servers” on page 22. See “About deduplication server requirements” on page 24.</td>
</tr>
<tr>
<td>Move the storage to the new host.</td>
<td>See the storage vendor's documentation.</td>
</tr>
<tr>
<td>Install the NetBackup media server software on the new host</td>
<td>See the NetBackup Installation Guide for UNIX and Linux. See the NetBackup Installation Guide for Windows.</td>
</tr>
<tr>
<td>Delete the NetBackup Deduplication Engine credentials</td>
<td>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 106.</td>
</tr>
<tr>
<td>Add the credentials to the storage server</td>
<td>Add the NetBackup Deduplication Engine credentials to the storage server. See “Adding NetBackup Deduplication Engine credentials” on page 106.</td>
</tr>
<tr>
<td>Get a configuration file template</td>
<td>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 99.</td>
</tr>
<tr>
<td>Edit the configuration file</td>
<td>See “Editing a storage server configuration file” on page 99.</td>
</tr>
</tbody>
</table>
Table 9-1  How to replace the deduplication storage server host  (continued)

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
</table>
| 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 100. |
| Configure the load balancing servers     | If you have load balancing servers, add them to the configuration.  
   See “Adding a load balancing server” on page 77. |
| Change configuration settings            | If you edited the deduplication configuration file, make the same changes to that file.  
   See “About the deduplication pd.conf file” on page 78.  
   See “Editing the deduplication pd.conf file” on page 79. |
| Change the disk volume state and disk pool state to UP | See “Changing the deduplication disk volume state” on page 110.  
   See “Changing the deduplication pool state” on page 110. |
| Change the disk volume state and disk pool state to UP | See “Changing the deduplication disk volume state” on page 110.  
   See “Changing the deduplication pool state” on page 110. |
| 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 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.
After recovery, your NetBackup deduplication environment should function normally. Any valid backup images on the deduplication storage should be available for restores.

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 9-2  Process to recover from media server disk failure**

<table>
<thead>
<tr>
<th>Step</th>
<th>Instruction</th>
</tr>
</thead>
</table>
| 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 58. |
| Delete the configuration file on deduplication servers | If you use load balancing servers in your environment, delete the storage server configuration files on those servers.  
See “Deleting a deduplication host configuration file” on page 101. |
| 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 106. |
| 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 106. |
| 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 99. |
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 9-3**  Process to recover from a permanent media server failure

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change the disk volume state and disk pool state to DOWN</td>
<td>See “Changing the deduplication disk volume state” on page 110.</td>
</tr>
<tr>
<td></td>
<td>See “Changing the deduplication pool state” on page 110.</td>
</tr>
<tr>
<td>Configure the new host so it meets deduplication requirements</td>
<td>Use the same host name as the failed server.</td>
</tr>
<tr>
<td></td>
<td>See “About deduplication servers” on page 22.</td>
</tr>
<tr>
<td></td>
<td>See “About deduplication server requirements” on page 24.</td>
</tr>
</tbody>
</table>
Table 9-3  Process to recover from a permanent media server failure (continued)

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move the storage to the new host.</td>
<td>Ensure that the storage and database are mounted at the same locations. See the storage vendor's documentation.</td>
</tr>
<tr>
<td>Install the NetBackup media server software on the new host</td>
<td>See the <em>NetBackup Installation Guide for UNIX and Linux.</em> See the <em>NetBackup Installation Guide for Windows.</em> See “About the deduplication license key” on page 58.</td>
</tr>
<tr>
<td>Delete the credentials on media servers</td>
<td>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 106.</td>
</tr>
<tr>
<td>Add the credentials to the storage server</td>
<td>Add the NetBackup Deduplication Engine credentials to the storage server. See “Adding NetBackup Deduplication Engine credentials” on page 106.</td>
</tr>
<tr>
<td>Get a configuration file template</td>
<td>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 99.</td>
</tr>
<tr>
<td>Edit the configuration file</td>
<td>See “Editing a storage server configuration file” on page 99.</td>
</tr>
<tr>
<td>Configure the storage server</td>
<td>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 100.</td>
</tr>
<tr>
<td>Configure the load balancing servers</td>
<td>If you have load balancing servers, add them to the configuration. See “Adding a load balancing server” on page 77.</td>
</tr>
<tr>
<td>Change configuration settings</td>
<td>If you edited the deduplication configuration file, make the same changes to that file. See “About the deduplication pd.conf file” on page 78. See “Editing the deduplication pd.conf file” on page 79.</td>
</tr>
</tbody>
</table>
Table 9-3 Process to recover from a permanent media server failure (continued)

<table>
<thead>
<tr>
<th>Task</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change the disk volume state and disk pool state to UP</td>
<td>See “Changing the deduplication disk volume state” on page 110.</td>
</tr>
<tr>
<td></td>
<td>See “Changing the deduplication pool state” on page 110.</td>
</tr>
<tr>
<td>Restart the backup jobs</td>
<td>If any backup jobs failed, restart those jobs. Alternatively, wait until the next scheduled backup, at which time the backup jobs should succeed.</td>
</tr>
</tbody>
</table>

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 100.

Symantec recommends that you save your storage server configuration.

See “Save the storage server configuration” on page 48.

Information about recovering the master server is available.

See the NetBackup Troubleshooting Guide.

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 9-4  Disable media server deduplication

<table>
<thead>
<tr>
<th>Disable client deduplication</th>
<th>Remove the clients that deduplicate their own data from the client deduplication list. See “Disabling deduplication for a client” on page 114.</th>
</tr>
</thead>
</table>
| 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 111. |
| Delete the deduplication storage server | See "Deleting a deduplication storage server" on page 97. |
| 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. |
| Reset the deduplication registry | See “Resetting the deduplication registry” on page 104. |
| On Windows, delete accounts and files | On Windows storage servers and load balancing servers, delete the following:  
- The `purediskdbuser` account. The account is for the deduplication database administration.  
- The `purediskdbuser` folder.  
See the operating system documentation. |
| On UNIX and Linux, remove files | On UNIX and Linux storage servers and load balancing servers, remove the following files:  
- `etc/pdregistry.cfg`  
- `opt/pdag`  
- `opt/pdshared` |
### Table 9-4 Disable media server deduplication (continued)

<table>
<thead>
<tr>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete the storage server configuration file</td>
<td>The storage server and every load balancing server contain a deduplication configuration file. Delete that file from every server that you use for deduplication. See “Deleting a deduplication host configuration file” on page 101.</td>
</tr>
</tbody>
</table>
| 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.* |
Uninstalling media server deduplication
Deduplication architecture

This chapter includes the following topics:

- Deduplication server components
- Media server deduplication process
- Deduplication client components
- Deduplication client backup process
- About deduplication fingerprinting
- Data removal process

Deduplication server components

Figure 10-1 is a diagram of the storage server components.
**Figure 10-1** Storage server deduplication components

**Table 10-1** describes the components.

**Table 10-1** NetBackup deduplication components

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PureDisk plug-in</td>
<td>The PureDisk plug-in is the data interface to the NetBackup Deduplication Engine on the storage server. The PureDisk plug-in does the following:</td>
</tr>
<tr>
<td></td>
<td>▪ Separates the file’s metadata from the file’s content.</td>
</tr>
<tr>
<td></td>
<td>▪ Deduplicates the content (separates files into segments).</td>
</tr>
<tr>
<td></td>
<td>▪ Controls the data stream from NetBackup to the NetBackup Deduplication Engine and vice versa.</td>
</tr>
<tr>
<td></td>
<td>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.</td>
</tr>
<tr>
<td>NetBackup Deduplication Engine</td>
<td>The NetBackup Deduplication Engine is one of the storage server core components. It stores and manages deduplicated file data.</td>
</tr>
<tr>
<td></td>
<td>The binary file name is <em>spoold</em>, which is short for storage pool daemon; do not confuse it with a print spooler daemon. The <em>spoold</em> process appears as the NetBackup Deduplication Engine in the NetBackup Administration Console.</td>
</tr>
</tbody>
</table>
Table 10-1  NetBackup deduplication components (continued)

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetBackup Deduplication Manager</td>
<td>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.</td>
</tr>
<tr>
<td>Catalog plug-in</td>
<td>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.</td>
</tr>
<tr>
<td>Deduplication database</td>
<td>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 PostgreSQL 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.</td>
</tr>
</tbody>
</table>

Media server deduplication process

Figure 10-2 shows the backup process when a media server deduplicates the backups. The destination is a media server deduplication pool. A description follows.
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) on the client generates the backup images and moves them to the media server bptm process. The Backup/Archive Manager also sends the information about files within the image to the Backup/Restore Manager (bpbrm). The Backup/Restore Manager sends the file information to the bpdbm process on the master server for the NetBackup database.
- 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 in 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 10-3** shows the backup process when a media server deduplicates the backups. The destination is a PureDisk storage pool. A description follows.

**Figure 10-3** 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 *bpkar* process on the client.
- The Backup/Archive Manager (*bpkar*) generates the backup images and moves them to the media server *bptm* process. The Backup/Archive Manager also sends the information about files within the image to the Backup/Restore Manager (*bpbrm*). The Backup/Restore Manager sends the file information to the *bpdbm* process on the master server for the NetBackup database.
- 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 PureDisk storage pool. The list is used as a cache so the plug-in does not have to request each fingerprint from the storage pool.
The PureDisk plug-in compares the file fingerprints and the segment fingerprints against the fingerprint list in its cache.

- 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 10-2 describes the client deduplication components.

<table>
<thead>
<tr>
<th>Component</th>
<th>Host</th>
<th>Description</th>
</tr>
</thead>
</table>
| 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 10-4 shows the backup process of a client that deduplicates its own data. The destination is a media server deduplication pool. A description follows.
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 Backup/Archive Manager also sends the information about files within the image to the Backup/Restore Manager (bpbrm). The Backup/Restore Manager sends the file information to the bpdbm process on the master server for the NetBackup database.
- 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 10-5 shows the backup process of a client that deduplicates its own data. The destination is a PureDisk storage pool. A description follows.

**Figure 10-5** 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 \((\text{nbjm})\) starts the Backup/Restore Manager \((\text{bpbrm})\) 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 \((\text{nbostpxy})\) on the client and the data moving processes \((\text{bpbkar} \text{ on the client and}\text{ bptm on the media server})\). NetBackup uses the proxy plug-in on the media server to route control information from \text{bptm} to \text{nbostpxy}.  

---

**Deduplication architecture**

**Deduplication client backup process**
The Backup/Archive Manager (bpbkar) generates the backup images and moves them to the client nbostpxy process by shared memory. The Backup/Archive Manager also sends the information about files within the image to the Backup/Restore Manager (bpbrm). The Backup/Restore Manager sends the file information to the bpdbm process on the master server for the NetBackup database.

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 its 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 114.
A
about NetBackup deduplication 13
about NetBackup deduplication options 14
about the deduplication host configuration file 101
appliance deduplication 15
attributes
  clearing deduplication pool 109
  clearing deduplication storage server 96
  OptimizedImage 30
  setting deduplication pool 108
  setting deduplication storage server 95
  viewing deduplication pool 112
  viewing deduplication storage server 104

B
backup
  client deduplication process 156
big endian 145
byte order 145

C
cache hits field of the job details 87
capacity
  adding storage 55
capacity and usage reporting for deduplication 87
  CIFS 44
clearing deduplication pool attributes 109
client deduplication
  about 26
  components 156
  disabling for a specific client 114
  host requirements 27
  limitations 28
  requirements 27
  sizing the systems 19
Common Internet File System 44
compression
  and deduplication 29
configuring a deduplication pool 62
configuring a deduplication storage server 61
configuring a deduplication storage unit 64
configuring deduplication 60
container files
  about 90
  viewing capacity within 90
CR sent field of the job details 87
credentials
  about NetBackup Deduplication Engine 28
  adding NetBackup Deduplication Engine 106
  changing NetBackup Deduplication Engine 106

data removal process
  for deduplication 160
database system error 126
dedup field of the job details 87
deduplication
  about credentials 28
  about fingerprinting 159
  about the license key 58
  adding credentials 106
  cache hits field of the job details 87
  capacity and usage reporting 87
  changing credentials 106
  client backup process 156
  compression 29
  configuration file 78
  configuring 60
  configuring optimized synthetic backups 70
  container files 90
  CR sent field of the job details 87
  data removal process 160
  dedup field of the job details 87
  encryption 29
  event codes 136
  garbage collection 115
  how it works 15
  license key for 58
  licensing 58
  limitations 25
  maintenance processing 114
deduplication (continued)
media server process 153
network interface 28
node 24
optimized 31
performance 42
planning deployment 18
requirements for optimized 31
scaling 44
scanned field of the job details 87
storage capacity 54
storage destination 19
storage paths 54
storage requirements 53
storage unit properties 65
stream rate field of the job details 87
supported systems 19
deduplication configuration file
editing 79
settings 80
deduplication database
about 153
log file 122
purediskdbuser account on Windows 153
deduplication disk pool. See deduplication pool
deduplication disk volume
changing the state 110
determining the state 111
deduplication host configuration file
about 101
deleting 101
deduplication hosts
and firewalls 29
client requirements 27
load balancing server 23
server requirements 24
storage server 23
deduplication logs
about 121
client deduplication proxy plug-in log 123
client deduplication proxy server log 124
deduplication database 122
NetBackup Deduplication Engine 122
NetBackup Deduplication Manager 122
configuration script 121
PureDisk plug-in log 123
VxUL deduplication logs 124
deduplication node (continued)
adding a load balancing server 77
removing a load balancing server 102
deduplication optimized synthetic backups
about 30
deduplication pool
about 62
changing properties 108
changing the state 110
clearing attributes 109
configuring 62
deleting 111
determining the state 111
properties 63
setting attributes 108
viewing 112
viewing attributes 112
deduplication port usage
about 29
troubleshooting 134
deduplication processes do not start 129
deduplication rate
how file size affects 43
monitoring 85
deduplication registry
resetting 104
deduplication servers
about 22
components 151
host requirements 24
deduplication storage capacity
about 54
monitoring 87
viewing capacity in container files 90
deduplication storage destination 19
deduplication storage paths 54
deduplication storage requirements 53
deduplication storage server
about 23
changing properties 94
clearing attributes 96
components 151
configuration failure 127
configuring 61
deleting 97
determining the state 97
editing configuration file 99
getting the configuration 99
reconfiguring 82
deduplicate storage server (continued)
- recovery 145
- replacing the host 141
- setting attributes 95
- setting the configuration 100
- viewing 103
- viewing attributes 104

deduplicate storage server configuration file
- about 98

deduplicate storage type 19

Deduplication storage unit
- Only use the following media servers 66
- Use any available media server 66

deleting backup images 113

deleating deduplication host configuration file 101

disaster recovery
- protecting the data 48
- recovering the storage server after catalog recovery 147

disk failure
- deduplication storage server 143

disk logs 91

disk logs report 89

disk pool. See deduplication pool
- cannot delete 132

disk pool status report 88, 91

disk storage unit report 91

Disk type 65

disk volume
- changing the state 110
- determining the state of a deduplication 111
- volume state changes to down 130

E

encryption
- and deduplication 29

endian
- big 145
- little 145

event codes
- deduplication 136

F

fingerprinting
- about deduplication 159

firewalls and deduplication hosts 29

G

garbage collection
- for deduplication 115
- invoke manually 115

H

host requirements 24

how deduplication works 15

I

images on disk report 91

initial seeding 42

iSCSI 44

L

license information failure
- for deduplication 127

license key
- for deduplication 58

licensing deduplication 58

limitations
- media server deduplication 25

little endian 145

load balancing server
- about 23
- adding to a deduplication node 77
- for deduplication 23
- removing from deduplication node 102

logs
- about deduplication 121
- client deduplication proxy plug-in log 123
- client deduplication proxy server log 124
- deduplication configuration script log 121
- deduplication database log 122
- disk 91
- NetBackup Deduplication Engine log 122
- NetBackup Deduplication Manager log 122
- PureDisk plug-in log 123
- VxUL deduplication logs 124

M

maintenance processing for deduplication 114

Mapped full VM backup 70

Maximum concurrent jobs 66

Maximum fragment size 66

media server deduplication process 153
media server deduplication (continued)
sizing the systems 19
Media Server Deduplication Option
    about 21
media server deduplication pool. See deduplication pool
migrating from PureDisk to NetBackup deduplication 50
migrating to NetBackup deduplication 51

N
NetBackup Client Deduplication Option 15
NetBackup deduplication
    about 13
    license key for 58
NetBackup Deduplication Engine
    about 152
    about credentials 28
    adding credentials 106
    changing credentials 106
    logs 122
NetBackup Deduplication Manager
    about 153
    logs 122
NetBackup deduplication options 14
NetBackup Media Server Deduplication Option 15
    network interface
        for deduplication 28
NFS 44
node
    deduplication 24

O
OpenStorage appliance deduplication 15
optimized deduplication copy
    about 31
    about the media server in common 32
    configuring 71
    configuring behavior for 75
    guidance for 33
    limitations 32
    pull configuration 39
    push configuration 33, 35
    requirements 31
    throttling traffic 77
optimized synthetic backups
    configuring for deduplication 70
    deduplication 30

P
pd.conf file
    about 78
    editing 79
    settings 80
pdde-config.log 121
performance
    deduplication 42
    monitoring deduplication rate 85
port usage
    and deduplication 29
    troubleshooting 134
provisioning the deduplication storage 53
PureDisk deduplication 15
PureDisk Deduplication Option
    replacing with media server deduplication 49
PureDisk plug-in
    about 152
    log file 123
purediskdbuser account 153

Q
queue processing
    invoke manually 115
    queue processing for deduplication 114

R
reconfiguring deduplication 82
recovery
    deduplication storage server 145
    from deduplication storage server disk failure 143
Red Hat Linux
    deduplication processes do not start 129
    replacing PDDO with NetBackup deduplication 49
    replacing the deduplication storage server 141
reports
    disk logs 89, 91
    disk pool status 88, 91
    disk storage unit 91
resetting the deduplication registry 104
restores
    at a remote site 117
    how deduplication restores work 48
    specifying the restore server 118
reverse host name lookup
    prohibiting 127
reverse name lookup 127
S
scaling deduplication 44
scanned field of the job details 87
seeding
  initial 42
server not found error 127
setting deduplication pool attributes 108
storage capacity
  about 54
  adding 55
  for deduplication 54
  monitoring deduplication 87
  viewing capacity in container files 90
storage paths
  for deduplication 54
  reconfiguring 82
storage requirements
  for deduplication 53
storage server
  about the configuration file 98
  changing properties for deduplication 94
  components for deduplication 151
  configuring for deduplication 61
  deduplication 23
  deleting a deduplication 97
  determining the state of a deduplication 97
  editing deduplication configuration file 99
  getting deduplication configuration 99
  reconfiguring deduplication 82
  recovery 145
  replacing the deduplication host 141
  setting the deduplication configuration 100
  viewing 103
  viewing attributes 104
storage server configuration
  getting 99
  setting 100
storage server configuration file
  editing 99
storage type
  for deduplication 19
storage unit
  configuring for deduplication 64
  properties for deduplication 65
  recommendations for deduplication 66
Storage unit name 65
Storage unit type 65
stream rate field of the job details 87
supported systems 19

T
counting deduplication 44
scanned field of the job details 87
setting deduplication pool attributes 108
storage capacity
  about 54
  adding 55
  for deduplication 54
  monitoring deduplication 87
  viewing capacity in container files 90
storage paths
  for deduplication 54
  reconfiguring 82
storage requirements
  for deduplication 53
storage server
  about the configuration file 98
  changing properties for deduplication 94
  components for deduplication 151
  configuring for deduplication 61
  deduplication 23
  deleting a deduplication 97
  determining the state of a deduplication 97
  editing deduplication configuration file 99
  getting deduplication configuration 99
  reconfiguring deduplication 82
  recovery 145
  replacing the deduplication host 141
  setting the deduplication configuration 100
  viewing 103
  viewing attributes 104
storage server configuration
  getting 99
  setting 100
storage server configuration file
  editing 99
storage type
  for deduplication 19
storage unit
  configuring for deduplication 64
  properties for deduplication 65
  recommendations for deduplication 66
Storage unit name 65
Storage unit type 65
stream rate field of the job details 87
supported systems 19

U
uninstalling media server deduplication 147

V
viewing deduplication pool attributes 112
viewing storage server attributes 104