Backup for VMware ESX Environments with Veritas NetBackup PureDisk™ from Symantec

A comparison of three different methods of backing up ESX guests using deduplication

Mark Salmon, Sr. Principal Systems Engineer
Contents

1.0 Executive overview ................................................................. 1
2.0 NetBackup Enterprise Client overview ........................................ 3
3.0 NetBackup PureDisk overview .................................................. 4
4.0 Test environment and procedures .............................................. 5
5.0 NetBackup PureDisk for VMware: configuration options ................... 6
5.1 NetBackup PureDisk client in a guest VM ...................................... 6
5.1.1 Observations ................................................................. 8
5.1.2 Pros ......................................................................... 9
5.1.3 Cons ...................................................................... 9
5.1.4 Best practices ............................................................. 9
5.2 NetBackup PureDisk client on VCB proxy server ............................. 10
5.2.1 Observations ................................................................. 11
5.2.2 Pros ......................................................................... 11
5.2.3 Cons ................................................................. 12
5.2.4 Best practices ............................................................. 12
5.3 NetBackup PureDisk Deduplication Option (PDDO) ......................... 13
5.3.1 Observations ................................................................. 14
5.3.2 Pros ......................................................................... 15
5.3.3 Cons ................................................................. 16
5.3.4 Best practices ............................................................. 16
6.0 Conclusions ............................................................................ 17
Glossary ..................................................................................... 18
1.0 Executive overview
The explosive growth of VMware Infrastructure in organizations both large and small has enabled corporations to fully exploit their hardware investments. With multiple virtual guest systems running on few physical hardware nodes, hardware costs are reduced, as well as space, power, and air-conditioning requirements. The necessity for regular backup functions still exists, however, and is consequently magnified. Long-established backup processes usually run in off-hours when both network and CPU resources are readily available but the efficient use of virtual infrastructure hardware often provides no window for backup. The successful backup of virtual guests without adversely affecting the guest operating systems or the VMware ESX server itself is a challenge. These challenges have driven the development of new methods and technologies, each featuring certain strengths and limitations, to back up and protect virtual infrastructure guests.

Today there are two fundamentally different methods to back up guest hosts residing on a VMware Infrastructure server. The first method is to install a client agent in each virtual guest. Client agents can be further divided into traditional backup agents or deduplication client agents. The traditional Veritas NetBackup from Symantec agent is prevalent in data centers around the world, where it is used on both physical and virtual hosts. The Symantec deduplication backup client agent, called the PureDisk client, is renowned for exceptionally light server impact and low network usage during backups.

The second method of backup is to take a snapshot of the virtual guests. With hardware snapshots, successfully creating a consistent image of the virtual guest presents a real challenge. The snapshot method that is dependable time after time is VMware Consolidated Backup (VCB), which was introduced with VMware Infrastructure 3.0. VCB is a snapshot technology developed by VMware to minimize the impact on the ESX server and the guests during the backup process.

The NetBackup client agent is easily installed in a virtual guest just as it would be on a physical host and it supports many features. Support for databases, encryption, compression, open files, and bare metal restore (BMR) to dissimilar hardware are some of the compelling reasons to deploy this client. With any VMware ESX server, the deployment of the NetBackup client must be carefully architected because the additional load added to virtual hosts will impact the overall environment. A fully utilized ESX server rarely provides an adequate backup window for simultaneous backup of multiple NetBackup clients, but still, successful backups can be performed on lightly loaded machines. A very effective backup of virtual guests with the NetBackup client can be achieved by integration through VCB outside of the virtual guest.

The challenge created by the rapidly growing VMware environments can be overcome by integrating within the backup process technology that will perform capacity optimization, single-instance storage, more commonly called data deduplication. Veritas NetBackup PureDisk from Symantec is an easy-to-deploy backup solution featuring low CPU and memory consumption with light network usage, making it an optimal choice for virtual guest backup. With a NetBackup PureDisk client, only one full backup is required, and all successive backups are effectively incrementals having only small changes; yet a full recovery is available at any time from any backup. Additionally, NetBackup PureDisk technology provides storage reduction for backups performed by NetBackup software. NetBackup PureDisk, the deduplication engine, can be deployed to enhance VMware backups in three functional methods:
1. NetBackup PureDisk client in each guest operating system (OS)
2. NetBackup PureDisk client on the proxy server utilizing VCB
3. NetBackup PureDisk as a disk storage unit for NetBackup software integrating with VCB
2.0 **NetBackup Enterprise Client overview**

The NetBackup Enterprise Client, which may be used independently of NetBackup PureDisk, incorporates support for VMware "off-host" backup through its seamless integration with VMware Consolidated Backup (VCB). The Enterprise Client is installed on a VCB proxy server and can be instructed to initiate a backup of any guest OS on any ESX server that is accessible through the SAN to the proxy server. Configuration is performed via the graphical user interface (GUI) on the master server where the Snapshot Policy Wizard assists in creating a policy using the following steps:

1. Identify a proxy server
2. Select the snapshot method
3. Display available ESX servers
4. Display all the configured guest OSs (whether running or not)
5. Test full functionality once completed

The VCB technology provides two ways to present the guest OSs to be backed up: the first is a file system backup whereby individual files can be restored with ease, and the second is a full virtual machine disk format (VMDK) backup. For almost all backup vendors, restoring individual files requires a file system backup, and restoring a full virtual machine requires a full VMDK backup—thus requiring two separate and complete backups. The NetBackup Enterprise Client, however, has eliminated the need for two separate backup passes with embedded snapshot and image-mapping technology so that a full VMDK backup can be restored as a full image or individual files can be extracted from that image. Subsequent incremental backups can be performed without difficulty, ensuring that the "one pass, two types of restore" functionality does not preclude the potential for advanced incremental technology to exist on top of it.

Beyond VCB, NetBackup Enterprise Client includes other technologies, which are included in the "bundle" of features created to support advanced and non-traditional backup methods, utilizing client-based snapshots (host or array) for applications, databases, and file systems. Enterprise Client can perform SAN backups in the role of a SAN media server, where the host performs its own backup to disk or tape storage units that can be shared with other media servers. The SAN client data moves over the SAN to a media server configured to receive data over the SAN, and the media server will manage the data movement to disk or tape. These technologies are explained and described in other Symantec white papers.
### 3.0 NetBackup PureDisk overview

NetBackup PureDisk is a backup solution initially developed to address the backup challenges of small or remote offices. Traditionally, small islands of remote servers have been backed up by deploying a standalone backup product in each location. The challenges of remote administration, offsite management of tape, and more stringent guidelines in the protection and retention of corporate data have driven the need for NetBackup PureDisk technology. The NetBackup PureDisk client can be deployed on a wide variety of operating systems, and it includes support for commonly used applications such as Microsoft® Exchange and SQL. The feature that ensures it is optimal over the WAN is the integral deduplication technology that totally eliminates the transmission of unneeded data to the NetBackup PureDisk storage pool. Only unique data is transferred from each client as data is deduplicated across all backup clients from all remote locations and safely held (compressed and encrypted) in the central NetBackup PureDisk storage pool. The NetBackup PureDisk solution can be configured in a high-availability mode with several "nodes" protecting the data, and disaster recovery is ensured through replication to one or more NetBackup PureDisk storage pools. The first time a client runs a backup, all the unique files are copied to the NetBackup PureDisk storage pool based on the file selection for that client. Subsequent backups will back up only the changed "segments," creating little data to back up and small backup durations. Each backup can be considered a "full backup," but only the granular changes are copied, making for fast and efficient backups.

Apply the NetBackup PureDisk technology in the data center and the powerful deduplication engine eliminates long agonizing LAN backups and large disk repositories, saving both time and money. NetBackup PureDisk technology tightly integrates with NetBackup software an intelligent disk target that deduplicates data across all existing NetBackup clients.
4.0 Test environment and procedures

All test backups were performed on the same hardware and infrastructure except where noted. Results for deduplication, effective backup performance, and backup duration were recorded for successful evaluation and conclusions.

<table>
<thead>
<tr>
<th>Component</th>
<th>Hardware</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMware ESX server</td>
<td>• Dell Power Edge 2950</td>
<td>• VMware ESX Server 3.5 Update 3</td>
</tr>
<tr>
<td></td>
<td>• (2) Quad-Core Intel® Xeon e5410 @ 2.33 GHz</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 16 GB RAM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• (2) onboard Intel Pro/1000 GbE NIC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• (1) onboard Intel Pro/1000 GbE NIC service console</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• (8) 73 GB 10k RPM disk</td>
<td></td>
</tr>
<tr>
<td>VMware Consolidated Backup proxy server</td>
<td>• Dell Power Edge 2950</td>
<td>• Windows® 2003 R2</td>
</tr>
<tr>
<td></td>
<td>• (2) Quad-Core Intel Xeon e5410 @2.33 GHz</td>
<td>• VMware Consolidated Backup 1.5</td>
</tr>
<tr>
<td></td>
<td>• 16 GB RAM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• (4) 73 GB 10k RPM disk</td>
<td></td>
</tr>
<tr>
<td>VMware Virtual Center server</td>
<td>• Installed in a guest OS</td>
<td>• Windows 2003 R2</td>
</tr>
<tr>
<td></td>
<td>• 2 GB RAM allocated</td>
<td>• VMware vCenter 2.5.0</td>
</tr>
<tr>
<td>Storage</td>
<td>• 4 Gb/s SAN</td>
<td>• N/A</td>
</tr>
<tr>
<td>VMware guest(s)</td>
<td>• 1 CPU 2.33 GHz allocated</td>
<td>• Windows 2003 R2</td>
</tr>
<tr>
<td></td>
<td>• 384 MB RAM allocated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 4.98 GB system drive allocated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 10 GB data drive allocated</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Hardware/Software environment used for testing
5.0 NetBackup PureDisk for VMware: configuration options

5.1 NetBackup PureDisk client in a guest VM

The NetBackup PureDisk client software may be quickly and easily deployed to any guest OS by remote connection via a browser to the NetBackup PureDisk storage pool. The client is then registered with the NetBackup PureDisk storage pool and added to the local database. The client backup can be run immediately after registration, or a policy with a schedule can be created. Installing the client in a virtual machine (VM) is no different than in a physical host, and the ongoing management will be the same as well. The first time a client performs a backup, all the unique files are copied to the NetBackup PureDisk server based on the file selection for that client. When a client is backed up, each file is segmented, and the segments receive a unique fingerprint before they are copied to the server. Subsequent backups will copy only new files or the changed blocks of previously backed-up files. Because of the small segments of files being compressed and immediately sent to the NetBackup PureDisk storage pool, an overall smaller load is applied to the backup host. The segment size is configurable, but it should be noted that there is a trade-off in performance as the segment size decreases, and often very little gain is found lower than the default size of 128 KB segments, which function efficiently for file and folder backups.

![NetBackup PureDisk client in the guest OS](image)

Each backup can be considered a “full backup,” but only the granular changes are copied, making for fast and efficient backups. The initial backup load will use as much network bandwidth as available unless throttling has been implemented in the policy, which may be desirable in a virtual host environment. Incremental backups will use system resources for a fraction of the time required for the initial backup because only the small partial changes of files will be copied to the backup server.
The virtual machines used in the testing were identical Windows 2003 R2 hosts, each with a 5 GB system drive. A logical 10 GB data drive located in the direct attached storage was assigned to each, host which initially contained 100% unique data across all the hosts used for the tests. For the first test, both compression and encryption were enabled across all clients to maximize the overall CPU and memory consumption in each guest and across the ESX server overall. No throttling was set in the policies, so the maximum network bandwidth was available for use (a 100 Mb/s network for the NetBackup PureDisk client in the guest OS). The default segment size of 128 KB was used and not altered throughout all the tests for the NetBackup PureDisk agent in the guests. With each test criterion, a single guest was backed up, followed by two guests at the same time, then four guests, and finishing with eight simultaneous guests being backed up. A single CPU was allocated to each guest and the backup was a single thread from each host. Memory, CPU, and backup durations were recorded, and factors were captured as shown in table 2.

<table>
<thead>
<tr>
<th>Number of guests backing up simultaneously</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Guest network</strong></td>
<td>Throughput (MB/s)</td>
<td>1.95</td>
<td>3.76</td>
<td>3.23</td>
</tr>
<tr>
<td><strong>Guest disk</strong></td>
<td>Throughput (MB/s)</td>
<td>3.37</td>
<td>3.28</td>
<td>3.23</td>
</tr>
<tr>
<td><strong>Guest CPU</strong></td>
<td>Average usage</td>
<td>80%</td>
<td>80%</td>
<td>75%</td>
</tr>
<tr>
<td><strong>Guest memory 6.0</strong></td>
<td>Peak usage</td>
<td>86%</td>
<td>86%</td>
<td>82%</td>
</tr>
<tr>
<td><strong>ESX disk</strong></td>
<td>Throughput (MB/s)</td>
<td>3.37</td>
<td>6.71</td>
<td>12.9</td>
</tr>
<tr>
<td><strong>ESX CPU</strong></td>
<td>Average usage</td>
<td>11%</td>
<td>21%</td>
<td>40%</td>
</tr>
<tr>
<td><strong>ESX memory 7.0</strong></td>
<td>Average usage</td>
<td>23%</td>
<td>21%</td>
<td>23%</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>All backups completed</td>
<td>19 min 12 sec</td>
<td>19 min 53 sec</td>
<td>22 min</td>
</tr>
<tr>
<td></td>
<td>Scan file system</td>
<td>53 sec</td>
<td>60 sec</td>
<td>1 min 22 sec</td>
</tr>
<tr>
<td></td>
<td>Store files</td>
<td>17 min 50 sec</td>
<td>17 min 50 sec</td>
<td>21 min 50 sec</td>
</tr>
<tr>
<td><strong>Deduplication</strong></td>
<td>Overall average</td>
<td>62.82%</td>
<td>65.10%</td>
<td>64.5%</td>
</tr>
<tr>
<td><strong>Modified files</strong></td>
<td>Data drive</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 2. NetBackup PureDisk client in the VMware guest(s)

For the subsequent full backups, the data drive information was modified by 10% on each virtual host using a script that would randomize a percentage of files and further randomize a percentage of the contents of in the selected files. Backups were performed with two, four, and eight guests; and the data changes were reset for each backup set, creating incremental amounts of changed data to simulate the backups.
5.1.1 Observations

**Initial full backup**—As the number of hosts backing up increased from one to eight, the backup times increased over 67% on the initial backups as the available network was found to be a bottleneck. It must be noted that the memory in each guest was a mere 384 MB and the network for these tests was 100 Mb/s. The 90% usage of memory may seem high—but less so considering the total memory available to each guest. The NetBackup PureDisk application uses about 33 MB of memory and VSS use exceeds 110 MB. The additional memory usage was the OS and the compression and hashing processes. The guest CPU usage decreased as more hosts were added to the backup queue, and the memory usage diminished over the same time period because the network was throttling the performance. The compression overhead was verified to be an additional 30% CPU load when additional backups were performed with compression disabled. The ESX statistics for disk throughput were almost exactly in line with the number of hosts being backed up because the MB/sec approximately doubled as the server count doubled. The ESX CPU usage went from 11% with one host, to 68% with eight hosts, and followed that pattern. Twelve hosts would likely have consumed 100% of the available CPU. Each backup stream used almost 10% of the total ESX CPU resources.

**Second full backup**—Backup times increased, but only a marginal increase was observed as the number of hosts was increased. Guest CPU usage diminished when more guests were backing up unique data concurrently; and at the same time, network throughput decreased on a per-host basis (but only marginally). With eight hosts backing up, the best performance overall was achieved, backup time increased only slightly, and the ESX server CPU utilization peaked at 60% for a short time. The peak periods can be linked to the time that was spent creating new hashes for the new data and moving new data to storage over IP.

<table>
<thead>
<tr>
<th>Number of guests backing up simultaneously</th>
<th>2</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Guest disk</strong></td>
<td>Throughput (MB/s)</td>
<td>4.83</td>
<td>3.88</td>
</tr>
<tr>
<td><strong>Guest CPU</strong></td>
<td>Average usage</td>
<td>78%</td>
<td>70%</td>
</tr>
<tr>
<td><strong>Guest memory 8.0</strong></td>
<td>Peak usage</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td><strong>ESX disk</strong></td>
<td>Minimum usage</td>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td><strong>ESX CPU</strong></td>
<td>Range of usage throughout backup (MB)</td>
<td>165–334</td>
<td>165–334</td>
</tr>
<tr>
<td><strong>ESX memory 9.0</strong></td>
<td>Throughput (MB/s)</td>
<td>9.77</td>
<td>15.87</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>Usage (%)</td>
<td>21%</td>
<td>(20–30) 37% peak</td>
</tr>
<tr>
<td>All backups completed</td>
<td>Average usage</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Scan file system</td>
<td>Range of usage throughout backup (MB)</td>
<td>560–808</td>
<td>520–1400</td>
</tr>
<tr>
<td>Store files</td>
<td>Overall average</td>
<td>96.22%</td>
<td>98.23%</td>
</tr>
<tr>
<td>Deduplication</td>
<td>Data drive</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Modified files</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Incremental backup with NetBackup PureDisk client in each virtual guest
5.1.2 Pros
- Browser-based installation is easy.
- Very high deduplication is achieved on daily backups.
- The Windows guest OS supports both SQL and Exchange.
- Bandwidth can be throttled if required.
- Restoration directly into the guest OS is supported.
- The impact on the ESX server is low.

5.1.3 Cons
- Additional software in the guest OS needs to be maintained.
- Clients compete for limited resources, so use caution in scheduling.
- Full VMDK restores are not possible.
- Disaster recovery is not quick and requires the recovery of the guest OS, followed by installation of the NetBackup PureDisk client, and completed with restoring all files.
- Guests must be powered up for backups to occur.

5.1.4 Best practices
The NetBackup PureDisk client is quick and easy to set up and configure in a VM guest operating system. The deduplication is generally very good to excellent with average file system data. Use the default segment size of 128 KB unless large graphic, compressed, or CAD files require backup; and, in that case, use a much larger segment size because little deduplication will be achieved. With a smaller segment size, better deduplication can be achieved; and by choosing a larger segment size, greater throughput can be achieved. Consider throttling initial backups when large hosts with long backups are required. Subsequent backups will move only changed data blocks so no throttling is necessary, but use caution when multiple guests are scheduled to back up in overlapping backup windows. The network will likely be the bottleneck, as it is most often in traditional backup environments, since the NetBackup PureDisk client is light on both memory and CPU usage. When using multiple streams, a maximum of two to three streams per virtual guest is usually the best for top performance.
5.2 NetBackup PureDisk client on VCB proxy server

Another effective method of data protection for virtual infrastructure guests is the deployment of the NetBackup PureDisk client on the VMware proxy server. Since the proxy server is central to the VCB backup, installing a NetBackup PureDisk backup client on the proxy server will enable backup of VMware guests with little impact on the individual guests or the ESX server. In this configuration, no backup software resides in the guest operating system. Proxy server software must be installed on a separate physical machine running Windows 2003 Server SP1 or Windows Server 2003 R2. For enterprise installations, SAN storage is recommended, although iSCSI is supported. Either is required to enable the proxy to share common storage with the ESX server(s) it will back up. Hardware considerations for the proxy need to be based on the factors that affect performance—the number of virtual machines to back up, the total amount of data to be backed up, network speeds, HBA size, the number of backup streams, and the desired backup window.

For additional information, refer to the Symantec Proxy Server Sizing Wizard and the VMware document, VMware Consolidated Backup—Best Practices and Deployment Considerations for SAN Environments.¹

![Diagram of NetBackup PureDisk client on the proxy](image)

Figure 2. NetBackup PureDisk client on the proxy

Installation of the NetBackup PureDisk client is as simple as opening a browser connection on the proxy server, connecting to the URL of the NetBackup PureDisk server, and following the GUI-based instructions to register the client. Integration with VCB is achieved through the `vcbmounter` command executed from the proxy; this initiates the sync driver to ensure the operating system is synchronized before the snapshot takes place. The `vcbmounter` command for the test was executed in a pre-script configured in the NetBackup PureDisk client with the following syntax:

```
```

A simple bat file was created on the client that creates and mounts the VCB snapshot. The script was specified in the **Pre Script** field in the client policy. The script used was

```
@echo on
vcbmounter -h dcdell1122.veritas.com -u vmadmin -p password -a ipaddr:vm802.veritas.com -r h:\mnt\test -t fullvm -m san
```

The NetBackup PureDisk policy was configured to back up only files that were located in the snapshot mount location; and immediately after the snapshot of the fullvm, the backup was started.

<table>
<thead>
<tr>
<th>Windows 2003 client with OS drive of 5.0 GB and a 10 GB data drive for total of 14.1 GB</th>
<th>NetBackup PureDisk client on proxy server</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of VMDK file utilized</td>
<td>Initial full backup</td>
</tr>
<tr>
<td>ESX statistics</td>
<td>85%</td>
</tr>
<tr>
<td>Backup speed</td>
<td>Throughput (MB/s)</td>
</tr>
<tr>
<td>Backup duration</td>
<td>hh:mm:ss</td>
</tr>
<tr>
<td>After deduplication</td>
<td>Effective backup size</td>
</tr>
<tr>
<td>Deduplication rate</td>
<td>PureDisk report</td>
</tr>
</tbody>
</table>

**Table 4. NetBackup PureDisk on the proxy server**

The first 15 GB virtual guest backup had initial source data of 14.1 GB; and when it was backed up, it compressed and deduplicated down to 8.62 GB for a deduplication rate of 39%. Subsequent backups with 6% of changed data resulted in an additional 0.8 GB of data to back up and an effective deduplication rate of 94.6% with the full backup taking only 9 minutes, 2 seconds. An additional 10% of the files were changed, and the backup then had a deduplication rate of 96.53% and a backup duration time of only 6:55. Using the ESX performance tools, the load on memory and CPU was observed and there was negligible impact on the guest operating system; plus, the changes to the ESX host itself were insignificant and were not recorded.

### 5.2.1 Observations

The deduplication was excellent with the NetBackup PureDisk client on the proxy server. Utilizing VCB virtually eliminates any extra load on both the ESX server and the virtual guests. The highly utilized VMDK files (91%) deduplicated very well (>94%), which shows that NetBackup PureDisk was not just compressing white space but deduplicating and compressing all the new data.

### 5.2.2 Pros

- Since the proxy is Windows-based, the client installation is easy and straightforward using a browser.
- Deduplication rates for both full and incremental backups are high.
- Storage optimization for full (snapshot-based) backups is good.
- There is very little load on the ESX environment.
5.2.3 Cons

- The use of scripts makes this method less user-friendly than alternatives.
- Granular restore of files requires installation of the agent within the virtual machine.

5.2.4 Best practices

To effectively back up virtual guests utilizing VCB with NetBackup PureDisk, install the NetBackup PureDisk client on the VMware proxy. The best way to initiate the snapshot creation with VCB is to have a pre-script that runs in the NetBackup PureDisk client before the backup job is started. Simple scripts and explanations are available on www.VMware.com and in several VMware documents. Once a working script is created for one guest, it can easily be repurposed for other virtual host backups. The command set is vcbmounter and associated switches. The configuration of the backup job needs to be only the partition where the VCB snapshot has been copied.
5.3 NetBackup PureDisk Deduplication Option (PDDO)

When a NetBackup PureDisk environment is configured to be used by NetBackup software as an intelligent disk destination with compression, encryption, and deduplication capabilities, it is referred to as the NetBackup PureDisk Deduplication Option. To install and configure PDDO within a NetBackup environment, a functioning NetBackup PureDisk environment is the first requirement. A license key is needed to enable the use of NetBackup PureDisk storage within NetBackup software.

The Windows 2003 VMware Proxy also had the NetBackup Media Server software loaded onto it and then configured as a NetBackup Server. Through a browser connection on the Media Server to the NetBackup PureDisk storage pool, the NetBackup PureDisk agent with the PDDO option was installed and configured on the Media Server, thereby establishing a secure connection between the Media Server and the PureDisk storage pool. With the NetBackup GUI, a new storage pool is created of the type called "PureDisk." With PDDO installed on the VMware proxy, the data flow is over the IP network to the NetBackup PureDisk server and stored in NetBackup PureDisk storage on the SAN. A NetBackup policy was created to back up the proxy as a NetBackup client snapshot. Within NetBackup software there is a wizard that assists in creation of the policy and selection of the ESX host and the associated guests. Once the policy is created, NetBackup software even attempts to execute the customized scripts that were generated through the wizard. If the policy fails, the error reported will assist in diagnosing the problem. The fullMapped VM backup is the option most often chosen since a single-pass backup will not only capture the full VM, but a mapping process within NetBackup software also builds a map of the files.
in the volume, which then facilitates single file recovery. For more on VCB and NetBackup software integration, refer to *Backing Up VMware with NetBackup* by George Winter.

<table>
<thead>
<tr>
<th>PureDisk Deduplication Option Backup Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td># of PureDisk Nodes</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

Assumes multi-stream NBU backups
Backup data set consisting of typical enterprise mix (file and folder, databases, mail)
Best possible performance achieved over 10GigE Ethernet
Hard drives configured in a RAID 6 configuration

Table 5. PureDisk deduplication option backup throughput

5.3.1 Observations
When backing up VMware guests through VCB, the data is moved to or mounted from a proxy server from where it is captured and moved to its final media destination (disk, tape, or NetBackup PureDisk). The methodology when capturing VMDK files on the proxy server has direct impact on the deduplication rate that is achieved. Backing up VMDK data with the standard Windows file system policy "MS-Windows-NT" will produce a high deduplication rate capability because the PDDO plug-in receives the standard NetBackup tape archive (tar) image that contains file-level information. The file-level information enables PDDO to deduplicate files within the tar image efficiently. Deduplication rates tend to be slightly lower when this same data set is captured with the snapshot policy of the type called "FlashBackup-Windows." When creating a new NetBackup snapshot policy to leverage the NetBackup VCB integration, four virtual machine backup options are available and listed in Table 6.
VCB provides an option that optimizes the VMDK files by removing white space contained in the VMDK file. This option is available in the NBU policy. Deduplication is high when VMDK files are backed up as files with an NBU policy of the type "MS-Windows-NT" (option 0). Deduplication may be slightly lower when backing up VMDK files with a policy of the type "FlashBackup-Windows" (option 1, 2, and 3). Enabling compression in the PDDO plug-in will increase the storage optimization. This option is by default disabled and can be enabled in the NetBackup PureDisk configuration file on the proxy server. A full VMDK file backup deduplicates better than optimized VMDK files, and the more empty space within the VMDK file the better the deduplication rate.

5.3.2 Pros
- The NetBackup interface is familiar: Seamless integration with NetBackup software.
- Deduplication is performed on all data as it is passed to the media server.
- Setup and configuration are easy.
5.3.3 Cons

- Deduplication may be lower for full backup.

5.3.4 Best practices

NetBackup for VMware makes the backup of virtual guests easy to set up, configure, and execute because the entire process can be done through a wizard. When backing up individual files and executing incremental backups, deduplication is outstanding and is comparable to standard NetBackup PureDisk client in VM backups. When backing up full VMDK files (fullvm), however, the deduplication may be slightly lower unless the non-optimized backup option is selected. NetBackup software, in optimizing the snapshot for backup, compresses the image, making high deduplication rates difficult to achieve. In general, non-optimized backups are recommended for achieving higher deduplication rates.
6.0 Conclusions

Traditional backup solutions are able to back up virtual guests on a VMware ESX host; however, the resource utilization on the host and the backup time required quickly become areas of concern. The NetBackup PureDisk client running in a virtual guest provides excellent deduplication at the source and can back up with less continuous load and much shorter backup durations. All clients that back up to the NetBackup PureDisk storage pool will realize the benefits of global deduplication. If a lesser load is desired on the VMware infrastructure, the NetBackup PureDisk client deployed on the proxy provides the same excellent deduplication, but the trade-off is the minor scripting that is required to produce a snapshot of each virtual guest. The NetBackup for VMware option available to NetBackup software users provides a quick and effective way to back up virtual guests through seamless integration with VCB. When coupled with PDDO, NetBackup for VMware will deduplicate all virtual guest data as it is stored on the NetBackup PureDisk storage unit.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 NetBackup PureDisk client installed in VMware Guest</td>
<td>• Excellent deduplication rates&lt;br&gt;• Simple individual file recovery</td>
<td>• Agent install required on each guest&lt;br&gt;• No VM image-level restore</td>
</tr>
<tr>
<td>NetBackup PureDisk client on proxy server utilizing VCB technology</td>
<td>• Excellent deduplication rates&lt;br&gt;• VM image-level restore supported</td>
<td>• Requires additional scripting&lt;br&gt;• No GUI support&lt;br&gt;• Individual file restore difficult</td>
</tr>
<tr>
<td>VCB with integrated deduplication on proxy server (using NetBackup PureDisk Deduplication Option—PDDO)</td>
<td>• GUI-driven deduplication through automated backup policies&lt;br&gt;• VM image-level restore supported&lt;br&gt;• Individual file recovery supported&lt;br&gt;• High deduplication rates for incremental files</td>
<td>• Deduplication rates may be lower on full VM backups</td>
</tr>
</tbody>
</table>
### Glossary

**Deduplication**

The elimination of redundant data through the storing of only the unique data from a single data source or more often multiple data sources. Data is sliced up into “segments” so small changes in a data source will generate a very small amount of data to store on disk. Deduplication dramatically reduces power, space, and cooling requirements for storing vast amounts of data.

**Disk Storage Unit**

A disk location where NetBackup software will store backup images.

**DSU**

Acronym for disk storage unit.

**Guest VM**

Guest virtual machine that is running on VMware, often Windows or Linux OS.

**NetBackup**

Leading enterprise data protection backup product.

**PureDisk**

A backup application that can back up multiple clients running various operating systems and deduplicates, compresses, and encrypts the data in a storage repository. NetBackup PureDisk software can be configured to be a multi-node clustered device supporting up to 100 TB of compressed deduplicated data, with replication capabilities and export images to tape in native NetBackup format.

**PureDisk Deduplication Option**

This is the option to enable the NetBackup PureDisk engine within a NetBackup media server. A customer configures a NetBackup disk storage unit that points to NetBackup PureDisk for one or more NetBackup media servers. NetBackup software can back up all supported clients, OSs, and applications to a NetBackup PureDisk storage pool. It performs deduplication on the data as it passes through the media servers.

**VCB**

See VMware Consolidated Backup.

**VM**

Acronym for virtual machine, which represents a guest OS running on VMware.
Virtual machine disk format is a VMware file format that represents a virtual machine.

A backup tool that resides on a separate server, called a proxy, which performs off-host copies of guest hosts that reside on ESX servers. Backup vendors integrate with VCB to create enterprise backup capability of VMware guests without impacting ESX or the guest OS.
About Symantec

Symantec is a global leader in providing security, storage and systems management solutions to help consumers and organizations secure and manage their information-driven world. Our software and services protect against more risks at more points, more completely and efficiently, enabling confidence wherever information is used or stored.

For specific country offices and contact numbers, please visit our website.

Symantec World Headquarters
20330 Stevens Creek Blvd.
Cupertino, CA 95014 USA
+1 (408) 517 8000
1 (800) 721 3934
www.symantec.com