

Lab Validation Report

NetBackup PureDisk from Symantec Efficient, Storage-Optimized Data Protection

By Tony Palmer
With Brian Garrett

June, 2008

Table of Contents

Table of Contents	i
Introduction	1
<i>Background</i>	1
<i>Symantec NetBackup PureDisk 6.5</i>	2
ESG Lab Validation	3
<i>Getting Started</i>	3
<i>Performance</i>	6
<i>Data De-Duplication</i>	9
<i>Disaster Recovery and High Availability</i>	11
<i>Remote Office Protection</i>	13
ESG Lab Validation Highlights	18
Issues to Consider	18
ESG Lab's View	19
Appendix	20

ESG Lab Reports

The goal of ESG Lab reports is to educate IT professionals about emerging technologies and products in the storage, data management and information security industries. ESG Lab reports are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objective is to go over some of the more valuable feature/functions of products, show how they can be used to solve real customer problems and identify any areas needing improvement. ESG Lab's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments. This ESG Lab report was sponsored by Symantec.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. Copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482.0188.

Introduction

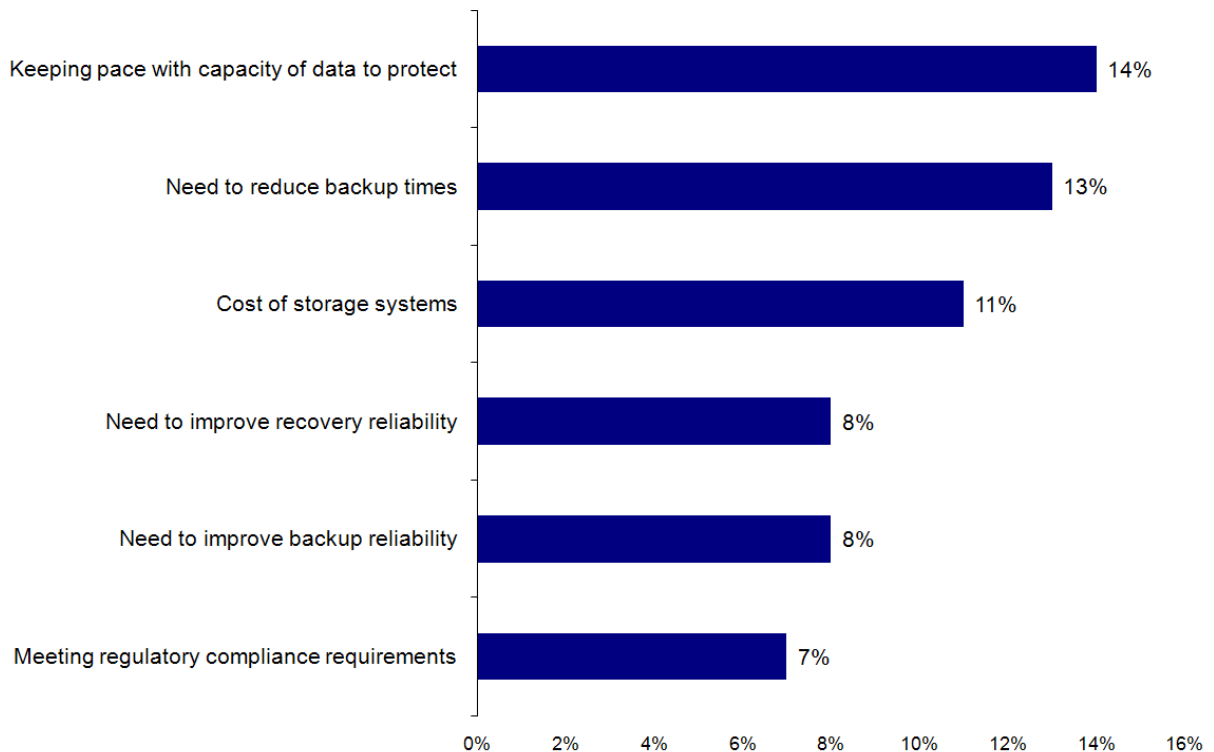
A growing number of organizations are deploying disk to disk backup and data de-duplication technologies to improve the speed and reliability of their backup and disaster recovery operations. This ESG Lab Validation Report examines Symantec’s family of backup and recovery solutions that combine the power of the PureDisk global data de-duplication engine with industry-leading NetBackup backup and recovery software to improve the speed and reliability of disk-based data protection while reducing the cost of disk capacity and network bandwidth.

Background

Backup and recovery challenges have plagued IT professionals for decades. As shown in Figure 1, a recent ESG survey of 398 storage professionals indicates that keeping pace with the capacity of data to protect tops the list of challenges associated with current data protection processes and recovery technologies.¹ Reducing both backup times and the cost of storage systems is also a concern. Backup and recovery reliability and regulatory compliance concerns indicate that backup administrators, and their managers, often go home at night wondering whether their current backup and recovery solutions are placing their organizations at risk.

FIGURE 1. FACTORS DRIVING WAN BACKUP SOLUTIONS

Which of the following would you characterize as the primary challenge with your organization's current data protection processes and technologies? (Percent of respondents, N = 398)



¹ Source: ESG Research Report, *Data Protection Survey*, October, 2007

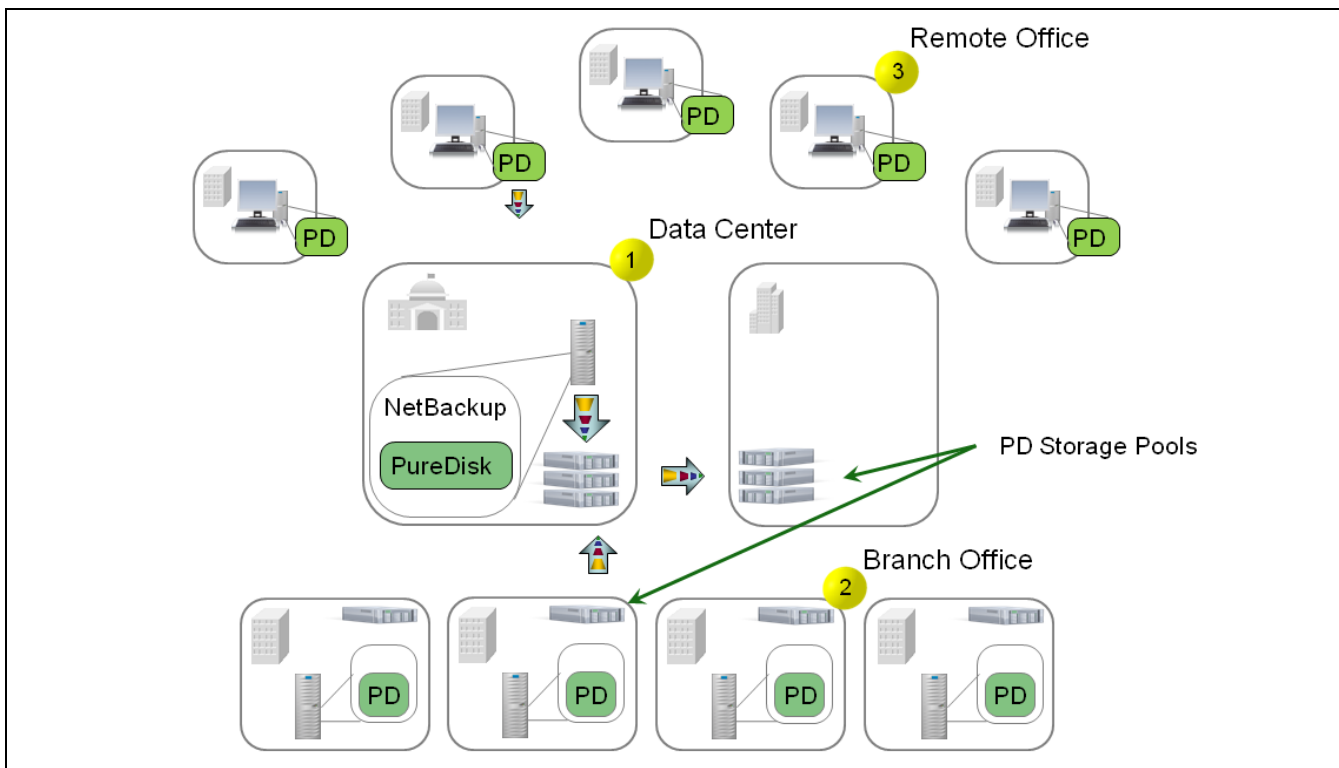
More and more organizations are turning to disk-based backup and recovery methods to improve the speed and reliability of backup and restore operations. As a matter of fact, ESG research indicates that 66% of organizations have already deployed some form of backup to disk solution. While disk-based systems are inherently faster and more reliable than tape, they can be significantly more expensive to acquire. This cost concern is driving the adoption of data de-duplication technologies. Data de-duplication reduces backup to disk capacity requirements by eliminating the need to store multiple copies of the same data over time. Eliminating extra copies reduces the capacity, and cost, of a disk-based backup and recovery solution. It can also be used to retain more backups on disk for quick and reliable recovery and reduce the cost of sending backup data over a WAN for disaster recovery.

Symantec NetBackup PureDisk 6.5

Initially introduced as a standalone platform for protecting corporate data residing in remote and branch offices, PureDisk has since been tightly integrated with Symantec NetBackup version 6.5. As shown in Figure 2, PureDisk is a software-based data de-duplication engine that can be deployed in a number of locations within an organization, such as:

1. A de-duplication engine installed on NetBackup servers within a data center (called PureDisk De-duplication Option). Data de-duplication (shown as an arrow) reduces the amount of disk capacity in the primary data center by up to 50 times and can be used to reduce the WAN bandwidth required to replicate data to a secondary data center by up to 500 times. Backup data is stored on a centralized PureDisk pool running on industry standard servers with direct- or network-attached storage.
2. A de-duplication engine in large branch offices with a smaller PureDisk pool for quick and reliable local restores and WAN efficient replication to a larger PureDisk pool in a data center. This option is also well-suited for the protection of virtual server data.
3. A software-only agent running on application servers, workstations and laptops within smaller remote offices for cost-effective, hands-off backup and recovery operations.

FIGURE 2. SYMANTEC NETBACKUP PUREDISK



ESG Lab Validation

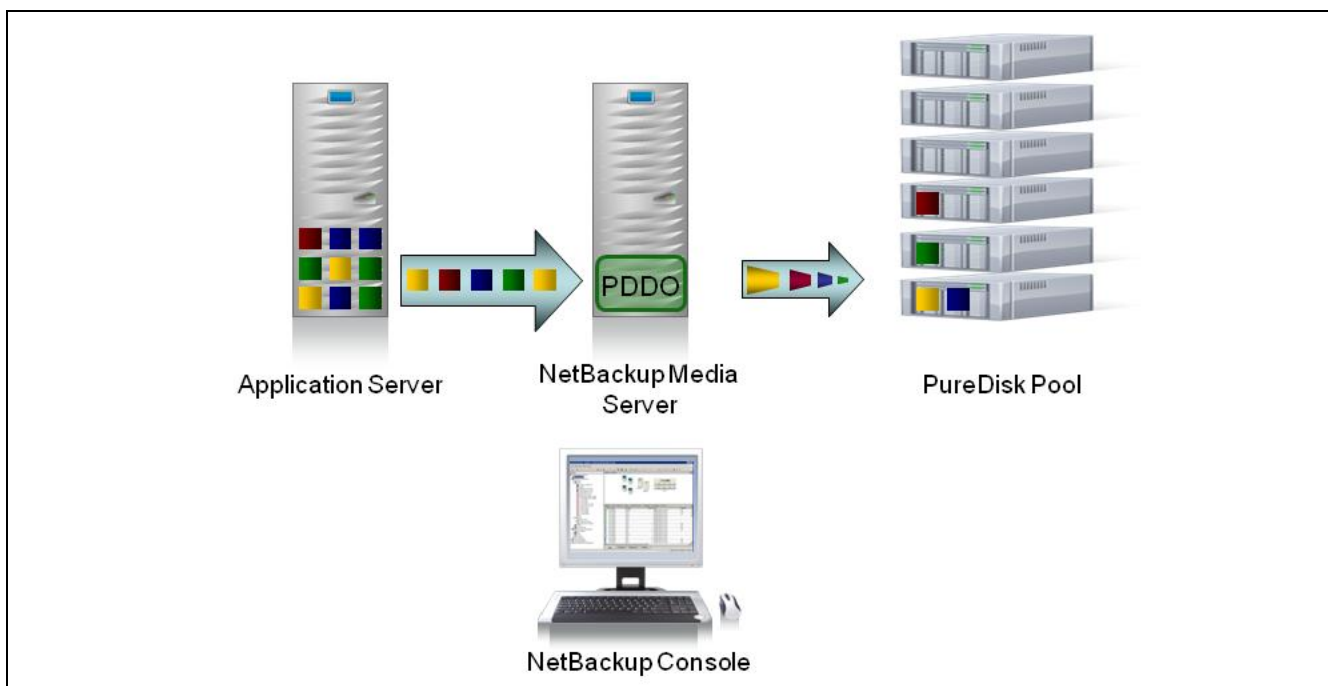
ESG Lab performed hands-on evaluation and testing of NetBackup PureDisk version 6.5 at Symantec's Mountain View facility. Testing was designed to validate the benefits of data de-duplication along with the performance, scalability, availability and recoverability of NetBackup PureDisk solutions.

Getting Started

PureDisk employs global data de-duplication across sites and servers within a thin layer of software. The software identifies and filters repeated data segments stored in files within a single system and across multiple systems over time to capture only net new segments. In addition, data compression algorithms are applied to eliminate space and redundant file patterns, further shrinking data volumes.

At the heart of PureDisk is the content router. Content routers are hardware and storage agnostic software modules that are installed on identical servers or blades. They can utilize SAN, DAS, NAS or any combination thereof for their back-end storage. Agents interact directly with the content router that owns the data, which enables parallel communication between agents and content routers. As long as the system is online and operational, content routers can be added to the PureDisk Storage Pool at any time. The system automatically balances content in the background for optimal performance and recoverability. Figure 3 illustrates how systems in a central data center can leverage NetBackup's PureDisk De-duplication Option (PDDO) to reduce the amount of data being transferred during backup by sending only unique blocks of data to a centralized pool of storage.

FIGURE 3. NETBACKUP PUREDISK DE-DUPLICATION OPTION

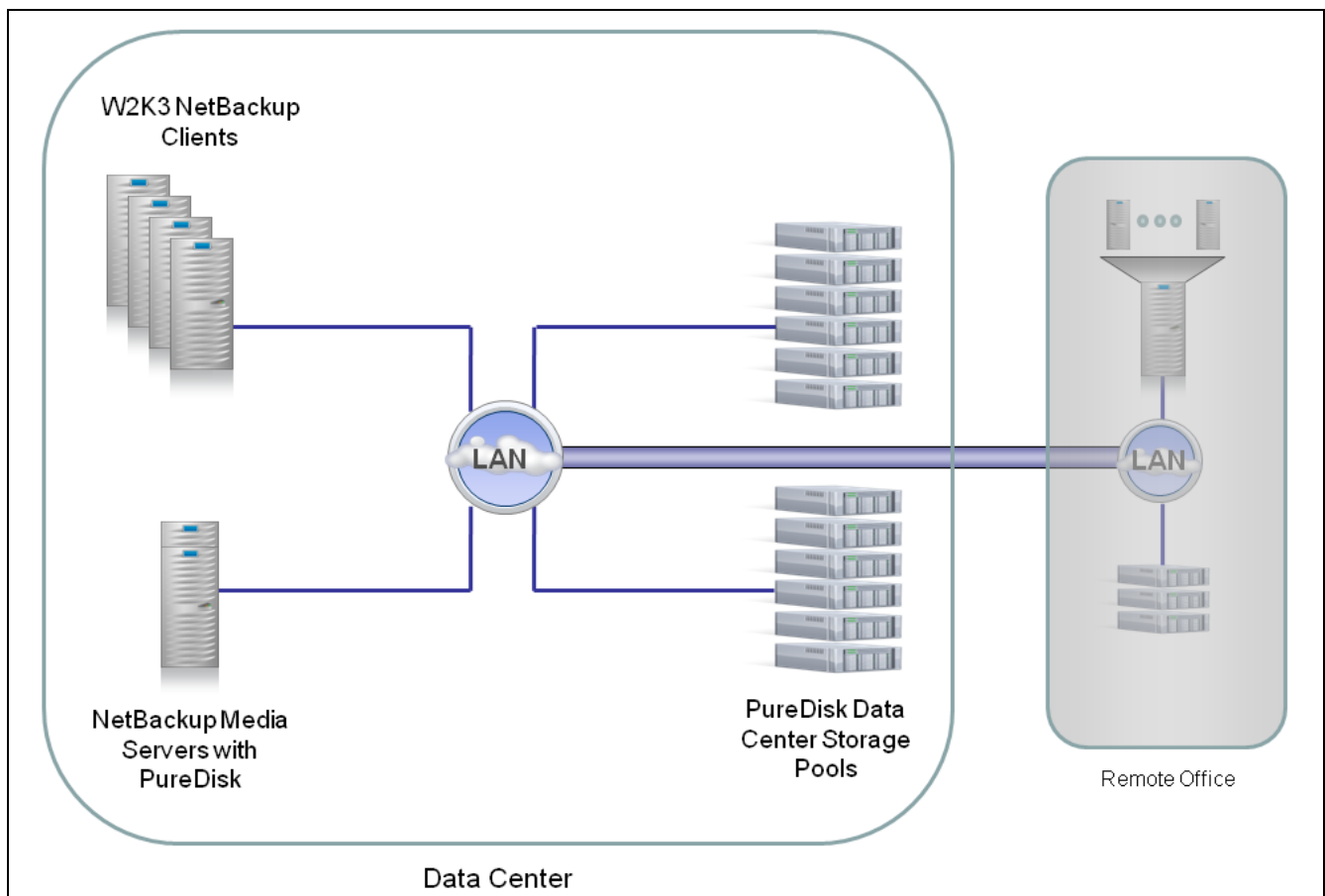


In this example, data residing on an application server to be backed up is processed by a NetBackup media server equipped with a PDDO engine. The PDDO engine identifies and eliminates redundant data before it is sent over a LAN or a WAN to a PureDisk pool. At smaller remote offices, PureDisk software agents are deployed directly on application servers, while larger remote offices and data centers typically deploy a local

PureDisk pool to improve backup, de-duplication and recovery performance. PureDisk also supports disaster recovery and offsite archival by replicating backup data over a wide area network to a remote data center. Global data de-duplication ensures that only unique (and optionally compressed) sub-file data segments are replicated. This reduces the cost of network bandwidth and improves the performance of backups and replication across the WAN.

ESG Lab tested PureDisk in a lab environment designed to simulate a central data center, a second data center for DR and a remote office.² Two PureDisk systems were configured to test central data center backup using Veritas NetBackup (see Figure 4). In this test, ESG Lab used PureDisk to protect Microsoft Windows 2003 servers running in the data center.

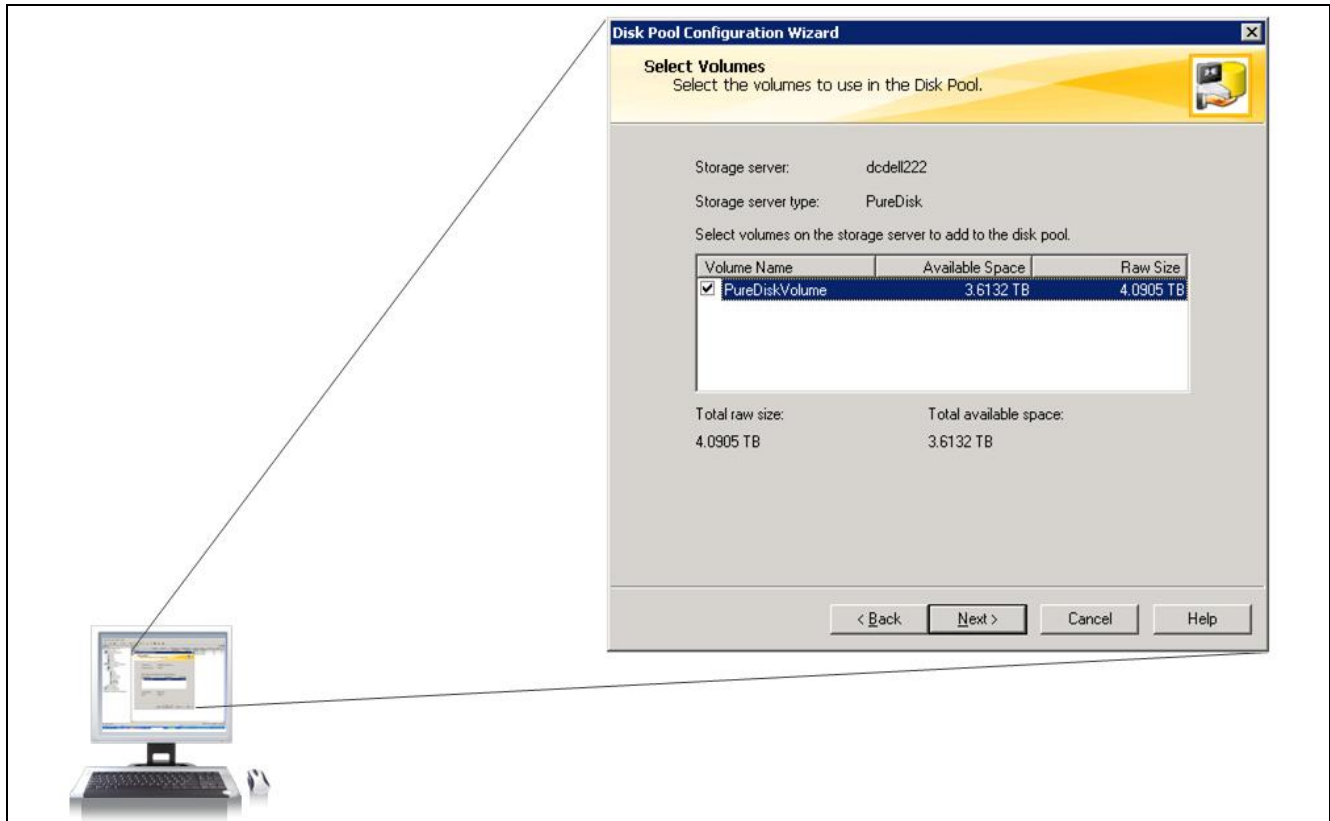
FIGURE 4. THE ESG LAB TEST BED: PUREDISK IN THE DATA CENTER



The installation of the PureDisk system with approximately 5 TB of de-duplicated backup capacity began with a PureDisk specific Linux distribution and was completed in about 30 minutes. The NetBackup Management console was used to create a disk pool and allocate capacity as shown in Figure 5.

² Configuration details can be found in the Appendix.

FIGURE 5. CREATING A PUREDISK DISK POOL



Within 40 minutes from first power on and five minutes after the PureDisk installation was completed, ESG Lab was using NetBackup to backup and de-duplicate the data associated with a Windows 2003 system.

Why This Matters

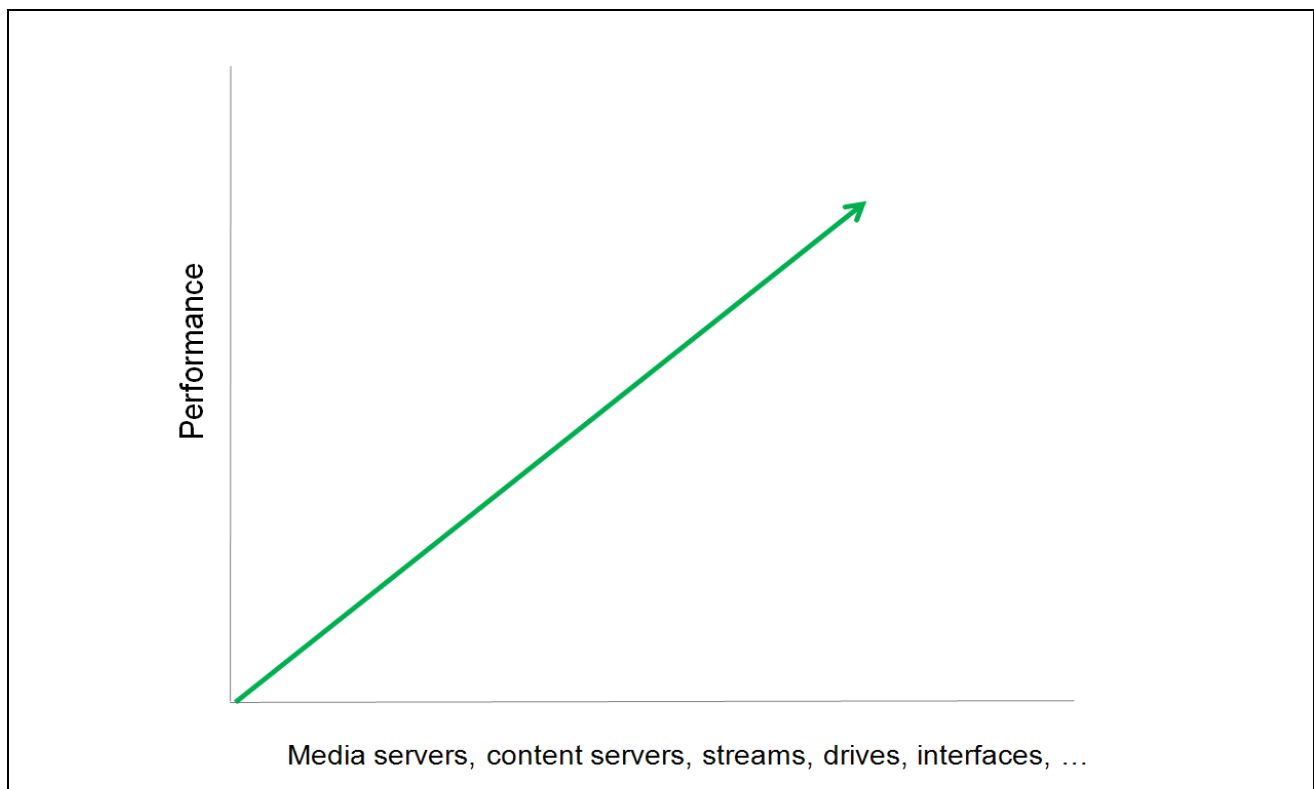
ESG Research found that 60% of early adopters of disk based backup solutions indicated ease of deployment as the single most important factor in purchasing a disk based solution. This is especially important for large, complex environments where backup policies span hundreds of physical and virtual servers running dozens of applications—stretching IT resources to the limit. ESG Lab has confirmed that PureDisk is extremely easy to configure and integrate into an enterprise NetBackup environment. ESG Lab was able to add PureDisk to an existing NetBackup installation, leveraging the existing infrastructure, policies and procedures in five minutes.

Performance

From a small remote office up to an enterprise-class data center with dozens of NetBackup servers connected to terabytes of SAN-attached disk pools, Symantec's software approach to data de-duplication can scale to meet a wide variety of performance requirements. Backup and restore performance scales as industry-standard servers and storage are added to a global pool of capacity-optimized storage.

PureDisk pools can be created using the latest server and storage technologies including DAS, NAS, FC- and iSCSI-attached disk arrays. The software-based PureDisk approach can also take advantage of future technology advancements (e.g., 8 Gbps FC). As the number of drives added to the global pool increases, the aggregate performance capabilities of the NetBackup PureDisk solution increases. Performance can also be increased by adding media servers, content router servers, streams, drives and network interfaces as shown in Figure 6.

FIGURE 6. SOFTWARE-ENABLED PERFORMANCE SCALABILITY



ESG Lab Testing

ESG Lab ran a series of tests to compare the performance of a traditional backup to disk solution to a NetBackup PureDisk solution. The exact same server full of direct-attached SAS drives was used as the backup target for each of these tests. In other words, the only difference between the "traditional" and "de-duplicated" results shown in Figure 7 and Table 1 is PureDisk data de-duplication.

FIGURE 7. PUREDISK DE-DUPLICATION OVERHEAD ANALYSIS

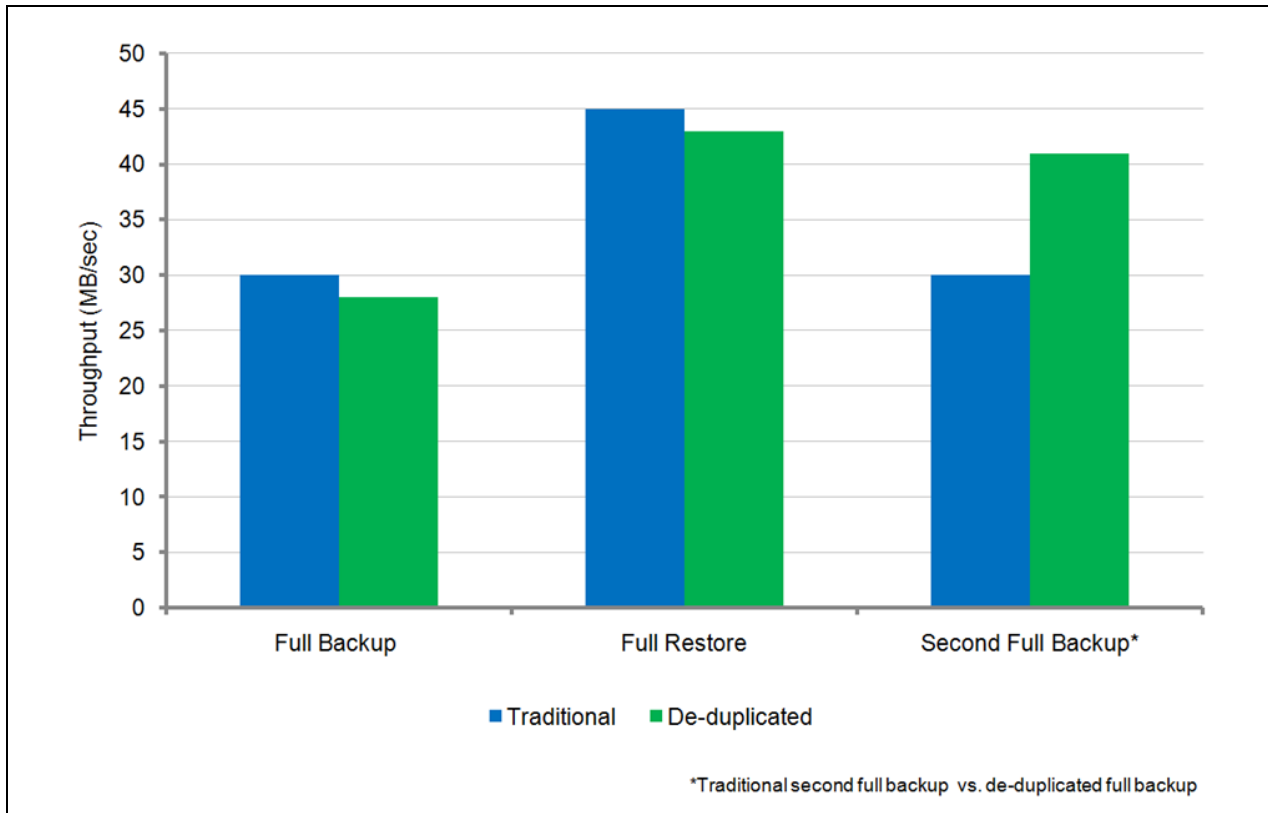


TABLE 1. PUREDISK DE-DUPLICATION OVERHEAD ANALYSIS

Backup Target	Operation	Data Set size (GB)	Throughput (MB/sec)
Traditional to staging disk	Initial Full backup	36	30
PureDisk de-duplication	Initial Full backup	36	28
Traditional	Full restore	12	45
PureDisk de-duplication	Full restore	12	43
Traditional	Second Full backup	12	30
PureDisk	Second De-duplicated full backup	12	41

What the Numbers Mean

- PureDisk provided 94% of the performance of a non-de-duplicated traditional disk-to-disk backup even as it identified and eliminated redundant data (a de-duplication rate of 8:1 was observed for the first full backup). This is a manageably small performance difference given the CPU intensive nature of data de-duplication processing.
- The difference in performance between full restores was also manageably low (less than 5%). Note that this minor difference in performance was measured as de-duplicated data was being reconstituted on the fly.
- It should be noted that the de-duplicated second full backup ran faster (41 vs. 30 MB/sec), and finished faster (five minutes vs. almost seven). The end result is a capacity-efficient full backup image for quick restores.

- The difference in performance between backup methods has some interesting implications. First, it should be noted that ESG Lab compared a traditional full backup to a second full backup as it was being de-duplicated on the fly. This was done to reflect the fact that a number of early adopters of data de-duplication technology are changing from a default policy of weekly full and daily incremental/differential backups to a daily full backup strategy. This is intended to provide the best of both worlds: only new data is written to disk, yet a full image is logically ready at all times for quick and reliable restores. The potential downside of a daily full strategy is the quantity of I/O required to move full backups across the LAN. In these cases, a de-duplicated differential backup strategy would be a better approach with similar storage benefits.

Having witnessed single client/single stream performance over the course of several backups and restores, ESG Lab audited a series of tests conducted by Symantec to determine the performance scalability of NetBackup PureDisk solutions. The number of PureDisk content routers, NetBackup media servers, client and streams were varied as indicated in Table 2. The PDDO option was installed and configured for inline data de-duplication during each of these tests. Configuration details including server and storage pool details can be found in the Appendix.

TABLE 2. BACKUP AND RESTORE PERFORMANCE SCALABILITY						
Backup						
Target	Operation	Content Routers	Media Servers	Clients	Streams	Throughput (MB/sec)
SAS	Full Backup	1	1	7	7	90
SAS	Full Backup	3	2	9	9	125
SAS	Incremental Backup	3	2	9	9	220
SAN	Full Backup	1	1	1	6	495
Restore						
SAS	Full Restore	1	1	1	2	48
SAS	Full Restore	3	3	1	3	60
SAS	Full Restore	1	1	1	4	76
SAN	Full Restore	1	1	1	3	110

What the Numbers Mean

- Backup throughput scales smoothly as clients, content routers and media servers are added.
- Using fast disk and SAN connectivity, a single content router can back up 1.7 TB per hour.
- Restore throughput scales smoothly as content routers and media servers are added.
- Using fast disk and SAN connectivity, a single content router can restore nearly 400 GB per hour through a single NetBackup Media Server.
- During the SAN Full Restore a single content router saturated the Gigabit Ethernet network between the Media Server and the restore target.

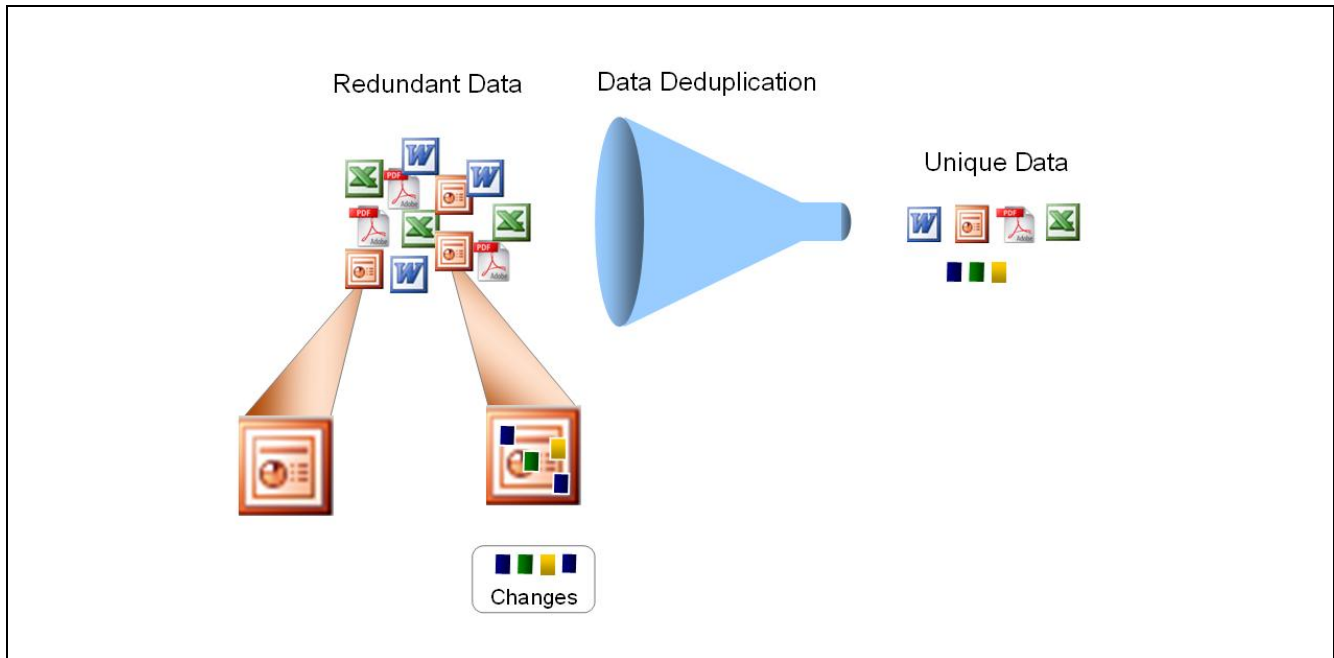
Why This Matters

ESG research indicates that performance is the top priority for data center managers considering a disk-based backup solution. Jobs that don't finish in an allotted backup window introduce risk and increase costs. PureDisk uses a scalable, software-based de-duplication approach to optimize backup and recovery performance. As an organization's data grows, the PureDisk approach scales backup and restore performance right along with it. Using a combination of hands on testing and audits, ESG Lab has confirmed that NetBackup PureDisk solutions deliver dramatic capacity and bandwidth savings with a manageably negligible performance impact. ESG Lab has also confirmed that NetBackup PureDisk performance scales as industry standard servers and storage are added to a global pool of de-duplicated disk capacity.

Data De-Duplication

Data de-duplication technology reduces the capacity required to store redundant data. To illustrate the power of data de-duplication technology, consider the example shown in Figure 8. Redundant copies of files (Word, PowerPoint, Excel and PDF) have been created. Data de-duplication recognizes and eliminates the redundant data and stores only the data blocks that make up the unique files.

FIGURE 8. BACKUP CAPACITY SAVINGS



Consider what happens when one of the copies of the PowerPoint file has been sent via e-mail to a number of team members. One of the recipients makes changes and forwards the updated presentation to the team, block level data de-duplication technology recognizes the changes and saves only the changed blocks. Pointers are used to recreate the file from the new and previously saved blocks. The capacity required for many copies of the same PowerPoint file, and the changed version, has been reduced significantly.

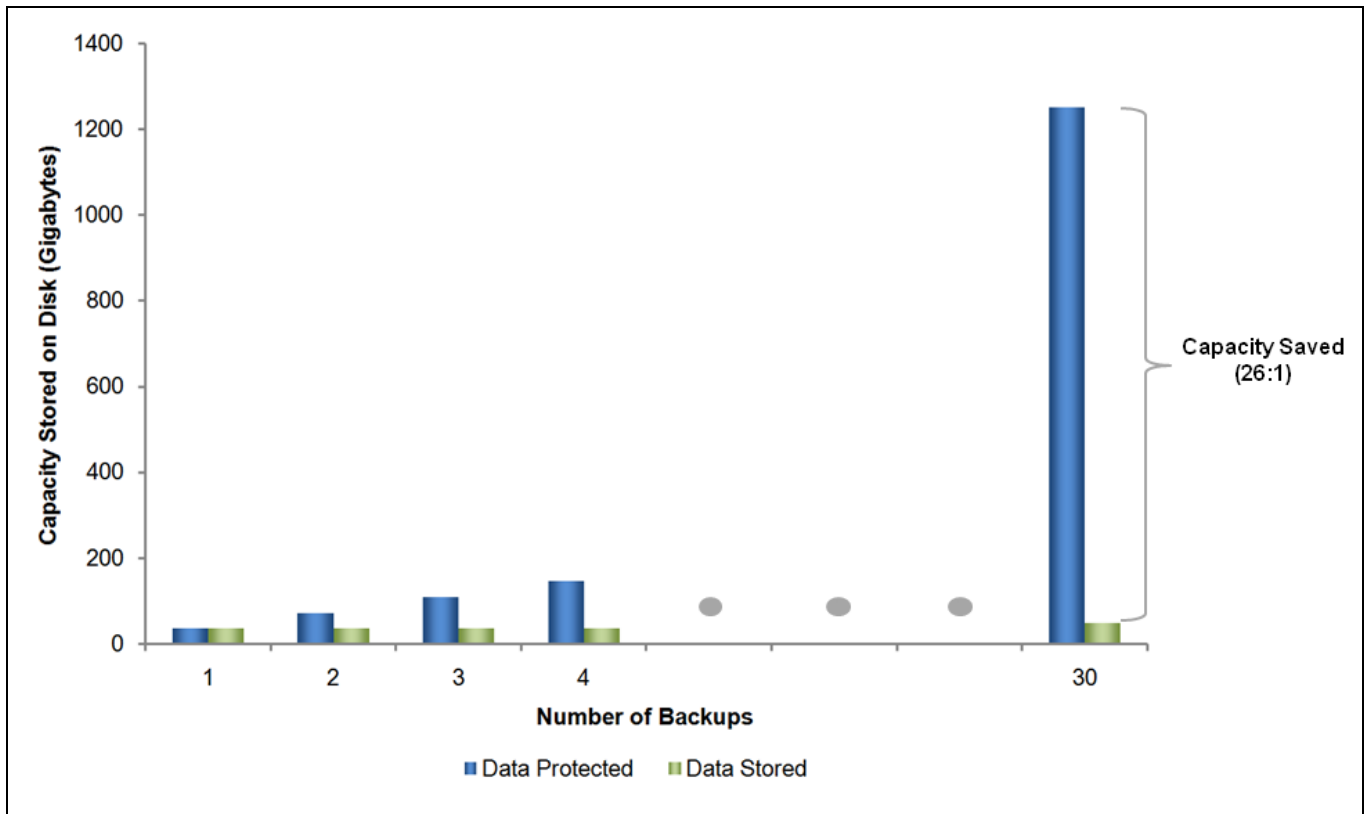
Block level data de-duplication technology has become popular in recent years due its ability to significantly reduce the capacity required to store backup data on disk. Block level de-duplication works well with backup data because a series of nightly backups tends to store copies of data that is mostly the same.

ESG Lab Testing

ESG Lab measured the capacity savings that can be achieved over a series of backup operations using Symantec PureDisk and NetBackup backup software. The initial data set backed up was composed of 36 GB of randomly generated file data. A Symantec-developed file creation utility was used to modify 10% of the data by 10% and grow the dataset by .5% to simulate changes between each full backup.

ESG Lab measured the capacity consumed before and after running each full backup. After four backup jobs had been run, the capacity savings for a 30 day retention policy were projected. The results are shown in Figure 9, with capacity savings reported by the PureDisk GUI interface for the first four backup jobs shown in Columns one through four. The projected de-duplication ratio for a month of full backups retained on PureDisk is approximately twenty-six to one.

FIGURE 9. BACKUP CAPACITY SAVINGS



What the Numbers Mean

- The total amount of data that NetBackup software would have sent to disk over the course of 30 full backups was calculated to be 1,252 GB. This capacity, shown in blue in Figure 9, is the amount of disk capacity that would be needed to retain a month’s worth of full backups on a backup to disk solution that didn’t support data de-duplication.
- PureDisk reduced the disk capacity from 1.2 terabytes (1,252 GB) to only 48 GB.
- The amount of capacity reduction that can be achieved with PureDisk will vary according to the backup policy in effect, the number of backups retained on disk and the type of data. In this scenario, capacity was reduced by a factor of twenty-six to one (26:1).

Why This Matters

ESG research indicates that cost is the leading obstacle to disk-based backup deployment. Data de-duplication changes the economics of backup to disk by reducing the amount of data retained on disk. ESG Lab testing confirmed that de-duplication can be used to reduce disk capacity by a factor of twenty-six to one for a daily full backup policy retained for 30 days with a one percent daily change rate. ESG Lab is confident that PureDisk can achieve de-duplication ratios of up to fifty to one depending on the type of data being backed up, the backup policies in effect and the number of backups retained on disk.

Disaster Recovery and High Availability

Symantec PureDisk leverages Veritas Cluster Server to provide continuous access to PureDisk storage pools. PureDisk also supports disaster recovery and offsite archival by replicating backup data over a wide area network to a remote data center. Symantec calls this feature Optimized Duplication. Data de-duplication and replication capabilities ensure that only changes since the last backup comprised of unique and optionally compressed data segments are replicated. This reduces the cost of network bandwidth and improves the performance of backups and replication over the WAN.

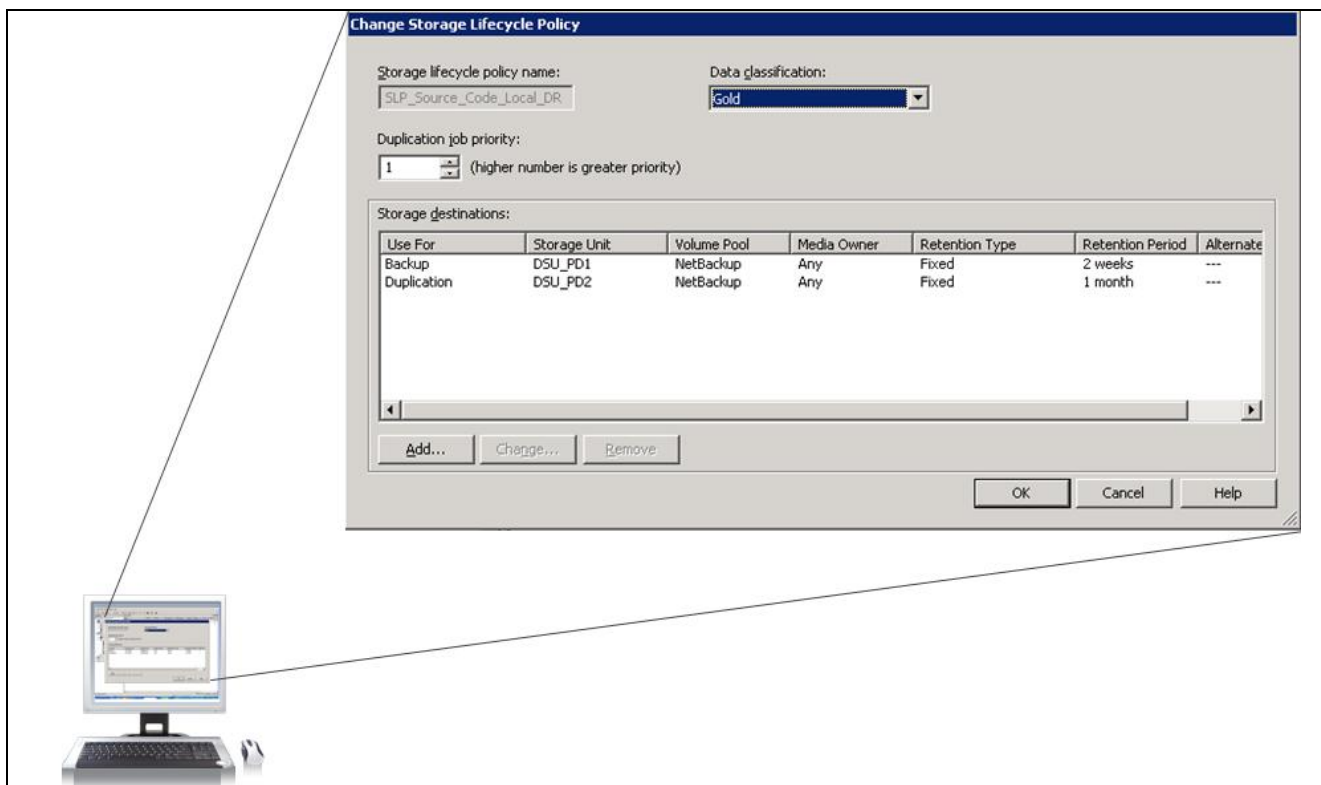
PureDisk's integration with NetBackup is enhanced by a feature called storage lifecycle policies. Using Storage lifecycle policies (first available with NetBackup 6.5) enables NetBackup to create a template to track multiple backup copies in its catalog and to manage different retention times for each copy. This differs from a traditional backup replication process which ties expiration of a replica to the original data.

Because a storage lifecycle policy behaves like a template, it can be applied to many different NetBackup policies. If a storage or retention plan changes (e.g. a new regulation is imposed requiring changes to retention periods or the number of copies created), only the relatively small number of Storage Lifecycle Policies need to be changed, as opposed to all backup policies.

ESG Lab Testing

ESG Lab configured a second PureDisk Pool as a simulated Disaster Recovery site. A local LAN was used to emulate a WAN connection between the local data center and the DR site. As seen in Figure 10, storage lifecycle policies were created to retain local backups for two weeks and include the duplication of data to a remote site with a one month retention period.

FIGURE 10. STORAGE LIFECYCLE POLICIES



Next, backups were conducted at the local site. After these completed, the optimized duplication process ran and transferred the globally unique data to the remote site as seen in Figure 11. No performance difference was detected.

FIGURE 11. PUREDISK OPTIMIZED DUPLICATION

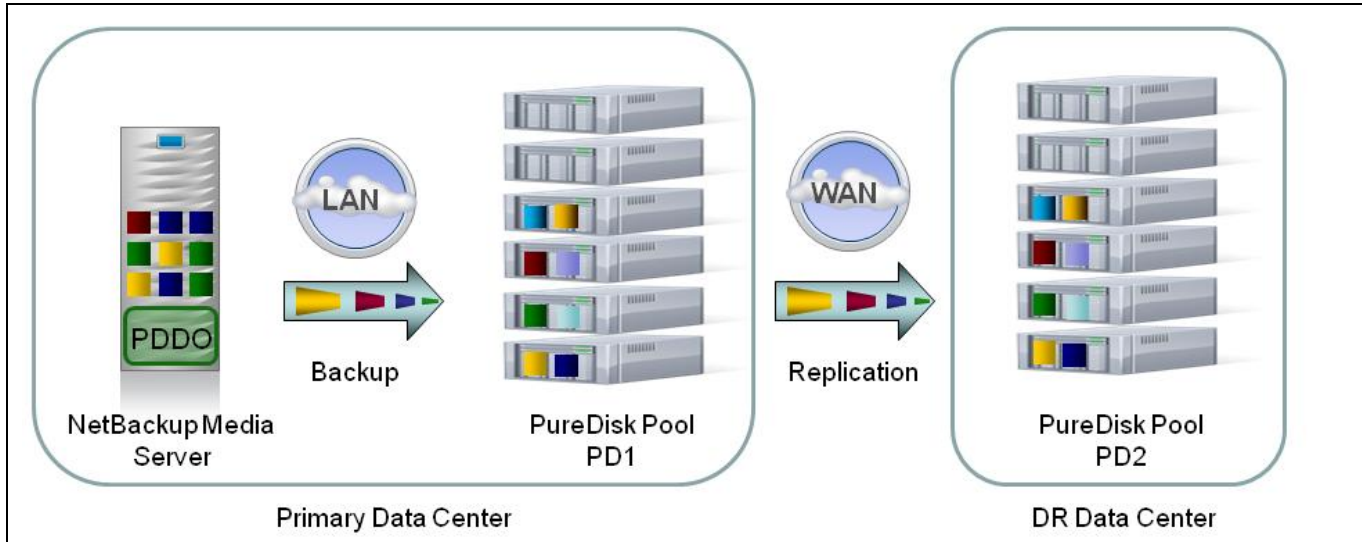


TABLE 3. REPLICATION PERFORMANCE

Replication	Data set	Elapsed Time	Effective Throughput	De-duplication Ratio
First off site copy	12 GB File Data	6:03	33.8 MB/Sec	33.85% (2.95:1)
De-duplicated second copy (1% change)	12 GB File Data	0:37	332 MB/Sec	99% (100:1)

What the Numbers Mean

- The first full replication to the DR site was moderately de-duplicated and transferred the equivalent of a full 12 GB backup by moving 4 GB of data in just over six minutes. When protecting multiple, similar clients, users could expect the initial de-duplication ratios and effective throughput to be higher as there will be more common data to eliminate.
- The second replication to the DR site, after one day’s worth of changes were made to the test dataset, transferred the equivalent of a full 12 GB backup by moving only 125 MB of data in just 37 seconds.

Once the data was replicated to the remote site, ESG Lab used NetBackup to perform a recovery operation. Comparison of the recovered data showed that the files were identical to the original data at the local site.

ESG Lab also saw PureDisk running in a four node VCS (Veritas Cluster Server) cluster. All four nodes were virtual machines under VMware ESX. No special VCS setup or configuration is required. ESG Lab added a node to the cluster with just three clicks and observed a node failover with minimal impact. The failover took about two minutes and the backup jobs running at the time of the failover were terminated. As the failed backup jobs were restarted and new backup jobs were initiated, the replacement node handled them automatically and transparently, having taken over the complete identity of the failed node.

Why This Matters

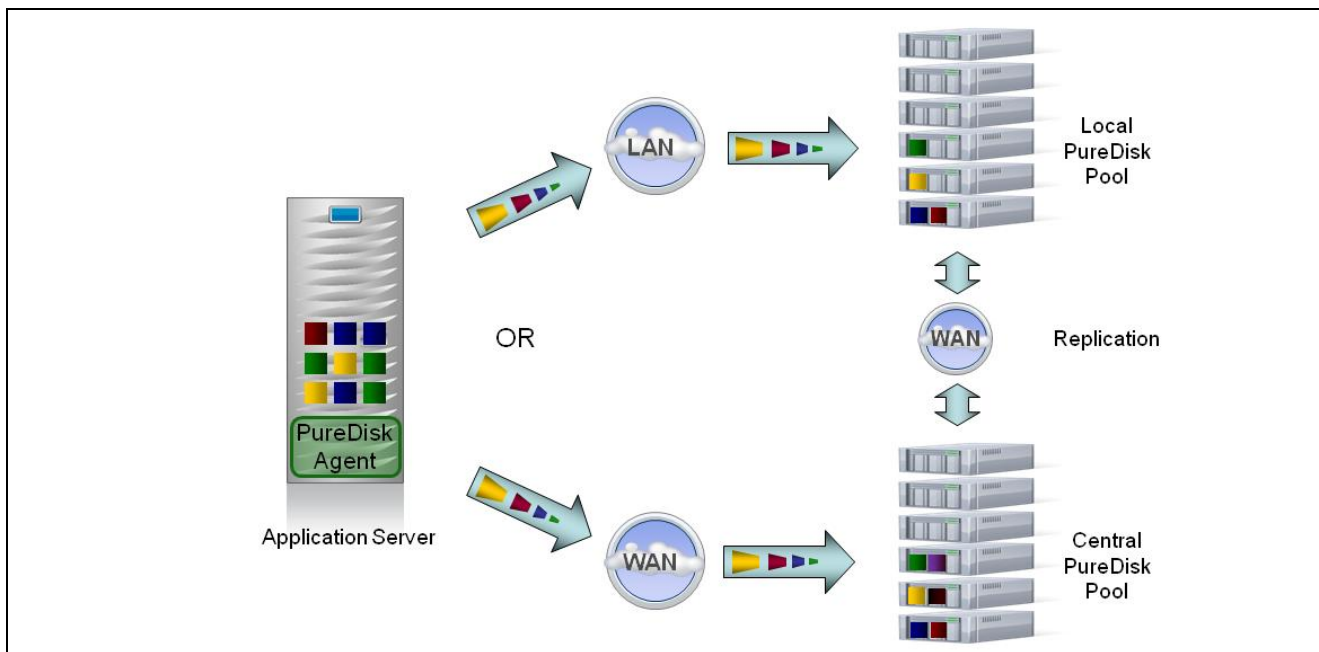
Offsite copies of backup data are needed to ensure that an organization can recover from a disaster. Organizations with large amounts of backup data protected by legacy backup and recovery solutions haven't been able to afford to make offsite copies of backup data electronically due to the high cost of WAN bandwidth. ESG research³ indicates that 79% of organizations cannot tolerate more than 12 hours of downtime for their most critical applications; restoring from tapes that have been moved offsite can take 24 hours or more as tapes must be shipped from the storage center before restores can even begin.

ESG Lab has verified that PureDisk data de-duplication significantly reduces the amount of data that needs to be transferred over the WAN, enabling DR without requiring prohibitively costly network upgrades. ESG also confirmed that PureDisk delivers enterprise class fault tolerance using HA clustering, providing seamless failover protection for the PureDisk environment.

Remote Office Protection

Remote and branch offices (ROBOs) house a large portion of enterprise data, yet most still rely upon tape-based backup methods managed by non-IT staff. Non-IT staff performing backups and handling tapes increases both the risk of data loss and the cost of data protection, while backups, snapshots, staging and archiving all significantly increase the amount of storage under management. In a distributed environment, existing network capacity makes it impractical to centralize backup operations with traditional backup architectures. Moving just daily incremental backup data sets across the WAN requires so much bandwidth and time that even this simple process can quickly become prohibitively expensive and inefficient.

FIGURE 11. REMOTE AND BRANCH OFFICE DE-DUPLICATION



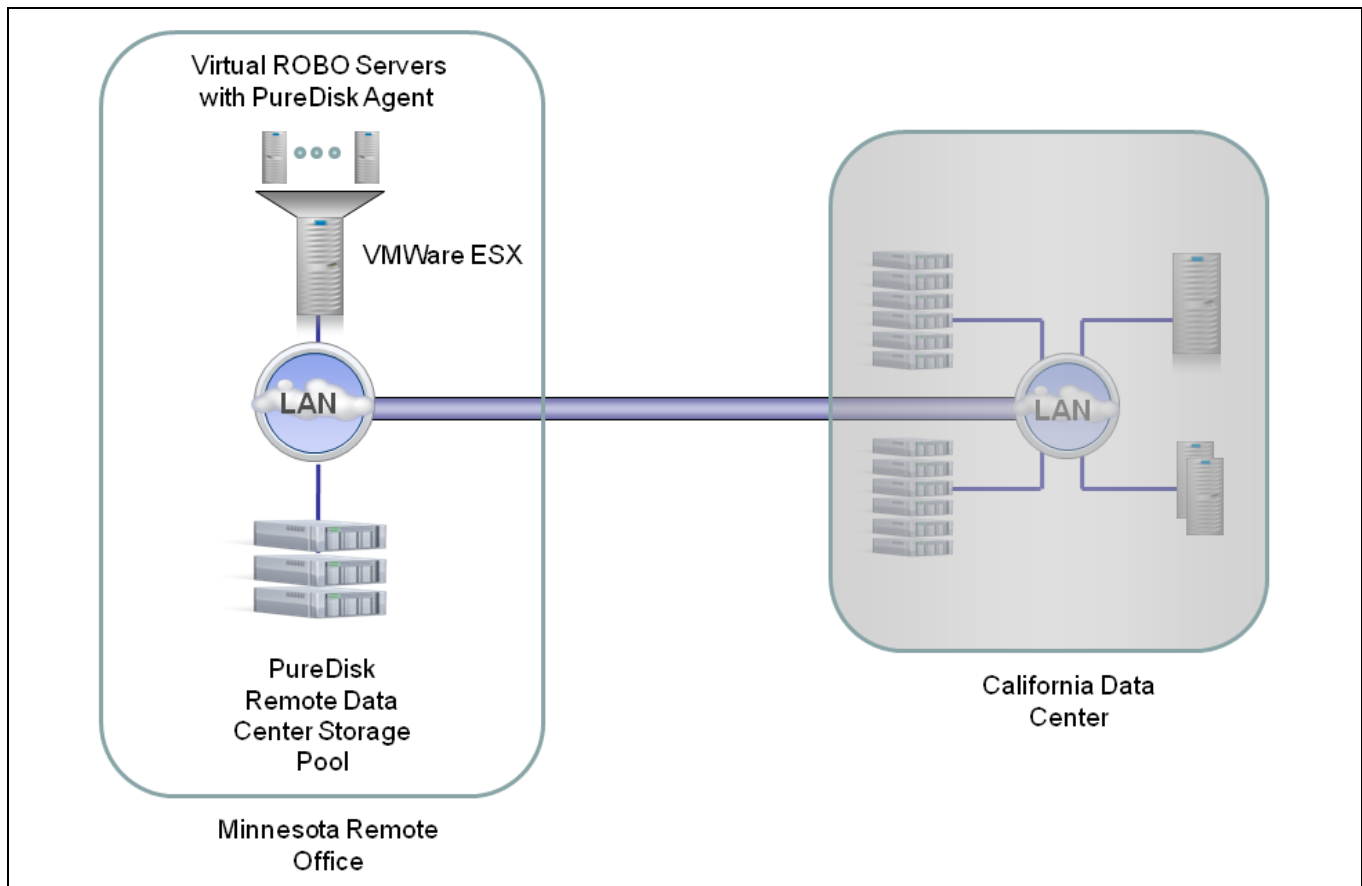
³ Source: ESG Research Report, *Data Protection Survey*, October, 2007

Whether remote and branch offices using PureDisk utilize a local PureDisk Pool or back up directly to the central data center, PureDisk reduces the amount of data being transferred during backup by sending only unique blocks of data to the PureDisk Pool. At smaller remote offices, only PureDisk agents are deployed on the systems, while larger remote offices and data centers typically deploy a local PureDisk server to improve replication and recovery performance.

ESG Lab Testing

A real remote office in Minneapolis was configured to test remote office backup to a corporate data center (see Figure 12). An 8 Mbps wide area network connection linked the two sites during this phase of testing. In this test, ESG Lab used PureDisk to protect a virtualized Microsoft Windows file server running under VMWare ESX at the remote office.

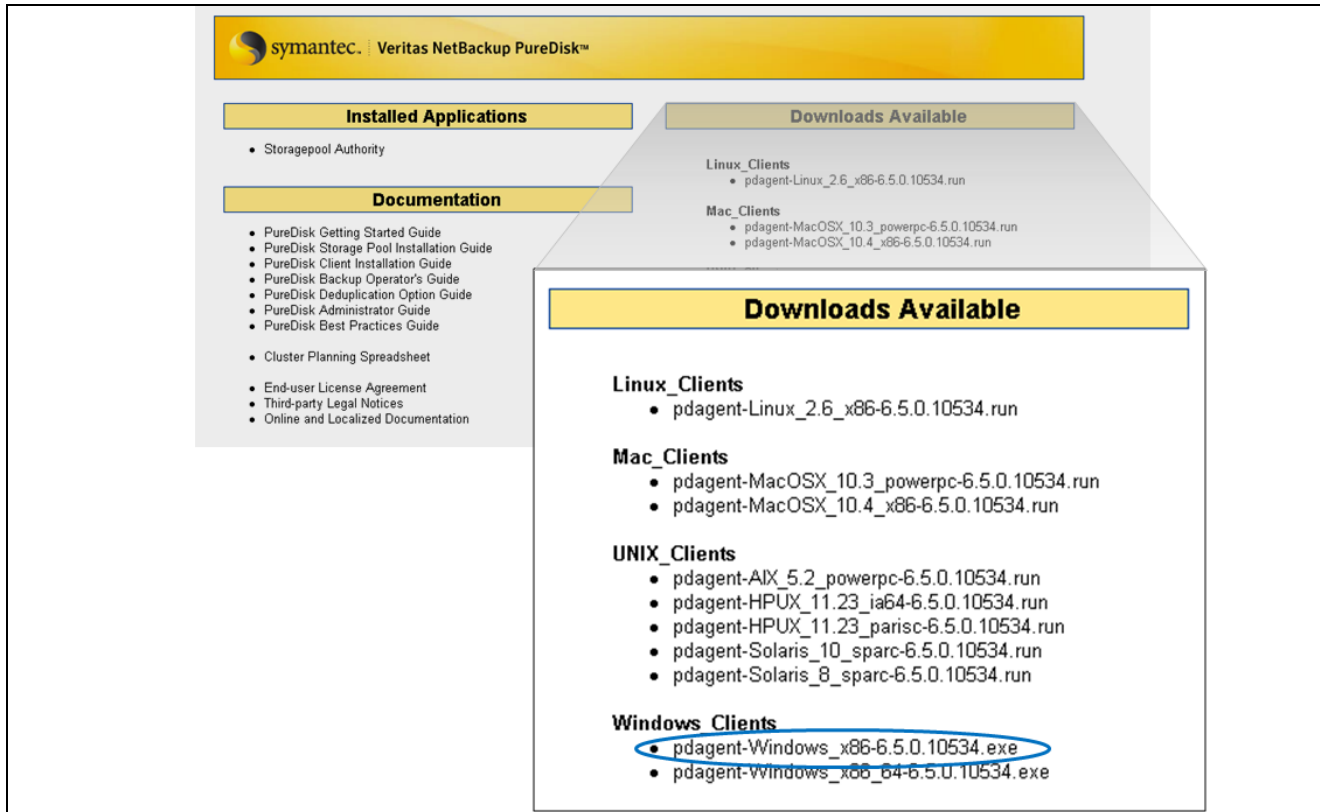
FIGURE 12. PUREDISK IN THE REMOTE OFFICE



Testing began by installing the PureDisk client on the remote Windows server. ESG Lab used a remote desktop connection to the Windows server and navigated to the PureDisk web administration page. Next, ESG Lab selected the appropriate PureDisk Client for the machine under test as seen in Figure 13, downloaded the executable and began the installation.

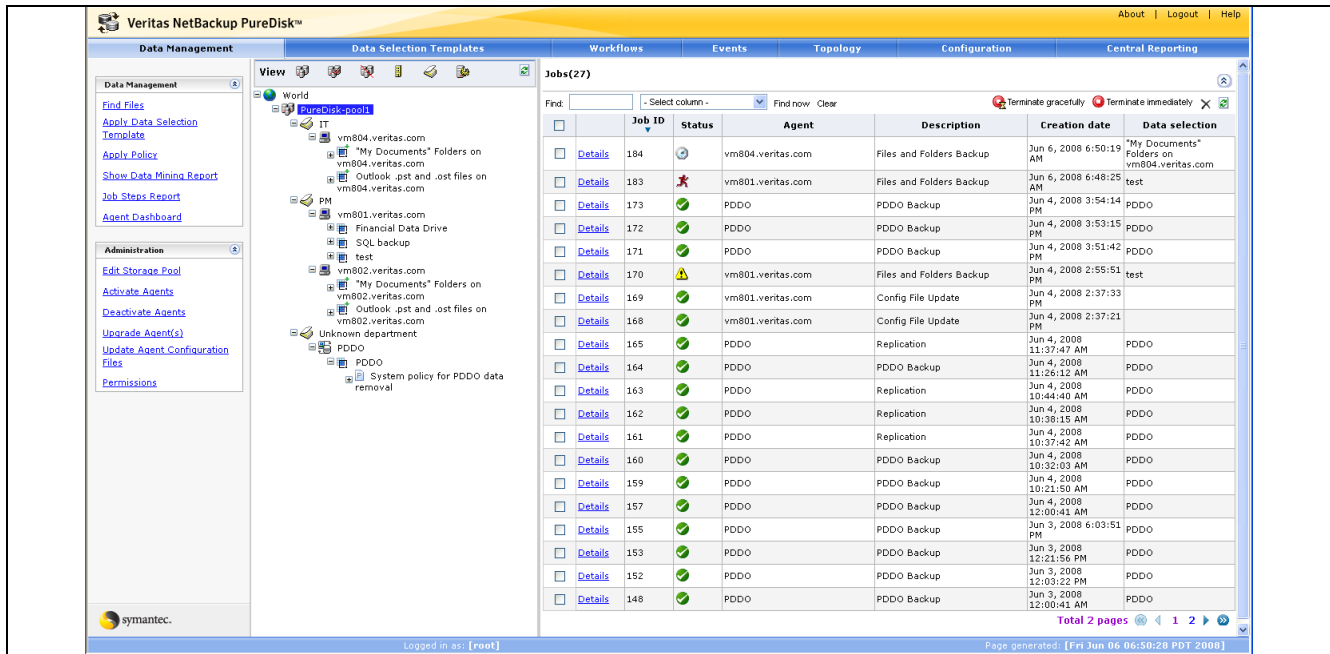
The installation had the familiar look and feel of a Windows InstallShield wizard. When prompted, ESG Lab entered the address and login information for the Storage Pool and let the installation complete. Seven clicks and sixty seconds later, the installation was complete.

FIGURE 13. INSTALLING THE PUREDISK CLIENT



Within less than five minutes of the first click to download the client, ESG Lab had installed the client, configured and started a manual backup of the operating system files of the server.

FIGURE 14. THE PUREDISK MANAGEMENT CONSOLE

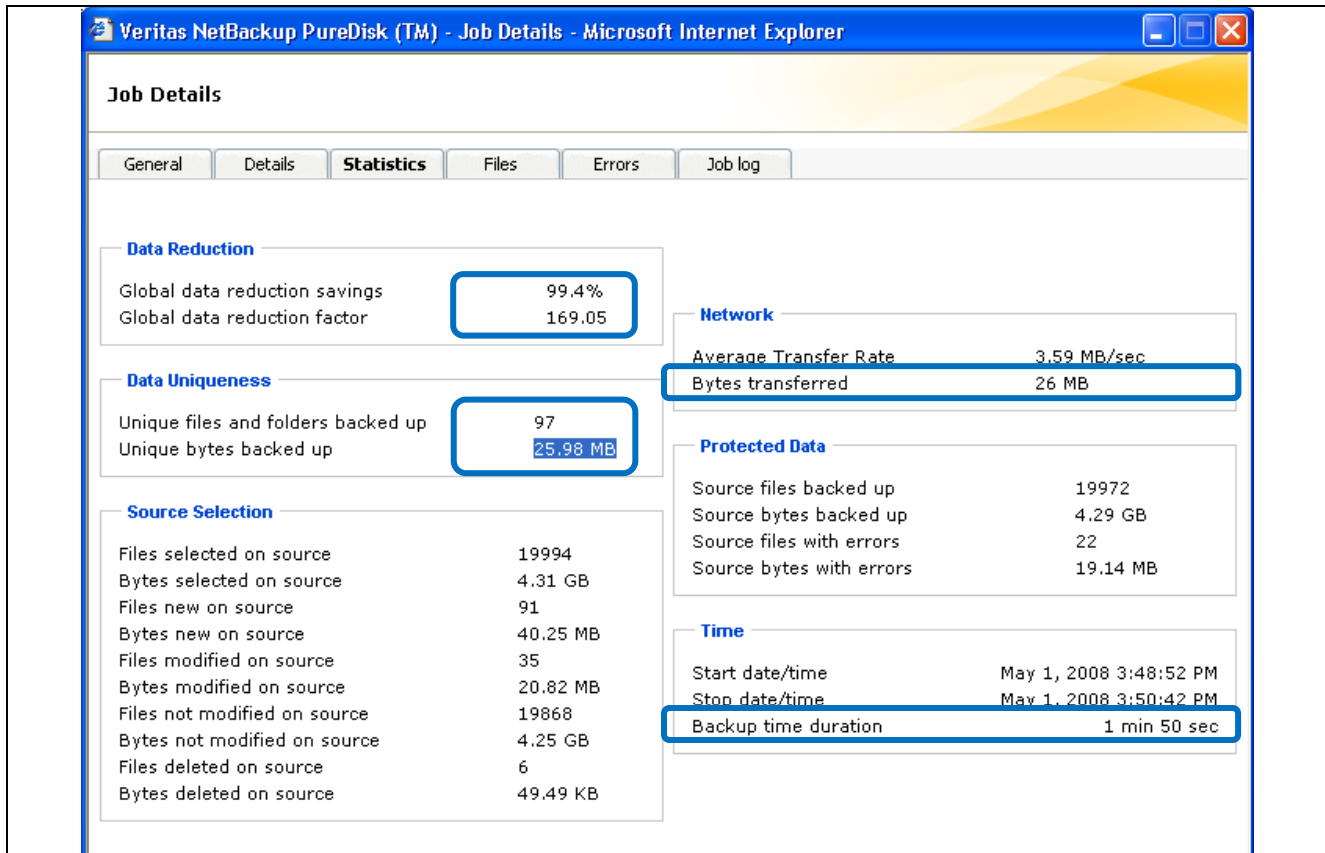


ESG Lab found that managing backups and restores using the PureDisk management console to be easy and intuitive. As shown in Figure 14, the backup history for all jobs appears in a clear list format in a pane on the right of the screen. A directory tree interface that looks and feels a lot like Windows Explorer was used to monitor and manage the individual components of the PureDisk system.

Two backup jobs were run with the remote client backing up to the PureDisk Pool in the central data center: an initial full backup and a de-duplicated full backup after additions and changes were applied to the data set.

Once the first full backup was complete, ESG Lab examined the logs to determine how much data was transferred between the PureDisk client in the remote office and the PureDisk Storage Pool in the Central Data Center. The initial backup showed a De-duplication rate of 65%, with 1.4 GB of the total 4.27 GB being transferred across the link in 42 minutes.

FIGURE 15. BACKUP JOB DETAILS



Prior to running the second job, 1% of the total data set (4.3 GB) was added in the form of 40 MB worth of randomly generated new files using a random file generation utility provided by Symantec. In addition, 35 existing files were modified; 20 MB of data was added to them, simulating another 1/2% of change to the data set being backed up. As seen in Figure 15, PureDisk completed this full backup in just under two minutes, transferring just 26 MB of de-duplicated data and metadata across the WAN.

Why This Matters

Much of the information critical to the success and efficiency of an organization resides at remote and branch offices. For example, ESG Research⁴ found that 61% of remote and branch offices house e-mail servers and data that needs to be protected. Protecting remote offices using traditional backup methods is often costly and risky—especially when performed by non-technical local staff. Replicating data over a WAN instead of shipping tapes reduces the risk of human error, but can be cost prohibitive because of the significant amount of data that needs to be transferred across the network.

ESG Lab has confirmed that PureDisk source-based data de-duplication significantly reduces the amount of data sent over the WAN, enabling fast, daily full backups for remote office data. Backups are centrally managed from a corporate data center using the intuitive PureDisk Management interface. Both data and metadata can be encrypted for added security.

⁴ Source: ESG Research Report, *Branch Office Optimization Survey*, January, 2007

ESG Lab Validation Highlights

- ☑ PureDisk inline data de-duplication performed backups and restores as quickly and efficiently as standard disk without de-duplication.
- ☑ Performance for backups and restores scaled smoothly to server and network limits by adding components to the cluster (NetBackup media servers and/or content routers) with PureDisk never becoming the bottleneck.
- ☑ ESG Lab observed PureDisk optimized duplication drastically reduce the amount of data required to transmit full backups across the WAN.
- ☑ ESG Lab noted the ease with which NetBackup storage lifecycle policies could be combined with PureDisk as a way to manage multiple copies of data on disk with different retention times.
- ☑ With its industry standard Linux operating system, PureDisk can be run in a Virtual Server environment and can be protected by Veritas Cluster Server software.
- ☑ Remote office backups across the WAN were easy to configure and ran very efficiently, making excellent use of global data de-duplication.

Issues to Consider

- ☑ There are distinct differences between hardware- and software-based D2D de-duplication products; a software solution—such as NetBackup PureDisk—can offer more flexibility, including the type of disk storage hardware and the size and scale of the solution over time. With flexibility can come complexity—more options can equal more moving parts. When PureDisk is deployed in typical user configurations such as distributed, data center and virtual server environments, the benefits versus those of hardware-based solutions become evident.
- ☑ PureDisk offers users the option of inline data de-duplication which occurs as the backup is taking place or post-process de-duplication. Symantec offers two options for running a post-process de-duplication job. With Storage Lifecycle Policies the job will be run immediately after a backup job completes. Customers can also use NetBackup Vault to schedule a de-duplication of a traditional backup image. Only the former (Storage Lifecycle Policies) was reviewed for this report. ESG Lab believes that incorporating the ability to schedule de-duplication jobs as part of Storage Lifecycle policies would be a useful enhancement. Symantec indicated that it had several improvements planned for Storage Lifecycle policies in the future, but could not elaborate on the nature or exact timing of these roadmap items.
- ☑ While ESG was able to generate a static report showing global de-duplication ratios over time, a summary of aggregate, enterprise wide statistics such as data stored, data protected and overall de-duplication ratio in a graphical depiction would be a great addition to the PureDisk administration interface. Currently, graphical reporting and trending are available to users who install Veritas Backup Reporter with the PureDisk product.

ESG Lab's View

Data de-duplication is a game-changing technology that drastically reduces the cost of backup to disk and disaster recovery. The first solutions to come to market supporting data de-duplication were delivered as appliances that could be used as targets for backup jobs. These appliances implement data de-duplication at the end of the wire as data is stored on disk. Data de-duplication has also been implemented within backup software solutions at the other end of the wire for WAN-efficient protection of remote and branch office data. Early solutions in this category required organizations to rip and replace their existing backup software and processes. PureDisk is a flexible, software-based solution that can be implemented at either end of the wire while preserving existing investments in NetBackup—a market leading backup and recovery solution.

Data de-duplication can be implemented inline as data is backed up, or as a post-processing task after backup jobs have completed. Each approach has merits, depending on the amount of data that needs to be backed up nightly. The software-based NetBackup PureDisk approach provides the flexibility to implement either approach—or both—depending on the backup window requirements within a data center, branch or remote office.

Scalability, performance and high availability are key concerns when deploying an enterprise-wide data protection solution. NetBackup PureDisk uses a modular, software-based approach that leverages industry standard server, storage and clustering technology to provide a solution that can scale to meet the data protection needs of organizations of any size.

ESG lab was impressed with the power and flexibility of the NetBackup PureDisk approach to data protection. Inline data de-duplication integrated seamlessly with NetBackup with negligible impact to backup performance and processes. PureDisk in remote offices enabled efficient, secure backups and restores both locally and across the WAN all managed by centrally located IT staff.

A single software-based data protection platform with integrated global data de-duplication has a number of compelling advantages, but the bottom line with a NetBackup PureDisk solution is clearly cost savings. Delivering disk capacity savings of up to 50 times, WAN bandwidth savings of up to 500 times, and remote office operational savings up to five times, ESG is confident that organizations can realize significant savings with a NetBackup PureDisk solution.

Appendix

Table 1. DATA CENTER PROTECTION TEST CONFIGURATION	
PureDisk Servers: - 2x Dell 2950 servers, 8GB RAM Storage: - 2 x 4TB Dell MD1000 array with 3.5-inch 7200 RPM SATA drives	PureDisk OS 6.5
Clustered PureDisk environment 4 nodes: 3 active / 1 passive - 8x Windows 2003 Virtual Machines 2GB OS Data and 5-10GB File Data - 2 x DELL 2950 servers	VMWare ESX 3.5 PureDisk 6.5 with Veritas Cluster Server 4.1 MP3 embedded Windows 2003 Enterprise Server SP1
NetBackup Windows 2003 Master/Media Server: - 2x Dell 2950 8GB RAM, 1.5 TB Internal 15K RPM SAS Storage for staging area	NetBackup 6.5.1 - 1 Master/Media server on Windows Enterprise Server SP1 - 1 Media server on Linux Enterprise Server (SLES 10-SP1)
Data Center LAN 1Gb switched Ethernet	

Table 2. PERFORMANCE TEST CONFIGURATION	
PureDisk Servers: - Dell 2950 – 8GB RAM -2Ghz (5130) Storage: - 4 GB/s Fiber attached SAN, ADTX and Winchester arrays. - 20TB total storage capacity	PureDisk OS 6.5
Clients: 10 physical clients Dataset: Randomly generated data	Windows 2003 Enterprise Server SP1 Solaris 10
NetBackup Master/Media Servers: - Dell 2950 – 4GB RAM – 3Ghz - Dell 2950 – 4GB RAM – 3Ghz - SUN T5120 - 8GB RAM	NetBackup 6.5.1 - Master/Media on Windows 2003 Enterprise Server SP1 - Media server on Linux SLES 10 SP1 - Media server on Solaris 10
LAN/SAN connectivity: - LAN: 1 NICs with 1 Gb/s - SAN: 4Gb/s	

1 PureDisk Server: <ul style="list-style-type: none">- Sun x4100, 4GB RAM, 1TB Internal SATA Storage	PureDisk 6.5
2 Windows 2003 Virtual Machines under ESX 3.5 5GB OS Data and 5-10GB File Data	Windows 2003 Enterprise Server SP1
WAN Connection	8Mb/s



20 Asylum Street
Milford, MA 01757
Tel: 508-482-0188
Fax: 508-482-0218

www.enterprisestrategygroup.com