Analysis: Minimizing System Risk via End-to-End Backup and Recovery Process Automation
Analysis: Minimizing Virtual System Risk via End-to-End Backup and Recovery Process Automation

Author: Jack Fegreus, Ph.D.
Managing Director
openBench Labs
http://www.openBench.com
September 10, 2009

Jack Fegreus is Managing Director of openBench Labs, consults through Ridgetop Research, and is on the Board of Advisors of IT Brand Pulse. He currently contributes to InfoStor and Virtualization Strategy Magazine. He has served as Editor in Chief of Open Magazine, Data Storage, BackOffice CTO, Client/Server Today, and Digital Review. Previously Jack served as a consultant to Demax Software, CTO of Strategic Communications, and IT Director at Riley Stoker Corp. Jack holds a Ph.D. in mathematics and worked on the application of computers to symbolic logic.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive Summary</td>
<td>04</td>
</tr>
<tr>
<td>Data Protection Test Bed</td>
<td>07</td>
</tr>
<tr>
<td>Policy-Driven Processes</td>
<td>12</td>
</tr>
<tr>
<td>Process Performance</td>
<td>17</td>
</tr>
<tr>
<td>Data Deduplication Cache</td>
<td>22</td>
</tr>
<tr>
<td>Data Protection Process Value</td>
<td>25</td>
</tr>
</tbody>
</table>
In a turbulent economic environment, the need to establish highly efficient operations is a priority for all enterprises. For IT, the path to higher efficiency begins with consolidating resource management and optimizing resource utilization. At a time when more processors, greater storage volume, and expanding portfolios of applications combine to drive greater complexity and higher labor costs, CIOs are responding with a two-pronged strategy that focuses on managing IT as a service within a virtual operating environment, such as that created by VMware® ESX Server.

That strategy is based on two complementary solutions. First, working with virtual resources allows system administrators to focus on a limited number of abstract device pools that can be centrally managed. For small-to-medium enterprise (SME) sites, a virtual operating environment is an ideal platform on which to scale out applications and take advantage of the enhanced reliability, availability, and serviceability characteristics typical of large data centers.

The road to nirvana, however, is not without some significant complications. Convergence of backup and disaster recovery technologies, growth of server virtualization, and expansion of government regulations on risk management create a maelstrom for IT departments. Now, ten, twenty, or more virtual servers may depend on a single physical host. Worse yet, different physical hosts may use different types of storage connectivity, which increases the complexity of data and system recovery.

Executive Summary

“To give IT a vital edge in Service Level Agreement (SLA) compliance, NetBackup provides a completely hardware-independent policy-based backup and restore framework to implement storage efficient backup as a robust end-to-end automated process.”

In a turbulent economic environment, the need to establish highly efficient operations is a priority for all enterprises. For IT, the path to higher efficiency begins with consolidating resource management and optimizing resource utilization. At a time when more processors, greater storage volume, and expanding portfolios of applications combine to drive greater complexity and higher labor costs, CIOs are responding with a two-pronged strategy that focuses on managing IT as a service within a virtual operating environment, such as that created by VMware® ESX Server.

That strategy is based on two complementary solutions. First, working with virtual resources allows system administrators to focus on a limited number of abstract device pools that can be centrally managed. For small-to-medium enterprise (SME) sites, a virtual operating environment is an ideal platform on which to scale out applications and take advantage of the enhanced reliability, availability, and serviceability characteristics typical of large data centers.

The road to nirvana, however, is not without some significant complications. Convergence of backup and disaster recovery technologies, growth of server virtualization, and expansion of government regulations on risk management create a maelstrom for IT departments. Now, ten, twenty, or more virtual servers may depend on a single physical host. Worse yet, different physical hosts may use different types of storage connectivity, which increases the complexity of data and system recovery.

Executive Summary

“To give IT a vital edge in Service Level Agreement (SLA) compliance, NetBackup provides a completely hardware-independent policy-based backup and restore framework to implement storage efficient backup as a robust end-to-end automated process.”

In a turbulent economic environment, the need to establish highly efficient operations is a priority for all enterprises. For IT, the path to higher efficiency begins with consolidating resource management and optimizing resource utilization. At a time when more processors, greater storage volume, and expanding portfolios of applications combine to drive greater complexity and higher labor costs, CIOs are responding with a two-pronged strategy that focuses on managing IT as a service within a virtual operating environment, such as that created by VMware® ESX Server.

That strategy is based on two complementary solutions. First, working with virtual resources allows system administrators to focus on a limited number of abstract device pools that can be centrally managed. For small-to-medium enterprise (SME) sites, a virtual operating environment is an ideal platform on which to scale out applications and take advantage of the enhanced reliability, availability, and serviceability characteristics typical of large data centers.

The road to nirvana, however, is not without some significant complications. Convergence of backup and disaster recovery technologies, growth of server virtualization, and expansion of government regulations on risk management create a maelstrom for IT departments. Now, ten, twenty, or more virtual servers may depend on a single physical host. Worse yet, different physical hosts may use different types of storage connectivity, which increases the complexity of data and system recovery.
As a result, a single ESX server host is a high availability issue as the risk of its failing cascades to multiple virtual machines (VMs) running multiple applications. This complicates a number of important IT operations, including the once simple notion of backup, which is rapidly becoming inextricably intertwined with disaster recovery. This puts backup and recovery among the biggest challenges that IT organizations must resolve when implementing a virtual operating environment.

In a virtual operating environment, IT system administrators are immediately confronted by a daunting question: What should be backed up? Should system administrators concentrate their efforts on the logically exposed virtual machines (VMs) running important business applications. Or should they focus on the virtual operating environment applications and files that create those logical VMs. At the heart of this problem is the fact that a VM has two distinct personas. First, there is the IT-centric persona of a virtual operating environment application that needs to run on a virtual operating environment server and migrate to another virtual operating environment servers to satisfy load balancing or business continuity issues. Second, there is the logical line-of-business persona of a VM as a standard computer system.

That dichotomy in VM perception has the potential to disrupt IT operations. To resolve those perceptual issues, Symantec NetBackup employs patented technology that integrates deep into VMware infrastructures to leverage the VMware Consolidated Backup (VCB) API, the vStorage API, and vCenter Server (a.k.a. Virtual Center). Using NetBackup, IT administrators have the ability to dynamically restore the backup set associated with a VM running a version of Windows Server to either its native ESX Server state—using vmdk and other VMFS files—or as a logical Windows system with NTFS formatted files. As a result, NetBackup is able to provide the data protection and disaster recovery support needed for an enterprise-class VMware virtual operating environment.

More importantly, all NetBackup data protection processes fit perfectly into any IT service management initiative. Consistently characterized as understaffed, underfunded and very busy, IT departments have long tried to improve productivity by acquiring software tools designed to resolve ad hoc resource problems. That “better firefighting” approach, however, was destined for failure from inception. Studies consistently cite changes in the computing environment that were introduced by IT as the cause of 80 percent of all computing disruptions.

Given that statistic, the goal of IT service management is to automate the standard tasks of systems and storage administrators by building upon classic quality-control (QC) practices for process management—schemes highly favored by CEOs and line of business executives. While the adoption of QC process management constructs by IT has the potential to eliminate the 80 percent of problem incidents attributable to IT, there is a significant hurdle to that adoption. The lion's share of the software used by IT greatly complicates any application of process management.

Current IT software tools are almost universally device-centric. As a result, every
storage device has its own data recovery services, which are managed through the device’s GUI and are replete with their own logins and passwords. In such an environment, the certification of backup procedures for risk management becomes a daunting task that often has to span multiple technologies. Worse yet, that scenario runs counter to all notions of quality control, which attempts to define work flows across generic devices to improve process management. Symantec NetBackup, on the other hand, provides IT with a data protection platform that consolidates the plethora of device-centric agents, policies, and reporting tools into a cohesive management framework.

To give IT a vital edge in Service Level Agreement (SLA) compliance, NetBackup provides a completely hardware-independent policy-based backup and restore framework to implement backup as a robust end-to-end automated process, which extends to the NetBackup PureDisk™ data deduplication environment. NetBackup even provides for quality control and process improvement via reporting tools that allow IT administrators to define and monitor service level compliance for SLAs entered into with line of business managers.
NetBackup offers IT administrators a powerful combination of policies and templates that allow them to quickly assign a storage life cycle for backup data—from creation to expiration—on all storage resources across physical locations.

For a better perspective on the ability of NetBackup to enhance critical virtual operating environment and IT service management initiatives, openBench Labs set up a data protection test scenario using VMware Virtual Infrastructure. We focused our tests on eight hosted VMs that were configured as application servers running Windows Server 2003 along with SQL Server and IIS.

To host the eight VMs, we utilized a quad-processor HP ProLiant®
DL580 server with 8GB of memory. To configure and manage the overall virtual operating environment, we used a dual-processor HP ProLiant DL360 running Windows Server 2003 and vCenter Server.

IT backup loads are dependent on a number of factors, including data retention requirements and the nature of the data in terms of compressibility and redundancy. This makes a virtual operating environment the perfect microcosm to examine all factors impacting backup load processing. Of particular importance is the growing practice of IT, at small to large enterprises alike, to use multiple VMs to establish the system availability and scalability that is characteristic of a large data center. This practice of utilizing multiple VMs, each dedicated to running a particular application, generates a prodigious amount of duplicate data within a virtual operating environment.

To support all of NetBackup's data protection services, including data deduplication via NetBackup PureDisk, openBench Labs configured three additional servers. The first server, a dual-processor HP ProLiant DL360, ran Windows Server 2003 and functioned as the NetBackup Enterprise Master Server. The Master Server maintains the NetBackup catalog of internal databases, handles the creation of data protection policies, manages device and media selection, and can also be utilized as a media server.

Given all of the focus surrounding IT service management as a way to institute better quality control within IT operations, it is important to note that NetBackup requires system and storage administrators to create policies in order to initiate and manage all data protection processes. From choosing media servers to enforcing life-cycle constraints on backup files, even ad hoc unscheduled backup and restore actions require an IT administrator to invoke a NetBackup policy.

To simplify the creation of a backup policy—or at least jump start the initial definition of a backup policy—NetBackup provides a wizard, which applies common defaults as it creates a policy.

NetBackup provides a wizard to create policies, which is especially helpful in getting started using NetBackup. Essentially the wizard walks an administrator through four basic questions: What clients are to be backed up, where is the data on the clients, where will the backup data be stored, and when will the backup process occur. Once we had created an initial set of policies, we found that cloning policies through a copy-and-edit process was a simple and efficient method to expand detailed policies.

NetBackup Policy Wizard
NetBackup offers IT administrators a powerful combination of policies and templates that allow them to quickly assign a storage life cycle for backup data—from creation to expiration—on all storage resources across physical locations.

Under the NetBackup framework, when a data protection process needs to access a storage device, the enterprise master server assigns a media server to handle the task. The enterprise master server uses the selection of media servers to optimize storage resource utilization of storage resources. To load balance and scale data protection, NetBackup 6.5 enables enterprise master servers to dynamically allocate the control of storage devices based on media server OS factors, such as CPU usage, memory usage, I/O load, and the number of active jobs.

On our third NetBackup data protection server, we ran NetBackup PureDisk 6.5.2 in order to provide our test bed with a data deduplication service. Like EMC's Avamar, NetBackup PureDisk runs its data deduplication process on the client system rather than the server. Client-side deduplication has benefits for both physical and virtual server backups that occur over a local area network (LAN) because it lowers both network traffic and backup storage requirements.

With NetBackup, IT has two powerful data deduplication options for VM backups in a VMware environment—similar benefits are being added for Hyper-V. IT can either run a PureDisk client on a VM or run the deduplication process in the NetBackup media server with the NetBackup PureDisk Deduplication Option (PDDO).

From an IT operations perspective, backup administrators continue to work with the resource virtualization and policy-driven operations of NetBackup, while PDDO transparently adds a data deduplication service. What's more, PDDO adds a copy of the backup meta data sent to PureDisk to the NetBackup Catalog, which enables IT administrators to restore backup images processed with PDDO very quickly.

**Data Deduplication in a Virtual Operating Environment**

For each VM in our test environment, we provisioned a 12GB logical system disk and a 25GB logical work disk. On each VM, the system disk was configured as a vmdk file stored within an ESX datastore—DS1_Xio, a 1TB logical disk that was exported by the Emprise 5000.

On the other hand, the work disk on each VM was configured as a Raw Device Map (RDM) volume, which is a logical disk volume formatted by the VM's operating system as a native file system. That makes an RDM volume directly accessible by other physical, as well as virtual, systems that are running the same OS. The RDM volume remains manageable through VMFS, through the creation of an optional vmdk mapping file by the ESX server. While other backup packages require virtual compatibility to back up RDM volumes, NetBackup 6.5.4 is able to handle any RDM volume, whether or not it has been configured with a vmdk mapping file.
From a data-content perspective, each VM system disk in our virtual operating environment contained 4-to-5 GB of highly redundant data in the form of common OS and application files. On the other hand, each work disk contained 5-to-6 GB of relatively unique structured and unstructured work data. As a result, backup images from our test virtual operating environment would present multiple layers of duplicate data.

Each system disk was represented by a physical 12GB vmdk file, which contained about 7GB of “empty” free space. In addition, the data in the used space of both the system and work disks contained a wide spectrum of data redundancy. That makes NetBackup deduplication an important tool to improve storage utilization of virtual machine backups, especially image-level backups.

Storage Options

Storage infrastructure for our virtual operating environment test bed included iSCSI and Fibre Channel (FC) disk and tape devices. Shared storage plays a critical role in the functionality of a number of key virtual operating environment features including backup via VMware Consolidated Backup (VCB). With low-cost iSCSI SANs becoming a mainstay at SMB sites, openBench Labs employed an iQstor iQ2850 to provide iSCSI storage.

In addition, we set up a Xiotech Emprise 5000 to provide FC SAN storage. Our Xiotech Emprise 5000 sported two ISE datapacs, each of which provided 3.2TB of RAID-5 or 1.7TB of RAID-1 storage. More importantly for high I/O throughput, the two 4Gbps FC controllers on the Emprise 5000 support active-active MPIO for both VMware ESX Server and Windows Server 2003. This gives the Emprise 5000 a significant advantage in...
throughput for applications that depend on streaming data, such as a VCB backup.

The linchpin in a VCB configuration is a Windows-based server that shares access to the VMFS datastores used to hold the vmdk files associated with VMs. Dubbed the VCB proxy server, this server uses a logical lun driver to copy VM snapshot images from the VMFS datastore to a local directory on the server. The files in that local directory are the files that get backed up and archived. As a result, I/O throughput for that local directory has to be at least as fast as reading from the ESX datastore and writing to the backup device in order to avoid creating a bottleneck.

We also provided each NetBackup media server with SAN access to a Spectra Logic T50e robotic tape library. For our tests, the T50e was configured with two LTO4 drives, two 4Gbps FC controllers, and fifty tape cartridge slots. What's more the T50e supports internal soft partitions of drives and cartridge slots as a way to virtualize the presentation of multiple tape libraries to NetBackup.

Managing the cartridge inventory and optimally balancing I/O throughput manually for the Spectra Logic T50e tape library could add significant overhead for an already burdened IT staff. NetBackup, however, greatly simplifies those tasks by managing all of the library's standard features. NetBackup is able to inventory each partition as a virtual tape library and apply unique lifecycle management policies to the media in each partition. Moreover, through NetBackup's ability to load balance multiple media servers and share tape drives among media servers, IT administrators have a mechanism to balance the throughput loads on a high-end multi-drive library.
When an IT administrator creates or edits any policy, including policies for backing up VMs, NetBackup interactively validates that the new policy is able to access client files and, more importantly, create and copy snapshots.

**VM Backup Automation**

For CIOs, the rationale for a virtual operating environment rests on improved resource utilization and streamlined IT management. An important outcome is the garnering of the system availability typical of large data centers without incurring the costs associated with the physical infrastructure of a large server farm. To ensure success, IT must adapt operating procedures, such as the running of clients in VMs that fit a virtual environment’s special needs. Applying current IT backup procedures can fail to improve disaster recovery preparedness and may introduce other unintended consequences that significantly degrade the virtual operating environment.

In particular, procedures designed to backup files resident on a physical system only address the logical persona of a VM as an instance of a system running a particular OS. A VM also has a physical persona as an application running on a host virtual operating environment server. A complete data protection process must be able to restore a VM as a functioning application on a virtual operating environment host server.

What’s more, in a virtual operating environment, the CPU processing load on one VM impacts the host server and that in turn impacts processing on every other VM running on that host. As a result, the standard IT strategy of running independent streaming backup jobs in parallel may not scale for multiple VMs that are running on a single virtual operating environment host server.

VMware recognized the issue of backups impacting other VMs and introduced an API called VMware Consolidated Backup (VCB) for off-host backup. The VCB API allows backup applications to move a snapshot copy of a VM to an alternate host for processing backup operations in order to eliminate the impact of backup operations on the primary VM host server. To enhance seamless integration with VCB, Symantec introduced unique patented technology into NetBackup 6.5 that indexes the contents of a virtual machine during a VM image-level backup—also known as a VMDK level backup.

For IT, this technology is manifested in the backup and restore policy dubbed FlashBackup-Windows. Using the FlashBackup-Windows policy type, IT administrators are able to configure a backup policy that supports a full VM backup, which backs up all of the files associated with a VM, including the VMware system files. By backing up all VM files resident in an ESX datastore, an IT administrator is able to restore a complete
working virtual machine. What's more, this method of backup also supports the restoration of a VM in its logical OS-specific format from the same backup image. Using the drill-down capabilities of the NetBackup Restore module, an IT administrator can even search for and restore an individual file within a VM.

Via integration with VCB, NetBackup avoids the overhead associated with a backup agent. To provide agent-less off-host backup, VCB is installed on a “proxy server,” that runs Windows Server and shares access to the ESX server's datastore volumes. To keep all backup data traffic on our SAN, we made each NetBackup media server a proxy server. The key VCB component for minimizing ESX server involvement and keeping data traffic on a SAN is a VLUN driver, which the proxy server uses to mount and access a VMFS-formatted datastore.

For virtual operating environment backup, NetBackup can leverage VCB integration to establish account credentials with multiple servers running ESX Server, vSphere, or vCenter Server. By establishing credentials with vCenter Server, NetBackup gains access to all ESX servers that the vCenter Server manages. In addition, NetBackup is able to leverage the management capabilities of vCenter Server to power on any VM in a powered off state during a VCB-based backup process.

The VM backup process starts with the proxy server sending the host ESX Server a VM snap command to initiate a snapshot. This creates a point-in-time copy of a virtual disk on a VM. In particular, the ESX Server freezes the vmdk file associated with the VM drive. Next, the ESX server sends a list of disk blocks for the frozen vmdk file to the proxy server. The proxy server then uses the block list with the VLUN driver to read the VM snapshot.
Most importantly, the exchange of data between the ESX server and the VCB proxy server is done entirely over a SAN: There is no LAN traffic. As a result, a VCB-based backup has minimal impact on production processing. The proxy server copies the frozen vmdk file to a local directory and NetBackup backs up on that local directory without involving the ESX host.

Once the proxy server finishes copying all of the frozen vmdk files to its local directory, the proxy server dismounts the frozen vmdk files and the ESX Server removes the snapshot from the VM. From the perspective of the VM, processing was interrupted for only the precious few seconds it took as the ESX Server executed the snapshot and then when the ESX Server removed the snapshot and consolidated any changes that were made while the snapshot existed.

**Script-free Policy Automation**

Normally, all of the directives for VCB are hard coded in JavaScript files, which are edited outside of the backup application. Backup operators must run their backup package in conjunction with the JavaScript files to execute a backup of a VM. Worse yet, since the JavaScript is external to the backup application, there is no way for an IT administrator to know in advance whether the JavaScript is correctly configured without running a complete backup.

On the other hand, when system administrators employ NetBackup, they never have to configure an external JavaScript file. Thanks to Symantec’s integration efforts, all interaction between NetBackup and VCB is established and managed internally through NetBackup policies. What’s more, when an IT administrator creates or edits any policy,
including policies for backing up VMs using VCB, NetBackup interactively validates that the new policy is able to access client files and create and copy snapshots. As a result, NetBackup takes all of the guesswork out of the process of configuring a backup for a VM.

For IT management, NetBackup's policy-driven framework bolsters IT operational consistency and credibility. The seriousness of maintaining credibility for IT was underscored in a Gartner Research Note, "Key Competencies for Service-Based IS Organizations." Gartner warned that IT departments were in danger of "losing their franchise to deliver IT solutions" to a variety of specialized consulting and outsourcing...
firms due to a lack of consistent performance and inability to meet a Service Level Agreement (SLA).

A good deal of the problem stems from a lack of technical understanding among line-of-business (LoB) managers, who see IT only as a cost-generating black box. As a result, making IT credible in the eyes of LoB managers is all the more critical because of the misconceptions of this constituency about IT simplicity, which are fostered by their desktop-PC-centric frame of reference. Nonetheless, it is precisely the complexity of the infrastructure required to scale out applications over multiple servers—whether physical or virtual—that significantly contributes to IT’s inability to quickly identify and resolve many of the root causes of computing problems.

That puts IT under a very unflattering corporate spotlight and helps drive corporate mandates for IT to become as predictable, manageable, and consistent as any other business process. NetBackup allows IT to adopt the principles of classical quality control with respect to operations management and begin automating common data protection processes.
In our virtual operating environment testing scenario, we ran D2D backups of eight VMs in parallel at upwards of 650MB per second—roughly 2.25TB per hour.”

**Backup, DR, and D2D Convergence**

Media servers play an important role in the way that NetBackup creates a unified framework of end-to-end data protection processes in order to address the convergence of backup and disaster recovery. An important characteristic of that convergence is the use of disk storage as both a backup and a disaster recovery medium. Old-line strategies for data backup all shared one simple directive: Back up everything to tape.

Today’s best IT practices, however, call for the implementation of new data protection approaches that utilize both disk and tape. In particular, the need to simplify the restoration of end-user data has spurred the growing use of disk-to-disk (D2D) backup. On the other hand, the necessity to maintain data over long-term periods measured in years, as well as, the need to keep data secure at off-site locations ensures the continued viability of tape.

NetBackup refers to physical storage targets for backup images as a storage unit. Storage units can be a large variety of devices including, robotic tape libraries, disk directories, or disk pools, which can be assigned to one or more media servers. For D2D backups, NetBackup permits an unlimited number of disk storage units, which can be consolidated and virtualized into a smaller number of storage groups for simplified management.

When an IT administrator creates a disk-based storage unit, the administrator also selects a type of disk from eight possible choices. In our tests we used three disk storage types:

1) Basic Disk storage units, represented by disk volumes imported from iSCSI and FC SAN arrays.
2) Storage pools, aggregate multiple disk volumes—any mix of DAS or SAN units—into a single logical unit that can be assigned to one or more storage policies.
3) A PureDisk storage unit, represented by a volume on our PureDisk server with data deduplication.

**Policy Driven Storage Management**

Continuing the policy-management paradigm, NetBackup 6.5 introduced a new type of policy template called a storage lifecycle policy. A storage lifecycle policy can be applied to any number of backup policies which significantly improves automation.
Lifecycle policies simplify data management by creating a template that states where every copy of a backup image should be stored and for how long. A single policy can be created that specifies the lifecycle of backup data including where all duplicate copies should be stored and when they should be expired—different images can be expired at different times.

To further refine the ability for IT to address the pain points associated with long-term backup retention, a NetBackup disk or storage pool can be assigned a staging policy. The staging policy provides for the automatic migration of backups through a storage hierarchy by creating a process in which a backup is written to a storage unit and later duplicated to a second storage unit. In a staging process, backup sets retain their original characteristics. When a backup operator restores a backup image, NetBackup automatically follows the pointers to the actual location.

In addition to ensuring consistency, IT can use storage policy options to fine tune backup performance. A key storage policy parameter for performance sets the maximum number of concurrent backup jobs for a storage device. For disk storage, this parameter is analogous to disk queue length. For tape libraries, this parameter defaults to the number of tape drives in a library. An IT administrator must reset the parameter in order to multiplex several job streams onto the same backup tape.

IT can use storage policies with staging to optimize both the performance and the costs associated with data protection processes. For mission critical systems that require a minimal backup window, backup processing can be accelerated with multiple streaming disk-to-disk (D2D) backup jobs that employ high-speed storage units. In our tests, we streamed backup jobs on fast SAN-based storage units at 650MB per second with one media server. Once the initial backup is complete, a staging policy can automatically move the backup images to less expensive media, such as tape or low-cost high-capacity disks, as a background task.
More importantly, NetBackup will automatically attempt to decompose a backup process into the maximum number of concurrent jobs allowed by a storage policy. We set the maximum concurrent job parameter for each disk storage unit to support at least ten concurrent jobs. As a result, NetBackup was able to decompose any D2D backup that included all of the virtual machines running on our host ESX server into eight parallel backup processes.

In addition, there is a similar parameter that enables NetBackup to decompose a single client backup into multiple streams. By setting the maximum data streams per client, an IT administrator enables NetBackup to decompose a single-client job that involves multiple volumes or directories into multiple parallel processes just as easily as a job from multiple clients.

This parameter is particularly important in combination with backup staging with NetBackup 6.5 in a virtual operating environment. The staging process is treated as a backup process and the VM backup images are readily identifiable by their originating client VM. As a result, a single staging process can be decomposed just as if there were multiple processes originating from multiple clients.

More importantly, these settings are made with respect to the distinct capabilities of each media server. This enables an IT administrator to provide NetBackup with all the information it needs to automatically tune and leverage the capabilities of a media server based on the real-time system load. What’s more, the NetBackup master server can now load balance across media servers that are themselves automatically tuned. As a result, each media server will be utilized optimally and the entire NetBackup environment will scale to the maximum amount possible.
DEALING WITH DATA REDUNDANCY

A crucial technology facilitating the move to disk-based data protection schemes is efficient and reliable data deduplication. As backup needs increase, backup processes stress servers with more than high I/O bandwidth demands. For any site embarking on a D2D backup initiative, provisioning enough storage resources to warehouse backup images is a pivotal issue.

From 1TB of primary disk storage, a traditional Grandfather-Father-Son (GFS) retention plan for daily incremental and weekly full backups consumes about 25TB of archival storage. In terms of a traditional tape library, that’s over 16 LTO-4 tape cartridges. For disk storage, that 25-to-1 metric has even more profound implications as IT heuristics peg 25TB as the maximum amount that an IT administrator can efficiently manage.

For an SMB site with 4TB of active data stored on-line, embarking on a D2D backup initiative carries with it the need to provision for 100TB of on-line archival storage to maintain backup images. Managing that expansion in backup-media has the potential to be a serious drain on any IT budget. To deal with the storage provisioning issues that a D2D backup scheme imposes on IT, customers can use the NetBackup PureDisk Deduplication Option (PDDO) to substantially reduce both storage and management requirements for storing data on disk instead of tape.

During a backup, the PureDisk deduplication engine on each media server segments all of the files in the backup image and replaces any duplicate segments with pointers to a single instance of the data. NetBackup then transfers a unique set of segments and meta data to the PureDisk server, which reduces LAN traffic to help optimize overall throughput.

Data Deduplication Cache

“Data deduplication rates with the PureDisk Deduplication Option were directly related to the media server cache hit rate: As we built up backups over time and as staging server jobs increased in the number of backups to transfer, the data depllication savings ratio went from 4-to-1 to 25-to-1.”

Data Retention and Archive Expansion

Data retention is a fundamental to any secure backup plan. Typically, data retention is expressed in terms of a tape rotation schedule designed to minimize the number of tapes used. The most common tape rotation schedule is the 7-year Grandfather-Father-Son (GFS) scheme.

1) **Grandfather**: A monthly full backup that is immediately stored off-site and held for 12 months.

2) **Father**: A week another full backup is created and held-on site for one week and then held off-site for three weeks and then recycled.

3) **Son**: A daily incremental backup that is held on-site for one week, sent off-site for three weeks, and then recycled.

4) **Last weekly backup each month** is promoted as a monthly and enters the 12-month rotation cycle.

5) **Last monthly backup each year** is promoted as an annual backup and enters the 7-year rotation cycle.

**Bottom Line**: GFS rules put 7 annual full backups, 12 monthly full backups, 4 weekly full backups, and 20 incremental backups in play at any time. As a result, every terabyte of primary disk storage can require up to 25TB of archival storage depending on retention details.
The architecture of PDDO enhances the efficiency of both restore and backup operations on data from a deduplication storage target. In particular, PDDO leverages the NetBackup catalog to cache segment meta data in order to optimize deduplication and speed the restoration of data. More importantly, this segmentation entirely changes the notion of a full backup as only the changed segments of a file are now transferred in a backup job.

In our testing scenario of a VMware host, our first objective with NetBackup was to minimize the time needed to run a backup of the eight VMs hosted on our ESX server. To meet this objective, we ran standard D2D backups of the eight VMs in parallel over our 8Gbps SAN at upwards of 650MB per second—roughly 2.25TB per hour. While that was ideal for minimizing the backup window, it left our backup images stored on relatively expensive high-performance storage.

To alleviate that pain point, we extended the NetBackup storage policy for that SAN-based storage unit with a staging policy. The staging policy scheduled an automatic process that duplicated the backup images on our high-performance SAN-based storage unit to a highly cost-effective LAN-based PureDisk storage pool on a regular basis.

From the perspective of a NetBackup administrator, we were running a staging job that used the load balancing of NetBackup to launch four duplication processes: one process for each of the four CPU cores in our server. From the perspective of a PureDisk administrator, however, we were running four separate PDDO jobs with data deduplication.
More importantly, the data deduplication rate for each PDDO process was directly related to the caching of metadata in the NetBackup catalog. On the first VM backup process, the media server cache hit rate was 0% and that produced a data reduction ratio of 3.6-to-1 as we transferred 5.88 GB of data to the PureDisk server for a backup job that scanned 21.16 GB of raw data. After running a number of VM backup jobs, the cache hit rate rose dramatically. As cache hits began to exceed 90%, the data deduplication ratio approached 25-to-1.

Average aggregate I/O throughput for staging jobs with four sub processes on our media server was 125MB per second—the throughput limit for the Gigabit Ethernet connection between the NetBackup media server and the PureDisk server. That level of throughput in our staging process easily validates the popular strategy to directly backup the VMs on an ESX host to a PureDisk pool without introducing media staging. This is a particularly effective strategy when using an iSCSI SAN fabric.

Media server caching also significantly benefited our ability to quickly restore VM backup images as either VM applications or logical servers in NTFS format. When we restored VMs from backups stored on our NetBackup PureDisk server, locally stored meta data enabled NetBackup to setup the directory structure of the desired VM before the data transfer process from NetBackup PureDisk began. We were able to find files more quickly because we could browse a VM backup image locally. In our test configuration, typical throughput when restoring deduplicated backup images was around 65MB per second. As a result, IT can increase backup and deduplication throughput by adding media servers or increase throughput when restoring deduplicated data by adding PureDisk servers.
Data Protection Process Value

“With NetBackup 6.5, IT is able to address both the issues of data growth and storage resource utilization, as businesses struggle to comply with government regulations for protecting data.”

**Process Perspective**
Corporate executives think in terms of business processes. Furthermore, they expect the services that support those processes to address issues such as availability, performance, security, and business continuity with specific agreed-upon support levels. For IT to create the policies and procedures needed to support such a service level agreement (SLA), IT must establish rigorous data center process control.

Unfortunately, the task of developing service-centric processes, especially for storage assets, is significantly burdened by the current set of IT storage management tools. The dearth of process-oriented tools leaves most IT organizations without a clear set of links that bind storage resources, applications, and business value. Compounding that problem, IT decision makers are now under pressure to improve the delivery of IT services and simultaneously increase the underlying level of IT resource utilization.

With NetBackup 6.5, IT is able to address both the issues of data growth and storage resource utilization, as businesses struggle to comply with government regulations for protecting data. Whether it is Sarbanes-Oxley, the Health Insurance Portability and Accountability Act, or the Federal Rules of Civil Procedure, which set discovery rules for email, fear of compliance violations has IT adopting defensive policies to store and save all information by default.

That conservative data protection strategy has sent the amount of managed data at IT sites skyrocketing. As a result, reducing storage costs linked to out-of-control capacity growth is a universal priority for IT. What’s more, rigorous backup procedures are also fundamental to any regulatory compliance policy for data security. Further complicating the issues of data protection, the burgeoning use of server virtualization as a means to improve IT resource utilization and reduce operating costs adds greater complexity to backup processes from both a technology and a business continuity perspective.

More importantly, backup as an IT process has not historically benefited from

---

**NetBackup & NetBackup PureDisk Quick ROI**

1) All data protection processes are policy driven, which reduces operator errors and promotes automation.
2) Federated media server configuration provides failover, checkpointing for backup jobs, and maximal throughput via load balancing across media servers using metrics such as memory and CPU consumption.
3) Data deduplication with PureDisk recognizes version changes in files and produced data deduplication rates of 25-to-1 on VM image backups over time.
efficiencies of scale. While storage technology improvements continue to shrink the footprint, increase the performance, and mitigate the impact of disk drives on capital costs, no factors passively mitigate storage management costs. Storage capacity expansion inevitably requires more people and storage management resources. A poorly implemented D2D backup strategy will simply accelerate the need for more capacity and management resources.

By load balancing data protection processes across media servers NetBackup is able to provide process scalability limited only by hardware capabilities. In addition, patented technology greatly simplifies backup and recovery of a VMware virtual operating environments. What's more, these benefits can be increased with Symantec PureDisk, which provides powerful data deduplication features that can be kept transparent for backup operators.

By utilizing PureDisk as either an alternative or offset to long-term storage of backup data on tape, IT can also reduce the amount of tape consumed for secure storage purposes. The need to maintain off-site storage of critical data can be better optimized at significantly lower costs. By using far fewer tapes that cover regulatory-based reporting periods, operations benefit directly by requiring far less media, which in turn reduces all of the ancillary costs associated with scheduled off-site media rotation schedules.