Data Center Backup and Recovery Automation

OPENBENCH LABS

Analysis: Maximize Storage Utilization and Network Bandwidth Automating Virtual System Backup

Analysis: Maximize Storage Utilization and Network Bandwidth Automating Virtual System Backup

Author: Jack Fegreus, Ph.D.
Managing Director
openBench Labs
http://www.openBench.com
September 08, 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

Executive Summary

> "Service recovery in a virtual operating environment, is an extremely important issue, as the risk of a single host server failing cascades to multiple virtual machines (VMs) running multiple applications. "

Today, IT needs to establish highly efficient operations. Working with virtual resources, lets IT administrators limit the devices they manage to a small number of abstract pools. Introducing multiple levels of logical abstraction and redirection of storage resources, however, can obscure and complicate important IT operations, including the once simple notion of backup.

openBench Labs Test Briefing:
CommVault® Simpana® 8.0

1) **Minimize disk and tape storage with data deduplication:** Simpana software supports deduplication policies that allow administrators to create special data stores containing signatures of unique data blocks or objects in a disk-based library along with reference pointers to the unique data, which means backup jobs only add new data or pointers to existing data.
2) **Maximize LAN throughput with data compression:** Simpana software iDataAgents can compress backup data at the client or the media server to reduce LAN data traffic and avoid I/O bottlenecks, and improve storage utilization of devices without hardware compression.
3) **Eliminate backup overhead for VMware® ESX and Microsoft® Hyper-V™ servers:** The Simpana Virtual Server iDataAgent configures a proxy server that automatically initiates, mounts, and backs up VM snapshots to minimize backup overhead on VMs and host servers.
4) **Backup and restore full VMs or individual files:** A disk-level backup created with the Virtual Server iDA provides for individual file or entire VM restores in either VMFS or NTFS format from a single backup job.
5) **Ensure data protection consistency with end-to-end policies:** Simpana guides IT administrators through the creation of simple narrowly focused policies that are linked to form a complete sophisticated end-to-end policy for data protection.

Simpana® from CommVault® enables IT to leverage a single solution that integrates backup, archive, replication and storage resource management across physical and virtual servers. Simpana helps IT deal with the challenges associated with critical backup recovery issues within a virtual infrastructure by providing a Universal Virtual Server Agent that helps reduce the impact of backups on the physical hosts and virtual machines (VMs), while simplifying administration with innovative features such as VM auto-discovery, which automatically discovers newly added virtual machines and places them into existing backup policies.

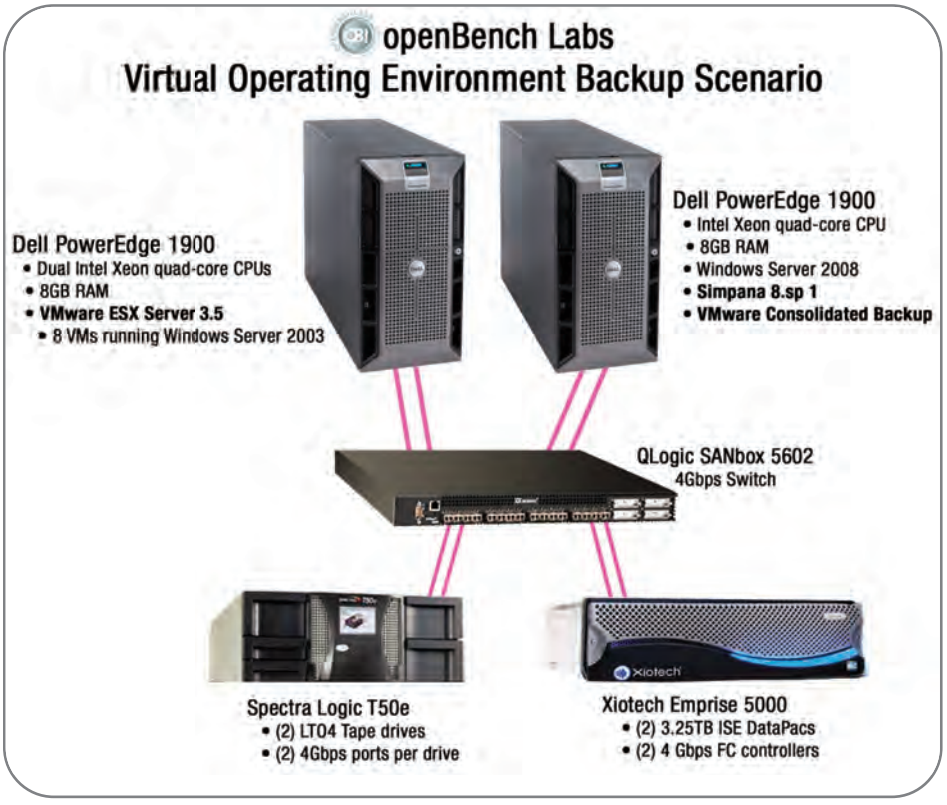
Simpana also simplifies a major issue for every VM: two distinct personas. First, there's the IT-centric persona of an application running on a host. Second, there's the logical line-of-business persona of a VM as a standard computer. As a result, service recovery in a virtual operating environment is an extremely important issue, as the risk of a single host failing cascades to multiple virtual machines (VMs) running multiple applications. To resolve the dichotomy in VM perception, CommVault offers a variety of options that align with recovery needs. From a single backup job, administrators can restore a Windows-based VM as either a set of VMware container files or a set of Windows system files. Simpana can also be deployed in a number of ways, including as an agent inside a virtual machine or on a proxy server with the Universal Virtual Server Agent for off-host backups of virtual operating environment servers and disk-level recovery.

Data Protection Test Bed

"We built and optimized four end-to-end backup and recovery processes using CommVault's File System and Universal Virtual Server iDataAgents configured to leverage policy-driven disk- and tape-based storage libraries with data deduplication."

Backup, Compress, Dedupe, and Restore

To get a better perspective on Simpana's ability to enhance strategic IT initiatives, such as the adoption of a virtual operating environment or an IT service management operations model, openBench Labs set up a data protection test scenario using VMware Virtual Infrastructure to host eight active VMs running Windows Server 2003.



Each test VM was configured as an application server running SQL Server and IIS. This virtual operating environment was hosted on a Dell PowerEdge® 1900 Server with two quad-core 3.13GHz CPUs, 8GB of RAM and a dual-port QLogic QLE2462 Fibre Channel HBA.

To provide IT with the end-to-end processes needed to move, store and retrieve critical business data, CommVault offers a suite of license-able data management modules, which leverage common services to provide enterprise-class data protection, archive, replication, search and e-Discovery. For backup and recovery processes, IT is able to control and monitor all end-to-end processes in a single pane of glass via the CommCell® Console.

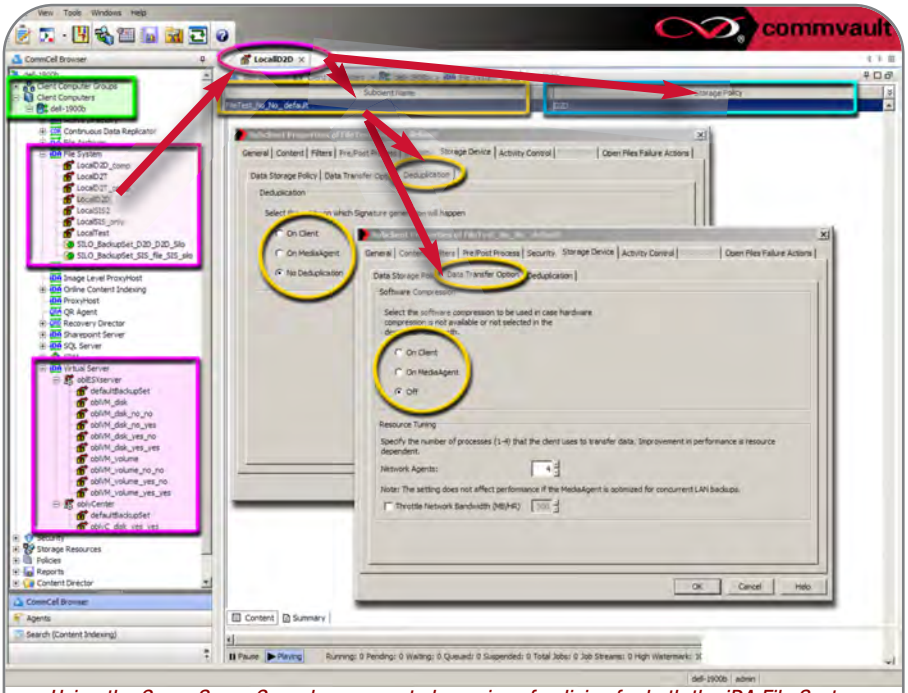
Within every CommCell configuration, a CommServe server coordinates all communication among CommCell components. The CommServe server also maintains

a database of information relating to the CommCell configuration. One or more media servers, also known as CommVault MediaAgents, manage external storage resources for the CommCell. By adding multiple servers to run the MediaAgents, IT is able to enhance both scalability and availability of storage services.

Intelligent Data Agents

Client Agents, also known as Intelligent DataAgents or iDAs, are software modules that provide various data protection services for client systems running a broad array of specific operating systems or applications. These agents are used to perform a very wide range of data protection services. Backup and recovery processes are handled by a family of iDAs, which are content-aware and designed for specific types of data. Other Client Agents support such data protection services as replication, compliance archiving, Storage Resource Management (SRM), content indexing, and search.

Client iDataAgents



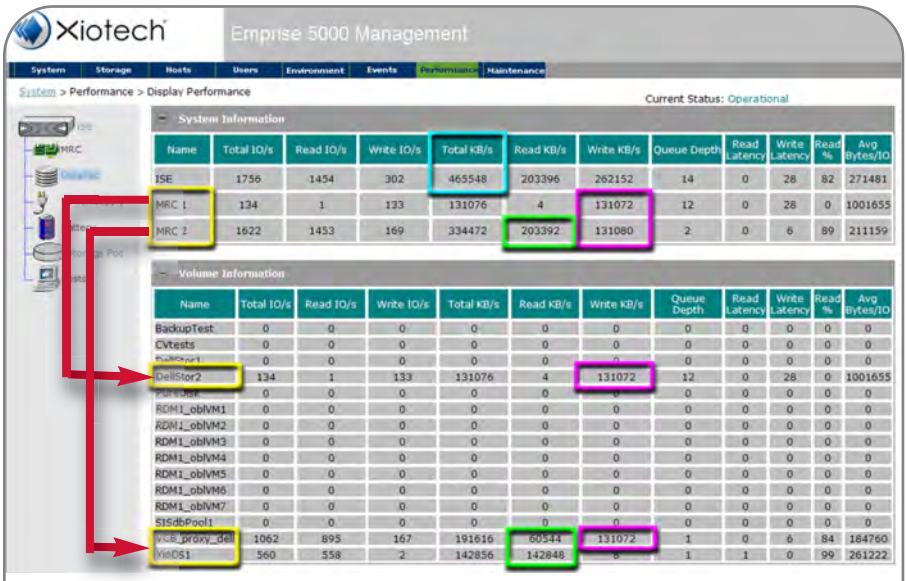
Using the CommServe Console, we created a series of policies for both the iDA File System and the iDA Virtual Server Agent. These policies were linked to policies for storage elements and policies for "subclients," which list the specific targets for backup. Options for iDataAgents include whether to implement software compression and where the hashing function for deduplication should be run—on the client or the media server..

Using Simpana, IT administrators build end-to-end processes by linking a series of policies that include backend rules for how media servers work with storage resources and client-side rules that govern how content-aware agents interact with media servers. For this analysis, openBench Labs explicitly focused on backup and recovery processes. In particular, we built four end-to-end backup and recovery processes using CommVault's File System and Universal Virtual Server iDataAgents configured to leverage policy-driven disk- and tape-based storage libraries with data deduplication.

To secure and minimize data traffic between clients systems and media servers within a CommCell, CommVault iDAs are able to compress and encrypt data on the client systems before transmitting it to a MediaAgent. More importantly, iDAs may share some of the deduplication burden by generating hash signatures for data blocks at the client. The data is then compressed and optionally encrypted before it is sent to the

MediaAgent, and the MediaAgent performs the deduplication.

VIRTUAL SERVER iDA STORAGE PERFORMANCE



Total I/O throughput for eight VMs with the Simpana Virtual Server iDA was typically about 500MB per second. From the perspective of our Xiotech FC storage server, the disk DellStor2 was used for the Simpana D2D library, to which all backup writes were directed. All snapshots of VM virtual disks were read from the ESX datastore disk, XioDS1. Finally, VCB_proxy_dell was the disk on the Simpana proxy server to which VM snapshots were written and from which backup data was read.

For the Universal Virtual Server iDataAgent, (VSA) the client system may be a physical server or a virtual server, depending on the virtual infrastructure. For Microsoft Hyper-V, the VSA installs on the physical Hyper-V server and uses Microsoft Volume Shadow Copy Services (VSS) to quiesce and snapshot the VM.

When VSA is installed on a VMware virtual server, it leverages hot-add mode to back up virtual machines. When deployed on a physical server, it can work in both SAN and LAN mode to access VM data. For our testing, we utilized VSA exclusively on a Windows-based Proxy Server in SAN mode."

For VMware, VSA leverages the VMware Consolidated Backup (VCB) framework to quiesce virtual machines, create a virtual machine snapshot, and make the snapshot available on the Proxy Server using the selected transport mode. VSS quiescing can also be enabled in the VMware tools to ensure that VSS-aware applications running within a VM, such as Active Directory, MS Exchange, or SQL Server, are in a consistent state before the VM snapshot is created.

Once the snapshot is created and made available, VSA backs up the VM data to the MediaAgent. If VSA and the MediaAgent are on the same server, the process is considered to be LAN-free backup. In addition, during the backup of Windows guest operating systems, VSA may track metadata to catalog and index individual files and folders inside the VM. This enables a single-pass backup to provide both full VM recovery, as well as granular recovery of files.

From the perspective of a VM, backup overhead is limited to taking and rolling back a snapshot. To support the powerful performance features of CommVault's VSA, openBench Labs utilized a second Dell PowerEdge 1900 Server with a single quad-core CPU, 8GB of RAM, and a dual-port QLogic QLE2462 HBA.

File System Backup Performance

> "There was no penalty when restoring jobs characterized by a 20-to-1 data deduplication ratio: Simpana proceeded to reconstruct and rewrite the original files back to disk at upwards of 300MB per second (480GB per hour)."

Storage Policies

Backup began as a disk-to-tape (D2T) process, which made maximizing streaming I/O throughput and minimizing the time to complete a backup the driving issues. In a D2T environment, storage utilization was addressed with a static tape rotation plan designed to limit the number of physical tapes needed to retain backup images at a secure off-site storage location. With the introduction of disk-to-disk (D2D) backup, however, a problem surfaced with backup retention policies: Storing 7 years of backup images consumes much more storage space—25 times more—than the original data.

To help organizations deal with massive data growth and meet more aggressive retention and recovery service level agreements (SLAs), Simpana lets IT enhance data protection policies with deduplication to meet recovery SLAs from disk without utilizing excessive disk capacity. What's more Simpana deduplication does not require specialized storage or appliances, and can be implemented using commodity disk and off-the-shelf servers.

Nonetheless, from a deduplication perspective, changed data is unique data and that's especially problematic with a database where small changes can cascade across multiple tables. What's more, external regulatory bodies continue to extend the time that companies must hold specific datasets to well beyond seven years. As a result, even a small 5 TB production dataset can expand over time to cover a disk footprint of 300TB.

To resolve that problem, CommVault Simpana can extend the deduplication tier to offline media for efficient and cheaper long term retention. Backup images that have been created using a policy with deduplication can be stored on disk to support near-term restoration and then moved to tape without re-hydration of the data for long-term storage. Moreover, sites that need to retain static data for long periods should also consider specialized CommVault Simpana options for compliance and migration archiving.

We began our assessment of Simpana's backup and restore capabilities using the File System iData Agent. This would serve as a baseline for analyzing functionality and performance with the Virtual Server Agent.

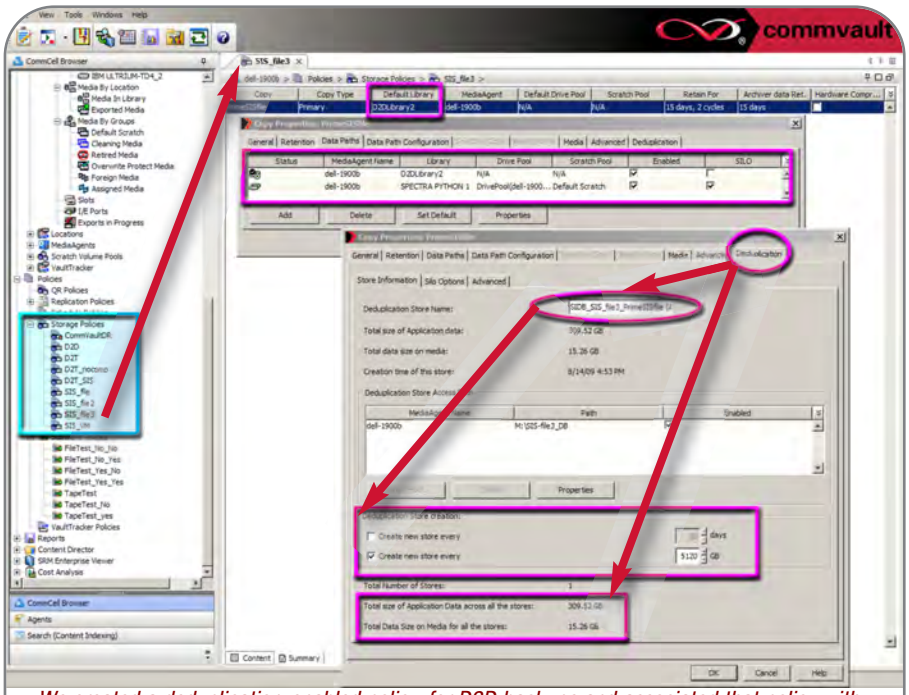
Data compression has long been a staple hardware function on tape drives to reduce the quantity of data written to tape. Simpana provides an option within iData Agent (iDA) policies to apply software compression at either the client system or at the media server. In our tests of Windows-based clients, the Simpana File System iDA typically

compressed data at a 4-to-1 ratio, which meant the volume of data transferred to the media server and then to the storage library was just 25 percent of the original data set.

Saving Space with Deduped Storage

Even more impressive were the storage utilization gains provided by adding data deduplication to a D2D backup along with the ability to pass the storage gains to tape media. For administrators, set up of deduplication starts with the creation of a special data store for the signature associated with each unique block and reference pointers to the unique elements.

Deduped Storage Policy



We created a deduplication-enabled policy, for D2D backups and associated that policy with File System iDA policies that applied data deduplication at the media server and data compression at the client: A strategy that garnered a 20-to-1 reduction in storage utilization. In addition to a default library, D2DLibrary2, we created a special store for the signatures of each deduped block of data that was backed up. We also designated our tape library as a secondary silo, for transferring deduped backups to tape for off-site storage.

Using policies invoking data compression on clients and data deduplication on the media server, we measured 20-to-1 reductions in the amount of storage required for file-system backups. We were able to backup 309.5 GB of file data using just 15.3 GB of storage capacity for unique data and pointers.

What's more, there was no penalty when restoring jobs characterized by a 20-to-1 data deduplication ratio: Simpana proceeded to reconstruct and rewrite the original files back to disk at upwards of 300MB per second (480GB per hour).

For long-term retention and off-site storage, deduplication-enabled backups created with a File System iDA can be assigned a special silo tape library to which deduplicated backup images can be directly migrated without re-hydrating (re-expanding) the data. This gives off-site tape storage the same utilization ratios as local disk-based libraries with deduped File System iDA backups.

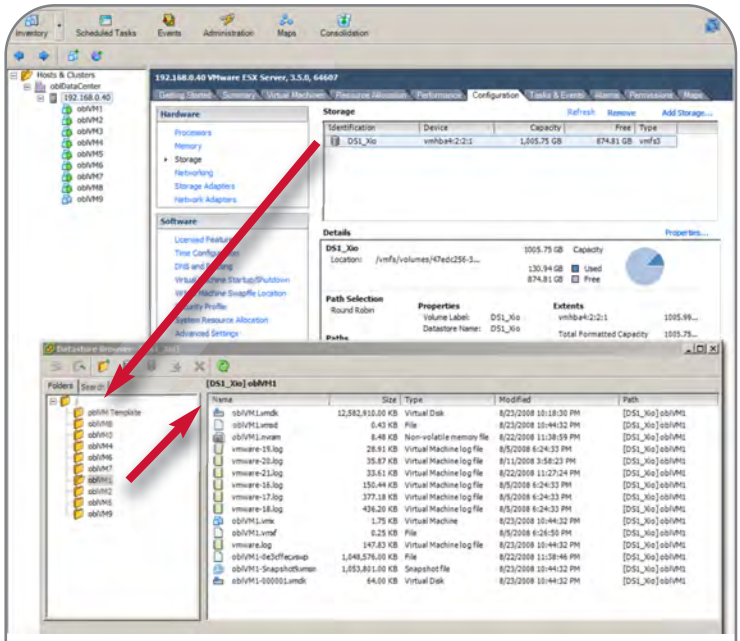
Virtual Server Backup Performance

> "Backing up eight VMs in parallel with the Virtual Server iDA, total I/O thoughput on simultaneous reads and writes reached 500MB per second, CPU processing on each VM never exceeded 5 percent, and memory usage never exceeded 10 percent."

Making Virtual Universal

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, along with about 7GB of "empty" free space. As a result, backup images from our test environment would present multiple layers of duplicate data.

ESX Server Storage Configuration



Each VM was provisioned with a 12GB logical system disk, which was stored as a vmdk file within an ESX datastore and shared with the Simpana proxy server. Moreover, each VM utilized about 70 percent of its system disk, which means 30 percent of each vmdk file was blank.

For CIOs, the rationale for a virtual operating environment rests on improved resource utilization and streamlined IT management. An important outcome is garnering the reliability, availability, and scalability (RAS) levels of a large data center without incurring the costs associated with the physical infrastructure of a large server farm. Attempting to manage a virtual operating environment with the tools that are not designed to fit a virtual environment's special needs can be a fast track to disastrous results.

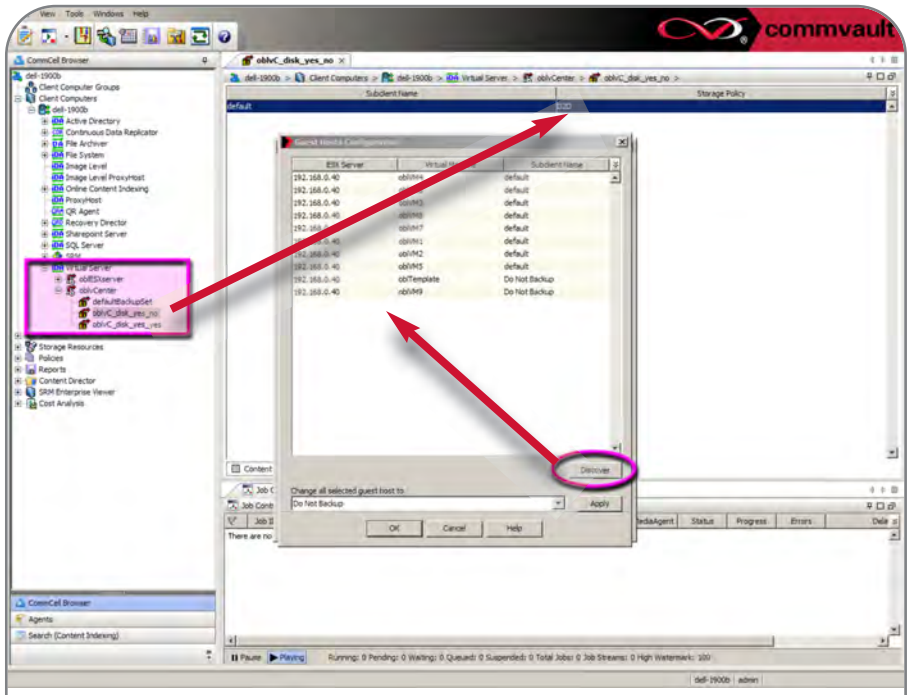
Backup and restore procedures are simplified with physical servers, since each system is isolated by its physical host. In a virtual operating environment, the CPU processing load on one VM impacts the host server and that in turn impacts processing on every other VM on that host. The IT strategy of running independent backup jobs in parallel may not scale for multiple VMs running on a single host server. What's more, a backup process must address more than just the logical persona of a VM as an instance of a system running a particular OS.

A complete data protection process must also address the physical persona of a VM as an application running on a host server. As an analog to bare metal restore for physical

servers, complete data protection for a VM requires the ability to restore a VM as a functioning application on a virtual operating environment host server.

The Universal Virtual Server iDataAgent supports three backup modes for VMware environments. With Disk Level, the Virtual Server iDA backs up the entire virtual machine disk image and provides the capability to recover the full VM or, for Windows-based guests, individual files inside the VM. With Volume Level, the Virtual Server iDA provides full VM protection for Windows-based guests, without the VCB overhead of additional cache space, as well as granular file level recovery. Finally, with File Level, the Virtual Server iDA backs up only the files inside the virtual machine, which is useful when combined with Simpana content indexing capabilities.

VIRTUAL SERVER IDA VM DISCOVERY



We created a Virtual Server iDA linked to our D2D storage policy that accessed our vCenter Server to gain account credentials for the VMs. We then set up a default subclient policy that selected eight VMs for backup.

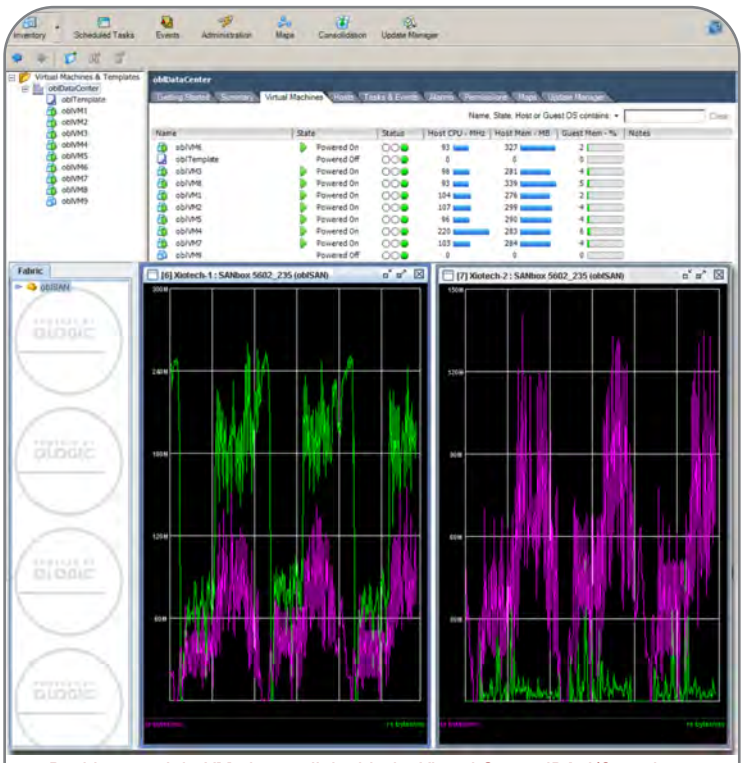
When creating backup and restore policies for the Virtual Server iDA, IT administrators can establish credentials with servers running ESX Server or vCenter Server. The latter option extends access to all ESX servers and objects, such as templates as well as VMs, under the control of the vCenter Server. Once a connection with an appropriate server is established, Simpana provides one-button VM discovery. The process identifies and configures all VMs that run in the configured scope as subclient contents inside the virtual server iDA."

Most importantly, in the SAN mode tested by openBench Labs, the exchange of data between the ESX server and the Simpana proxy server was done entirely over a SAN by the proxy server: There is no LAN traffic. As a result, a Virtual Server iDA backup has virtually no impact on production processing. The proxy server copies a snapshot of the VM's data files to a local directory and then the Virtual Server iDA backs up on that local directory.

Once the proxy server finishes copying all of the VM's files to its local directory, the proxy server dismounts the snapshot files and the ESX Server removes the snapshot from the VM. From the VM's perspective, processing was interrupted only for the precious

few seconds it took to execute and later roll back the snapshot.

Virtual Server Backup Throughput and Overhead



Backing up eight VMs in parallel with the Virtual Server iDA, I/O total throughput on simultaneous reads and writes reached 500MB per second: At the same time, CPU processing on each VM never exceeded 5 percent and memory usage never exceeded 10 percent on any VM.

The isolation of the ESX server from the Simpana proxy server running the Virtual Server iDA is highly significant for the data deduplication stage of our testing. Under the most favorable conditions, data deduplication is highly resource intensive in terms of memory, disk, and CPU. When dealing with foreign resource objects such as ESX vmdk files, an already resource intensive process becomes even more so. In a virtual operating environment, the deduplication process must be configured to take a very fine-grain approach to the division of large virtual disk files into a maximal number of minimal small-block chunks, which must be compared with each other.

In our final testing, we added data compression and data deduplication to a Virtual Server iDA policy. In these tests, data compression was 2-to-1. With data compression and block deduplication set to a 32KB block size, our data deduplication storage policy provided an initial overall data reduction ratio of 8-to-1 in stored data. In particular, we utilized only 19.1 GB of data storage to hold backup images for 151.5 GB of VM virtual disk data.

Achieving this level of storage saving with the Virtual Server iDA, I/O throughput was reduced by 90 percent. Nonetheless, the added processing exhibited no measurable impact on either the VMs or the ESX server. More importantly, there was no penalty on job restorations, which proceeded to reconstruct and write the original VMware files or Windows back to disk at upwards of 300MB per second (480GB per hour).

Data Protection Process Value

"Simpana 8 provides scalable data protection processes for physical and virtual clients that are policy driven and limited only by hardware capabilities."

Process Perspective

Corporate executives think in terms of business processes and 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 data protection policies and procedures needed to support a service level agreement (SLA), IT must establish rigorous data center process control.

CommVault Simpana 8 Quick ROI

1) **All data protection processes are policy driven, to reduce operator errors and promotes automation.**
2) **Software compression reduced backup data traffic and storage requirements by a factor of 4-to-1 in tests using Windows file data and 2-to-1 in tests of VMware Virtual Machines.**
3) **Block-level data deduplication immediately generated reductions in storage requirements for file-based backups of 20-to-1 and 8-to-1 for multiple VMs hosted on a VMware ESX server.**

The task of developing service-centric processes, especially for storage assets, can be daunting. Lacking process-oriented tools, most IT organizations are left 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 the CommCell technology that underlies Simpana 8, IT is able to address the issues of data growth, resource utilization, and compliance with government regulations for protecting and archiving data, as well as deal with the more prosaic needs of backup and restore.

Whether it is Sarbanes-Oxley, the Health Insurance Portability and Accountability Act, or the Federal Rules of Civil Procedure, 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. 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.

Simpana 8 provides scalable data protection processes for physical and virtual clients that are policy driven and limited only by hardware capabilities. Through flexible data compression and deduplication options IT administrators can configure end-to-end backup processes that maximize data throughput and minimize data storage for all types of data. What's more the Simpana Virtual Server iDataAgent greatly simplifies data protection operations across VMware and Hyper-V virtual operating environments.