White Paper

Symantec Corporation

June 2010

By Jerome M Wendt

DCIG, LLC
7511 Madison Street
Omaha NE 68127

O 402.884.9594

Best Practices for Breaking Down the Barriers to Centralized Virtual Server Backup and Recovery

Best Practices for Breaking Down the Barriers to Centralized Virtual Server Backup and Recovery

Table of Contents

DCIG

Best Practices for Breaking Down the Barriers to Centralized Virtual Server Backup and Recovery

Executive Summary

The benefits of server virtualization are becoming so well-known that it has become a mandate in many organizations to virtualize all new application servers. But as this occurs, new challenges are emerging. Better known issues include trying to control virtual machine (VM) sprawl, optimizing the performance of VMs on individual physical servers and identifying new techniques to backup and recover VMs.

But a barrier that is sometimes invisible to enterprises is how to best manage their virtual servers teams and which, if any, of their responsibilities, should be centralized.

Chief among these responsibilities is the backup and recovery of these virtual servers. Many enterprises have already consolidated the backup and recovery of their other platforms but adding the backup and recovery of virtual servers is a decision enterprises cannot take lightly.

The obstacles to do this go beyond just the lack of the backup team's familiarity with virtual server operating systems. They also encompass technical hurdles that are sometimes made invisible by the virtual nature of this environment which can preclude this team from successfully accomplishing this function.

To understand what it takes for an enterprise to successfully centralize the backup and recovery of their virtual server environment, this paper examines:

- Why the virtual server team reporting structures within enterprises are still in flux
- The technical barriers that have precluded the centralization of virtual server backup and recovery
- Why corporate backup teams sometimes resist taking on this responsibility
- Why virtual server teams sometimes resist letting it go
- Best practices for breaking down the barriers to centralized virtual server backup and recovery

Best Practices for Breaking Down the Barriers to Centralized Virtual Server Backup and Recovery

The Organizational Barriers Inhibiting Centralized Virtual Server Backup

The benefits and corresponding drawbacks of server virtualization are becoming well-known. Some of its many benefits include improved utilization of physical server resources, decreased numbers of physical servers, drops in power consumption and the elimination of the need to deploy a physical server every time a new application is deployed.

At the same time server virtualization creates new challenges. Controlling virtual machine (VM) sprawl, optimizing the performance of VMs on individual physical servers and developing new techniques to backup and recover VMs are new obstacles that organizations need to account for as part of managing their virtual infrastructure.

But an organizational barrier emerging within enterprises is preventing the effective centralization of virtual server management. The team or teams responsible for managing this growing VM infrastructure are often still trying to find their rightful place in the organizational hierarchy.

The enterprise organizations that DCIG contacted reflect these struggles. Some virtual server teams are only loosely managed by the central IT. Others do not directly report to a VP or Senior VP of IT operations but rather through IT managers that report to them, such as Windows or Unix IT managers.

Still other enterprises have virtual server teams that are part of specific business units. These teams either do not report to the corporate IT department or, if they do, are given a fair amount of latitude to manage and make decisions about their VM environment.

What happens in many of these cases is that enterprise organizations inadvertently end up creating a virtual server team that operates outside of corporate IT guidelines. These teams may then receive minimal or no guidance about established corporate IT practices for data and server management. This leaves them adopting solutions that meet the specific challenges associated with managing and supporting the VM environment but which are not easily transformed into corporate IT processes.

In some cases, giving virtual server teams the freedom to make these decisions is not a problem. For instance, a decision to deploy a solution such as VMware's vCenter Server to manage vSphere servers will likely not adversely affect the rest of the enterprise's IT operations. Conversely some decisions around VMware support, such as the selection of a backup and recovery solution, conflict with an established IT process.

The VM Backup and Recovery Conundrum

The conundrum that enterprises face is that backup and recovery is typically already a consolidated function handled by a dedicated enterprise IT team.

This team transcends departmental boundaries to backup and recover data for the various operating systems that an enterprise may possess, including Linux, Windows and the different versions of UNIX. By doing it this way, all backup data is centrally managed and maintained.

However enterprises cannot automatically assume that this team can take on the responsibility for virtual server backups and recoveries. The obstacles go beyond just their lack of familiarity with the virtual server environment. They also encompass technical hurdles that are sometimes made invisible by the virtual nature of this environment that precludes this team from successfully accomplishing this function.

Due to the nuances associated with effectively protecting VMs coupled with how close virtual server teams still are to the VMs they support, the virtual server teams may be in a better position to make decisions regarding the backup and recovery of VMs and then carry them out.

So while it may be a corporate objective to centralize the backup and recovery of their VMs, an enterprise needs to first identify what specific technical hurdles exist that precludes them from doing so. Once identified, they then need to select a solution that addresses these issues and enables them to successfully centralize and manage the backup and recovery of VMs.

The Invisible Barriers to Centralized VM Backup and Recovery

Server-based backup agents

The primary method traditionally used by the central backup team to protect servers under their management was to install a backup agent on each physical server. Since many production applications either use less computer resources or are dormant at night, this left ample processing power for the backup agent to use to complete the backups. However in the virtual environment, this agent-based approach to backup can, in some cases, become difficult to carry over.

One potential problem is that physical servers will have multiple VMs on it. In these cases, backup administrators have to take care not to schedule backup jobs so they overlap or conflict with production jobs running with other VMs. Since the backup job on each individual VM can consume up to 20% of

a physical server's hardware resources (CPU, memory and network), overlapping multiple jobs can result in prolonged backup jobs or adversely affect production applications running on other VMs.

This situation is one example of why it is difficult to centralize VM backup and recovery and argues for localized control in this circumstance. Server virtualization calls for the central backup team to identify which application is running on the VM (which it would have to do anyway) plus it introduces new levels of complexity.

The backup team now needs to understand what application is hosted by each VM, confirm there is no overlap in terms of the length of backup jobs and then schedule backup jobs accordingly. Further, it must track when new VMs are introduced to that physical server so the timing of backup jobs can be adjusted accordingly.

Virtual server backup utilities

Virtual server operating system providers such as VMware recognize this backup challenge and provide their own solution to this problem. Available as the vStorage API in VMware vSphere, VMware makes it possible for enterprises to more seriously entertain centralizing backup and recovery.

The vStorage API does help to alleviate the overhead and potential scheduling conflicts associated with VM backups by taking snapshots of individual VMs that can then be backed up. It removes the need for a proxy server that the VMware's backup utility in VMware ESX 3.5 required. It adds features such as incremental VM image level backups and restores.

But even with this improved backup functionality for VMs, enterprises cannot assume that they can centralize the backup and recovery of virtual servers using these native virtual backup utilities. For example the vStorage API only facilitates the creation of the snapshots and some limited management capabilities. So if organizations want to move a copy of that image off to tape, replicate it to an offsite location or put a legal hold on it, third party software is still required.

Resistance from central backup team

These types of limitations can lead to continued resistance from enterprise backup teams in wanting to fully assume the responsibility of VM backups. Traditional approaches to backup and recovery do not always translate well when implemented in virtual server environments. Even new native backup and recovery utilities available from virtual server operating system providers such as VMware lack the robust backup and recovery capabilities to which they are accustomed.

However enterprise backup software now better integrates with these native virtual server backup and recovery utilities. This integration is helping to remove these obstacles to the centralized protection of VMs.

New Backup Techniques Facilitate Centralized Protection of VMs

Enterprises are no longer limited to only applying the traditional methods of backup and recovery to VMs nor do they have to rely solely upon native utilities found in VMware to protect their VMs. The last few years have seen entirely new sets of VM data protection methodologies emerge that are specifically designed to facilitate the centralized backup and recovery of VMs.

Client-side deduplication

A recent advancement is the introduction of client side deduplicating backup software. Like the traditional agent-based backup software model, a backup agent is still installed on each guest VM. The new twist is that the agent no longer backs up all of the data on the VM but rather first deduplicates the data and only backs up net new chunks of data.

Deduplicating the data can reduce the time needed for backups on individual VMs to minutes or even seconds. In so doing, it reduces or eliminates concerns about overlaps in backup jobs and the need to do the balancing act of scheduling backups since they occur more quickly. Further, backing up less data reduces network bandwidth, server overhead and backend storage consumption.

This technique can be more efficient than other forms of partial backup such as differential or incremental. Those two backup methods look for differences in data from prior backups and then backs up those differences.

The deduplicating backup agent differs in that it seeks to eliminate redundant data before it is transmitted while the other two techniques do not. Like both the differential and incremental forms of backup, deduplication first checks to see what data is new. However if the data is classified as 'new' but that chunk of data already exists in the deduplicated backup data store, it does not transmit it again.

Integrated snapshots

The initial implementation of snapshots within VCB only supported recoveries of the entire VM image. This made it more difficult to recover individual files or folders within a VM as backup administrators had to first recover the entire VM and then go outside of the backup software to recover these files or folders.

Now because of enhanced backup software integration with the vStorage API in vSphere, backup administrators can more easily assume the central management of these snapshots.

Enterprise backup software applications now integrate with these features and can centrally schedule VM snapshots to occur on vSphere servers in such a way that they do not conflict with each other or with production applications hosted by that server.

Finally, most virtual server deployments use externally attached storage systems to store the data associated with the VMs. The good news is that many of these external storage systems provide their own native snapshot functionality. Enterprise backup software packages now recognize and work with these snapshot features which they can leverage to perform application consistent, off-host snapshots of VMs which can then again be independently backed up.

Multiple deduplication options

Centralizing the backups of VMs carries another risk with it as well: added cost. Backing up the data on all of these new virtual servers can mean more data to protect, potentially a lot more. Since it is the corporate backup team that needs to manage and store all of this data, it inevitably introduces new costs that their budget may not be ready to absorb.

The good news on this front is two-fold. First, VM backups can deliver high deduplication ratios. Since much of their data is redundant, this is conducive for deduplication with a recent study showing that the deduplication of VM data can save 80% of more the space that is required without it.[1]

The other piece of good news is that deduplication is no longer limited to just the client. Deduplication technology is available on both the backup software media server as well as the backup appliance target. These multiple options to perform deduplication enable enterprises to implement deduplication in such a way that is optimal for the application, minimizes backup data stores and keeps costs under control.

Why Barriers to Centralized Virtual Server Data Protection Persist

Despite these advancements that facilitate the centralization of virtual server data protection, they alone are still insufficient. Implemented individually, they become difficult to centrally monitor, manage and schedule for the following reasons:

- These techniques alone provide no clear picture of how the overall virtual server backup environment is configured or performing.
- They do not track the data growth on individual VMs or across the enterprise.
- Each tool requires its own set of processes for installation, management and maintenance.
- They do not leverage deduplication of other applications that is occurring elsewhere in the enterprise
- Organizational breakdowns can occur in terms of defining responsibility for specific tasks. For example, once a VM snapshot is created and moved off to tape, who is then responsible for tasks like defining data retention or restores?

Best Practices for Centralizing VM Data Protection

Therefore in order for enterprises to truly optimize VM data protection within their environment as part of their broader corporate data protection strategy, there are some best practices that they can follow to achieve this ideal:

- *Assign a single owner that is responsible for the protection and recovery of VMs.* This individual's responsibilities should be clearly communicated throughout the organization. Further, individuals within specific departments and business units should be designated as points of contact who can ensure that their specific backup, recovery and data retention needs are meet.
- *Automated, centralized monitoring and reporting.* The creation of a VM can now occur as quickly as in a few minutes by an individual and their creation can even be automated though the use of scripting. This makes it a necessity that organizations have centralized reporting software that can monitor and report on the creation of these VMs. This is needed to identify what VMs are being backed up, if they are being backed up successfully and how fast backup data is growing. Exception reports also need to be created so organizations can identify what VMs are not yet being protected so these can be identified and added to the backup schedule.
- *Centralize the protection of physical and virtual machines using a single software platform.* A common software platform enables organizations to centrally schedule backup jobs, manage recoveries, monitor the success and failure of backup jobs and provide a common console that backup administrators can use to administer all backup jobs.
- *Choose a platform that integrates with data protection features in the leading virtual server platforms.* Integrating with leading virtual server platform features such as the vStorage API in VMware vSphere should be viewed as a prerequisite in order for them to successfully and expeditiously protect and recover these platforms. For example, the vStorage API

1 Keren Jin, Ethan L Miller, "The Effectiveness of Deduplication on Virtual Machine Disk Images", http://www.ssrc.ucsc.edu/Papers/jin-systor09.pdf, Website last referenced on April 23, 2010.

introduced the concept of Change Block Tracking (CBT) that can be used to identify used blocks in the VMDK file so only those blocks are backed up instead of needing to backup the entire VMDK file. This same feature then can also be used when doing incremental backups so only used blocks that have changed from the previous backup need to be backed up.

- *Deduplicate all backup data at all levels across the enterprise's physical and virtual environments.* Deduplicating just VM backup data helps to reduce data stores but centralizing all backup data from both physical and virtual machines and then deduplicating it can result in even higher storage savings. Further, the flexibility to deduplicate data at the client, media server or target in both physical and virtual environments enable enterprises to select and implement the form of data deduplication that is most appropriate for the application.
- *Provides granular levels of recovery for virtual machines.* One of the challenges when recovering data on VMs is to recover just the data that is needed from within the VM (a file, a user account, an email message, etc.). The method that a number of backup software products take for recovery is to restore the entire VM and then require the backup administrator to traverse the directories and files or, if recovering an email message, to comb through the entire email repository. To avoid this laborious method, the backup software should be knowledgeable about the content of the VM images that it is protecting so restores of individual items within the VM can occur quickly and easily.
- *Manages the replication and recovery of backup data for both physical and virtual environments to deliver centralized DR.* The management of the replication of the backup data can be as simple as making a copy of the data to tape or it can be as sophisticated as managing and tracking the replication of data that is natively occurring between two disk-based backup targets. Enterprise organizations will use any and all means of replication for both their physical and virtual environment as part of their day-to-day data protection strategies as well as their more strategic disaster recovery initiatives. By centralizing and managing the replication and recovery for both their physical and virtual environments, organizations can apply common sets of policies to each.

A Successful Backup and Recovery Strategy Depends Upon Both the Right Technology and the Right Processes

Server virtualization solves many of today's most pressing business problems, but it creates its own set of issues, with backup and recovery chief among them, that enterprises need to be equipped to handle.

Putting in place the right technology is certainly an important part of the equation to achieve this new reality of a successful centralized backup and recovery plan. To accomplish this, organizations should look to leaders such as Symantec NetBackup 7 that has made a significant investment to ensure that VMs can be protected and are part of the broader data management and protection solution that enterprises put in place as they bring both their physical and virtual servers under centralized management.

However organizations need to do more than purchase the right technology to successfully achieve this end. They also need to put the right internal processes in place that will both facilitate the introduction of these centralized backup technologies as well as ensure their successful implementation and ongoing management.

By implementing an appropriate data protection solution and setting up the right processes to support them, they should be able to not only realize the initial benefits that server virtualization delivers but successfully avoid some of the challenges that can accompany it. ■

DCIG, LLC | 7511 Madison Street | Omaha NE 68127 | 402.884.9594
dciginc.com