White Paper

Improving Backup and Recovery for VMware vSphere Environments

By Mark Bowker

July, 2010

This ESG White Paper was commissioned by EMC and is distributed under license from ESG.
Contents

Introduction .................................................................................................................................................. 3

Server Virtualization ..................................................................................................................................... 3
  Impact on Backup and Recovery .............................................................................................................................. 4
  Backup and Recovery Approaches ........................................................................................................................... 5
  Considerations for Protecting Virtual Environments ............................................................................................... 6

EMC Avamar Efficiently Addresses VMware vSphere Data Protection Needs ............................................ 7

The Bigger Truth ........................................................................................................................................... 8
Introduction

Data deduplication adoption has made impressive strides over the past two years, likely fueled by interest in disk-based backup and recovery, the impact of the economy on IT budgets, and the proliferation of server virtualization. EMC incorporates data deduplication as a native feature to deliver highly efficient backup and recovery—a must for rapidly expanding virtual server environments. EMC can accelerate the adoption of VMware since data protection is an imperative part of any virtualization deployment strategy. VMware vSphere 4 and the vStorage API for Data Protection introduced several enhancements including changed block tracking (CBT) and virtual proxy servers. This paper details the challenges of protecting VMware vSphere environments, but more importantly examines the potential for backup improvements when compared to physical environments. Specifically, it offers insight into how EMC and its integration with vSphere technology provide more efficient data protection.

Server Virtualization

Server virtualization enables multiple virtual machines to run on a single physical server and share numerous computing resources, eliminating the economic and operational issues of infrastructure silos. Deploying a dedicated physical infrastructure for an application isolates it from outside impact. However, it also results in lower resource utilization, the need for high availability safeguards, workload spikes, and inflexibility to accommodate changes. Employing server virtualization addresses the issues of siloed infrastructure by improving hardware utilization, consolidating physical servers (which helps to lower operating costs through optimized power, cooling, and space efficiency), increasing operational agility, and facilitating higher availability.

Due to the many benefits it provides, server virtualization is gaining wider adoption. ESG research found that 61% of research respondents are currently using server virtualization and another 30% have plans to (see Figure 1). Those using server virtualization—even with only a fraction of their total production servers—have plans to expand. IT organizations plan to increase the use of server virtualization, ranking this initiative as the number one most important IT priority over the next 12-18 months. This entails driving the consolidation ratio higher and deploying more applications in virtual machines.

Figure 1. Server Virtualization Use and Interest

<table>
<thead>
<tr>
<th>Is your organization currently utilizing a server virtualization solution? (Percent of respondents, N=304)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, we are currently using server virtualization, 61%</td>
</tr>
<tr>
<td>We have not yet deployed server virtualization but we plan to do so, 30%</td>
</tr>
<tr>
<td>We have not yet deployed server virtualization and we have no current plans to do so, 9%</td>
</tr>
</tbody>
</table>


---

One of the byproducts of server virtualization is often a net growth in total storage volume due to proliferation and redundancy. As shown in Figure 2, ESG found the difference in average capacity between those using and not using server virtualization to be compelling.  

Figure 2. The Impact of Server Virtualization on Overall Volume of Storage Capacity

Impact on Backup and Recovery

Server virtualization creates an opportunity to improve backup and recovery. Encapsulating the virtual machine workload into a single file containing the operating system, applications, and data enables mobility and removes hardware dependency. The ability to capture the virtual machine system state enables bare metal-like recovery.

Organizations deploying server virtualization often struggle with the backup challenges resulting from the transition from a physical to a virtual environment. The aforementioned increase in the volume of data could result in more time required for backup processes—often stressing already-stressed backup windows.

There are implications for sharing physical resources as well. A physical server environment has low resource utilization, but plenty of bandwidth for backup. The new paradigm, on the other hand, has higher resource utilization, but less bandwidth for backup. Simultaneous resource-intensive processes occurring on one physical ESX host system could cause resource contention—potentially impacting the pool of application workloads sharing the common physical resources and causing performance issues.

IT organizations see backup and recovery of virtual machines as an area of concern. In fact, improving virtual machine backup and recovery was ranked second by respondents to ESG’s 2010 IT Spending Priorities research survey when asked about their top server virtualization initiatives for 2010 (see Figure 3).  

---

In ESG’s Data Protection Market Trends survey, 55% of respondents cited use of separate backup solutions for virtual and physical environments. It’s likely that these organizations believe that implementing virtualization-specific backup and recovery solutions better address their backup and recovery needs in virtual environments.

**Backup and Recovery Approaches**

There are a few ways to back up and recover virtual machines, including:

**Installing an Agent in the Guest Operating System**

Applying agent-based, file-level backup strategies to virtual server backups is similar to how backup is done in the “physical” world. This approach has the advantages of familiarity and simplicity, and when backup agents support applications, it also allows for assured, application-consistent backup and recovery.

The most significant drawback to the guest backup approach is resource contention. Backups demand significant processing power; concurrent or overlapping virtual machine backups can place a burden on the VMware host server’s CPU, memory, disk, and network components and often make it difficult or impossible to complete backups within available windows.
Virtual Machine Image-level Backup via a Proxy Server

This method leverages VMware tools to capture the virtual machine disk image and, in conjunction with the backup application, deliver it via a proxy server to the target backup storage. It is less disruptive to the virtual machine-resident applications and is less likely to overwhelm the host’s CPU. This method eliminates the backup window by off-loading the backup process to a proxy server. However, this method typically doesn’t support file-level or application object-level recovery without at least a two-step recovery process.

Image backup makes it easier to recover the whole system (no need to re-create the virtual machines or system settings), including the entire application—but it’s important to understand the application environment in order to be assured that backups will be consistent and recoverable. A concern with image backup is its ability to provide fully application-consistent backup and recovery. For application-consistent backup of virtual machines running Windows, VMware relies on Microsoft Volume Shadow Copy Service (VSS). VSS quiesces applications such as Microsoft SQL and Exchange before initiating backups. At present, VMware supports VSS “Copy Backup” mode rather than “Full Backup” mode. VSS Copy Backups do not automatically truncate application transaction logs whereas VSS Full Backups do. Applications (databases in particular) need to know when to truncate their transaction logs (typically, at backup time) in order to avoid over-running available space and also to support log-shipping capabilities.

Considerations for Protecting Virtual Environments

When virtualizing servers, several factors should be considered to optimize data protection:

Supporting multiple methods of (file- or image-level) backup and recovery. The business value of applications and data dictates protection strategies. For each class of application and data, risks must be assessed. Backup solutions that can support different methods of backup will be desired.

Overcoming challenges of resource contention. As the ratio of virtual to physical servers increases, there is a greater opportunity for the aforementioned resource contention between virtual workloads to compete for limited shared resources. Solutions should be evaluated to ensure that the backup load is minimal.

Ease of management. While every data protection solution has a management interface, vCenter forms the foundation for virtualization management in a VMware environment. Integration, therefore, with this central management console simplifies administration of virtual workloads and introduces efficiency.

Cost-efficiency. Vendors that offer the features to enable more effective and cost-efficient data protection are favored. Straightforward management, simplified licensing, rapid and non-intrusive backup and recovery, and technology that reduces bandwidth and storage requirements allow IT organizations to deliver improved service levels, lower costs, and reduced operational overhead.

Fully leverage hypervisor features. VMware vSphere introduced several features that greatly improve data protection processes:

- One of the more impactful enhancements is the vStorage API for Data Protection (VADP). VMware replaced VMware Consolidated Backup (VCB) with new APIs for data protection, greatly improving the implementation of backup for the platform. Backup vendors integrate with vSphere VMFS drivers to access data on VMFS volumes. The VADP allows a live system image snapshot to be captured without impacting applications or over-taxing the host’s CPU. A physical proxy server is no longer required (as it was with VCB), which reduces the infrastructure commitment because the proxy system can now be a virtual machine.

- New to vSphere is changed block tracking (CBT). This feature tracks changed blocks of a virtual machine’s virtual disk. This allows backup applications to immediately identify the blocks changed since the last backup and copy only those blocks, reducing backup time and network traffic. For recovery with most backup applications, a temporary full backup is required, to which the changes are applied.
EMC Avamar Efficiently Addresses VMware vSphere Data Protection Needs

EMC Avamar is a deduplication backup and recovery solution that delivers fast, daily full backups for VMware environments. Unlike traditional backup methods, Avamar eliminates redundant data segments at the client (source) before data is transferred across the network and stored to disk.

Traditional backup solutions are often inefficient because they back up too much. Daily incremental and weekly full backup schedules make the impact of duplicate data staggering. This is especially severe for virtual environments since each virtual machine represents an individual backup job and includes redundant operating system, application, and file data. As a result, backup processing for virtual machines can often overrun backup windows and tax shared resources, leaving data unprotected and creating management issues for backup administrators. Unlike traditional solutions, Avamar provides daily full backups that can be quickly recovered in just one step—eliminating the hassle of restoring full and subsequent incremental backups to reach the desired recovery point.

In vSphere environments, Avamar can take advantage of both guest and image-level backup approaches to best meet the requirements for particular backup workloads. In the case of VMware guest backup, Avamar can be configured for file- or application-level protection. Avamar deduplication within the virtual machine moves minimal backup data, speeds backup, and reduces resource contention. This involves installing Avamar agents in virtual machines; operating system-specific backup agents for Windows, Linux, or Solaris; or application-specific backup agents for application-level, transaction-consistent backup of SQL, Oracle, Exchange, SharePoint, DB2 or Lotus Domino.

Server consolidation drives higher utilization of server hardware, leaving very little headroom for IO-intensive backup processes. VMware backups with Avamar deduplication require less server resources for shorter durations, removing the bottleneck and freeing up server resources to run more virtual machines. More virtual machines per server results in greater server consolidation, lower costs, and better return on investment.

To deliver virtual machine image-level backup, Avamar leverages the vStorage APIs for data protection. Avamar uses a virtual proxy server, where an Avamar agent resides. The efficiency twist is that Avamar automatically assigns virtual machine backup jobs to available proxy servers using a “round robin” protocol. Backups are distributed evenly across available virtual proxy servers, creating more efficient use of available resources. Jobs can be directed to available proxies to handle high-priority workloads. This approach drives down costs by having fewer proxies to manage. Avamar deduplication creates greater efficiency, eliminating concerns with virtual proxy server sprawl.

Avamar has an architecture that was designed to reconcile daily changed block backups into full instances for recovery, differentiating it from most backup applications that require a periodic full backup. This functionality enables Avamar to process the changed blocks identified by vSphere CBT into a new full image on an ongoing basis. Recovery is faster and more reliable because Avamar always maintains an up-to-date full image. Second, taking advantage of vSphere CBT can greatly reduce backup processing time.

Avamar image-level backups leverage an architecture which includes a native metadata index for each backup instance. Unlike other backup software, which must create a metadata index during or after the backup is complete, Avamar’s native index provides end-users with a granular view of image backup content, which can be used to perform file-level recovery without first requiring the image backup to be mounted. In addition, files can be restored to their original location or to an alternative location.

Rounding out Avamar’s integration with vSphere is its monitoring and management capabilities via VMware vCenter. This integration allows IT organizations to monitor backup and recovery operations in the Activity Monitor and view virtual machine protection policies (“guest,” “image,” and “none”) as well as the date and time of the most recent backup. This makes it very easy for operators to identify newly created VMs that may not be adequately protected. From vCenter, virtual machines may be added and backup policies defined for them.
The Bigger Truth

As server virtualization deployments expand in breadth and depth, IT organizations experience firsthand many of the inefficiencies of deploying data protection in the manner applied to physical server environments. The challenges for data protection—such as an increase in the volume of data to protect, difficulty meeting backup windows, backup processes in virtual machines disrupting normal processing of virtual machines sharing the same physical resources, increases in secondary storage capacity requirements, and higher costs—make the VMware environment ripe for improvement.

VMware vSphere addresses many of these challenges. Data is captured, transferred, and stored more efficiently by backup applications that take advantage of vSphere enhancements. EMC Avamar adds another efficiency dimension by not only leveraging vSphere functions and features, but by taking innovative approaches to economize and streamline backup and recovery processes.