

White Paper

Backup and Recovery of Large-scale VMware Environments

*By Mark Bowker, Senior Analyst
and Jason Buffington, Senior Analyst*

February 2012

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

Contents

Introduction	3
Virtualization Priorities	4
Backup and Recovery Approaches	5
Considerations for Protecting Virtual Environments	6
Data Protection at Scale	6
VMware Enables Faster Recovery	7
EMC Avamar Eases Data Protection for Large Virtual Deployments	8
Management	8
Deduplication	8
Sidebar: Vblocks.....	9
Other Avamar 6 Features	9
Enterprise Deployment Considerations	9
Avamar and Virtual Computing Environment Vblock Infrastructure Platforms	10
The Bigger Truth	11

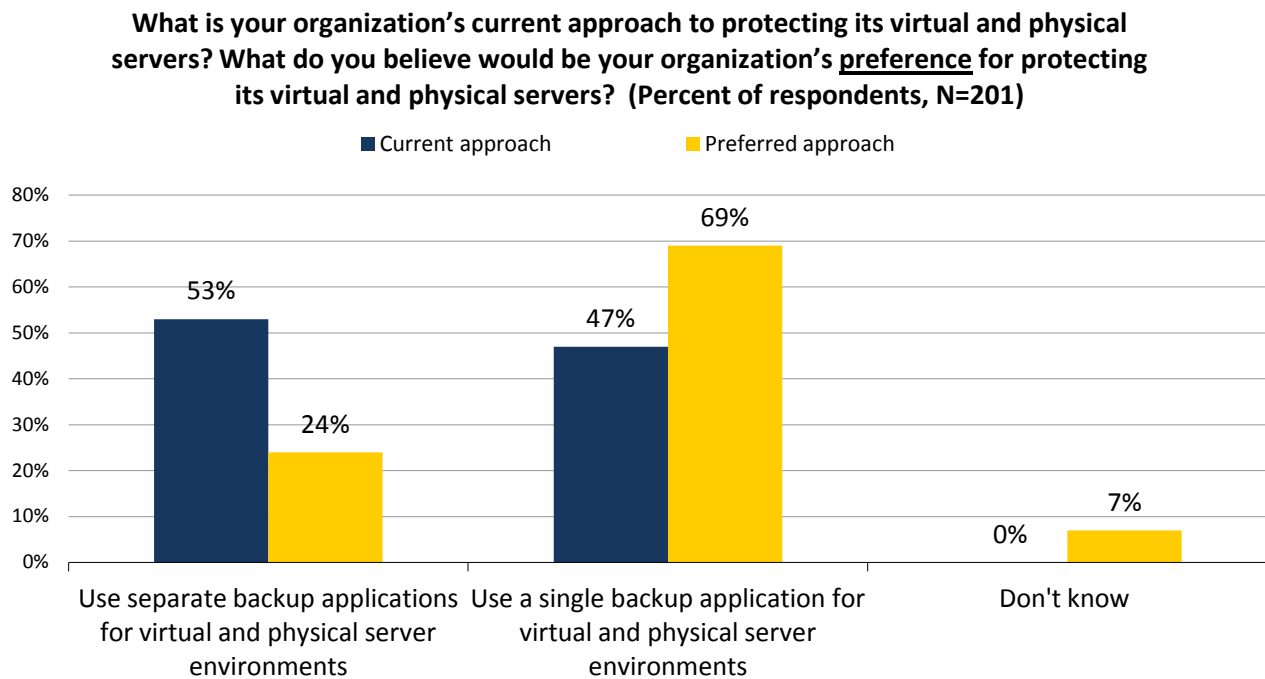
All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.

Introduction

Rapid growth in the adoption of server virtualization at scale is continuing, with more and more organizations deploying server virtualization as a strategy for consolidation, cost reduction, higher resource utilization, and greater efficiency. Phased adoption seems to be the standard. Organizations begin with IT-based applications residing in the data center and, after discovering the benefits and becoming more comfortable with the technology, they begin to expand virtualization deployments. As they start virtualizing user-facing applications, new challenges arise. For those applications, consolidation is not the focus; high availability and data protection are. As a result, any task that limits productivity or scalability, such as I/O-intensive backup, must be examined carefully.

Traditional backup processes aren't always the best solution in virtual deployments. They weren't built for the extreme data redundancy and hefty server loads of consolidated environments. The level of virtual and physical heterogeneity often requires environment-specific backup solutions for efficient protection. In fact, recently conducted ESG research revealed that more than half of virtualization users (53%) currently employ separate backup applications for virtual and physical server environments¹ (see Figure 1).

Figure 1. Data Protection for Virtual and Physical Servers, Current Approach Versus Preferred Approach



Source: Enterprise Strategy Group, 2012.

These users believe they need a hypervisor-specific or virtualization-only solution to protect virtual machines. In a traditional data center, physical servers host individual applications. Server resources are significantly underutilized, but that leaves plenty of processing power for backup. In a consolidated virtual server implementation, however, multiple virtual machines (VMs) share the same physical hardware. Resources are better utilized.

This is cost efficient, but it leaves fewer resources for backup. Also, more data must be backed up due to redundant operating system images, application profiles, and data. With fewer backup resources, an obvious bottleneck arises. It is exacerbated as virtual environments scale to hundreds, even thousands, of virtual machines sharing a common resource pool. These deployments need a backup/recovery solution designed to handle their needs.

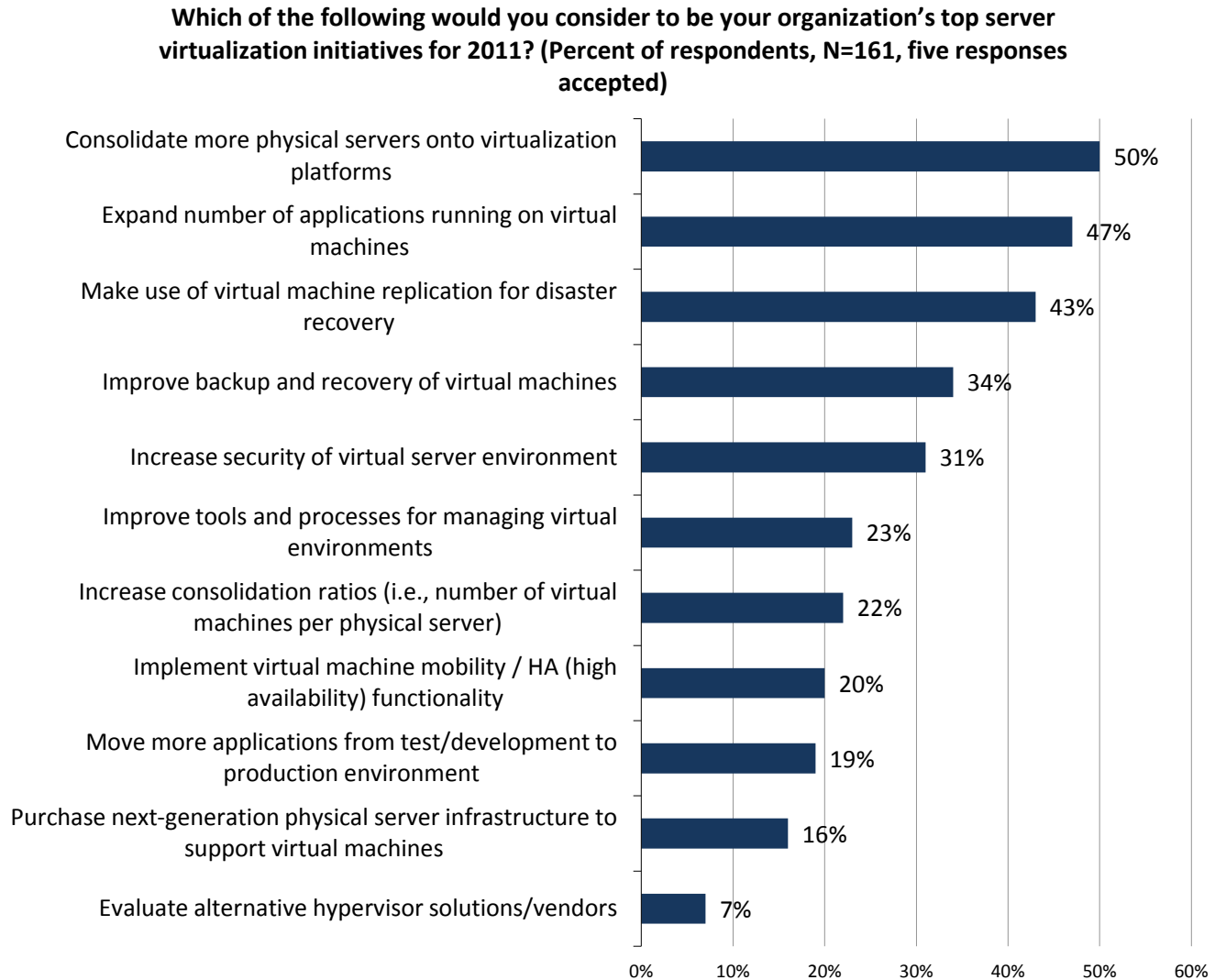
The best advice is to *plan* for data protection when strategizing for future virtualization growth instead of waiting to face problems and risking a breakdown of virtualization goals.

¹ Source: ESG Research Report, *The Modernization of Data Protection*, report to be published.

Virtualization Priorities

ESG asked survey respondents what they considered to be their top server virtualization initiatives for 2011. As Figure 2 shows, the top four responses were focused on expanding virtualization deployments and improving data protection, including the backup and recovery of virtual machines.² As the ratio of VMs to hosts increases and more applications are virtualized, backups will take longer and consume more resources. Reducing the amount of data being backed up is essential to ensuring uncompromised production operations. ESG believes that most organizations will struggle to expand virtualization deployments without first improving virtual machine backup and recovery.

Figure 2. Top Server Virtualization Initiatives for 2011



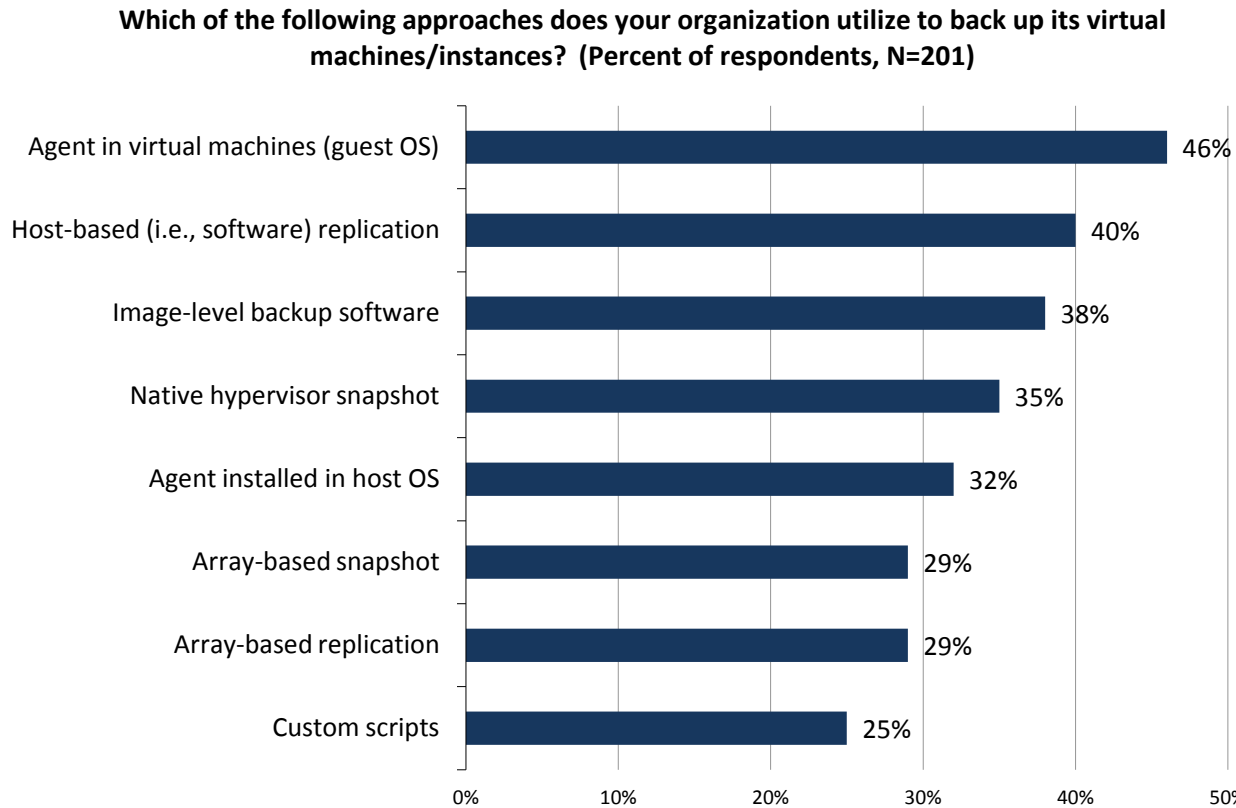
Source: Enterprise Strategy Group, 2012.

² Source: ESG Research Report, [2011 IT Spending Intentions Survey](#), January 2011.

Backup and Recovery Approaches

As Figure 3 shows, there are several ways to back up and recover virtual machines.³

Figure 3. Virtual Machine Backup Methods



Source: Enterprise Strategy Group, 2012.

Installing an Agent in the Guest Operating System

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

The most significant drawback to the guest backup approach is resource contention. Backup processes typically 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 they often make it difficult or impossible to complete backups within available windows.

Virtual Machine Image-level Backup via a Proxy Server

This method leverages the VMware vStorage APIs for Data Protection (VADP) to capture the virtual machine disk image(s) and, in conjunction with the backup application, deliver the image(s) 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 offloading the backup process to a proxy server. However, the method typically doesn’t support file-level or application object-level recovery without at least a two-step recovery process.

³ Ibid

Image backup makes it easier to recover the whole system (no need to recreate virtual machines or system settings), including the entire application. But it's important to understand the application environment enough 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 Server 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) to avoid overrunning available space and to support log-shipping capabilities.

Considerations for Protecting Virtual Environments

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

- **Supporting multiple methods of (file- or image-level) backup and recovery.** The business value of applications and data dictates the protection strategies. For each class of application and data, risks must be assessed. Backup solutions that support different backup methods are desirable.
- **Overcoming challenges of resource contention.** As the ratio of virtual to physical servers increases, a greater opportunity arises for the aforementioned resource contention among virtual workloads competing 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. Therefore, integration with this central management console simplifies administration of virtual workloads and introduces efficiency.
- **Cost-efficiency.** Vendors that offer 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 improve service levels, lower costs, and reduce operational overhead.
- **Fully leveraging hypervisor features.** VMware vSphere 5.0 introduced several features that greatly improve data protection processes:
 - ✓ One of the highest-impact enhancements is 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 Virtual Machine File System (VMFS) drivers to access data on VMFS volumes. VADP allows the IT organization to capture a live system image snapshot without affecting applications or overtaxing 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, then copy only those blocks, reducing backup time and network traffic. For recovery with most backup applications, a temporary full backup—to which the changes are applied—is required.

Data Protection at Scale

Traditional data protection models were designed for a physical infrastructure in which isolated applications were matched with hardware and backed up individually. If the backup were to interfere, only that one application would be affected. However, now that multiple virtual machines can reside on each physical server, the backup process becomes more complicated. Contention for resources can cause slowdowns and can interrupt multiple applications

simultaneously. Scaling virtualization deployments effectively requires overhauling traditional data protection models. Backup and recovery must transform from tape-based storage to capacity-optimized disk because tape simply cannot deliver the performance and availability that new consolidated data loads require.

Scaling a virtual environment can be easier if its impact on backup and recovery is considered in advance. By proactively building a backup and recovery strategy into the overall virtualization plan, an organization can avoid some headaches.

Figure 4. Considerations for Backup and Restore at Scale in VMware Environments

Cost Effective	<ul style="list-style-type: none"> Physical and virtual Integrated into converged infrastructure offerings
Management	<ul style="list-style-type: none"> Integration with virtualization platform Problem identification/resolution
Retention	<ul style="list-style-type: none"> Capacity efficiency Maintain multiple retention periods Structured and unstructured data
Restore Performance	<ul style="list-style-type: none"> File, image, and site recovery Application consistency Improved RTOs (recovery time objectives)
Backup Performance	<ul style="list-style-type: none"> Maintain backup windows with scale Improved RPOs (recovery point objectives)
Production Platform Infrastructure	<ul style="list-style-type: none"> Consistent performance at scale Virtualization integration Backup and recovery to disk

Source: Enterprise Strategy Group, 2012.

VMware Enables Faster Recovery

While backup in a virtual server environment can be challenging, virtualization actually enables faster recovery in terms of provisioning and getting data back online. In a VMware implementation, a virtual machine workload is encapsulated into a single file containing the operating system, applications, and data. This file can be moved or copied anywhere. If it is copied to another server while the first is backed up, very little production time is lost because the backup no longer interferes with productivity—a VM is not dependent upon particular hardware. Also, because the entire application stack the user needs is encapsulated in a VM, it is simpler and faster to recover. Instead of a sequential recovery process—operating system, then application, then data—everything is recovered in one step (using application-specific guest OS agents) or two steps (using image-level backup and virtual proxy servers).

In addition, advances in VMware offerings such as VADP and CBT have greatly improved backup processes. It also enables virtual (instead of physical) proxy servers. CBT monitors which disk blocks of the VM virtual disk are being written to, which enables backing up only those changed blocks (without having to compare or perform other I/O-intensive decision processes), reducing backup time and network traffic. These improvements can make a tremendous difference in backup time and storage capacity as an environment grows.

EMC Avamar Eases Data Protection for Large Virtual Deployments

As virtualized deployments scale, it's important to know what to expect in order to plan for and avoid problems. A key part of virtualization scaling involves dealing with a mix of business applications, not just IT system workloads. The farther along an organization gets in its virtualization journey, the greater its data protection needs will be because the applications tend to be larger, more difficult to manage, and focused on different priorities. A key to successful large-scale adoption involves virtualizing test beds and unstructured data applications. For organizations that rely on Microsoft Exchange, SAP, Oracle databases, and/or desktop virtualization solutions, backup is even more important. They need tools and solutions that meet the service level expectations of business users—applications must be fully protected, often remotely replicated, yet remain constantly available.

Expanded virtualization results in net growth of storage volumes. Daily incremental and weekly full backups create a lot of duplicate data as each VM's backup job includes OS, application, and file data. How do you continue to fully protect data and still minimize the CPU, bandwidth, and storage needed? Scaling creates more opportunities for resource contention as virtual workloads compete for limited, shared resources. So, the smaller the backup load, the better.

[EMC Avamar](#) delivers highly efficient, deduplicated backup and recovery, which is a must for rapidly expanding virtual server environments. Because reliable data protection is imperative for a virtualization strategy, Avamar can actually help accelerate VMware adoption as it meets all the requirements for optimizing backup and recovery in a virtual environment.

Management

EMC Avamar supports both guest- and image-level backups with support for VADP. It minimizes the backup load; enables cost-efficiency in management, licensing, bandwidth, and storage; and leverages the full hypervisor feature set. In addition, Avamar is integrated with VMware vSphere 5.0, so it can be monitored and managed via vCenter. As a result, IT can 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.

For image-level backup, Avamar employs a unique proxy pool algorithm that scales up and down based on backup and restore demands. This allows the backup job to occur on the first available proxy server as opposed to waiting for a dedicated proxy server. There are no hard assignments of virtual machines to proxy servers or mapping of storage LUNs to proxy servers. Avamar manages proxy load-balancing automatically, requiring no additional work by the IT team.

These management features:

- Help improve backup performance
- Scale up with performance upon demand to help quickly onboard new applications
- Scale out to meet the needs of large VMware environments
- Are tightly integrated with the management platform for rapid change integration

Deduplication

The enabling technology for Avamar is source-based deduplication. Instead of copying all of the data during every daily backup, Avamar only copies new, unique, sub-file, variable-length data segments. Compared with traditional whole-file backup methods, Avamar can reduce the daily impact on virtual and physical infrastructure. Traditional backup software moves approximately 200% of primary backup data weekly; Avamar moves far less over that same period. Avamar's global data deduplication can eliminate the cost of storing and moving redundant data across thousands of geographically separate VMs.

The implications for resource contention, network bandwidth, and storage requirements are obvious: Deduplication at the source reduces all of those factors. What makes this so important in a growing virtual environment is that it enables greater levels of server consolidation and actually makes expansion possible for many organizations. Without optimization to something better than “whole file” protection, virtualized environments suffer backup bottlenecks, bandwidth problems, contention for server resources, and missed backup windows. IT can also benefit from minimizing the resources required for the backup and restore process and confidently predict utilization based on application workload without having to factor in the impact of backups and restores on the virtualization platform.

What makes Avamar an efficient, client-based data deduplication engine is its intelligent algorithms for sub-file, variable-length data segments. Other solutions that use fixed-block or fixed-length data segments to deduplicate can be fooled by logical data shifts, such as inserting data into the beginning of a file. Avamar’s algorithms quickly determine logical boundary points and redundant data segments. Because it so dramatically reduces the amount of data copied, organizations can retain full backups on disk for longer periods of time, making data more quickly recoverable.

Other Avamar 6 Features

Avamar also provides single-step recovery, completely eliminating the tasks of restoring a previous full backup plus subsequent incremental backups. Avamar’s support for VMware CBT, coupled with its own inherent capability to reconcile changed data into recoverable backup images, gives it a recovery performance advantage over most traditional approaches.

In addition, Avamar can offer granular, file-level recovery from image-level backups by opening the image and leveraging its own integrated file system to present the directory structure to the administrator. Finally, with Avamar, administrators have the flexibility to recover data to the originating virtual machine or to create, provision, and recover data to a brand-new virtual machine—all managed from within the Avamar user interface. For disaster recovery, Avamar provides secure, efficient, virtual-to-virtual or virtual-to-physical replication.

Enterprise Deployment Considerations

Avamar itself scales using a grid architecture, offering linear increases in both performance and capacity as nodes are added. Each node increases CPU, memory, I/O, and capacity for the entire grid.

Avamar offers equivalent benefits to both physical and virtual environments and supports all major proprietary and open-source operating systems. It includes support for major database and messaging applications from IBM, Microsoft, and Oracle, as well as the network data management protocol (NDMP) for NAS filers. Avamar’s interoperability enhances its scalability; organizations are not restricted as they expand their virtual server environments.

Sidebar: VCE Vblocks

Large-scale virtualization requires tight integration among servers, networks, and storage. More organizations and service providers are using converged infrastructure units, such as VCE Vblocks from VMware, Cisco and EMC.

These bundled compute, network, and storage resources offer a consumption model that can dramatically simplify delivery and accommodate massive scaling of virtualized environments. IT gains visibility into the infrastructure and can maintain performance proactively. Capacity can also be served up much faster. Massive-scale virtualization results in data growth and in greater power consumption and management complexity; Vblocks reduce power consumption and simplify management. As server virtualization deployments focus on SAP, Oracle, and Windows workloads, Vblocks can help speed deployment and time to value.

Vblocks can contain thousands of virtual machines with a mix of application workloads and data sizes. The high-end configuration, Vblock Series 700, is intended for 3,000 to 6,000 VMs, while the midsize Vblock Series 300 is designed for 800 to 3,000 VMs. All Vblocks are built with VMware vSphere 5.0, Cisco compute and networking equipment, and EMC storage. The three industry leaders have gone to great lengths to provide integrated tools, services, and support to ensure that IT services are delivered fast, without hiccups and bottlenecks. However, if not considered in advance, the data protection needs of these Vblocks could cause performance problems—exactly what the VCE vendors have worked so hard to avoid.

Avamar and Virtual Computing Environment Vblock Infrastructure Platforms

Avamar is well-suited for Vblock configurations (see sidebar). Pre-integrated units consolidate a lot of IT resources; Avamar can reduce the amount of network bandwidth and backup storage required.

Avamar also adds another benefit: Business users often need to be convinced to allow their production applications to be virtualized to build their overall comfort in the virtualized platform. (They are wary of change and its potential impact on productivity.) These end-users have similar concerns about converged units such as the Vblock. Avamar, with its ability to eliminate backup resource contention, bottlenecks, and performance problems, can actually help demonstrate to business owners the value of the Vblock infrastructure and its absence of negative impact on application performance in large-scale environments.

In addition to protecting VM data, Avamar also protects infrastructure data. This enables bare-metal restoration of Vblock components, including the vSphere SQL database, the Cisco UCS profile, boot information and IP ranges, the external connectivity configuration files, the storage configuration, and EMC Ionix Unified Infrastructure Manager.

Whether the “infrastructure as a service” is coming from an internal IT department or a service provider, Avamar can help guarantee enough performance, availability, and data protection to build user confidence.

The Bigger Truth

Backup redesign is a must if customers are to reach their virtualization goals—without it, scaling one’s virtualization and/or private cloud environment is nearly impossible. Next-generation backup solutions must align with next-generation IT by doing more than just escalating server virtualization. They need to support resource and management consolidation, infrastructure as a service, cloud services, and pre-integrated infrastructure bundles.

Backup and recovery processes built for the physical world will not provide the required data protection, application performance, or storage optimization needed in these new environments. Only with a backup process built for this kind of infrastructure can organizations reap the operational benefits and ROI of large-scale deployments.

VMware recognized this. As a result, it introduced vSphere 5.0 APIs for data protection and change block tracking to keep data protection of larger systems from becoming a bottleneck. The combination of VMware enhancements and Avamar architecture creates an exceptional solution. With Avamar, data is captured, transferred, and stored more efficiently because it takes advantage of vSphere enhancements. Avamar adds another efficiency dimension by taking innovative approaches to economize and streamline backup and recovery processes. It drives higher VM-to-ESX server ratios, supports all methods available (including guest, image, and remote office backup), and optimizes capacity.

When it comes to data protection, IT organizations can either proactively build a solution such as Avamar into their virtualization plans or wait for problems to emerge, then fix the backup process *and* restore user confidence in the virtual infrastructure at the same time. In standard environments, a backup and recovery solution based on deduplication is extremely helpful; in a very large virtualized infrastructure, it’s an absolute necessity.



Enterprise Strategy Group | **Getting to the bigger truth.**