Managing the information that drives the enterprise

STORAGE

ESSENTIAL GUIDE TO

Virtual server backup technologies

Get the lowdown on VMware and Hyper-V backup technologies including how deduplication, enterprise backup modules and third-party applications can ease the burden of virtual server backup.

INSIDE

TechTarget

VEEAM

1 VMware Backup

vPower™ : Virtualization-Powered Data Protection™

SureBackup™ + InstantRestore™ + SmartCDP™

Reliability	RTO	RPO
100% Guaranteed Recovery	Fastest Recovery Time	Minimal Data Loss

5 Patents Pending!

NEW Veeam Backup & Replication v5

vPower is enabled by 5 patent-pending technologies, including:

- **Instant VM Restore** - restore an entire VM IN MINUTES by running it directly from backup
- **U-AIR™ (Universal Application-Item Recovery)** - recover individual application items from ANY application and ANY OS
- **SureBackup™ - 100% Guaranteed Recovery** – automatically verify the recoverability of EVERY backup, EVERY VM, EVERY time

To learn more, visit www.veeam.com/vPower

Plenty of alternatives for backing up virtual servers

What's been a boon to the server side of the shop has been the bane of storage. But new tools and new approaches are making backing up virtual servers a lot easier.

SOMETIMES A VERY delicate balance is required to keep an IT shop running smoothly. Although there are clearly disparate disciplines at work in the data center, the gears have to mesh at key points to ensure that all the moving parts are working together well enough to press on in the same direction. When server virtualization technology matured and was embraced wholeheartedly by the systems side of the house, storage managers may have taken little notice. But when server virtualization began to approach critical mass, it was clear that storage had some catching up to do, and that backup was the first area that had to be addressed.

Server virtualization changes the environment in both subtle and profound ways. Operationally, things may not be so different and traditional data management methods—as with backup—may appear to fit in nicely. Such is the case in many shops today where backup methods have been barely altered to accommodate virtualized servers. Running a standard backup agent in each virtual machine may work okay today but, sooner or later, that approach is bound to stumble.

The problem is that traditional backup apps see things on a one-to-one basis—one OS running on one physical server—and count on having those physical resources to their jobs. But as multiple virtual machines (VMs) begin to share a single hardware platform, all bets are off and contention for the resources is inevitable. Without proper management, performance will suffer and basic activities—like backup—can be stopped in their tracks.

Citrix, Microsoft, VMware and other hypervisor vendors recognized the problem and offered server-centric remedies like VMware Consolidated Backup (VCB). There's no doubt that applications like VCB did provide remedies for some of the backup problems, but they did so with the cost of added complexity. New data protection apps also emerged from vendors like PHD Virtual, Quest Software (formerly Vizioncore) and Veeam that were designed from the ground up to operate in often unpredictable virtual environments. These new backup apps do an admirable job, but they're often shackled by the weight of having to elbow

their way into—and fit into—enterprise environments that are set in their ways.

Traditional backup vendors were a little late to the game in responding to the new requirements of virtual server architectures. But with some help from the hypervisor vendors who are exposing APIs, nearly every backup app vendor now offers hooks specifically designed to back up data from virtual servers.

Backup in a virtualized server environment—or even a virtual desktop environment—is definitely still a work in progress. But the good news is that it has already progressed significantly in a relatively short time and while the backup solutions available today may not be perfect for all use cases, there's enough choice and variety of approaches to serve most needs.

But, as we all know, backup is never easy. You'll probably need to study and test the available options before you find the right one for you shop. In this *Storage* magazine Essential Guide, we provide a crash course in virtual server backup, describe the alternatives and offer some best practices for implementing a successful virtual server backup operation. ⊙

Rich Castagna (rcastagna@techtarget.com) is the editorial director of TechTarget's Storage Media Group.

EMC²
where information lives®

Next-generation backup for virtualization starts now.

Virtualization benefits your data center. But your current backup infrastructure simply can't keep up. Now is the time to transform your backup environment: improve data protection, increase efficiency, reduce TCO. As the leader in disk-based backup and recovery, EMC lets you backup virtual environments to disk more economically than tape. See payback in just 12 months: using up to 98% less storage and 99% less bandwidth, while achieving greater server utilization and consolidation.

Solving today's backup challenges for VMware starts now with EMC.

To learn more download the free white paper: *Optimizing Data Protection for Virtualized Environments*

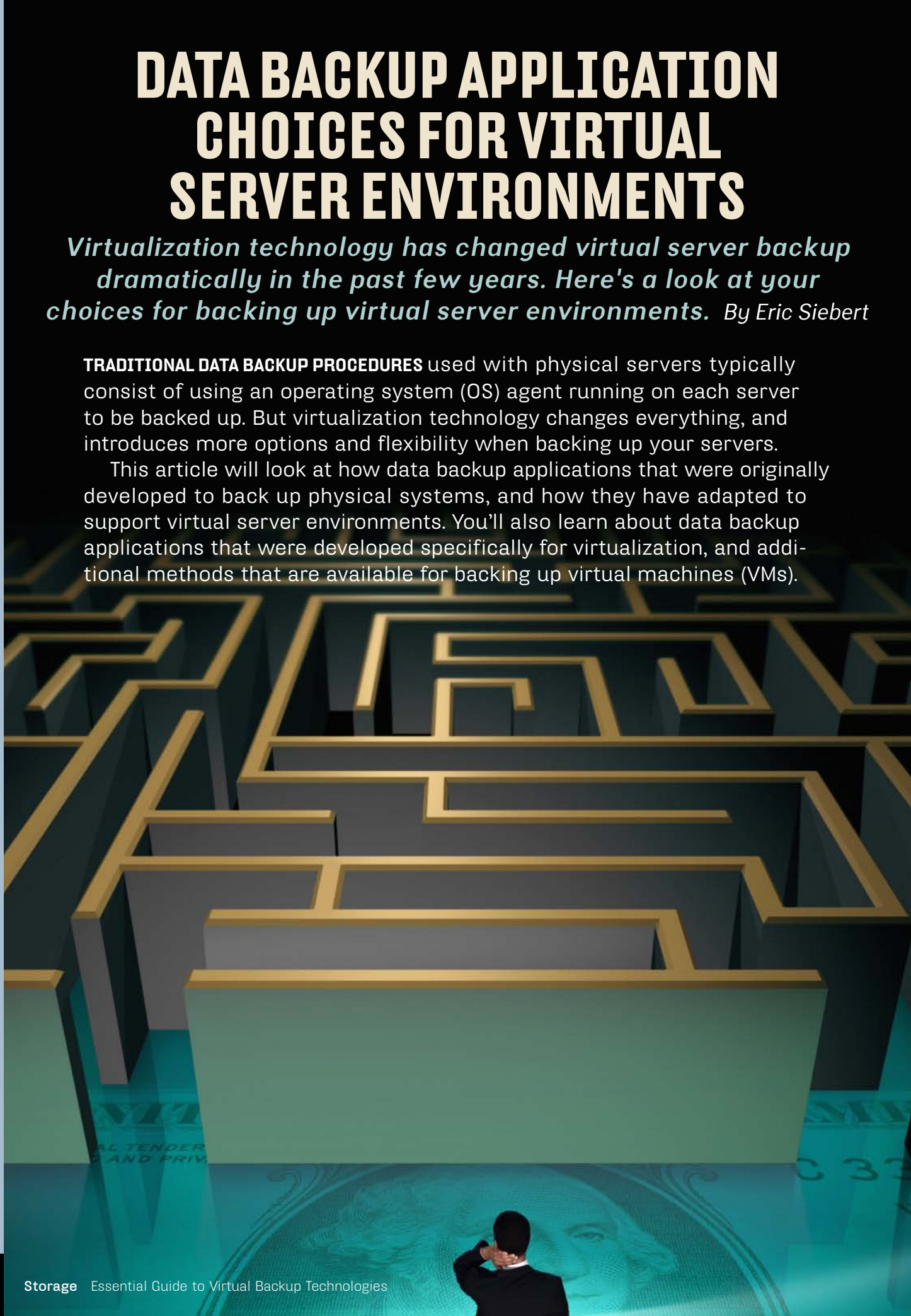
NEXT starts now.

DATA BACKUP APPLICATION CHOICES FOR VIRTUAL SERVER ENVIRONMENTS

Virtualization technology has changed virtual server backup dramatically in the past few years. Here's a look at your choices for backing up virtual server environments. *By Eric Siebert*

TRADITIONAL DATA BACKUP PROCEDURES used with physical servers typically consist of using an operating system (OS) agent running on each server to be backed up. But virtualization technology changes everything, and introduces more options and flexibility when backing up your servers.

This article will look at how data backup applications that were originally developed to back up physical systems, and how they have adapted to support virtual server environments. You'll also learn about data backup applications that were developed specifically for virtualization, and additional methods that are available for backing up virtual machines (VMs).



TRADITIONAL DATA BACKUP METHODS

Traditional data backup methods usually operate inside the operating system where a backup device communicates over the network with a backup agent running on the OS to back up the contents of the disks on the server. This method worked well for physical servers, but virtual servers are much different than physical servers. For one thing, the entire contents of a guest operating system on a virtual machine is encapsulated into a single virtual disk file located on a host server's file system. In addition, a VM must contend for host resources with many other VMs and each VM can have an impact on the performance of other VMs running on the host.

Traditional data backup methods usually operate inside the operating system where a backup device communicates over the network with a backup agent running on the OS to back up the contents of the disks on the server.

As a result, backing up a VM using a backup agent running inside the OS is not very efficient because it increases network and disk I/O, as well as CPU utilization on the host while the backup is running. This results in fewer resources for the other VMs on that host. Additionally, if multiple backups are running on the host, the problem will be even worse and can seriously degrade the performance of the host.

DATA BACKUP AT THE VIRTUALIZATION LAYER

Backups that use OS agents on VMs must navigate through the virtualization layer to get to the guest operating system layer. A more efficient way of doing backups in virtualization is to perform the backup at the virtualization layer and never enter the guest operating system. VMware recognized this and released its VMware Consolidated Backup (VCB). VCB acted as a proxy server to offload the backups from within the virtual machine by mounting the virtual disk on the VCB server and then doing an image-level backup of it without involving the host or the virtual machine. This shifted the backup overhead from the VM and the host to the proxy server instead. While this was a step in the right direction, it required a middleman between the backup device and the target disk, adding a step to the process.

With the vSphere release, VMware eliminated VCB and the proxy that was used, and instead leveraged APIs and software development kits (SDKs) so data backup vendors could directly connect to virtual storage targets to back up VMs. The new vStorage APIs for Data Protection include the functionality that was available in VCB and also added new functionality such as Changed Block Tracking (CBT), and the ability to directly

IS YOUR BACKUP AS SUPER AS THE SYSTEM IT'S PROTECTING?

INTRODUCING Rectiphy ActiveImage Protector™.

It's a backup software, sure - but **it's not just any backup software**. Rectiphy AIP can recover your entire PC from a system crash and restore your data before you even break a sweat. **It runs a lightning-fast (yet utterly reliable) disk imaging** of online volumes, including OS and applications, **while the Windows® machine is up and running**.

- **It's tailor-made to back up virtual environments.**
- **It's the only backup and recovery software that can hot image and restore your entire Windows Hyper-V™ host virtual environment without shutting down guest virtual machines.**
- **With new ReZoom™ technology, you're able to backup and restore as many VMs as you need to, with just one AIP installation on the Hyper-V host - or even ReZoom VMs to host servers with dissimilar hardware.**
- **It's super-infused with piles of features to make your life easier.**
- **It rises to challenging times with customizable licensing plans that are intelligent, adaptable and economical.**

Advanced, agile, count-on-it backup. It's about time.
The Backup Revolution Will Be Virtualized.

Rectiphy

www.rectiphy.com • sales@rectiphy.com • 951-200-5660 • 41146 Elm St. Suite H, Murrieta, CA 92562

interact with the contents of virtual disks. Doing this was much more efficient and offered more features than using VCB to back up virtual machines.

IMAGE-LEVEL BACKUP

Backing up a virtual machine at the virtualization layer involves backing up the virtual machine disk file (VMDK) that is added to a virtual machine. This is referred to as an image-level backup. This differs from traditional data backups that are done inside the OS where files are backed up individually (file-level backup). While it may seem more efficient backing up just one large file instead of thousands of smaller files, this is not the case.

Backing up a virtual machine at the virtualization layer involves backing up the virtual machine disk file (VMDK) that is added to a virtual machine.

The reason is that image-level backups cannot see inside the operating system and are backing up the whole virtual disk. They're also backing up empty disk blocks and deleted files. If a virtual machine has a 50 GB virtual disk file and only 10 GB is in use, 50 GB is backed up with an image-level backup. With file-level backups, only the 10 GB in use is backed up. To get around this, backup vendors have gotten creative and rely on technologies like data deduplication, synthetic backups and empty block recognition that ignore empty and duplicate blocks as well as blocks that are no longer active when files are deleted.

You might wonder that if an image-level backup cannot see inside a guest operating system, how can it handle open files and avoid corruption from files that change when the backup occurs. This is done by first quiescing the VM using a special driver (either VMware Tools or a backup vendor-supplied driver) that runs inside the guest operating system that momentarily pauses the running processes on a guest and forces the operating systems and applications to write any pending data to disk. Once that is complete, a snapshot of the VM is taken at the virtualization layer that creates a new temporary virtual disk file (delta) for any new disk writes that occur on the VM, which prevents the original disk from being written to while the backup is running. Once the backup is completed, the temporary virtual disk file is merged back into the original disk file and the snapshot is deleted.

Many data backup application vendors that back up at the virtualization layer use deduplication to detect duplicate blocks and ignore them. They also detect empty disk blocks that have not been written to the operating system yet and ignore those as well. Quest Software Inc. (formerly Vizioncore) has taken it a step further and uses a technology called

Active Block Mapping (ABM) to recognize disk blocks that once contained data but no longer do because files were deleted. Normally when a file is deleted within an operating system, only the pointer to the file is removed while the data still resides on the hard disk.

IMAGE-LEVEL BACKUPS AND FILE-LEVEL RESTORES

A common misconception with image-level backups is that you cannot do incremental backups because you're only backing up one large file, and if any disk block changes the whole file must be backed up again. With traditional file-level backups, only the files that have changed are backed up on incremental backups. This is noted by setting a flag called an archive bit that indicates a file has changed since the last backup. Once the file is backed up, the archive bit is cleared until it changes again. With an image-level backup, a backup application has to keep track of all the blocks that have changed since the last backup so it knows which ones to back up when doing incremental backups. This process can increase the time of backups because the backup application must calculate a hash for each block, scan the entire virtual disk and compare it against a hash table to see what has changed since the last backup. To speed up incremental backups, most backup vendors have taken advantage of the new Changed Block Tracking feature accessible via the VMware vStorage API for Data Protection. This allows the backup application to simply query the VMkernel to find out which disk blocks have changed since the last backup, and this greatly speeds up incremental backups.

With an image-level backup, a backup application has to keep track of all the blocks that have changed since the last backup so it knows which ones to back up when doing incremental backups.

Another common misconception when doing image-level backups is that you cannot do individual file restores. This is possible, but doing image-level backups has one drawback: Since you are only backing up the large virtual disk file, it changes the way individual file restores are done. Traditional file-level backups simply create a catalog of all the files as they are backed up, and then that's used to restore individual files.

Image-level backups also have this capability because they simply mount the virtual disk file that is backed up, and look inside the guest operating system to see the file layout. As a result, individual file restores with image-level backups are possible. With file-level backups, individual file restores are simple. You choose the file to restore from the backup media, and the backup server connects to the agent as the target server locates the file and copies it back to the original source. With image

backups, because there is typically no agent on the target server, it's slightly different. What happens is the virtual machine disk file from the backup media is mounted by a restore application that allows the file to be copied from it to either a local disk or back to the original server. Then, once the file is copied the virtual disk is un-mounted. The process for individual file restores is different but the end result is the same.

While an image-level backup may change the way you do file-level restores, it has the advantage of making a bare-metal restore of a VM a simple process. Since the VM is encapsulated into one big file, all you have to do is copy that file back to a virtual host and you have a complete copy of the server from the point in time of the backup. Another big advantage of this is that virtual machines all have the same type of virtual hardware regardless of the underlying physical host hardware. This eliminates any hardware incompatibilities that may occur when performing a bare-metal restore to a different host. With traditional backups, if you do a bare-metal restore to a different server you have to do a lot of pre- and post-restore steps to make sure hardware drivers, disk partitions and system configurations are all correct for the new hardware.

Image-level backups offer some other advantages over traditional file-level backups of physical servers. Having the server encapsulated into one big file makes for easy portability; the virtual disk file can easily be copied to any other storage device. For example, one could easily copy a VM from a host server to another storage device, external hard drive or flash drive for safekeeping. This makes creating ad hoc backups of virtual machines a simple process.

TESTING DATA BACKUP RESTORES IN VIRTUAL SERVER ENVIRONMENTS

While having good data backups is very important, having good data backup restores is even more important. Testing restores with traditional backups of physical servers can be difficult, time-consuming, disruptive and may require extra server hardware. Virtualization can make this a much simpler process because virtual machines can be restored and isolated on hosts without overwriting and affecting the original virtual machines. This makes testing a restoration of individual files or whole virtual machines an easy process, so you can verify that your backups are working properly.

Veeam Software has taken this even further by introducing a new feature called SureBackup. SureBackup automates the verification of virtual machines in a separate environment on a host so data and applications can be verified that they'll function when restored. Normally, to do this you have to copy the virtual disk files to a host so you can power on the VM and test it. Veeam figured out a way to avoid these extra steps by being able to run a VM directly from the target backup store

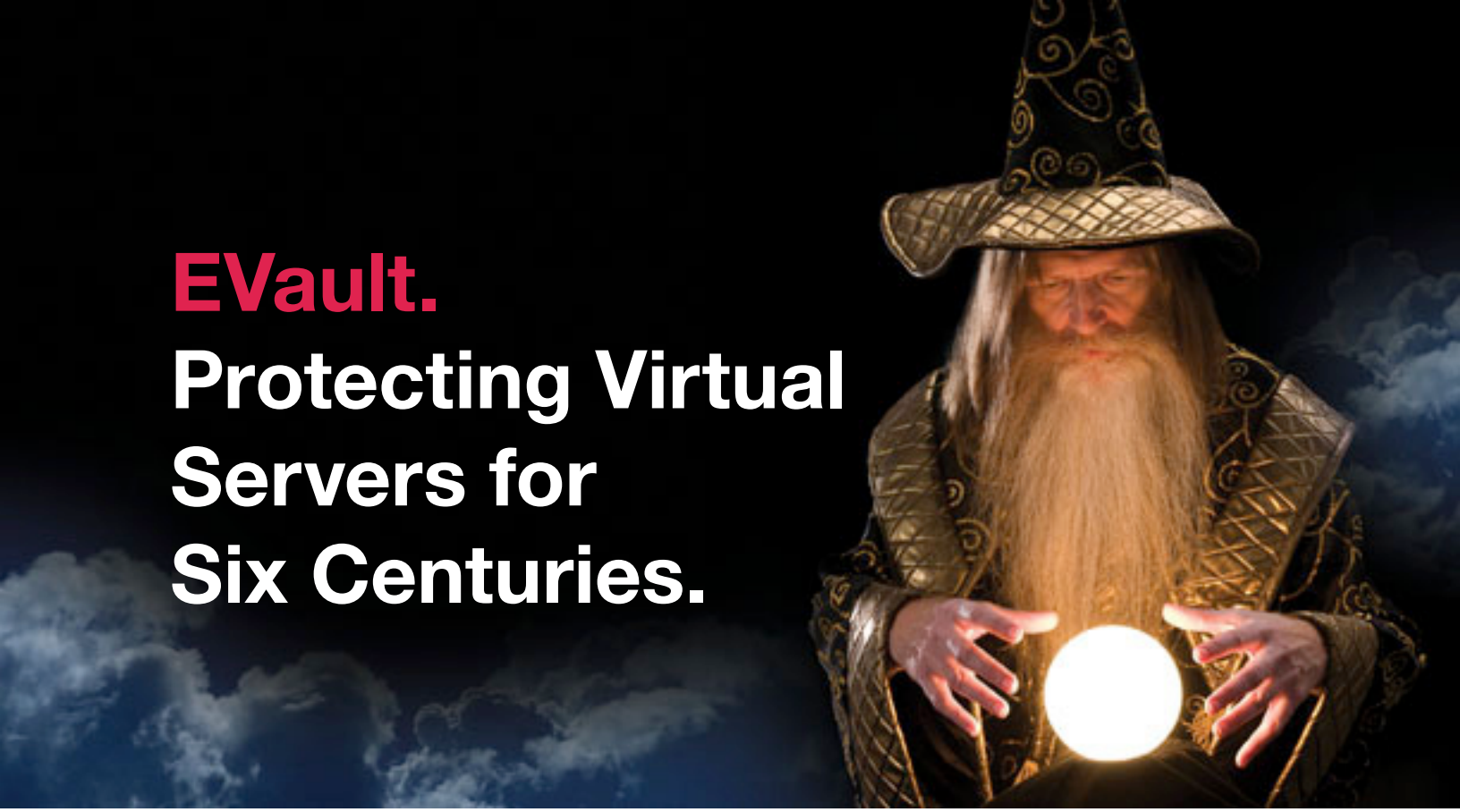
without having to extract it. SureBackup publishes the contents of a backup file as a datastore that a virtual host can connect to. Virtual machines are automatically created from the datastore in an isolated environment where they can be powered on and tested to ensure that applications are functioning properly and data is intact. Using this method reduces the host resources that are required and doesn't require the extra storage that you would normally need to copy a virtual disk back to a host server. This capability automates and simplifies the verification of backed up VMs and also makes application-item level restorations possible.

Virtual machines are automatically created from the datastore in an isolated environment where they can be powered on and tested to ensure that applications are functioning properly and data is intact.

While you can still use file-level agent backups running inside the VM, it's not efficient and you should consider changing your backup method to one that is optimized for virtualization. Traditional backup vendors like EMC, IBM and Symantec all have adapted their products to better integrate with virtualization.

In addition, there are several vendors that have developed data backup applications specifically for virtual environments including Veeam Backup & Replication, PHD Virtual Backup for VMware ESX (formerly esXpress) and Quest Software vRanger Pro Data Protection Platform (DPP). These vendors recognized the need for better backup solutions for virtual environments and developed products that are optimized for virtualization. When backing up your virtual environment, you should avoid using traditional methods and applications that are not aware of the virtualization layer. Instead, leverage ones designed specifically for virtual server environments so you can achieve maximum efficiency and flexibility with your backups and restorations. ⊙

Eric Siebert is an IT industry veteran with more than 25 years experience covering many different areas but focusing on server administration and virtualization.



EVault.
Protecting Virtual Servers for Six Centuries.

Actually, since 1997. But in the IT realm, 13 years is a very, very long time.*

Today more than 30,000 small and mid-size businesses rely on EVault® on-premise, cloud-based, edge, and hybrid products and services for professional-grade data protection and anywhere, anytime access—seamlessly across all their physical and virtual machines.

Hot backups. WAN-optimized performance. End-to-end encrypted security. Assured recovery from site disasters. Our integrated data protection solutions for mixed IT environments will exceed your organization's wildest dreams. Now and well into the future.

*One IT year = (1100 pots of coffee x 27 weekends worked in row) ÷ (99.999% uptime + you).
Give or take a few weekends.

i365.com | 1.877.901.DATA

i365
A Seagate Company

To learn more about i365 and EVault storage solutions for virtualization, please visit us at www.i365.com.

New features in VMware vSphere that benefit data backup and recovery

Here's a look at the changes in the most recent version of VMware vSphere that will impact your backup and recovery environments. By Eric Siebert

When VMware Inc. released vSphere, the successor to VI3, many new enhancements to existing features were included that are very beneficial to the backup and recovery of virtual machines (VMs). In this article, I'll discuss what's new and changed and how those features make for better backup and recovery of virtual machines in vSphere.

vSTORAGE APIs

Perhaps the biggest benefit to backups and storage in vSphere is the new set of vStorage APIs that VMware developed. These APIs allow third-party applications to directly interface with the VMkernel without the need for scripts or agents. The vStorage APIs existed in VI3, but were referred to as the VMware Consolidated Backup (VCB) Backup Framework. However, unlike VMware Consolidated Backup, they are not a separate standalone application and built directly into ESX(i), and require no additional software installation. While the VCB Backup Framework still exists in vSphere and can also be used by backup applications, the vStorage APIs are the successor to VCB and will eventually completely replace it. The vStorage APIs are broken into four groups that have different types of functionality. They include:

vStorage APIs for Array Integration: Currently being developed with specific third-party storage vendors (i.e., EMC Corp., Hewlett-Packard [HP] Co. and NetApp), these APIs will allow vendors to leverage their storage array-based capabilities directly from vSphere. This includes things such as array-based snapshots, hardware offloaded storage device locking, integration between VMware and array-level thin provisioning, storage provisioning, replication, and more. This will enable vSphere to act more efficiently for some storage-related operations by allowing the storage array to perform certain operations.

vStorage APIs for Multipathing: These APIs enable third-party storage vendors to leverage array multipathing functionality through plug-ins that they can develop. These plug-ins allow for more intelligent storage multipathing to achieve better storage I/O throughput and storage path failover for a specific vendor storage array.

vStorage APIs for Site Recovery Manager: These APIs are part of VMware's Site Recovery Manager (SRM) and are used to integrate SRM with array-based replication for block and network-attached storage (NAS) models. This allows SRM to seamlessly handle both virtual machine, host failover and storage replication failover, and also enables SRM to control the underlying array-based replication that it relies on.

vStorage APIs for Data Protection: These APIs are the ones that are very important to third-party backup and replication vendors as they enable better and more seamless integration to virtual machines disks. While designed to be the successor to the VCB, they include the functionality that was available in VCB and also added new functionality such as Changed Block Tracking (CBT) and the ability to directly interact with the contents of Virtual Disks via the VDDK.

The vStorage APIs are not really a single API and the term is basically just a name for a collection of interfaces that can be used by third-party applications to interact with storage devices in vSphere. These interfaces consist of various SDKs that exist in vSphere and also their Virtual Disk Development Kit (VDDK). The VDDK is a combination API and SDK that enables vendors to develop applications that create and access virtual disk storage. The VDDK is used in conjunction with other vStorage APIs to offer a complete integrated solution for management of storage in vSphere. For example, while VM snapshots can be managed using the SDK functionality, other operations like mounting virtual disks are handled through the VDDK.

CHANGED BLOCK TRACKING

The vStorage APIs for Data Protection are most beneficial to backup and replication applications and vendors seem to be most excited about the new Changed Block Tracking feature that is included in it. This feature allows third-party applications to query the VMkernel to

find out which disk blocks have changed in a virtual machines disk file since the last backup operation. Without this feature, applications would have to figure this out on their own which can be time-consuming. Now with CBT they can instantly find this out so they know exactly which disk blocks need to be backed up. This enables much faster incremental backups and also allows for near continuous data protection (CDP) when replicating virtual disk files to other locations. In addition, point-in-time restore operations are much quicker as CBT can tell exactly which disk blocks need to be restored to the virtual machine.

Changed Block Tracking is supported on any storage device and datastore in vSphere except for physical mode Raw Device Mappings; this includes iSCSI, VMFS, NFS and local disks.

Changed Block Tracking is supported on any storage device and datastore in vSphere except for physical mode Raw Device Mappings; this includes iSCSI, VMFS, NFS and local disks. It also works with both thin and thick disk types. CBT is a new feature to vSphere, so it does require that the virtual machine hardware be version 7, which is the default in vSphere. By default, CBT is disabled as there is a very slight performance penalty that occurs when using it. It can be enabled on select VMs by adding parameters (ctkEnabled=true and scsi#:#.ctkEnabled=true) to the configuration file of the virtual machine, backups applications can also enable it using the SDKs. Once enabled, a VM must go through something called a stun/unstun cycle for it to take effect; this cycle happens during certain VM operations including power on/off, suspend/resume and create/delete snapshot. During this cycle, a VM's disk is re-opened, which allows a change tracking filter to be inserted into the storage stack for that VM.

The Changed Block Tracking feature stores information about changed blocks in a special "-ctk.vmdk" file that is created in each VM's home directory. This file is fixed length and does not grow and the size will vary based on the size of a virtual disk (approximately .5 MB per 10 GB of virtual disk size). Inside this file the state of each block is stored for tracking purposes using sequence numbers that can tell applications if a block has changed or not. One of these files will exist for each virtual disk that CBT is enabled on.

The vStorage APIs for Data Protection and the CBT feature make backups quicker and easier in vSphere and are a big improvement over VCB. VMware has provided third-party vendors with a much improved backup interface in vSphere, now it's up to them to adapt their products to take advantage of them.

THIN PROVISIONING AND BACKUPS

Thin provisioned disks are virtual disks that start small and grow as data is written to them. Unlike thick disks where are all space is allocated at the time of disk creation, when a thin disk is created its initial size is 1 MB, (or up to 8 MB depending on the default block size) and it then grows up to the maximum size that was defined when it was created as data is written to it by the guest OS. The benefit of thin provisioned disks is that they allow for the over-allocation of storage on a VMFS volume to make use of the often wasted unused space inside of a VM's disk. Thin provisioned disks are not new to vSphere and also existed in VI3, however, there were numerous changes to make them more usable in vSphere.

Thin provisioned disks are virtual disks that start small and grow as data is written to them.

Why are thin disks important to backups? Many backup applications for virtualization do not operate inside the guest operating system but rather they operate outside of it at the virtualization layer. Instead of backing up individual files inside the guest OS, they back up the single large virtual disk files (VMDK) that contain the encapsulated VM. Because of this, backup applications must search for empty disk blocks contained inside the virtual disk file so they do not back them up. This process of identifying empty blocks takes additional time and resources to complete. With thick disks all space is allocated at once, so a 40 GB virtual disk will actually take up 40 GB of disk space on a datastore regardless of how much space is used by the guest OS running on it. So, if only 10 GB of disk space is in actual use by the guest OS you will want to avoid backing up the extra 30 GB of empty space inside the virtual disk file.

Thin disks only take up as much space on a datastore as what is actually used by the guest OS, so if only 10 GB of a 40 GB virtual disk is in use the virtual disk file will only be 10 GB in size. Because of this, backup applications no longer have to worry about searching for those empty disk blocks because there are none in a thin disk. Not having to do this results in faster and more efficient backups which is just one of the advantages of using thin disks.

HOT-ADD OF VIRTUAL DISKS

The hot-add of virtual disks feature allows a virtual machine to mount the disk of another virtual machine while it is running so it can be backed up. This is similar to what was first introduced in VCB where a virtual disk can be mounted by another server to be backed up. The hot-add feature in vSphere allows one virtual machine running a backup application to mount the disk of another so it can read the data from it and write it to destination media. Doing this removes the backup traffic from the network as the VM running the backup application uses the VDDK to access the disk

and all I/O requests to it are sent directly down the VMkernel I/O path.

The hot-add feature works by taking a snapshot of the virtual disk that deflects writes to a separate delta file. Once this is complete, the now read-only disk can be mounted by another VM so the data can be copied from it. Hot-add takes advantage of the SCSI specification that allows for SCSI devices to be added/removed from a server without powering it down. It works with disks on any type of storage supported by vSphere as long as the VM running the backup application is on a host that can access the storage of the target VM (i.e., shared storage). However, it does not work with VMs that have IDE virtual disks that are now supported in vSphere.

Several data backup applications have already taken advantage of the hot-add feature including VMware Data Recovery and Veeam Backup & Replication. The use of the hot-add feature is not available in all editions of vSphere and requires the more costly Advanced, Enterprise and Enterprise Plus editions.

iSCSI IMPROVEMENTS

VMware made significant improvements to the iSCSI storage protocol in vSphere that resulted in increased performance and greater efficiency of virtual machines on iSCSI datastores. This is also beneficial to backup applications as the increased efficiencies with the iSCSI protocol are a direct benefit to heavy disk I/O operations that occur during virtual machine backups. The improvements to iSCSI in vSphere included the following:

- In vSphere, VMware made significant updates in iSCSI for both software and hardware initiators. The software initiator that is built into ESX was completely rewritten, tuned and optimized for virtualization I/O. The result of these efforts includes a marked improvement in performance as well as greater CPU efficiency which resulted in a significant CPU usage reduction when using software initiators.

VMware made significant improvements to the iSCSI storage protocol in vSphere that resulted in increased performance and greater efficiency of virtual machines on iSCSI datastores.

- Support for Jumbo Frames was introduced in VI3.5, but was not officially supported for use with storage protocols. With vSphere, VMware officially supports the use of Jumbo Frames with the iSCSI and NFS storage protocols. In addition, they now support 10 Gb Ethernet with iSCSI that results in much greater I/O throughput.
- Easier provisioning of iSCSI storage due to the iSCSI stack no longer requiring a Service Console connection to communicate with an iSCSI target. Configuration steps for iSCSI have been made easier and

global configuration settings will now propagate down to all targets. Additionally bi-directional CHAP authentication is now supported for increased security.

These improvements make the use of iSCSI a more attractive choice over the more expensive Fibre Channel storage area network (SAN) for either virtual machine datastores or backup targets.

VMWARE DATA RECOVERY

VMware also introduced VMware Data Recovery (VDR) in vSphere. VDR is a disk-to-disk backup application that provides basic backup capabilities natively in vSphere. VMware Data Recovery provides an alternative method for backing up virtual machines instead of the traditional OS agent methods that are used in physical environments. While not as feature rich as some of the other third-party backup applications, it does provide some advanced features such as inline data deduplication and compression, and a centralized management console that is integrated into the vSphere Client. In addition, VDR takes full advantage of the new features in vSphere such as Changed Block Tracking and hot-add of disk to ensure more efficient and faster backups. VDR is available as part of the Essentials Plus, Advanced, Enterprise and Enterprise Plus editions, or can be purchased a la carte with the Standard edition.

All the new or improved features covered here make upgrading to vSphere very compelling, especially for much improved backup and recovery. The vStorage APIs provide much better integration for third-party backup applications and enable vendors to develop more efficient products to safeguard virtual machine data. If you have been putting off upgrading to vSphere, these new data backup-related features, along with the many other new or improved features in other areas, may persuade you to upgrade. ⊙

Eric Siebert is an IT industry veteran with more than 25 years experience covering many different areas but focusing on server administration and virtualization.

Data dedupe technology helps curb virtual server sprawl

Learn about how data deduplication technology can help ease the burden of virtual machine data backup and recovery. *By Christine Cignoli*

VIRTUAL SERVERS and data deduplication technology have both kept the data storage industry buzzing in the past few years. But how virtualization and dedupe technology work together is something vendors and users are still fine-tuning.

"We saw the tipping point last year when the number of virtual servers exceeded the number of physical servers," said Steve Scully, research manager, continuity, disaster recovery and storage orchestration at IDC. "The biggest challenge is around virtual machine backup."

"Virtualization has caused server sprawl," said Eric Pitcher, VP of technology strategy at CA Technologies. "People say virtual machines get thrown away, but the reality is it doesn't happen. Typically you just keep creating them." Data deduping is a way to battle virtual server proliferation, he said.

With a traditional, non-virtualized backup scheme, said Scully, a company buys a license for each server, runs the backup app on each server, backs up all files and sends them to disk or tape. But when it comes to virtual servers, "if you do that times 50 or 100, you're paying a lot for those licenses and not getting the potential advantage of dedupe technologies," said Scully. "It's identical processes running without any knowledge of what the other guy is running." Virtual machines are often backed up as complete images as opposed to a set of individual files. Some backup apps can do dedupe across multiple VM images, said Scully. But "you don't get the granularity of file-level backup," he said. "You have to recover the entire virtual machine."

VIRTUAL SERVER BACKUP MORE COMPLICATED THAN TRADITIONAL BACKUP AND RECOVERY

A common challenge of virtualized servers is that all machines are sharing physical CPU, bandwidth and disk, according to Rob Emsley, senior director of product marketing, EMC backup recovery systems division. “You have to make the physical resources more efficient, which becomes a challenge for doing traditional backup and recovery,” he said.

Backing up virtual servers is more complicated than other backups, said Pitcher. “You take a snapshot of the server, move it to a temporary location and do backup from that location,” he said. CA’s strategy for improving and deduping virtual backups included cutting out the temporary storage location to back up the virtual space directly from the virtual machine.

“Server virtualization and dedupe is the same concept: consolidation, optimizing storage, reducing power and cooling and retaining data for longer periods of time,” said Mike DiMeglio, product marketing manager at FalconStor, which offers data dedupe technology with its DiskSafe and FileSafe product features. They currently use a proxy server to run source deduplication for virtual machines, supporting various backup software options, but DiMeglio said snapshots are a big part of FalconStor’s roadmap for virtual machines and data dedupe. “Then you can back up from a snapshot, and dedupe applies to that,” he said.

"Server virtualization and dedupe is the same concept: consolidation, optimizing storage, reducing power and cooling, and retaining data for longer periods of time."

—MIKE DiMEGLIO, product marketing manager, FalconStor

EMC’s Avamar software dedupes at the source and from within the virtual environment with tight VMware integration, said Emsley. When deduping virtual machines with an appliance, using target dedupe, like EMC’s Data Domain product, virtual servers are seen as just another workload, said Shane Jackson, senior director of product marketing, Data Domain and Disk Library at EMC. Dedupe rates can be very high for virtual machines because the level of redundancy is so high, said Jackson.

TARGET DEDUPLICATION vs. SOURCE DEDUPE TECHNOLOGY

There are advantages to both source deduplication and target dedupe technology with virtual machines, said IDC’s Scully. One thing to consider is whether the backup application is doing the incremental backups on full backups of individual VMs. “I guarantee the image is going to look different every time,” said Scully. “If some file on that entire image has changed, that entire image as a set of files is going to be

different. So it isn't seen as an incremental because the whole file has changed." In that case, it might make sense to dedupe that entire image at the source. But others might want to get data out of the production environment without an extra load on the servers or analysis on the source side. In a true disaster recovery situation, image-level backups of VMs can be "very powerful" for getting systems up and running, said Scully.

Deduping backup data at the source can transfer data off VMs quickly, said Mathew Lodge, senior director of product marketing at Symantec. "It's important to move data off virtual servers," said Lodge, as virtual servers can consume a lot of CPU. Symantec's recently released new versions of NetBackup and Backup Exec do granular recovery of virtual machines, and allow for dedupe in several locations throughout the backup process, including at the source. Symantec recommends dedupe within each virtual machine if there are bandwidth constraints or in a data center using Microsoft Hyper-V. Otherwise, VMware users should use the vStorage APIs for Data Protection to send entire VMware images to a NetBackup media server for deduplication, said Lodge.

> "It's important to move data off virtual servers."
>
> —MATHEW LODGE, senior director of product marketing, Symantec

OTHER DEDUPE TECHNOLOGY OPTIONS FOR VIRTUAL SERVERS

Other creative options for deduping virtual servers are out there. At BlueLock LLC, a cloud computing provider, they've approached deduplicating virtual server data from a different angle, said Pat O'Day, chief technology officer. BlueLock uses VMware-linked clones to reduce duplicate data. They create a server in VMware as a template, put it in the cloud and let users provision servers from that template. When the user renames that server, just the one block changes.

"The linked clone only tracks the differences between the machine the user spun up and the original template," said O'Day. "It's essentially dedupe." The downside is that in the long-term, "it never reconciles changes like a dedupe solution would," he said. Though O'Day said BlueLock is looking at dedupe options, they're hoping to incorporate both technologies. "I don't think linked clones are going to go away in favor of deduplication."

As virtual server use has moved beyond testing and development, the related backup and recovery continues to mature. "There's a lot to get done," said EMC's Jackson. "Even now we're pushing to a greater degree of server virtualization in the data center, getting to 80% virtualization." Data dedupe is letting companies get there, he said, as part of a greater "backup redesign."

When choosing how to dedupe virtual server data, “it really comes down to understanding what your needs are, what you want to recover at the file level and what you want to recover at the image level,” said IDC’s Scully. “There are various knobs and dials you can tune of what you want to recover and what level of granularity.” ⊙

Christine Cignoli is a Boston-based technology writer.

MICROSOFT HYPER-V BACKUP FOR VIRTUALIZED SERVERS

In this Q & A with W. Curtis Preston, learn about the pros and cons of backing up Microsoft Hyper-V for virtual servers. By W. Curtis Preston

MANY COMPANIES incorporate Microsoft Hyper-V virtualized servers into their data backup environments. But backing up Microsoft Hyper-V virtualized servers involves different steps than with more familiar VMware servers. Hyper-V can be backed up with either agent or host-based approaches, and there are specific things you need to enable in Hyper-V, such as Volume Shadow Copy Service (VSS) and Hyper-V integration services to ensure good performance and effectiveness of the data backups.

W. Curtis Preston, executive editor at TechTarget and independent backup expert, discusses the challenges of Microsoft Hyper-V back-up for virtualized servers, the differences between backing up Hyper-V and VMware servers, and more in this Q&A.

What are the challenges associated with backing up Hyper-V virtualized servers?
I'd say the biggest challenge is the predominance of VMware. Generally you run into a blank stare when you ask someone from your backup software company about Microsoft Hyper-V. Hyper-V is certainly increasing its market share, and data backup and recovery is at least a small reason as to why that's the case. The other challenges associated with Hyper-V are similar to the ones you run into with VMware, which are basically the laws of physics. For example, you've taken 20 physical servers and you've put them

inside one physical server. So you have all of this data that needs to be moved around for the purposes of data backup and recovery, and it can only go through one physical machine. That's pretty much the main challenge that you have to deal with.

What are the different approaches for backing up Hyper-V? For example, what's the difference between host-based and agent-based?

Agent-based is the most common method for backing up virtual servers. Basically, you put an agent in each of the virtual machines (VMs), and then you kind of put your head in the sand and pretend that everything's physical. And that gives you a lot of benefits where you're able to back up a lot of databases and things using the same agents that you're used to. So the biggest advantage to that approach is familiarity.

The other approach is host-based, which is looking at the actual Hyper-V server. In this case, that's just a Windows server. It's not quite the same as VMware where there's this whole other world and you don't have a host to connect to. With Hyper-V, it's a Windows host and you can talk to it like any other Windows host, but if you want to connect and back up those virtual machines outside of the VM world, then you need to interface with their infrastructure, mainly VSS, so that you can get good backups of the VMs. Any data backup product that interfaces with Hyper-V should be able to do that.

Agent-based is the most common method for backing up virtual servers.

Are there specific things you need to enable in Hyper-V to ensure data backup performance/effectiveness? Can you discuss VSS and Hyper-V integration services?

Volume Shadow Copy Service is the biggest thing that needs to be enabled because it is the overall infrastructure that allows a backup app to quiesce an application such as Exchange or SQL Server, and the file system upon which that application is storing its data. And so, when a backup app wants to back up a virtual machine that's inside Microsoft Hyper-V, it needs to connect to the Hyper-V integration services, which talk to VSS. VSS then talks to the VSS writers. Each application has its own VSS writer, and once all the writers have communicated that they've done their job, then VSS creates a snapshot or a shadow copy of the volumes that are part of that VM. At this point, VSS turns around and tells the backup app that it can back up that snapshot. Once that backup has been completed, the snapshot can be released, and if the backup app used the right backup type, then the applications will be notified via their writers that a backup was just taken, and that they can do the right thing after the backup, which typically means they're going to truncate their transaction log so that you have a nice clean system after the backup.

How does backing up Hyper-V differ from backing up VMware servers?

Everything I just described about VSS needs to happen in VMware. Unfortunately it does not. VMware only talks to the application VSS writers in Windows 2003. And remember how I said that if they use the proper backup type, the applications will do things like truncate their logs? Well, VMware chose not to use the VSS backup type. It uses VSS copy. So basically it's telling VSS and the applications that it's creating just a snapshot of the data. It's not creating a backup, but rather a copy. And because it's telling VSS that it's just creating a copy, VSS does not truncate the logs, because you wouldn't do that if you're just making a copy of the data, you would only do that if you're making a backup. So the concept is the same, but it's not as full of an implementation as what Microsoft Hyper-V does.

So overall, VMware only talks to the applications in Windows 2003; it does not talk to them at all in Windows 2008, which means you're creating a crash consistent copy. With a crash consistent copy, when you restore the VM from that backup, the application is going to have to go through a crash recovery process in order to bring the app to a consistent point. That's why it's called a crash consistent—it's as consistent as a crash, which frankly is a phrase that scares me. So if you're not running Windows 2003, you're only getting a crash consistent copy of the apps. Even if you're running Windows 2003, you're not doing the step of truncating the logs. So overall, Hyper-V's implementation of VSS is much more advanced than in VMware.

With a crash consistent copy, when you restore the VM from that backup, the application is going to have to go through a crash recovery process in order to bring the app to a consistent point.

What are the benefits or drawbacks of using Microsoft System Center Data Protection Manager (DPM) to back up Hyper-V virtualized servers instead of another backup software product?

Microsoft Data Protection Manager is a near-CDP product. Near-CDP is close to continuous data protection (CDP), where it's not truly continuous, it's something like once an hour. First off, DPM fully integrates everything I described about VSS implementation—for obvious reasons—it's put out by Microsoft. That's the first advantage. You can be assured of complete integration and that they're going to talk to all the VSS writers.

The second thing is that when you compare it to traditional third-party backup apps, those are traditional full and incremental backup apps. Even when we look at IBM Corp. Tivoli Storage Manager (TSM), which offers a progressive incremental feature, it does that only for

file systems. When we look at doing data backups of databases and applications inside VMs, VMware also does full and incremental backups. So if you compare Data Protection Manager to the typical backup app, the typical backup app is going to create full and incremental backups. The only thing DPM does each time you create a snapshot is that it transfers the bytes that changed from the snapshot that was taken previously. And so it's a very incremental-forever block-level technology. So there are two things to remember about using DPM with Hyper-V. First, it should have very little impact on the performance of Hyper-V, and second, it should have very tight integration with Windows because it's made by Microsoft.

What are the challenges with using another backup tool?

The biggest thing to remember when using a third-party data backup tool is to make sure that it fully integrates with VSS. Also, make sure that it's talking to all the VSS writers. Everyone's going to want to talk to Exchange and SQL Server, but what about Active Directory, and Oracle? There are several other smaller VSS writers that aren't as popular. Are there other applications that are present in the virtual machines, and do they have a VSS writer? And is my backup app talking to all the VSS writers? The proper way for a backup to behave is to do a metadata query of VSS. VSS should give you a list of all the VSS writers that it has. It should then talk to each of the VSS writers that it discovers in the process. And so hopefully it's able to discover all those, and hopefully it's able to talk to them so each of the VSS writers can do the right thing. These are all things that you simply need to verify with documentation and testing. ⊙

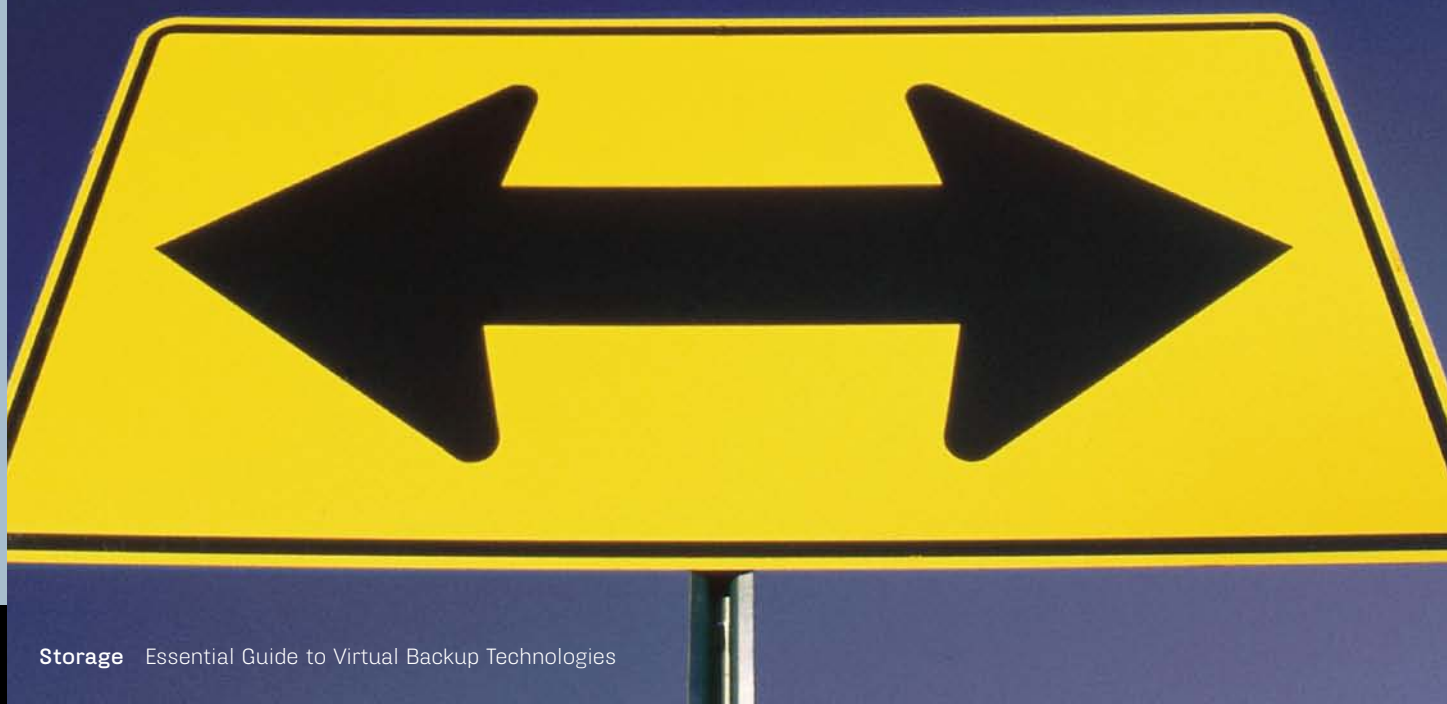
W. Curtis Preston is an executive editor at TechTarget and an independent backup expert.

VM backup: VM-specific software vs. traditional enterprise data backup software

In this storage channel tip, learn whether or not you should provide the module for VMware that your enterprise data backup partner provides, or if you should you sell a VM-specific software application.

By George Crump

Sometimes storage integrators find themselves between a rock and a hard place. In the case of VMware, that often happens when the discussion turns toward virtual machine (VM) backup. Should you provide the module for VMware that your enterprise data backup partner provides, or should you sell a VM-specific software application?



Most storage integrators have a strong relationship with a particular enterprise data backup software vendor. It's the one they recommended 99.9% of the time. This relationship may have developed because at one point the organization felt it was the best available application or it was the one that the engineers knew the most about. Over the past five years or so, these applications have become more similar than they are different, and because of the sheer size of the code, vendors have struggled to keep them in sync with the very latest initiatives in the data center.

In the server virtualization space, VMware was slow to develop an acceptable API set for data backup software vendors to tap into to support applications such as Exchange, Oracle and SQL Server (though the company has made significant changes to that API set in its latest release). The major data backup software vendors, in turn, were slow to add support for VMware.

With a lack of good data backup tools for a VM environment, companies like Quest Software (formerly Vizioncore), Veeam and PHD Virtual were born and developed, respectively, vRanger Pro, Veeam Backup & Replication and Backup for VMware ESX (formerly esXpress). These tools are licensed per socket rather than per server (though experts and users caution that it doesn't always equate to lower costs), and they enable recovery of the virtual machine disk (VMDK) image for greatly simplified disaster recovery (DR) preparedness, as well as recovery of individual files within the VMDKs. (Traditional backup tools operate within the guest OS, so early on, they were adept at file-level restore but required multiple steps to restore entire VMDKs.)

With these tools in the mix for VM backup, life was pretty easy for storage resellers addressing the issue with their customers. They could push back on their enterprise vendors because their VMware support was either weak or nonexistent. Providing a VMware-specific backup solution to customers caused little friction with the big data backup software vendors.

Providing a VMware-specific backup solution to customers caused little friction with the big data backup software vendors.

Now, though, the data backup software vendors have either closed the gap or are starting to, leaving resellers with a more difficult decision to make: provide the probably still feature-advanced VM-specific data backup product, or provide their data backup software vendor's "pretty close" VM module.

FACTORS IN A VM BACKUP TOOL DECISION

When confronting the question of whether to recommend VM-specific tools or the enterprise data backup tools, there are a few points to consider. The first point relates to the effect on the environment. Often,

VMware-specific backup applications are significantly easier to get from installation to point of first protection. Enterprise data backup applications that are protecting VMware as a module, even assuming that all you are adding is the module, are typically more complex to integrate. The point from install to first backup is almost always faster with an application-specific solution. The downside to a VMware-specific solution, of course, is that it adds yet another wrinkle in the overall data protection process that needs to be accounted for and managed.

The downside to a VMware-specific solution is that it adds yet another wrinkle in the overall data protection process that needs to be accounted for and managed.

In making the decision, you need to consider whether the customer has a significant investment in its current data backup software and whether that application's VMware module provides adequate protection for the environment. If either is lacking, it may make sense for you to propose a VMware-specific data backup solution to manage both protection strategies within a single process, via a management tool like those available from Bocada, SolarWinds (Tek-Tools) or Aptare.

Another alternative comes into play when virtualization is the primary infrastructure, meaning that the customer is reaching 100% virtualization. Ironically, we are seeing this not only in large businesses but also in small businesses, where because of the small number of servers, total virtualization can be accomplished quickly. In this environment, a VMware-specific backup tool may be the only solution needed, eliminating the need for additional enterprise backup products.

The second point to consider is whether there is a significant capability offered by either the VMware-specific tools or the enterprise data backup software that you or your customer thinks is a "must have." It's important to remember that point solutions typically end up in a game of leapfrog, delivering very advanced features in a new space, which are eventually matched by the more comprehensive systems. The VM backup technology area is no different. The leading edge of functionality will change quickly. As an integrator, you should interview and keep a scorecard of which vendors can provide which functionality and then weigh that against what your customers' needs. ⊙

George Crump is president and founder of Storage Switzerland, an IT analyst firm focused on the storage and virtualization segments.

Check out the following resources from our sponsors:

EMC² where information lives®

How VMware Improved Backup Performance Using EMC
Fast, Efficient Backup and Deduplication for VMware Environments
Better Backup in 4 Real World VMware Environments

i365 A Seagate Company

Developing a Data Protection Strategy for Virtual Environments
Gateway Funding Trades Tape for Fast EVault Backups and Restores
VMware Solution Brief

Rectiphy

What's Rectiphy ActiveImage Protector? It's your new hero. Read the manifesto.
ActiveImage Protector for Hyper-V makes VM backups easy: ReZoom your VMs with no additional hassles or costs.
All the features. Sixty days. Thirty seconds of your time. Sound good?

VEEAM #1 for Virtualization™ Data Protection and Management

Deduplication and VMware Backup Sprawl
5 Ways VMware vSphere Improves Backup and Recovery
3 Options for VMware Disaster Recovery and CDP