

5 Ways VMware vSphere Improves Backup and Recovery

Sponsored by Veeam Software



Written by:

Eric Siebert

VMware vExpert and
Senior Systems Administrator

CONTENTS

Overview 2

Five Significant Improvements 3

Four Enabling Technologies 3

Thin Provisioning 4

vStorage APIs 5

iSCSI Software Initiator 8

Hot-add of Virtual Disk Files 9

Summary 9

About the Author 10

About Veeam Backup & Replication 11

OVERVIEW

VMware's latest enterprise-level hypervisor product suite, vSphere, includes many new features and technological enhancements over its predecessor, VMware Infrastructure 3 (VI3). While there are several reasons to upgrade to vSphere, this white paper focuses on how VMware improved various storage technologies, how these relate to data backup and recovery, and how the new vStorage application programming interfaces (APIs) allow for better storage integration with third-party backup applications.

How virtualization improves the backup and recovery process

One of the key benefits of virtualization is that it provides unique and lower-cost options for data protection and disaster recovery. The reason for this is that a VM is encapsulated in a single disk file, which allows for image-level backups instead of traditional file-level backups. By using image-level backups, a VM can be backed up much faster. Image-level backups also allow for greater flexibility when restoring a VM, as either the whole image can be restored or just the individual files.

Encapsulation of a VM in a single disk file also provides the ability to easily replicate the VM for application or disaster recovery purposes either on-site or off-site. On-site replication is used for faster application recovery as the VM is located on a spare ESX(i) server in the native, ready-to-restore format, while off-site replication is used for disaster recovery purposes. A spare ESX(i) server can hold many more VMs than the production ESX(i) server, making data protection and disaster recovery really affordable.

vSphere builds on the inherent benefits of virtualization for backup and recovery to provide organizations of all sizes with advanced data protection that leverages the best recovery time objective (RTO) and recovery point objective (RPO) techniques, such as replication and continuous data protection (CDP).

FIVE SIGNIFICANT IMPROVEMENTS

vSphere improves backup and recovery in five important ways:

1. **30% average reduction in storage space** on both the primary VM datastore and on backup storage due to thin provisioning.
2. **50% average increase in backup speeds when performing full backups** also due to thin provisioning, which results in smaller backup windows.
3. **90% average increase in backup speeds when performing incremental backups** and replication due to the new Changed Block Tracking technology of the vStorage APIs.
4. **Near continuous data protection (CDP) at 20% of the cost of traditional CDP** solutions. Faster replication cycles are possible due to Changed Block Tracking, with increments as small as five minutes.
5. **Increased reliability.** The vStorage APIs offer great reliability over VMware Consolidated Backup (VCB).

As you can see, VMware vSphere provides several technological advances for storage that become compelling reasons to upgrade. Let's take a look at the technology behind some of the storage-related features that makes these benefits possible.

FOUR ENABLING TECHNOLOGIES

Of the many new storage-related technologies in vSphere, the following four offer direct benefits to backup and recovery of VMs by improving backup speeds, reducing backup windows, and reducing the amount of storage required for backups:

- Improved support for thin-provisioned virtual disk files
- New vStorage APIs for third-party integration
- Improved iSCSI software initiator
- Hot-add of virtual disk files

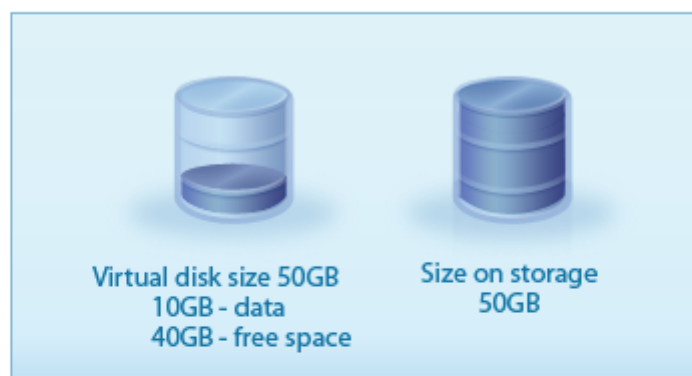
The remainder of this paper is devoted to a closer examination of each of these new or improved technologies.

THIN PROVISIONING

What is thin provisioning?

There are two types of virtual disk files: thick disks and thin disks. When you create a VM with a thick disk, all space is allocated at once. If you select a disk size of 50GB, the resulting disk file takes up 50GB of space on the storage volume. Those blocks of space are empty, but “used” and unavailable for other VMs.

Thick Disk



Thin Disk

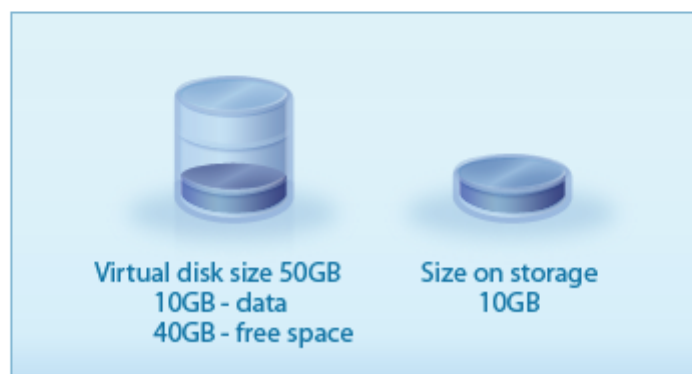


Figure 1. A thick disk with 10GB of data uses the full amount of storage allocated to it, in this case 50GB. If configured as a thin disk, it uses just 10GB of storage.

When a VM with a thin disk is created, the initial disk size is only 1MB regardless of the size that is specified when it is created (or up to 8MB, depending on the default volume block size). Using the same example as before, a 50GB disk uses 1MB of space initially and grows in 1MB increments as data is written to it.

How did thin provisioning change in vSphere?

Thin disks existed in VI3, but they were not widely used because they were difficult to create, maintain, and monitor. VMware made some important changes in vSphere that make using thin disks easier. Firstly, in VI3 a thin disk could only be created using a command-line utility after creating a VM. In vSphere, a thin disk can be created while creating a VM via the vSphere client. Secondly, in VI3 you had to use the command line to convert a thick disk to a thin disk or shrink a thin disk. In vSphere, you can use Storage VMotion to convert a thick disk to a thin disk or shrink a thin disk (by zeroing out the disk blocks in the OS and then doing a thin-to-thin conversion). Finally, in VI3 it was hard to figure out the true size of a thin disk and monitor its growth. vSphere provides built-in reporting and alarms to keep track of thin disk size.

Why are thin disks important?

Thin disks allow you to over-commit disk space, similar to how you can over-commit memory. Most VMs do not use all the disk space allocated to them, which is a wasteful use of valuable disk resources. Many disk-to-disk backup applications use image-level backups when backing up a VM. What this means is they back up the single large virtual disk file of a VM no matter how much space is in use inside it. Veeam Backup & Replication looks for empty disk blocks to avoid backing them up, but this process takes additional time and resources.

When a thin disk is used with a VM, it is no longer necessary to seek out empty disk blocks to avoid backing them up as the thin disk contains no empty blocks. While thin disks help to reduce backup times and the amount of space used on host datastores, Veeam Backup & Replication can also reduce the amount of space needed on the target backup storage device using built-in compression algorithms. Using thin disks in conjunction with Veeam Backup & Replication can provide an average of 30% disk space savings on host datastores and backup storage devices.

VSTORAGE APIS

What are the vStorage APIs?

An application programming interface (API) is an interface that allows external applications and scripts to communicate with an operating system or application such as vSphere. APIs leverage function calls

built in to vSphere, each of which links to a subroutine that executes when the function is called.

vSphere has numerous APIs for different areas. The vStorage APIs allow for specific storage-related functions and are composed of several sections:

- vStorage APIs for Multi-Pathing (VAMP) – extend the multi-pathing capabilities in vSphere.
- vStorage APIs for Site Recovery Manager (VASRM) – extend the remote replication capabilities of Site Recovery Manager.
- vStorage APIs for Array Integration (VAAI) – allow array vendors to offload certain tasks from the VMkernel (not available yet).
- vStorage APIs for Data Protection (VADP) – provide enhanced functionality and greater flexibility for third-party backup and data protection applications.

The vStorage APIs are similar to VCB as they both allow for direct access to virtual disk files. The vStorage APIs, however, are not a standalone application like VCB and do not rely on VCB to function. The vStorage APIs are essentially VMware's evolution of VCB to allow for greater flexibility, greater reliability, and more capabilities for backup applications.

How did the vStorage APIs change in vSphere?

The vStorage APIs existed in VI3 but were referred to as the VCB Backup Framework. VMware renamed the vStorage APIs with the vSphere release and also added new features to them. One of the more notable new features is Changed Block Tracking (CBT), which allows the VMkernel to track changed blocks of a VM's virtual disk.

The vStorage APIs also have the ability to access Virtual Machine File System (VMFS) volumes directly on a storage area network (SAN) without having to use VCB. Assuming their backup application supports the vStorage APIs, this is a benefit to those who want to stop using VCB for backups, even if they have not yet upgraded to vSphere, since this feature of the vStorage APIs is available for both vSphere and VI3.

Why are the vStorage APIs important?

By using Changed Block Tracking, backup applications do not have to scan or keep track of changed blocks on their own and can simply query the information using an API call to the VMkernel. This greatly reduces the amount of time and resource usage when backing up VMs. By leveraging Changed Block Tracking, incremental backups are much faster because it's no longer necessary to scan the whole VM image for changes since the last backup.

This equates to:

- Less I/O on the storage device and over the network
- Less I/O and CPU usage on the host server

The amount of resource reduction with Changed Block Tracking and where it occurs depend on the backup mode. Veeam Backup & Replication offers several different modes: SAN mode, Virtual Appliance mode, and Network mode.

The different modes provide different benefits. When using the SAN mode, the backup server connects directly to the SAN using the vStorage APIs, which allows it to directly access virtual disk files. With this method, Changed Block Tracking results in less I/O on the SAN storage device.

In Virtual Appliance mode, the backup appliance VM connects to virtual disk files by hot-adding them from the shared storage volume. With this method, Changed Block Tracking results in less I/O on the shared storage volume and less I/O and CPU usage on the host of the VM being backed up. (Note: Only one appliance is needed per cluster.)

In Network mode, the backup server connects over the network to the host of the VM that is being backed up. Changed Block Tracking results in less I/O on the shared storage volume and less network I/O.

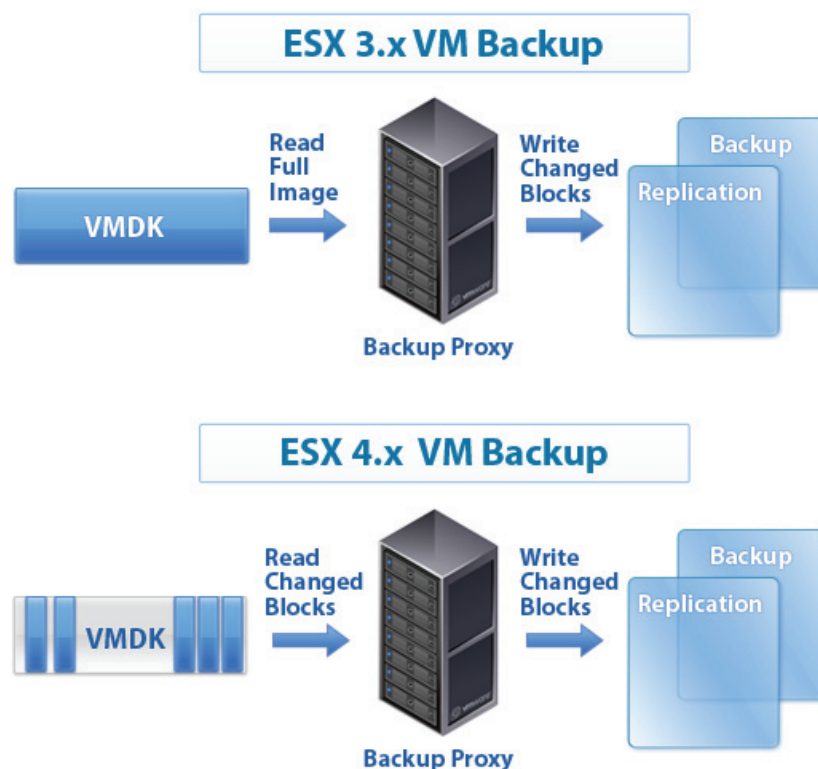


Figure 2. Incremental backup with VMware vSphere requires one-tenth the time since there is no need to scan the disk for changed data, which is instead provided automatically by native vStorage APIs.

As you can see, the Changed Block Tracking feature can result in much less resource consumption when performing backups, as well as smaller backup windows.

ISCSI SOFTWARE INITIATOR

What is the iSCSI software initiator?

ESX and ESXi both have an iSCSI software initiator built in to the VMkernel that allows hosts to connect to iSCSI storage devices using a network interface card (NIC) in the host server. A software initiator differs from a dedicated hardware initiator by using software code as a device driver that leverages the network stack to provide block-level access to storage devices for VMs. Because it is a software implementation of iSCSI there is some CPU overhead, but this can be eliminated by using a dedicated hardware initiator adapter. Software iSCSI initiators provide a cheaper shared storage alternative to Fibre Channel SAN storage while still providing decent performance.

How did the iSCSI software initiator change in vSphere?

The iSCSI software initiator was introduced in VI3 and improved in vSphere. VMware completely rewrote the iSCSI software initiator stack in vSphere to reduce the amount of CPU overhead. In addition, VMware optimized the cache affinity, enhanced the VMkernel TCP/IP stack, and made better use of internal locks. By doing this, VMware has optimized the iSCSI software initiator for virtualization I/O and provided significant reductions in CPU overhead and increased throughput from the previous VI3 release.

Why is the iSCSI software initiator important?

Many organizations use software iSCSI initiators as a more affordable method for implementing shared storage in their virtual environments. The enhancements in vSphere provide more efficiency in the storage subsystem, which directly benefits all storage operations, including I/O-intensive backup and replication operations. The result is more disk I/O throughput, which is important when backing up the environment and helps reduce backup windows.

HOT-ADD OF VIRTUAL DISK FILES

What is hot-add of virtual disk files?

Many people identify hot-add as the ability to add new hardware, such as virtual disks, to a VM that is powered on. There is also a hot-add API feature that allows one powered-on VM to access the disk file of another powered-on VM so it can read the data from the other VM's disk file. This is not something that you can normally do in the vSphere client, but by leveraging special API calls it is possible for a backup appliance VM to access the read-only snapshot disk of a VM so it can read the data and back it up. To use this feature, the VM to be backed up must reside on shared storage, and the backup appliance VM must be able to see that shared storage. The hot-add feature requires that your hosts have either a vSphere Advanced, Enterprise, or Enterprise Plus license.

How is hot-add of virtual disk files different in vSphere?

Hot-add of virtual disk files was available in VI3, but you could not add a disk of a running VM to another through an API. In VI3 the hot-add feature was available in all editions, even free ESXi, but this was changed in vSphere, so only the more expensive versions support it.

Why is hot-add of virtual disk files important?

Providing backup and data protection applications that run as virtual appliances with the ability to access data from other running VMs allows them to access the data directly at the storage layer and not have to go through the network layer, thus reducing the load on the network during backups. It also allows for much faster backups.

SUMMARY

It's often challenging to find available backup windows in a busy production environment. The greater efficiency available from the technological advancements in vSphere can greatly reduce backup windows, making scheduling easier, and reduce the amount of stress the backups place on resources in your virtual environment.

The table on the following page summarizes the major enhancements of VMware vSphere over VI3 in the storage and data protection area.

Technology	VMware VI3	VMware vSphere	Benefits
Thin provisioning	Partial	YES	- 30% average reduction in storage space - 50% average increase in backup speeds for full backups
Changed Block Tracking (CBT)	NO	YES	- 90% average increase in backup speeds for incremental backups
Near continuous data protection (CDP)	NO	YES	- Affordable near CDP at 20% of the cost of traditional CDP
Reliability	Partial	YES	- Highest reliability due to next-generation vStorage technology
iSCSI software initiator	Partial	YES	- Affordable shared storage
Hot-add of virtual disk files	Partial	YES	- Faster backups and less load on network in Virtual Appliance mode

The thin-provisioning improvements are leveraged by Veeam Backup & Replication, reducing the amount of resources used during backups and the amount of storage needed to store VM data. The improved iSCSI software initiator provides greater throughput for backups to virtual machines on iSCSI datastores.

Veeam Backup & Replication 4.0 leverages each of vSphere's new storage-related features so organizations can achieve the five benefits mentioned earlier in this whitepaper: a 30% average reduction in storage space, a 50% average increase in backup speeds when performing full backups, a 90% average increase in backup speeds when performing incremental backups, near continuous data protection (CDP) at 20% of the cost of traditional CDP, and increased reliability.

To experience these benefits for free, visit Veeam's website at: <http://www.veeam.com/backup> to download a complimentary trial copy of Veeam Backup & Replication 4.0.

ABOUT THE AUTHOR

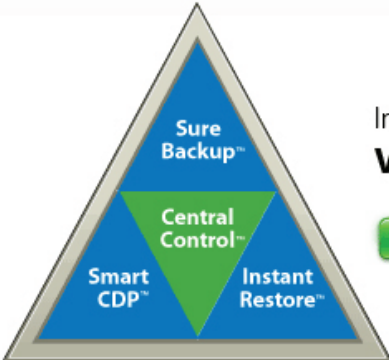
Eric Siebert is a senior systems administrator at a major restaurant company. He is a 25-year IT veteran whose primary focus is VMware virtualization and Windows server administration. He is one of the 300 VMware vExperts for 2009, author of the book *VI3 Implementation and Administration*, and a frequent TechTarget contributor. In addition, he maintains vsphere-land.com, a VMware information site, where you can find a post about this article: <http://vsphere-land.com/news/new-white-paper-on-5-ways-vmware-vsphere-improves-backup-and-recovery.html>.

ABOUT VEEAM BACKUP & REPLICATION

Veeam Backup & Replication is an evolution of the proven Veeam FastSCP engine, providing unmatched reliability verified by nearly 100,000 VMware professionals over the last three years.


Through continued, rapid innovation in response to customer feedback, Veeam Backup & Replication offers many industry-firsts, including full native support for advanced functionality in VMware vSphere. It helps ensure business continuity in virtual environments and allows customers to reduce costs, minimize risks, and enhance service levels.

Virtualization-Powered Data Protection™



Integrated with FastSCP
Veeam Backup & Replication

[Learn More](#) [Download](#)



SureBackup™

- 2-in-1: backup and replication
- 10x faster: vSphere and vStorage
- Proven architecture; FastSCP engine
- SmartDedupe™

InstantRestore™

- Best RTO and RPO
- VM or file level (Windows and Linux)
- Self-restoring backups
- Advanced VSS support

SmartCDP™

- Replicate on-site for HA, off-site for DR
- Fraction of cost of traditional CDP
- vSphere Changed-Block Tracking
- Replica rollback

CentralControl™

- Federation of backup servers
- Reporting and alerting
- Unlimited scalability
- Automation via SDK

#1 for Virtualization™

Data Protection and Management