

Realtime  
publishers

The Essentials Series: New Techniques  
for Creating Better Backups

# How Virtualization Changes the Standard Approach to Backups

*sponsored by*

**ARC**serve®  
More than Backup

by Eric Beehler

---

How Virtualization Changes the Standard Approach to Backups .....	1
Virtual Backup Methods .....	2
Agent-Based Backup.....	2
Host-Based Backup .....	3
Consolidated Backup.....	3
How to Approach These Options.....	4
Host-Based Backup .....	5
Consolidated Backup.....	6
Summary.....	7

---

## Copyright Statement

© 2009 Realtime Publishers. All rights reserved. This site contains materials that have been created, developed, or commissioned by, and published with the permission of, Realtime Publishers (the “Materials”) and this site and any such Materials are protected by international copyright and trademark laws.

THE MATERIALS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. The Materials are subject to change without notice and do not represent a commitment on the part of Realtime Publishers or its web site sponsors. In no event shall Realtime Publishers or its web site sponsors be held liable for technical or editorial errors or omissions contained in the Materials, including without limitation, for any direct, indirect, incidental, special, exemplary or consequential damages whatsoever resulting from the use of any information contained in the Materials.

The Materials (including but not limited to the text, images, audio, and/or video) may not be copied, reproduced, republished, uploaded, posted, transmitted, or distributed in any way, in whole or in part, except that one copy may be downloaded for your personal, non-commercial use on a single computer. In connection with such use, you may not modify or obscure any copyright or other proprietary notice.

The Materials may contain trademarks, services marks and logos that are the property of third parties. You are not permitted to use these trademarks, services marks or logos without prior written consent of such third parties.

Realtime Publishers and the Realtime Publishers logo are registered in the US Patent & Trademark Office. All other product or service names are the property of their respective owners.

If you have any questions about these terms, or if you would like information about licensing materials from Realtime Publishers, please contact us via e-mail at [info@realtimepublishers.com](mailto:info@realtimepublishers.com).

---

# How Virtualization Changes the Standard Approach to Backups

---

Virtualization is the hot technology trend for business and it's easy to see why. It aims to gain efficiencies and take advantage of performance, which are goals that fall in line with an effort to reduce costs in data centers and be more energy efficient. But with virtualization comes certain challenges. It's testing the limits of many traditional approaches to data center management. From extreme heat from consolidated blade servers racks to increased need for network bandwidth, we have seen virtualization relieve our challenges in some ways, namely simplification and consolidation of servers, but put other technologies on a critical path to the breaking point.

One such technology that requires new thinking is data backup, a critical piece of any IT strategy. Although backup is considered essential to keep business up and running and is a well-ingrained requirement to most IT departments and administrators, it is often given little attention in many virtualization projects. There are challenges to be met on the backup front when delving head-first into virtualization. Understanding the specific challenges and the methods to meet those challenges will allow success and limited interruption when undertaking the task of migrating a backup environment to the virtual world.

When considering a standard physical server environment, server backup has been handled in a straightforward manner. Most solutions rely on an agent installed on the server and then a separate backup server and tape library that accepts the bits from that server to tape over the network. There is usually a set period for backups to occur at night or at some lower volume period when network utilization is low and the servers can be dedicated to the backup task. Next, consider a virtual machine, which could be sharing a host with 8, 20, or even more virtual machines. These virtual machines also share disk storage and network bandwidth. With every virtual server pushing the full contents of their hard drives over the network, often compressing the data as they do so, there can be contention for those host resources. When all machines are running at a high utilization for CPU, disk I/O, and network, there can be contention for resources that is multiplied by the number of servers. In addition, you may be backing up unnecessary data because virtual machines are so easy to deploy. When planning virtualization backup, you must look at the available solutions and reconsider the backup solution as well as the backup policies in place.

---

## Virtual Backup Methods

If a traditional physical server-and-agent method is used without modification of the backup jobs, one would likely see contention on the CPU of the host machine. Also common are saturation of the physical network interfaces that are shared by those virtual machines and by the backup server and library, which will be contending with many more bits than before. This setup can end up pushing the backup period outside the expected backup window, which can affect prime-time processing. Virtualization's biggest benefit, multiple servers on a single host, can also be its downfall when considering agent-based backup.

### Agent-Based Backup

There are distinct benefits to agent-based backup on the virtual machines. First, you don't have to throw away a process that is working today. As many systems administrators will attest, tuning a backup environment takes time; once it's in place, it's difficult to justify implementing an entirely new system.

Another big benefit is granularity. When backing up the individual files of each machine, the files from all volumes are saved and cataloged in the backup software. With the advances of modern backup systems, finding a specific file and restoring it from a specific time is relatively pain free. The method is usually just a click and go.

Another consideration is the number of virtual machines in your environment. In many environments, virtual machines are easy to bring online, which might lead to *agent sprawl* if you are installing the agent on every virtual machine. Then the question becomes, What are you using the backup capacity you have available to back up? If the answer is "C:\Windows" 500 times, the efficiency of backups can be called into serious question. Tracking the configuration of every virtual machine and what should be backed up would certainly help and should be done, but configuration management at that level is a serious undertaking.

In addition, you must consider the status of servers that are in a suspended state or just powered down. This was hardly an issue in the traditional data center, as all machines would typically be on all the time. Not the case with virtual machines, which can be easily suspended and even moved between hosts.

---

## Host-Based Backup

You may still want to use agents, but instead of backing up from the virtual machine, you could back up at the host. Host-based backup allows you to back up just the virtual hard disk files that make up each virtual machine, and contains all the data from that virtual machine. There are several benefits to this method. First, you will reduce the number of backups running at once. For example, if you have 16 virtual machines on a single host, you'll be reducing the number of backups by one-sixteenth. This also reduces the number of agents you have to install, track, and possibly purchase licensing for. You are still dealing with features such as agent compression, but not on every single guest virtual machine. This option also allows you to back up virtual machines that are in a suspended or powered-off state.

One of the downsides to host-based agent backup is the absence of granularity, even though you are using an agent. When backing up the actual snapshot of a machine via its virtual hard disk, the files that may need to be restored are in that container. So, instead of having a nice, cataloged set of file systems, you now must know where the file will be and undertake the extra step of mounting a virtual machine in order to grab files from within the larger container file. Some backup software does make this easier, though, and provides a simpler one-step recovery solution, eliminating the need to stage the machine first.

In addition, tracking server configuration becomes an especially critical issue because machines are often moved between host machines in a virtual environment using features such as VMware's VMotion feature. If this kind of migration feature is used, systems administrators will find the flexibility great for daily operations but an extra challenge when trying to track down a virtual machine in the backup set.

Another challenge is SAN- and NAS-connected storage. When backing up from the host, the machine will include local drives but not logical disks on a SAN. If the server is mapping to Fibre Channel- or iSCSI-connected volumes, they are not saved in the virtual machine file as local drives and will not be included in the host-based backup. You may also have to script automatic shutdown of virtual machines in order to back up the virtual hard disk (VMDK or VHD) files.

## Consolidated Backup

Another option is consolidated backup. Consolidated backup, sometimes referred to in varying terminology such as *hot backup* or *off-host backup*, is a virtualization-specific backup solution. It works by first taking a snapshot of the virtual machine on the host. Then, instead of moving that snapshot of the virtual machine over the network, the snapshot can be transferred over the Fibre Channel or the network to a backup proxy. From there, it will be transferred to backup disk or backup tape. The network is left out of the picture in most cases, so the time to back up can decrease from hours to minutes. The backup proxy also has the ability to mount the snapshots and gain file-level access to the backup. The speed benefit also means less demand on host resources. In addition, most backup software will understand and work with a consolidated backup proxy server, allowing you to integrate into your existing infrastructure for backups.

---

The downside to consolidated backup is the dedicated hardware and software you have to place into your infrastructure. The cost for these solutions can be significant, and if management isn't ready to sign off on an additional component for backup, there isn't much to talk about technically. The benefit of a consolidated proxy server is really just mounting a snapshot or point-in-time image of the server and backing up the individual files for granularity. If file-by-file backup isn't required, snapshots of servers can be restored very quickly and brought online without much thought. A whole server restored to the state it was in 24 hours ago goes a long way to covering the backup bases in many cases—especially for servers that don't function as file servers. And in many cases, servers that may be running another solution, such as an Exchange mail server or database server, will have their own dedicated agent method for backup anyway. Take into account the extra costs of a consolidated backup solution before dismissing the other solutions because of the appearance of consolidated backup's technical superiority.

## How to Approach These Options

Keeping a traditional agent-based backup in place can be an excellent low-cost, low-hassle option if your environment can support it. To really know whether this will be the right solution, some benchmarking should be done. Deploy the agents to the virtual machines and take CPU and network throughput measurements using your virtualization tools or Performance Monitor (perfmon) counters in Windows for Hyper-V. (Hyper-V has its own set of performance counters listed under the Hyper-V set of counters.) You might be most interested in the overall performance of the host, though, so standard counters for CPU, disk, and memory are a good start.

For VMware, you can use specific ESX service console counters from the host to look at overall system performance. For a basic view, from the service console utility, run `Esxtop`. Doing so will give you real-time performance metrics for CPU, memory, disk, and network performance. If you have access to the VMware Infrastructure (VI) client, open it and select the Performance tab. Gather statistics such as CPU usage to build a proper chart. You can also use some of the other available performance monitors, such as VirtualCenter and other third-party utilities for VMware to gather these statistics. Have the virtual machines run their backup schedule according to normal parameters. If there is little effect and the backups finish on-time, there may be little to no tweaking involved.

If the results are not optimal, especially compared with the normal backup of physical-only servers, first consider looking at the backup job scheduling. Move jobs around so that the number of simultaneous backups is reduced. For example, if you have eight virtual machines hosted on a single server, consider running only one or two backups from that host at any one time and rerun your performance counters to compare. In addition, consider full versus incremental or differential backups. Consider staggering full backups. Full backups always use the most resources, so running them staggered on the same host would likely result in a performance gain.

---

Of course, there is the option of adding more bandwidth with additional network cards. In addition, try separating the VMDK or virtual hard disk files onto separate volumes of storage. If the virtual hard disk files are locally stored, move them to separate volumes. When hosted on a SAN or iSCSI LUN, separate them by LUN or volume as well. This will reduce contention from the disk by separating file systems.

Another clever option is to run the backup server on a virtual machine that is connected to the host of the servers being backed up. An internal network to the virtual host can be set up that doesn't interact with the physical network so that all network traffic is passed on a virtual network inside the host machine itself. This setup eliminates the need for network bottleneck concerns but is only appropriate for simple virtual environments with available backup hardware to connect in a direct manner.

### Host-Based Backup

Running the agent on the host computer instead of the virtual machine comes with its own set of requirements. First, when considering VMware, a Linux agent must be used. For Hyper-V, obviously Windows Server 2008-compliant agents are required. The backup agent will be dealing with fewer big files instead of many smaller files, which should improve performance. Remember to back up all volumes that contain the virtual machine files. For example, if there are five separate volumes that contain virtual hard disks, include them all in the backup job.

In order to ensure consistency, unless the agent is virtual machine-aware and can back up running virtual machines, the virtual machines will need to be shut down. Backing up a running virtual machine may lead to inconsistencies that may not be recoverable when mounting a virtual hard disk for restore purposes. Consider instituting scripts that will shut down and restart a virtual machine when a backup is about to run, but be aware that this kind of requirement can certainly impact service level agreements (SLAs) for uptime for these virtual servers.

You will need to consider uptime requirements and understand the capabilities of the agent in regards to virtual machine backup. In VMware, a snapshot of a virtual machine can be created using VCB utilities on the ESX Service Console. For example, vcbMounter will allow you to back up the virtual machine into a specific location, either local or remote over the LAN. This will take up additional disk space and require more scripting but would avoid downtime. With Hyper-V, if the guest operating system (OS) is Windows Server 2008 or 2003, Volume Shadow Copy can be utilized to prevent inconsistencies. Oftentimes, a virtual machine-aware backup agent will handle snapshot duties.

Restores are handled differently from the virtual machine guest installed agent. In essence, there is no file-level restore when backing up entire virtual hard disks. The benefit is that a full machine can be brought back from a crash or from a change that caused instability just by shutting down the existing virtual machine and replacing the virtual hard disk file with the version from backup. VMware offers the vcbRestore utility, part of ESX Service Console, to restore a virtual machine to its original or remote location.



---

In addition, in order to access individual files and directories inside a virtual machine backup, the virtual hard disk must be mounted. This can introduce several issues. First, if the virtual machine is brought online on the same network as the running virtual machine, it will cause network conflicts. In addition, there is the possibility of Active Directory (AD) conflicts when a machine tries to register with duplicate SIDs. You also run the risk of keeping the incorrect version of a machine online and in production.

Oftentimes, when doing snapshot-style backups, administrators will not support restoring individual files because of the complexities as well as the fact that individual files within those machines are not cataloged in the backup software. Also, finding the proper machine from host-based backup can be difficult if virtual machines are moved between hosts; VMware VMotioned machines require you to track the backup by host.

Another consideration is data that exists outside of the virtual machine's local disk. If virtual machines are connecting to a remote volume such as a SAN, you must make plans to back up that data separately. You may want to run a virtual machine-based agent backup for those volumes. Although the mappings will be restored with the snapshot, the data on those volumes is not.

### Consolidated Backup

The limitations of these other backups have resulted in consolidated backup. These solutions are vendor-specific and require additional hardware and software. They provide the framework necessary for a virtual machine to be backed up online, without shutting down, as a snapshot. That snapshot is then stored on the SAN, eliminating the issue of network utilization. This architecture enables the restore process for a whole virtual machine to be exceptionally fast, using storage connections to restore these large files after a virtual machine failure.

One must be conscious, however, with application supportability, particularly with specialized applications such as databases and email systems such as Exchange. These systems require additional backup methods, such as a specific backup agent for utilization of built-in backup to a local file on the virtual machine's hard disk that can be restored through normal restore functions.

As mentioned earlier, third-party backup software is integral to the concept of consolidated backup. And although the framework of consolidated backup allows a snapshot to be taken and creates a place to put that snapshot, it's still the agent and related backup software/hardware doing the real backup of those files.

Consolidated backup features being implemented include better catalogs of files and direct restores of virtual machines. Some solutions even include compression on the snapshot files themselves at the host. Others are bypassing the need for the proxy server and providing direct access to files inside of snapshots, and still others are allowing incremental versions of snapshots (normally an entire snapshot is backed up every time).

A final area of concern is consistency of Windows snapshots. Vendors taking advantage of Volume Shadow Copy are now able to help guarantee that consistency. Some are also

---

approaching the issue by installing their own virtual machines on each host to handle backup duties.

## Summary

Virtualization is quickly becoming the production norm instead of the test bed. As companies look to save money on hardware and electricity and to contribute to green standards, virtualization offers a solution with many benefits. However, there are challenges to be faced to keep backups a non-issue in a virtualized environment. Consider this an opportunity to look at the backup environment with a new view, understanding that current practices may have to be dropped in favor of modern approaches. It's likely many will end up approaching the problem with a hybrid solution, using the best approach for a certain set of servers. However you look at the issue of backup, virtualization has changed the game.