Agent and Agentless Virtual Machine Backup and Recovery – Unraveling the Myths

Data Sheet: Backup and Disaster Recovery

Agents. Agentless. VMware® vStorage™ Application Programming Interfaces (APIs) for Data Protection (VADP) integration. Microsoft® Volume Shadow Copy Service (VSS) integration. Image-based backups. File-based backups. Hypervisor-based snapshots. Array-based snapshots. Host-based backups. Guest-based backups. It's no wonder why backup professionals are confused about what the best approach is for backing up their virtual machines. With a myriad of vendors, all positioning their own way as the "best" way, it leaves in its wake ambiguity and a good dose of confusion about what an agent is, or does.

With the help of my technical experts here at Symantec, this paper distils the confusion with unbiased information so you can make the right choice for your environment. By looking at each method and highlighting the pros and cons of each, you can make informed decisions without the distraction of smoke and mirrors.

Caution: Before we dive in, it's important to mention that the phrases *agentless backup* and *agent-based backup* can mean different things to different vendors. To truly determine what the best approach for your organization is, you need to look under the cover and weigh in the pros and cons of each method. Don't worry; we have done the hard work for you at Symantec. Now let us jump in and take a look at each one in turn.

1) Traditional agent-based backup (also known as guest-based backup)

Traditional agent-based backup is also known as guest-based backup. An agent is installed in every virtual machine and treats each virtual machine as if it was a physical server. The agent in this scenario is reading data from disk and streaming the data to the backup server. This method must not be confused with agent-assisted backups that we will cover later.

There are many people today using this approach to protect their virtual machines. According to ESG, 46 percent of all environments are utilizing guest based protection methods with a backup agent running inside the guest operating system (OS).[1] Although there are newer methods available, you may be asking yourself why so many people are still using this method. The main reason is because it's a time tested and proven solution.

Pros:

- Both physical and virtual machines are protected using the same method
- Application owners can manage backups and restores to guest OS
- Time tested and proven solution
- Meets their recovery needs
- This is the only way to protect VMware Fault Tolerant virtual machines and virtual machines with Physical Raw Disk Mappings RDMS

Cons:

- Significantly higher CPU, memory, I/O, and network resources utilization on virtual host machines when backups run.
- Need to install and manage agents on each virtual machine
- Cost may be high for solutions that license on a per agent basis as opposed to per hypervisor based licensing
- Cannot accommodate virtual machine sprawl, lack of visibility into changing virtual infrastructure

[1] ESG Research Report, 2012 Trends in Data Protection Modernization, August 2012.

- No visibility for backups from virtual machine administrators' point of view; for example, backups are not visible at VMware vSphere™ client level
- May need multiple kinds of backups and recovery methods; for example separate backup policies may be needed for file and folders backups, Microsoft® Exchange backups, bare metal recovery, etc.
- Complex disaster recovery strategies
- Lack of storage area network (SAN) transport backups to offload backup processing job from virtual infrastructure
- No protection for offline virtual machines and virtual machine templates
- Slow file by file backup by agent sending the even unchanged data over and over again

Verdict: A cumbersome, traditional backup and recovery method, but offers flexible recovery features.

2) Agentless backup (also known as host-based backup)

Agentless backup, also known as host-based backup, refers to solutions that do not require an agent to be installed on each virtual machine by the administrator. However, it's important to note that the software may be injecting an agent onto the guest machines without your knowledge.

These solutions integrate with VADP or VSS, which creates fast, high performance snapshots of the virtual disks attached to virtual machines. The backup software communicates with VADP or VSS and tells it what it wants to backup. VADP and VSS carry out a number of steps and in turn prepare the data to be backed up. The VSS/VADP provider snaps the volume and gives the backup solution access to this snapshot by feeding the file to the backup server. The backup solution then backs up the snapshot.

While it provides recovery for full virtual machines, files, and folders, the recovery of applications and application data can be complex and time consuming. This is because it requires additional processing that engages resources external to the virtual machines. Applications on these hypervisors won't truncate their transactions logs or perform other database maintenance tasks. An Exchange Server is a perfect example. Without an agent or agent-like executable in the virtual machine gathering metadata about the Exchange information store, there is a need for additional processing external to the exchange virtual machine in order to map mailbox data. If you ignore this process, it can result in unmanaged transactional applications that must be manually managed by the application owner, and data that might only be recovered by first restoring the entire virtual machine and its virtual disks. Therefore, one of the key differences between agentless and agent-assisted backups is how the transactional post-processing happens.

Pros:

- Virtual machines can be backed up online or offline
- Less CPU, memory, I/O, and network impact on the virtual host
- An agentless architecture doesn't require the management of agent software
- No per virtual machine agent licensing fees

Cons:

- Extremely difficult to recover granular object data—first restore the entire virtual machine and its virtual disks
- Traditional login techniques to log into the server
- Temporary "injected" drivers can destabilize the system and compromise data integrity
- Troubleshooting is more complex when using injected (temporary) agents
- A centralized controller is a single-point-of-failure

- Requires a fully-virtualized environment. Physical machines still require agent-based backup. If you have physical and virtual you will need two backup solutions—one for physical and the other for virtual.
- Additional processing e.g. post backup scripts and truncation engages resources external to the virtual machines

Verdict: Good method for protecting file and print servers, but not an optimal solution for virtual machines with applications. Recovery is operationally painful for applications and application data.

3) Agent-assisted backup: Next generation backup (also known as host based backup)

Agent assisted backups are also known as host based backup and integrate with VADP and VSS to provide fast and efficient online and offline backups of VMware ESX®, vSphere, and Microsoft Hyper-V®. The primary difference between agentless and this design is its perspective: it pairs the VADP or VSS with an agent that gathers application metadata to enable multiple avenues of recovery (full virtual machine, applications, databases, files, folders, and granular objects). The agent-like executable in this instance is not carrying out the backup and thus does not impact the performance of the virtual machine. It's simply handling metadata and necessary post backup processing like log truncation.

Pros:

- The backup is for the entire virtual machine. This is important because it means the entire virtual machine can be recovered from the image. It also means that products like Symantec Backup Exec™ and Symantec NetBackup™ can offer "any-level" of recovery from the image contents: Files/folders, databases, granular database contents, like email and documents.
- The backup can be offloaded from both virtual machine as well as the hypervisor. This means that Backup Exec and NetBackup have the flexibility to offload virtual machine backup onto an existing backup server, instead of burdening the hypervisor. It also means that users have the option of deploying a dedicated virtual machine, e.g. a virtual appliance, when a physical backup server is not practical.
- Application owner can self serve restore requests: The application owner can request restores directly back to the application.
- Enhanced security: The agent installed for assisting with virtual machine backup can be managed by the application owner. Thus you are avoiding the need to share guess OS credentials with backup administrator.

The most resource efficient backups are the backup operations done at the hypervisor level and provide some of the following advantages:

- Backup is still image based. Leverages VADP/VSS.
- Ability to directly recover files and folders directly back to a virtual machine.
- Enable automatic discovery of application inside virtual machine
- Granular Application Recovery
- For VSS compliant applications, backup is application consistent via VSS integration
- For non-VSS compliant applications, backup is crash consistent
- Less performance and I/O impact on the virtual machines
- Can be on local area network (LAN) or SAN interface

Verdict: Excellent method for virtual machines with applications like Microsoft Active Directory®, Microsoft® Exchange, SQL®, and SharePoint®.

It is very important to understand that backup vendors who utilize VADP or VSS to perform backups are not the same. Some are better than others. How good the backup software is will depend on the validation phase. So don't be fooled in thinking that all solutions with VADP or VSS integration offer the same functionality and benefits. A quality backup application that integrates with VADP or VSS to perform backups should at the very minimum:

1. Validate snapshots prior to backup
2. Use VMware change block tracking (CBT) mechanism to have smaller storage footprint by backing up only changed data
3. Verify backup data
4. Backup virtual machine directly from storage location for example, SAN, iSCSI, network attached storage (NAS), without having to install any software a.k.a agent inside the virtual machines
5. Offer flexible recovery options (Full virtual machine recovery, file/folder level recovery, application level recovery, and granular object recovery)
6. Centralized backups for virtual machines
7. Dynamic inclusion of virtual machines
8. Ability to transport your data offsite for disaster recovery

Conclusion

As you embrace virtualization or increase your virtual footprint, selecting a backup solution that integrates with VADP and VSS will provide fast snapshot-based image backups of online and offline guest virtual machines. For superior recovery capabilities a solution that gathers metadata and executes post processing tasks is a must. So we've learned that the question isn't whether you have an agent or other differently named binary in the guest. The question is what is the agent's function. Jason Buffington, a Senior Analyst at ESG, wrote a great blog on "good agents" and "bad agents". If you would like to learn more, you can check out his blog here: http://www.esg-global.com/briefs/agent-best-practices-for-host-based-backups/.

About Backup Exec

Backup Exec provides a perfect solution for virtual environments with technology that was designed for VMware and Hyper-V powered by patented V-Ray technology. Not only does Backup Exec provide superior data protection for virtual environments, it also provides market-leading technology for physical environments too. With Backup Exec you get it all in a single solution. Backup Exec 2012 dramatically reduces the time to recover from small or major disasters by protecting all of your virtual machines and/or physical servers through a single pass backup, while still allowing for individual file, folder, and granular object level recovery. In short, it's powerful, efficient, reliable, and fast.

More Information

Visit our website

http://enterprise.symantec.com

To speak with a Product Specialist in the U.S.

Call toll-free 1 (800) 745 6054

To speak with a Product Specialist outside the U.S.

For specific country offices and contact numbers, please visit our website.

About Symantec

Symantec protects the world's information and is the global leader in security, backup, and availability solutions. Our innovative products and services protect people and information in any environment—from the smallest mobile device to the enterprise data center to cloud-based systems. Our industry-leading expertise in protecting data, identities, and interactions gives our customers confidence in a connected world. More information is available at www.symantec.com or by connecting with Symantec at go.symantec.com/socialmedia.

Symantec World Headquarters

350 Ellis St.
Mountain View, CA 94043 USA
+1 (650) 527 8000
1 (800) 721 3934
www.symantec.com