STORAGE SWITZERLAND

5 MISPERCEPTIONS OF AGENTLESS VMWARE BACKUP

Storage Switzerland, LLC

George Crump, Senior Analyst

Agents are small applications that are installed on a server to perform a particular function. Common examples are the agents that a backup application will install on a server to back up that server and provide specific support for the applications that run on it. Over the years agents have taken the blame for 'misbehaving' software applications and have become 'persona non grata' in many data centers.

This anti-agent movement reached new levels with virtual server infrastructures. The marketing tirades against agent-based backups are exacerbated in the virtual environment where multiple virtual machines with specific resource requirements exist on the same host which has finite resource requirements. Agentless VM backup is a result of single-point VM backup tools performing VMware backup through the vStorage API. This creates a perception that agentless is lighter and better compared to agents, which are claimed to consume more CPU cycles for overhead and limiting the host resources needed to service requests for service.

Is all the anti-agent sentiment fair? And is the agentless approach so much better that no one should even consider agent-based backup solutions? Most of the arguments in favor of agentless backup software are based on comparisons agentless vendors paint between their agentless offerings and older, legacy agent-based backup schemes. Those older applications, some of which have origins going back to the 1990s, have not been updated to take advantage of new capabilities within operating systems or hypervisors.

But such legacy technologies are not the only agent-based schemes you can choose from. The days where agents had to "hack" their way into the operating system are past. Look instead to the crop of newer "smart agent"-based applications, built from the ground up to support virtual and physical machines. Smart agents offer the functionality and ease of use found in an agentless approach, taking advantage of APIs or 'hooks' that they can reliably and safely snap into.

But that's not enough of a reason to go with smart agents. *This* is: *modern smart agent backup offerings deliver tremendous performance and functional advantages over agentless schemes*. Let's look at some agentless claims and see how they stand up to the new crop of smart agent-based backup and recovery tools.

Misperception 1: Agentless VM Backup uses less CPU
Fact: vStorage API (i.e. "Agentless") can use the same amount of CPU as agents use, and in some cases it can use more

The number-one misconception about the use of agents is that they consume more CPU cycles than an agentless approach. Unlike backups of standalone servers, backups of virtualized servers require careful attention be paid to CPU use because the host processor is a more finite resource, typically shared with six or more virtual machines. To understand this myth, look at the two main causes of a CPU spike when doing a backup.

First, CPU spikes occur when the backup agent has to scan the entire file system to find files that qualify to be backed up, usually those files which have changed since the last backup. The exception is a traditional full backup, since all files need to be backed up, so no scan is needed. During incremental or differential backups this walking of the directory tree is time consuming and processor intensive.

Some, but not all, hypervisors have the ability to report to a backup application just the blocks of data that have changed since the last backup. No scan needs to occur—the information is in a sense "pushed out".

The net impact is that scan times can be eliminated with modern software code. It can be done with an agent, either inside the virtual machine, with change block tracking created by the backup application, or outside of the VM leveraging an API set by the hypervisor provider.

The second cause of CPU spikes in the backup process is the actual transfer of data. This leads us to the next myth about agentless VM backup.

Misperception 2: Agentless backups are faster
Fact: They can be slower and cannot backup more than 3+ VMs concurrently without slowing production applications

In either the agentless or agented use case, data, once identified for backup, has to be transferred across a network to a backup server or target device. Agentless solutions claim an advantage because they use changed-block tracking, so there is less data to transfer. But a modern agent-based solution can also transfer just the changed blocks. The agentless approach is more of a blind sweep and requires a "pull" approach to the backup process which can slow the CPU, whereas a smart agent approach that pays proper attention to CPU utilization can send many backup jobs to the backup server simultaneously. Compare this to agentless backup vendors that suggest a limit on the number of VMs (as few as two) that should be backed up at the same time per host.

Misperception 3: Agentless is truly agentless
Fact: "Agentless" has agents for both Application Item Recovery and for VSS

Many virtual machines run some form of an application like Exchange, Oracle or MS-SQL. Each of these applications has levels of granularity and there are times when only a sub-component of the application needs to be recovered. A common example is the need to recover an email message for a user who accidentally deleted it.

Agentless backups have a challenge when it comes to understanding the data that they are backing up. After all they're just backing up blocks of information, and there's nothing in the virtual machine that understands what those blocks represent. At some point they either need to "break protocol" and deploy application-aware processes on the host or add the same to their backup server. The only other alternative is to require the full restore of a VM as a clone to extract specific application sub-components. Some of the VM backup tools that claim to be agentless actually use agents in order to perform application-item recovery. Thus by default they are not truly agentless.

Additionally, for VSS… single point VM backup tools also use an agent which can be clearly seen and identified in the Task Manager. For this case as well they are, by default, not truly agentless.

Agented backup applications, on the other hand, are already resident in the virtual machine and in most cases support for a specific application merely needs to be turned on.

Misperception 4: Agentless is easier to deploy
Fact: Smart Agent backups are equally easy to deploy and manage

In theory agentless technology should be easier to deploy. After all there are no pieces of software that must be manually installed on each virtual machine. When compared to legacy agented technology that had limited ability to "push install", that's true. But a modern agented application supports an automated, built-in push installation feature available in almost all operating systems. Additionally, in the virtual environment the backup agent can be made part of the "VM Template" so that the agent is auto-installed as each new virtual machine is created.

Using a push install or installing as part of a template also gives insight to a very common problem within the virtual environment—understanding which VMs are being protected and which are not. With an agented approach it's very easy to identify those VMs with an agent installed and those without.

Misperception 5: Agentless permits faster recoveries
Fact: Agentless VMware backup tools are can be slower on recovery

Agentless VM Backup is supposed to have an advantage because the process of recovering a virtual machine simply involves creating another virtual machine and restoring data to it. Some products do automate this process; the only downtime is the time required to copy data across the network. Other products take this a step further by allowing recovery of data directly from the backup store, so there's no downtime due to data transfer.

Agented solutions like those from AppAssure can also recover directly from the backup store but they also have several advantages. In the agentless scenario all data is on the backup target and although the recovered VM is up and running, data still remains on the backup device. Few users in a production environment will want to execute the application from a backup device for very long. At some point a migration back to primary storage will be required, which may involve logging out users and shutting down the VM. Or it may require the purchase of Storage vMotion, which of course adds to the cost of the solution. More importantly, all of these are separate steps that need to be performed one after the other, in the right order, by the virtual administrator.

An agented solution like AppAssure's restores the original via a VMware creation wizard. If there are separate data volumes, products like AppAssure can bring those back via a live recovery feature. In parallel the restore process begins. As users log in to request information, the agent determines if the data has been restored to the VM or not. If it hasn't that data is instantly recovered from the backup with no interruption to the user's request. This feature takes the above multi-step process and integrates it into a single-step with the end result being a fully recovered machine.

Summary

Agentless backup solutions have enhanced their collective reputations by comparing themselves to older legacy backup agents written before operating systems had provisions for agents that can communicate and execute within the OS. But when compared to modern agent-based solutions, like those from AppAssure, a less favorable impression of agentless solutions can emerge.

Modern agented solutions have most of the advantages of an agentless design and they have proven to run faster *and* recover more data with applications like Exchange, SQL Server and Sharepoint. A backup solution should not be judged solely on whether it uses an agent, but on whether or not it can reliably and quickly back up and recover applications in the event of a failure.