Tivoli Storage Management   IBM

Leveraging Tivoli Storage Manager FastBack™ in Virtual Environments

Oren Wolf, IBM Tivoli Storage Product Manager

May 2009

Abstract

Server virtualization is one of the most common scenarios in today's data center and with the latest price drops for hypervisors it is making its way into the SMB market as well. Tivoli Storage Manager FastBack offers some unique technologies that make it a primary candidate to be the data protection and recovery solution in such environments.

This paper briefly describes specific ways in which TSM FastBack can solve business and operational problems in a virtual environment.

Leveraging TSM FastBack in Virtual Environments

What are the basic challenges in data protection for virtual environments?

In an effort to reduce infrastructure and energy costs, organizations are moving aggressively to virtualize servers in order to compact as many virtual machines (guests) into a single physical host (host) as possible. The minimal ratio is usually 1 physical server to 5 virtual ones, but in order to meet more aggressive ROI some data centers will reach a 1:10 ratio and more. While the cost savings from this consolidation are substantial, there are trade-offs. For example, traditional "file level" protection running on the "guest OS" usually does not work in a virtual environment due to the high resource consumption they require to perform file system scans and file transfer operations.

Let's assume that a normal backup process on a typical machine takes 10% CPU utilization and runs for 4 hours. Now, in a virtual environment with 5 guests, let's assume that all these machines need a daily backup and that we went through the burden of scheduling them separately to minimize resource utilization. In this case, the backup will run for 20 hours straight consuming 10% CPU (with the 1:5 ratio). Of course the problem gets bigger when you consider the network bandwidth, physical IO and most important of all, the side effects that this behavior will have on the production work of these servers.

One of the things that get virtualized together with the server is the storage that is being used, but the guest OS still sees a 5/50/500 Gigabyte disk. The host, however, sees this as a flat file (.vmdk for VM, .vhd for Hyper-V).

In order to avoid common "resource consumption" that the backup operation presents VMware introduced the concept of Virtual Consolidated Backups (VCB). Think of this as a "hardware snapshot" that allows the users to create a snapshot of their virtual disks and then expose it to a proxy server.

So by utilizing VCB technology one can move the resource consumption to a proxy server avoiding resource collisions on the production host; however this introduces a different set of challenges:

- The host will still copy the entire .VMDK file to the proxy server. Although the data can be moved over the SAN, it's still a lot of data to move around on a daily/weekly basis.

- Up until the release of vSphere v4.0 (April 2009), the whole .VMDK file would be moved, and the only option to back it up was to back up the entire file, increasing the size of the backup repository.

- Another option is to expose the virtual disk as a file system to the proxy server. The downside is that now the proxy server needs to scan all those "new" filesystems which makes it inadequate for large environments.

- Both methods lag behind when it comes to backup frequency, resulting in relatively poor Recovery Point Objectives (RPO), leaving a large amount of new data at risk between backup operations.

Positioning TSM FastBack in a virtual environment

As IT departments and datacenters continue to embrace virtual environments for their Windows operating systems (e.g. VMware and Hyper-V), solutions are needed to enable the fast and granular recovery of these unique environments. TSM FastBack may be installed directly on each guest operating system, allowing for full protection of the filesystems and applications of these individual, virtual environments. By virtue of its block-level, incremental design, FastBack does not burden the host server with significant I/O or processor overhead… even if many of the guest operating systems were performing backups simultaneously on the same host. This is in stark contrast to traditional, file-level backup products, which must often rely on special integration with the host to coordinate the backup and recovery.

The biggest concern that administrators have is resource utilization. The fact that one physical machine hosts multiple guest machines exacerbates this problem: if the backup is run from the guest machine and consumes 10% CPU and you have 5 machines running in parallel, that means that 50% of your PHYSICAL CPU is dedicated to the backup process.

When considering Tivoli Storage Manager FastBack (FastBack) for use within existing Tivoli Storage Manager (TSM) datacenters, it is important to remember that TSM will typically remain the primary backup and recovery solution for the virtual environment. TSM transparently integrates with VMware's VCB or allows for image level backups directly from the host console. FastBack offers a different approach that is suitable for a subset of the virtual servers.

For example, TSM FastBack is ideally suited for virtual servers running a Windows Server Operating System, with a large file system, where granular file-level recovery is desirable. Or for virtual servers host critical applications and cannot be down for more than a few minutes without causing pain to the organization.

Due to its block-level, progressive incremental design, TSM FastBack can supplement TSM in the datacenter in very significant ways. *Typical challenges in the datacenter include:*

- inadequate performance when backing up virtual machines using VCB

- virtual machines sprawl

- high resource utilization when using TSM (or other products) as "guest side" backups

- the need for more granular application recovery options than are currently available with TSM using VCB backups

Solutions for common recovery needs in a virtual environment are described below:

TSM FastBack's "guest side" approach avoids full-image backups and reduces the size of the storage required to store those backups. And since FastBack is a "guest based" solution it allows very aggressive RPO and even Continuous Data Protection (CDP). These technologies are not viable through the "host-based" approach.

The TSM FastBack Server can also run inside a virtual machine allowing the entire data protection solution to be virtualized. This might appeal to customers that are looking to virtualize the entire environment.

Large File Servers or Web Servers

Usually the physical servers that are handling these tasks are underutilized which makes them perfect candidates for virtualization; however, many of these servers manage large amounts of data or huge numbers of files (or both).

High numbers of files can cause most "file level" solutions to consume more resources and run longer, due to the need to scan through the whole file system looking for new and changed files. This leads to VM sprawl and affects other virtual machines running on the same host.

When coupled with VCB or Image Backups these high volume images create a huge burden on the storage space required for backup, so most users will default back to a weekly or monthly backup which is probably not adequate for their data protection and recovery needs.

TSM FastBack does not perform a filesystem scan as part of its backup operations. Therefore, FastBack may be used tactically on these Windows servers to provide block-level, incremental backup of these massive filesystems, while still allowing the administrator to perform file level recovery. The backup process will typically be performed in minutes (as opposed to hours with TSM and other file-level backup solutions), and the amount of data moved to the FastBack repository server will be even less than TSM's file-level backup. On the recovery side, the story gets even better. Instant recovery allows the administrator to make the fileserver immediately available to end users, while allowing the data to be recovered in the background.

And individual files may be recovered by mounting a snapshot of the filesystem, then dragging/dropping required files over to their recovery point.

Development and Test

The first servers that are virtualized are usually the development and test workload. FastBack can play a key part in their virtualization by utilizing its migration capabilities (see below); more importantly FastBack can be a real time saver to customers who perform recoveries of databases or applications in order to run tests, perform queries or validate functions. FastBack also allows the database administrator to quickly recover from a data corruption issue, test multiple snapshots of a database, and even expose the production databases to the development and test environments without the need for any storage for recovery (by utilizing the mount function).

VMotion Awareness

While FastBack is not "VMotion aware" it is definitely indifferent to VMotion's action. A virtual machine can float around from one server to another during a backup without interfering with the backup process. From that perspective the FastBack client is just another application running on that VM and it should not be interrupted by utilizing VMware's VMotion capabilities.

P2V & V2P Migrations

In addition, when FastBack is combined with FastBack for Bare Machine Recovery, it may be used for system recovery in a variety of ways… either for the initial consolidation of systems (Physical to Virtual), the recovery of a host (Virtual to Virtual), or the unique recovery of a virtual guest back to a dedicated system (Virtual to Physical). The flexibility of FastBack in supporting virtual environments is unsurpassed.

<u>Microsoft SQL, Oracle, DB2 databases (and other applications)</u>

First line database servers are considered to be resource hogs and consume high levels of IO, CPU, memory etc. and therefore are rarely virtualized. However, you'll find that the 2nd line database servers can and will be virtualized.

In this case, most experts agree that the "agent based" backup is still the only viable approach for the following reasons:

1. It allows for full application integration, letting the user select the quiescing methods and verifying that the integrity of the backup is intact.

2. It allows a better granularity of both backup and restore processes.

3. Image (or full VM) backups are too large.

Once again, TSM FastBack's block-level design will eliminate the need to perform regular full backups, while providing full database recovery. In addition, and more importantly… administrators (especially those performing nightly backups) are no longer limited to a recovery point of up to 24 hours ago. FastBack may be configured to capture block-level changes in MS SQL, Oracle or DB2 as often as the administrator wishes: hourly, several times an hour or even continuously. The administrator retains the ability to perform a single-step restore, and may choose their recovery from any "point in time" available on the FastBack server. Combined with the Instant Restore feature, FastBack delivers superior recovery points (RPO) and recovery times (RTO).

FastBack also enables the recovery of individual database objects directly from a mounted snapshot, whereas TSM for Databases must recover the entire database or volume. This technique may be extended to other applications such as Microsoft SharePoint, Lotus Domino, FileNet, etc.

Strength & Weaknesses:

Strengths

- The only "agent side" approach that provides volume- and file-level recovery while consuming minimal resources on the guest.

- Allows very aggressive RPOs: the more frequent the snapshots the less resources will be consumed (due to smaller incremental differences)

- Instant Recovery - system volumes can be recovered in minutes, no matter how big they are

Weaknesses

- No integration with VM management. FastBack does not display Virtual Machines in a different manner, nor does it integrate with VMware's virtual center or Microsoft's System Center, so there is no "auto discovery" of new virtual machines.

- No integration with other backup API available from the vendors (such as VCB)

Other Resources:

Tivoli Field Guide - Running TSM FastBack Server on a Virtual Machine (http://www-01.ibm.com/support/docview.wss?uid=swg27014144)
Deployment Guide Series: IBM Tivoli Storage Manager FastBack (http://www.redbooks.ibm.com/redpieces/abstracts/sg247685.html?Open)
IBM Tivoli Storage Manager (TSM) guest support for Virtual Machines and Virtualization
 (http://www-01.ibm.com/support/docview.wss?rs=663&context=SSGSG7&uid=swg21239546&loc=en_US&cs=utf-8&lang=en)