Executive Brief

Sponsored by VEEAM

Executive Brief on VMware Backup and Recovery: Challenges and Solutions

W. Curtis Preston ("Mr. Backup")

Introduction

Downtime from virtualized applications costs just as much money as downtime from physical servers, and yet most backup systems are unable to recover them quickly in the event of failure. Unfortunately, many attempts to make VMware systems better have done little more than add cost and complexity, and have not significantly enhanced recovery speed.

> *Full recovery of a virtual machine (VM) takes nearly 5 hours. Small improvement on the 6 hours required to recover a physical server.*[1]
>
> ***2010 VMware Data Protection Report***

To understand why this is the case, this paper needs to explain the challenges of a traditional backup and recovery system—especially when it comes to backing up VMware. The paper will also explain what this author believes to be the solution to these challenges.

Since solving this problem often requires an executive-level decision, this paper is written especially for the executive that must make that decision. Technical jargon will be kept to a minimum, and if a technical term is necessary, it will be explained. This paper will also explain each important concept as it relates to costs—something executives must keep a reign on. Look for the phrase Cost Factor.

The Status Quo

Typical backup and recovery systems perform their job by performing a nightly process of copying large amounts of data from the system being backed up (the backup client) to a server performing the backup (the backup server). Most customers perform a weekly full backup (which copies everything), followed by daily incremental backups (which copy all files that have changed on the client since the last backup). The full backup places a significant load on the storage system of the backup client and the backup server.

"The legacy backup approach simply does not scale well, and cannot handle the increased demands of protecting more data in less time, ideally with faster recovery."[2]

Dave Russell, Vice President, Gartner

The incremental backup creates a significant load as well, because it copies an entire file when any part that file has changed. Even if someone just changed the spelling of a single word, the entire file will get backed up. This nightly batch process typically takes several hours, and some customers are actually performing this process around the clock in order to finish it.

Cost Factor: Since this nightly process places a significant load on the systems being backed up—and system those same systems often have other jobs to do at the same time—this typical method of performing backup often requires companies to buy more expensive server with faster CPUs and more RAM in order to compensate for this load.

The backup system then stores multiple versions of these full and incremental backups to provide the ability to restore each server to multiple points in time. For example, if someone wanted to restore a server to how it looked just over two weeks ago, one would need to restore the full backup from three weeks ago, followed by restores from a few incremental backups.

The average environment will store about 25 GB on tape for every 1 GB they have on their servers.[3]

Deni Connor, Senior Analyst, Storage Strategies NOW

The challenge that this method of storing backups creates is that it requires a significant amount of storage.

Cost Factor: Storing 20 copies of data is not an inexpensive thing to do.

Another important thing to realize about today's backup and recovery systems is that they are very error prone. Whether it's the large amount of data that must be copied each night or the fact that most systems are still using tape as their primary storage mechanism, typical backup systems require a significant amount of attention to ensure that they are doing their job.

Cost Factor: This complexity requires companies to hire additional personnel to give the backup system the attention it requires.

Speaking of personnel, most backup experts agree that these same personnel are also the single largest cause of backup and restore failures. This phenomenon is caused by a number of factors, the first of which is that backup systems are complicated and customizable.

By 2013, at least 20% of organizations will have changed their primary backup vendor due to frustration over cost, complexity and/or capability, up from the typical single-digit percentage shifts today. [4]

Dave Russell, Vice President, Gartner

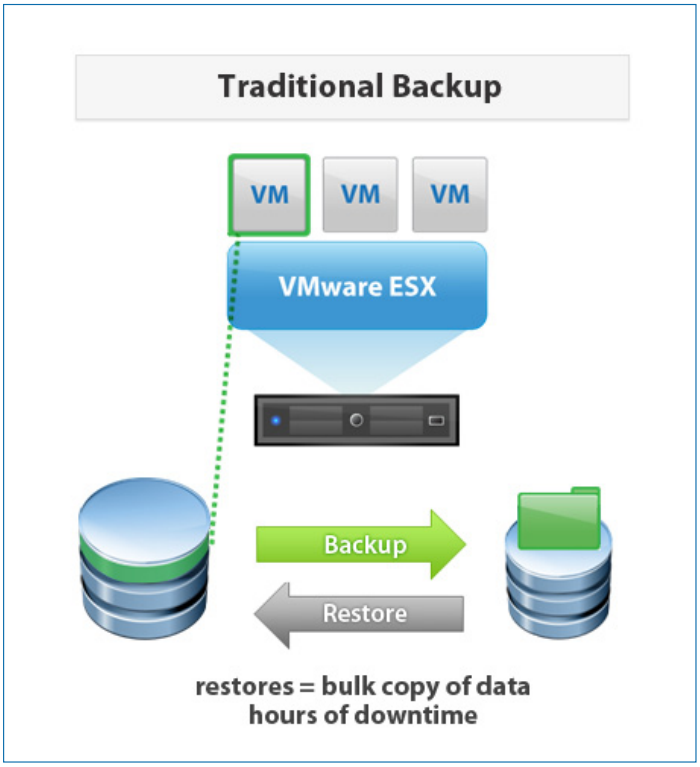
Unfortunately, it is very easy to customize them in such a way that backups or restores are actually less reliable than they were before the customizations. Another factor is that administering the backup system is considered to be a lowly or junior task, causing those responsible for it to seek positions elsewhere in the company as soon as possible. This creates a revolving door of personnel that results in the most junior people in the company administering the backup system.

Most backup experts agree that most of these restore failures could have been avoided with a proper backup configuration, but it's hard to have such a configuration when the most junior people in the company are configuring one of the most complex systems you have. Even if a restore goes well, it is a batch process like backup. It performs a bulk copy of data from the backup server to the system being restored.

Failed recoveries cost the average enterprise more than $400,000 every year. [5]

2010 VMware Data Protection Report

Cost Factor: While the cost of downtime is different for every company, it is a very real cost that must be calculated, and time spent restoring data from backups is time your company will never get back. Even successful restores contribute significantly to the total downtime of any outage requiring such a restore.



It's even worse if a restore doesn't go well, the entire process must be repeated, usually with a different version of the backup. While the first backup may have been stored on-site, it is very common for the "older" backup to be stored off-site at an offsite vaulting vendor (e.g. Iron Mountain). This means that those tapes must be recalled before the restore can even begin. A two-hour or four-hour response may sound acceptable when you're negotiating the price of an Iron Mountain contract, but when you're sitting there waiting to start a large restore, it feels like forever. Once the tape finally arrives, you can start the restore. Then you wait several more hours, hoping the entire time that this restore will work.

The amount of time necessary to perform a restore often creates a desire in system administrators to attempt multiple ways to fix things without having to perform a restore.

Companies lose $42,000 on average when a major application is down.[6]

Alinean ROI Report

They know that a restore is going to take hours, so they try to do everything they can without actually using the backup system. This again adds to the overall downtime of the system needing the restore.

Cost Factor: All of this downtime costs real money that can't be recovered. The loss of a supply chain management application costs $11,000 per minute, for example, while the loss of an electronic commerce platform costs $10,000 per minute.

The annual U.S. cost of data lost from PCs alone is $18.2 billion. The average cost of a data loss incident is $3,957. [7]

David Smith, PhD, Associate Dean of Academic Affairs and Associate Professor of Economics, Graziado Business Report from Pepperdine University

Assuming the system is finally recovered, the nightly nature of backup systems also means that there are gaps in the data. You usually lose any data that was created between the time the backup was created and the time the outage began. This lost data also has a cost element to it that can be difficult to quantify. It's one thing to ask personnel to redo their work from the last 24 hours; what if that work consisted of customer orders that were only recorded in the restored system? How do you re-enter that data if you don't know who the customers are?

Cost Factor: The cost of lost data is different with each company, but it is a very real cost that can be measured, and is often measured in millions.

VMware Backups

When looking at VMware, the problem is compacted due to the laws of physics. VMware allows one physical machine to pretend to be dozens of physical machines. It makes perfect sense for applications that do not require a constantly high level of computing resources—applications that have bursts of activity followed by long periods of relative inactivity. Such applications rarely place much stress on the storage system for any extended period of time.

Traditional backup and recovery applications, on the other hand, are completely different. This paper explained how they place a significant load on the computing and storage resources of the system being backed up. Combine the resource load of one VM's full and incremental backups with the load of dozens of other VM's on the same VMware server and you see why traditional backup applications are fundamentally incompatible with virtualized servers.

Cost Factor: **The nature of traditional backups in VMs creates the need to reduce the number of VMs that can be placed in a given ESX server, increasing the number of ESX servers that must be licensed.**

Another problem is the use of tape drives when backing up virtual servers. Modern tape drives are high-speed devices that must be sent data at a rate much higher than is possible for a VMware environment. For example, an LTO-5 tape drive is the latest iteration of the most popular tape drive in open systems, and the minimum speed at which it can reliably write data to tape is 40-80 MB/s. VMware VMs are not capable of creating a stream of data this fast.

Proposed Solutions

A number of attempts have been made to address one or more of all of the challenges mentioned above, but none of them actually addresses the core problem—the bulk copy nature of traditional backups and restores. The following are three proposed solutions to these challenges:

Disk staging

Using disk as the initial target for backups helps the tape performance problem mentioned above, but it doesn't fix the issues with restores. The bulk of restores still come from tape.

Deduplicated disk as a backup target

Deduplicating the backups does allow one to keep more backups on disk, making restores faster and more reliable. However, it still takes a significant amount of time to perform restores.

Image backup of VMware

Performing backups at the VMware level, using (VMware vStorage APIs and Changed-Block Tracking, is referred to as image backups.

> *The vStorage API provides the ability to revolutionize backups in the data center. The entire model of backups can now change from one of dealing with millions upon millions of files to dealing with a few hundred server images, all being backed up at a changed block level.* [8]
>
> ***George Crump, Lead Analyst, Storage Switzerland***

This does help the backups to run much faster, but—again—image backups do not solve the bulk restore problem. They also make backup configuration much more complex. Finally, restores can still take several hours to perform and may need to be repeated, costing hours of downtime.

A Ray of Hope

Two types of products that have addressed the challenges of both backup and recovery (and especially those of VM machines) are called continuous data protection (CDP) and near-continuous data protection (near-CDP). Backup problems are solved by removing the need for repeated fulls and by reducing the amount of data that incremental backups transfer as well. Restore problems are solved by removing the need for a restore altogether!

In place of the weekly full backup and nightly large incremental, CDP and near-CDP systems continuously copy new blocks of data from the backup client to the backup server. Where a typical backup system would transfer an entire file if you change the spelling of one word, a CDP or near-CDP system would copy just the

blocks of data that the spelling change modified. This block-level incremental nature of CDP and near-CDP is the main feature that removes the challenges this paper previously mentioned about backup. In addition, these systems use only disk as a backup medium, which removes all of the challenges associated with tape.

The real beauty of such systems, however, is what happens when a machine needs to be restored. As discussed previously, a typical backup system restores data by copying it from one format to another, and this process can take several hours or even days to perform. In contrast, a CDP or near-CDP system can immediately present backup data to the application that needs it in an activity we will call "instant restore." As difficult to believe as this may sound, a CDP or near-CDP system can literally provide a recovery time of zero seconds; the longest portion of the restore is the time taken to point the application to the recovery system.

In addition to instant restores, CDP and near-CDP systems can also significantly reduce the amount of data that is lost during a restore (referred to as the recovery point objective, or RPO). A typical backup system can only offer a 24-36 RPO, since the backup only happens once a day. CDP systems can offer RPOs measured in seconds, and near-CDP systems can offer RPOs measured in minutes. This difference in granularity also comes with a difference in cost; near-CDP systems tend to cost much less than their CDP counterparts.

Summary

The rules of backup and recovery need to change. We need to be able to back up incrementally throughout the day rather than via a typical nightly batch process.

In addition, we need to be able to restore servers in seconds rather than hours. CDP and near-CDP represent the best way to do all of the above, where near-CDP is the more affordable of the two. [10]

> *While backup/recovery solutions certainly have a great deal of room for improvement, it is also the case that most organizations have yet to fully embrace and deploy data protection approaches and techniques that have been available for years. Such concepts as incremental forever, synthetic backups and virtual full backups are now offered by several vendors, and are robust enough for production implementations.* [9]
>
> ***Dave Russell, Vice President, Gartner***

About the Author

W. Curtis Preston (a.k.a. "Mr. Backup"), executive editor and independent backup expert, has been singularly focused on data backup and recovery for more than 17 years. From starting as a backup admin at a $35 billion dollar credit card company to being one of the most sought-after consultants, writers and speakers in this space, it's hard to find someone more focused on recovering lost data. He is the webmaster of BackupCentral.com, the author of hundreds of articles, and the books "Backup and Recovery" and "Using SANs and NAS."

1 2010 VMware Data Protection Report. Veeam Software. October 2010. http:// www.veeam.com/survey

2 Russell, Dave. "Best Practices for Addressing the Broken State of Backup." Gartner. August 27, 2010.

3 Connor, Deni. Storage Strategies Now. Personal conversation. October 2010.

4 Russell, Dave. "Best Practices for Addressing the Broken State of Backup." Gartner. August 27, 2010.

5 2010 VMware Data Protection Report by Veeam Software. October 2010 http://www.veeam.com/survey

6 "ROI for the Automated Enterprise." The Alinean ROI Report. January 2004. http://www.alinean.com/Newsletters/2004/2004-1-Jan.asp

7 Smith, David. "The Cost of Lost Data." David. Graziado Business Report. 2003, Volume 6, Issue 3. http://gbr.pepperdine.edu/033/dataloss.html

8 Crump, George. "vStorage API Spreads Its Wings" InformationWeek's Storage Blog. September 2, 2010. http://www.informationweek.com/blog/main/archives/2010/09/vstorage_api_sp.htm

9 Russell, Dave. "Best Practices for Addressing the Broken State of Backup." Gartner. August 27, 2010.

About Veeam Software

Veeam Software, a premier-level VMware Technology Alliance Partner, develops innovative software to manage VMware vSphere. Veeam vPower™ provides advanced Virtualization-Powered Data Protection™ and is the underlying technology in Veeam Backup & Replication™, the #1 VMware backup solution. Veeam ONE™ provides a single solution to optimize the performance, configuration and utilization of VMware environments and includes: Veeam Reporter™—VMware capacity planning, change management, and reporting and chargeback; Veeam Business View™—VMware business service management and categorization; and a choice of VMware monitoring options including the nworks Management Pack™—VMware management in Microsoft System Center, the nworks Smart Plug-in™—VMware management in HP Operations Manager, and Veeam Monitor™—framework-independent VMware monitoring. Learn more about Veeam Software by visiting www.veeam.com.