Optimizing Backup and Data Protection in Virtualized Environments

5 Rules for Better VMware Backups

Introduction: Moving to VMware Server Virtualization Brings a Variety of Challenges

The promise of maximizing IT investments while minimizing complexity has resulted in widespread adoption of server virtualization. In fact, VMware users are now deploying more than half of their new servers as virtual servers, rather than as physical servers. That said, rapid adoption of server virtualization has brought with it a variety of challenges in backup and data protection, and many organizations still have a long way to go when it comes to successful backup and recovery of virtual servers:

- It still takes 5 hours on average to recover a virtual machine.
- 59% of organizations still use tools designed for physical servers to perform their virtual server backups.
- 63% of organizations experience problems every month when attempting to recover a server, (physical or virtual servers).

In addition to these challenges, VMware server virtualization often leads to large increases in storage utilization, as moving from physical to virtual servers typically results in a larger number of combined physical and virtual servers in the environment. This can occur because the cost of generating a new virtual server is now much lower than was the cost of procuring an additional physical server. This in turn increases the amount of primary storage needed to support the additional servers, and increases the amount of secondary storage needed to back up these additional servers. As storage needs increase with greater numbers of virtual servers in the environment, the need for better visibility into the storage utilization of individual virtual machines also becomes important.

Another challenge brought about by server virtualization is related to the way traditional physical server backups have typically been performed. IT administrators have traditionally backed up their physical servers with backup agents on each server, with the backup agent installed on top of the operating system (OS) running on the physical server. Each backup agent had its own physical server on which to run, and could access all the resources belonging to that physical server without undue concern for over-utilizing the physical server.

However, this old paradigm of using a single backup application agent on each physical server can break down with large numbers of virtual machines running on a smaller set of physical servers. With the adoption of server virtualization, where multiple guest operating systems now run on a single physical server, the traditional one-backup-agent-per-OS approach presents problems. Multiple guest OS's may now require multiple backup agents to be run on a single physical server, potentially taxing the resources of the physical server and therefore the resources available to all of the virtual machines and all of the applications running on those virtual

machines, on that single physical server. Additionally, as multiple virtual machine backups with potentially overlapping backup windows are sent to one or more backup targets, landing all of those backups so that they can be protected as soon as possible is an increasing concern.

Server virtualization also brings advantages in creating second-site server backups, since all that is required for a virtual server backup is a copy of the virtual server image. However, managing these virtual server copies in a way such that the virtual servers can be retrieved and brought online as fast as possible can be a challenge.

So how can an IT administrator best take advantage of the benefits of server virtualization without getting mired in the issues of increased storage utilization, longer backup windows, and scaling up in the face of an exploding amount of data to manage? Below we discuss several ways to manage VMware backups while mitigating these issues.

Five Rules for Better VMware Backups

Rule #1: Deduplicate Your Data

While server virtualization does result in additional backup and data protection challenges, many of these challenges can be mitigated via a disk-based backup solution with data deduplication. Because multiple backups of a given virtual machine are likely to contain highly redundant data, virtual machine backups lend themselves particularly well to data deduplication. With data deduplication, virtual machine backups are compressed and deduplicated, yielding ratios as high as 1000 to 1, and allowing the customer tremendous savings in backup storage.

Data deduplication also makes second-site protection more efficient, as only changed data is transmitted across the WAN from the primary site to the second site. Because the amount of data that changes in a VMware virtual server is generally minimal, keeping a second-site copy of a VMware virtual server up-to-date can be done easily and efficiently.

Rule #2: Back up Fast

Data deduplication is by nature a processor-intensive activity, and how a vendor implements this can have a definite impact on the length of a customer's backup window.

With in-line data deduplication, which some vendors have implemented, data deduplication occurs as data is flowing into the system, before the data is written to disk. Because data is being processed and deduplicated as it makes its way to disk, performing VMware backups using an in-line approach can result in slower backups

and longer backup windows, potentially putting your VMware virtual server backups at risk.

Post-process deduplication, by contrast, allows the data to land to disk first, so that the backup can occur as fast as possible. Deduplication is then performed after the VMware virtual server backup has safely landed to disk, without interfering with the backup window.

The net result is that post-process deduplication yields high performance backups with the shortest backup windows. This means that your virtual servers can be protected faster so that there is minimal intrusion into your normal IT operations.

Rule #3: Restore Fast

While restoring VMware virtual servers is typically not something that needs to be done on a daily basis, it is important to make sure that you can get back those backups as quickly as possible when you do need them. Some approaches to deduplication are far better than others at achieving this goal, depending on how the deduplication is performed.

In a post-process deduplication approach, the most recent virtual server backup can be kept in its entirety, ready to be restored as quickly as possible. This gives you the ability to rapidly restore a VMware virtual server so that you can get that server back up and running right away. This stands in contrast to in-line approaches, where, if you need to restore a VMware virtual machine that has just been backed up, the contents of that virtual machine would need to get re-assembled through the deduplication algorithm before it could be restored. This is a much more time consuming approach that will lead to much longer restore times.

The ability to quickly restore a VMware virtual server from a second-site location is also critical, in the event that the primary site becomes inoperable and the VMware virtual servers that are backed up at the second site need to be brought on-line. Maintaining a complete copy of the most recent VMware virtual server backup on the second site as well, so that when needed, the second site copy can also be restored as quickly as possible, is key in rapidly recovering from a primary site outage or disaster.

Rule #4: Don't Scale with Disk Alone

One thing you can count on is that your VMware virtual server data will continue to grow over time – that is just a fact. And while deduplication can mitigate the impact of all that data growth, it is inevitable that the solution you adopt will need to grow as your data grows. Therefore, an important area to look at when evaluating how to best back up your VMware virtual servers is scalability. The most scalable solutions

can grow along with the environment in which they are working, in a manner that is less disruptive, while maintaining performance characteristics over time.

In some implementations, however, the customer installs an appliance that is sized and specified for their environment, but as the amount of data grows, the solution scales up either by only adding additional storage capacity, or by replacing the appliance with a larger, higher performance unit. This approach to scalability can create potential problems. Simply adding storage to an existing unit, for example, means that a greater amount of data is being managed by the same amount of processing, memory, and bandwidth. This can result in slower backups and longer backup windows. On the other hand, swapping out a lower-performing appliance for a higher-performing one, to better deal with the greater amount of data, is disruptive to the backup environment.

Better approaches to scalability allow the system to maintain all the elements needed for performance as the amount of data grows. Grid-based approaches combine multiple servers that each contain processing power, memory, and bandwidth, in addition to disk. This type of approach allows you to add additional disk when needed, combined with additional processing power, memory, and bandwidth, so that you can maintain your short backup windows over time, without the need for wholesale upgrades to your environment.

Rule #5: Improve RTO with Technologies like Veeam's Instant VM Recovery

With technologies such as Veeam's Instant VM Recovery, dramatic improvements in Recovery Time Objective (RTO) are possible, as VMware virtual machines can be run directly from a backup appliance itself, even before the VMware backup has been restored to primary storage. This is useful in the event that the primary virtual machine becomes unavailable, and there is an immediate need to get that virtual machine up and running again. Once the primary storage environment has been brought back to a working state, the virtual machine running on the backup appliance can then be restored to primary storage for continued operation.

In addition, when combined with a backup appliance using a post-processing deduplication approach, technologies like Instant VM Recovery become even more efficient and valuable. This is the case because it is much faster to get the most recent VMware backup running and operational when it is being restored from an un-deduplicated version of that backup, as is the case when using post-process deduplication, instead of having to piece that backup together as occurs with inline deduplication.

The ExaGrid Advantage

The current range of backup and data protection issues facing organizations that have moved to a VMware virtual server environment are numerous. With the right approach, however, companies can enjoy the gains achieved with VMware server virtualization and gain better control of their backup environment.

The ExaGrid system greatly reduces the amount of storage needed for your backups through data deduplication. ExaGrid's post-processing approach provides the fastest possible backups and restores, and extends this capability to second sites via Instant Disaster Recovery – the ability to rapidly restore your most recent backup on the second site as well. ExaGrid's backup aware reporting provides insight into how much data your backup jobs are using, how well they are being deduplicated, and how quickly they are being replicated to a second site. Finally, ExaGrid's grid-based architecture allows the system to scale easily and with minimal disruption to your environment. In the next section, we discuss various ways you can execute VMware virtual server backups with an ExaGrid system.

VMware Backup Methods with ExaGrid

The challenges of backing up servers in a virtualized environment opened the door for the adoption of many new backup methods over the traditional backup apps that dominate the market. Some IT administrators devised their own script-based and other home-grown methods for backing up virtual machines, and various third-party vendors emerged by productizing these types of approaches. In the meantime, the traditional backup application vendors reacted by improving their VMware integration in an attempt to make VMware backups easier with their products. The result is that today there are now numerous methods for performing virtual machine backups. IT customers that have adopted server virtualization often have multiple backup procedures in place within the same data center – procedures to handle physical machine backups, and procedures to handle virtual machine backups.

Regardless of which backup methods are used, and regardless of whether a single backup method or multiple backup methods are employed, ExaGrid gives the customer a single backup system to which both physical and virtual machine backups can be targeted, while allowing the customer the flexibility to choose among the various VMware backup options that are available.

VMware backup methods supported by ExaGrid include the following:

Support for backup application agents running within virtual machines

This method, where a backup application agent is installed in each virtual server, is carried over from the way that physical servers are ordinarily backed up. A backup

agent is installed on each VMware virtual server, even if there are multiple virtual servers installed on a single physical server. These backup agents are managed by the backup server, while the ExaGrid system sits behind the backup server, as a target for the backups. Virtual server and physical server backups are essentially treated the same way. This method is often recommended by backup application vendors for mission critical applications where a backup agent with application-specific optimization is needed.

Support for backup agents running with VMware Consolidated Backup (VCB)

VMware Consolidated Backup (VCB) is a backup application extension typically used to facilitate VMware backups with traditional backup application agents. While VMware has updated this technology with vStorage API (below), it is still in use by some older versions of various backup applications. VCB allows a backup agent to run "off-host," or off of the VMware ESX Server itself, on what is called a "proxy server". The backup agent uses VCB to run a script of VMware command lines, on the ESX server, to quiesce a given virtual server to be backed up and then creates a snapshot of that virtual server. The virtual server snapshot is then copied to the proxy server and made available to the backup application (typically running on the proxy server as well). The backup application then performs a backup of the virtual server from the proxy server to the ExaGrid system (the backup target).

Support for backup agents running with VMware vStorage APIs for Data Protection (VADP)

VMware vStorage APIs for Data Protection (VADP) is the newest VMware technology for integrating VMware backups with third-party backup applications. Like VCB, VADP also allows backups to run "off-host," or off of the VMware server itself. It is simpler to use VADP instead of VCB, however, in that the backup application extensions that were used with VCB are now integrated within the third-party backup agent itself – mitigating the need for these additional extensions and the command line scripts that were often required. In addition, with VADP, virtual server backups no longer have to be copied to a proxy server first; backups can instead be copied directly from the source to the target (ExaGrid) appliance. ExaGrid also supports vendor implementations of VADP's change block tracking, allowing for incremental virtual machine backups.

Support for direct backups of VMware virtual machines (.vmdk files)

This method allows users who do not wish to use a traditional backup application the ability to copy virtual server backups directly to the ExaGrid system. The user simply copies the .vmdk and other accessory files that correspond to a given virtual machine, directly to the ExaGrid system.

Conclusion

The return on investment of server virtualization is unquestioned, but the impact this has on backups must be addressed, as there are a variety of backup and data protection issues facing IT administrators that have moved to a virtualized environment. However, with the right disk-based backup approach, IT administrators can enjoy the gains achieved with server virtualization and gain better control of their backup environment. ExaGrid's approach works extremely well for IT administrators in this position.

.

About ExaGrid

ExaGrid offers the only disk-based backup appliance with data deduplication purpose-built for backup that leverages a unique architecture optimized for performance, scalability and price. The product was named “Product of the Year” for Backup and Recovery Hardware in 2010 by Storage magazine-SearchStorage.com

ExaGrid’s unique combination of post-process deduplication, most recent backup cache, and GRID scalability enables IT departments to achieve the shortest backup window and the fastest, most reliable restores, tape copy, and disaster recovery without performance degradation or forklift upgrades as data grows.

With offices and distribution worldwide, ExaGrid has more than 3,500 systems installed and more than 250 published customer success stories and testimonial videos available at www.exagrid.com.

ExaGrid Systems, Inc | 2000 West Park Drive | Westborough, MA 01581 | 1-800-868-6985 | www.exagrid.com