

STORAGE SWITZERLAND

DESIGNING DISK FOR VMWARE AND HYPER-V BACKUPS



George Crump, Senior Analyst

One of the key advantages of a virtualized environment is how well it can be protected. VMware and Hyper-V specific applications like Veeam provide efficient changed block and deduplicated backups, as well as flexible in place recoveries. A challenge remains, however, in selecting the backup storage for these environments. Designing the right backup target is critical to realizing the maximum benefit from VM specific backup applications.

The Unique Requirements of VM Backup

A major dilemma is properly sizing a disk backup infrastructure to support VM data protection. While change block transfers improve network efficiency and deduplication reduces the physical storage required, the VMware and Hyper-V environments will in all likelihood, grow at an unpredictable pace. Furthermore, all the features that comprise a VM specific backup application, like off-host snapshots, replication and data in place recoveries, require additional disk capacity. As a result, the ability to incrementally scale the disk backup infrastructure is critical.

Without the ability to incrementally scale the disk storage back end the back infrastructure either becomes more complex or too costly. The most common of which are

either the over provisioning for storage which means buying too much storage too soon or the under provisioning of storage which means buying too little storage and leads to disruptive forklift upgrades.

The second demand, and one that often takes IT Planners off guard, is the need for random I/O backup performance. Unlike traditional backup, where most of the data movement is large block sequential transfers, VM specific backups are characterized by very randomized I/O. Transfers are the unique blocks of each VM and blocks can be originating from dozens of VMs at a time from multiple hosts. The end result is very small block transfers arriving in a very random pattern. It may not be noticeable in the early days, but as the volume of backups grow, this performance hit will become severe if the performance of the target can not scale with the volume.

The third demand is high performance restoration. Just as in traditional recoveries, when a virtual machine needs to be restored, it has to be completed as quickly as possible. But products like Veeam add a new requirement to recovery due to their popular “in place recovery” feature. Now what was once backup storage can temporarily become primary storage. An earlier decision to use low cost, low performing disk can come back to hurt you right at the worst possible time.

A fourth demand is reliability. Unlike traditional disk backup that supports physical environments, products like Veeam rely on the disk storage system for more than just making an initial copy of the backup. In most cases, the backup storage is the only copy of data. Therefore high reliability is of paramount importance and the assurance of zero data loss is absolutely critical.

The fifth and final demand is cost. Backup is now more than an insurance policy because applications like Veeam allow you to use the backup data for more than just recovery, however, the backup infrastructure is still a secondary copy of data and it must be cost effective. In other words, it is not practical to use a primary storage system for secondary data. What is needed is a cost effective storage platform that can meet backup needs almost as well as primary storage.

Meeting the Demands of VM Backup

To meet the first demand, accurate sizing, requires a storage system that can start at relatively small capacities and then grow into a system that can support very large storage capacities. It is important that the storage system can scale capacity within a single platform so that the management of the backup target does not become too time consuming.

Wide range scaling means that the traditional scale up storage architectures, where all capacity is scaled behind a single set of controllers, is impractical. These systems will either cost too much upfront or not scale large enough to meet backup growth demands. These systems also don't have the ability to scale network bandwidth as the size of the backup grows. This naturally leads many VM backup customers toward scale out storage systems for backup storage.

The challenge with using scale out storage systems for backup storage is that these systems are best suited to streaming I/O and often require high performance backplane networks to deliver optimum performance. Also the scale out backup devices are typically laden with features that are either redundant to the capabilities of the VM backup solution or the hypervisor itself. Finally some of these systems data protection techniques use replicas that may require 3X the raw backup capacity.

Companies like [Gridstore](#) have addressed this demand through a software defined storage model that virtualizes the storage stack and uses a grid based architecture to scale capacity, I/O bandwidth and controller processing. By deploying virtual controllers to each host, each virtual controller can optimize independently to absorb and offload the random I/O produced from many backup jobs. On the capacity side, while similar to scale out storage, a grid is able to provide capacity per node at a price point that is ideal for the typical midrange virtualized data center without requiring a backplane network or 3-way replica for reliability. A grid provides the ability to “scale small” initially and then to add nodes as capacity demands increase. Also, since these are modern era storage systems, they do not include features that VMware/Hyper-V or the VM backup application already provides; thus obviating the need to pay for the same feature twice.

These systems also have the ability to stripe data across their nodes. This increases aggregate performance while eliminating the need for the use of replicas.

The second and third demands of high performance backup and restoration again, require a storage system whose compute and network bandwidth capabilities can scale along with capacity. In scale out and grid architectures, these two resources automatically scale as nodes are added. A grid architecture, however, also leverages the compute resources of attaching hosts - in this case the backup servers. As stated earlier, this type of architecture is ideal for the random I/O; small block transfer type of backups that changed block tracking creates.

This type of system also provides an excellent platform to be “temporary” primary storage when an in-place recovery is performed. They can provide more than adequate performance until the VM can be safely migrated back to primary storage.

A grid architecture like Gridstore’s also addresses the fourth demand of a VM backup device, reliability. Grid volumes can be configured for a variety of high availability options which is different than a scale out architecture that typically only offers one level. The grid system can be configured to continue to provide data access if one or more nodes fails. In fact, that setting is user selectable and can be changed on the fly.

Since this is backup data, most environments will be fine with a one or two node level of redundancy. Ensuring there is no single point of failure in your back target is critical. If the backup device is suddenly pressed into service as temporary primary storage because of multiple in-place recoveries, then it is extremely valuable if the redundancy level can be temporarily raised until the VMs can be restored back to the primary storage device.

For scale out storage systems, one of the biggest challenges is cost, especially in the backup environment. The cost of replicating three times what is already replicated data (your backup) as well as requiring a high performance backplane across to replicate the data is wasteful and expensive.

As an example, Gridstore, with its grid based architecture, avoids the potential cost problems of scale out storage systems by leveraging the parallel compute of the connecting backup servers. Through its vControllers, data is encoded before it leaves the source and is written in parallel chunks across the storage nodes. This eliminates the cost of a backplane network and the overhead 3-way replicas. This combination provides the same or higher levels of data protection while ensuring the highest possible throughputs due to the direct parallel I/O.

Conclusion

VM specific backup brings a unique set of new requirements and demands on storage targets, such as having to deal with small block random I/O and potentially being used as a primary storage if a host fails. It also shares the scalability and affordability demands that any backup process requires of its storage system. A grid architecture may be uniquely positioned to address these demands and can be the perfect compliment to a VM specific backup product like Veeam.

About Storage Switzerland

Storage Switzerland is an analyst firm focused on the virtualization and storage marketplaces. For more information please visit our web site: <http://www.storage-switzerland.com>

Copyright © 2013 Storage Switzerland, Inc. - All rights reserved