The Power of Deduplication-Enabled Per-VM Data Protection

SimpliVity's OmniCube Aligns VM and Data Management

DeepStorage.net

Prepared for
SimpliVity

Contents

The Bottom Line 1
Introduction 2
Per LUN Problems 2
- Can't I use Hypervisor Snapshots 3
- Snapshots as Metadata 3
- The Next Step Virtualization Optimized Storage 4

Introducting SimpliVity's OmniCube 4
SimpliVity's Per VM Protection 5
- Offsite Replication 6
- VM Migration 7
- VAAI NFS Integration 7
- Conclusions 7

Appendix - OmniCube Models 9
About DeepStorage 10

The Bottom Line

Many organizations have sufficiently virtualized their compute workloads that the virtual server has become their default unit of management. Their attempts to manage, and of course protect, their assets on a per VM basis, giving each the level of protection it deserves, has been thwarted by a storage infrastructure that provides its services to an entire volume.

A group of new solutions led by SimpliVity's OmniCube, a hyper-converged platform for virtual environments that combines storage and compute resources with a set of advanced management and data protection capabilities all of which operate on individual VMs, can now deliver all the data management and protection services of a SAN array to individual virtual machines.

OmniCube's high performance, metadata based data management on a per VM basis has several significant advantages:

- *Global deduplication*
  - *Provides significant capacity and performance efficiencies on local backup, and bandwidth savings with replication for DR*
- *Application consistent snapshots/backups without waste*
  - *Protected against node failure*
- *Offsite replication for critical VMs*
  - *Including cloud backup*
- *Fast template and VM cloning*
  - *Especially for VDI*

Introduction

While it has simplified our lives and saved us money, in many ways server, and to a lesser extent desktop, virtualization has made data protection more complicated. A new generation of virtualization optimized storage systems, some of them software defined, are addressing these complexities by managing their data protection features on a virtual machine by virtual machine, rather than volume by volume, basis.

Back when a logical volume on your SAN array held the data from a single server, snapshots were a great data protection feature. A snapshot provides a frozen image of a server's disk(s) you can revert to in the event of a system problem or extract data from when users call in a panic because they overwrote last month's closing spreadsheet with this month's. Even better you could make those snapshots application consistent through Windows VSS or Linux scripts that quiesce your database engine forcing it to flush its buffers to make the data on the disk internally consistent.

The array could also replicate the data for a volume, which belonged to a single server, to your disaster recovery site. The frequency of the replication was based on the importance of that server's data.

As we virtualized our servers, we also pooled their data together putting five, ten or more virtual machines in a single datastore that resided on a single SAN volume or LUN. Really, we had no choice. As the number of virtual servers grew, individual LUNs for each of them became a management nightmare. Even if we were willing to have thousands of LUNs, vSphere only supports 255 LUNs in a cluster.

Per LUN Problems

Once we merged all our data into datastores, we lost the ability to have the storage system protect the data from a single server. SAN arrays, and many NAS appliances, can only manage storage at the volume level. When your SAN takes a snapshot of one VM, it saves all the changes all the other VMs in that datastore wrote to their disks as well. This means you'll use a lot of storage capacity storing snapshots of servers that may have little value.

Then there's the consistency problem. The consistency problem is that you can't practically create application consistent snapshots of multiple servers at the same time. It's just too hard to get multiple applications across multiple server to all quiesce at the same time and then take the snapshot. You can make sure you only install one application that needs consistent snapshots in each datastore, or you can take multiple snapshots where one application is consistent in each snap.

This data inflation effect, where the disk updates for all the other VMs in a datastore have to be included with the critical application's data, also significantly boosts the amount of data that has to be replicated to a disaster recovery site increasing the amount of expensive bandwidth required.

Can't I use Hypervisor Snapshots?

Hypervisor snapshots are a poor substitute for their storage system cousins. Hypervisor snapshots, which use redo logs to track the changes between a snapshot and its parent, are only really useful for specific situations; like serving as a backup source where the snapshot only exists for a limited period of time. The longer a snapshot exists the larger its logs get and the slower access to that .VMDK gets.

Hypervisor snapshots are a poor substitute for their storage system cousins

The performance penalty is even worse when working with volumes that have multiple generations of snapshots. Even deleting a snapshot causes a lot of disk IO as the hypervisor posts the updates to the parent file.

Snapshots as Metadata

Over the years, storage vendors have used several different snapshot mechanisms. Old-fashioned Copy-On-Write snapshots slow things down by turning every write IO into three operations to read the old data, write the old data to the snapshot, and finally write the new data.

The old storage systems that used COW snapshots were originally designed for spinning disks. Disk-based systems try to keep data that's logically adjacent physically adjacent on the disk so it can be read sequentially.

Modern storage systems, which use flash memory for some or all of their storage, process random IO as quickly as they do sequential. Consequently, there's no advantage to keeping logically adjacent data physically adjacent. To get the best performance and endurance from their flash, these systems virtualize their storage. They use pointers and other metadata to track which physical block holds each logical block.

In this type of storage system a volume is, in reality, just a set of metadata that describes which logical data blocks make up the volume. A snapshot is a copy of that metadata. Since a typical volume's metadata is many times smaller than the data itself, making a copy is quick and has almost no performance impact.

When each snapshot is created, it is an independent copy of the metadata, which brings several advantages. Unlike with COW and log-based snapshots, there's no chain or tree of snapshots that needs to be referenced or updated when applications access the volume. This means administrators can create multiple snapshots and retain snapshots for extended periods without worrying about any resulting performance.

Since keeping multiple snapshots doesn't affect storage performance, organizations can use snapshots as their first line of defense satisfying the daily restore demand from snapshots.

Metadata copies can also be used as read-write replicas. With read write replicas, the test and development groups can work against a full, and fully independent, version of the production database.

The Next Step Virtualization Optimized Storage

While metadata based snapshots are great, in virtualized environments they have all the same problems as any other volume level snapshot – insufficient granularity. Taking a snapshot of a SAN volume to protect a single VM is like swatting flies with a bulldozer, overkill and likely to create some collateral damage in the form of rapid storage space consumption.

Taking a snapshot of a SAN volume to protect a single VM is like swatting flies with a bulldozer

Recently, engineers have developed storage systems that are optimized for use in virtual environments. Rather than serving up dumb LUNs to the hypervisor, these systems manage their own file systems providing access to the hypervisor via a file sharing protocol like SMB 3.0 or NFS. This gives them enough context to manage their data protection features, including snapshots, one virtual disk or virtual machine at a time.

Introducing SimpliVity's OmniCube

SimpliVity's OmniCube is a 2u hyper-converged system that provides both scale-out storage and compute resources to a VMware vSphere cluster. There are three models of OmniCube each with Dual Xeon processors and both solid state and spinning disk drives. Readers will find their specs at the end of this report.

When connected together, and running VMware's vSphere, a number of OmniCubes form a federation. In addition to the ESXi hypervisor, each OmniCube also runs the OmniCube virtual controller, a virtual machine that manages the nodes' hard and solid state disk.

Within the federation the virtual controllers pool the storage resources, in what Simplivity has dubbed the OmniStack architecture. Within each OmniCube, data is parity protected when on SSDs and double parity protected when on spinning disks. All data is also replicated to a second OmniCube within the data center so an OmniCube federation can suffer a node failure and multiple disk failures without losing data.

The virtual controllers in all OmniCubes in a data center pool their storage to create a single distributed file system through what SimpliVity calls the Virtual Resource Assimilator. vSphere datastores on the OmniCube file system are shared via NFS. The capacity of the entire pool can be presented as a single datastore or multiple datastores that can be resized on the fly.

As data is written to the OmniCube file system, it's deduplicated inline at a very fine grain of 4-8KB, compressed, serialized and written to the storage pool. In addition to the obvious saving of storage capacity efficient data reduction like SimpliVity's brings additional advantages to a storage system. First it increases the efficiency of the SSDs increasing the percentage of storage IOs that are satisfied from SSD and therefore application performance.

Data reduction also reduces the amount of data the system has to transmit to replicate a virtual machine, or set of virtual machines, to another location reducing the amount of expensive WAN bandwidth needed to keep a remote site up to date. In addition to reducing the data locally the OmniStack architecture SimpliVity's implemented global deduplication further reducing replication traffic.

The whole OmniCube federation, including the virtual controllers, is managed through a vCenter plugin to make it easy for VMware administrators to use automating tasks like mounting datastores on hosts. The federation model allows administrators to add or remove nodes from the federation in just a few clicks with no disruption to the workloads.

Most providers of hyper-converged infrastructure systems want their solutions to stand alone. Regardless of whether it's as a total infrastructure or an isolated cluster running a demanding workload like VDI, the hyper-converged system provides all the storage and compute resources for the cluster.

SimpliVity recognizes that some users may need to add more compute resources to their cluster. Rather than requiring them to buy more storage to get it they support using an OmniCube federation as the storage resource for compute servers from the major vendors. VMs running on Dell, HP and IBM servers can still take advantage of SimpliVity's advanced storage functionality.

SimpliVity's Per VM Protection

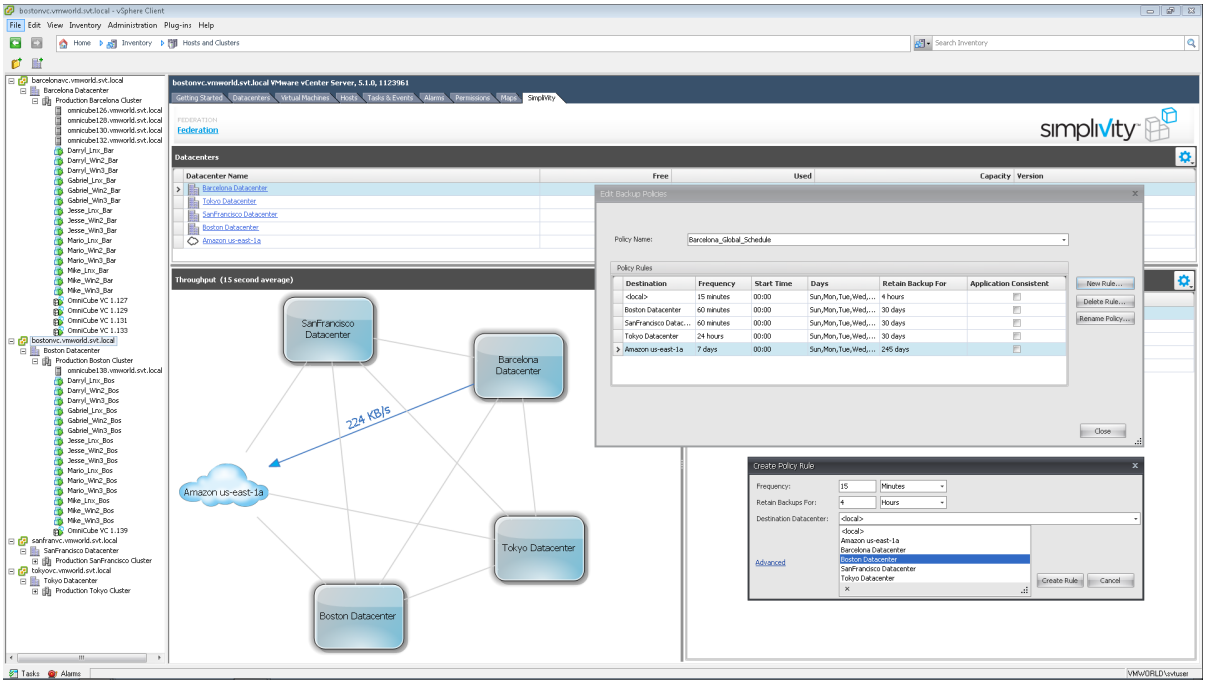
SimpliVity refers to their metadata-based snapshots as backups. We at DeepStorage would ordinarily cry foul at calling a snapshot a backup because conventional snapshots are dependent on the primary data. Since in the OmniCube file system the data is replicated and no data is lost in the event of a node failure, we'll allow it.

The key to data protection in the OmniCube system is backup policies. A backup policy is made up of a set of rules, each of which defines how frequently the OmniCube makes a backup, where that backup should be stored, and how long it should be retained. A simple policy might call for daily backups with 30 day retention and weekly backups with a retention period of a year.

Where conventional backup software uses a tape drive or a disk array as its destination OmniCube backups are sent to an OmniCube cluster in another data center. When creating a backup policy, the user simply uses the drop down menu to select the backup destination datacenter. Since the OmniCubes are part of the same federation, and the data center is defined in vCenter administrators can skip the step of defining a backup destination and it's configuration details like IP addresses, LUNs, and shares or mount points.

Administrators can define a basic protection policy that creates a backup once a day and stores it in the local cluster, then assign that policy to be the default policy for their datastores insuring that VMs can't completely fall through the cracks and be unprotected. Since all the storage operations are performed VM by VM, virtual machines running critical applications can be assigned a more aggressive backup

policy that creates more frequent backups, or replicates the data offsite. That offsite replication can be to an OmniCube in another data center or even to Amazon Web Services.



Defining an OmniCube Backup Policy

If the Application Consistent option is selected, the system will quiesce the source server through VMware Tools, which acts as a VSS provider on Windows systems and invokes scripts for Linux hosts.

Remember that the system deduplicates across all the OmniCube systems in a datacenter minimizing the amount of space the backups take up. Regardless of how many Windows Server 2008 R2 virtual machines you backup, there's always just two copies of that operating system stored in the data center regardless of how many servers you're backing up or how many backups you retain.

Offsite Replication

If your federation of OmniCubes extends over more than one datacenter, you can specify that the backup be stored on the file system of the OmniCube cluster in another datacenter. The snapshot will be taken locally and then replicated to the OmniCube cluster in the other datacenter.

SimpliVity's data deduplication extends across the replication link sending only those fine grained chunks that don't exist at the remote site reducing the need for expensive bandwidth. If Windows Server 2008 exists in the OmniCubes in San Francisco, when you send the first backup of a Windows Server 2008 VM the common operating system data won't be sent over the link. Before actually sending data, the source and destination OmniCubes negotiate which data blocks need to be sent.

OmniCubes can also use Amazon Web Services as their destination. Virtual machines running the OmniStack software receive the deduplicated compressed replication traffic and store the data at AWS. Organizations without multiple OmniCube equipped data centers can use AWS storage as an offsite backup location, while larger organizations may find AWS a good place for their long retention just-in-case machine images.

> SimpliVity's data deduplication extends across the replication link sending only those fine grained chunks that don't exist at the remote site reducing the need for expensive bandwidth

The only things keeping us from saying that OmniCube eliminates the need for conventional backups are the lack of a catalog or an easy way to retrieve individual files from OmniStack backups. Restoring a lost file today requires the administrator to restore a snapshot as a new VM and spin it up and copy the file out.

VM Migration

The SimpliVity vCenter plug-in makes it easy to migrate or copy virtual machines from one data center to another. These migrations use the OmniCube's deduplicated replication mechanism to use minimal bandwidth. In addition to copying the virtual server's data, these functions also add it to the system's inventory. We can see organizations that centrally develop their VM templates and then distribute them to remote data centers will really like this feature.

VAAI NFS Integration

With the latest version of their OmniStack software, SimpliVity has added support for VMware's VAAI (vSphere API for Array Integration) for NFS. Taking full advantage of the fast file clone primitive means that whenever vSphere wants to make a copy of a VM, or .VMDK file, it offloads that process to the OmniCube file system which just makes a metadata replica and lets vSphere use it as if it were a full copy.

This operation is so fast that in many cases, especially when creating multiple clones as when provisioning VDI images, the storage system creates clones faster than the vCenter server can add them to inventory and update its database.

Conclusions

While virtualization has allowed data center operators to consolidate workloads and simplified server management it has complicated the process of data protection. Conventional storage systems data protection features work at the volume or LUN level they can't give each virtual machine the level of data protection it deserves.

As if converging the compute and storage functions for a virtual environment into a single system weren't enough SimpliVity's OmniStack technology is also at the forefront of a movement to optimize storage for virtual workloads by making the virtual machine, rather than the storage volume, the unit of storage management.

Each OmniCube adds SSDs for performance and spinning disks for capacity as it joins the federation. The resulting storage pool is compressed and deduplicated for storage efficiency and performance.

OmniCube's greatest strength is in data protection. Administrators can easily assign policies to virtual machines that define the frequency at which the system snapshots and replicates the VM's data. The system's global deduplication reduces the amount of data that must be replicated reducing demand for expensive WAN bandwidth whether replicating to an OmniCube at a remote site or Amazon Web Services for cloud backup.

Appendix

OmniCube Models

SimpliVity currently offers three OmniCube models each targeting a different market and set of workloads. All include:

- Dual 6 or 8 core Xeon Processors
- The OmniCube accelerator card that offloads data compression and some other storage related tasks from the main processor
- PCIe slots for optional 1 or 10Gbps Ethernet

CN-2000

SimpliVity's Entry Level OmniCube for SMEs and Remote Offices

Each CN-2000 comes with:

- Four 100GB SSDs
- Eight 1TB 7200 RPM hard drives
- 48-128 GB of main memory
- Two 10Gbps Ethernet
- Additional two 1Gbps Ethernet connections for management

CN-3000

Simplivity's General Purpose OmniCube

Each CN-3000 comes with:

- Four 200GB, 400GB or 800GB SSDs
- Eight 3TB 7200 RPM hard drives
- 128-768 GB of main memory
- Two 10Gbps Ethernet
- Additional two 1Gbps Ethernet connections for management

CN-5000

The Ultra High Performance OmniCube

Each CN-5000 comes with:

- Four 400GB or 800GB SSDs
- Twenty 900GB 10000 RPM hard drives
- 384-768 GB of main memory
- Two 10Gbps Ethernet
- Additional two 1Gbps Ethernet connections for management

About DeepStorage

DeepStorage, LLC. is dedicated to revealing the deeper truth about storage, networking and related data center technologies to help information technology professionals deliver superior services to their users and still get home at a reasonable hour.

DeepStorage Reports are based on our hands-on testing and over 30 years of experience making technology work in the real world.

Our philosophy of real world testing means we configure systems as we expect most customers will use them thereby avoiding "Lab Queen" configurations designed to maximize benchmark performance.

This report was sponsored by SimpliVity. However, DeepStorage always retains final editorial control over our publications.