The Power of Deduplication-Enabled Per-VM Data Protection

SimpliVity's Hyperconverged Infrastructure Aligns VM and Data Management

A DeepStorage Technology Report

DeepStorage.net

Prepared for
SimpliVity

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 our client. However, DeepStorage always retains final editorial control over our publications.

Contents

The Bottom Line

Many organizations have sufficiently virtualized their compute workloads such 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, to provide each the level of protection it deserves, have been thwarted by a storage infrastructure that provides its services to an entire volume.

SimpliVity's hyperconverged infrastructure provides an integrated scale-out solution that delivers the compute and storage resources to support a modern virtualized environment. SimpliVity also includes a set of advanced management and data protection capabilities, all of which operate on individual VMs.

SimpliVity'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 disaster recovery (DR)*
- *Application-consistent backups with minimal space overhead*
  - *Protects 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, server, and to a lesser extent desktop, virtualization has, in many ways, 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), which 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 that 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 servers 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, or COW, 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 that have modern storage systems use snapshot technology as their primary backup mechanism, satisfying the daily restore demand from those image backups.

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 backups 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; it's 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 Hyperconverged Infrastructure

Organizations can bring SimpliVity's hyperconverged infrastructure into their data center in the form of SimpliVity's own OmniCube appliance or SimpliVity's OmniStack technology integrated into servers of their OEM partners, such as Cisco's Unified Computing System (UCS).

Either way, a SimpliVity hyperconverged infrastructure environment provides both scale-out storage and compute resources to a VMware vSphere or KVM cluster. Since vSphere is most widely deployed, the rest of this paper will focus on vSphere. Regardless of which SimpliVity appliance an organization prefers, they can choose between multiple configurations to fit their compute and storage requirements across a wide range of use cases.

When connected together, and running VMware's vSphere, the customer's preferred hypervisor, a number of SimpliVity hyperconverged infrastructure appliances form a federation, which pools the compute, primary storage, and backup storage resources in a location.In addition to the hypervisor, each node also runs the OmniStack virtual controller, a virtual machine that manages the nodes' hard and solid-state disks.

The virtual controllers in all the SimpliVity hyperconverged infrastructure nodes in a data center pool their storage to create a single distributed SimpliVity file system. vSphere datastores created by the SimpliVity hyperconverged infrastructure 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.

Within each SimpliVity HCI appliance, data is parity protected when on SSDs and double parity protected when on spinning disks. All data is also replicated to a second SimpliVity appliance within the data center, so a SimpliVity federation can suffer a node failure and multiple disk failures without losing data.

Since data is protected in each SimpliVity hyperconverged infrastructure appliance, as well as being distributed across the network, users can put one or two appliances in a remote or branch office where many other hyperconverged infrastructure solutions require a minimum of three or more nodes per location.

Before data is written to the media, it's deduplicated inline at a very fine grain of 4-8KB, compressed and serialized, 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, improves 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, SimpliVity implements global deduplication to further reduce replication traffic.

The whole SimpliVity federation, including the virtual controllers, is managed through a vCenter plug-in 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.

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 hyperconverged 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 hyperconverged 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 a SimpliVity hyperconverged infrastructure 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

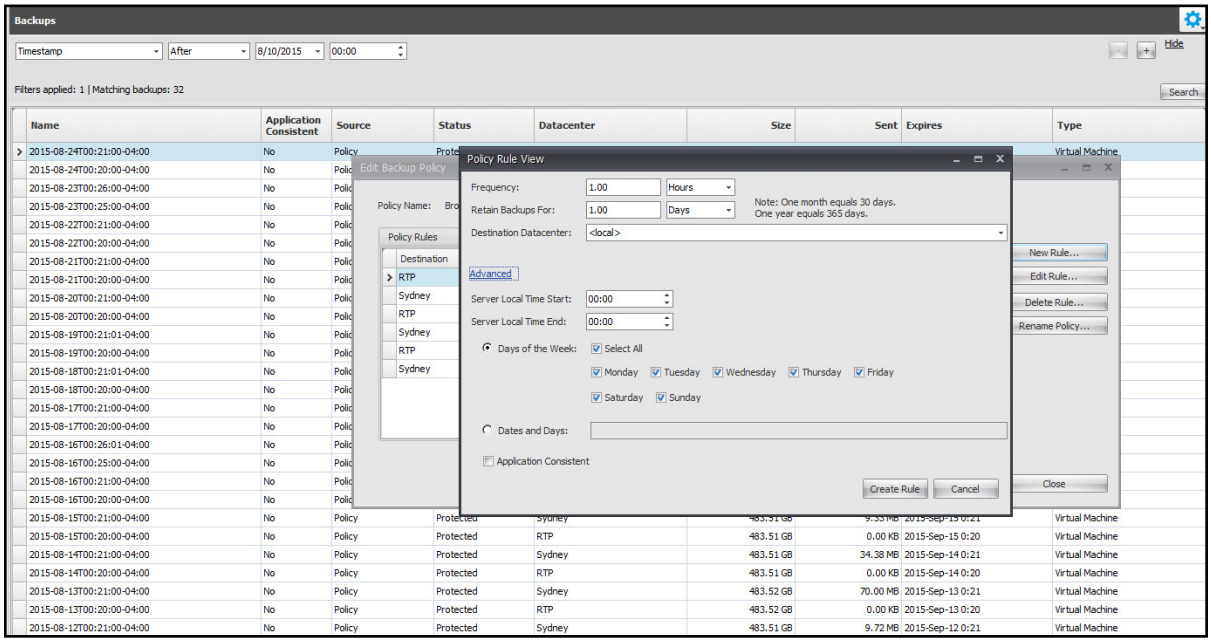
While the creators of most other hyperconverged infrastructure solutions are satisfied with integrating storage and compute into the same appliance, SimpliVity goes a step further by integrating full-functioned data protection into its systems. By using metadata to capture the changes to a VM's state since the last backup and data deduplication-optimized replication, the SimpliVity system simply and efficiently protects user data, without the need for a dedicated backup system.

Backup policies are the backbone of SimpliVity's data-protection model. A backup policy is made up of a set of rules, each of which defines how frequently the SimpliVity Federation creates 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, SimpliVity backups are sent to a SimpliVity 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 SimpliVity hyperconverged infrastructure appliances are part of the same federation, and the datacenter is defined in vCenter, administrators can skip the step of defining a backup destination and its 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, ensuring 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 off site. That off-site replication can be to a SimpliVity hyperconverged infrastructure in another data center or even to Amazon Web Services' S3 cloud storage service.

SimpliVity's backup policy management now allows administrators to not only create a backup policy and associate it with a VM, but also to manage VM backup policies in bulk and manipulate the retention policies for existing backups and create ad hoc policies for manual backups.



Defining a SimpliVity hyperconverged infrastructure appliance 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, since SimpliVity uses inline data deduplication and compression to optimize the data across all the nodes in a datacenter, it minimizes the amount of space the backups take up. Regardless of how many Windows Server 2012 R2 virtual machines you back up, there's always just two copies of that operating system stored in the datacenter regardless of how many servers you're backing up or how many backups you retain.

Off-Site Replication

If your SimpliVity federation extends over more than one datacenter, you can specify that the backup be stored on the file system of the SimpliVity cluster in another datacenter. The snapshot will be taken locally and then replicated to the SimpliVity 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 2012 exists in the SimpliVity cluster in San Francisco, when you send the first backup of a Windows Server 2012 VM, the common operating system data won't be sent over the link. Before actually sending data, the source and destination SimpliVity hyperconverged infrastructure nodes negotiate which data blocks need to be sent.

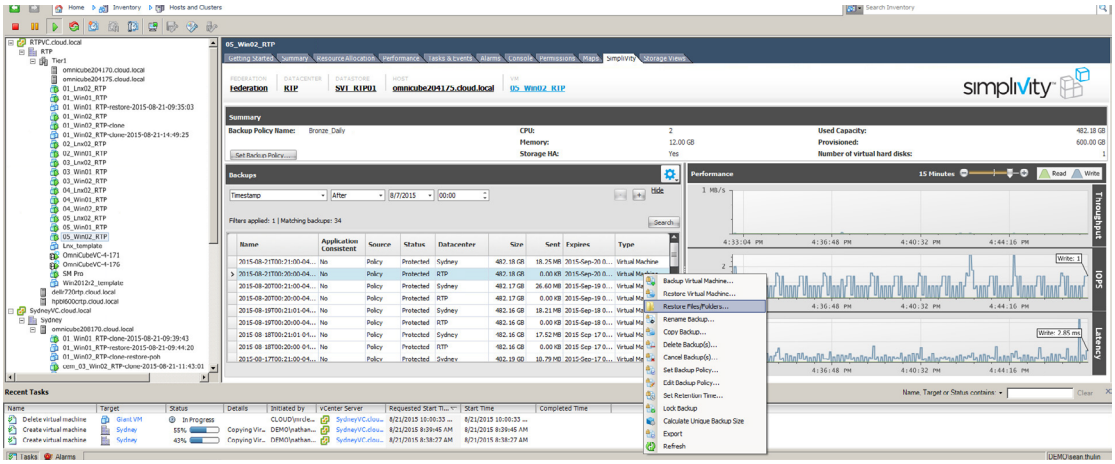
Organizations can manage the replication topology to match their geographic or network requirements. While many organizations may use a full-mesh design over internet VPNs, others may prefer a more centralized hub-and-spoke system.

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

File-Level Restore

While we tend to think of our backups as primarily protection against system failures of one type or another, the sad truth is that most restores are due to user error, not system failure. It's not at all uncommon for the help desk, or backup administrator, in a large organization to get a phone call saying, "Hi, this is Bill in accounting. We lost the Excel spreadsheet that we use to close the quarter. Can you restore it for me?" or, "I opened the Smith proposal and it has information for the Jones project in it. I know it was OK last month."

With some data-protection systems based on system images, satisfying one of these requests for a single file restoration can be a nightmare. The admin has to restore the whole VM, isolate it from the network so it won't cause a conflict with the production server, power up the VM, and then copy the file from the now isolated system back to the production system.



Administrators Can Restore Individual Files From Simplivity Backups

SimpliVity provides a simple wizard that allows the administrator to select a backup and mount it like a DVD. He or she can then easily copy the file the accounting department swears they need right away or they can't close the quarter, and be a hero.

VM Migration

Some vendors of hyperconverged solutions simply provide the storage layer as a virtual machine without really integrating it into the hypervisor it supports. The SimpliVity vCenter plugin, on the other hand, makes it easy to migrate or copy virtual machines from one datacenter to another. These migrations use SimpliVity'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 datacenters will really like this feature.

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

VAAI NFS Integration

SimpliVity's OnmiStack software now supports 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 SimpliVity 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 simplify server management, it has complicated the process of data protection. Since 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 hyperconverged infrastructure 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 SimpliVity hyperconverged infrastructure appliance adds SSDs for performance and spinning disks for capacity as it joins the federation. Data in the resulting storage pool is compressed and deduplicated inline for storage efficiency and performance improvement.

SimpliVity's greatest strength is in data protection. Administrators can easily assign policies to virtual machines that define the frequency at which the system backs up 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 SimpliVity hyperconverged infrastructure at a remote site or Amazon Web Services for cloud backup.