THE CLIPPER GROUP
Navigator ™

Navigating Information Technology Horizons^{SM}

EMC NetWorker Offers Eight Ways to Back Up a VMware Virtual Machine

Analyst: Michael Fisch

Management Summary

Information backup in VMware virtual environments involves a unique set of challenges and capabilities. It is different from backing up dedicated physical servers because of the additional virtualization layer and multiple, self-contained virtual machines running on each server. For its part, VMware has made great strides over the years in developing its backup facilities. Most notably, the *vStorage APIs for Data Protection (VADP)* deliver a functional and streamlined capability for off-host backup.

EMC *NetWorker* now is integrated with VMware VADP. In fact, NetWorker has expanded from a traditional enterprise backup solution to one that embraces next-generation capabilities for data protection, including extensive support for VMware virtual environments:

- **VMware *vCenter*** – Integration with vCenter allows NetWorker to automatically discover and track virtual machines and their backup status, so administrators can see what they need to protect.
- **VMware in-guest backup and VADP** – Enterprises can choose in-guest or VADP off-host backup based on their workload requirements.
- **EMC *Avamar* and *Data Domain*** – Through tight integration, NetWorker brings the capabilities of two industry-leading data deduplication technologies to VMware.
- **Traditional tape backup** – For lowest-acquisition cost.
- **Un-deduplicated disk backup** – For backup data sets that do not duplicate well.

This list is long. In fact, if you calculate the permutations, you get eight ways NetWorker can back up your VMware virtual machine. Well, *at least* eight. The point is that NetWorker brings a broad toolset of capabilities, each with particular strengths and suitable use cases that allow you to implement an optimized backup strategy for your VMware environment. Read on to learn more about these options and how you might use them to protect your virtual environment.

VMware vCenter Integration

Prior to considering how to back up a virtual environment, the question to answer first is, *What data needs to be backed up?* Unlike physical servers that you can actually see while strolling through a data center, virtual machines are intangible and invisible. These dynamic server partitions are created, deleted, and moved at will. As such, the virtual environments are continually changing, and software management tools are necessary to see and track virtual machines. Moreover, the ease of provisioning, which is a major benefit of this technology, has created a problem on the other end known as *virtual machine (VM) sprawl*. This situation occurs when

virtual machines are created in abundance – in response to user requests, for testing and development, and for new applications – without controls in place to ensure they are properly provisioned, managed, and protected. In other words, they proliferate out of control.

NetWorker has a ready answer to the challenge of VM sprawl and visibility into virtual environments. It integrates with VMware *vCenter* and automatically discovers virtual machines. NetWorker even tracks them when they move between servers via *vMotion*, *Distributed Resource Scheduler* (DRS), or *High Availability (HA)*. Virtual machine discovery can be scheduled on a regular basis.

Then, NetWorker Management Console graphically maps the virtual environment and displays the backup status of each virtual machine, including new ones that have not yet been backed up. Administrators can understand at a glance the relationships between *ESX* Servers, clusters, virtual machines, and backup configurations, and then take appropriate action to ensure virtual machines are protected adequately. For example, if a server administrator creates a new virtual machine in response to a user request, the backup administrator can automatically discover it and back it up. With this baseline protection in place, he can inquire with the server administrator about the level of data protection it should have going forward.

So, NetWorker provides a good understanding of what data to back up in a virtual environment. The next question is *how you want to back up each virtual machine, because there are numerous options available*.

VMware In-guest and Off-host Backup

In-guest Backup

NetWorker supports VMware in-guest backup. In-guest backup is the same as traditional backup for physical servers, except the backup agents run in each virtual machine on an ESX server. So, if there are 20 virtual machines on a physical server, then 20 backup agents are also running. That might seem like too many agents for one server, and it may well be, which is why other approaches have been developed.

The advantages of in-guest backups are:
- **Familiar operations** – As mentioned, in-guest backup basically is the same as traditional backup. It uses the same agents. In the interest of continuity and ease of transition into virtual servers, an enterprise may choose

the in-guest backup approach.
- **Application-consistent recovery** – The only way to achieve an immediately restartable, application-consistent recovery is to use in-guest backup agents that integrate with the specific applications like those from SAP or Oracle. These backup agents also enable granular recovery, such as mailbox-level restores involving Microsoft *Exchange*.
- **File-level recovery** – In-guest backup delivers straightforward file-level recovery for operating systems supported within virtual machines.

Possible limitations of in-guest backup are:
- **Slow production applications** – Traditional backup agents compete for CPU, memory, and I/O resources on the server. When multiple agents run simultaneously during the backup window, they can overload the system and slow production applications.
- **Limit physical server consolidation** – A primary benefit of server virtualization is the ability to run more workloads per physical server, which drives up utilization, facilitates server consolidation, and ultimately reduces the costs of acquiring and operating the infrastructure. However, multiple traditional backup agents running on a server may create resource bottlenecks that effectively reduce the potential for consolidation.

VADP (Off-host) Backup

NetWorker now supports *vStorage APIs for Data Protection* (*VADP*), the significantly improved successor to *VMware Consolidated Backup* (*VCB*). Both VADP and VCB are technologies that address the problem of backup overload on VMware ESX Servers by enabling non-disruptive, off-host backup. Backup agents are not required to run in each virtual machine, as the case with in-guest backup. Rather, the ESX Server takes a snapshot copy of each virtual machine and presents the copies to a backup proxy. Backup jobs flow directly over the network from the storage array to the backup server, bypassing the source ESX Server altogether. This off-host approach facilitates faster backups and is non-disruptive to production applications. It also allows VMware to consolidate more workloads per physical server.

However, VCB suffered from limitations that hindered broad-scale adoption. Users had to install the VCB software on a dedicated physical Windows proxy server, which seemed counterproductive to the server consolidation they hoped

to achieve. Backup was a two-step process from the source to the VCB proxy and from the VCB proxy to the backup target. It also required two passes for backing up virtual machines – at the image and file levels.

The good news is that its successor VADP has corrected these limitations and added new, useful backup functionality:

- **Integration with backup application** – The VADP software library is integrated into the NetWorker backup application and now users no longer will need to download and install a separate integration module.
- **Streamlined data transfer** – Off-host backup of virtual machines is a single-step process from the storage system over the SAN to the NetWorker storage node. Restores are also much faster.
- **Non-disruptive snapshots** – VADP leverages the differential snapshot facility in *VMFS*[1] to create virtual machine images non-disruptively for use as the backup source.
- **Change Block Tracking** – VMware time-stamps changes to block data in *VMDK*[2] files and makes this information available through VADP. NetWorker leverages it to back up changed data only incrementally. Therefore, backups run more quickly with less I/O traffic. NetWorker can restore incremental or full virtual machine images.
- **Single pass with file level recovery** – With a single-pass backup of virtual machine guest images, NetWorker offers both image and file-level[3] restores.

When choosing between VADP and in-guest backup for your workloads, VADP is preferable because it is non-disruptive and offers greater efficiency, but there are limitations such as application consistency where in-guest is the viable alternative.

Data Deduplication

NetWorker supports data deduplication through cohesive integration with EMC *Avamar* and EMC *Data Domain*. Deduplication technology scans for repetitive data segments in the backup stream and replaces subsequent occur-

rences with pointers to the original. Since backup data is highly repetitive, deduplication is especially effective at reducing storage requirements (by an average of 40-60%, depending on the data) and paves the way for greater adoption of disk backup. It becomes economical to store more backup data for longer periods on disk, and the business benefits from faster and more reliable backups and recoveries.

EMC Avamar

Avamar is the industry's leading next-generation deduplication backup software. It reduces the amount of backup data at the source (client) – before it is transferred across the network and stored to disk. By sending only new, unique sub-file variable length data segments, Avamar enables fast, daily full backups despite congested networks or infrastructure. As a result, Avamar can reduce the required daily network bandwidth by up to 500 times and cumulative backend storage can be reduced by up to 50 times across sites and servers. It also offers replication to a remote site for disaster recovery.

EMC's integration of NetWorker and Avamar has produced a tightly coupled solution.[4] By combining the NetWorker and Avamar clients into a single backup agent and managing Avamar within NetWorker resources, administrators can manage the entire backup environment through a familiar user interface and common workflow. This reduces complexity by simplifying day-to-day tasks and further consolidating the backup infrastructure. Backup procedures do not have to change. Administrators can recover deduplicated data from the same user interface as traditional backup data. NetWorker automates the entire process, so there is no need for guesswork at recovery time, in terms of how data was protected or where it is located. You have one go-to vendor for backup implementation and support. Avamar also can be licensed separately in non-NetWorker environments.

Avamar's client and global deduplication provides daily full backups and is particularly effective for:

- **VMware environments**, because it reduces the amount of data backed up at the source ESX Server, streamlines and speeds up the backup, and alleviates the problem of system

---

[1] VMFS (Virtual Machine File System) is a cluster file system used in VMware ESX Servers.
[2] A VMDK (virtual machine disk) encapsulates an entire virtual server or desktop environment in a file.
[3] File-level recovery from images currently is available for Windows only. Linux file recovery requires file-level backup.

[4] See **The Clipper Group Navigator** dated June 8, 2010, entitled *EMC NetWorker and Avamar — An Integrated Pair for Traditional and Deduplication Backup* and available at http://www.clipper.com/research/TCG2004028.pdf.

resource overload during backup on highly utilized servers.

- **Remote office backup and desktop/laptop PCs**, because Avamar provides daily full backups via existing network links and enables end-user recovery without the need for IT staff intervention
- **File servers**, because of the nature of their data, with relatively low rates of data change and the space and times savings from only backing up altered file segments.
- **NAS systems,** because Avamar provides daily full backups without the need for recurring, lengthy level-zero full backups that often exceed available windows.

EMC Data Domain

Data Domain is the industry-leading deduplication storage system for disk-based backup. A Data Domain appliance slips into an enterprise backup environment by connecting to a backup system as a file server (*CIFS*, *NFS*) over an Ethernet network, as a VTL[5] over a Fibre Channel network, or as EMC *Data Domain Boost* (*DD Boost*) with NetWorker. It delivers high-performance throughput and an average of 10 to 30 times data reduction. Data Domain also offers network efficient replication of deduplicated data to a disaster recovery site.

EMC NetWorker with DD Boost is the ultimate integration of these two solutions.[6] DD Boost is like a turbocharger for deduplication storage. This option simultaneously speeds up the backup process, reduces the impact on the NetWorker storage node, and simplifies deduplication management within the NetWorker Management Console. It is accomplished through software only, so no additional hardware is necessary.

According to EMC's figures, DD Boost delivers significantly faster backup performance: up to 50% faster backups, 20% to 40% less impact on the backup server, and 80% to 99% less backup traffic on the LAN. DD Boost operates by offloading parts of the deduplication processing to the NetWorker storage node while performing the remainder on the Data Domain system. (Without DD Boost, the Data Domain system handles all of the deduplication processing.) The NetWorker storage node breaks the backup

data into variable-length segments. It checks with the Data Domain system to determine if the segment is unique and then compresses and sends any unique data segments over the network. The Data Domain system on the receiving end completes the deduplication process and writes the unique, compressed data to disk.

DD Boost reduces the overall resource utilization on the NetWorker storage node because the added deduplication processing is more than offset by a reduction in I/O processing. It also optimizes the communication between the storage node and Data Domain system by only sending unique data over the LAN.

DD Boost enables system management from the NetWorker Management Console, giving administrators a single point of management, monitoring, and reporting for deduplication backup and recovery. NetWorker automatically discovers the systems and assigns them their own "Data Domain" device type. Once a Data Domain device is configured, it can serve as a target for client backups and administrators can set policies for deduplication backup in NetWorker, just like other target devices. The cloning workflow in NetWorker also automatically manages and controls replication between Data Domain systems.

Data Domain systems are particularly effective for:

- **Existing backup processes and policies**, because storage deduplication really is a plug-and-play solution for existing backup environments. This is a major reason why it is so popular for enterprises that want to enjoy the benefits of data deduplication without overhauling their backup infrastructure.
- **Transactional applications with a high rate of data change**, because deduplication storage systems can be faster than deduplication software when the database is heavily utilized and has a high rate of data change.
- **High-priority recovery**, because host severs or virtual machines with large volumes of data that require the fastest recoveries may be served better by storage deduplication.

Traditional Tape and Disk Backup

Lastly, but still relevant, are traditional backups to disk and tape. Tape is the least expensive backup media. It is also compact and does not consume power when archived. So, tape is still a good choice for backup when low-cost storage is the overriding factor, as opposed to RPO or

---

[5] Virtual Tape Library.

[6] See **The Clipper Group Navigator** dated October 4, 2010, entitled *EMC NetWorker Rounds Out Deduplication Support with EMC Data Domain Boost* and available at http://www.clipper.com/research/TCG2010046.pdf.

RTO, e.g., for long-term archiving.

Un-deduplicated backup to disk is another option. While deduplicated *backup to disk* is usually faster, there are certain data types that do not deduplicate well. In those cases (image and video files or other pre-compressed data types), traditional backup to disk might be justified.

Assessment Questions

In summary, NetWorker supports numerous backup options for VMware virtual environments:

- VMware in-guest
- VADP off-host
- Avamar deduplication
- Data Domain deduplication
- Traditional tape
- Un-deduplicated disk

The question is, *Which should you employ in your environment?* If your IT environment is small or homogeneous, the straightforward simplicity of a single technique might be sufficient if it meets the recoverability requirements of all workloads. In larger and more sophisticated IT environments, we recommend a nuanced approach employing multiple backup techniques that balance the cost and recoverability requirements of individual workloads.

To get you started, here are some questions to help you characterize your VMware backup environment:

- *What are the RPO, RTO, and archiving requirements of the various workloads? Budget constraints?*
- *Is the data rate of change of the applications high or low? File servers usually have a low rate of change, but transactional applications sometimes are high.*
- *Is your organization geographically distributed, and would you like to consolidate backups at a central location?*
- *Are you creating a backup system from a blank slate, or do you have an existing system with established procedures that you need to evolve?*
- *Do you want to transition to all-disk backup, or will tape continue to play a significant role?*
- *Do you need application-consistent recovery, or is crash-consistency sufficient?*
- *Do you need to maintain non-disruptive operations?*

Conclusion

As we hear so often, there is more than one way to skin a cat – and there is more than one way to backup a VMware virtual machine. For many enterprises, using multiple techniques to back up multiple, unique workloads in your virtual environment is the best approach. It allows you to meet the business requirements for backup speed, RPO and RTO, asset utilization, and cost-effectiveness more precisely and flexibly.

The good news is you do not have to commit to a single approach if you use EMC NetWorker as your enterprise backup application. You have options, because, in conjunction with EMC Avamar and Data Domain, NetWorker supports "virtually" everything.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

➢ *The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.*

About the Author

Michael Fisch is a Senior Contributing Analyst for The Clipper Group. He brings over 14 years of experience in the computer industry working in marketing, sales, and engineering, the last nine of which he has been an analyst with Clipper. Before Clipper, Mr. Fisch worked at EMC Corporation as a marketing program manager focused on service providers and as a competitive market analyst. Before that, he worked in international channel development, manufacturing, and technical support at Extended Systems (since acquired by Sybase). Mr. Fisch earned an MBA from Babson College and a Bachelor's degree in electrical engineering from the University of Idaho.

➢ *Reach Michael Fisch via e-mail at mike.fisch@clipper.com at 781-235-0085 Ext. 211. (Please dial "211" when you hear the automated attendant.).*

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager,** Clipper Notes, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.