

## BUYER CASE STUDY

---

### VMware Improves Network Utilization and Backup Performance Using EMC Avamar Deduplication

---

Laura DuBois

---

#### IDC OPINION

---

In physically constrained datacenter environments, deduplication technology yields significant savings in storage optimization and datacenter footprint. However, a client-side approach to deduplication can yield even further benefits in terms of improved network bandwidth utilization and smaller backup windows. For firms making use of server virtualization or protecting remote branch office data, the benefits of a client-side approach include reducing backup windows and network I/O, eliminating local remote office backups and tape infrastructure, and offloading congested WAN links. This IDC Buyer Case Study:

- Explores the benefits of client-side deduplication technology at VMware Inc.
  - Discusses VMware's implementation of EMC Avamar within its corporate datacenter
  - Highlights future plans for the use of the technology within VMware's remote and branch offices
- 

#### IN THIS BUYER CASE STUDY

This IDC Buyer Case Study — based on a March 2009 interview with Chanh Chi, IT architect at VMware, and Michael Pate, manager of Technical Operations at VMware — examines the company's use of client-side deduplication. In 2007, VMware evaluated and deployed the Avamar client-side deduplication and replication offerings as a means of protecting and recovering over 300 physical and virtual systems across its infrastructure.

This document examines VMware's implementation of the Avamar technology, including a discussion of its organizational challenges and evaluation requirements. VMware's results are highlighted and future implementation plans are summarized. End users can use this information to assess whether client-side deduplication technology is applicable within their environment and what benefits can be expected using the Avamar solution set.

## SITUATION OVERVIEW

---

### Organization Overview

VMware (NYSE: VMW) is a leader in virtualization solutions from the desktop to the datacenter. Firms rely on VMware to reduce costs, ensure business continuity, strengthen security, and go green. With 2008 revenue of \$1.9 billion, more than 130,000 customers, and more than 22,000 partners, VMware is one of the fastest-growing public software companies and is majority owned by EMC Corp.

The overall IT environment at VMware includes hundreds of Windows and Linux physical and virtual servers running applications such as Microsoft Exchange, Microsoft SQL, and Oracle in addition to directory, file, and print servers. The organization has approximately 50TB of data across all its physical and virtual servers. The IT group supports over 6,000 users.

As members of the VMware IT organization, Chi and Pate provided insights into VMware's challenges, evaluation of new backup technology, and implementation results using Avamar. In his role as IT architect and his previous position as backup administrator, Chi was instrumental in laying out a new architectural direction and strategy for the firm to address its backup and recovery challenges.

---

### Challenges and Solution

The IT organization at VMware faced several challenges. With the growth in infrastructure, its existing datacenter was running into physical space constraints. The organization's existing environment utilized EMC NetWorker for backups to both physical tape and CIFS-attached Data Domain deduplication appliances, implemented to eliminate the use (and offsite removal) of 200 physical tapes weekly. The organization was also replicating to a Savvis colocation datacenter in Palo Alto, California, for its disaster recovery site. It clearly needed a larger, more stable, and environmentally efficient datacenter. But it also wanted a solution that would meet its current and future data protection needs, so it decided to reevaluate its current backup strategy and rearchitect its environment.

During the evaluation of its backup environment, VMware's IT department found it was facing severe network bottlenecks. As expected, the organization utilizes its VMware solutions to consolidate and maximize the use of available servers. VMware was over 95% virtualized and the virtual machine guest systems shared the same underlying physical network interface on the physical servers. VMware's existing backup methods consumed 100% of the available network during its 24 x 7 backup window, creating severe backup bottlenecks. With too much data to protect and shrinking backup windows, the organization was faced with staggering full backups for each client once a week and then doing incremental backups the rest of the week at best. According to Chi, "With expected data growth, we knew this approach would not scale."

In 2006 (prior to the EMC acquisition), the VMware organization considered three options to address its backup challenges. It considered augmenting its backup environment with a new deduplication-enabled application for its virtualized environment. It also considered the use of EMC Disk Library solutions in its backup

environment. However, since its paramount problem was the shrinking backup window and network congestion in its virtualized environment, the organization focused mainly on client-side deduplication solutions, evaluating with Symantec NetBackup PureDisk and EMC Avamar. At the time, the Symantec NetBackup PureDisk product did not support all the clients VMware had in its environment such as Oracle servers and NAS filers, which ruled out this solution.

The VMware IT team performed an onsite proof of concept comparing the performance of all three upgrade scenarios: EMC NetWorker backups to tape, EMC NetWorker backups to Data Domain deduplication appliances, and EMC Avamar backups to generic disk. The team requested an Avamar evaluation unit, and it tested Avamar in its own environment for about two months. The VMware IT team conducted several tests comparing results from backups of the same 12 Windows and Linux file servers and an Exchange backup, totaling about half a terabyte of data. The tests clearly showed that Avamar's backup performance was superior to the alternatives, despite running on a slower network connection. According to Chi, "The deduplication ratios varied, but on average, we saw a 3:1 optimization ratio on the first backup of most clients. Some Windows file system clients yielded optimization ratios as high as 500:1 on subsequent backups."

---

## Results

With strong results from its proof-of-concept tests, VMware purchased the Avamar product and deployed it in December 2007. Although VMware was part of EMC at that time, the Avamar deployment was made after the VMware IPO, which meant VMware had to purchase the product. The new production datacenter in Santa Clara, California, went live in late 2007. According to Chi, "An important feature for us in choosing Avamar was its remote replication. This allowed us to replicate backup data between our new production datacenter and DR datacenter, eliminating tape handling."

VMware purchased two Avamar systems consisting of 12 data nodes each. EMC Professional Services helped VMware with the Avamar setup, which occurred over two days. One Avamar system is located in Palo Alto, California, and data is replicated bidirectionally between the two Avamar systems every day. VMware started with 12 data nodes in each Avamar "grid," with the first grid protecting approximately 12TB of primary data and the second grid protecting about 7TB of primary data. Across both grids, the Avamar solution protects over 300 virtual and physical file server systems. The organization has kept EMC NetWorker in place for backup of the EMC Celerra filers straight to tape and for backup of the Exchange environment to existing Data Domain devices. As more clients come online, VMware IT has since added another four storage nodes per grid for a capacity of 16TB per grid and purchased an additional higher-density grid with 32TB of capacity:

- ☒ **Virtual machine backups.** VMware deploys Avamar software agents within each VM guest and also deploys Avamar in its VMware Consolidated Backup (VCB) environment for a handful of clients that require extra protection. In all cases, Avamar provides fast, daily full backups.
- ☒ **Backup window.** According to Chi, "Within the first 10 minutes, 50% of the 360 clients will have completed their daily full backup. After an hour, only a handful are still left." Prior to implementing Avamar, backups would take a full day or longer.

- ☒ **Backup performance.** With one virtual client with 12 LUNs, each with 415GB of data, backup with NetWorker would take a week. Now, an Avamar backup of 5TB and 40 million–50 million files will take approximately 11 hours.
- ☒ **Network performance.** Overall, the organization does approximately 18TB of backups nightly and has 200GB of traffic over the network, utilizing approximately 1% of the network bandwidth needed for a traditional backup.
- ☒ **Recovery performance.** Using Avamar to recover data is dramatically easier and faster than its previous solutions. According to Chi, "To provide recovery of individual files because of corruption or user error previously meant waiting a few hours for a 200KB file. With Avamar, it's basically done after you start it. The data is rehydrated on the client and we have seen an 8GB file restored in just 30 seconds."
- ☒ **User interface.** According to Chi, "The Avamar user interface is very intuitive and no real backup experience is necessary, enabling some backup tasks to be offloaded to a broader set of IT staff. For example, release engineering teams can perform their own backup and recovery of source code."

VMware currently keeps approximately 30 days of backups on the disk-based Avamar systems and does weekly backup dumps to tape for long-term retention and as a last resort, third copy.

---

## Future Plans

According to Pate, "We have doubled our Avamar capacity in the last week. We have purchased 52TB of Avamar to back up our remote R&D and engineering sites internationally. We have placed a grid in a new datacenter in the state of Washington to back up 10 remote and branch office locations." VMware will rely upon the Avamar local software agent to back up to a larger datacenter. Here, the Avamar solution has helped in dealing with limited WAN links back to the larger regional datacenter. This approach has allowed VMware to back up primary file servers in remote sites to a Avamar data store grid within a larger datacenter and eliminate the use of tape changers at remote branch locations and also eliminate on the cost of approximately 1,000 tapes per quarter.

---

## ESSENTIAL GUIDANCE

### Advice for End Users

End users should consider the following recommendations to help them evaluate the potential benefits of data deduplication technology within their environment:

- ☒ **Not all deduplication is equal.** There are several different places where deduplication technology can be used such as in backup software, on backup appliances, and so forth. Users should outline the main problems they are trying to solve and map those challenges to the deduplication approach best suited to provide results.

- ☒ **Proof-of-concept testing is needed.** Deduplication is in strong demand today, and many suppliers offer technology that claims to offer deduplication. However, IDC has found that during proof of concepts, some deduplication solutions do not work as advertised.
- ☒ **Test with actual data you plan to back up.** Deduplication ratios will vary based on the type of data that is being backed up. Some data deduplicates well, while other types of content do not. Understanding real-world deduplication ratios as outlined in this document as well as the deduplication ratios obtained from your proof of concept will help ensure no surprises once the technology is in production.
- ☒ **Other optimization workloads need to be considered.** Optimization technologies such as compression and deduplication improve the economics of using disk for data protection and enable more backups to reside on disk for extended periods of time. Firms should ensure that compression and deduplication (as well as encryption) are complementary and done in the right order to achieve the technology benefit optimization affords.

## LEARN MORE

---

### Related Research

- ☒ *Worldwide Storage Software 2009–2013 Forecast: Quest for Efficiency Driving Investments in Storage Software* (IDC #217529, March 2009)
- ☒ *Worldwide Data Protection and Recovery Software 2007 Vendor Shares: Market Rebounds* (IDC #213320, July 2008)

### Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit [www.idc.com](http://www.idc.com) to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit [www.idc.com/offices](http://www.idc.com/offices). Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or [sales@idc.com](mailto:sales@idc.com) for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights.

Copyright 2009 IDC. Reproduction is forbidden unless authorized. All rights reserved.