IDC
Analyze the Future

BUYER CASE STUDY

EMC IT Increasing Efficiency, Reducing Costs, and Optimizing IT with Data Deduplication

Sponsored by: EMC Corporation

Robert Amatruda Laura DuBois

February 2011

EXECUTIVE SUMMARY

Data deduplication is being deployed within datacenter, remote/branch office, and endpoint environments as a means to reduce storage cost and footprint, improve network bandwidth utilization, improve service levels, and reduce backup windows. Like many large enterprises, EMC Corporation was experiencing a number of challenges in its legacy backup environment that impacted data availability and recovery. These hurdles included increasing backup volumes, constrained network bandwidth, poor backup reliability, and inadequate recovery times. As a result of these issues, EMC IT embarked on a mission to transform its backup and recovery operations, with data deduplication technology at the core of this transformation.

This IDC Buyer Case Study:

- Explores the benefits EMC realized from the use of a range of EMC's own backup and recovery solutions that leverage deduplication technology.
- Identifies the unique backup challenges for different computing environments, including datacenters, remote offices, and laptops/desktops, and how data deduplication can address these environments.
- Highlights EMC's legacy backup environment and the changes EMC made as part of a transformation process to increase efficiency, reduce cost, and optimize IT — as part of its journey to the private cloud.

IN THIS BUYER CASE STUDY

This IDC Buyer Case Study is based on interviews with EMC IT staff and outlines the benefits and cost savings of deploying EMC backup and recovery solutions in concert with VMware virtualization to manage a rapidly growing worldwide organization. This case study is unique in that it illustrates EMC's usage of its own backup and recovery technologies. With discretion to use any products in the industry to achieve its objectives, EMC IT elected to leverage the same products that EMC sells to its customers.

EMC had been experiencing rapid data growth resulting from its more than 48,000 internal users in approximately 400 offices in 80 countries worldwide. Faced with diminishing returns of aging tape solutions, EMC's IT staff embarked on a mission to completely transform its data protection and recovery operations both in core datacenters and in distributed locations and endpoints.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

This document examines the implementation of EMC backup and recovery solutions and the business value and specific metrics gained from this deployment. End users can use this information to benchmark whether deduplication technology is applicable within their environments and what benefits can be expected using the EMC solution set.

SITUATION OVERVIEW

Organization Overview

EMC Corporation (NYSE:EMC) is a world leader in information infrastructure solutions with 2010 revenue of $17 billion. EMC solutions are designed to help customers grow and operate efficiently, and the company shares the same strategic objectives for its own global IT environment. Like all IT organizations, EMC IT faces the challenge of balancing cost, risk, and agility in its operations. The functionality, interoperability, and performance requirements of its internal customers must be satisfied — without compromising the security and manageability of IT systems and processes. EMC IT must also justify all of its investments with strong, metrics-based business cases that demonstrate return on investment (ROI) and total cost of ownership (TCO) before receiving management approval.

EMC faced challenges familiar to many large IT organizations. Its distributed employee and office network posed many challenges to adequately safeguard rapidly growing critical data — especially with its growing virtualized environment. EMC IT supports more than 48,000 users in approximately 400 offices, and hundreds of partners, across 80 countries and in excess of 500 business applications. The IT organization is managing over 7PB and backing up over 110TB of data nightly on more than 5,000 servers, totaling 40PB annually. Its IT staff is supporting a wide array of storage infrastructure in its five datacenters worldwide. EMC is managing five storage tiers — three primary, one for backup, and one for archive. In addition, EMC has started on a journey to the private cloud and has begun to standardize on the x86 architecture for its server platform. According to the company, it has, to date, achieved 75% virtualization of its server environment. EMC expects to achieve the goal of 100% server virtualization by 2012.

Challenges and Solution

EMC needed to address enormous annual data growth while safeguarding its growing distributed environment and reducing the number of failed backups. EMC's data growth was exceeding 60% annually, driven by a combination of rapid business expansion and acquisitions of more than 40 firms. Backup and recovery was becoming increasingly critical to ensure business continuity and meet compliance requirements. Meanwhile, EMC was struggling with growing backup windows and failures and was having difficulty safeguarding its growing distributed environment.

Remote Offices and Branch Offices

As EMC grew, remote backup operations eventually spread across 149 sites worldwide. These sites used a variety of tape- and disk-based solutions — commonly deployed and managed by the local office personnel — and were outside the purview

of EMC IT. Seventy-one percent of remote locations did not have IT personnel onsite. This led to approximately 6,000 backup-related trouble tickets issued annually due to missing tapes, media errors, and backups that timed out. Tapes often were stored offsite and could take days to locate, resulting in restoration times ranging from a couple of hours to several weeks.

With regulated, business-critical data at risk, EMC IT decided that it needed to implement a centralized, global backup solution. The company required an easy-to-use automatic, disk-based system able to support both offsite and local backup. Due to the limited bandwidth available on existing networks, it also became evident that a data deduplication solution would be advantageous and would eliminate the need to send redundant backup data over congested networks. EMC IT reviewed many vendors and selected Avamar (prior to its acquisition by EMC). EMC Avamar backup software with integrated deduplication reduced the size of backup data before it was passed across the network, enabling faster daily full backups over existing bandwidth.

Today, EMC's remote sites — 72 in the Americas, 49 in Europe, and 28 in Asia/Pacific — are automatically backed up to one of four Avamar storage grids. Two of the Avamar grids are located on the east and west coasts of the United States. The other two Avamar grids are located in Bangalore, India, and Cork, Ireland. Data that is backed up is continuously replicated to alternate grids, ensuring recovery in the event a grid is disabled or taken offline. Each EMC Avamar grid contains 16 data nodes (plus a utility and spare node) and can store more than 52TB of data.

EMC Data Protection Advisor software provides real-time monitoring and reporting, and EMC Ionix for IT Operations Intelligence software provides automatic alerts on issues relating to other components of the backup infrastructure, such as servers and network equipment.

Enterprise Datacenters

Prior to 2005, all of EMC's backup data was sent to physical tape. There was a high failure rate associated with restore from physical tape. Additionally, EMC's IT staff was spending an inordinate number of man-hours creating custom scripts for business continuance volumes (BCVs) and other types of replicas to support tape backup. Specifically, many man-hours were spent on manual scripting for Microsoft Exchange email protection. EMC utilized its Replication Manager to reduce its reliance on custom scripts and took advantage of replication technology to safeguard its environment.

EMC IT introduced the EMC Disk Library as the primary target for writing backup data sets. This improved its backup success rate relative to tape. However, it did not address management or availability issues as the growth of data increased and EMC continued to make acquisitions. Furthermore, the EMC IT staff had to manage many antiquated technologies and processes that came as part of some of EMC's acquisitions. Much time and effort was put toward the management of backup data in multiple platforms, formats, and geographic locations.

EMC was not effectively mitigating its risks from unexpected downtime, loss of data, and increasing time for backups. EMC undertook an application alignment process to determine the recovery time and point objectives and SLA expectations companywide.

EMC's goals and objectives were to reduce lengthy backup windows, reduce or delete expired or unwanted data, archive data to a lower-cost tier of storage, deduplicate redundant data, as well as automate operations. In addition, any potential solution had to be operable with EMC's rapidly virtualizing server infrastructure.

Because of the successful remote backup project, EMC IT installed an Avamar storage grid in one of its primary datacenters for VMware backup and recovery. Avamar is integrated with the EMC NetWorker backup application, which enables EMC IT to centrally manage Avamar deduplicated backups, as well as other disk and tape backup operations in the datacenter.

EMC IT is using EMC Data Domain deduplication storage systems to manage the backups of Oracle, SQL, and Exchange. Data Domain was selected based on its performance, scale, and deduplication efficiency for these workloads. In each case, NetWorker is used as the backup application to deliver the backups to a Data Domain system. All Data Domain systems replicate backup data to EMC datacenter facilities in Durham, North Carolina, and Hopkinton, Massachusetts. In this manner, EMC is utilizing each technology for its specific differentiating function: Avamar to deduplicate and minimize backup data transmitted across the network, Data Domain for massive scale and performance, and NetWorker for single-pane-of-glass backup management.

Specifically, EMC is now supporting 32 Disk Library systems, eight Avamar grids, and four Data Domain storage systems.

Desktops and Laptops

EMC had no formal, established policy for end-user client system protection. With nearly 48,000 desktops and laptops worldwide, 7% were protected with an online backup/restore solution that was very difficult to scale and manage effectively and was no longer supported by EMC IT, 36% had the ability to protect up to 500MB on network shares, and 57% had no means or ability to recover data at all. Users were creating their own ad hoc backup methods (typically with removable media) that could potentially expose EMC information outside the company infrastructure.

EMC IT knew it needed a strategy and road map for desktop and laptop protection that could accommodate both mobile and tethered end-user PCs utilizing scalable, cost-effective technologies and that would promote compliance with corporate standards and policies. Further, it needed to minimize the support overhead by providing a solution with which end users could do their own restores.

Two Avamar storage grids were used in a proof of concept (POC) with Avamar end-user client software on the desktops and laptops of a small user base of approximately 200 IT users. This POC was quickly expanded into a pilot with 800 end users, with multiple backup policies, across the Hopkinton, Massachusetts, and Cork, Ireland, campuses. EMC IT will have completed the Avamar client deployment across EMC by the end of the first quarter of 2011. Furthermore, the company will be using Avamar to protect all of its approximately 48,000 desktops and laptops.

Results

EMC has transformed its data protection environment in its core datacenters, in remote offices, and on laptops/desktops and gained measurable benefits in backup performance, recovery time, and storage optimization and experienced significant improvements in operational as well as capital costs. EMC's implementation and approach yielded the following results:

- **Remote office protection.** Overall data backup volume decreased by an average of 95% and backup times decreased by 90%, there is no network degradation, and EMC is using only 40% of the storage space previously needed. Backup failures have almost completely disappeared; the number of backup trouble tickets decreased from over 6,000 to nearly zero. The restoration success rate is now 100% with most restores taking only minutes as opposed to hours or days. Annual savings of $1 million (including $250,000 in tape media and $450,000 for offsite storage) have been realized, and eight full-time equivalent (FTE) positions that had been dedicated to backup have been redirected to other tasks.

- **Desktop/laptop protection.** Data reduction averaged 96% per backup, providing consistently faster backups than previous online service with negligible performance and network impact. Self-service restores (no IT help desk intervention) are intuitive and relatively quick. The solution has seamlessly scaled from a beta user base of approximately 200 IT users to over 23,000 current production users. EMC has closed the data protection gap; it is no longer dependent on end-user initiated, ad hoc processes and point solutions.

- **Journey to the private cloud.** Overall data backup volumes have been reduced by 50%, decreasing backup times by 75%. The number of physical tape libraries was reduced by 60%, and the number of virtual tape libraries was reduced by 53%. EMC now supports backup of over 2,000 virtual machine clients. EMC's IT administrators are taking advantage of key tools such as Data Protection Advisor and Ionix for IT Operations Intelligence to monitor, manage, and automate data protection processes.

- **Backup volumes significantly reduced.** Backup volumes decreased by an average of 95% using Avamar in concert with Data Domain systems. The amount of data protected increased by well over 40%.

- **Measureable cost savings.** Moving from an all-tape backup strategy to a mix of disk and tape reduced the cost of tape media from $1 million to $20,000 annually. In the long term, more cost savings will be realized with the elimination of tape infrastructure and media vaulting.

FUTURE PLANS

EMC backup and recovery solutions — EMC Avamar, EMC Data Domain, EMC Disk Library, and EMC NetWorker — enabled a number of significant cost and efficiency improvements while improving service levels and reducing risk in EMC's own remote office, datacenter, and desktop/laptop environments.

As EMC moves toward 100% virtualization of its x86 server environment by late 2012 with VMware, EMC's backup and recovery solutions will be key enablers in this transition because of support for both physical and virtual server environments. EMC IT will be able to more effectively scale to support business growth and enable standardized backup policies.

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:

- **Understand that not all deduplication is equal.** Deduplication technology can be deployed in several ways, such as in backup software, on backup appliances, and so forth. In addition, product maturity, architecture, and differentiated features such as data integrity assurance should be assessed. Users should outline the main problems they are trying to solve and map those challenges to the deduplication approach best suited to provide results.
- **Test proof of concept.** 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 does not work as advertised. Users should seek vendors with established products, a solid reputation, and application- or industry-specific references.
- **Assess the backup environment.** Deduplication ratios will vary based on the type of data, change rate of data, amount of redundant data, type of backup performed (full, incremental, or differential), and retention length. Understanding real-world deduplication ratios as well as the deduplication ratios obtained from the proof of concept will help ensure no surprises once the technology is in production.
- **Consider other optimization workloads.** 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 that optimization affords.