

## BUYER CASE STUDY

### Fortune 50 Bank Division Implements Deduplication Technology for Storage Optimization Across Their IT Infrastructure

Laura DuBois

#### IDC OPINION

In today's already constrained datacenters, deduplication technology yields storage footprint and cost savings. IT executives are reporting significant reductions in storage costs from several perspectives including capacity, physical footprint, power, and cooling. Additionally, deduplication is an extremely relevant and timely technology as datacenter consolidation and virtualization often result in the rearchitecture of traditional backup processes. In particular, virtualization drives changes across the IT infrastructure encompassing server, network, storage, and data protection initiatives. As firms look to replace tape with disk-based approaches to data protection, the economic advantages of data deduplication cannot be overlooked.

The leading supplier of data deduplication technology from a real-world implementation and market penetration perspective is Data Domain. Although it is difficult to define a typical installation (because Data Domain has installations across many industries and within many types of application workloads and capacities), the customer represented in this case study in many ways has characteristics that make it "ripe" for data deduplication. These include a technology refresh for backup hardware, a requirement to improve backup windows, a move to disk-based backup and archive, and initiatives to virtualize the environment.

#### IN THIS BUYER CASE STUDY

This IDC Buyer Case Study describes the implementation of Data Domain's deduplication storage systems at a division of a Fortune 50 financial services firm. In 2007, this firm evaluated deduplication technology from several vendors in an effort to optimize storage infrastructure and reduce backup disk capacity while also improving backup processes and data protection/recovery. The bank selected Data Domain to do its superior dedup ratio (in proof of concept [POC] testing and also once deployed) and starting in March 2008 did a several month "phase-in" implementation of these storage systems across 13 datacenters (both primary and disaster recovery sites).

This document provides a detailed examination of this bank's evaluation and implementation of Data Domain's deduplication technology and products (and the competitive evaluation process) within one of its divisions. The bank's IT, network, and application infrastructure are discussed as well as the IT department's objectives within this specific division. Benefits and the value from both a business and

architectural perspective of this technology deployment are examined and future initiatives are summarized. End users can use this information to benchmark whether deduplication technology is applicable within their environments and what potential hurdles might exist to deployment.

## **SITUATION OVERVIEW**

---

### **Organization Overview**

At the request of the end user, this Buyer Case Study does not include detailed corporate or organizational information. In broad terms, this Buyer Case Study represents a division of a Fortune 50 bank located within the United States with several hundred million dollars in assets.

IDC recently conducted an in-depth interview with IT personnel at this bank. The IT director and staff members that we spoke with are part of a divisional IT organization (separate from the bank's main IT staff). This IT group supports the division, running its own datacenters and supporting its own infrastructure.

Topics of discussion included a description of the bank's IT infrastructure, storage and server environment, and application portfolio. Additionally, information regarding backup windows, recovery times, ingestion performance, and deduplication ratios relative to the Data Domain implementation was sought out. Specific information on the deduplication installation includes the evaluation and implementation process; the resulting benefits yielded to the organization; and future IT initiatives across the bank's primary and recovery IT datacenters.

### **IT Overview**

The IT department of this division of the bank directly supports approximately 6,000 employees, with ancillary support for another 5,000 employees from another area within the bank. The IT infrastructure of this division spans more than 150 geographically dispersed locations in a dozen or so states from the West Coast to the Midwest. There are nine total IT personnel each with shared duties and many working interchangeably, creating a very "hands on" department. For backup services, there is a lead engineer with additional resources that provide backfill as necessary.

The environment includes 13 datacenters — many of which consist of IT "server rooms" located at a remote branch or office. All 13 datacenters have backup infrastructure with Data Domain technology installed for deduplication and CommVault installed for backup. Of these datacenters, four are considered primary sites; two are in California, one is in Idaho, and one is in Arizona. The four primary datacenters serve as "hubs" to the other smaller sites and provide Web, data recovery, and remote replication services.

The main application that this IT group supports is the file and print environment (basic file services), which consists of approximately 80 Novell servers running on HP hardware with HP storage. At the four primary sites, the architecture is three Novell servers that are clustered and attached to a SAN, while the smaller datacenters

typically have a single Novell server with internal disk storage. The storage capacity for its file and print servers is currently 14TB, growing approximately 10–20% per year.

It should also be noted that within this environment, there are approximately 250 Windows servers (characterized as one-off servers with a variety of installed applications), as well as NetApp and Hitachi storage that support the division's Oracle data warehouse applications. Data deduplication is currently being used within the Novell environment. One final note on the infrastructure is that this environment is not currently virtualized. The IT department is just starting to evaluate virtualization with a proof of concept for one of its applications.

---

## **Challenges and Solution**

The IT department has a three-year technology depreciation policy where depreciated hardware is evaluated every three years for either a refresh to newer or next-generation equipment or for a switch to new products/vendors and technologies. In late 2007, the bank's backup systems (virtual tape library [VTL] technology from Sepaton) were coming off its books, and as part of this process, the IT department decided to evaluate deduplication technology. New management wanted to move away from VTL as the team did not see any advantages of VTL over disk-based backup.

Additionally, several of the bank's applications had grown in storage capacity size to the point that they were not hitting their backup windows. The bank runs full backups (CommVault is the backup vendor) outside of regular banking business hours during a 12-hour window from 6:00 p.m. to 6:00 a.m. Some of its backup jobs were not complete by 6:00 a.m. when business was ready to resume for the next day.

With these challenges as a backdrop, deduplication technology was evaluated to help with storage capacity efficiency (how much disk storage capacity the backups were consuming) and to subsequently help the IT department meet its backup windows. The evaluation criteria for the bank was strictly based on technology (the dedup ratio) and less on budgetary factors because this refresh project was already planned from a budget standpoint based on the three-year depreciation of the bank's backup systems.

Initially, the bank planned on evaluating four vendors: Sepaton, EqualLogic, Quantum, and Data Domain. Sepaton and EqualLogic were quickly eliminated — Sepaton because of credibility issues and EqualLogic because of a disjointed sales process due to the Dell acquisition. A proof of concept for Quantum and Data Domain products was implemented in one of the bank's datacenters, with identical data running on both products. The POC test was run longer than expected, with Quantum submitting several revisions of microcode. The results came out strongly in Data Domain's favor, with a dedup ratio in the high teens/low 20s, while Quantum's results were in the low teens. According to the bank's division IT director, "When we added it all up, Data Domain came out on top in terms of performance. We were extremely satisfied with its performance in the proof of concept, and it continues to get even better on our production data."

A few other factors to note: Quantum's pricing was more aggressive (both vendors came down in price somewhat during the evaluation and negotiations), but Data Domain's better dedup results ultimately were the deciding factor. Additionally, the IT department gave Quantum high marks for ease of use based on a better graphic user interface (GUI), and both vendors received praise for ease of installation and system stability.

---

## Results

The IT department purchased 13 Data Domain systems in the January/February 2008 time frame, and the first systems were installed in March of this year. This was actually the third evolution of the bank's backup infrastructure — as the IT team went from standalone tape several years ago, then to VTL and raw, non-deduplicated disk and now to deduplicated disk-based backup. The Data Domain systems complement two systems (one of which is a backup) and 13 media agents from CommVault, and together these systems have completely replaced the technology from Sepaton and EqualLogic.

The current Data Domain environment consists of one system at each datacenter site; eight of which are the smaller 530s and five of which are the larger 565s. All of these systems do local backups for their specific sites, while the larger datacenters serve as a disaster recovery site (remote replication) for the smaller datacenters. The typical installation for the smaller datacenters is a standard fully configured one-shelf unit with expansion cards at the four larger datacenters. With respect to expansion, the bank's IT director indicated, "Within our division, Data Domain is our standard installation for our backup environment. We tell other divisions about Data Domain — we think it's really good technology and we'd like to see it grow throughout the bank."

The following specific results are reported by the IT team:

- ☒ Average dedup ratios of the backup jobs are 20:1, which exceeds the requirements (and initial expectations) by the IT team. The ratios have gotten slightly better over time because the bank has older data.
- ☒ Sustained throughput is within a range of 330GB per hour for the smaller systems up to 610GB per hour at the high end in its production environment.
- ☒ Backup windows are comparable with the windows the bank was experiencing prior to the installation of Data Domain as the backup is performed prior to deduplication (and also prior to encryption/transit/de-encryption process).
- ☒ Storage capacity savings are estimated to be roughly 3.5TB across the 13 sites.
- ☒ Recovery results have yet to be fully tested as the bank has not had to do a full recovery, but several file-based restores (55 in total since May) have all gone smoothly. SLAs for recovery are one business day, recovering to the previous business day's data.

## Future Direction

The following initiatives at the bank will potentially impact the backup/archive and deduplication environments:

- ☒ **Server virtualization.** This technology will play a significant role in the bank's IT infrastructure in the near term. Currently, the bank is in the planning stages relative to implementation, with several labs set up with VMware. The IT department has not yet made final decisions as to how the backup environment will change relative to the virtualized server infrastructure. As a default, an agent per virtualized server is a possible (but not ideal) configuration. The team continues to research possible backup configurations and will use the upcoming VMworld conference to discuss different scenarios.
- ☒ **Encryption.** The IT team would like to see encryption on Data Domain's future road map as this would provide additional security to the environment.
- ☒ **System stability.** The bank reports a number of drive failures in its newly installed systems and is working with Data Domain to resolve the issue (possibly a programming code issue). Although this has surfaced as an issue, the IT department gives Data Domain's service team very high marks as all drives have been replaced quickly and support has been very responsive.
- ☒ **Recent acquisition and growth of already installed systems.** The bank recently acquired another bank in Arkansas and will purchase a Data Domain system to support this new IT infrastructure. Additionally, the IT team will add drive shelves to already installed systems as needed, based on the growth of the business unit and associated storage capacity growth.
- ☒ **Potential expansion of Data Domain to the corporate IT environment.** Based on the recommendation from this IT division, corporate IT is evaluating Data Domain as a potential standard backup configuration across all of its datacenters. Currently, it is using an HP backup solution for its Unix environments.

## ESSENTIAL GUIDANCE

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

- ☒ **Extensive proof of concept is mandatory.** 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, other deduplication does not work as advertised, if at all.
- ☒ **Test with actual data you plan to backup.** 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 deduplication ratios you get from your proof of concept will help you with expectation setting 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. Firms should ensure that compression and deduplication (as well as encryption) are complementary and done in the right order to achieve the technology benefit it affords.

---

## Summary

In today's already constrained datacenters and growing distributed remote and branch office locations, field-proven deduplication technology is yielding significant storage cost, footprint, and power/cooling savings. As firms look to make use of disk in the data protection path, the use of deduplication is becoming a default requirement. This default technology need is being met by Data Domain, a leading if not *the* leading supplier of data deduplication solutions for the data protection market.

## LEARN MORE

---

### Related Research

- ☒ *Worldwide Data Protection and Recovery Software 2008–2012 Forecast* (IDC #212213, May 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 2008 IDC. Reproduction is forbidden unless authorized. All rights reserved.

**Published Under Services:** Storage Software; Data Protection and Recovery;  
Archiving and Hierarchical Storage Management; Storage Replication Software;  
Storage Management