EDISON Group

89 Fifth Avenue, 7th Floor
New York, NY 10003
www.TheEdison.com
212.367.7400

# White Paper

## The Emergence of a New Generation of Deduplication Solutions:

## Comparing HP StoreOnce vs. EMC Data Domain


Printed in the United States of America

First Publication: November 2011
Produced by: Renee Boucher Ferguson, Analyst; Barry Cohen, Editor-in-Chief; Manny Frishberg, Editor

# Table of Contents

# Executive Summary

The HP StoreOnce deduplication technology, launched in 2010, helps IT organizations address the challenge of protecting and recovering exponentially growing amounts of data in the face of stagnant or incrementally increasing IT budgets.

On November 29, HP launched the latest iteration of its StoreOnce deduplication portfolio. The HP B6200 StoreOnce Backup System provides enterprise-class scale-out capabilities and autonomic restart of backup jobs for high data availability. The autonomic restart feature, an important differentiator, is designed to eliminate failed backups by pairing nodes within a couplet – there are two nodes in each couplet – so the surviving node can take over when its companion node fails.

To help current and potential customers understand the value of the appliance and the StoreOnce strategy, technology research firm Edison Group compared HP's B6200 StoreOnce offering to its nearest competitor, EMC Data Domain 890 and Data Domain Global Deduplication Array. Edison considered a number of criteria that are of critical concern to today's data center IT managers in evaluating products. These include scalability (including capacity and performance), high availability, architectural approach, pricing, and licensing.

In the course of its research Edison found that HP B6200 StoreOnce meets, and in many cases exceeds, Data Domain's published specifications. Notably, Edison also found the HP B6200 StoreOnce to be the only enterprise-class deduplication appliance to offer an autonomic restart feature, which provides industry-leading availability for big-data backups.

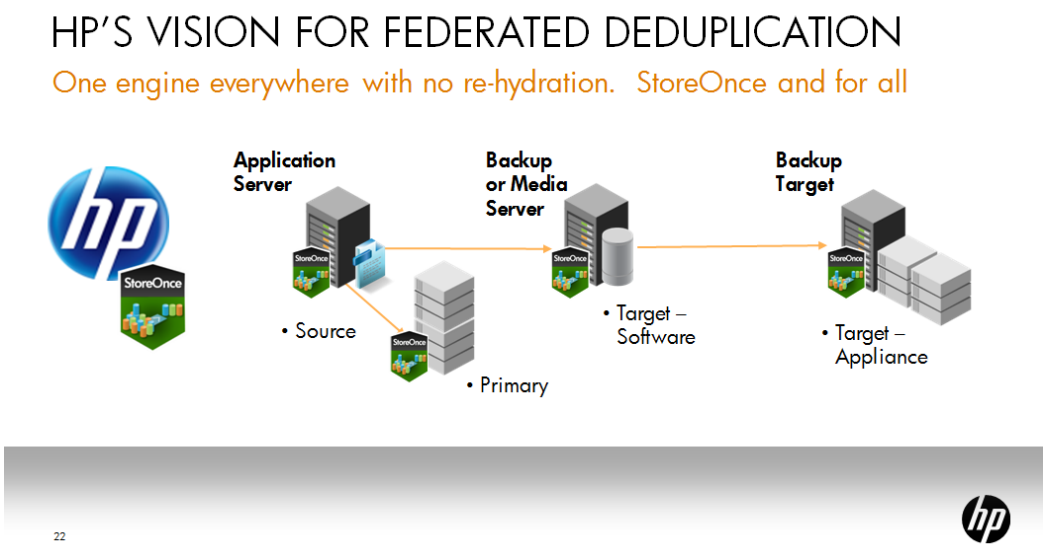
Edison believes that the HP B6200 StoreOnce appliance has leveled the playing field, providing a number of key competitive capabilities, including:

- Scale-out design that delivers enterprise class capacity, twice that of the Data Domain 890 (768TB vs. 384TB) and equivalent to the Data Domain GDA which, like the HP B6200, provides 768TB of system capacity.
- High performance for both ingest and recovery. The B6200 StoreOnce Backup System will ingest and restore at a published rate of 28TB per hour. That is faster than the published specifications of any other system.
- Automated node failover using HP Autonomic Restart to ensures high availability of data and completion of backup jobs.
- A considerable price advantage. The HP B6200 StoreOnce is comparable to Data Domain's DD890 and GDA models but costs 20 percent less.

- A simplified licensing structure that licenses applications by "target," with no associated license fees for "source sites."

Finally, Edison found HP to have a compelling Federated Deduplication vision with StoreOnce. By providing a deduplication engine that can be embedded across hardware and software platforms, it enables organizations to scale their data centers to meet increasing demands. This common deduplication engine enables data deduplicated in StoreOnce to be moved anywhere – to a data center, remote office, or a regional center – with limited bandwidth, and without having to be "rehydrated" in between deduplicated states. We believe this meets the definition of Federated Storage, which boils down to the native communication and movement of data between homogeneous systems.

With the introduction of the HP B6200 StoreOnce appliance (and the integration of StoreOnce into HP's Data Protector software), Edison believes HP has taken a significant step towards the Federated Deduplication vision launched in 2010 (illustrated below).



# Introduction

## Objective

In this white paper, Edison Group is providing an independent, third-party perspective and evaluation of HP's new B6200 StoreOnce Backup System, following its launch on November 29, 2011. Edison's objective as an independent research firm is to demonstrate the technological advantages and business benefits of the new, highly scalable HP B6200 StoreOnce solution, and to evaluate its deduplication technology versus its nearest competitor, EMC Data Domain.

## Audience

The competitive white paper is a public report offering an objective third-party perspective. It provides evidence that validates claims, and educates customers about the strengths and benefits of HP B6200 StoreOnce – particularly the value of high-availability backup solutions in the enterprise.

# Background

The one thing that is certain in IT departments today is that data will continue to increase at exponential rates. The numbers are sobering. Recent research found that by the end of 2010, the amount of digital information generated reached 1.2 million petabytes. By 2020, the amount of digital information is expected to be 44 times as much as it was in 2009.[1] Enterprise capacity, on average, is growing 40 to 60 percent, year over year,[2] as the result of a number of convergent factors. These include a massive increase of unstructured data from a multitude of sources – from email and metadata to electronic health records, audio and video files.

Facing this data explosion, IT organizations are under mounting pressure to get a handle on their storage infrastructure to meet ever-changing business requirements. Being able to store massive amounts of data (often a regulatory mandate) quickly is an imperative so enterprises can meet shrinking backup window requirements. Equally important, enterprises must to be able access backup data. In fact, a recent survey of more than 1,000 enterprises from eight countries found that 47 percent of survey respondents said their biggest data center hardware infrastructure challenge is data growth. The next closest concern is system performance and scalability, cited by 37 percent of respondents.[3]

System performance and scalability are paramount concerns in this data deluge for one reason: business continuity. There are two interdependent scenarios at play. First is the ability to backup data quickly to ensure availability in the event of a disaster. (While CIOs hope a disaster scenario never occurs, to sleep at night they need disaster recovery assurance.) Secondly, IT organizations need to be able to scale to keep up with data inundation. The ever-increasing volume of data, coupled with the need for 24/7 access to mission critical applications, has all but eliminated the almost leisurely overnight backup windows of the past. Today, companies are faced with shorter backup windows and more data to protect within those windows.

At the same time, on a day-to-day operations level, IT managers must be able to respond to evolving data-driven business demand, and to do so cost effectively. This is because, while IT budgets are remaining stagnant in the current economy, the costs of storage

[1] "A Digital Universe Decade – Are You Ready?" IDC, April 26, 2010

[2] "Data growth remains IT's biggest challenge, Gartner says," ComputerWorld, November 2, 2010

[3] Gartner Survey Shows Data Growth as the Largest Data Center Infrastructure Challenge," Gartner, November 1, 2010

continue to rise, from capital expenses to energy costs, maintenance, administration, and services.

To help combat these backup window issues, IT organizations are supplementing or replacing traditional tape libraries with speedier disk-based technologies as well as implementing source-side data reduction tools, so they send less redundant data across the network. As a result data deduplication has emerged as the fastest growing market segment in storage today.[4] Deduplication works by storing each unique data sequence only once, so only the first instance of data is retained for backup. Subsequent iterations of data are replaced by a pointer to the original data.

There are two architectural approaches to deduplication that determine how data is parsed: fixed and variable. A fixed deduplication algorithm "chunks" data into a fixed block size. Variable chunking groups the data into chunks based on patterns in the data itself. If a subsequent backup adds new information to the file or backup stream, there is a shift in the data pattern. The new information is written to disk and all other information is resynchronized and deduplicated accordingly. Variable chunking—the predominate method utilized by deduplication vendors today – has proved more effective in recognizing duplicate data where small changes merely shift the data that follows the change.

Variable chunking drives 20-times or higher deduplication ratios on average. HP StoreOnce in-line Backup System takes an additional step it calls Adaptive Micro-Chunking – chunking data into a 4K block size (the industry average is from 8K to 32K). This increases deduplication ratios through the ability to "match" smaller data more easily and quickly.

[4] HP StoreOnce video www.youtube.com/watch?v=CFTZ3NyNwPU

# Technology Solutions

## HP B6200 StoreOnce Overview

HP StoreOnce technology is part of HP's StoreOnce Backup Systems family of storage solutions. The D2D products in the StoreOnce family are an example of Converged Storage.

HP describes Converged Storage as an architecture that utilizes industry standard platforms (in this case HP ProLiant) and federated scale-out software. HP has developed a storage application stack that includes RAID management and storage pooling, along with Virtual Tape Library (VTL) and Network-Attached Storage (NAS) host protocols. HP's StoreOnce deduplication technology is integrated into this stack.

The vision, or strategy, behind HP StoreOnce modular architecture is what HP describes as "Federated Deduplication", enabled by a single deduplication engine that can be deployed across a storage infrastructure, from virtual machines and D2D backup systems (client and backup servers) to enterprise storage systems and servers. (See the Data Protector Appendix to read about the StoreOnce software integration with HP Data Protector software). The use of a common deduplication engine enables the native communication and movement of data across the various systems without rehydrating the data.

Three separate appliances have made up the StoreOnce family: the D2D25xx, D2D41xx and D2D43xx. The new HP B6200 is their enterprise-class offering. The HP B6200 StoreOnce appliance adds a number of key features to the D2D family, including industry-leading enterprise-level scale out performance – up to 28 TB/hr. – and scale out capacity that starts at 48TB and grows up to 768TB raw (512 TB usable). Significantly, the HP B6200 also adds Autonomic Restart with automated node failover, for high availability with no single point of failure. No other deduplication product currently on the market offers this capability.

Developed by HP Labs, the StoreOnce deduplication technology has a dozen patented technologies within the algorithm, including: Sparse Indexing and Container Matching. Sparse Indexing reduces memory and disk I/O requirements, enabling higher performance, while Container Matching defragments data after the deduplication process, reducing I/O and improving restore performance.

HP Labs have continued their algorithmic innovation with the B6200 StoreOnce appliance, adding a Sliding Window Assigner. This advance further improves restore

performance by placing more data in containers, multi-threaded chunking that delivers higher ingest performance, and larger containers that improve dedupe ratios for multiplexed backups. This latest development enables the B6200 to offer restore performance that equals the backup or ingest performance. As the amount of protected data grows, restore rates become more and more important. Offerings from other vendors have a significant restore-performance penalty.

## EMC Data Domain Overview

### *Data Domain 890*

EMC's Data Domain technology, acquired in 2009, is a member of EMC's family of in-line deduplication appliances. In addition to deduplication technologies from Data Domain, EMC also has Avamar-based deduplication, and deduplication appliances from Quantum. The Data Domain 890 is EMC's system for large enterprise data centers, with throughput of 8.1 TB/hour native and up to 14.7 TB/hour with DD Boost technology, which is a separate license from both DD 890 and GDA.

Data Domain utilizes EMC's Data Invulnerability Architecture that, in a sense, prioritizes data protection over performance. According to EMC, Data Domain's architecture emphasizes several key areas, including:

- End-to-end verification of backup time: data is read after it is written, and verified that it is the correct data and reachable through the file system to disk.
- Most restores happen within a day or two of backups, according to EMC.
- Fault avoidance and containment: new data does not overwrite good data, fewer complex data structures and no partial stripe writes.
- Continuous fault detection and health. Data Domain RAID-6 provides double disk failure protection and read-error correction.
- File-system Recoverability that enables the file system to be recreated by scanning the log and rebuilding it from metadata stored with the data.

### *Data Domain Global Deduplication Array (GDA)*

EMC's Data Domain GDA is EMC's largest system for enterprise data centers, in terms of scale and capacity, with throughputs of up to 10.7 TB/hour native and up to 26.3 TB/hour with additional DD Boost technology, according to EMC.

EMC says GDA presents a single deduplication storage pool to the backup application across two DD890 controllers. Multi-terabyte data sets are load-balanced across

controllers for capacity and performance management, and backup administration. GDA utilizes several EMC software packages to maximize its deduplication capabilities, including Data Domain Boost to manage file replication and lower the required replication bandwidth, Data Domain Virtual Tape Library, and Data Domain Global Deduplication software. Data Domain Replicator software enables GDA to be included in replication topologies for multiple remote sites and large data centers, according to EMC's GDA data sheet,

As demonstrated in Table 1 (below), HP B6200 StoreOnce appliance outperforms Data Domain 890 and GDA in scalability and native performance.

**Table 1: HP StoreOnce and Data Domain Product Comparison**

| | HP B6200 StoreOnce | Data Domain 890 | Data Domain GDA (2xDD890) |
|---|---|---|---|
| **Performance** | | | |
| Aggregate write (TB/hr) | 28 TB | 8.1[5]/14.7[6] | 10.75/26.36 |
| Aggregate read/restore (TB /hr) | 28 TB | Not published | Not published |
| **Capacity** | | | |
| Base system capacity (raw) | 48 TB | 16 TB | 32 TB |
| Max system capacity (raw) | 768 TB | 384 TB | 768 TB |
| Expansion shelves | 24 TB or 48 TB per couplet (+3 expansion shelve pairs with 24 or 48 TB each) | 16 TB or 32 TB | 16 TB or 32 TB |
| **FC/Ethernet Host Connect** | 8G FC, 10 GbE | 8G FC, 10 GbE | 8G FC, 10 GbE |
| Interface | CIFS, NFS, VTL (FC) | CIFS, NFS, VTL (FC), OST, DD Boost | CIFS, NFS, VTL (FC), OST, DD Boost |

[5] Native performance

[6] Performance using EMC Boost software

| | HP B6200 StoreOnce | Data Domain 890 | Data Domain GDA (2xDD890) |
|---|---|---|---|
| **Scaling** | | | |
| Max Number of controllers/nodes | 8 (4 couplets with two nodes each) | 1 | 2 |
| Supports expansion of nodes | From 2 to 8 | N/A | Yes, up to 2 |
| Single dedupe domain across system | Within each 2-node couplet | N/A | Across the 2 nodes |
| **Availability** | | | |
| No single point of failure | Yes | No | No |
| **Replication** | | | |
| Fan-in | 384 (48 per node) | 180 | 270 |
| Multi-hop | Yes | Yes | Yes |

# Matters of Scalability

As data continues to proliferate at astonishing rates, protecting critical business data is a key concern for IT organizations today. With more data to protect, IT organizations are hard pressed to meet backup windows. Tape libraries, the traditional means of backup storage, have proven to be unreliable and costly. Restores from tape can be slow, causing IT organizations to miss the mark on SLAs (service level agreements). At the same time, fragmented backup solutions add management complexity – and costs – while increased capacity requirements can result in expensive "fork-lift" upgrades.

The fact is, most medium and large companies have remote backup requirements – over 50 percent of data can reside outside of the data center – yet data growth can be unpredictable. In light of these factors, IT managers need to address data protection and recovery with a number of objectives in mind: storage cost reductions, speedier backup windows, performance, resiliency, future scalability, manageability, and, at the end of the day, less demand on IT resources.

IT organizations have found a number of quantifiable business benefits in implementing HP's StoreOnce data deduplication technology, rather than competing technologies. For example, in a published case study, Ultra Flex Packaging Corp.[7] discussed how it explored a disk-based backup system with deduplication technology to reduce storage costs and speed backup times. The company evaluated competitive backup solutions from Data Domain, ExaGrid Systems and HP, based on a number of criteria, including cost, performance and future scalability. The Brooklyn-based company ultimately chose HP StoreOnce for its ability to provide a way to grow capacity without purchasing additional systems.

Ultra Flex deployed two HP StoreOnce backup systems, each containing 4.5 terabytes of SATA disk space and StoreOnce deduplication software. One system is deployed in the company's main data center, the other in an off-site location for disaster recovery.

Since implementation, the company has found a number of IT improvements and business benefits, including:

- A 12-fold reduction in weekend full backup time, with the ability to protect more data, faster.
- Simplified administration with consolidated backup platforms.
- Five-fold faster data recovery for better business continuity.

[7] http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA3-6973ENW.pdf

- 20 hours of reclaimed productivity for design engineers.
- A decrease in upfront capacity purchases.
- Reduced storage costs through increased capacity capabilities.

Similarly, a leading professional services firm in the food services and facilities management industry had been a Data Domain customer for years. However, the firm found that it was necessary to have a multitude of products to accomplish its deduplication tasks – Data Domain in some locations, EMC Avamar in others. After some due diligence looking into the competition, the firm implemented four HP StoreOnce appliances, at 20 percent less net cost than Data Domain.
Of particular importance to this professional services firm is the fact that it is able to standardize on the HP StoreOnce technology moving forward, based on HP's Federated Deduplication vision and roadmap for StoreOnce. The firm is particularly interested in the StoreOnce-powered Data Protector deduplication client (please see Appendix for more information) and the scale-out capabilities of the B6200 StoreOnce Backup System.

With its primary concern being unmanageable backups and a complex tape environment, a Midwestern public power district went in search of a more reliable and streamlined approach to its backup environment. The district already knew it wanted to include deduplication as part of its solution so, when it released an RFP, it fully expected a response from Data Domain. It did not even consider HP, even though it was already an HP customer; the utility was under the impression HP did not have a product that could meet its needs. Fortunately, an HP partner heard about the deal and worked with HP to prepare a response to the utility's RFP. The utility, focused on cost savings as well as performance, quickly realized HP StoreOnce would meet its needs. The utility installed seven StoreOnce units in three separate locations, with replication between all three sites.

The customer found that it was able to achieve every one of its business and IT goals, while saving over $500, 000, compared to the Data Domain solution. The customer is particularly happy with the StoreOnce performance, and its ability to come in under budget expectations.

An Australian IT business solutions provider needed a data protection solution to complete their managed Cloud Platform services offering. It also required a solution that would support IBM AIX and VMware-based services with top-notch data security and rapid recovery. After considering the competition, the Australian consultancy chose HP Data Protector software and two HP StoreOnce appliances. The customer found IT and business benefits to be considerable, including automated processes for backup, restore, and copying backups to tape for longer-term retention or off-site vaulting. With

automated processes, the consultancy has achieved maximum efficiency and reduced the resource time normally required to manage backup and restore.

## What Makes HP B6200 StoreOnce Backup System Unique

In its research, Edison Group found that the HP B6200 StoreOnce Backup System brings a number of key features to the enterprise deduplication market – 2nd generation deduplication features that are available only from HP, including:

- Federated Deduplication, a single common StoreOnce deduplication algorithm across the enterprise to enable data mobility without rehydration.
- HP Labs algorithmic innovations like Predictive Acceleration and Adaptive Micro-Chunking that enable leading capacity and performance.
- High availability with Autonomic Restart, which means that the HP B6200 node failover is *automatic* and enables backup applications to retry jobs, so they always complete.
- New StoreOnce GUI and Replication Manager, which provides management capabilities in a replication environment, enabling customers to manage up to 384 D2D devices replicated across multiple sites.

## Break it Down: HP B6200 StoreOnce In Detail

### *Scale-Out, Future Ready Design*

HP B6200 StoreOnce is the company's first enterprise-class deduplication appliance that adds enough scalability – in terms of capacity and performance – to enable even the largest data centers to back up a tremendous amount of data in a short time window. Edison Group found that, with up to 768 TB of capacity and backup performance of up to 28TB/hr., the B6200 enables enterprises to meet increasingly challenging backup window requirements. This level of scalability is critically important in large scale virtual and cloud datacenters, where massive consolidation of applications and content are seen.

Comparing HP and EMC published capacity and performance results, Edison found the HP B6200 StoreOnce appliance outperformed Data Domain 890 in aggregate native write speeds (28TB/hour vs. 8.1TB/hour). It also exceeds the DD890, even when it employs DD Boost (28TB/hour vs. 14.7TB/hr). The B6200 also exceeded the Data Domain GDA in native performance (28TB/hr vs. 10.7TB/hr) and when the GDA employs Boost as well (28TB/hr vs. 26.3TB/hr).

In terms of capacity, HP B6200 StoreOnce outperformed Data Domain 890 in both base system capacity (48TB vs. 16TB) and maximum system capacity (768 TB vs. 384 TB). Similarly, the HP B6200 outperformed the Data Domain GDA in base system capacity (48TB vs. 32TB), and matched GDA in maximum system capacity (768TB). The B6200 StoreOnce Backup System is a game changer for HP in terms of its future-ready, scale-out capabilities.

Because the B6200 offers a clustered system with performance and capacity scale-out, IT organizations can now easily add additional capacity and performance within a single clustered backup system, without adding additional smaller "boxes." In fact, the B6200 improves deduplication performance six-fold, compared to HP's previous generation deduplication appliance, and improves capacity scalability seven-fold compared to the previous generation.

The idea with the HP B6200 is that organizations have the ability to start small with a single B6200 48 TB StoreOnce Backup System, consisting of two servers/nodes connected in failover configuration as a couplet. Then, they can build on top of that by adding upgrade kits to the couplet. Each upgrade kit contains two shelves and each shelf contains 24x1 TB or 2TB disks, providing 24TB or 48TB raw capacity (16-32TB usable). This allows up to 192TB of raw capacity (128 TB usable) associated with the original couplet.

Notably, capacity upgrade kits can be added to the couplet while the Backup System is online, so there is no scheduled downtime associated with adding capacity or performance.

To scale out to the top end of raw capacity available on the market, users simply add another B6200 rack (which is the same approach Data Domain utilizes to achieve its capacity rating with GDA), as illustrated in the graphic below.


**Figure 1: HP B6200 Configurations -** This image shows B6200 scalability from the base couplet of two compute and storage nodes, through a fully expanded single couplet in a rack with network switches, to the fully scaled system with four complete couplets.

This capacity increase enables users to:

- Keep pace with data growth.
- Protect over 10PB based on 20:1 deduplication – two times the DD890 capacity.
- Back up 224 TB in an eight-hour window – three times the DD890 native throughput.
- Reduce backup-window impact by streaming backups in parallel.

From a customer perspective, the scale-out design of HP B6200 essentially means that the appliance is fast, backing up a lot of data in a short period of time. In addition, customers can expand performance and capacity by adding additional couplets, rather than adding another appliance.

Overall, Edison found that the HP B6200 StoreOnce Backup System delivers performance and capacity greater than the DD890 and DD GDA. At the same time, the B6200 was found to be the only scale-out deduplication appliance currently on the market with high availability ensured by node failover with autonomic restart.

### *High Availability with Autonomic Restart*

It is an understatement to say that data protection – and, in turn, data availability – is a critical concern for business continuity today. According to the National Archives and Records Administration, 93 percent of companies that lost their data center for 10 days or more due to a disaster filed for bankruptcy within one year. Putting aside the concerns of a potential disaster and data recovery assurances, IT organizations require a high availability of data, regardless of its age or backup location, to ensure daily business challenges are met.

To enable high availability of data, an autonomic restart feature of the HP B6200 StoreOnce is designed to eliminate failed backups across nodes, controllers, cache, disks, paths and power cooling. This is enabled by pairing nodes within a "couplet" so the surviving node can take over when its companion node fails. With certain backup applications, including HP Data Protector, the autonomic restart feature within the B6200 provides the self-managing characteristics to adapt to unpredictable failure modes while hiding the intrinsic complexity of that protection from operators and users. That is, StoreOnce automatically detects certain failures and deals with them—something the competition does not.

The HP B6200 StoreOnce Backup System – the only scale out dedupe solution with automated node failover – enables no single-point-of-failure, from a number of perspectives. The automated node failover design ensures that if one node in a couplet – or failover cluster – fails, the "service set" automatically swaps to the remaining node in the couplet.

The HP B6200 dual controllers ensure that if one front-end controller – or node – within a couplet fails, all critical non-replicable aspects of that controller are moved to another controller and the failed controller is disabled (as seen in the graphic below). Cache mirroring between the RAID controllers ensure that if one controller fails, data is still written to media by the second controller.

Built-in hardware redundancy – dual-path disk arrays, internal network and power supplies throughout – ensures system availability with full redundancy across couplet nodes, controllers, cache, disks, paths, or power and cooling.

Edison found the B6200 StoreOnce Backup System to be very resilient, compared to Data Domain. Even in Data Domain's highest-end product, the GDA, the loss of one of the two nodes means customers lose access to that node's data unless it was previously replicated elsewhere. Even if that is the case, manual intervention is required.

From a customer perspective, Edison found that HP B6200 brings two features to the table that are important to ensure data protection and high availability. One, the move

from single-node in previous generations of StoreOnce to multi-node ensures high availability. Two, dual controllers in the B6200 ensure high reliability. In both instances data protection is assured through a full, clean backup —a key concern to customers in vertical industries such as financial services, where data viability is mandated.

### *Federated Deduplication*

When it launched its HP StoreOnce deduplication technology roadmap in 2010, HP advanced a new paradigm for deduplication – Federated Deduplication, an architectural approach it believes is required for second-generation deduplication or Dedupe 2.0. It addresses some of the issues inherent in existing deduplication solutions: high costs, complexity, scalability, and siloed data.

Federated Deduplication, in HP parlance, amounts to a single, common StoreOnce deduplication algorithm (developed by HP), which works across the company's hardware and software platforms. In other words, HP's StoreOnce deduplication engine can be deployed across multiple HP platforms, including: the B6000 series, the D2D 2500 and 4000 series, and the new release of Data Protector software that is integrated with StoreOnce.

Looking at EMC's approach to deduplication, Edison found they rely on products from three acquired companies (Data Domain, Avamar, and Celerra) to meet the needs of large organizations. HP, with its Federated Deduplication strategy, designed a single architecture to meet the same needs of the organization, from small data backup and recovery needs to the largest enterprise data requirements.

HP's StoreOnce architecture enables companies to move deduplicated data from device to device without rehydration. StoreOnce technology is independent of file system, operating system or hardware platform. This independence enables data deduplicated in StoreOnce to be moved anywhere – a remote office, a home office, a large data center or a regional center, using limited bandwidth links and without having to "rehydrate" the data in between deduplicated states. At the same time, smaller, deduplicated data can be stored among locations; the data only needs to be expanded when it is accessed. This provides both scale-out capabilities and high availability of data from anywhere.

The HP Federated Deduplication capability not only reduces the time that data protection process takes, it also enables data availability anywhere, from a distant location to a data center. Because data is portable in its deduplicated state, HP B6200 StoreOnce improves both backup/ingest and recovery performance rates. At the same time ingest/recovery performance is up to two times competitors' restore rates, all but eliminating the restore "performance penalty."

In contrast, Dedupe 1.0 "point solutions" are, again in HP parlance, first-generation tactical solutions that require data to be rehydrated when it is moved across different

platforms. In this scenario, the performance focus from vendors tends to be only on backup/ingest rates, ignoring recovery rates. With the waves of explosive data growth, Edison agrees with HP that Dedupe 2.0 should include designs such as Federated Deduplication to meet modern data center requirements.

### *Price Points*

Edison research found that the HP B6200 StoreOnce appliance wins out against Data Domain in the particularly sensitive area of pricing. While HP StoreOnce products have traditionally been priced lower than EMC's comparable product, Data Domain often beat out StoreOnce in its ability to scale. For example, if a customer needed to backup a large amount of data, they might be able to accomplish the task with a single DD890 appliance or, on the higher end, GDA. Before the B6200, that same customer would have had to add multiple HP StoreOnce appliances to enable the same backup.

However, even though capacity was an issue at times, HP often won out over Data Domain on pricing (and licensing) alone. This competitive pricing scenario holds true now more than ever with the HP B6200 StoreOnce appliance, particularly given HP's considerable storage might.

Edison found the HP B6200 to be comparable to Data Domain's DD890 and GDA models at 20 percent less cost. This price-performance is even more striking when one considers that the HP B6200 can be more than 50 percent less expensive than Data Domain.

Edison also found in its research that when EMC Data Domain does provide a competitive introductory price, customers are often significantly up-charged to expand or upgrade their environments. Similarly, EMC's approach to adding capacity is to "add a lot to get a little," according to one source in an EMC deal. For example, should a customer require an extra shelf of disk space, they are often required to purchase more functionality than they need, at a significant additional cost.

Similarly, Edison found that Data Domain offers customers different product lines based on different needs. For example, if a customer has a small or remote office, EMC offers a deduplication solution from within another EMC product line – Avamar or Celerra, for example. This results in a heterogeneous deduplication environment where data is deduped in various incompatible formats. In this circumstance, rehydration becomes cumbersome at best. Many vendors take this approach – tailoring sometimes incompatible product lines to meet customer needs. Edison found that HP differs is in its vision and roadmap for StoreOnce, Federated Deduplication, which utilizes one deduplication engine across platforms, enabling data availability anywhere, any time.

### *Licensing: A Matter of Simplicity*

Edison found that prior to the HP B6200 launch, one way HP StoreOnce sales won out against Data Domain was with its competitively priced, flexible licensing structure, at a significant cost savings for customers.

IT managers utilize deduplication for replicating data to consolidate operations and improve disaster recovery. However, each replication site typically requires an additional license. Because it licenses replication by target alone (not by source site and target) HP has vastly simplified the replication license structure. It has also made it very cost effective for customers — particularly those with a large number of remote offices that need to replicate data into a head office. On the other hand, EMC Data Domain is licensed per array.

When HP customers acquire a replication license HP StoreOnce Replication Manager software is included at no extra charge. Replication Manager provides basic management capabilities in a replication environment, as well as an easy way for customers to manage up to 384 D2D devices that are being replicated across multiple sites. EMC also bundles its version of replication management software with Data Domain, supporting up to 20 sites at no charge.

For customers using HP Data Protector software, HP allows unlimited backup of servers, a license model that saves customers money as they add new clients. Similarly, because online backup agents work for any supported application, Data Protector licenses can be redeployed as customers retire or change applications. Finally, because Data Protector licenses are priced significantly lower than competitive products, the savings are carried forward into annual renewals.

### *The Importance of High Availability*

According to Meta Group, over 90 percent of all mission-critical systems – including financial, ERP and CRM – store their data within a relational database. IT organizations, by necessity, have to put in place backup and recovery systems that enable high availability of that essential data. And backup, of course, is all about restore. Not only do IT managers have to assure data is backed up during allotted windows, but once data is backed up, it must be available when needed. To accomplish this many IT organizations implement deduplication appliances to enable both faster backup and availability of data.

Edison's research showed that HP B6200 StoreOnce is the only enterprise-class deduplication appliance to offer high-availability (HA). HP typically talks about couplets (each couplet has two nodes) as the basis of its HA functionality. HA, combined

with scale-out architecture, eliminates forklift upgrades and provides industry-leading availability for big-data backups. With a restore rate of 28TB per hour —faster than any stated rate —HP has the clear advantage when it comes to restoring data efficiently. This is an essential point of high-availability deduplication as companies back up more and more data quickly, and have to be able to access that data, as well.

# Conclusions

There are no accessible published test reports validating EMC performance and capacity claims. However, taking them on face value and comparing them to HP's performance and capacity claims, Edison found that both systems provide high degrees of storage utilization efficiency.

Edison also found that, with the addition of HP B6200 to its family of StoreOnce deduplication family, HP has leveled the playing field by adding a number of key benefits, including:

- Scale-out design that delivers enterprise class capacity and performance as compared to Data Domain 890 (768TB vs. 384TB) and Data Domain GDA, which like the HP B6200, provides 768TB of system capacity. The scalability of the HP B6200 StoreOnce appliance enables IT organizations to:
  - Protect over 10PB based on 20:1 deduplication – two times the DD890 capacity.
  - Back up 224 TB in an eight-hour window – three times the DD890 native throughput.
  - Reduce backup-window impact by streaming backups in parallel.
- HP's B6200 StoreOnce Backup System is currently the only solution on the market that provides automated node failover. With Data Domain's highest end product, the GDA, the loss of one of the two nodes means customers lose access to that node's data unless it was previously replicated and stored in another location. Data loss necessitates manual intervention.
- HP's Federated Deduplication vision provides a holistic, extensible deduplication strategy that organizations can utilize to meet ever-growing data demands and ever-changing business requirements. As part of this strategy, the HP StoreOnce dedupe engine can be deployed across multiple HP platforms.
- Price is a tremendous advantage with the HP B6200 StoreOnce. While this enterprise class appliance is comparable to Data Domain's DD890 and GDA models, it is available at 20 percent less cost. The B6200's price is more than 50 percent less than Data Domain. In addition to providing a compelling price advantage with HP B6200 StoreOnce, HP has provided a relatively simple replication licensing structure that licenses applications by "target," with no associated license fees for "source sites."

In conclusion Edison found that the HP B6200 StoreOnce Backup System delivers price and performance over the DD890 and DD GDA. The B6200 was found to be the only scale-out deduplication appliance currently on the market that can offer high availability ensured by node failover with autonomic restart.

# Appendix

## HP Data Protector Software: Integrated with HP StoreOnce Technology

With the HP B6200 StoreOnce Backup System appliance HP has enabled simple, fast, highly available data deduplication at even the largest data centers. However, it is the addition of the latest version of HP Data Protector to the StoreOnce Backup System family that really propels HP's vision for federated deduplication forward. The reason: Data Protector is now integrated with StoreOnce technology to provide software-enabled, target-based deduplication capabilities that can be installed virtually on any hardware.

Essentially, HP's vision for federated deduplication is that backup data should be accessible wherever it is required – data center, remote office, branch office or on physical tape – without the need to "rehydrate" data between deduplicated states. With the integration of HP Data Protector software and HP StoreOnce technology, customers have the choice of implementing a high-performance StoreOnce Backup System or software-only solution.

HP Data Protector is automated backup and recovery software that provides a single pane of glass to manage data protection in virtual and physical environments. The software simplifies data (and application) protection through a unified recovery platform and advanced automation, across both disk and tape devices. HP StoreOnce deduplication technology, on the other hand, is currently available in two ways. One is as target-based deduplication through HP StoreOnce Backup System (D2D appliance), with StoreOnce Backup System configurations managed via the Data Protector console. Also, as a StoreOnce software library that is deployed via the Data Protector console.

Target based deduplication is currently the easiest way to implement deduplication into any existing backup infrastructure. With this in mind, the StoreOnce library can be installed on any existing hardware—for example Network-Attached Storage [NAS], virtual servers or physical servers. A single StoreOnce library can also be shared among multiple clients, for example across a LAN. Or it can be centrally deployed and managed from the data center—as long as there is a WAN connection, there is no need for IT staff onsite. Backup and restore for small offices can be managed entirely remotely.

Target-side deduplication is a good fit for small data centers and offices enabling IT organizations to store more data locally, for faster restores. Data Protector is fully replication-aware and, together with HP StoreOnce Backup System appliances, provides deduplication-enabled replication between sites — for locally or geographically distributed environments. This allows organizations with multiple locations to take control of the data at its furthest outposts and bring it to the data center in a cost-

effective way. Replication enables better disaster tolerance without the operational costs associated with transporting data offsite on physical tape.

Storage optimization needs can also be met by implementing synthetic or virtual full backups, which can provide up to 95 percent reduction in space and are simple and cost-effective. For customers who are already using Data Protector Advanced Backup to Disk, synthetic and virtual full backup technology can be implemented without any additional cost —enabling customers to leverage what they already have.

Synthetic full backup consolidates the most recent and all previous incremental backups into a new full backup to reduce impact on production file systems. Data Protector enables customers to copy a synthetic full backup from disk to tape. The image is then stored on a single tape for disaster recovery. A full restore from a synthetic full backup is as fast as from a conventional full backup as there is no need to retrieve data from incremental backups.

Virtual full backups offer the same benefits as a synthetic full without copying the data. Virtual full backup performs a full backup of file systems on disk using pointers to existing data for space savings and faster backup and recovery of file systems to/from disk.

HP Data Protector is a "legacy" technology, with more than a decade on the market and, it is interesting to note, a customer base that includes nearly half of the Global 500. HP has a solid, well thought out roadmap for Data Protector integrated with StoreOnce. In the next release, for example, HP plans to offer StoreOnce based source-side deduplication that can be installed on the client or on a virtual machine at a remote location. This capability is an excellent option for customers that have multiple data protection and recovery needs, such as a really fast in-line deduplication in their headquarters and multiple offsite locations, where an appliance is appropriate, others where software deduplication is more appropriate. This single deduplication engine approach across the enterprise furthers even more HP's federated deduplication vision.

4AA3-8537ENW, Created November 2011