White Paper

HP StoreOnce Extends its Industry Leadership in Deduplication 2.0

An objective, competitive analysis
# Table of Contents

**Executive Summary** .................................................................................................................. 1

**Introduction** ............................................................................................................................... 3
  - Objective ...................................................................................................................................... 3
  - Audience ...................................................................................................................................... 3

**Background** ................................................................................................................................. 4

**Technology Solutions** .................................................................................................................. 6
  - HP StoreOnce ............................................................................................................................. 6
    - StoreOnce Catalyst ....................................................................................................................... 6
    - HP Data Protector ...................................................................................................................... 7
  - EMC Deduplication Technologies ............................................................................................... 8
    - Data Domain DD860, DD890 ...................................................................................................... 8
    - Data Domain DD990 .................................................................................................................... 9
    - Data Domain Boost Software .................................................................................................... 9

**Performance Does Matter:** ......................................................................................................... 10
**Time is of the Essence** .................................................................................................................. 13

**For Simplicity’s Sake** .................................................................................................................. 15
  - Flexibility Matters ...................................................................................................................... 15
  - Simplicity Matters ..................................................................................................................... 16
  - Order out of Chaos .................................................................................................................... 18
  - Bottom Line: Value .................................................................................................................... 18

**Dedupe 2.0: Revolution vs. Evolution** ...................................................................................... 20
  - Where HP Started ....................................................................................................................... 20
  - What’s Next for HP ..................................................................................................................... 20

**Conclusions** ............................................................................................................................... 22
Executive Summary

Disk-based backup with deduplication has emerged as a way for enterprise customers to deal with explosive data growth. Using it, they can accelerate backup and recovery processes and reduce total capacity requirements. The HP StoreOnce deduplication technology, announced in 2010, helps IT organizations provide data deduplication where it’s needed, either at a big data center or small remote office.

With the introduction of StoreOnce Catalyst in June, HP has further advanced the speed at which companies—regardless of size or locations—can protect data. By distributing processing tasks across application servers, backup servers, and target appliances, StoreOnce Catalyst enables lightening fast data backup and recovery. As an API it also rounds out and expands the StoreOnce family of deduplication solutions, which includes hardware (appliances), software (Data Protector) and partner solutions, integrated through StoreOnce Catalyst.

To help current and potential customers understand the value of the HP’s StoreOnce strategy—including StoreOnce Catalyst technology—research firm Edison Group compared HP’s StoreOnce family of products to its nearest competitor’s products, EMC Data Domain, Avamar, and NetWorker. In its research Edison considered a number of criteria critical to today’s data center IT managers, including performance (backup and recovery performance rates), availability of deduplication across globally distributed storage infrastructures, architectural approach and bottom-line value.

In the course of its research Edison found that HP StoreOnce meets, and in most cases exceeds, EMC’s published backup and recovery specifications. HP has increased the native performance capability of its StoreOnce B6200 appliance. With the June, 2012 release of StoreOnce Catalyst, Edison found that StoreOnce Catalyst itself significantly increases performance, above and beyond that of its nearest competitor, Data Domain Boost.

To pose a question that illustrates the point: What do you get when you add StoreOnce Catalyst to the StoreOnce B6200 Backup system? 100TB/hour of backup performance, which is 3.2 times faster than the EMC DD990 with Data Domain Boost. You also get 5.3 times faster data restoration. There is a reason for this: architecturally, the DD990 is a single-node appliance, while the equivalent StoreOnce B6200 Backup system is multi-node, consisting of two servers/nodes connected in failover configuration as a couplet. This means both native performance and software-assisted performance is faster than EMC’s newest – and fastest – backup appliance.
Notably, Edison research also found the StoreOnce family of products is the only solution currently available that enables deduplication at the application source, backup server, and the target appliance using the same deduplication algorithm. This enables management of the process through a single “pane of glass” and eliminates the “rehydration” of data required when you mix deduplication technologies. At the same time, HP has opened its StoreOnce family of deduplication products to a large ecosystem of partners. Benefits include:

- The ability to manage directly from the backup application with HP or an open ecosystem of ISV partners.
- To utilize StoreOnce Catalyst with HP Data Protector 7 and Symantec NetBackup now, and Symantec Backup Exec in August, 2012.
- A new StoreOnce Catalyst SDK, available to HP ISV partners.
- The ability to do all this at almost 75 percent better price/performance than EMC.

At the same time, HP has added integrated Data Protector 7 with its Autonomy IDOL 10 platform, brought over from HP’s 2011 acquisition of Autonomy Corp. This means HP Data Protector 7 is currently the only meaning-based information-protection software on the market. Edison believes this complex search capability can be an important, sustainable point of differentiation for large enterprises.

HP announced its vision for the StoreOnce Federated Deduplication technology in 2010 and has consistently executed on that path through a well thought out architectural approach developed by HP Labs. Today, HP StoreOnce can be deployed across a storage infrastructure—from virtual machines to enterprise data centers—using a single deduplication engine, as seen in the graphic below.

![HP StoreOnce Federated Deduplication](image-url)

Figure 1 - HP StoreOnce Federated Deduplication
Introduction

Objective

In this paper Edison Group is providing an independent, third party perspective and evaluation of HP’s new StoreOnce Catalyst product. As an independent research firm, Edison’s objective is to demonstrate the technological advantages and business benefits of the new, highly scalable HP StoreOnce Catalyst solution, and to compare its deduplication technology to its nearest competitor, EMC Data Domain Boost.

Audience

The competitive white paper is a public report that offers an objective third-party perspective, provides evidence that validates claims, and educates customers, as well as technical and non-technical decision makers, about the strengths and benefits of HP StoreOnce.
Background

In its oft-cited 2011 Digital Universe Study: *Extracting Value from Chaos*,¹ IDC posits that the world’s information is doubling every two years, with a staggering 1.8 zettabytes of data created last year alone. By 2020 the world will generate 50 times that amount of information and, as a result, 75 times the number of “information containers” will be needed to store it. Yet the IT staff to manage all that data will grow less than 1.5 times, hardly keeping up with the explosive data growth the world’s businesses will be expected to be managed.

This new reality of exponentially increasing digital data—in contrast to barely growing IT staffs—leaves IT organizations in a quandary, with two main data protection goals: fast backup and fast recovery. As more businesses turn to a 24-hours-a-day, 7-days-a-week business model, there is more pressure on IT organization to maintain access to data, reliably meeting SLAs, and to securely backup data within constrained backup windows. But not only is data proliferating at unheard of rates, it’s distributed around the globe, both in centers of business and in data centers.

According to a recent study² of the North American data center market by Digital Reality Trust, 92 percent of survey respondents (North American-based large enterprises) will definitely or probably expand their data centers in 2012. Of those respondents with definite plans to expand in 2012:

- 38 percent expect to expand in three or more locations.
- 92 percent plan to expand in the U.S.
- 50 percent also expect to expand in Europe or the Asia Pacific region.
- 21 percent reported plans for projects in South America.

This is the fastest pace of expansion in the six years that the survey has been sponsored. And it comes as no surprise given the current rate of data explosion. But for IT organizations—faced with shrinking budgets and shrinking backup windows—new challenges are presented. While it is a given that IT organizations of today require the

---

¹ “Extracting Value from Chaos,” John Gantz and David Reinsel, IDC, June 2011

ability to backup more data in less time, they also must have the ability to find data quickly, and to restore it quickly. Ensuring rapid access to data following an outage, or in a worse case, a disaster, not only enables business productivity to return to normal levels, it also ensures that revenue generating activities are able to resume as quickly as possible. This is a paramount concern, particularly in today’s economy.

To help ensure business continuity, IT organizations are utilizing tools and technologies that enable capacity optimization or data reduction. Data deduplication, which fits into this category, reduces the amount of data that has to be stored, often with a 20-to-1 or better reduction in the total amount of data to be stored. As a result, data deduplication has emerged as the fastest growing market segment in storage today.3

However early generation deduplication technologies—which debuted on the tech scene around 2007—offered limited backup performance and even more limited restore performance. They were designed to solve point problems and, as a result, were often isolated from each other. This meant data was reduced and expanded again, or “rehydrated,” as it moved between incompatible technology silos, forcing customers to purchase extra storage capacity and waste network bandwidth.

HP innovation—work being done at HP Labs—is overcoming the gaps left by early deduplication technologies with HP StoreOnce, which utilizes a federated deduplication approach. This approach consists of one deduplication algorithm across the organization, regardless of whether it is applied at the target, where data is stored, or at the source, close to where data is created. The StoreOnce approach is certified with all of the major backup applications, so it is easy to install the technology into existing environments.

With the introduction of StoreOnce Catalyst, HP StoreOnce is providing even deeper integration for customers who use HP Data Protector, Symantec NetBackup, or Symantec Backup Exec (supported as of August, 2012). StoreOnce Catalyst increases the speed at which IT organizations can securely backup and restore data by orders of magnitude. At the same time, it is also making it easier to deduplicate data at remote locations. A StoreOnce Catalyst SDK (software development kit) is available to enable other backup ISVs to take advantage of StoreOnce Catalyst’s speed-inducing capabilities.

3 HP StoreOnce video:
www.youtube.com/watch?v=CFTZ3NyNwPU
Technology Solutions

HP StoreOnce

HP StoreOnce technology is part of HP’s StoreOnce Backup system family of storage solutions. Four separate appliances make up the StoreOnce family: the StoreOnce 25xx, StoreOnce 41xx, the StoreOnce 43xx and the already industry leading HP StoreOnce B6200, which was released in November 2011. The StoreOnce B6200 is HP’s enterprise-class scale-out deduplication appliance.

The HP StoreOnce B6200 appliance adds a number of key features to the StoreOnce family, including industry-leading enterprise-level scale out performance—up to 100 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.

StoreOnce Catalyst

StoreOnce Catalyst is a new technology that is part of the StoreOnce family of products. An API, or licensed software interface, StoreOnce Catalyst deepens the integration between the HP StoreOnce B6200 Backup appliance and HP Data Protector, or Symantec backup software. The StoreOnce Catalyst API enables IT organizations, to more quickly backup and restore more data across more usage scenarios. The StoreOnce Catalyst API also enables IT organizations to “float” parts of the deduplication procedure and algorithm out to the client side, which means that backup transfers are deduplicated. In turn, this reduces network load on IT infrastructure, and reduces the requirements for CPU and memory on the B6200 appliance itself. The end result is magnitudes faster data backup.

This distributed deduplication capability is strongest when StoreOnce B6200 appliances are used with Data Protection 7 backup software. In this configuration, deduplication can occur at the application source, at the backup server, and at the target appliance. When Symantec backup software is used, deduplication can occur at the backup server and at the target appliance. With both backup applications, a remote office handling a relatively small amount of data can run backups over a WAN link with relatively low bandwidth. The backup can go into a big central data center without requiring an appliance at the remote site.
The StoreOnce Catalyst API accepts Open Storage OST commands from Symantec’s backup application—with no changes on the Symantec side—as shown in the graphic below. Similarly, HP StoreOnce Autonomic Restart is supported. StoreOnce Catalyst also supports many-to-one, one-to-one, multi-site and multi-hop data transfers. Likewise the StoreOnce Catalyst API supports variable data expiration control.

**HP Data Protector**

HP Data Protector is integrated with StoreOnce technology to provide software-enabled target-based deduplication capabilities that can be installed on virtually any hardware. This means customers have the choice of implementing a high-performance StoreOnce Backup system or software-only solution.

HP Data Protector 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 is currently available in two ways: HP StoreOnce Backup system (appliance), managed via the Data Protector Console and as a StoreOnce software library that is deployed via the Data Protector console.
EMC Deduplication Technologies

EMC has three separate deduplication technologies that span client (source) deduplication and storage (target) deduplication. Those solutions include Data Domain, Avamar, and NetWorker. EMC’s Data Domain (DD) technology, acquired in 2009, is a member of EMC’s family of in-line deduplication appliances for large enterprises.

NetWorker (formerly Legato NetWorker) automates backup to tape, disk-based and flash-based storage media across physical and virtual environments. It is, essentially, a data protection software suite that backs up client file systems, which can be sent to a remote NetWorker storage node or stored on a locally attached appliance, and OS environments. Avamar is EMC’s remote office backup solution. It supports source deduplication of databases and is integrated with VMware for virtual machine backup. It also exports deduplicated backup data to tape for longer-term archiving.

Edison found that, although EMC has a well-rounded complement of deduplication technologies, because they are three separate products—Data Domain, Avamar, and NetWorker—customers must utilize three separate user interfaces. At the same time, the solutions, even combined, do not match the capabilities of StoreOnce B6200 and Data Protector 7 with StoreOnce Catalyst integration.

Data Domain DD860, DD890

The DD860 is EMC’s deduplication appliance for data recovery and replication to a primary data center, with a throughput of 9.8 TB/hour and up to 1.44 TB/hour of single stream throughput, according to EMC. The Data Domain DD890 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 DD 890.

Data Domain utilizes EMC’s Data Invulnerability Architecture which, in a sense, prioritizes data protection over performance. According to EMC, Data Domain’s architecture emphasizes two key areas:

- According to EMC, most restores happen within a day or two of backups.
- File-system Recoverability that enables the file system to be recreated by scanning the log and rebuilding it from metadata stored with the data.

Edison believes these points are crucial to bear in mind, particularly when one considers the implications of data-restore performance.
Data Domain DD990

EMC’s Data Domain DD990, announced in May 2012 as a replacement for DD GDA (Global Deduplication Array), is the company’s largest system for enterprise data centers, in terms of scale and capacity. According to EMC, it offers throughputs of up to 15 TB/hour native and up to 31 TB/hour with additional DD Boost technology.

The new DD990 appliance, in addition to serving as a regular target appliance, also has the capability to migrate older backups to one or more retention tiers for long-term data retention, as does the DD860. Both appliances migrate older backups via the DD Extended Retention Software. The DD990 also combines capacity with some archiving capacity.

Edison found that HP offers similar archiving functionality using a StoreOnce B6200 Backup system with IBRIX X9000 active archiving and Data Protector or other backup applications that support Object Copy.

At the same time, with the performance enhancements to the HP StoreOnce B6200, (see Table 1, below), HP’s enterprise-class appliance eclipses DD990’s backup speeds, with or without DD Boost. The reason is simple: DD990 is a single-node appliance, while the equivalent StoreOnce B6200 Backup system is multi-node, consisting of two servers/nodes connected in failover configuration as a couplet.

Even though the DD990 controller uses the new Intel Sandy Bridge processor, which enables practically double the performance of the DD890 (15 TB/hr with OST and 31 TB/hr. with DD Boost), Edison found that the single node design is limiting to both capacity and performance, as compared to the StoreOnce B6200 Backup system, with (100 TB/hr) or without (40 TB/hr) StoreOnce Catalyst.

In other words, the single node DD990 on the latest Intel processor is still one-third the speed of StoreOnce with Scale-Out on older hardware.

Data Domain Boost Software

DD Boost is software developed by EMC, that works by moving part of the deduplication process away from the Data Domain appliance, and out to the source of where the backup starts. According to EMC, Data Domain Boost extends the backup optimization of Data Domain deduplication storage solutions by distributing parts of the deduplication process to the backup server or application client. In doing so, Boost (like StoreOnce Catalyst) increases throughput speeds, minimizes backup LAN load, and helps improve backup-server utilization.
Unlike StoreOnce Catalyst—which relies on one algorithm across the storage environment, resulting in only one version of StoreOnce Catalyst—Boost comes in two EMC-specific flavors: Data Domain Boost for Avamar and Data Domain Boost for NetWorker. The Avamar system can be deployed in two ways: single node and multi-node, where one Avamar server is designated as the utility node that handles the backup application, and the other nodes house data. Avamar is often used in remote backup scenarios.

Data Domain Boost works in NetWorker by embedding the Data Domain Boost library in the NetWorker storage node, moving some of the processing from DD systems out to the NetWorker storage node. It also allows the storage node to run a comparison leveraging the Boost library, with the DD system, to look for unique data. Since only unique data is sent to the DD system, load is decreased and performance increased.

**Performance Does Matter**

In a recent Data Growth and Protection survey, IT organizations were asked what is considered critical data, and by extension, what data is being backed up the most. The response:

- 70 percent said financial data is critical data to the organization.
- 65 percent said customer records are critical.
- 65 percent said email and other communications are critical.

While these results are not surprising, they bring to bear an important point: the data being backed up in today’s enterprises is indeed critically important. Not only is this data necessary to ensure day-to-day operations, it is imperative to ensure innovative new approaches that tackle both business and customer issues. This brings up another important point: performance matters.

This is where HP StoreOnce shines. With the use of HP StoreOnce Catalyst, IT organizations can achieve backup speeds of up to 100 TB/hour—nearly 3.2 times faster than Data Domain Boost on DD990. Large amounts of data can be restored at up to 40 TB/hour. To put this in perspective, fully configured, with StoreOnce Catalyst enabled, StoreOnce B6200 appliance can save up to 5.5-hours-per-day of backup overhead for use in business operations. That is over three extra months per year. What can IT organizations, and in turn businesses, do with three extra months of operational time per year? Innovate and increase revenue through high-value business projects.
One way to think about this extra operational time is in monetary terms. For example, assume that each hour adds $10,000 in business value, 5.5 extra hours provides nearly $20 million per year in incremental revenue. Doing that math, that’s a less than one-and-a-half-month payback period for a fully configured StoreOnce B6200 appliance with StoreOnce Catalyst.

HP’s backup speeds are enabled, in part, because of the scale-out architecture introduced with the launch of the B6200 appliance in November, 2011. This permits backups and restores to continue even if an appliance node fails. Compared with the industry-leading performance established in November 2011, HP has increased both the performance rates of the native B6200 appliance from 28TB/hr. to 40TB/hr., and of the B6200 with StoreOnce Catalyst from 28TB/hour to 100TB/hour. This puts HP StoreOnce squarely ahead of its nearest competitor, EMC Data Domain, and Data Domain Boost.

As demonstrated in Table 1 and Table 2 (below), HP StoreOnce B6200—with and without StoreOnce Catalyst—outperforms Data Domain DD890 and DD990.

**Table 1: Data Domain vs. StoreOnce Backup Performance**

<table>
<thead>
<tr>
<th>Native</th>
<th>Data Domain</th>
<th>Backup Performance</th>
<th>StoreOnce B6200</th>
<th>Backup Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>192 TB capacity</td>
<td>DD 860</td>
<td>5.1 TB/hr.</td>
<td>One couplet 192</td>
<td>10 TB/hr.</td>
</tr>
<tr>
<td>Native Performance</td>
<td></td>
<td></td>
<td>TB capacity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Native Performance</td>
<td></td>
</tr>
<tr>
<td>384 TB capacity</td>
<td>DD 890</td>
<td>8.1 TB/hr.</td>
<td>Two couplets 384</td>
<td>20 TB/hr.</td>
</tr>
<tr>
<td>Native Performance</td>
<td></td>
<td></td>
<td>TB capacity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Native Performance</td>
<td></td>
</tr>
<tr>
<td>576 TB Capacity</td>
<td>No DD model</td>
<td>n/a</td>
<td>Three couplets 576</td>
<td>30 TB/hr.</td>
</tr>
<tr>
<td>Native Performance</td>
<td></td>
<td></td>
<td>TB capacity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Native Performance</td>
<td></td>
</tr>
<tr>
<td>768 TB Capacity</td>
<td>DD 990</td>
<td>15 TB/hr.</td>
<td>Four couplets 768</td>
<td>40 TB/hr.</td>
</tr>
<tr>
<td>Native Performance</td>
<td></td>
<td></td>
<td>TB capacity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Native Performance</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Data Domain Boost vs. StoreOnce Catalyst Backup Performance

<table>
<thead>
<tr>
<th>Boost/StoreOnce Catalyst</th>
<th>Data Domain</th>
<th>Backup Performance</th>
<th>StoreOnce B6200</th>
</tr>
</thead>
<tbody>
<tr>
<td>384 TB capacity</td>
<td>DD890</td>
<td>14.7 TB/hr.</td>
<td>Two couplets 384 TB capacity</td>
</tr>
<tr>
<td>With Boost/StoreOnce Catalyst</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>576 TB capacity</td>
<td>No DD model</td>
<td>n/a</td>
<td>Three couplets 576 TB capacity</td>
</tr>
<tr>
<td>With Boost/StoreOnce Catalyst</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>768 TB Capacity</td>
<td>DD990</td>
<td>31 TB/hr.</td>
<td>Four couplets 768 TB capacity</td>
</tr>
<tr>
<td>With Boost/StoreOnce Catalyst</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Edison Group found that, with up to 768 TB of capacity and backup performance of up to 100TB/hr., the HP StoreOnce B6200 with StoreOnce Catalyst enables enterprises to meet increasingly challenging backup window requirements. This level of scalability is critically important in large scale virtual and cloud data centers, where massive consolidation of applications and content are seen.

Comparing HP’s and EMC’s published capacity and performance results, Edison found the HP StoreOnce B6200 appliance outperformed Data Domain 890 in aggregate native write speeds (50TB/hour vs. 14.7TB/hour). It also exceeds the DD890, even when it employs DD Boost (100TB/hour vs. 14.7TB/hour). The B6200 also exceeded the recently introduced Data Domain 990 in native performance (40TB/hour vs. 15TB/hour). It also outperforms DD990 even when it employs boost (100 TB/hour vs. 31 TB/hour).

Edison believes the dramatic performance increase of the StoreOnce B6200 appliance, both native and the StoreOnce Catalyst, provides customers with a deduplication solution that helps to combat rapid data growth in the face of constantly shrinking backup windows.
Time is of the Essence

Natural disaster damages for 2011 were five times higher than the past 10 years’ average, and doubled 2010’s total of $130 billion. And 2012 is turning out to have its share of natural disasters as well, including flooding in Australia and the deadly tornado outbreak that occurred over a large section of the southern United States in early March. The reason these statistics are important is that regardless of where a company or data center is located, there is a potential for data loss. With data being absolutely essential to companies—and indeed the economy as a whole—the importance of ensuring the rapid recovery of enterprise data in the event of an outage, or a disaster, is paramount.

The HP StoreOnce B6200 appliance mitigates risk by deduplicating data anywhere, and by creating consistent disaster recovery copies. It enables companies to manage disaster recovery utilizing a single pane of glass, and provides disaster recovery options, including any-to-any data movement across multiple sites.

More specifically, the StoreOnce B6200 appliance adds a Sliding Window Assigner that improves restore performance in three ways. It places more data in containers, provides multi-threaded chunking that delivers higher ingest performance, and larger containers improve dedupe ratios for multiplexed backups. This is important because as the amount of protected data grows, restore rates become more and more important. For a full understanding of how B6200’s restore stacks up against Data Domain’s restore rates, please see the table below.

Table 3: Data Domain vs. StoreOnce Restore Performance

<table>
<thead>
<tr>
<th></th>
<th>Data Domain</th>
<th>Restore Performance</th>
<th>StoreOnce B6200</th>
<th>Restore Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>192 TB capacity</td>
<td>Fully configured DD 860</td>
<td>Not published</td>
<td>One couplet 384 TB capacity Native Performance</td>
<td>10 TB/hr.</td>
</tr>
<tr>
<td>Native Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>384 TB capacity</td>
<td>DD 890</td>
<td>Not Published</td>
<td>Two couplets 384 TB capacity Native Performance</td>
<td>20 TB/hr.</td>
</tr>
<tr>
<td>Native Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>576 TB Capacity</td>
<td>No DD model</td>
<td>Not Applicable</td>
<td>Three couplets 576 TB capacity</td>
<td>30 TB/hr.</td>
</tr>
<tr>
<td>768 TB Capacity</td>
<td>DD 990</td>
<td>Not published</td>
<td>768 TB capacity</td>
<td>40 TB/hr.</td>
</tr>
</tbody>
</table>

### Table 4: Data Domain Boost vs. StoreOnce Catalyst Restore Performance

<table>
<thead>
<tr>
<th>Boost/StoreOnce Catalyst</th>
<th>Data Domain</th>
<th>Backup Performance</th>
<th>StoreOnce B6200</th>
<th>Backup Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>384 TB capacity</td>
<td>DD 890</td>
<td>Not Published</td>
<td>Two couplets 384 TB capacity</td>
<td>17 TB/hr.</td>
</tr>
<tr>
<td>With Boost/StoreOnce Catalyst</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>576 TB capacity</td>
<td>No DD model</td>
<td>Not Published</td>
<td>Three couplets 576 TB capacity</td>
<td>25.5 TB/hr.</td>
</tr>
<tr>
<td>With Boost/StoreOnce Catalyst</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>768 TB Capacity</td>
<td>DD 990</td>
<td>Not published</td>
<td>Four couplets 768 TB capacity</td>
<td>34 TB/hr.</td>
</tr>
<tr>
<td>With Boost/StoreOnce Catalyst</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For Simplicity’s Sake

A recent survey of CIOs found that capacity demand is the top factor driving 37 percent of all data center enhancements—more than doubled from just 16 percent in 2011. Additional capacity requirements are being driven by data center investments in cloud computing, virtualization and power management technologies, not to mention the data proliferating throughout the world’s enterprises at a fantastic rate. The bottom line—with sometimes overwhelming capacity demand, IT organizations must strive for hardworking backup systems that are both capable and simple. This means having infrastructure that has the ability to backup a lot of data quickly, simply, flexibly, and (relatively) inexpensively.

Flexibility Matters

The days of most employees going to work at a central office are over. Today’s enterprises are global, highly distributed, and require access to data 24x7. In response, IT administrators must have inherently flexible systems that enable them to backup and restore data from where it is to where it needs to be—from anywhere, to anywhere.

The beauty of the HP StoreOnce appliances with Data Protector backup software is that deduplication can occur anywhere in the data protection process—at the application source, at the backup server or at the target appliance—using a single deduplication algorithm. The StoreOnce B6200 appliance, for example, works well in a big central data center, where StoreOnce Catalyst (integrated with Data Protector 7) provides remote office backup capabilities.

Edison believes a tremendous differentiator for StoreOnce Catalyst is that it provides IT administrators the ability to do branch office backup on the application server—without the aid of an appliance—using Data Protector 7. This in turn provides a local copy of data for fast restore when necessary, using the software target in Data Protector 7. To accomplish the equivalent with Avamar would require a DD “box” in each remote site.

As mentioned, this flexibility is enhanced with StoreOnce Catalyst’s ability to work with other backup solutions, including Symantec NetBackup currently and Backup Exec in August, 2012. The SDK opens up the StoreOnce environment even more.

Edison found the StoreOnce approach differs from Data Domain in a very specific critical area: Avamar and Data Domain are not fully integrated, making communication between the systems difficult. Essentially Avamar integrates with Data Domain through Data Domain Boost. In this scenario Boost functions in two ways: as a deduplication plug-in to support backup applications, and as an OST integrator. This integration allows Avamar a foothold into larger enterprise-class environments—Boost provides scalability. However, client-side deduplication is sacrificed with Data Domain integration, as dedupe is performed inline at the Data Domain appliance. At the same time, replication of the Avamar data housed on a Data Domain appliance can only be sent to another Data Domain appliance—and still can’t be scheduled separately.

**Simplicity Matters**

Despite a massive trend toward data center consolidation over the past several years, shrinking a data center footprint doesn’t mean shrinking capacity. HP found this out in its own consolidation efforts. In 2010 HP completed a consolidation effort that brought 85 data centers in 29 countries down to six data centers in the U.S., reducing the number of servers by 40 percent (while delivering a 250 percent increase in processing power). However, HP found that, because of consolidation, its databases were getting larger and larger. This presented a challenge in backup and recovery scenarios.

Lowell Dale, a technical architect in HP’s IT organization, explained how HP overcame this issue in a 2010 interview.⁶

> “In our environment at HP, we’re running about 93,000-95,000 backups per week with an aggregate data volume of about 4 petabytes of backup data and 53,000 run-time hours. That’s about 17,000 servers worth of backup across 14 petabytes of storage. It’s pretty much every application that HP’s business is run upon. It doesn’t matter if it’s enterprise warehousing or data warehousing or if it’s internal things like payroll or web-facing front-ends like hp.com. It’s the whole slew of applications that we have to manage.

> “One of the first things we had to do was simplify, so that we could scale to the size and scope that we have to manage. You have to find and simplify configuration and architecture as much as possible, so that you can continue to grow out scale.”

---

⁶ “HP Data Protector, a case study on scale and completeness for total enterprise data backup and recovery,” Dana Gardner, ZD Net, June 15, 2010
Part of that simplification process for HP is using its own technology, including HP Data Protector, which enables HP to manage and safeguard petabytes of storage per week across its data centers. One of the helpful features in Data Protector, according to the Dale interview, is autonomic failover, which helps keep backup success rates up and tickets down.

What is important about this example is that HP StoreOnce technology is available as a StoreOnce library that is deployed via the Data Protector software. StoreOnce Catalyst is available with Data Protector 7, which increases the backup scenarios where StoreOnce and Data Protector can be utilized. The graphic below provides a good example of how HP StoreOnce Catalyst is used in a typical multi-site implementation.

StoreOnce Catalyst-compatible backup applications also provide a single pane of glass to manage all data movement. This centralized management across the enterprise helps to eliminate failed backups and business risk at remote sites, while eliminating complexity. It also means that backup applications can control all backup and replication between StoreOnce devices, which means IT organizations spend less time managing infrastructure.

Similarly, IT organizations are able to simplify operations with management integration into backup applications, enabling IT administrators to manage backups directly from an HP backup application, or through an open ecosystem of ISV partners.
Order out of Chaos

That enterprises are inundated with data is well documented, here and throughout the blogosphere. But what happens in a scenario where data must be retrieved quickly and accurately and there are potentially petabytes of data to wade through? For instance, in the event of a lawsuit, where specific data is mandated. With the acquisition of Autonomy in 2011 and the subsequent integration of Data Protector 7 with the Autonomy IDOL 10 platform, HP is delivering functionality that traditional backup software can’t begin to match.

Autonomy makes software that processes human information, or unstructured data—including social media, email, video, audio, text and Web pages. Importantly, this provides “meaning based” search capabilities that enable Data Protector 7, integrated with the IDOL platform, to discern the difference between apple the fruit and Apple the company when searching across various file types, regardless of language.

Data Protector 7 is powered by Autonomy’s IDOL platform that, through adaptive pattern recognition technology and probabilistic modeling, forms a conceptual and contextual understanding of digital content. What that means for IT administrators utilizing Data Protector 7 is that they can recover the files they want—or need—by content, rather than by file name. This search feature can be a critical time and liability saver in eDiscovery scenarios.

HP Data Protector 7 is currently the only meaning-based information protection software on the market (with Data Protector 7 customers can also acquire additional Autonomy modules). Edison believes this search capability can be an important, sustainable point of differentiation for large enterprises.

Bottom Line: Value

In previous research Edison found the HP B6200 to be comparable to Data Domain’s DD890 and GDA models at 20 percent less cost. However, with the June release of StoreOnce Catalyst, HP is upping the ante on price/performance. The StoreOnce B6200 now has up to 75 percent better price/performance than DD890, according to HP.

Similarly:
- Extended simplicity reduces expenses with a pay-as-you grow, manpower-efficient architecture. Addressing the limitations in Dedupe 1.0 products, the StoreOnce B6200 utilizes scale-out architecture. This means that StoreOnce customers can start small and grow in more easily digestible increments preventing expensive forklift upgrades.
• Edison found in early feedback from customers that they are particularly excited about StoreOnce Catalyst from a cost perspective. Backing up remote office data using Avamar is not cheap. Customers see StoreOnce Catalyst as a welcome competitor into that market.

• Similarly, StoreOnce Catalyst enables branch office backup on the application server utilizing Data Protector 7, which provides a local copy of data for fast restore. Avamar, conversely, requires a Data Domain appliance at each remote site for the same functionality.

• StoreOnce Catalyst licensing is competitively priced next to Avamar. All of the functionality enabled by StoreOnce Catalyst only requires a license for each StoreOnce Catalyst-enabled couplet. At the same time, the B6200 and Data Protector prices have not increased with the launch of StoreOnce Catalyst.

Finally, Edison found in previous research that, when EMC Data Domain does provide a competitive introductory price, customers are often significantly up-charged to expand or upgrade their environments. 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.
Dedupe 2.0: The Evolution

Where HP Started

In 2010 HP launched its Federated Deduplication vision, borne out in the HP StoreOnce modular architecture. This enables a single deduplication engine to be deployed across a storage infrastructure, from virtual machines to enterprise storage systems and servers. With the introduction of the HP StoreOnce B6200 appliance (and the integration of StoreOnce into HP’s Data Protector software) in November, 2011, HP took a significant step toward rounding out that initial Federated Deduplication vision.

Now, with the enhanced native output of the B6200 and the introduction of StoreOnce Catalyst, Edison believes HP has taken another huge step in fulfilling its “federated deduplication” vision. By using a single dedupe algorithm, the HP StoreOnce deduplication technology helps IT organizations protect and recover exponentially growing amounts of data—at every step along the data protection continuum, even in the face of stagnant IT budgets.

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.

Edison found this federated “Dedupe 2.0” vision to contrast with the “Dedupe 1.0” approach taken by HP’s nearest competitor, EMC, whose deduplication strategy to date has been to bolt together acquired technologies—from Data Domain to Avamar to NetWorker.

What’s Next for HP?

As HP continues to deliver on its Federated Deduplication vision, it is also looking to the next step in data protection. With the move towards meaning-based information protection through the acquisition of Autonomy, HP is pushing the boundaries of its federated deduplication strategy beyond just deduplication.

Autonomy’s IDOL 10 platform determines contextual meaning regardless of format, language, location, or quantity of data. In doing so, IDOL 10 is able to detect patterns,
emotions, sentiments, intent, risks, and preferences as they happen. Put in a data protection context, companies have the ability to contextually search data archives in a manner never before deemed possible. This opens up a world of potential not only for customers, but for the HP StoreOnce family of deduplication products, as well.

At the same time, HP is broadening the horizons of StoreOnce technology, and StoreOnce Catalyst in particular. In addition to faster native processing power in the HP StoreOnce B6200 appliance, StoreOnce Catalyst—essentially an integration API—connects the B6200 and Data Protector 7, and integrates with other backup products being in use in the data center.

Finally, as mentioned previously, HP has opened up its StoreOnce Catalyst integration to partners with a (no cost) SDK.

---

7 Autonomy IDOL 10 Product Brief, 20111202_PI_PB_AUTN_IDOL10_Web.pdf
Conclusions

There are no accessible published test reports validating EMC performance claims. However, taking them on face value, Edison found that, compared to HP’s performance claims for the native performance increase to the StoreOnce B6200 appliance, and with the addition of StoreOnce Catalyst, HP StoreOnce outperforms Data Domain, with and without Boost.

Importantly, the addition of the StoreOnce Catalyst also makes HP the only solution currently on the market that enables deduplication at the application source, backup server, and the target appliance using the same deduplication algorithm, enabling management through a single pane of glass.

With this June, 2012 release of StoreOnce Catalyst, HP has added a number of additional key benefits, including:

- The ability to manage directly from the backup application with HP or an open ecosystem of ISV partners.
- The ability to utilize StoreOnce Catalyst with HP Data Protector 7 and Symantec NetBackup now, and with Symantec Backup Exec in August, 2012.
- A new StoreOnce Catalyst SDK available to HP ISV partner
- Doing all of this at almost 75 percent better price/performance than EMC.

Finally, HP has added integrated Data Protector 7 with its Autonomy iDOL 10 platform, brought over from HP’s 2011 acquisition of Autonomy Corp. Edison believes this complex search capability can be an important, sustainable point of differentiation for large enterprises.