White Paper

EMC Bridges the Data Protection Gap

By David A. Chapa

January, 2011

This ESG White Paper was commissioned by EMC and is distributed under license from ESG.

© 2011, Enterprise Strategy Group, Inc. All Rights Reserved
Contents

Expansion and Growth.........................................................................................................................3
Relentless—and Infamous—Data Growth..........................................................................................3

EMC Helps Solve Customer Challenges ............................................................................................6
Data Domain Platform Refresh...........................................................................................................6
Data Domain Archiver.......................................................................................................................7

The Bigger Truth..................................................................................................................................8
Expansion and Growth

Relentless—and Infamous—Data Growth

One need not look far in the storage industry to find a headline, blog post, or white paper referencing the continued growth of data and its impact on customers' ability to protect and retain data effectively. ESG research echoes those sentiments: 58% of respondents to a recent survey see annual data growth between 11% and 30% annually with 28% citing annual data growth rate of more than 30% per year.¹

![Figure 1. Annual Data Growth Rates](source)

At approximately what rate do you believe your total volume of data is growing annually? (Percent of respondents, N=510)

To put this into perspective, 36% manage between 11 and 100 TB of data in their data centers, so a median annual growth rate of 20% for a customer who is managing 20 TB of data means adding a whopping 4 TB of data in one year. Project this out five years and by year four this customer has more than doubled their current data under management; by year five, they are nearly at 50 TB: not quite the proverbial hockey stick, but it is fairly close. Figure 2 is a simple chart visually representing 20 TB of storage with annual growth rates of 20, 30, and 40%.

![Figure 2. Sample Projected Data Growth](source)

For some, 2 TB of data may not seem like much, but to a customer who must manage, protect, and retain it, it certainly becomes a very serious issue. Incidentally, 31% of respondents to the same survey are managing more than 100 TB.

Backup configurations that once completed jobs within their windows are now struggling to meet expectations. Even at 10 GbE, the achievable speeds are just about a gigabyte per second which at the modest 20 TB with no hiccups could finish in six hours. However, by year five with a 20% growth rate, the backup job jumps to over 12 hours, the 30% growth to almost 23 hours, and the 40% growth to over 32 hours. No longer can IT be focused purely on the “speeds and feeds” of yesterday. Clearly, this is a discussion around the speed AND efficiencies one can create within the backup ecosystem.

Data growth is compounded by the need and/or desire to retain data longer to satisfy various requirements such as long term archive/retention, legal hold, internal governance, and regulatory compliance or simply to create a broader range of recovery options by retaining backups on disk for longer. Disk-based backup technologies that include deduplication have eased the pain felt by some customers who find themselves sorting through the aforementioned challenges. In fact, thanks to capacity-optimized disk-based backup solutions, more customers are retaining backups longer on disk before the backups expire or the data is migrated to a removable medium. As much as 27% keep backups on disk for one to six months and 20% are keeping backups for more than 12 months.
It’s no surprise that backup, recovery, and archive continue to be top of mind for customers from year to year. It can be hard to keep pace with these growth trends given requirements and solution capabilities. The story of customer challenges continues to be muddied when considering the digital archive market. ESG’s forecast for this market estimates the total worldwide archive capacity in the commercial and government sectors will increase from 33,217 petabytes (PB) in 2010 to 302,995 PB in 2015—a 56% compound annual growth rate (CAGR).

Certainly, “archive” can mean a lot of things to a lot of people, but suffice it to say that archiving is the act or process of keeping data or a series of data as it existed at a particular point in time. In some cases, this process includes removing that data or series of data from the primary storage area after it is archived. Though there are some purpose-built software applications designed to archive very specific data types, many users may lean toward taking advantage of their backup applications to archive some data for a period of time. While some may say this is not true archiving, it does serve a purpose for a customer on a budget faced with an immediate challenge for long-term retention of data. This approach will not necessarily support the immutability of data in the “archive,” but it does let the customer manage it, which is something they wouldn’t necessarily have been doing previously.

With the exception of the digital market forecasts, this information is not brand new: research over the past 10 years or more has been showing that data continues to grow. No matter what superlative is used to describe the heightened growth, the fact is customer challenges are consistently exacerbated by this phenomenon from year to year. Customers should be asking what vendors are doing to not only mitigate, but solve some of the problems associated with this ever-present trend.
EMC Helps Solve Customer Challenges

Shortly after its acquisition of Data Domain, EMC formed the new EMC Backup Recovery Systems Division. For many, this new division was perceived as EMC getting very serious about data protection and how the company relates to customers through its portfolio. The EMC backup and recovery solutions portfolio includes:

- Data Domain
- Avamar
- NetWorker
- Disk Library
- Data Protection Advisor

EMC has been able to tighten its message to customers who may be interested in rethinking their existing backup and recovery strategies. In a relatively short time, EMC completed acquisition of Data Domain and delivered on promises of tighter integration between products with the first being Data Domain integration with NetWorker, quickly followed by the most recent announcements from EMC.

- Data Domain platform refresh
- Data Domain Archiver for long term retention

EMC announcements remain consistent with the vision presented at the 2009 EMC World: to continue to unify and improve the products in its portfolio with integrations across solutions, creating higher value for customers and helping achieve their data protection goals.

Data Domain Platform Refresh

EMC touts Data Domain systems as the industry’s fastest deduplication storage systems. While it may not be all about “speeds and feeds” today, it is important to know the solution designed to store your backup jobs will not be the bottleneck in your performance. The faster data can be ingested, the faster it can be protected and the backup window closed.

From its introduction in 2004, the Data Domain SISL (Stream-Informed Segment Layout) architecture has been at the core of its linear performance scale and with the addition of the Nehalem processor in the new Data Domain DD860 and the Westmere processor in the new Data Domain DD890, EMC claims that performance has been driven up to 9.8TB/hr and 14.7TB/hr, respectively.

With the new Data Domain Global Deduplication Array, EMC is posting performance numbers with dual DD890 controllers of up to 26.3TB/hr while scaling to a logical capacity of 28.5 PB. EMC also announced VTL support for the Global Deduplication Array, enabling easy integration with all leading backup applications. For many larger customers, performance and scale have been the gating factors keeping them from adoption of these solutions. Now the largest customers who have been previously tentative about looking at disk-based solutions to replace their very tape-centric operations can entertain the new GDA with VTL to either eliminate or drastically reduce their dependence on tape.

In addition, many midsized enterprises rely on IBM i operating environments and have been seeking a better way to protect these systems. EMC now offers Data Domain VTL support for IBM i platforms. Data Domain device support ranges from the DD610 to the DD890, giving a broad range of options based on performance requirements, retention needs, and data protection objectives. The Data Domain systems can still be co-shared with other applications for consolidated protection of open systems and IBM i environments which alleviates the need for multiple point solutions and reducing a customer’s investment.
Customers never want the solutions designed to protect their data to become a bottleneck. The performance scale in this recent refresh means customers can eliminate any notion of a performance bottleneck in their backup environment by introducing one of the Data Domain devices. Now they can focus on maximizing performance, whether that means adding more backup servers to increase aggregate throughput or updating and tweaking their existing network environments to get the most optimal performance possible from their existing servers and Data Domain devices.

### Table 1. EMC Data Domain Performance and Capacity

<table>
<thead>
<tr>
<th></th>
<th>DD140</th>
<th>DD610</th>
<th>DD630</th>
<th>DD670</th>
<th>DD860</th>
<th>DD890</th>
<th>GDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed (DD Boost)</td>
<td>490 GB/hr</td>
<td>1.3 TB/hr</td>
<td>2.1 TB/hr</td>
<td>5.4 TB/hr</td>
<td>9.8 TB/hr</td>
<td>14.7 TB/hr</td>
<td>26.3 TB/hr</td>
</tr>
<tr>
<td>Speed (Native)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>450 GB/hr</td>
<td>675 GB/hr</td>
<td>1.1 TB/hr</td>
<td>3.6 TB/hr</td>
<td>5.1 TB/hr</td>
<td>8.1 TB/hr</td>
<td>10.7 TB/hr</td>
</tr>
<tr>
<td>Usable Capacity</td>
<td>0.86 TB</td>
<td>Up to 3.98 TB</td>
<td>Up to 8.4 TB</td>
<td>Up to 55.9 TB</td>
<td>Up to 142 TB</td>
<td>Up to 285 TB</td>
<td>Up to 570 TB</td>
</tr>
<tr>
<td>Logical Capacity</td>
<td>9-43 TB</td>
<td>40-195 TB</td>
<td>80-420 TB</td>
<td>0.6-2.7 PB</td>
<td>1.4-71 PB</td>
<td>2.9-14.2 PB</td>
<td>5.7-28.5 PB</td>
</tr>
</tbody>
</table>

*Source: Enterprise Strategy Group, 2011.*

With this refresh, EMC states that Data Domain systems are the fastest inline deduplication solutions on the market today, giving customers the flexibility of storage expansion and keeping pace with network performance requirements.

**Data Domain Archiver**

EMC introduced Data Domain Archiver, a bridge solution for customers who are dealing with longer data retention periods but want to avoid moving data to tape. As discussed earlier, customers are retaining data longer on secondary storage for a variety of reasons; DD Archiver is EMC’s response to this trend.

DD Archiver’s architecture is familiar as it uses a standard Data Domain DD860 controller, but it is cost optimized for long term retention and scales to a logical capacity of 28.5 PB while maintaining the performance characteristics of the DD860 with up to 9.8 TB/hr ingest. The solution provides the opportunity to create policies to move aging data from an “active tier” to an “archive tier” of storage managed under the DD Archiver controller. For example, data on the active tier may remain for 90 days—on day 91, data is moved to an archive unit in the archive tier. Once this unit is full, it becomes “sealed” for fault isolation but remains online for retrievals and a new archive unit is used for aging data. The tiers are independent logical storage units managed by the single Data Domain controller and namespace, driving the cost per GB with a DD Archiver down as low as 10 to 50 cents per GB.

This solution addresses a longer term retention gap that today is filled either by tape solutions or more expensive disk-based backup solutions. A very nice, cost effective model taking advantage of the technology already found in the Data Domain systems, a customer may optionally deploy replication for Disaster Recovery or Data Domain Retention Lock software for file locking to satisfy IT governance and compliance policies. EMC continues to look at the future of this product, including the introduction of drive spin-down and more aggressive data reduction techniques.
The Bigger Truth

This is a big announcement for EMC; it shows the company is making strong progress with its backup and recovery portfolio. Customers are faced with ever-worsening challenges compounded by data growth management issues and are clearly looking for a better way to not only protect their data, but manage it without going broke or getting lost in the process. Many solutions have served customers well and aptly met their needs, but there is a new dawn breaking and the solutions that once closed the gap in the past are now falling short. Many customers would like to redesign their data protection architectures if given the opportunity; this means looking at how all solutions— including disk and tape—are used and how they could contribute to the success of a new data protection strategy.

Recovery time is no longer just a bullet point on a slide: it is now truly part of the service level agreements to which customers must adhere to support their organizational needs. The good news for those who continue to evaluate the use of both disk and tape in their backup/recovery/archive environment is that EMC stands ready to help with the transition. Holding true to its stated vision and mission, EMC is filling the gaps in its portfolio through smart acquisitions and even smarter integrations that address customer needs. The EMC Backup and Recovery foundation has been laid and, from what EMC has demonstrated thus far, the building looks to be very strong.