

White Paper

# The Economic Value of Data Domain

Validating the Cost Efficiency of Data Domain Solutions

By Vinny Choinski, Senior ESG Lab Analyst;  
and Jason Buffington, Principal Analyst

May 2017

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



---

## Table of Contents

Table of Contents .....	2
Introduction .....	3
Data Domain Architectural Benefits .....	4
Data Domain Economic Benefits.....	5
Deduplication Efficiency Does Matter .....	6
Speed and Scalability .....	8
Cost Savings Considerations .....	9
The Bigger Truth.....	10

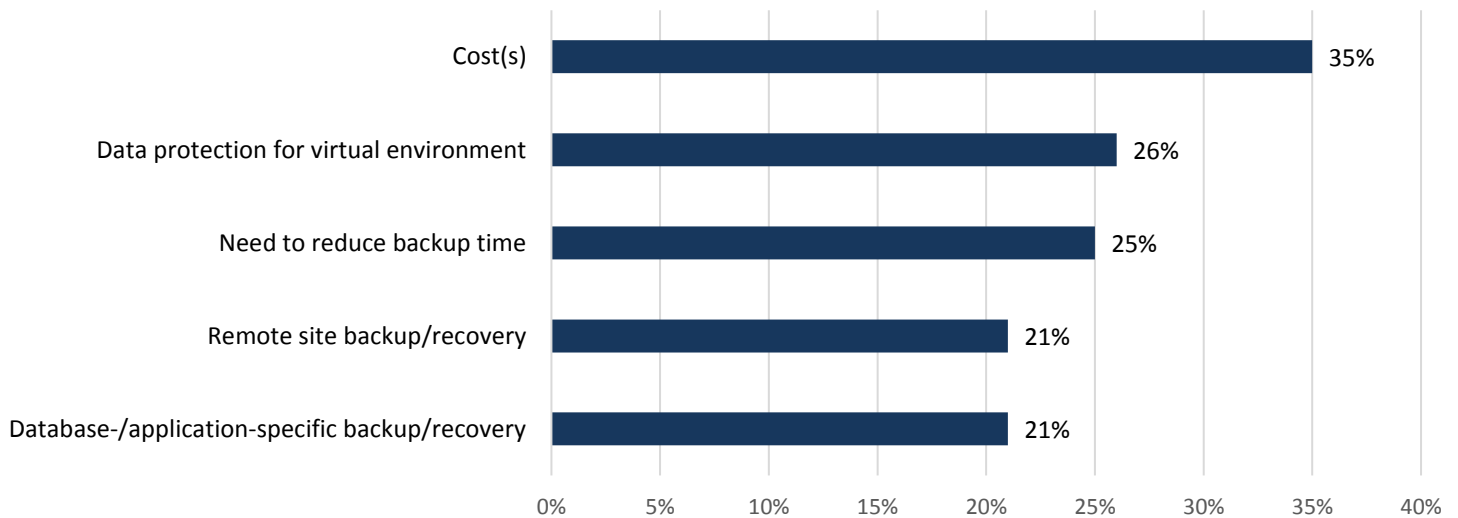
## Introduction

To out-innovate and out-pace their competition, organizations must be on a consistent path to keep their infrastructure modern. IT is under constant pressure to deliver optimized infrastructure for new business initiatives and supporting applications all while trying to contain or even reduce costs. In fact, respondents to ESG’s ongoing research consistently cite cost reduction as one of the top business drivers affecting their IT spending. When asked in a recent survey how their organizations intended to contain costs in the coming year, 27% of respondents said that they will be purchasing new technologies with better ROI.<sup>1</sup>

To drill down on one specific group, 35% of IT managers tasked with implementing data protection processes and technologies routinely cite cost as one of their top challenges (see Figure 1). This group seems to find workload-centric issues, particularly challenges tied to protecting virtualized environments and remote offices, especially problematic as well. However, when looking at the overall rankings of data protection challenges, a bigger story emerges. Costs and virtualization are top concerns, but the next three reported considerations are both *operational* and *functional* in nature. They are challenges related to performance, distributed architecture backup and recovery, and database and application protection.<sup>2</sup>

**Figure 1. Current Data Protection Processes and Technology Challenges**

**Which of the following would you characterize as challenges with your organization’s current data protection processes and technologies? (Percent of respondents, N=387)**



Source: Enterprise Strategy Group, 2017

These challenges are also consistent with the top mandates from IT leaders, which relate to better speed, agility, reliability, and cost control, all of which have been consistently reported in previous ESG data protection modernization reports. ESG analysis confirms that inconsistencies between what implementers struggled with and what leaders mandated often resulted in very little actual modernization or transformation occurring. Implementers were not able to act on leaders’ mandates due to technical impediments in their data protection environments.<sup>3</sup>

This paper discusses how Data Domain systems, and integrated solutions built on the Data Domain architecture, such as the Dell EMC Integrated Data Protection Appliance, help deliver the agility IT implementers need to transform their

<sup>1</sup> Source: ESG Research Report, [2017 IT Spending Intentions Survey](#), March 2017.

<sup>2</sup> Source: ESG Research Report, *Data Protection Modernization*, to be published.

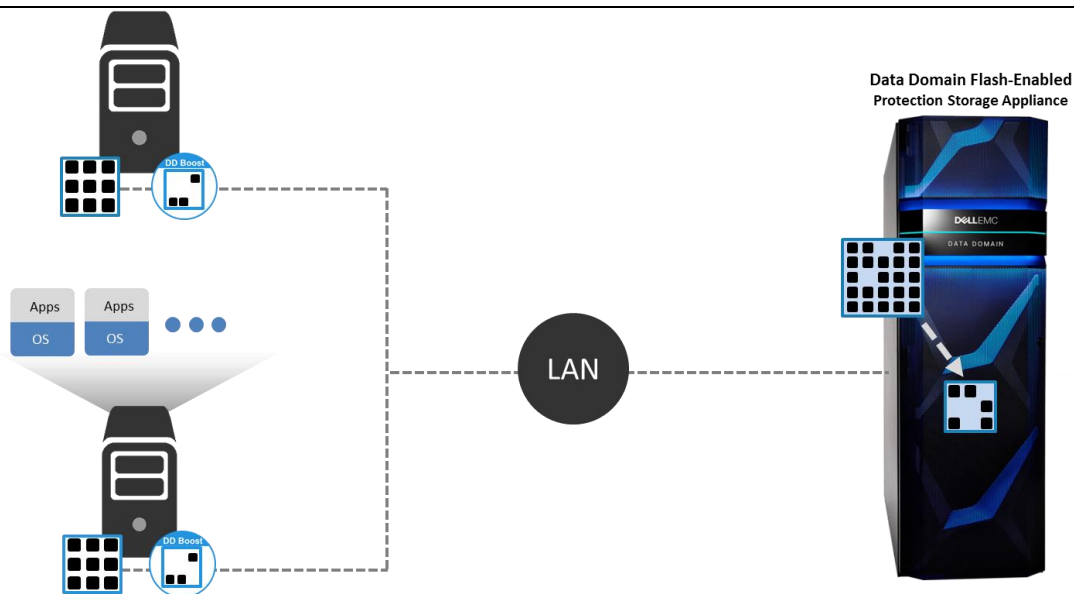
<sup>3</sup> Source: ESG Research Report, [2015 Trends in Data Protection Modernization](#), September 2015.

infrastructures to meet senior IT leader mandates, solve today's data protection challenges, and, most importantly, reduce the cost of storing protection data. ESG's findings are based on an audit and analysis of key performance indicators (KPIs): real-world data from deployed systems, including the original purchase price, environmental costs, capacity/utilization, and performance.

## Data Domain Architectural Benefits

In 2006, EMC made the decision to avoid bolting data deduplication code onto its existing data protection solutions, marking a strategic shift in its data protection strategy. This shift began by leveraging technology from two key acquisitions. The first purchase delivered source-based deduplication technology, and the second was Data Domain, with target dedupe technology. The technologies from these two acquisitions became fundamental components in EMC's data protection solutions. Now under the Dell EMC brand, Data Domain systems including the Integrated Data Protection Appliance, are flash-enabled, fully integrated, purpose-built data protection appliances designed to reduce the amount of disk storage needed to retain and protect data. With both source- and target-based data deduplication natively integrated into the architecture, these systems make it possible to complete more backups in less time, provide faster and more reliable restores, and reduce the amount of storage capacity needed for data protection. The most recent generation of Data Domain systems introduced the inclusion of flash for metadata, enabling performance at scale and 20x faster instant access and restore of virtual machines directly from Data Domain compared with previous generations.

**Figure 2. Data Domain Architecture Overview**



*Source: Enterprise Strategy Group, 2017*

Key Data Domain architectural features include:

- **Stream-informed Segment Layout (SISL):** SISL enables Data Domain systems to perform 99% of deduplication processing in CPU and RAM. This means the systems do not rely on the number of disks to increase performance.
- **Variable-length Segmentation:** For optimal deduplication rates, Data Domain systems leverage variable-length segmentation to break up streams based on the natural structure of the data for optimal deduplication rates. This allows the system to determine whether the segment is unique before compressing and storing it.

- **Inline Deduplication:** Data Domain systems perform deduplication in CPU and memory as the backup stream is received by the system. This means only unique data is ever sent to and stored on disk. This eliminates the need for a disk staging area and compute resources for post-process deduplication.
- **Data Domain Boost Software:** The Data Domain Boost software distributes parts of the deduplication process to the application clients or the backup server. With DD Boost, only unique data traverses the connection between the backup server or clients and the Data Domain system. This also helps free up resources on the customer's network and the Data Domain system for improved target-side deduplication performance.
- **Data Domain Data Invulnerability Architecture:** Data Domain is built to ensure data can be reliably recovered. The Data Invulnerability Architecture provides inline write and read verification, which protects against and automatically recovers from data integrity issues during data ingest and retrieval. Continuous fault detection and self-healing ensure data remains recoverable throughout its lifecycle on the Data Domain system.

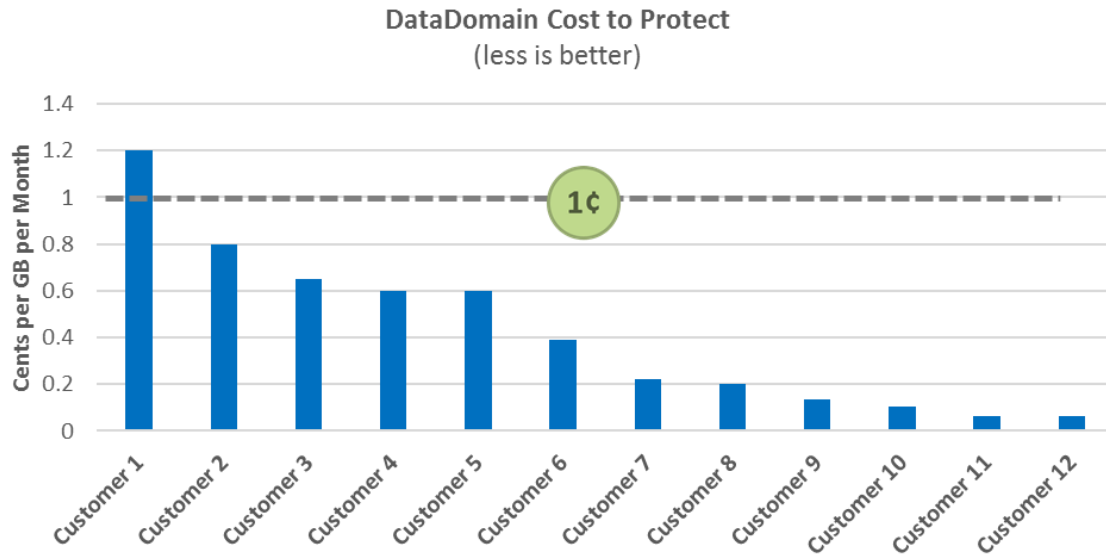
## Data Domain Economic Benefits

Let's take a closer look at how the Data Domain architecture translates into economic benefits for business stakeholders. Data deduplication enables customers to store more data on the same amount of physical disk space. This reduces storage capacity requirements and drives down cost. Data deduplication implemented at the source or client side also helps with capacity savings, but with the added benefit of improving backup performance. With source-side deduplication, only unique data blocks are sent from the source to the target during the backup operation, which significantly reduces network traffic. This improved network efficiency allows for backup data growth using existing network infrastructure and possibly eliminates or postpones the need for expensive network upgrades. Obviously, the less data that needs to be transferred, the faster the backup performance. Shorter backup durations also allow customers to increase the frequency of backups, reducing the risk of data loss, which can be extremely costly to an organization. With DD Boost software, Data Domain systems support both source- and target-side deduplication, giving customers the flexibility to deploy deduplication where it makes the most sense for their environments.

As shown in Figure 3, ESG's analysis of real-world data, including hardware, software, power, cooling, and deduplication, demonstrates that Data Domain systems are easily capable of serving storage to data protection environments for fractions of a penny per GB per month.

Note how the cost to protect ranges between 0.062 and 1.2 cents per gigabyte for the twelve customers that ESG examined. This relatively wide range in savings is primarily due to the amount of deduplication capacity savings that are being achieved by Dell EMC customers in production environments. The balance of this report takes a closer look at the cost to protect with a focus on how deduplication savings can be magnified with an end-to-end combination of Dell EMC Data Protection Suite software and Data Domain hardware.

**Figure 3. The Economic Value of Data Domain**

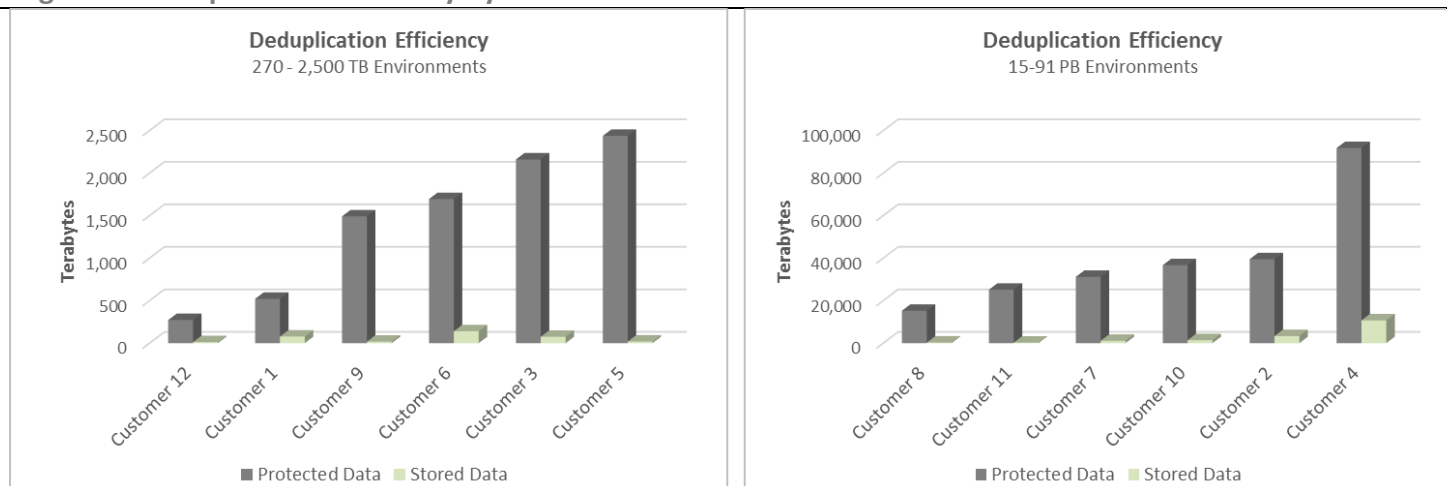


Source: Enterprise Strategy Group, 2017

**Deduplication Efficiency Does Matter**

ESG began its exploration of the economic value of Data Domain deduplication capacity savings by auditing and analyzing call-home support data from 12 active Dell EMC customers. As shown in Figure 4, the customer environments ranged from approximately 270 TB to more than 90 PB of data to be protected. The customers spanned multiple industries, including technology, manufacturing, insurance, and health care. The selection of customers in different industries was designed to capture deduplication results across different types of data sets. The gray bars in Figure 4 show the amount of data being protected for each environment. The green bars show the amount of data stored after deduplication.

**Figure 4. Deduplication Efficiency by Customer**



Source: Enterprise Strategy Group, 2017

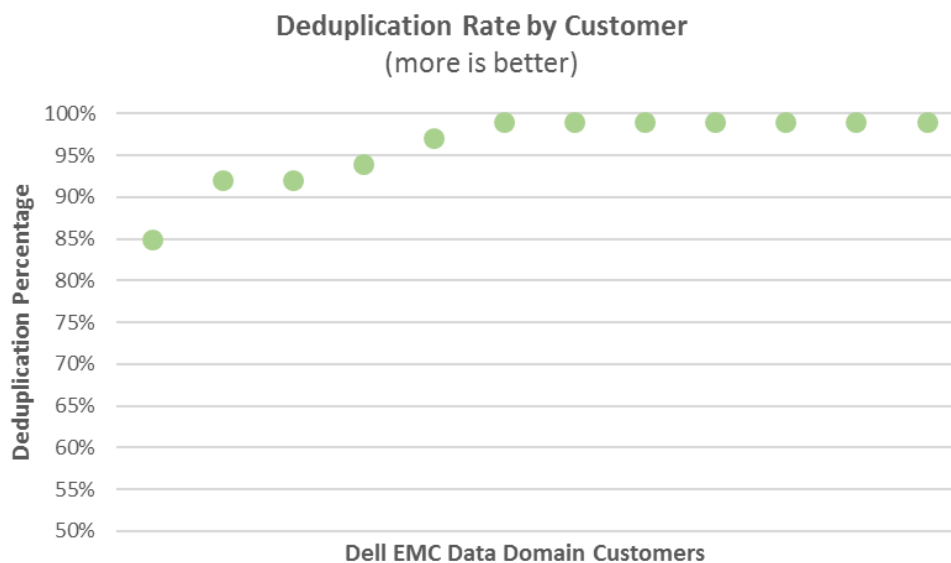
- Deduplication:** The largest deduplication rate observed in the analyzed data was 126:1. This result came from a customer leveraging Data Protection Suite backup and recovery software with a Data Domain appliance, and is identified by the first data point on the left side of Figure 4. The average deduplication rate for all analyzed customers was 41:1.

- **Protection:** Even the customer with the smallest observed deduplication rate of approximately 9:1 could protect almost 92 PB of data using just 10 PB of capacity. This customer is identified by the last data point on the right side of Figure 4. The customer is part of the manufacturing industry and has a highly distributed environment.

Figure 5 provides another lens on the Data Domain deduplication effect. This figure shows the percentage of deduplication achieved for each of the same 12 customer environments. Deduplication rates range from 85% to 99%. Typically, the longer the Data Domain solution has been receiving data in the environment, the higher the deduplication rate. This is because it is more likely to see the same data patterns over time and only needs to store unique data.

The variable-length segmentation feature of the Data Domain architecture is a key component to achieving this level of deduplication. With variable-length segmentation, the Data Domain solution can more easily align with natural patterns, including database timestamp markers, in the data structures being sent to the device for protection. Variable-length segmentation produces significantly higher levels of deduplication over fixed-length architectures. It also enables more effective scalability within a single storage pool. This also results in fewer devices to manage, saving on operational costs.

**Figure 5. Percentage of Deduplication by Customer**



Source: Enterprise Strategy Group, 2017

Another benefit of high variable-length segmentation deduplication comes when data is replicated between Data Domain devices. Less data means less data to replicate. This means not only better replication performance, but also lower WAN bandwidth requirements for the replication process.

- **Deduplication Rates:** The audited and analyzed deduplication rates ranged from 85% to 99%, with an average of 96%. This means that only a fraction (in best cases 1% or less) of the production data needed to be stored on the Data Domain system for protection.
- **Under 90%:** Only one customer from the analyzed data had a deduplication rate under 90%. Again, this customer represents the highly distributed environment with distributed protection pools.
- **90% and Above:** The remainder of the clients analyzed achieved deduplication rates above 90%, and most customer environments were running a deduplication rate of 99%.

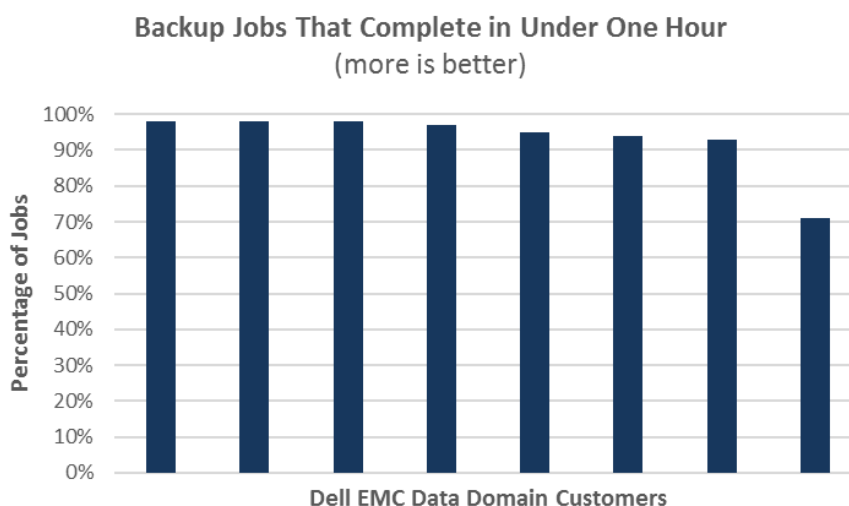
## Speed and Scalability

Increasing storage capacity through deduplication does not always go hand in hand with high performance. Data Domain addresses this challenge by leveraging two key performance elements: Stream-informed Segment Layout (SISL) and DD Boost software.

SISL is an architectural element of the Data Domain platform that enables 99% of the deduplication process to occur in CPU and RAM. This means the solution does not rely on a high disk spindle count for performance. As a result, smaller footprint solutions based on the same Data Domain controller can achieve the same performance as higher capacity configurations. Also, because the Data Domain controllers use Intel CPUs, they benefit from performance enhancements in each new release of the Intel processors.

DD Boost software also helps deliver performance by distributing parts of the deduplication process to the application clients or the backup server. With DD Boost, only unique data needs to be moved from the backup server or clients to the Data Domain system. DD Boost reduces the amount of data moved by up to 99%. The net benefit of quicker backup job completions was audited by ESG. As shown in Figure 6, most customers routinely complete backup jobs in an hour or less.

**Figure 6. Backup Job Performance**



Source: Enterprise Strategy Group, 2017

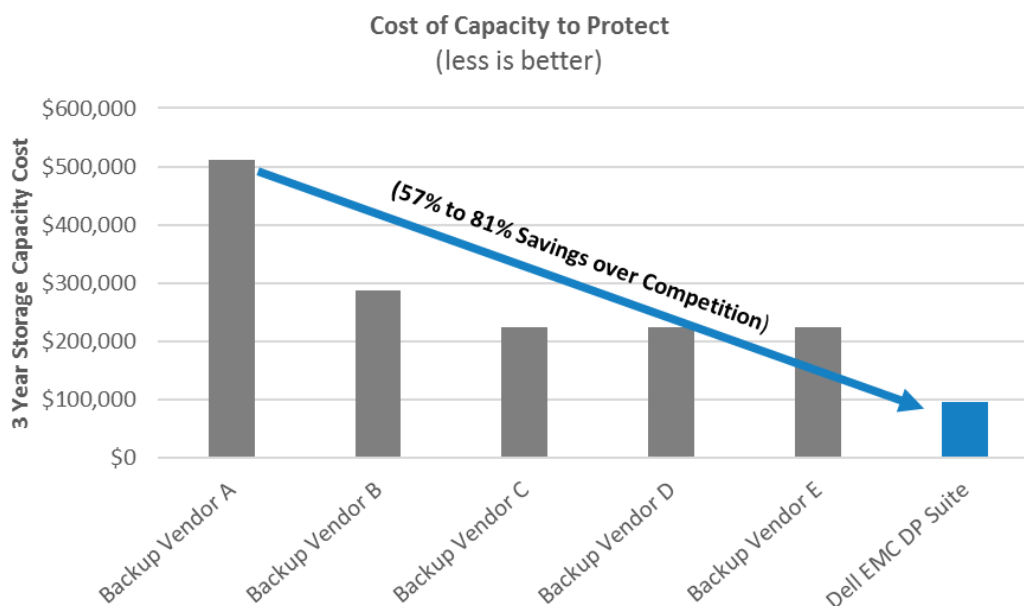
- **Under 15 Minutes:** For one customer in the technology industry, 76% of all backup jobs finished in less than 15 minutes.
- **Under 1 Hour:** For eight of the twelve customers analyzed, 70% to 98% of all backup jobs completed in less than an hour (as shown in Figure 6).
- **Under 4 Hours:** For one customer in the insurance industry, all backup jobs completed in less than four hours.
- **More than 4 Hours:** For two of the largest environments (in terms of the number and size of applications), it took more than four hours to complete all backup jobs.



## Cost Savings Considerations

ESG completed the exploration of the economic value of Data Domain deduplication capacity savings by auditing and analyzing call-home support data from over 15,000 Data Domain systems deployed worldwide. This sample was collected from Data Domain systems that record the backup software that is being used and the deduplication rates that are being achieved. The data shows that Data Domain’s deduplication efficiency can be taken even further when an end-to-end Dell EMC data protection solution is implemented. As shown in Figure 7, pairing Data Domain with Data Protection Suite magnified duplication savings and reduced the cost of capacity to protect by 57% to 81%.

**Figure 7. Magnifying Savings with Dell EMC Data Domain and Data Protection Suite**



Source: Enterprise Strategy Group, 2017

The Data Domain solution delivers other economic benefits beyond the cost of capacity savings shown in Figure 7. ESG’s analysis of real-world Data Domain customer data also revealed the following benefits:

- **Network Bandwidth Reduction:** Like most enterprise applications, backup and recovery solutions rely on network resources to function. This means both LAN resources for local backup and restore and WAN resources for DR and business continuance. Inefficient data protection solutions can quickly consume valuable network resources and even impact user connectivity and productivity. By leveraging Data Domain deduplication, a multi-national manufacturer was able to reduce its local and remote data protection bandwidth utilization by 98%.
- **Performance:** Efficient data deduplication can have a major impact on improving overall data protection performance. By implementing DD Boost, which reduces the amount of data that needs to be transferred between the client and the Data Domain system, a heavy equipment manufacturer was able to improve its backup and restore speeds by 50%. The same customer improved its DR readiness by 90%.
- **Data Center Footprint Reduction:** Real estate, especially the amount needed to support the space, power, and cooling requirements of a modern data center, is not cheap. It can be difficult, sometimes impossible, to physically expand a data center without relocation. With Data Domain, a national department store chain eliminated physical tape in its data protection environment and freed up three full rows of highly valuable data center rack space.

## The Bigger Truth

The top data protection mandates from IT leaders are focused on improving the fundamental reliability and agility of the solution(s) in use. The mandate that follows closely behind is cost reduction, which is also seen as a top priority among data protection implementers. These challenges should not be seen as contradictory or mutually exclusive; in fact, they can all be addressed by improved data protection solutions that are engineered as much for efficiency as they are for reliability and capability.

Efficiency comes in many flavors, often grounded in the cost of doing business and interoperability:

- The ability to economically deliver protection storage by reducing the cost to serve the data protection application.
- Interoperability across backup software solutions, since only one in four organizations has a single backup application.<sup>4</sup>
- The scalability and deduplication efficiency of the protection storage system, which can be used to reduce the number, and cost, of systems that need to be deployed and managed.
- The ability to leverage the same pool of capacity-optimized protection storage for backup and archive data.
- Interoperability (and integration) between protection storage and myriad platforms, so that application owners (e.g., database administrators and vAdmins) can utilize their own tools for supplemental protection/recovery, while still being responsible IT citizens and leveraging efficient centralized protection storage.

By providing what many in the industry consider synonymous with “protection storage” and/or “deduplication,” Dell EMC has expanded the Data Domain platform from simply a component of a data protection infrastructure to the basis of an ecosystem that includes production storage, backup vendors, archive vendors, and plug-in accelerants for applications (i.e., DD Boost), among other elements. The Dell EMC Integrated Data Protection Appliance is an extension of this ecosystem with protection software and analytics pre-configured. As such, there is significant impetus to continue innovating in front of where the current market demands lie, resulting in further efficiency gains within Data Domain’s deduplication mechanisms, enhancements with their integrations, and expansions of their capabilities beyond just on-premises deduplication, such as cloud-tiering, replication to/from cloud-hosted appliances, virtual appliances, disaster recovery to the cloud, etc.

One of the most important differentiable elements of any deduplication platform is the rigor of ensuring the integrity of the data within the repository, since so much relies on the integrity of every unique block within the system. As such, it is critical that organizations diligently investigate how zealous and how frequently the system checks the underlying storage for assured integrity. Dell EMC’s approach to this has been a hallmark of the Data Domain platform since its inception through what it refers to as Data Invulnerability Architecture (DIA).

Based on an audit of field data collected from more than 15,000 production environments, ESG has confirmed that the combination of Dell EMC Data Domain hardware and Data Protection Suite software reduces the cost of protection capacity by 57% to 81% compared to Data Domain environments with competitive backup software. Organizations that are hesitant to make a Dell EMC Data Domain investment based on “price” would be well served to reconsider the economic benefits over time including the cost of reduced downtime and data loss, improved performance, and, most important, a cost to protect of less than a penny per GB per month.

---

<sup>4</sup> Source: ESG Research Report, *Data Protection Modernization*, to be published.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of The Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.



**Enterprise Strategy Group** is an integrated IT research, analysis, and strategy firm that is world renowned for providing actionable insight and intelligence to the global IT community.

© 2017 by The Enterprise Strategy Group, Inc. All Rights Reserved.

