White Paper

Understanding the Economics of In-cloud Data Protection: Designed for Today with Tomorrow in Mind

Why In-cloud Data Protection Architecture Matters

By Vinny Choinski, Senior ESG Lab Analyst;
and Christophe Bertrand, Senior Analyst
April 2019

Table of Contents

Introduction

The meaning of data protection is the same today as it was two years ago. Data protection simply means backup, restore, and archive. As the percentage of cloud infrastructure service users running production applications in cloud platforms continues to grow, it's interesting to note that the most common cloud infrastructure use case just a couple of years ago was for applications running on-premises infrastructure.[1] While backup is still an established use case today, running production applications as a cloud infrastructure service, which was already starting to percolate in 2017, is now, as shown in Figure 1, the most-cited use case for cloud infrastructure services, followed closely by running business intelligent queries and running test and development environments.[2] As organizations continue to embrace digital transformation for both their business and IT groups, the use of public cloud infrastructure has become a key driver for that transformation. Organizations are using cloud infrastructure services to mitigate the capital and operational expenses associated with traditional IT hardware deployments. Clearly, the cloud infrastructure services paradigm is evolving.

Figure 1. Cloud Infrastructure Use Cases

For which of the following purposes does your organization use cloud infrastructure services (IaaS and/or PaaS)? (Percent of respondents, N=438, multiple responses accepted)

Running production applications	49%
Running business intelligence queries	47%
Testing and development	47%
Serving as a repository for backup and/or archive data	40%
Accommodating workload spikes	37%
Serving as a disaster recovery target	37%
Using as temporary compute resources for time-limited projects	34%

Source: Enterprise Strategy Group, 2019

Why In-cloud Data Protection Architecture Matters

As more organizations move production applications to the cloud, they often find that they still need to run the same data protection workflows to support the same business operations that they did when their applications were on-premises. These workflows commonly include application-aware backups with application-consistent restores that support other requirements such as disaster recovery, business continuance, test, and development. Organizations may not be able to achieve these types of recoveries or meet the SLAs of the business using the native data protection services provided by

---

[1] Source: ESG Research Report, *2017 IT Spending Intentions Survey*, March 2017.
[2] Source: ESG Research Report, *2019 Technology Spending Intentions Survey*, February 2019.

cloud providers. These services typically rely on volume-based snapshot technologies, which often cannot meet the consistency and granularity of data recovery required by key business applications.

To validate the cost, ESG did extensive auditing of a comprehensive model based on publicly available pricing and reference architecture guides for three industry-recognized data protection solutions. From an architectural perspective, each of these solutions can be deployed in the cloud as virtual appliances running on virtual machines, leveraging both block and object storage. The Dell EMC and Vendor A solutions represent a more traditional server/data mover model with data movers that can scale compute and storage resources independently. Vendor B's solution uses a server grid architecture with compute and storage scaling in parallel as the amount of workload increases. As shown in Figure 2, the Dell EMC solution starts with a slight edge and builds a significant cost advantage over the competition as the environment scales.

The modeling is based on the amount of production data to be protected and the resources required by each solution. It includes three critical cost components: compute, block storage, and object storage. ESG's analysis compares the Dell EMC solution against a solution (Vendor A) that does not efficiently handle storage repository requirements and a solution (Vendor B) that relies on significantly more block storage and compute capacity to support the targeted workloads. As shown in Figure 2, the Dell EMC solution delivers up to a 47% reduction in total monthly costs for in-cloud data protection infrastructure over its competition. The effect of each component, compute and storage, on the overall cost of the infrastructure is explained in detail in the **Storage Efficiency** and **Compute Efficiency** sections of this report.

Figure 2. Monthly Infrastructure Cost of In-cloud Data Protection by Solutions
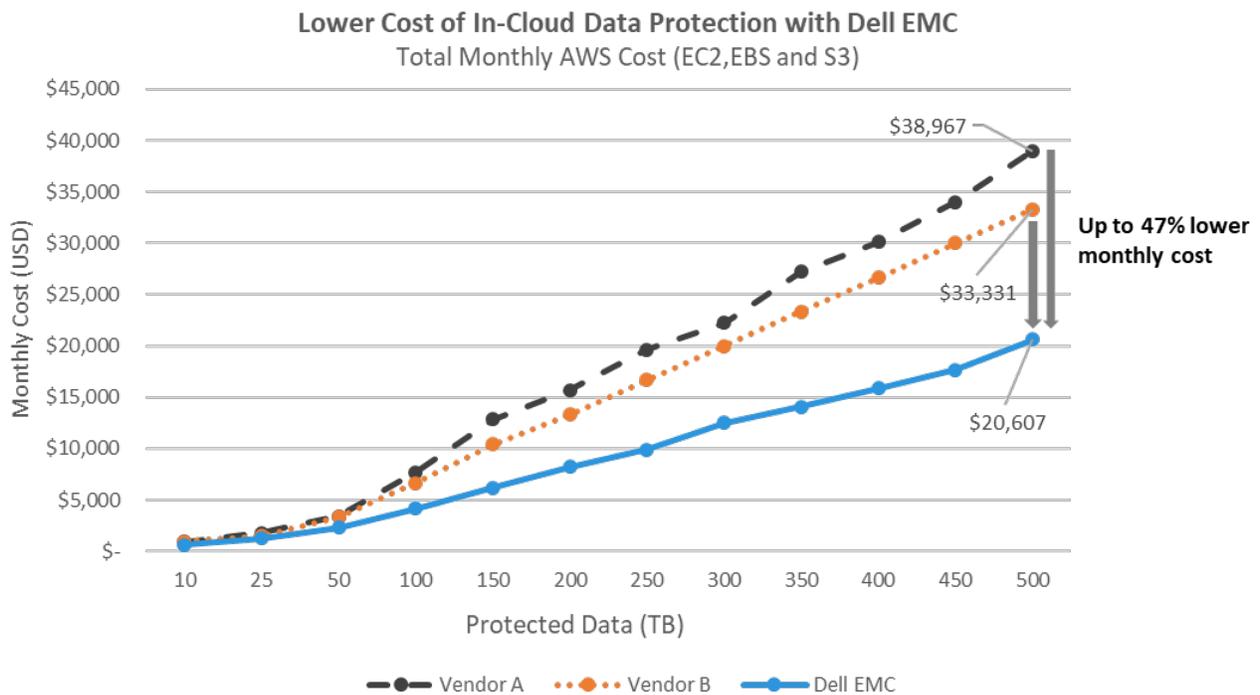
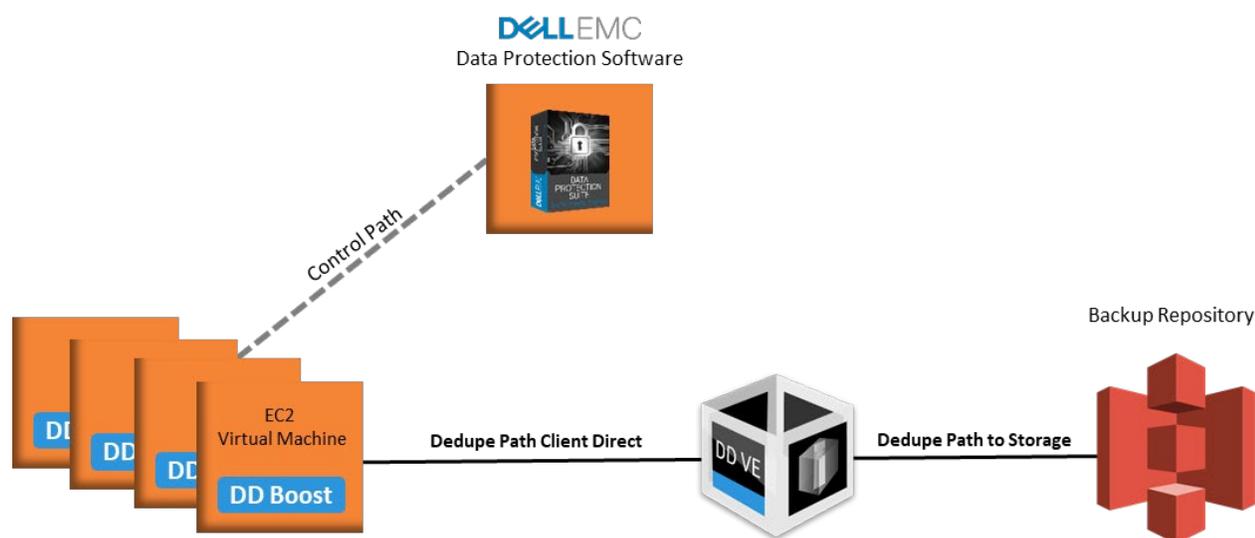*Source: Enterprise Strategy Group, 2019*

Dell EMC Data Protection Software with Data Domain Virtual Edition

Dell EMC Data Protection Software offers unified data protection for the enterprise that centralizes, automates, and accelerates backup and recovery across the entire IT environment. This includes providing the same quality of data protection for cloud-hosted applications as for applications running in on-premises data centers. Deployment of Dell EMC

Data Protection Software is accomplished by using virtualized appliances, both for the centralized server controlling policies and schedules and the Data Domain storage repository.

This integrated solution also includes a client-based feature called **DD Boost,** which integrates with file system and application backup and recovery running on Linux and Windows clients. DD Boost enables client direct backup to Data Domain Virtual Edition (DD VE) storage repository. This reduces the overall solution cost by distributing the data movement workload across the client virtual machines, which reduces the need to provision dedicated data mover virtual machines.

Figure 3. NetWorker with Data Domain Virtual Edition Architecture Overview in AWS Cloud

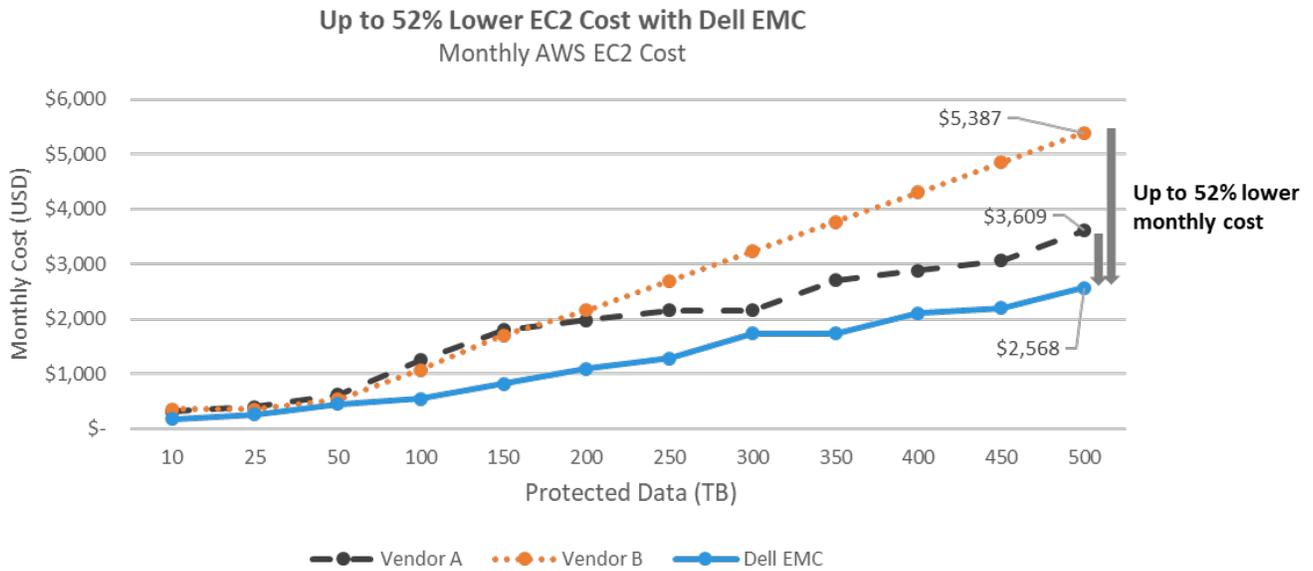*Source: Enterprise Strategy Group, 2019*

Key architectural features include:

- **NetWorker Virtual Edition (NVE) Server:** This is the core component of the solution and supports the policy, scheduling, and catalog functions to manage backup and recovery. It runs as a virtual appliance on an AWS EC2 virtual machine.

- **Client:** The client is a software component deployed on a host system to protect the operating system and application data. It is integrated with DD Boost and can send data to be protected directly to the DD VE in the cloud.

- **Data Domain Virtual Edition (DD VE) Server:** This is a software-defined data protection appliance running on an AWS EC2 virtual machine leveraging AWS EBS (block) and S3 (object) storage. DD VE emulates the data management functions of a physical Data Domain appliance, including: variable length deduplication, data integrity, standard management interfaces, and integration with Dell EMC Data Protection Software, NetWorker.

- **DD Boost:** DD Boost is a client-based software component that enables the client to efficiently back up and retrieve deduplicated data directly to and from the Data Domain virtual appliance.

Compute Efficiency

ESG analyzed the compute resources required to deliver enterprise-class data protection at scale for cloud-based production environments. The analysis included the cost of the compute, block, and object storage resources required for a data protection solution to handle the load of backup jobs as the production environment was scaled from 10 to 500TB. The resource modeling was based on reference architecture guides and publicly available pricing. Figure 4 represents the EC2 (compute) costs required to protect up to a 500TB production environment.

**Figure 4. Compute Requirements for In-cloud Data Protection**

Source: Enterprise Strategy Group, 2019

At 500TB of protected data, the Dell EMC solution required two (2) NVE backup servers and six (6) DD VE servers ("large" compute instances). Vendor A required a single backup and three (3) data movers to handle the backup data transport, deduplication, and index processing. These all were sized as "extra-large" compute instances, which explains the higher cost than Dell. Vendor B had to be scaled to 60 compute instances in the cluster in order to support 500TB of protected data. It should be noted that all EC2 pricing was based on 1-year, no up-front reserved pricing.

Storage Efficiency

A major component of any in-cloud, hybrid, or on-premises data protection solution is the backup repository. This is where the backup images of protected data are stored. If not managed efficiently, the backup repository can grow large very quickly as more backup jobs complete over time, systems are added to the protection schema, and the amount of production data grows naturally over time. Historically, backup architects leveraged a combination of tape volumes, simple disk pools, and disk storage with deduplication technology for backup repositories. More recently, object storage has been added to the mix as IT professionals start to leverage more cloud-based features in their data protection solutions. As shown in Figure 5, the Dell EMC solution uses proprietary deduplication technology between the EC2 virtual machines, DD VE and the Amazon S3 object storage container to deliver backup repository efficiency. For enterprise-level data protection, each EC2 virtual machine runs client software, which includes DD Boost agent libraries, for client to appliance data movement and deduplication.

**Figure 5. NetWorker with Data Domain Virtual Edition Deduplication Overview**
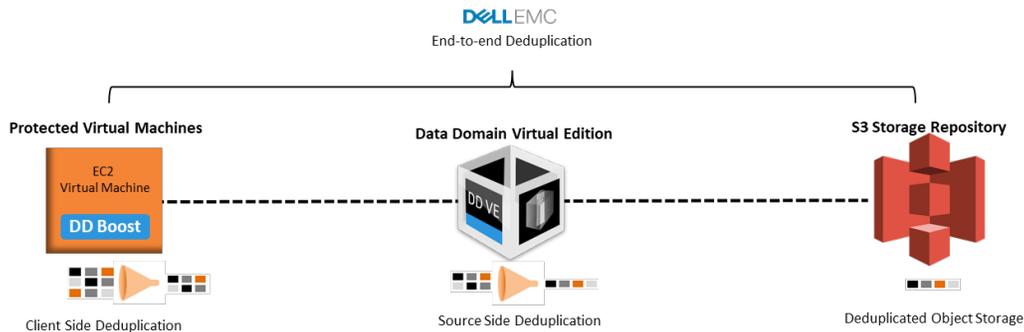
As shown in Figure 5, the solution enables each EC2 client to perform its own data deduplication and efficiently send only the unique data blocks directly to the DD VE appliance. Backup processing metadata, such as deduplication chunk indexing, is sent directly to the NVE server. This schema decouples backup data transport from backup processing tasks for improved performance and storage efficacy.

Figure 6 shows the storage cost (EBS and S3) modeling results. Dell EMC and Vendor A both support data deduplication for S3 object storage. Vendor B also supports deduplication, but that storage must be block storage (EBS) due to the way deduplication takes place. Based on reference architecture specifications and audited results, the Dell EMC solution was modeled at a 14.5:1 deduplication ratio, Vendor A at 3.5:1, and Vendor B at a 3.25:1. It should be noted that the deduplication ratios are based on a combination reference architecture recommendation and, where available, real-world data analysis.[3] The ratios were then applied to the cost model assumptions as the environment was scaled. The production environment model is based on a 70/30 mix of file system data and database data with a 3% DB and 1% file system daily change rate. The protection schema included initial fill backups and daily protection jobs with a retention of 28 days.

Figure 6. Storage Requirements for In-cloud Data Protection
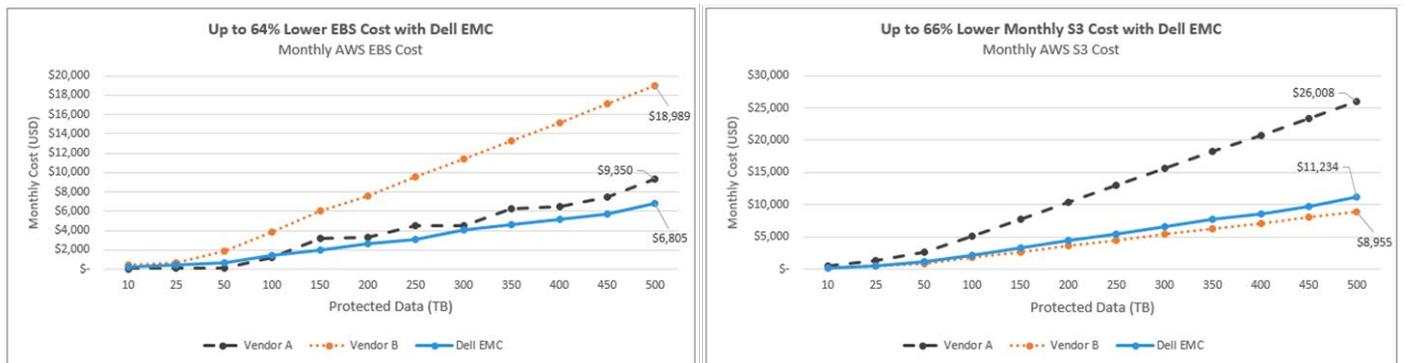
What the numbers mean:

- Dell EMC's solution requires significantly less block storage (EBS) then Vendor B, which results in a 64% cost savings. Dell EMC is also 27% less costly when compared with Vendor A's EBS costs.

---

[3] Source: ESG Lab Review, *Efficiently Protect Virtual Environments with Integrated Data Protection Appliance from Dell EMC*, February 2018

- The Dell EMC solution's efficient use of S3 storage delivered a 66% reduction in S3 costs when compared with Vendor A. Vendor B was 25% less costly for the S3 storage component than Dell EMC because it can leverage a lower tier of S3 storage (S3 Standard – Infrequent Access). Note that retrieving data from S3 Standard – Infrequent Access storage is subject to additional per GB costs, unlike retrieval from S3 Standard storage.

- When both EBS and S3 storage costs are combined, the Dell EMC solution results in a 49% costs savings when compared with Vendor A and a 35% savings when compared with Vendor B.

Figure 7 represents the EC2, EBS, and S3 costs required to protect a 250TB production environment. From this chart, it's easy to see where the associated costs are incurred for each of the solutions.

At 250TB of protected data, the Dell EMC solution required a single instance of NVE and three DD VE instances. Vendor A also required only a single backup server along with two data movers. Vendor B, on the other hand, required 30 data protection nodes. Figure 7 shows the economic benefits of compute- and storage-efficient solutions. While Vendor A was the costliest solution overall due to its inefficient use of S3 storage resources, it is efficient from a compute and block storage perspective. Vendor B, on the other hand, incurs significantly more cost due to the additional block storage and compute needs when compared with the Dell EMC solution.

Figure 7. Cost Analysis at 250 TB of Production Data
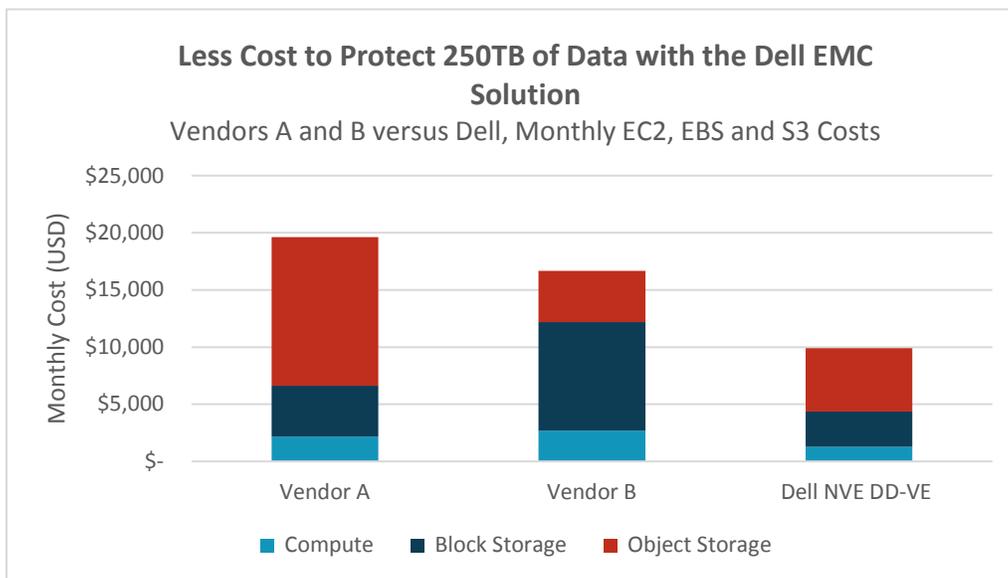
Source: Enterprise Strategy Group, 2019

What the numbers mean:

- At 250TB of production data to protect, roughly the mid-point of the model, the Dell EMC solution delivers a 50% cost reduction over Vendor A (41% cost reduction over Vendor B) based on the total data protection application resources required.

- These costs savings are due to the decoupling of metadata from the backup data transport process and the efficient metadata processing provided by the DD Boost running on the client systems. In addition, highly efficient deduplication processing (client-side via DD Boost and target-side on DD VE) resulted in a much higher deduplication ratio on S3 storage for the Dell EMC solution.

The Bigger Truth

ESG completed its annual IT spending intentions survey of 600 senior IT decision makers at midmarket (i.e., 100 to 999 employees) and enterprise (i.e., 1,000 or more employees) organizations across North America and Western Europe in early 2019. As part of that research, respondents were asked about their organizations' usage of public cloud services. The research found that 49% of respondents indicated they were currently running production applications on cloud infrastructure.[4] While previous ESG research has demonstrated a consistent synergy between public cloud infrastructure and data protection functions,[5] the use of these services to support production applications has been steadily increasing over the years. In fact, ESG research shows a steady year over year increase from 2015, when 27% of respondents indicated they were running production applications on IaaS, to 49% in 2019. As the amount of in-cloud, business-critical production applications increase, so too does the need for cost-efficient enterprise-class, in-cloud data protection. And it's important to remember that even though a cloud provider may have SLAs for the infrastructure the applications run on, protecting the data in those applications at the granular level the business requires is the user's responsibility.

ESG conducted in-depth auditing and analysis of the Dell EMC cost model for in-cloud data protection and found that the solution delivered the following advantages over the audited competitors:

- Up to 47% lower total monthly cost for in-cloud data protection infrastructure.

- Up to a 64% reduction in the amount of EBS storage required for data protection processing/storage.

- Up to 52% lower monthly cost for the required EC2 data protection application resources.

So, why does in-cloud data protection architecture matter? Because organizations are moving more business-critical applications to the cloud and they still need to provide enterprise-class data protection for these applications at a reasonable cost. Inefficient use of any one of the cloud resources can have a major impact on the cost of data protection. ESG analysis found that, with small cloud deployments, it's hard to find a cost advantage between the audited solutions. However, just as with on-premises data protection, choosing the wrong solution for the initial deployment can have a major impact on cost efficiency as the solution scales. ESG confirmed that the Dell EMC NVE with DD VE solution efficiently leverages EC2, EBS, and S3 cloud resources, providing a cost-efficient in-cloud data protection environment at scale.

---

[4] Source: ESG Master Survey Results, *2019 Technology Spending Intentions Survey,* December 2019.
[5] Source: ESG Research Report, *2017 Public Cloud Computing Trends*, April 2017.

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.