White Paper

Economic Value of In-cloud Data Protection with Dell EMC Data Protection Software

Why In-cloud Data Protection Architecture Matters to a Modern IT Infrastructure

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

June 2017

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

Table of Contents

Introduction

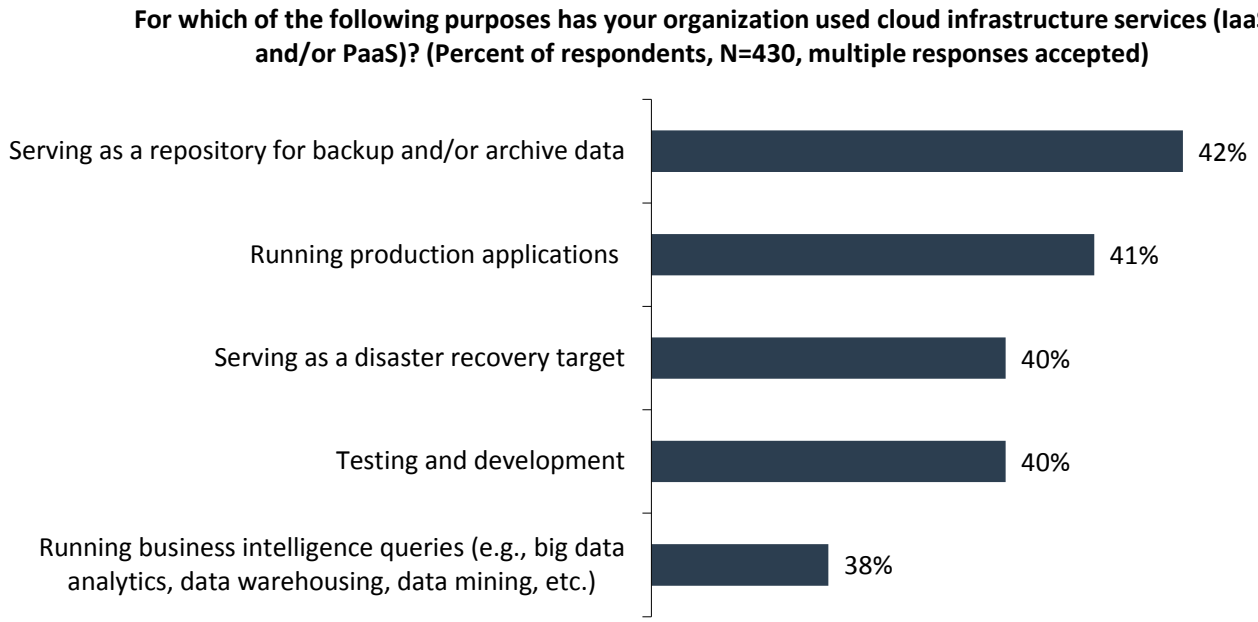
As the shift into a digital economy continues to accelerate, companies must out-innovate, outthink, and outpace their competition. Businesses must embrace change and transform to become the disruptors in their industries rather than wait to be disrupted by their competitors.

In the face of this pressure to evolve, organizations need to transform IT to reduce both the capital and operational costs of legacy IT. They must offload repeatable and time-intensive manual tasks like backup, disaster recovery, and service deployment to software and policy-driven automation.

To do this, organizations are using infrastructure-as-a-service (IaaS, also known as cloud infrastructure services) to mitigate the capital and operational expenses associated with traditional IT hardware deployments. Potential use cases for IaaS include providing target systems for remote backup and replication, supporting the compute and storage resource requirements for short-term test and development activities, and even accommodating temporary/event-driven spikes or peak workloads.

Consistent with previously conducted ESG research in the area of cloud-based data protection, data protection is the most common IaaS use case among those organizations currently leveraging cloud infrastructure services.[1] While data protection is clearly an established IaaS use case, it is worth noting that 41% of cloud infrastructure users leverage these services to run production applications, ahead of (albeit slightly) supporting test and development environments.[2]

Figure 1. Top Five Cloud Infrastructure Use Cases



Source: Enterprise Strategy Group, 2017

It is also worth noting that participants surveyed for the *Data Protection Cloud Strategies* research report selected overall cost-effectiveness of the solution and overall scalability of the solution as the top two most important considerations when leveraging storage-as-a-service/data protection (STaaS/DP) to add cloud-based protection functionality.[3]

[1] Source: ESG Research Report, *Data Protection Cloud Strategies*, December 2016.
[2] Source: ESG Research Report, *2017 Public Cloud Computing Trends*, April 2017.
[3] Source: ESG Research Report, *Data Protection Cloud Strategies*, December 2016.

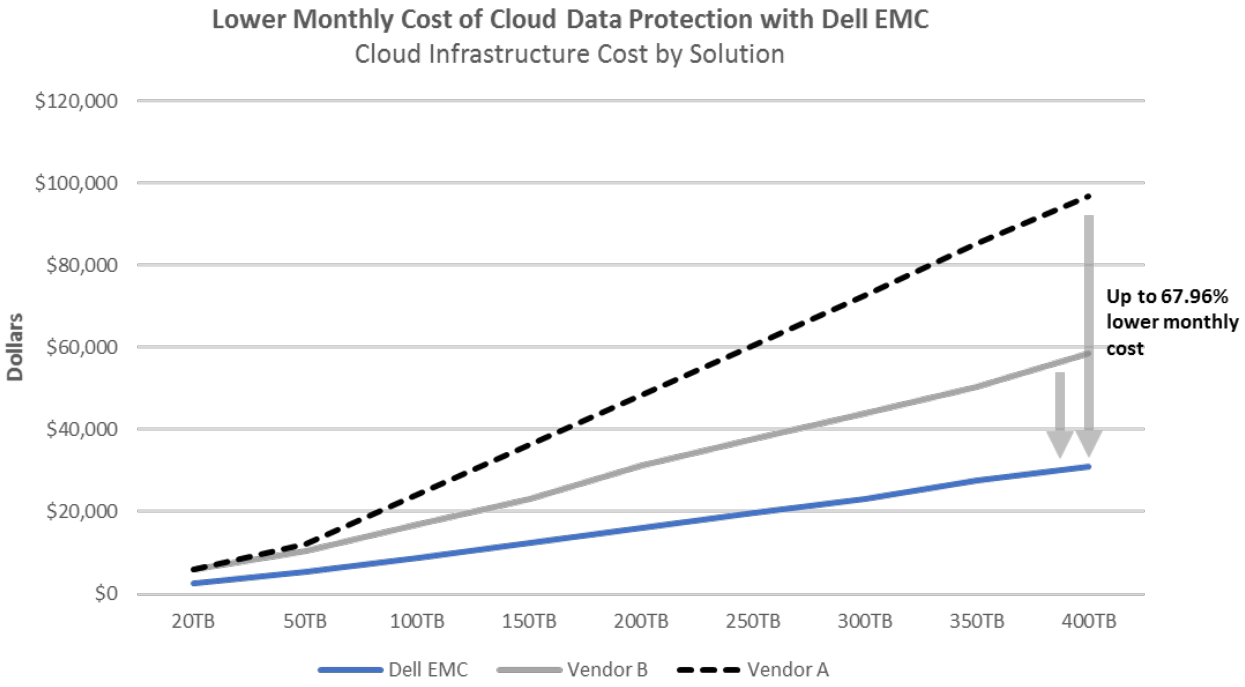
Why Cloud Data Protection Architecture Matters

As more organizations move production applications to the cloud, they often find that, to support their normal business operations, they still need to run the same data protection workflows they did when their environments were on-premises. These workflows commonly include application-aware backups with application-consistent restores that support other cloud initiatives such as disaster recovery, business continuance, test, and development. These types of recoveries may not be achievable to the consistency level required with the snapshot technology offered through the cloud provider.

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. 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 architecture. It includes two critical cost components, compute and storage. ESG 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 does efficiently handle indexing and backup data processing. The effect of each component, compute and storage, on the overall cost of the infrastructure is explained in detailed in 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, 2017

What the numbers mean:

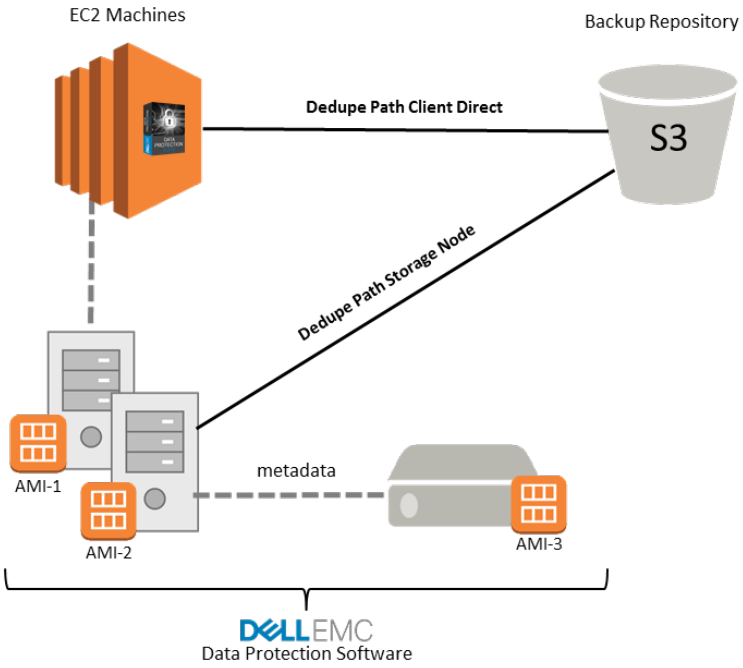
- Dell EMC Data Protection software delivers up to a 67.96% reduction in total monthly cost for in-cloud data protection infrastructure over its competition.
- Though the cost model starts at 5 TB of production storage, we present analysis from 20 to 400 TB based on architecture specifications for data center-class solutions from each vendor.

Dell EMC Protection Software with CloudBoost

Dell EMC Protection Software offers unified data protection for the enterprise that centralizes, automates, and accelerates backup and recovery across the entire IT environment. It includes a data protection-optimized cloud storage appliance with support that spans on-premises, hybrid, and in-cloud environments including full support for AWS.

The integrated solution includes a native device type called CloudBoost and integration with the CloudBoost library for Linux clients, Windows clients, and storage nodes. As shown in Figure 3, the solution is deployed in the AWS cloud using preconfigured Amazon Machine Images (AMIs). In the figure, AMI-1 is the server, AMI-2 is a storage node, and AMI-3 is the CloudBoost appliance. The figure also shows the data protection workflow with the efficient separation of HTTPS backup data transfer and its associated metadata processing.

Figure 3. NetWorker with CloudBoost Architecture Overview



Source: Enterprise Strategy Group, 2017

Key architectural features include:

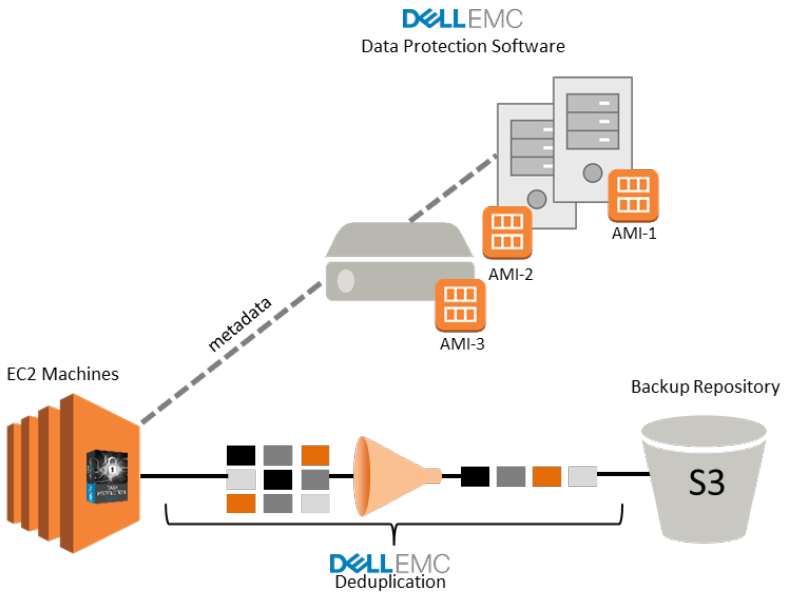
- **Server:** This is the core component of the solution and supports the policy, scheduling, and catalog functions to manage backup and recovery.
- **Client:** The client is a software component deployed on a host system to protect the operating system and application data. It sends data to be protected to the server, storage node, or directly to a storage device.
- **Storage Node:** The storage node is an optional component that acts as a data transport mechanism for those clients that do not support sending data directly to a storage device.
- **CloudBoost Appliance:** The appliance enables access to public, hybrid, and private cloud storage resources and decouples metadata from backup data for efficient data transfers.

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, the 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. This paper focuses on the efficient use of object storage as the backup repository—specifically, Amazon Simple Storage Service (Amazon S3), which is object storage designed to deliver nine nines of durability and scale past trillions of objects.

As shown in Figure 4, the Dell EMC Data Protection Software solution uses proprietary deduplication technology between the EC2 virtual machines and the Amazon S3 object storage container to deliver backup repository efficiency. In this figure, AMI-1 is the server, AMI-2 is the storage node, and AMI-3 is the CloudBoost appliance. For enterprise-level data protection, each EC2 virtual machine runs client software, which includes CloudBoost agent libraries for client to appliance communication.

Figure 4. NetWorker with CloudBoost Deduplication Overview

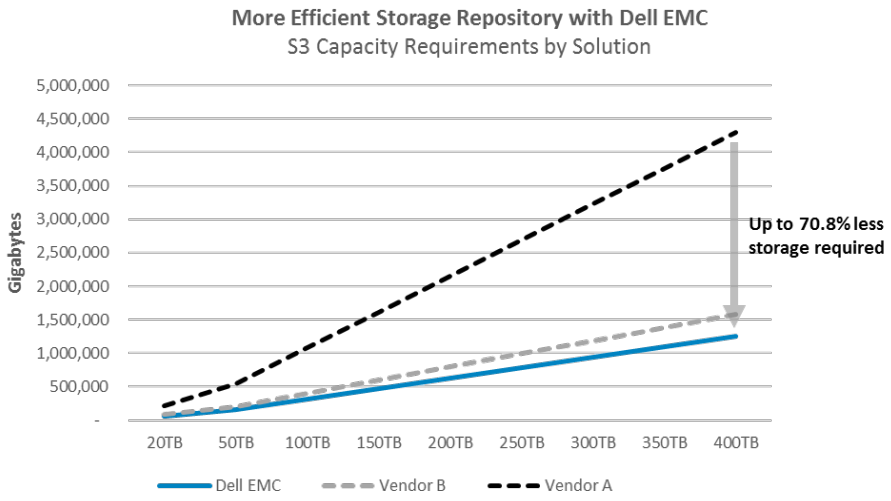


Source: Enterprise Strategy Group, 2017

As shown in Figure 4, the solution enables each EC2 client to perform its own data deduplication and efficiently send only the unique data blocks directly to the S3 repository via the HTTPS protocol. Backup processing metadata, such as deduplication chunk indexing, is sent directly to the CloudBoost appliance. This schema decouples backup data transport from backup processing tasks for improved performance and storage efficacy. It should be noted that the storage nodes handle data transport and CloudBoost appliance communication for those clients that do not support direct access to the S3 storage repository.

Figure 5 shows the storage utilization modeling results. Dell EMC and Vendor B both support data deduplication for S3 object storage. Vendor A only supports data compression. Based on reference architecture specifications and audited results, the Dell EMC solution was modeled at a 6:1 deduplication rate, Vendor B at 4.75:1 deduplication, and Vendor A at a 1.75:1 compression ratio.

Figure 5. Storage Requirements for In-cloud Data Protection



Source: Enterprise Strategy Group, 2017

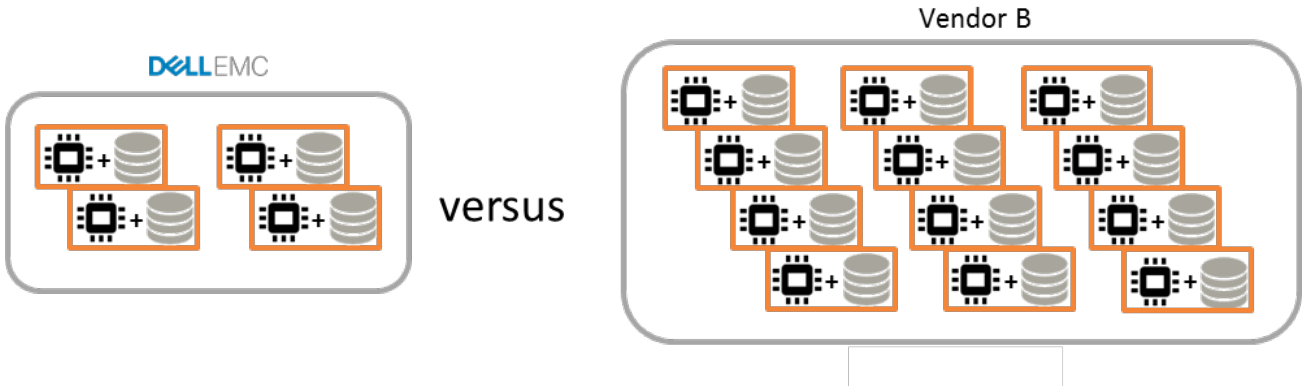
What the numbers mean:

- The Dell EMC solution's efficient use of S3 storage delivered a 70.8% reduction in backup storage repository resource requirements.
- This equals an almost 71% reduction in S3 cost due to the reduction in resource requirements.

Compute Efficiency

Finally, 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 and block storage resources required for a data protection solution to handle the load of backup jobs as the production environment was scaled from 20 to 400 TB. The resource modeling was based on reference architecture guides and publicly available pricing. Figure 6 represents the EC2 and EBS resources required to protect a 400TB production environment—the left side shows Dell EMC and the right side shows Vendor B.

Figure 6. Application Infrastructure Efficiency

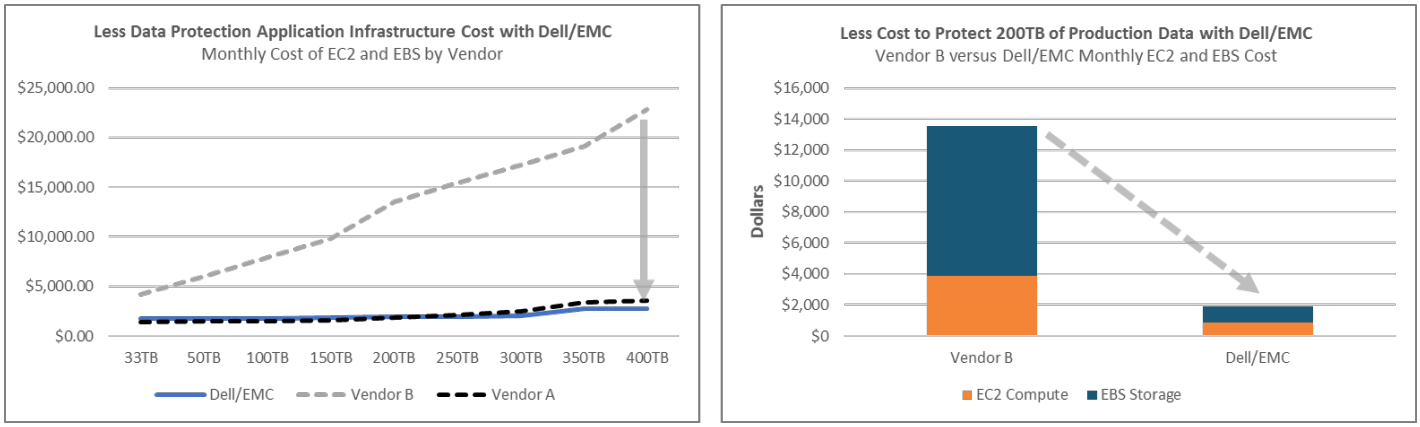


Source: Enterprise Strategy Group, 2017

The Dell EMC solution required a single backup server, one storage node, and two CloudBoost appliances. Vendor B also required only a single backup server—however, it also needed 11 data movers to handle the backup data transport, deduplication, and index processing.

Figure 7 shows the economic benefits of compute-efficient solutions. The left side of the figure shows the monthly EC2 and EBS cost of all three solutions. It should be noted that Vendor A, the costliest solution overall due to its inefficient handling of S3 storage resources, does a good job managing application resources. The right side of Figure 7 presents a detailed view of Vendor B versus Dell EMC Data Protection Software with CloudBoost EC2 and EBS resources mid-point in the cost model.

Figure 7. Data Protection Application Resource Analysis



Source: Enterprise Strategy Group, 2017

What the numbers mean:

- At scale, the Dell EMC Data Protection Software solution delivered an 87.8% reduction in cost of EC2 and EBS resources over Vendor B. This is due to the decoupling of metadata from the backup data transport process and the efficient metadata processing provided by the CloudBoost appliance.
- At 200 TB of production data to protect, roughly the mid-point of the model, the Dell EMC solution delivers an 85.8% cost reduction over Vendor B based on the total data protection application resources required. Broken down further, that equals a 77% reduction in EC2 cost and an 89% reduction in EBS costs.

The Bigger Truth

ESG research confirms that the majority of IT organizations are using cloud computing today, and there is significant interest in further adoption of these services as organizations look to control costs and drive agility in their businesses. Cloud computing will likely continue to serve as an extension of existing IT strategies, as opposed to a replacement of them, for the foreseeable future, with the use of both on-premises and public cloud services. However, key benefits such as reduced IT infrastructure costs, faster resource provisioning, and increased "time to value" for new applications and IT services, combined with the ongoing dissipation of concerns over public cloud security, reliability, and data availability, will give early adopters the confidence to entrust more critical applications and processes to cloud services over time.[4] And, as the amount of in-cloud, business-critical production applications increases, so too will the need for in-cloud, cost-efficient enterprise-class data protection.

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

- Up to 67.96% lower total monthly cost for in-cloud data protection infrastructure.
- Up to a 70.8% reduction in the amount of S3 storage required for the backup repository.
- Up to 87.8% lower monthly cost for the required EC2 and EBS 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. And inefficient uses of any one of the cloud resources can have a major impact on the cost of protection. ESG analysis found that with small cloud deployments it's hard to find a cost advantage between the audited solutions. However, choosing the wrong solution for the initially small deployment can lock a customer into a very expensive situation as its cloud environment scales. ESG confirmed that the Dell EMC Data Protection Software solution efficiently leverages EC2, EBS, and S3 cloud resources, providing a cost-efficient in-cloud data protection environment at scale.

[4] 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.

 www.esg-global.com

 contact@esg-global.com

 P. 508.482.0188