ESG Lab Review

Efficiently Protect Virtual Environments with Integrated Data Protection Appliances from Dell EMC

Date: February 2018 **Author:** Vinny Choinski, Senior Validation Analyst; and Edwin Yuen, Analyst

Abstract

This ESG Lab Review documents the results of extensive performance testing conducted on the Dell EMC Integrated Data Protection Appliance (IDPA) and a competitive server grid appliance technology solution. Testing focused on the ability of each solution to efficiently protect virtual environments at scale.

The Challenges

Protecting and recovering virtual environments continues to challenge IT organizations today. According to a recent survey of IT professionals, VM recoveries remain difficult to complete consistently. Eleven percent of ESG research respondents indicated that the most common cause of VM recovery failure is that the VM was never designated to be backed up.[1] When respondents to a separate ESG research survey were asked about their current data protection processes and technologies, 26% responded that data protection for virtual environments was one of the challenges they are dealing with, while 25% indicated that the need to reduce backup time was a particular challenge (see Figure 1).[2] If IT organizations could find solutions that would back up data more reliably and faster in their virtual environments, they could then improve their response to service level agreements (SLAs), allowing business operations to continue with little disruption.

Figure 1. Top Five Current Data Protection Processes and Technologies 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)
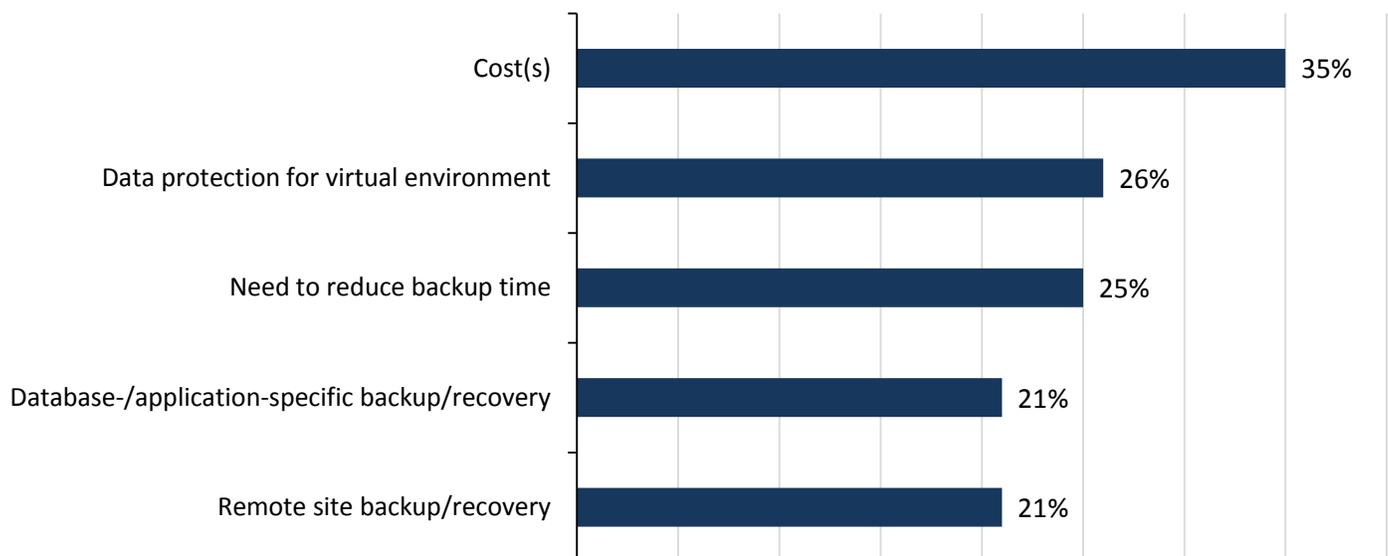
Source: Enterprise Strategy Group

---

[1] Source: ESG Brief, *Reliable Virtualization Protection Continues to Elude Many Organizations*, October 2017.
[2] Source: ESG Brief, *Why and How to Protect VMs within Hyperconverged and Converged Infrastructures*, January 2018.

Test Environment: Dell EMC IDPA Versus Vendor A

To validate the performance of each solution, ESG Lab constructed the test environment shown in Figure 2. As seen in the figure, the virtual environment to be protected consisted of nine ESXi servers attached via a pair of redundant network switches with multiple 10GbE connections to a Dell EMC VMAX All-Flash Array, functioning as primary storage. Each data protection solution to be evaluated had eight 10GbE connections to a dedicated network switch with a 20GbE link to the simulated virtual production environment.

The Dell EMC solution used in this performance review was an IDPA DP5800, consisting of three Dell servers running the latest vSphere Hypervisor hosting an Avamar Virtual Edition 7.4.1, and a physical Data Domain 6800 with two SSD devices and a tray of 4TB SAS HDDs. The competitive solution used in this review, Vendor A, was a server grid appliance with a proprietary file system that enabled backup data, metadata, and protection tasks to be distributed across each server node in the grid architecture. The Vendor A solution consisted of two physical 2U appliances. Each appliance had four Intel eight core 2.4GHz Haswell CPUs, 256 GB of DDR4 memory, and four dual port 10GbE network adaptors. The storage for each 2U appliance consisted of twelve 4TB HDD devices and four 400GB SSD devices.

Figure 2. Test Environment Overview
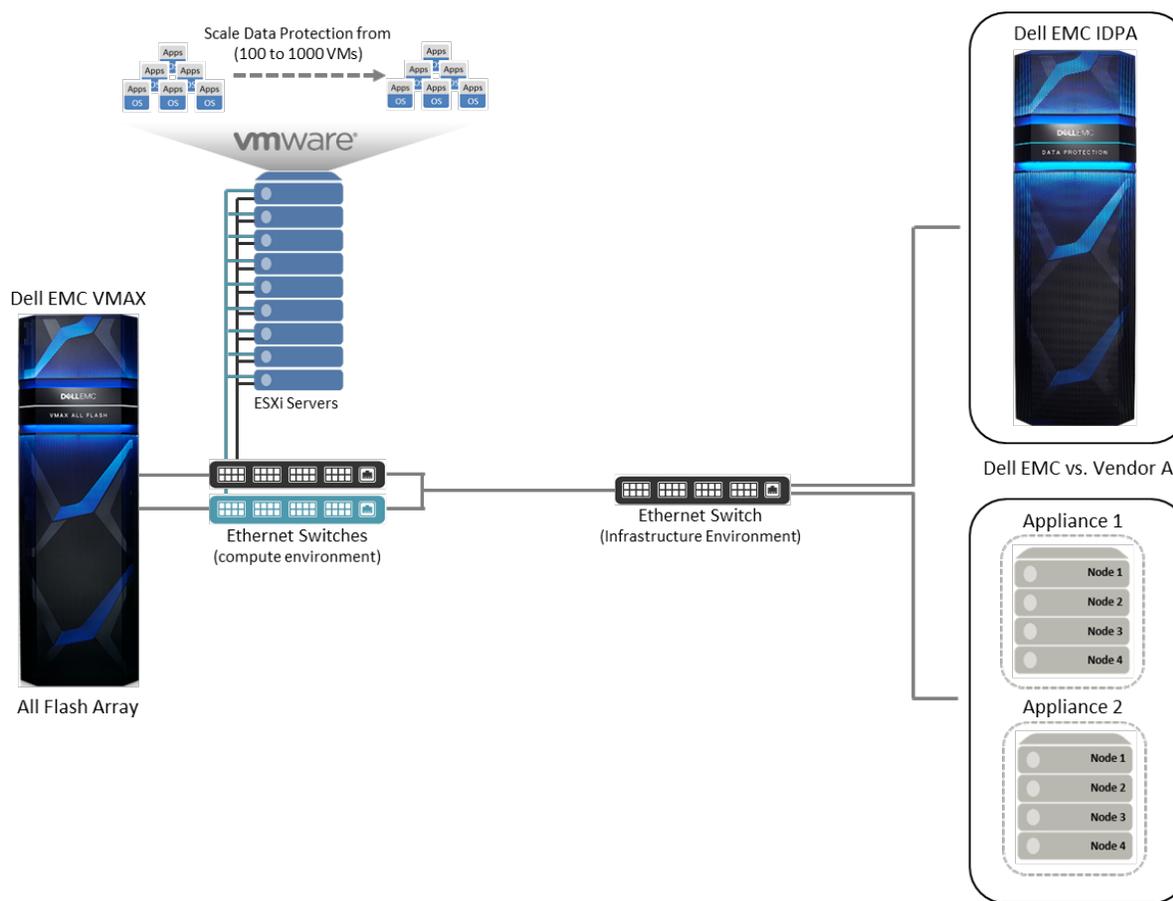
Source: Enterprise Strategy Group

ESG worked with Dell EMC to finetune a test plan for the architecture shown in Figure 2 that would demonstrate performance capabilities for data protection tasks critical to the protection of virtual environments. The IDPA hardware infrastructure was configured to match the Vendor A solution as closely as possible. It should also be noted that the IDPA solution was throttled down to 24 streams and three vProxy VMs to match the stream count of the Vendor A solution. Then, the rigorous test plan tasks were executed against each solution over a three-month testing period.

Testing Methodology

To demonstrate data protection performance of virtualized environments at scale, the following test methodology was leveraged. First, a group of virtual machines consisting of 80 Windows VMs, 10 Linux VMs, and 10 Oracle/MSSQL VMs was deployed on the ESXi server farm in the test environment. An initial full backup with a 14-day retention period was run on the first 100 VMs to both the IDPA and the Vendor A data protection solutions. Next, a file generation utility was used to create a 3% data change rate on the Windows and Linux file systems. The DBgen[3] utility was used to create a 5% data change rate on each database VM. Another 100 VMs were added to the test environments and a second backup was run. This process was repeated in increments of 100 VMs until the test environment was protecting 1,000 VMs. To complete the performance testing, four steady state backups of 1,000 VMs were run.

Backup Performance

ESG started its performance analysis by auditing the overall backup time each solution required to protect each group of 100 VMs. Figure 3 shows total duration times for each backup group as the test environment was scaled from 100 to 1,000 VMs. For the last four backup jobs shown in Figure 3, only the incremental changes to the file systems and database data needed to be backed up; no new VMs were added to the environment for these jobs.

Figure 3. Total Backup Durations
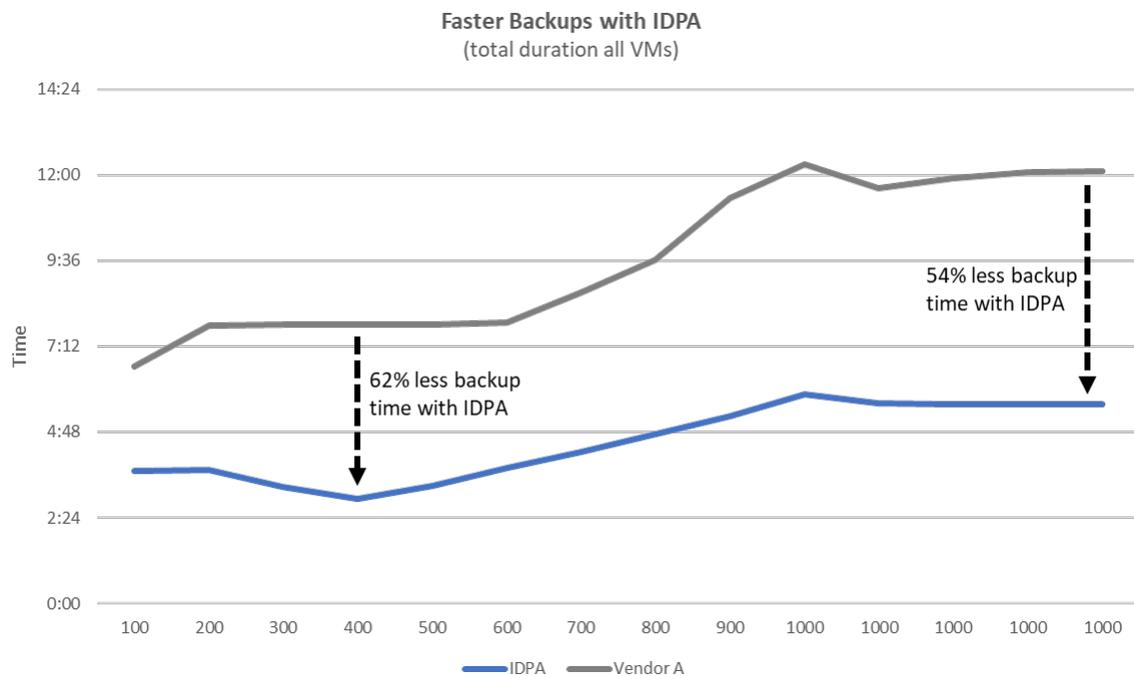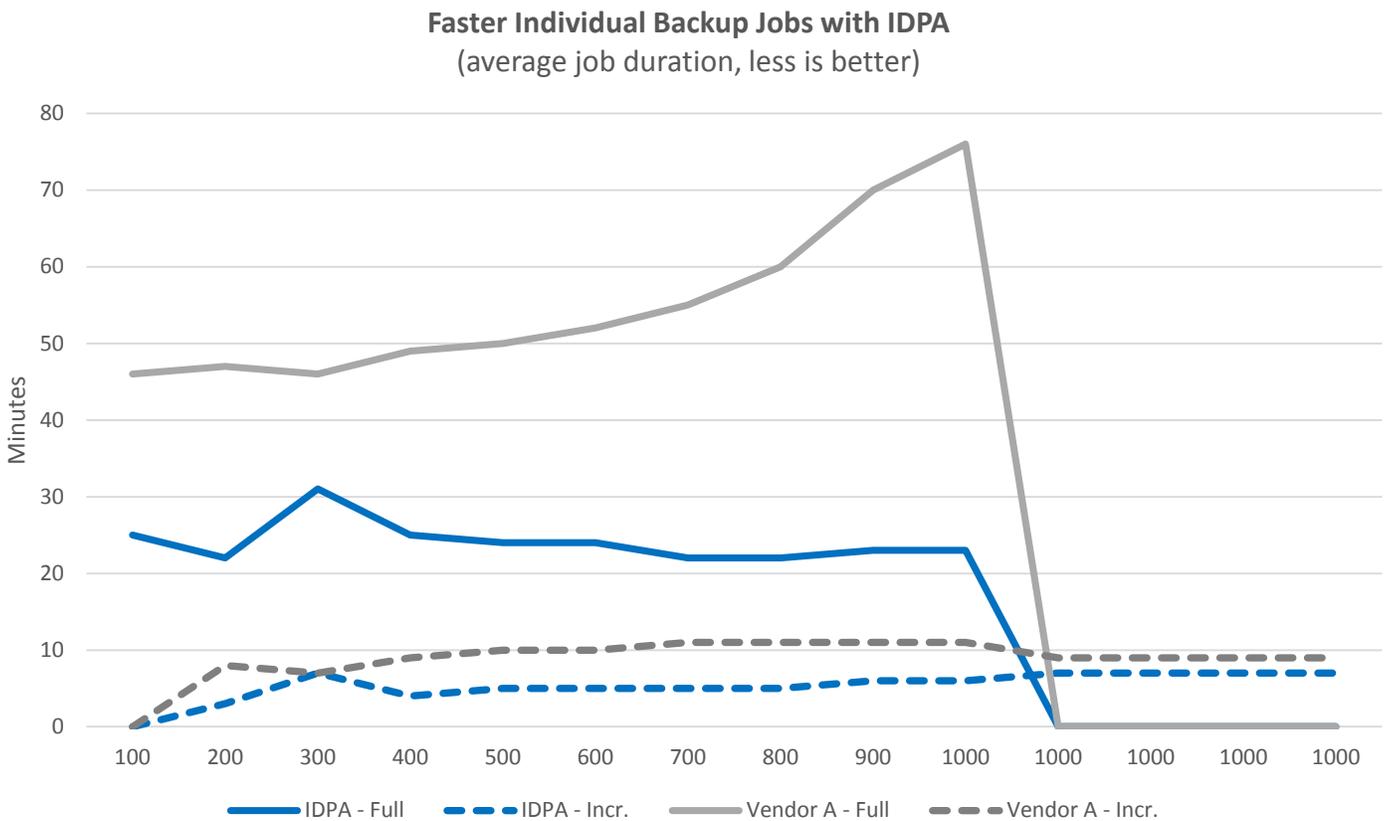

Source: Enterprise Strategy Group

What the Numbers Mean

- For steady state backups, the IDPA solution shrunk the backup window by 54% over the Vendor A solution.
- The largest IDPA advantage (64% less backup time) was observed with a 400-VM workload.
- The IDPA solution was able to complete a steady state backup in 29 minutes when throttling was removed.
- With throttling removed, the IDPA solution was 96% faster than the Vendor A solution.

---

[3] DBgen is a database population program that can be downloaded at the following website. http://www.tpc.org/tpch/

To further understand this cumulative effect, ESG also examined the individual backup times as we increased the VM workload in increments of 100. We observed that, with IDPA, for every 100 VMs added, backup times for individual jobs were consistently lower than those for Vendor A's solution. As shown in Figure 4, ESG separately graphed both full and incremental average backup durations to better understand the architecture benefits of each solution. The solid lines in Figure 4 represent average full backup duration, while the dashed lines represent average incremental backup duration. During analysis, ESG noted a continual increase in full backup duration for Vendor A, as demonstrated by the solid gray line in Figure 4. It should be noted that this duration increase was not observed with the IDPA solution. Based on this data, ESG can further verify that Dell EMC can achieve overall faster backup times with its IDPA.

Figure 4. Average Backup Job Duration

Faster Individual Backup Jobs with IDPA
(average job duration, less is better)

What the Numbers Mean

- Depending on the workload, IDPA has a two- to six-minute advantage over Vendor A for incremental backups.
- Both solutions maintained fairly consistent incremental backup performance as they were scaled during testing.
- Vendor A showed a slight incremental performance increase once the full backup cycle completed.
- Vendor A full backup durations increased significantly as the overall backup load increased during testing.
- IDPA maintained fairly consistent full backup performance as the overall backup load increased.
- Depending on the workload, IDPA has a 15- to 53-minute advantage over Vendor A for full backups.

Finally, ESG reviewed data transfer efficiency of both solutions and found that IDPA outperformed the alternative solution by orders of magnitude from 100 to 1,000 VMs. ESG attributes the ability of the Dell EMC solution to efficiently transfer

backup data to its extensive experience in deduplication given its mature and customer-proven Data Domain technology. The IDPA solution also uses client-side deduplication so that only the changed data on the client needs to be transferred over the network to the backup target where additional inline deduplication is performed.

Figure 5. Efficient Deduplication and Data Transfer
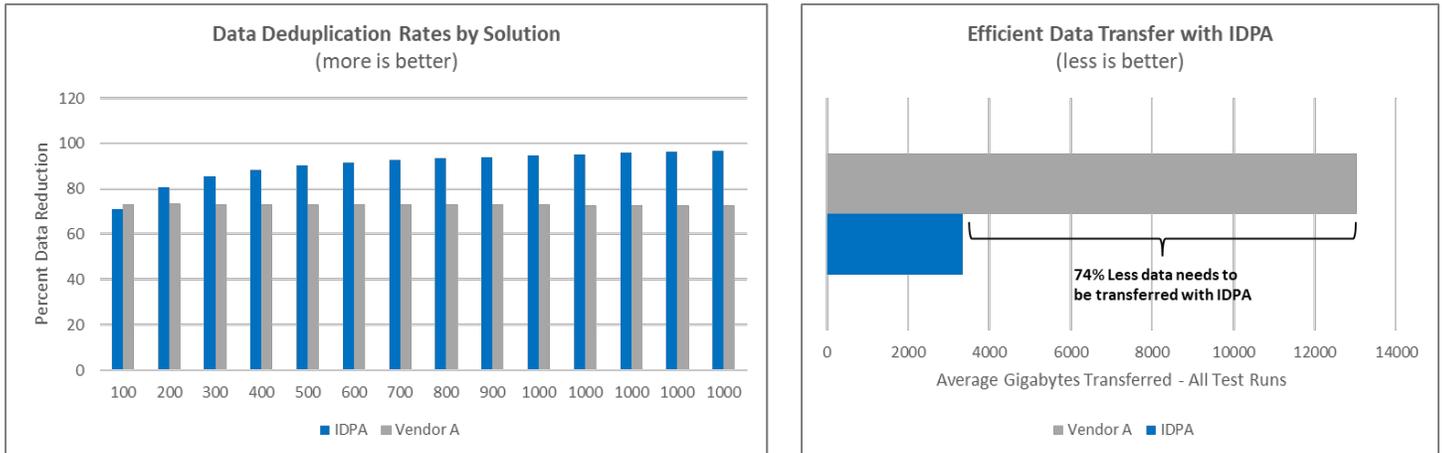
Source: Enterprise Strategy Group

What the Numbers Mean

- During scale testing, ESG observed a data deduplication reduction rate of nearly 97% for the IDPA solution.
- The Vendor A solution was only able to attain a data deduplication reduction rate of 73% for the same tests.
- This means the IDPA solution consumes less storage for the same amount of data protection over Vendor A.
- During performance testing, ESG also observed up to a 74% reduction in the amount of data that needed to be transferred during backup operations over the Vendor A solution.
- This reduction in the amount of data transferred for protection is due to the efficient source-side deduplication of the IDPA technology.

ⓘ Why This Matters

Backup performance is a primary concern for any organization that is concerned with minimizing disruption to ongoing business operations. However, performance levels become more critical in large enterprises as IT deals with numerous virtual workloads simultaneously. Unlike smaller enterprises, large enterprises face the additional challenge of backing up mixed virtual workloads at scale. Handling backups at scale can be effectively accomplished by technology that has been successfully deployed and operated in the field with numerous large enterprise customers.

ESG Lab validated that the Dell EMC IDPA delivers overall better backup performance than the alternative vendor's solution. We examined the cumulative backup times for VMs as the number scaled from 100 to 1,000 and saw that times for the Dell EMC IDPA were consistently lower than those of Vendor A's solution. At 1,000 VMs, the cumulative backup time achieved by the Dell EMC IDPA was 54% less than that of the alternative solution.

Recovery Performance

Data protection performance, especially when it comes time for recovery, is an important aspect of any data protection solution. It becomes even more important when the number of virtual machines and the amount of corporate data start to scale in leaps and bounds as virtual environments so often do. Meeting recovery requirements without impacting the system resources needed to maintain daily backup operations can be very challenging.

To analyze the recovery performance capabilities of the two architectures being validated, ESG audited the results of the following test scenarios. First, as shown in Figure 6, an instant access recovery of ten VMs was carried out from both data protection solutions. The instant recovery process runs each VM from a backup image stored on the respective IDPA and Vendor A data protection targets. Next, Microsoft's Diskspd utility was used to generate different I/O patterns to each of the ten VMs on both data protection platforms. Twelve different test profiles were simultaneously run on all 20 VMs for fifteen minutes each. The I/O test used random writes at block sizes that ranged from 8K to 64K with twelve simultaneous threads and read/write profile rates of 90/10, 80/20, and 70/30. Each test was first run with no daily backup operations running and then repeated while the daily incremental was active.

Figure 6. Instant Recovery Overview
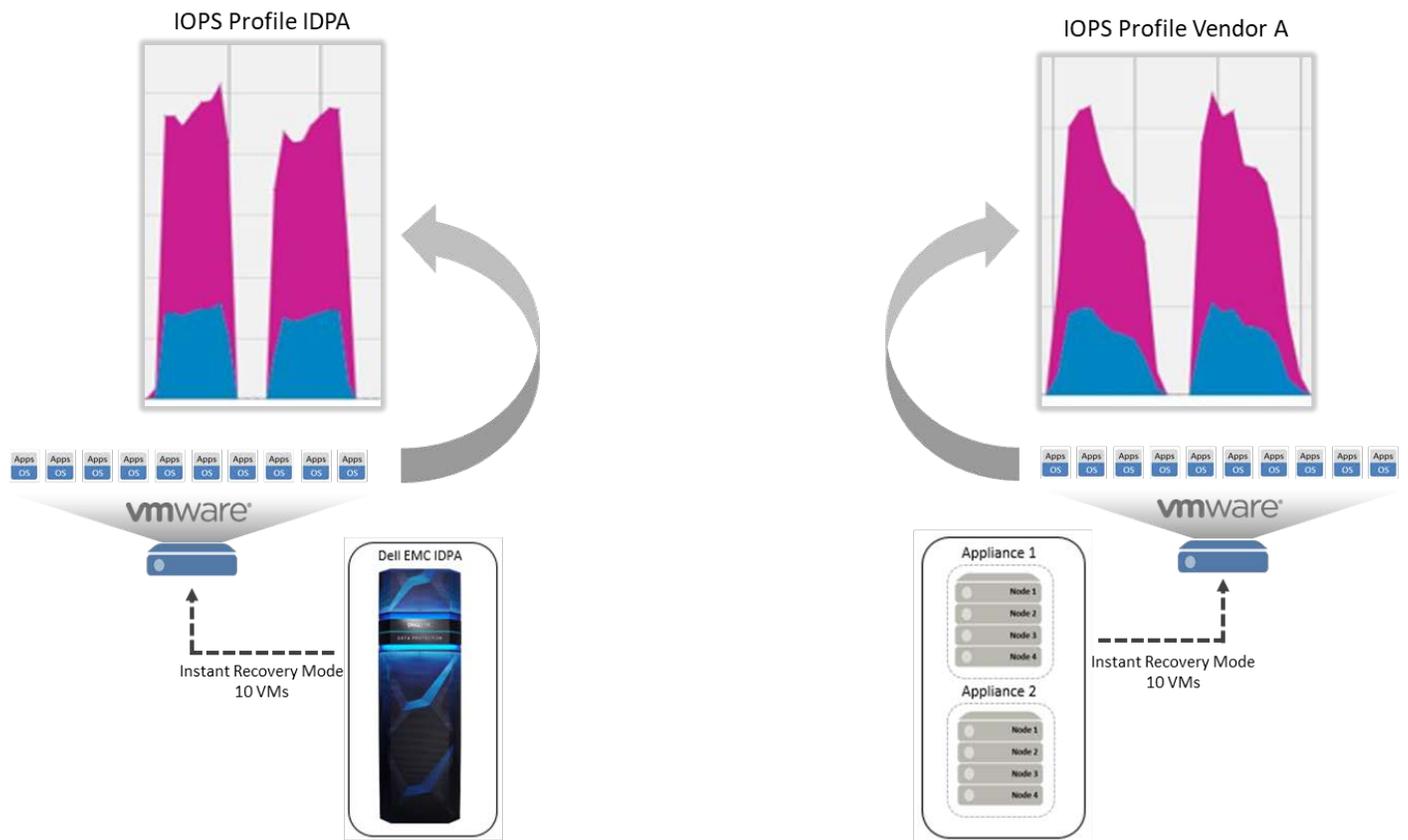
Source: Enterprise Strategy Group

What the Data Means

- The upper left and upper right screen captures show the read/write profiles in purple and blue respectively.
- Notice that the Vendor A IOPS profile on the right side initially peaked and then quickly tapered off the longer the test ran, while the IDPA profile on the left maintained or even gained performance the longer it ran.

Finally, Figure 7 shows the actual IOPS performance results of the instant recovery process for the two solutions.

Figure 7. Instant Recovery Performance

Instant Recovery Performance
(more is better)
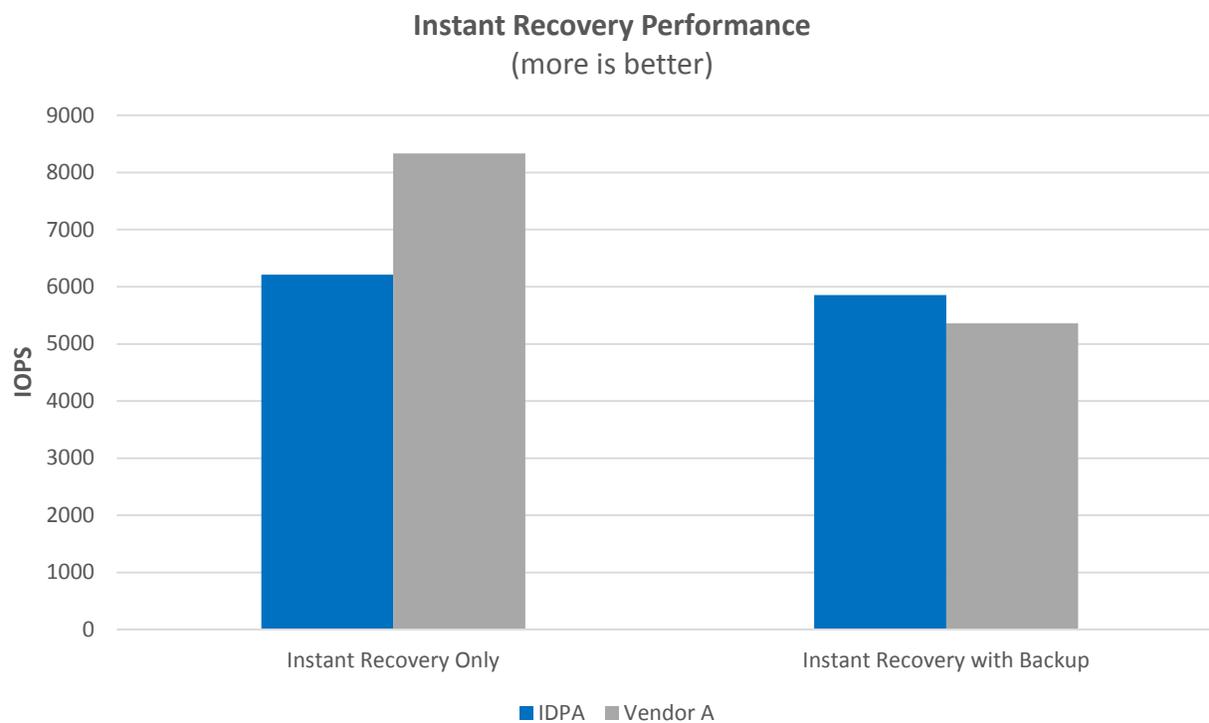
Source: Enterprise Strategy Group

What the Numbers Mean

- For instant recoveries with no backup operations running in the background, IDPA was able to deliver 6,210 IOPS while Vendor A delivered 8,336 IOPS.

- However, when the daily incremental operation was added to the workload mix, IDPA delivered 5,855 IOPS while Vendor A delivered 5,359 IOPS.

> ℹ **Why This Matters**
>
> Not only is it important for a large organization to reduce backup times for VMs, but it is also necessary that the data recovery is efficient. Both backup and recovery performance can affect how quickly an organization can minimize any disruption to normal business operations. In a large enterprise, backup and recovery performance becomes critical as it deals with other workloads such as backup operations at scale. IT cannot afford for backup and recovery operations to interrupt business operations, and affect end-user systems and data recoverability.
>
> ESG verified that the Dell EMC IDPA performs better than Vendor A's solution when backing up and recovering VMs simultaneously. While our testing saw that the alternative solution can recover VMs faster than the Dell EMC IDPA, the IDPA completes VM recovery with backup faster. The fact that Dell EMC IDPA can do both source- and target-side deduplication enables end-users to access data faster, allowing them to run applications even while backups are streaming in the background. Again, Dell EMC shows its proven technology can enable large enterprises to back up data while minimizing disruption to ongoing business operations.

The Bigger Truth

There is always going to be some shiny new technology, whose hype makes it appear as "the next big thing." Similarly, there will always be startups that sparkle like stars, particularly in traditional IT categories that have been historically plagued with antiquity. To be fair, the IT landscape continues to rapidly evolve, and yesteryear's data protection technologies will not meet tomorrow's requirements as-is.

Instead of hastily throwing away technology that has reliably proven itself in favor of the shiny new one, it may be better to reimagine and repurpose instead of trying to reinvent. Imagine a technology suite composed of building blocks bought in kits that originally worked well together, but today, isn't as functional as it used to be. Yet, each kit still has some unique elements to leverage. What if you took the best parts of those kits and built something new that was designed for what you want for the future? This "repackaging" takes advantage of what you know and have relied on, reimagined and repurposed for tomorrow.

That is one way to think about the Dell EMC IDPA solutions, whose building block backup and deduplication technologies have been relied upon by some of the largest enterprises in the world for decades. Why? As large enterprises needed heterogeneous protection and storage scalability, Dell EMC has continually met those needs, and with IDPA, now makes these technologies and capabilities available to a broad base of mid-size and smaller organizations. Aside from scalability, backup performance (to not encumber the production environment) and reliability (when recovery matters) are key requirements for IT leaders developing their future data protection strategy—and the results of ESG's testing show that Dell EMC continues that innovation.

Over the years, a lot of shiny new technology started out as the next big thing, was useful for a few years, and then failed to grow beyond its initial promise and hype. In contrast, key existing technologies are often overlooked simply because they are not new. The key to efficient innovation is to leverage all the technology available to best solve business problems. By providing solutions that can be used for a range of business needs, from existing to new, one particular building block company continues to be a household name. When it comes to data protection, you have to rely on first "building" a strong foundation out of "components" that matter before you have a solution that can truly exceed expectations.