Dell APEX outperformed comparable Amazon EC2 instances on a decision-support workload

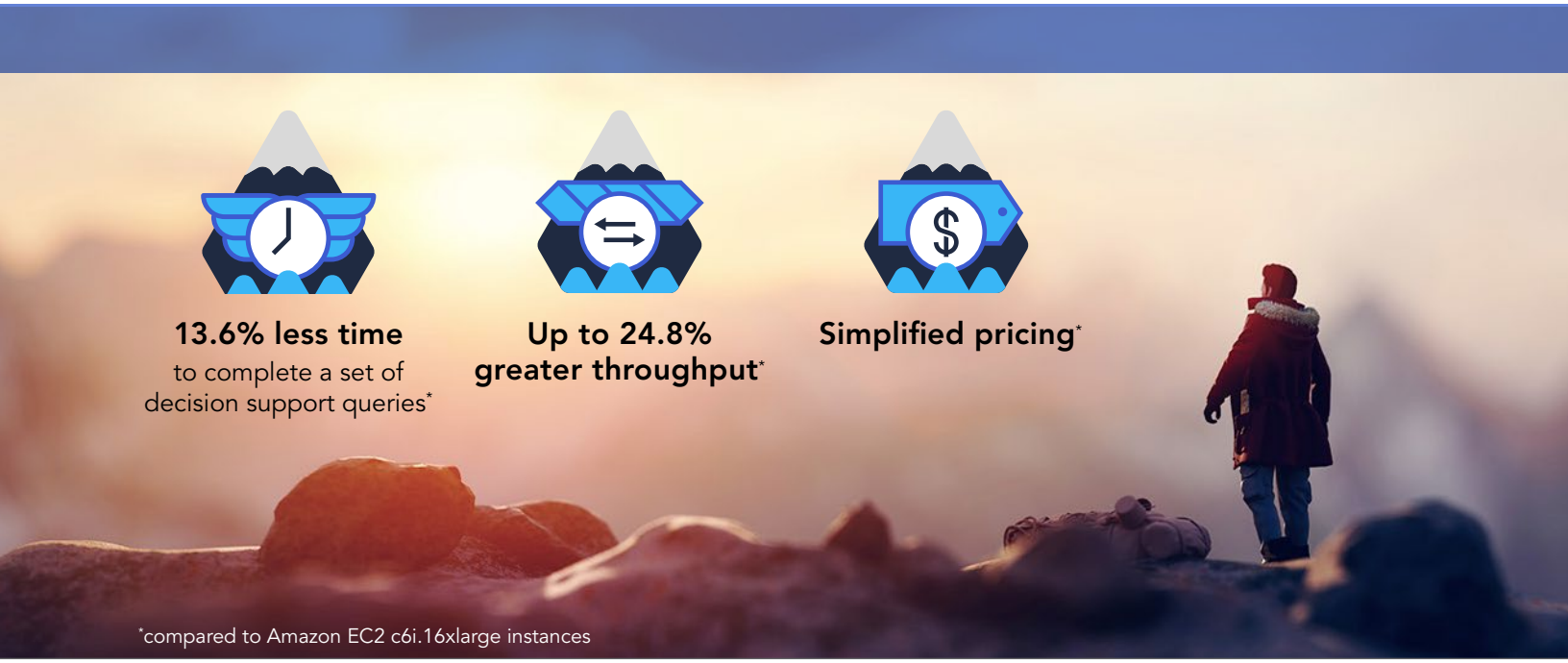
A solution using Dell APEX Private Cloud and Dell APEX Data Storage Services Block generated data-driven insights earlier than Amazon EC2 c6i.16xlarge instances with Amazon Elastic Block Store (EBS) storage

Big data has changed the way that enterprise decision-making takes place. When New Vantage surveyed 94 executives from Fortune 1000 and industry-leading organizations, 73.7 percent of the respondents said their organizations had appointed a Chief Data Officer (CDO), up from 12 percent in 2012, and 97.0 percent had invested in big data initiatives.[1]

The decision support systems (DSS) around which organizations build their big data initiatives typically analyze great quantities of information and generate recommendations for next steps based on this data. For DSS workloads to be most valuable, the data must be as current as possible. This means that performance is an important consideration as companies weigh their options for where to run DSS workloads. In addition to being compute-intensive, decision support workloads also demand a great deal from storage. Companies may consider instances from public cloud service providers such as Amazon Web Services, but a private cloud solution can deliver performance advantages.

We ran a decision support big data workload in two environments whose VMs had the same vCPU, memory, and storage drive capacity: (1) an eight-node Hadoop cluster running on Dell APEX Private Cloud customer-managed dynamic nodes, enabled by Intel® Xeon® Gold 6338 processors and using Dell APEX Data Storage Services Block, and (2) an eight-node Hadoop cluster of Amazon EC2 c6i.16xlarge instances using two 4,000-IOP io2 drives for the data storage and another 4,000-IOP io2 drive for the Spark temporary scratch space.

The Dell APEX solution completed a set of decision support queries in 13.6 percent less time than the Amazon EC2 solution. It also delivered 24.8 percent greater read throughput and 19.2 percent greater write throughput. These performance advantages could put actionable insights into the hands of decision-makers sooner.



Make your data-driven decisions earlier

The Dell APEX Private Cloud solution we tested consisted of 256 General Purpose compute instances paired with a 50TB Dell APEX Data Storage Services Block, Balanced performance tier. We used these resources to set up four dynamic node (no vSAN) VMware® vSphere® Dell APEX Private Cloud nodes, 256 Intel Xeon Gold 6338-based CPU cores, and 2,048 GB of RAM. We configured a Hadoop/Spark cluster of eight VMs. Each VM had 64 vCPUs, 128 GB of memory, 2 TB of storage to house the Hadoop Distributed File System, and 60 GB of storage to house the operating system. We also configured a manager VM with 16 vCPUs, 32 GB of memory, 4 TB of data storage, and 2 TB of temporary storage.

We also deployed a Hadoop cluster of eight compute-optimized Amazon EC2 c6i.16xlarge instances powered by 3rd Generation Intel Xeon Scalable processors. According to Amazon, these instances "deliver up to 15% better price performance than C5 instances for a wide variety of workloads…. support up to 128 vCPUs per instance, which is 33% more than C5 instances….and offer up to 200 Gbps of network bandwidth and up to 2x higher packet-processing performance than C5n instances."[2]

About the Dell APEX offerings we used in our study

Dell APEX delivers cloud resources as a service, wherever an organization needs them. We chose a customer-managed private cloud option for this study, but organizations can deploy Dell APEX offerings in their data center, a co-location facility, or an edge location.[3] IT staff configure and manage the lifecycle of their Dell APEX solutions through the Dell APEX Console, a unified web interface that we did not use in this study.

Dell APEX Private Cloud

Dell says APEX Private Cloud is "perfect for getting started with cloud or expanding your data center out to the edge."[4] Dell provides APEX Private Cloud users with a single point of contact to assist with their cloud maintenance and needs. Dell claims the service provides a small footprint and features built-in lifecycle automation.[5]

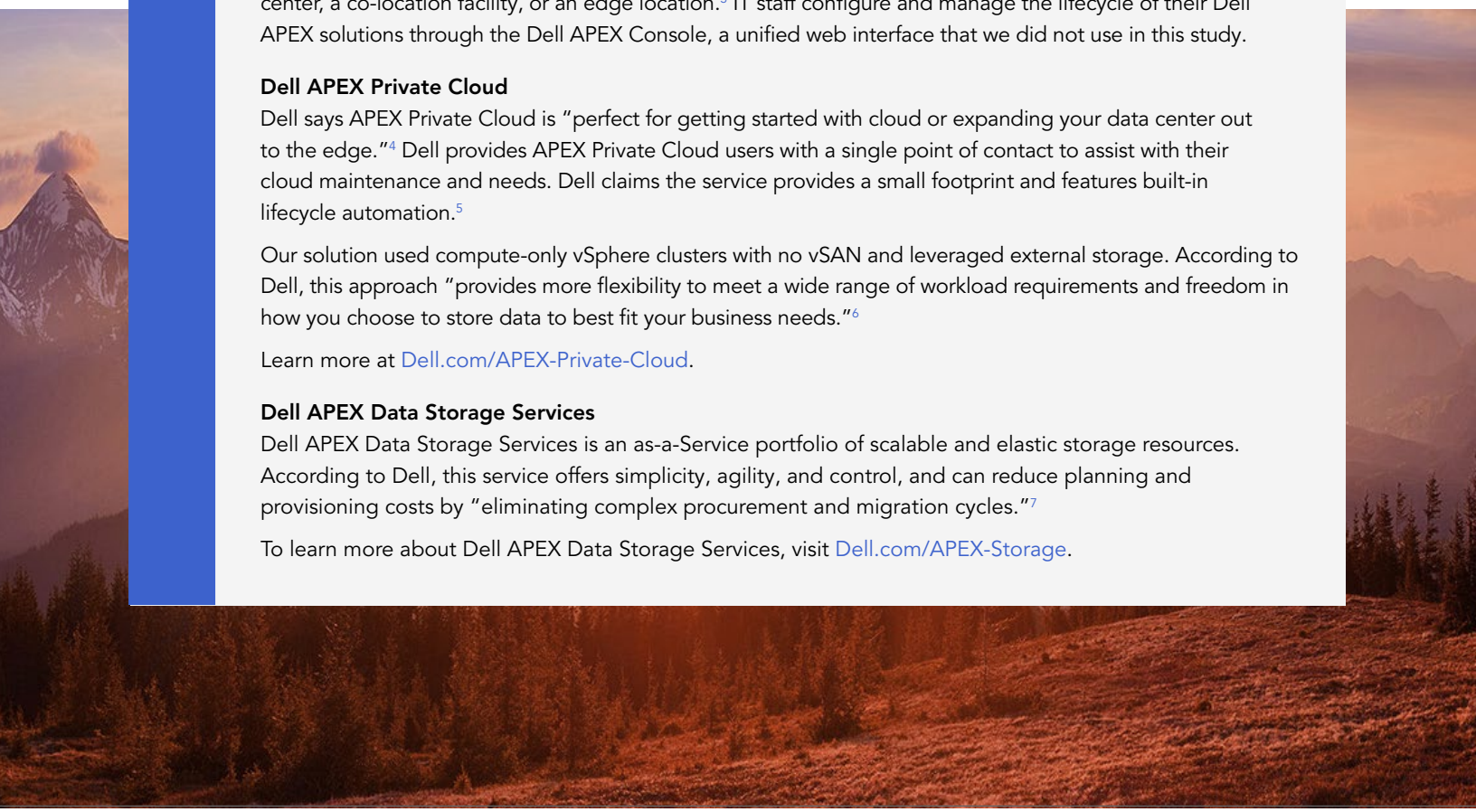
Our solution used compute-only vSphere clusters with no vSAN and leveraged external storage. According to Dell, this approach "provides more flexibility to meet a wide range of workload requirements and freedom in how you choose to store data to best fit your business needs."[6]

Learn more at Dell.com/APEX-Private-Cloud.

Dell APEX Data Storage Services

Dell APEX Data Storage Services is an as-a-Service portfolio of scalable and elastic storage resources. According to Dell, this service offers simplicity, agility, and control, and can reduce planning and provisioning costs by "eliminating complex procurement and migration cycles."[7]

To learn more about Dell APEX Data Storage Services, visit Dell.com/APEX-Storage.



What we learned

Figure 1 shows the total time to complete the decision support workload for the Hadoop cluster of database VMs in the Dell APEX solution and the cluster of Amazon EC2 c6i.16xlarge instances. The Hadoop/Spark cluster on the Dell APEX solution took 3 hours 12 minutes to run a set of queries on approximately 4 TB of data. That was 13.6 percent less time than the Amazon EC2 c6i.16xlarge Hadoop/Spark cluster, which needed 3 hours 43 minutes to analyze the same 4 TB of data, due to the EBS throughput limitation of 500 MiB per second at a 256KB I/O size.[8] The sooner your DSS applications can execute queries, the sooner the information will reach those who need it to make informed business decisions.

The Hadoop/Spark cluster on the Dell APEX solution also delivered greater read and write throughput while running the set of queries. As Figure 2 shows, the Hadoop cluster of database VMs in the Dell APEX solution achieved 24.8 percent greater read throughput and 19.2 percent greater write throughput than the cluster of Amazon EC2 c6i.16xlarge instances. This greater performance helps put actionable insights into the hands of decision-makers earlier.

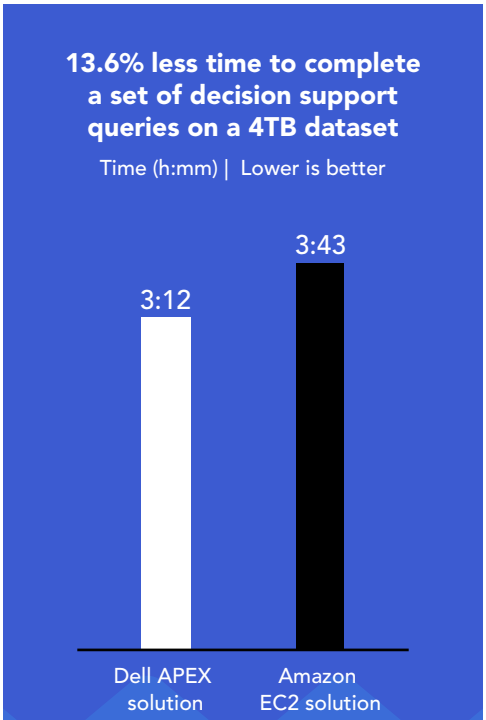


Figure 1: The total time that the Dell APEX solution and the Amazon EC2 c6i.16xlarge instances needed to complete the decision support workload we used in testing. Lower is better. Source: Principled Technologies.

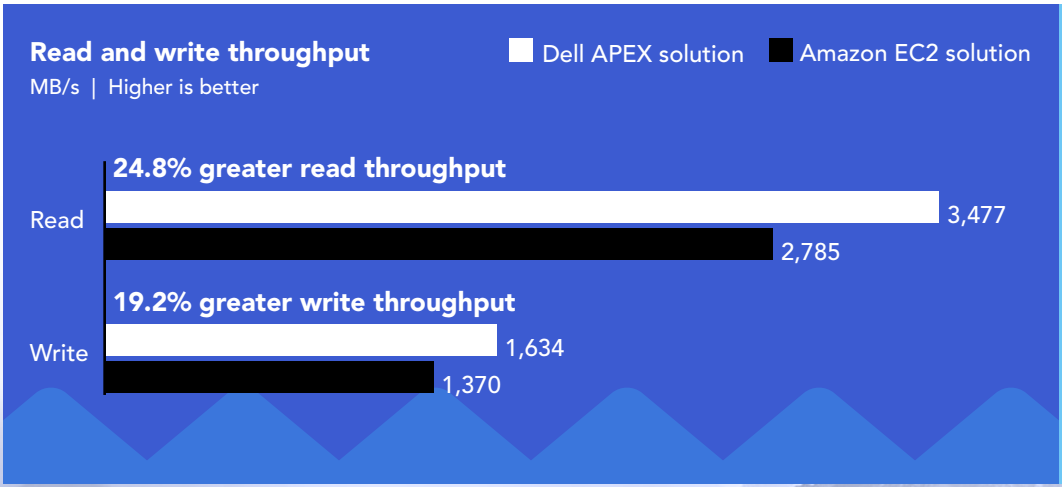


Figure 2: Read and write throughput the Dell APEX solution and the Amazon EC2 c6i.16xlarge instances achieved while running the decision support workload we used in testing. Higher is better. Source: Principled Technologies.

How the Dell APEX solution simplifies pricing

Our goal in this study was to explore the performance potential of two comparably priced solutions. As we said on page 2, the Dell APEX solution we tested used four dynamic node (no vSAN) VMware® vSphere® Dell APEX Private Cloud nodes, 256 Intel Xeon Gold 6338-based CPU cores, and 2,048 GB of RAM. We configured a Hadoop/Spark cluster of eight VMs. Each VM had 64 vCPUs, 128 GB of memory, 2 TB of storage to house the Hadoop Distributed File System, and 60 GB of storage to house the operating system. We also configured a manager VM with 16 vCPUs, 32 GB of memory, 4 TB of data storage, and 2 TB of temporary storage. We did not have to make any subscription changes to this setup, which cost $13,062 per month for both compute and storage.

We also deployed a Hadoop cluster of eight compute-optimized Amazon EC2 c6i.16xlarge instances powered by 3rd Generation Intel Xeon Scalable processors. Our first configuration included two 1TB driver per Spark worker node, which together formed a large cluster at a monthly cost of $13,788 (roughly 5 percent higher than the Dell APEX solution). We believed this would provide sufficient space for our data with additional extra room. Once testing began, we learned that Spark used up the extra room for temporary storage and we needed more space. (This was also the case for the Dell APEX solution, but we had the additional space.) In the interest of time, we added one more 1TB drive per Spark worker node, which increased the monthly cost of the Amazon EC2 solution to $16,892 (roughly 22 percent higher than the Dell APEX solution). This extra allocation of space was mostly likely higher than necessary, and the true monthly cost of the Amazon EC2 configuration would fall somewhere between. (Note: We present configuration and pricing details in the science behind the report.)

About the DSS workload we used

To investigate how Dell APEX Private Cloud and AWS EC2 instances handled DSS workloads, we used a TPC-DS-derived benchmark that simulates a decision support system. The benchmark ran 99 queries against a 4TB dataset and measured query response time. Because we derived our workload from the TPC-DS benchmark, it is not comparable to published TPC-DS results. To learn more about our workload, see the science behind the report.

We explain our experience because we believe it illustrates the two solutions' different approaches to pricing. Amazon uses a consumption-based pricing model. When you buy a service, the amount you pay is determined by the resources you select (instances, storage, networks, etc.) times the amount you use them in a given period. Dell APEX pricing is much simpler.

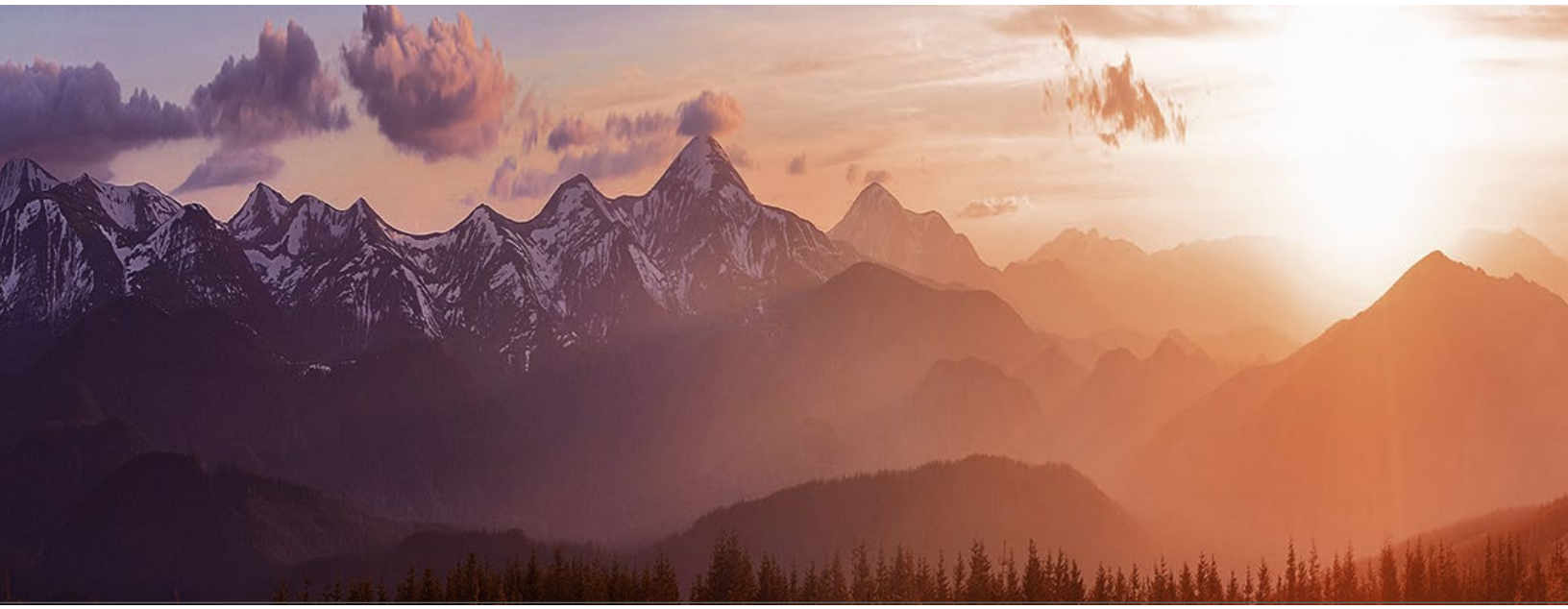
For APEX Private Cloud, customer pricing is based on the number of compute instances and terabytes of storage configured when initiating the subscription. The customer configures their APC clusters using predefined instance types (i.e., Compute Optimized, General Purpose, Memory Optimized, Large Memory Optimized or Accelerator Optimized instances). Customers pay a price per instance and a price per terabytes per month regardless of usage.

For APEX Data Storage Services, the customer selects a base capacity, a commitment to a minimum amount of storage. Dell provides additional on-demand buffer capacity on top of the base commitment, allowing for incremental growth and ensuring that the customer doesn't run out of available capacity. For both base and on-demand usage, customers pay a single rate on a dollars-per-terabyte-per-hour basis. The greater the base capacity amount, the lower the terabyte-per-hour rate. There are no overage fees.

Another advantage of the Dell APEX pricing model is that if your usage fluctuates from month to month, pricing is more predictable than it would be with Amazon.

About the Intel Xeon Gold 6338 Processor

Part of the 3rd Generation Intel Xeon Scalable Processor family, the Intel Xeon Gold 6338 Processor has 32 cores, 64 threads, a maximum turbo frequency of 3.20 GHz, a processor base frequency of 2.00 GHz, and a 48MB cache. According to Intel, this processor family offers optimization for "cloud, enterprise, HPC, network, security, and IoT workloads with 8 to 40 powerful cores and a wide range of frequency, feature, and power levels."[9]



Conclusion

For your company to reap the benefits of big data analysis can require a substantial investment, so it's crucial to determine which platform can deliver the vital insights you need quickly. In our performance testing of two solutions with comparable specifications, we found that the Dell APEX solution took 13.6 percent less time to run a set of DSS queries on a 4TB database than a solution based on Amazon EC2 c6i.16xlarge instances. Throughput on the Dell APEX solution was also greater, with an advantage of 24.8 percent for read throughput and an advantage of 19.2 percent for write throughput. This increased performance could help your business by giving decision-makers access to critical information sooner.

1. New Vantage Partners, "Data and AI Leadership Executive Survey 2022, Executive Summary of Findings," accessed December 8, 2022, https://c6abb8db-514c-4f5b-b5a1-fc710f1e464e.filesusr.com/ugd/e5361a_2f859f3457f24cff9b2f8a2bf54f82b7.pdf.
2. Amazon, "Amazon EC2 C6i Instances," accessed January 9, 2023, https://aws.amazon.com/ec2/instance-types/c6i/.
3. Dell Technologies, "Simplify your multicloud with a consistent experience everywhere," accessed December 8, 2022, https://www.dell.com/en-us/dt/apex/cloud-services/index.htm.
4. Dell, "APEX Private Cloud," accessed December 12, 2022, https://www.dell.com/en-us/dt/apex/cloud-services/private-cloud.htm.
5. Dell, "APEX Private Cloud."
6. Dell, "APEX Specification Sheet," accessed January 9, 2023, https://www.delltechnologies.com/asset/en-us/solutions/apex/technical-support/h19142-apex-private-cloud-spec-sheet.pdf.
7. Dell Technologies, "APEX Data Storage Services solution brief," accessed December 8, 2022, https://www.delltechnologies.com/asset/en-us/solutions/apex/briefs-summaries/apex-data-storage-services-overview.pdf.
8. AWS, "Provisioned IOPS SSD volumes," accessed December 14, 2022, https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/provisioned-iops.html.
9. Intel, "3rd Gen Intel® Xeon® Scalable Processors," accessed March 23, 2023, https://www.intel.com/content/www/us/en/products/docs/processors/xeon/3rd-gen-xeon-scalable-processors-brief.html.

Read the science behind this report ▶

▶ View the original version of this report at https://facts.pt/XJvpK9D

This project was commissioned by Dell Technologies.

Principled Technologies®

Facts matter.®