

Technical Review

Performance Testing of Dell EMC PowerScale for Google Cloud

Date: April 2021 **Author:** Kerry Dolan, Sr. IT Validation Analyst; Alex Arcilla, Validation Analyst; and Brian Garrett, EVP Validation Services

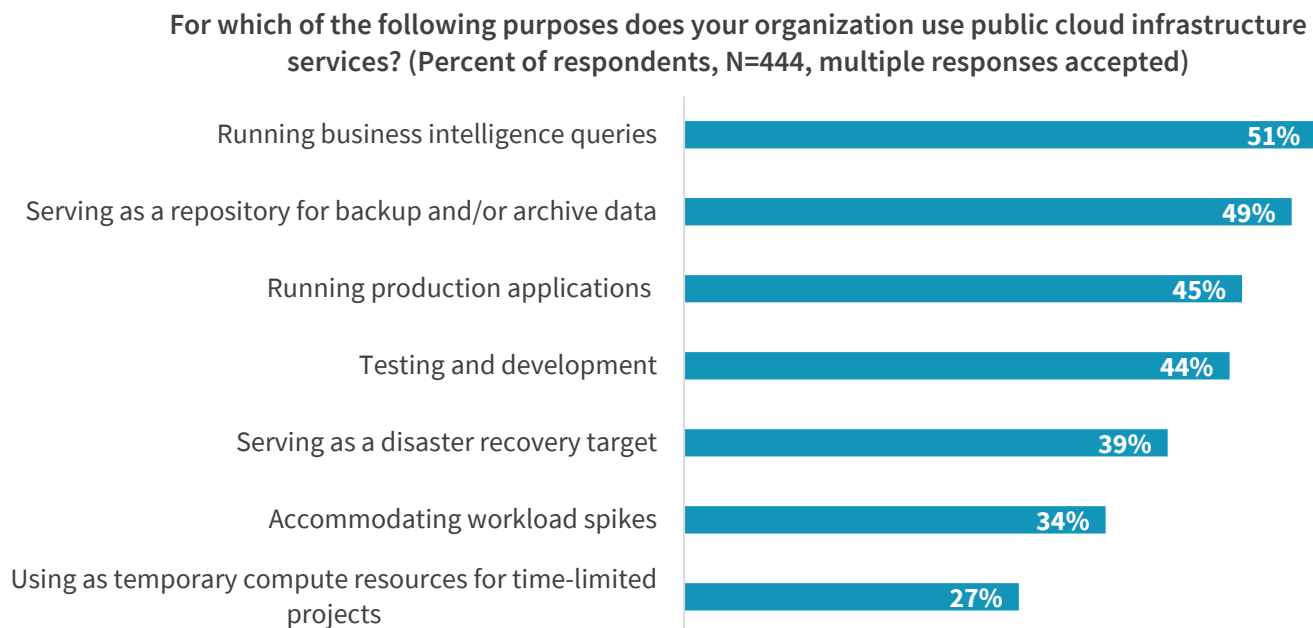
Abstract

This Technical Review documents ESG's audit of testing designed to demonstrate the high-performance capabilities of Dell EMC PowerScale for Google Cloud. This solution enables large capacity file applications such as those in analytics, media and entertainment, life sciences, and high-performance computing (HPC) to benefit from the flexibility and cost efficiency of Google Cloud while delivering high performance.

The Challenges

Organizations are increasingly moving workloads to the cloud to take advantage of agility, flexibility, and cost efficiency benefits. According to ESG research, 94% of respondents are currently using public cloud services for infrastructure- or software-as-a-service (IaaS or SaaS). They report that 74% of their current workloads could be targets to move to the cloud in the next five years. However, only 45% report that they are running production applications in public cloud IaaS (see Figure 1).¹

Figure 1. Use of Public Cloud IaaS



Source: Enterprise Strategy Group

While file data often accounts for at least half of an organization's on-premises data, very little file data is stored in the cloud. This is primarily due to performance and scale limitations of the solutions available today. In particular, large file workloads such as those used in media and entertainment, life sciences, and commercial HPC are not running in the cloud

¹ Source: ESG Master Survey Results, [2020 Technology Spending Intentions Survey](#), January 2020.

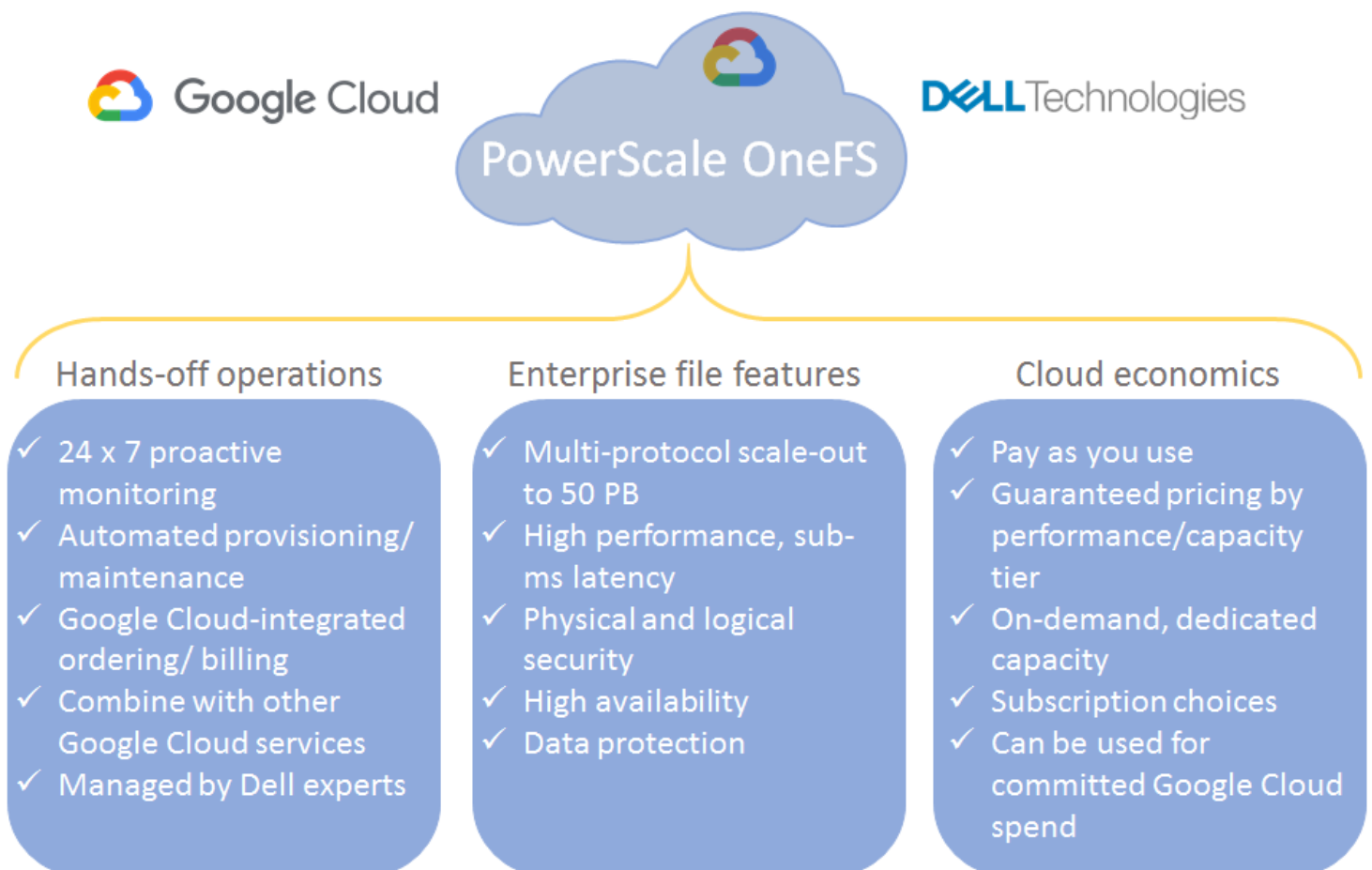
because to date there has been no solution that has offered the levels of performance and scalability these applications demand. And while cloud-resident data analysis tools can be applied to on-premises data, those processes suffer from performance and WAN network bandwidth bottlenecks, which make them both slow and expensive.

The Solution: Dell EMC PowerScale for Google Cloud

Dell Technologies has partnered with Google to offer the PowerScale OneFS file system as a service in Google Cloud. This offering enables customers to combine the performance, scale, and enterprise-class features of OneFS with the economics and flexibility of the cloud. Organizations can quickly and easily create file shares in the cloud and start using them with a service that is actively operated by Dell Technologies. Key benefits include:

- **Delivered as a service for simple, fast deployment.** The infrastructure is proactively managed by Dell Technologies experts, with automated provisioning and maintenance as well as 24 x 7 proactive monitoring.
- **Simple management.** No changes are needed to applications, and organizations can easily manage file shares while having visibility into capacity and performance metrics. There is no separate portal to use; customers create, expand, and manage shares through the native Google Console.
- **Native integration with Google Cloud.** Service ordering, billing, and support are provided by Google Cloud; organizations can easily combine PowerScale service with other Google Cloud compute and analytics services.

Figure 2. PowerScale for Google Cloud



Source: Enterprise Strategy Group

- **Enterprise features** include:
 - Multi-protocol support, including NFS, SMB, and HDFS.
 - Scale-out up to 50 petabytes in a single namespace.
 - Scale-out performance and sub-millisecond latency.
 - High availability and snapshots.
 - Native file system replication between on-premises and Google Cloud.
 - Enterprise-class availability and performance SLAs.
- **Security.** Dedicated systems provide data compliance and security.
- **Cloud economics.** Organizations gain multiple cost-efficiency features, including:
 - Pay-as-you-use.
 - No charges for I/O operations.
 - On-demand, dedicated capacity.
 - Predictable pricing and guaranteed performance based on performance/capacity tiers.
 - Choice of subscriptions.
 - Ability to apply charges to Google Cloud committed spend.

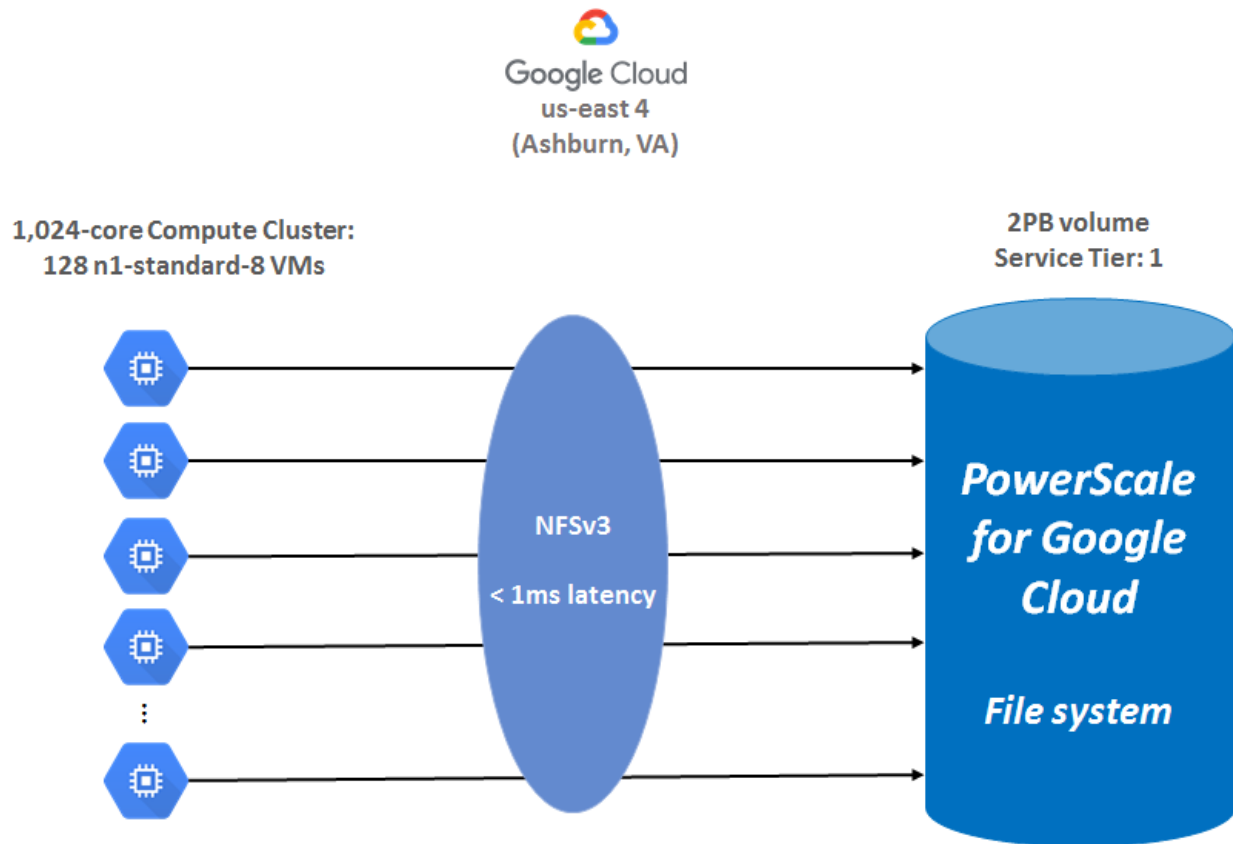
With this solution, organizations can use native Dell EMC SyncIQ file system replication to move file data to the cloud, saving them from the traditional data migration process that adds complexity and creates another management silo. They can also asynchronously mirror their on-premises configurations for a hybrid cloud deployment. Avoiding migration saves significant time and money.

While most cloud-based file services have focused on low- to mid-tier workloads, PowerScale for Google Cloud focuses on scale, offering the same high performance and capacity scalability in the cloud as it offers with on-premises Isilon. PowerScale for Google Cloud currently supports file system sizes of up to 50PB. The next section describes our performance testing.

ESG Tested

ESG measured the throughput of PowerScale for Google Cloud with the industry-standard IOzone benchmark and a publicly available genomics benchmark utility. As shown in Figure 3, the ESG test bed was configured in the us-east-4 Google Cloud region in Ashburn, Virginia. Benchmark load generating clients ran on a 1,024-node compute cluster that was deployed on 128 n1-standard-8 VMs. Clients accessed the file system using NFSv3 with less than one millisecond latency. The 2PB OneFS file system was provisioned with a Tier-1 service level designed for the highest performance applications.

Figure 3. ESG Test Bed



Source: Enterprise Strategy Group

ESG began testing with a goal of measuring maximum read and write throughput with the industry-standard IOzone benchmark.² Performance results were recorded by IOzone as the number of IOzone threads was scaled from 64 to 1,024. We tested a relatively large block size of 512KB with a goal of emulating the throughput processing characteristics of large file workflows such as high-definition video post-production, seismic data analysis, IoT, and genomics.

ESG monitored performance metrics in real time during testing with a Dell Technologies developed monitoring GUI that was built with Grafana, an open source analytics and monitoring solution.³ With the maximum load, we measured sustained read throughput of 200 GB/sec and write throughput of 120GB/sec on a single volume, with peaks encountered along the way. Figure 4 shows a screenshot that was captured during a read test with 1,024 threads. The dial and the graph below it show a peak when the PowerScale for Google Cloud volume was delivering more than 221 GB/sec of throughput.

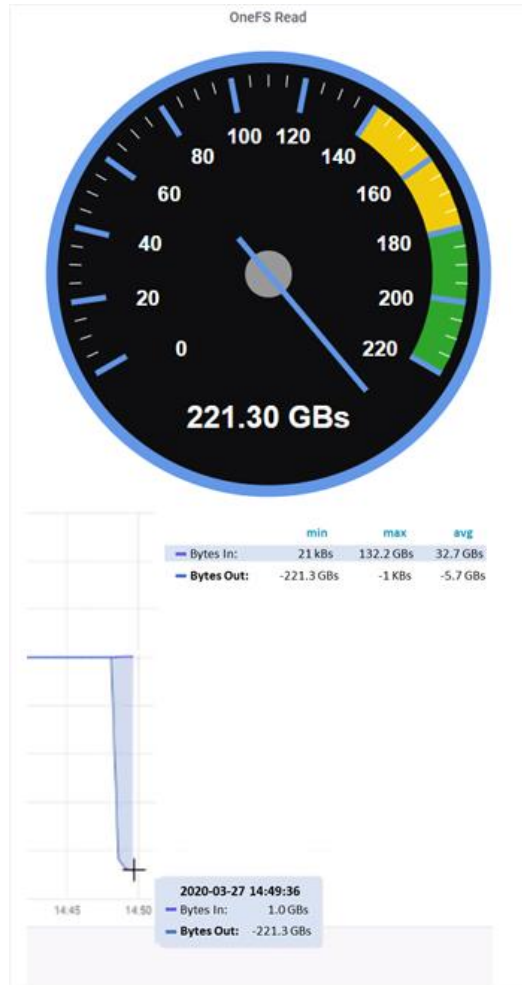
2 PB

ESG benchmark testing was performed on a single PowerScale for Google Cloud volume with a usable capacity of two petabytes.

² <http://www.iozone.org/>

³ <https://grafana.com/>

Figure 4. Grafana Real-time Performance Monitoring (Peak Read Test with 1,024 IOzone Threads)



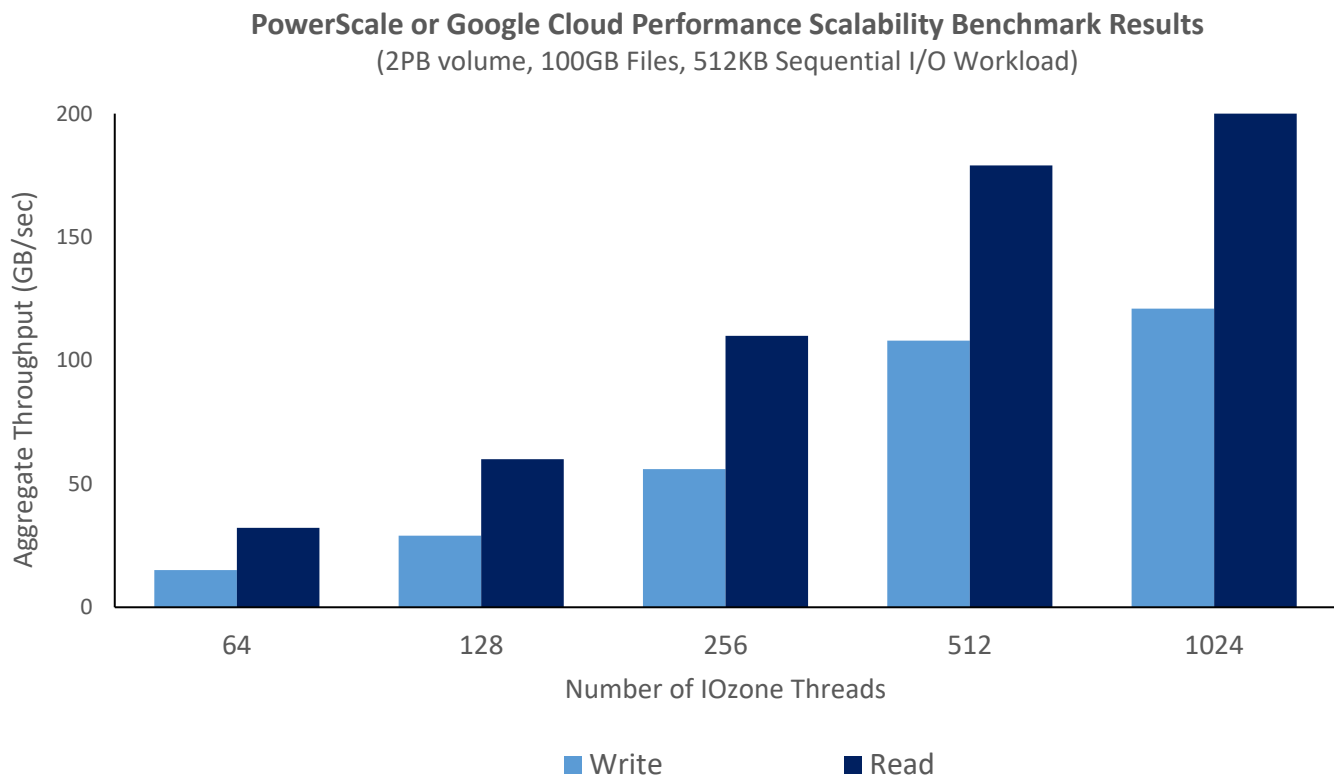
Source: Enterprise Strategy Group

I/O statistics were collected from each IOzone thread and aggregated. The results are summarized in Table 1 and Figure 5.

Table 1. PowerScale for Google Cloud IOzone Benchmark Results

Number of IOzone Threads	Write (GB/sec)	Read (GB/sec)
64	15.1	32.2
128	29	60
256	56	110
512	108	179
1,024	120	200

Source: Enterprise Strategy Group

Figure 5. PowerScale for Google Cloud Performance Scalability Benchmark Results

Source: Enterprise Strategy Group

What the Numbers Mean

- Aggregate throughput is a measure of the total amount of data that is flowing between the IOzone clients and the OneFS file system.
- Higher throughput means more data can be ingested or processed in a shorter amount of time.
- With the maximum load, sustained write throughput of 120 GB/sec and read throughput of 200 GB/sec were measured on a single volume.
- More than 100 GB/sec of aggregate throughput is an impressive level of performance for a single NFS volume in a public cloud.
- Due to the scale-out architecture of PowerScale for Google Cloud and the fact that it supports more than 2PB in usable capacity for a single NFS volume, ESG is confident that higher levels of performance can be achieved with a larger volume. On-premises, Isilon delivers linear performance scalability with file system scale-out, and we expect similar behavior in the Google Cloud.

200 GB/sec
IOzone throughput testing peaked at 200 GB/sec for reads and 120 GB/sec for writes.

ESG also reviewed publicly available specifications and performance benchmarking results for a competing NAS solution using Google Cloud from another vendor (Vendor X). It should be noted that these tests used a smaller workload and smaller file system; in addition, other test bed details were not published, so we cannot make a direct “apples to apples” comparison. However, we assume other vendors publish the best results they can. Comparing these specifications and results, we noted:

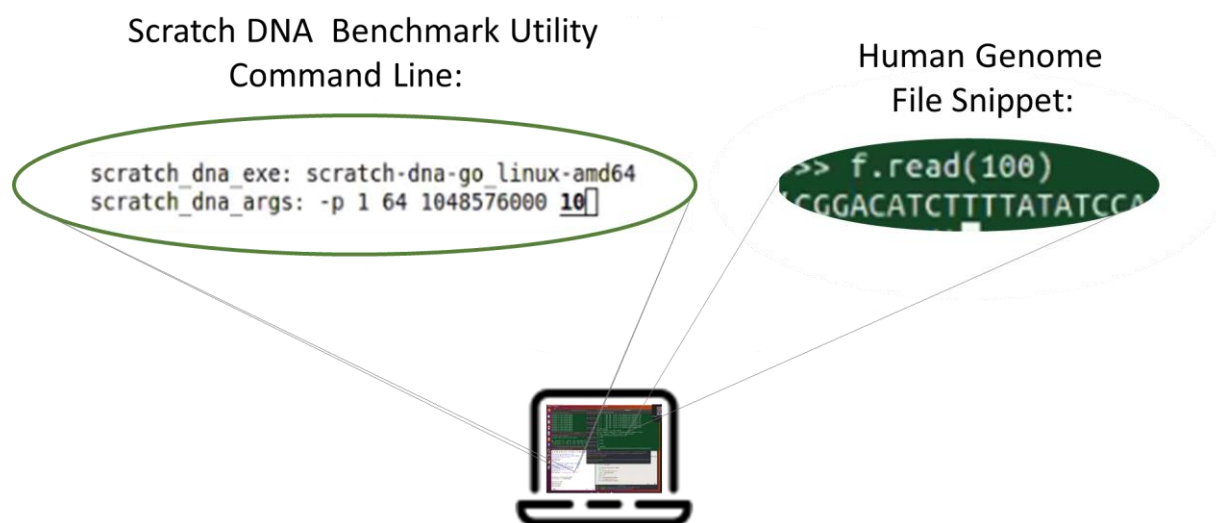
- PowerScale for Google Cloud offers up to **500x higher maximum file system capacity** compared with Vendor X.
- PowerScale for Google Cloud can deliver up to **46x higher maximum read throughput** compared with Vendor X.
- PowerScale for Google Cloud can deliver up to **96x higher maximum write throughput** compared with Vendor X.

Genomic Sequencing

Next, ESG measured the write performance of PowerScale for Google Cloud with the publicly available Fred Hutch Scratch DNA benchmark utility.⁴ Scratch DNA is a storage benchmarking tool that creates a random DNA string, concatenates that string a random number of times, and writes the results to multiple files with varying lengths. The goal of this benchmark is to emulate the write-intensive storage workloads that are typically generated by scientists as they ingest, rescan, and write genomics data. The load peaked when running 2,048 threads of the Scratch DNA in parallel, creating approximately 1 TB of genomic data with file sizes ranging in size between 1 GB and 10 GB. Figure 6 shows the Scratch DNA command line utility parameters and a snippet of the simulated human genome data that was created during testing.

A peak aggregate write throughput of 110 GB/sec was calculated by summing up the file system performance statistics that were reported by each Scratch DNA thread.⁵

Figure 6. Genomic Workload Throughput Testing



Source: Enterprise Strategy Group

i Why This Matters

For data-intensive workloads such as IoT and big data analytics, most cloud solutions cannot adequately scale the file system size or achieve sufficient performance. In addition, the storage costs to handle these workloads are often prohibitive.

Using the industry-standard IOzone benchmark utility, ESG validated that Dell EMC PowerScale for Google Cloud achieved a maximum read performance of 200 GB/sec and maximum write performance of 120 GB/sec against a 2PB storage volume. Because this is a scale-out service, expanding the file system would result in even higher performance; currently PowerScale for Google Cloud supports file system sizes of up to 50PB. ESG also observed a maximum write performance of 110 GB/sec with a simulated genome sequencing benchmark.

⁴ <https://github.com/FredHutch/sc-benchmark>

⁵ The 9% difference between the 110 GB/sec that was achieved with the Scratch DNA benchmark utility and the 120 GB/sec that was achieved with IOzone write is due to the real-world application overhead associated with generating simulated human genomic data patterns for 1 TB of file data.

The Bigger Truth

While many block and object storage workloads have found a home in the cloud, for the most part file workloads have been missing due to performance, scalability, and cost limitations. However, with the right capabilities, the cloud should work well for large file workloads that depend on a lot of data, including artificial intelligence (AI)/machine learning (ML)/deep learning (DL) and large commercial HPC workloads such as genomics, oil and gas simulations, media and entertainment rendering and animation, virtual reality, and gaming. Some of these workloads need only occasional bursts to scalable infrastructure—a common cloud use case—but only if performance objectives can be met. All of them could benefit from the economics of the cloud.

The PowerScale for Google Cloud service brings an industry-leading file system to the cloud with the levels of scalability and performance these applications demand. Organizations can take advantage of the cost efficiency and flexibility aspects of the cloud for these workloads, saving money while making data-driven decisions with predictable high performance. This solution brings together innovations from two industry leaders:

- **From Dell Technologies:** Dedicated OneFS file systems with the high performance, high scalability, and enterprise features that it offers. Workloads can be moved with native file system replication, eliminating difficult and time-consuming data migration to a different cloud platform than used on-premises.
- **From Google Cloud:** A native service that leverages Google Cloud's high-performance compute and GPU instances and high-bandwidth network. Instead of a capital expense in the data center, organizations get a service that can be used when needed and are billed only for what they use.

ESG validated throughput performance of PowerScale for Google Cloud using both IOzone, an industry-standard benchmark, and Scratch DNA, a benchmarking tool that emulates the write-intensive storage workloads typically generated in genomics research. Highlights include:

- The ability to test against a 2PB OneFS volume demonstrates the extreme scalability of the file system in the cloud. Performance scales with capacity, and since PowerScale for Google Cloud currently supports file systems of up to 50PB, it offers significant performance scale potential.
- IOzone aggregate throughput results showed high performance scalability as the workload increased, with a peak of 200GB/sec aggregate read throughput and 120GB/sec aggregate write throughput against a 2PB storage volume.
- A simulated genomic sequencing benchmark, Scratch DNA, demonstrated peak aggregate write throughput of 110 GB/sec.
- These results compared favorably against a competitive NAS vendor using Google Cloud, delivering 500x maximum capacity, 46x read throughput, and 96x write throughput.

These are impressive results for a single NFS file system in the cloud, and ESG is confident that OneFS scaled to a larger capacity would deliver even higher performance. These benchmarks demonstrate the ability of this solution to deliver high performance and scalability for large file workloads such as analytics, video, seismic data analysis, IoT, genomics, and more.

It should be noted that this solution also offers good price/performance with a simple \$/GiB/month pricing model based on performance tiers. Discounts are applied based on capacity and one-year or three-year commitments. There are no other variables or pricing add-ons, making it different from other public cloud solutions that include surcharges for throughput, I/O operations, etc. ESG reviewed publicly available pricing and noted that this solution offers 2x-10x lower cost/GiB/month than similar file solutions in the cloud.

The service is currently available in the Northern Virginia, Iowa, Singapore, and Sydney Google Cloud Platform regions, with additional regions expected to roll out in the future.

As always, organizations should look at their data and find the right methodology that provides the levels of performance, flexibility, and cost that meet each workload's needs. ESG expects that the Dell EMC PowerScale for Google Cloud will add an exciting option for large file workloads that need high performance, scalability, and cost efficiency.



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. 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.

The goal of ESG Validation reports is to educate IT professionals about information technology solutions for companies of all types and sizes. ESG Validation reports are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objectives are to explore some of the more valuable features and functions of IT solutions, show how they can be used to solve real customer problems, and identify any areas needing improvement. The ESG Validation Team's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments.

