

Technical Review

Hitachi Content Software for File

Date: April 2022 **Author:** Tony Palmer, Principal Validation Analyst

Abstract

This report documents ESG’s review of Hitachi Content Software for File performance testing that demonstrates predictable, scalable performance for HPC (high-performance computing), AI (artificial intelligence), ML (machine learning), and analytics. Results of testing with rigorous, industry-standard benchmarks were audited to evaluate Hitachi Content Software for File capabilities using complex, performance-critical workloads.

The Challenges

Organizations need more data, faster, to deliver insights that drive business decisions. ESG research reveals that 59% of organizations identify data as their business or part of their business; 23% have core products and services that are information-based, and 36% offer both tangible and information-based products and services.¹ It should come as no surprise that, according to a separate ESG research survey, organizations reported that they expected to increase spending on data-centric projects in 2021 (see Figure 1).² This is helping to drive ever larger data volumes in production. In fact, nearly half (48%) of organizations indicated that at least 1 PB of their storage capacity is associated with storing their active data and 60% of organizations anticipate their on-premises data growth to be at least 21%, with 27% reporting that they expect greater than 50% year-over-year growth over the next three years.³

Figure 1. Object Storage and On-premises Applications Workload Spending Intentions over Next 24 Months



Source: Enterprise Strategy Group

As analytics usage increases, organizations want faster performance to speed data-focused insights that can inform business decisions. A high-performance storage solution can enable additional legacy and modern/cloud-native applications to gain the benefits of its scalability, fast data retrieval, and cost effectiveness.

¹ Source: ESG Research Report, [Data Infrastructure Trends](#), November 2021.

² Source: ESG Research Report, [2021 Technology Spending Intentions Survey](#), January 2021.

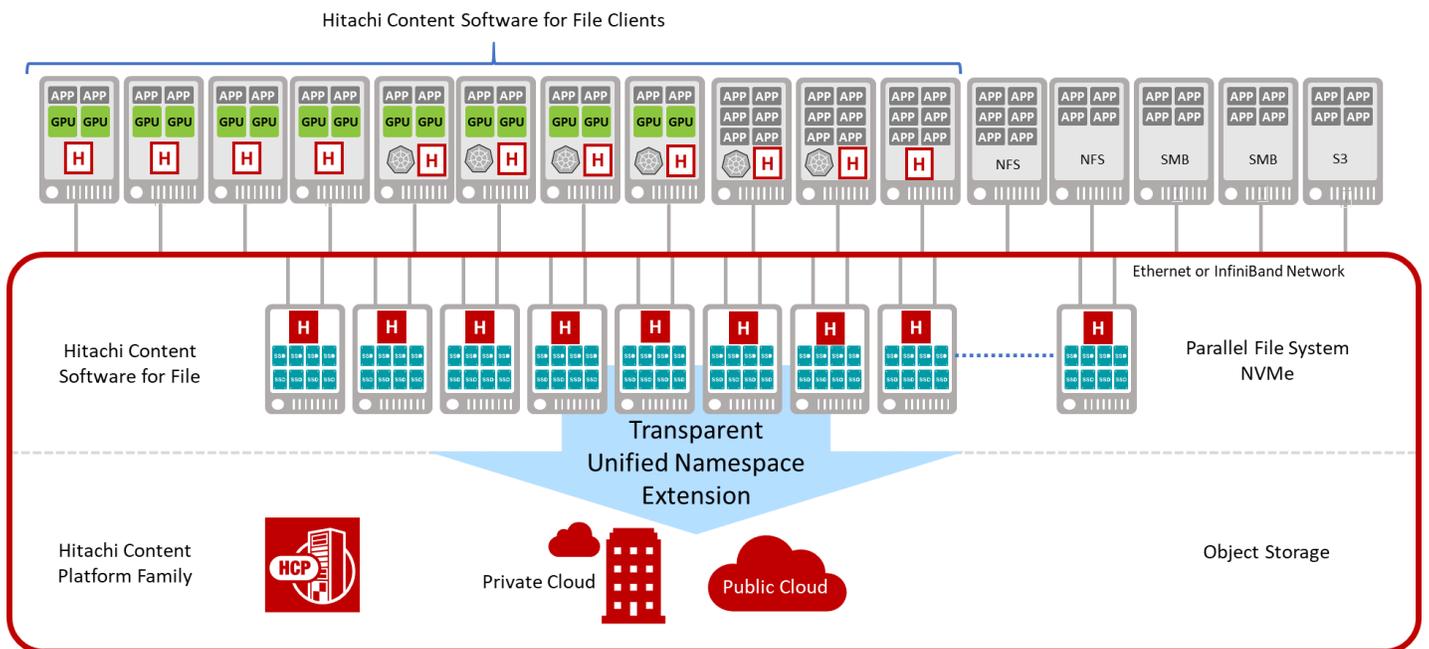
³ Source: ESG Survey Results, [2021 Data Infrastructure Trends](#), September 2021.

The Solution: Hitach Content Software for File

Hitach Content Software for File is a POSIX-compliant, scalable, distributed file system that integrates with Hitachi Vantara’s object storage portfolio. Content Software for File can be deployed on-premises, as a cloud platform, or as a hybrid cloud solution, utilizing either a single cloud vendor or a multi-cloud environment to store and manage the unstructured data associated with high-performance modern applications and workflows (e.g., AI, ML, and analytics). Supporting capacity requirements that can exceed 10+ Exabytes, a single cluster can scale the meet the demand of modern organizations’ data requirements.

Content Software for File is especially well-suited to the most performance-sensitive and resource-intensive applications across markets and industries—genomics and life sciences, finance, media and entertainment, and manufacturing/ engineering. Content Software for File delivers accelerated flash storage performance using modern NVMe SSDs powered by the latest in PCI Gen4 computer platforms, powering both the file-system and the applications consuming the storage technology. Organizations can scale the capacity of the platform by adding additional NVMe SSDs to a server, by adding more servers with NVMe devices, or by allocating S3 object storage capacity as a secondary tier. Performance can scale by allocating additional CPU resources or adding server instances to the distributed flash portion of the file-system. Content Software for File storage can be consumed by applications using conventional NAS mechanisms—NFS and/or SMB, or via Content Software for File’s accelerated POSIX client interface. Content Software for File also supports the NVIDIA GPUDirect Storage protocol to maximize the value of organizations’ GPU investment and S3 APIs for modern cloud/object storage aligned applications.

Figure 2. Content Software for File in a Production Environment



Source: Enterprise Strategy Group

Content Software for File distributed data protection is designed to run both on-premises and in the cloud and to deliver high resilience with minimal impact. Content Software for File distributes data across failure domains. Journaling is integral to Content Software for File, so recovery from a server failure or from a complete system failure is fast, never requiring a full file system check or any other traditional recovery mechanisms. As the volume of data increases, organizations can scale the file system beyond the boundaries of the flash layer, utilizing object storage for density and protection. As a customer consumes more storage, native file system tiering to an object storage platform like the Hitachi Content Platform can be utilized to improve the density and economics of the total solution. To improve data resilience, the file system can also

utilize the same tiering object storage or a different object storage target for data protection. This could allow for scenarios where the file system utilizes a local object storage as a tiering target and a remote on-premises or cloud-based object storage snapshot target, providing off-site data protection.

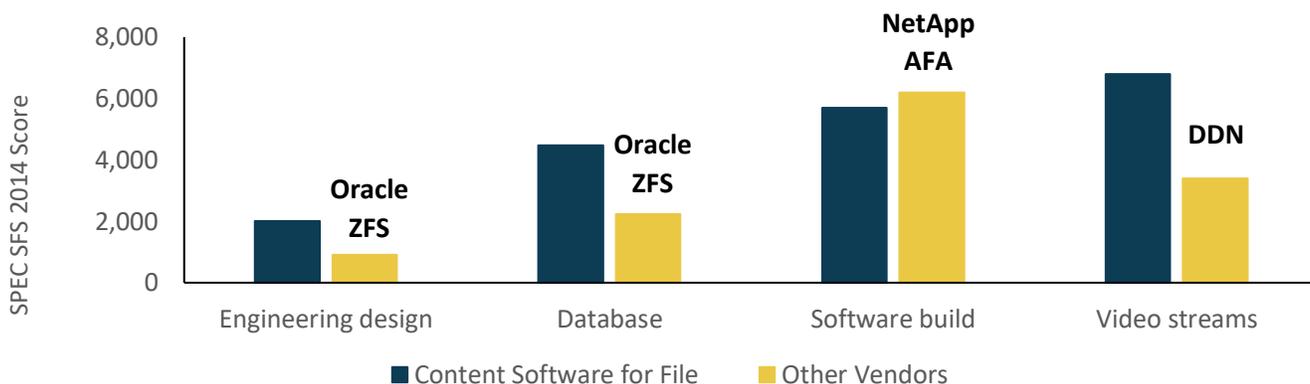
ESG Validated

ESG validated performance of the Hitachi Content Software for File storage solution leveraging platforms deployed on-premises and in AWS. ESG examined Content Software for File’s performance as compared to the market using three rigorous industry-standard benchmarks. In addition, time to deploy and extend a file system to AWS, deployment and management simplicity, and data durability and recoverability were also evaluated.

Performance

ESG began by auditing performance results from three industry-standard benchmarks to gauge complex workload performance capabilities for a variety of performance-critical workloads. First, we examined published SPEC SFS 2014 results. Content Software for File earned the top spot for multiple workloads. The results are summarized in Figure 3.

Figure 3. Content Software for File SPEC SFS 2014 Performance Compared to Other Solutions



Source: Enterprise Strategy Group

As seen in Table 1, performance was double the next closest result for three out of the four tests, with lower overall response time across the board.

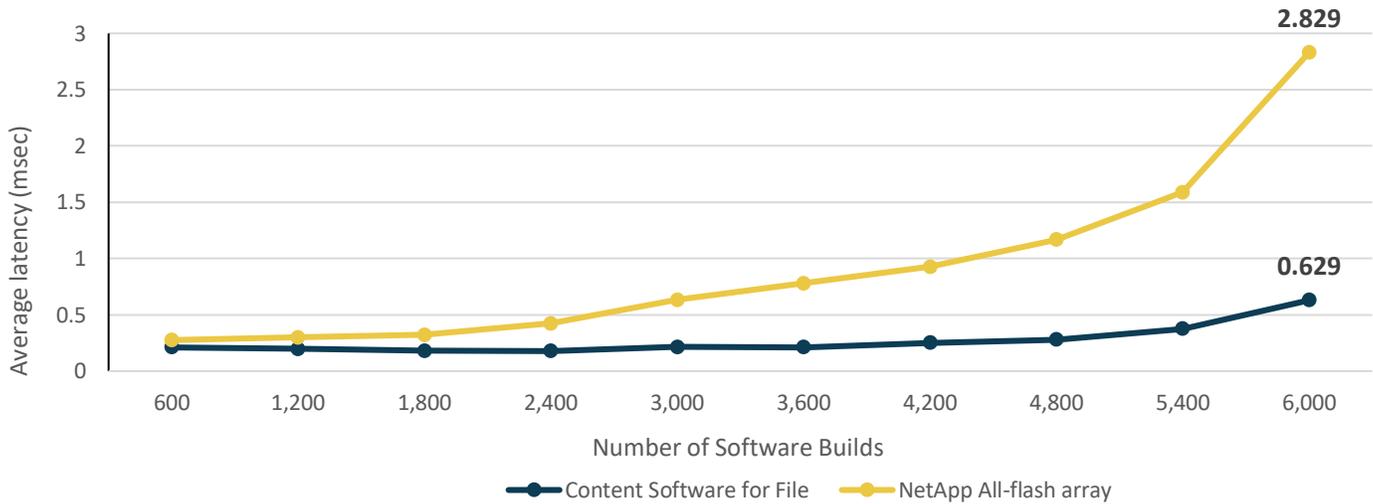
Table 1. Hitachi Content Software for File SPEC SFS 2014 Performance Across Multiple Workloads

Benchmark	Top Scoring Solution	SPEC SFS 2014 Score	Overall Response Time (ms)	Number Two Solution	SPEC SFS 2014 Score	Overall Response Time (ms)
Engineering Design	Content Software for File	2,000	0.26	Oracle	900	0.61
Database	Content Software for File	4,480	0.34	Oracle	2,240	0.78
Software Build	NetApp	6,200	0.83	Content Software for File	5,700	0.48
Video Streams	Content Software for File	6,800	1.56	DDN	3,400	50.07

Source: Enterprise Strategy Group

Of particular interest is the software build benchmark. The software build benchmark is a file system benchmark that mimics the behavior of software builds and traces. 70% of the operations performed by the benchmark are stat calls returning file attributes (metadata). As such, the software build workload has a high amount of small block reads with interspersed larger block file I/O. While the overall number of concurrent builds was a bit higher than Content Software for File, response times tell a more complete story.

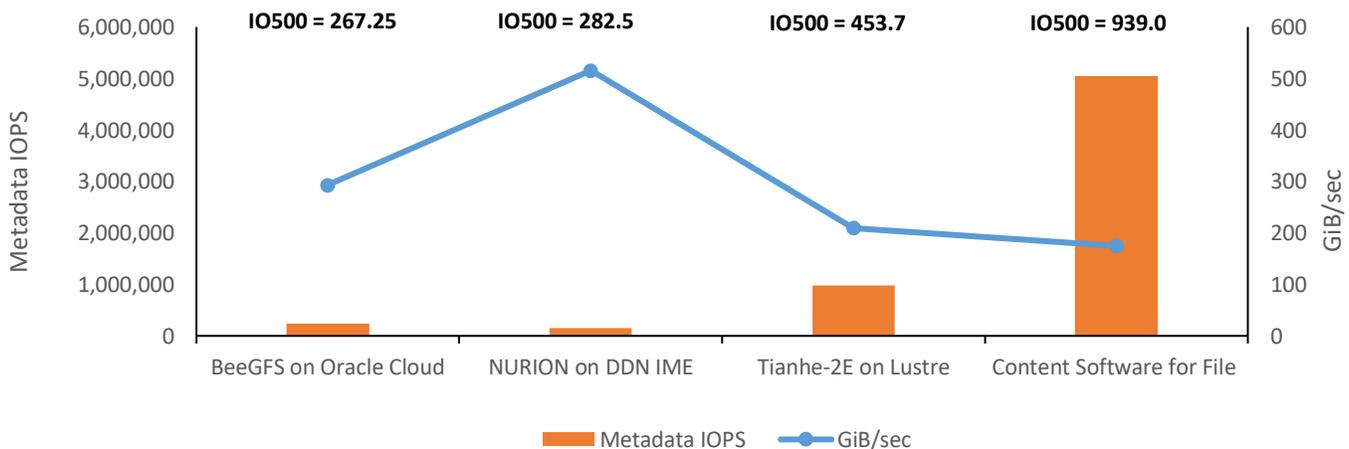
Figure 4. SPEC SFS 2014 Software Build Response Time Curve



Source: Enterprise Strategy Group

As seen in Figure 4, response time stayed low as the number of concurrent software builds was increased from 570 to 5,700 with an overall response time for the entire test run of 480 μs or .48 ms. The All-flash array tested by NetApp scaled to 6,200 concurrent builds, but with 4.5x greater response time than Content Software for File. Content Software for File posted the highest performance in three of the four categories we examined. Next, ESG analyzed the results of the IO-500 benchmark and compared Content Software for File performance on AWS to other file system solutions. The IO-500 benchmark suite was designed to enable comparison of high-performance storage systems with full disclosure of how the tests were performed and a methodology and ranking system designed to maximize credibility and usefulness. IO-500 scoring is derived from the square root of the product of metadata IOPS and throughput.

Figure 5. Content Software for File IO-500 Performance on AWS Compared to Other File System Solutions



Source: Enterprise Strategy Group

As seen in Figure 5, Content Software for File’s IO-500 results on AWS are significantly higher than other file system solutions whether on-premises or running in the cloud. It’s important to note the throughput and metadata IOPS are weighted equally in this benchmark, so a solution will be scored fairly. It’s particularly impressive that Content Software for File supported 5x the metadata IOPS of the next highest file system solution while sustaining 175 Gibibytes (GiB) per second of throughput. Note: IO 500 reports throughput in Gibibytes/sec. Gibibytes (GiB) differ from Gigabytes (GB) in that they are calculated in multiples of 1,024, rather than 1,000. 175 *Gibibytes/sec* is equal to 187.9 *Gigabytes/sec*. It’s important to look at the composition of the systems under test. Note that while the NURION on DDN IME bandwidth was 515 GiB/sec, that result was achieved with 2048 clients—about .25 GiB/sec per client. Content Software for File’s 175 GiB/sec was achieved with just 345 clients—about .5 GiB per client.

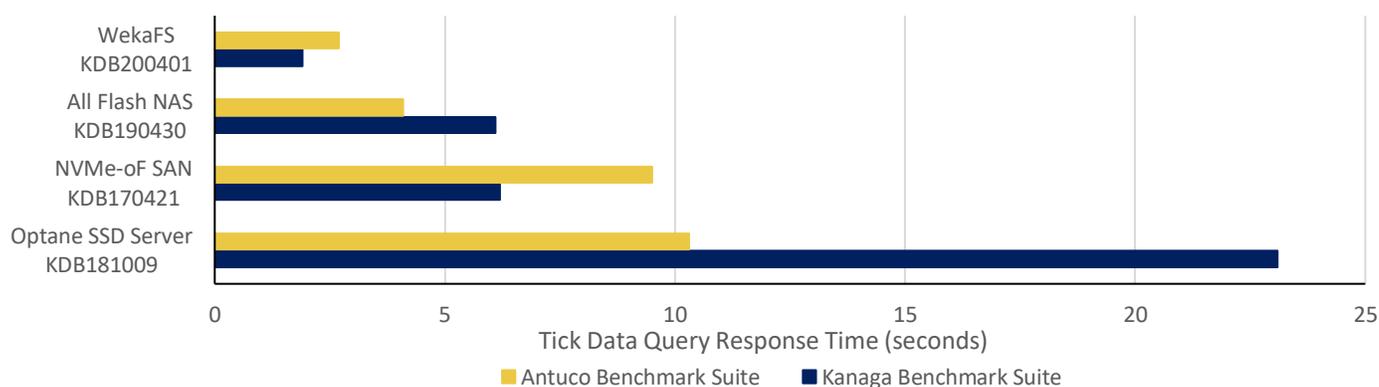
Another interesting comparison is with the Tianhe 2-E supercomputer running Lustre. Lustre is a parallel filesystem where metadata nodes and storage nodes are separate. Clients talk to metadata nodes then data is fetched directly to the client. Hitachi Content Software for File does not split metadata and handles it as part of the storage layer. While Tianhe-2E had a higher aggregated throughput with 40% more client nodes but performed 56% slower on a per-client process basis—important for modern workloads and 14% slower per client throughput—important for traditional HPC workloads.

STAC-M3 is a benchmark published by the Securities Technology Analysis Center (STAC), with the goal of helping end-user firms relate the performance of new technologies to that of their existing systems. The STAC Benchmark Council is an organization of leading financial institutions and technology vendors that specifies standard ways to assess technologies used in finance.

STAC-M3 is a set of industry-standard enterprise tick-analytics benchmarks for database stacks that manage large sets of time series market data—aka “tick data.” We examined results from the baseline benchmark suite (Antuco) and the optional scaling suite (Kanaga). STAC-M3 focuses on the complete system under test—compute, storage, and networking elements tested together. The Antuco benchmark suite simulates gathering data from 12 specific (not consecutive) days from one year’s worth of tick data. Clients are scaled from one to 50 to 100, performing all searches concurrently. The Kanaga benchmark suite includes multi-year high bid analytics that reads terabytes of data to answer a query, placing significant load on storage I/O, while other tests such as Theoretical P&L in the Antuco suite are computationally intense with less impact from the storage system.

This multidimensional model incorporating both the Antuco and Kanaga benchmark suites highlights the limits of solutions while articulating where another solution starts to make sense for the anticipated workloads. It also helps users understand the resources—compute, storage I/O, and networking—required to reach the performance demands for an organization.

Figure 6. Content Software for File STAC-M3 Tick Data Query Response Time (Shorter Is Better)



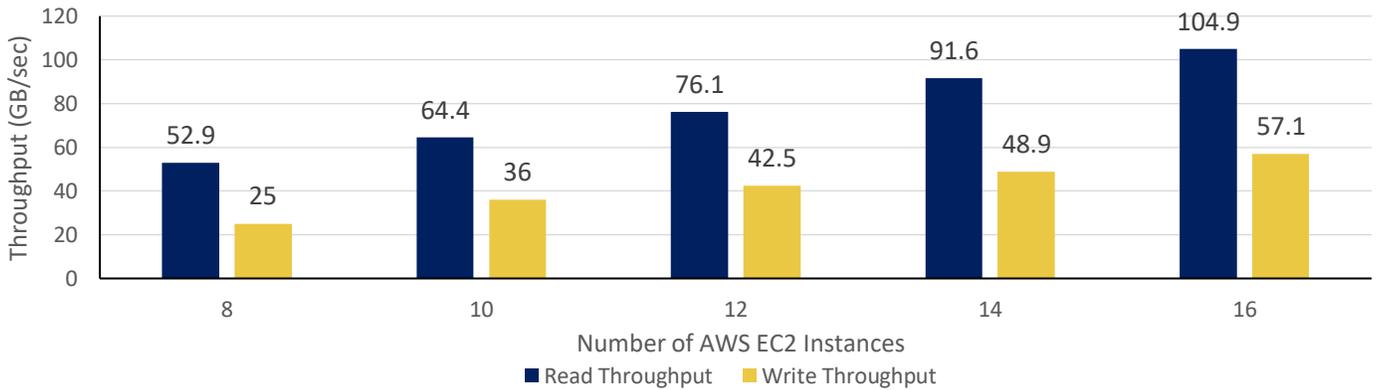
Source: Enterprise Strategy Group

The systems compared include a server with internal Optane SSDs, an NVME over Fabric SAN, and a high-end all-flash NAS array. As seen in Figure 6, Content Software for File posted the shortest query response time in both benchmarks by a significant margin.

FIO Testing in AWS

Finally, ESG tested performance and scaling of Hitachi’s platform in AWS. We deployed an eight-node Content Software for File cluster on AWS i3en.24xlarge all-flash instances. We used the fio load generation tool running on 16 AWS c5n.18xlarge instances to test IOPS, throughput, and latency. Each test was run for five minutes and repeated three times and results were averaged. After we completed each series of tests, we added two nodes to the cluster and tested again. The results are summarized in the next three charts. The first series of tests were used to stress the throughput capabilities of the cluster. Fio was configured to drive 1MB I/O with each of the 16 clients running 32 threads. As shown in **Figure 7**, throughput scales linearly as nodes are added to the cluster.

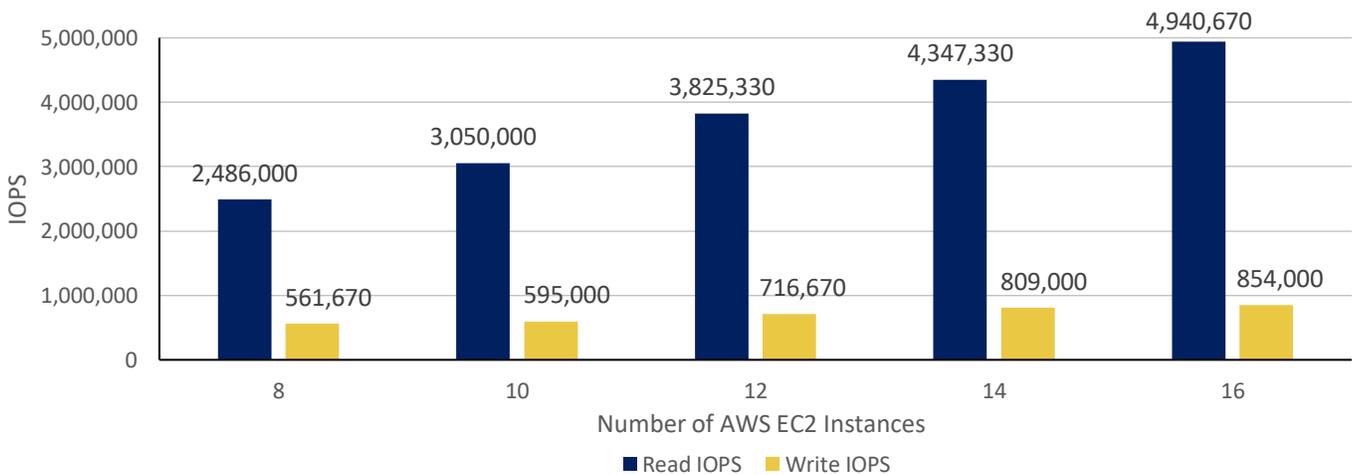
Figure 7. Content Software for File Linear Performance Scaling on AWS—Throughput



Source: Enterprise Strategy Group

The second set of tests focused on IOPS. Fio was configured to drive 4KB I/O with each of the 16 clients running 192 threads. As seen in **Figure 8**, read IOPS scaled with near-perfect linearity, while write IOPS increased consistently as nodes were added, adding an average of 73,000 IOPS as each pair of nodes were added. During these tests, latency averaged 748 microseconds across all IOPS tests for all cluster sizes.

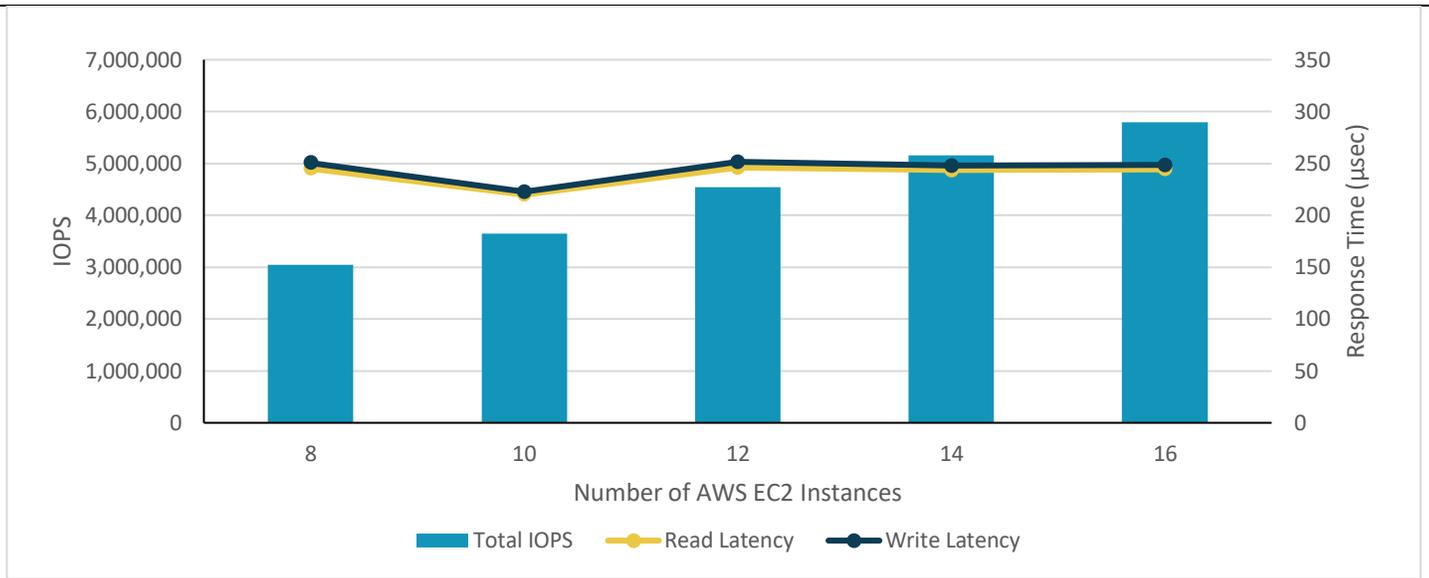
Figure 8. Content Software for File Performance Scaling on AWS—IOPS



Source: Enterprise Strategy Group

Finally, we looked at client latency in the IOPS test. As **Figure 9** shows, latency stayed steady as IOPS increased, never exceeding 250 µsec, even when the system was servicing nearly 6 million IOPS.

Figure 9. Content Software for File Performance on AWS—Latency



Source: Enterprise Strategy Group

It’s important to note that the latency Content Software for File can achieve in on-premises environments running on commodity hardware is considerably lower. ESG observed average response times of 450 µsec under similar IOPS load as seen in **Figure 8** and less than 100 µsec in single-threaded tests.



Why This Matters

Organizations running modern and traditional HPC workloads need performance. It’s not surprising that 74% of organizations are already using or plan to use NVMe-based solid-state storage technology, and more than half of the organizations that are using it say it has improved the performance of their storage infrastructure to support new, more demanding workloads (59%) or improved the performance of existing (62%) applications.⁴

ESG verified Content Software for File’s ability to outperform all-flash storage arrays and to use its distributed, parallel file system to expand performance even further. Content Software for File on AWS delivered nearly 5.8 million 4KB IOPS with sub-millisecond response times and more than 160 GB/sec of throughput. An on-premises cluster delivered the same performance with response times averaging under 500 µsec. Content Software for File has posted top results in multiple industry-standard benchmarks including SPEC SFS 2014, emulating multiple application workloads, and STAC-M3 running intensive financial analysis. In the IO-500 benchmark, Content Software for File posted the top result for solutions running on a file system.

⁴ Source: ESG Research Report, [2021 Data Infrastructure Trends](#), November 2021.

The Bigger Truth

Organizations are continuing to generate and store exceptionally large amounts of unstructured data. ESG uncovered that fully 60% of organizations expect their on-premises data to grow by at least 21% annually over the next three years.⁵ With the increasing adoption and use of data-intensive applications—artificial intelligence, machine learning, and data analytics, to name just a few—organizations require a solution that can efficiently store and process data with consistently high performance. The solution should also scale in a manner that enables organizations to increase processing power and capacity independently, on-premises or in the cloud.

Content Software for File is a software-defined storage solution that provides a distributed file system on-premises or in the cloud to cost-effectively provide extremely high-performance file storage using hyperscale techniques. It is also especially well-suited to the most performance-sensitive and resource-intensive applications. Content Software for File delivers flash storage performance using a tiered methodology that enables organizations to extend their unified namespace to private and public clouds.

Hitachi Content Platform offers object storage for today's workloads, including the latest analytics applications. Content Software for File delivers the high performance and scalability that enterprises demand of their business-driving workloads.

ESG validated the performance of Content Software for File, analyzing results that placed the file system head and shoulders above traditional all-flash storage arrays; Content Software for File achieved millions of IOPS and hundreds of GBs of throughput for common HPC workloads, while Content Software for File delivered linear scalability in an AWS cloud deployment. Providing such high throughput and IOPS with extremely low latency is a differentiator and enables organizations to use HCFS in place of servers with local drives while improving the overall time to completion of customers' workloads, saving time and money.

Hitachi Vantara has long been a trusted provider of solutions for enterprise customers, including large, complex environments with distributed employees. The company's solutions are known for reliability, security, availability, and enterprise-class features. This Content Software for File performance validation adds to the company's strong resume.

Of course, your mileage may vary, as these tests were run in controlled environments, and every organization should plan and test in its own data center to ensure the efficacy of the solution. But if you are looking for a storage solution that delivers scalable high performance with cost efficiency, ESG recommends that you take a good look at Hitachi Content Software for File.

⁵ Source: ESG Survey Results, [2021 Data Infrastructure Trends](#), September 2021.

Appendix

In July 2020, WekaIO and Hitachi Vantara entered into a partnership whereby Hitachi Vantara OEMs and re-brands WekaIO WekaFS platform as the Hitachi Content Software for File. As part of this ongoing relationship, both companies strive to achieve and maintain compatibility with third party hardware and software products spanning several areas of the storage eco-system.

For the purposes of certification and qualification with third party products, WekaIO and Hitachi Vantara are providing a statement of product equivalency for their respective products.

In all respects, except product branding, the WekaIO WekaFS and the Hitachi Content Software for File product are functionally identical.

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.