Technical Review

Hitachi Content Platform: High-performance Object Storage for Cloud-native Applications

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

Abstract

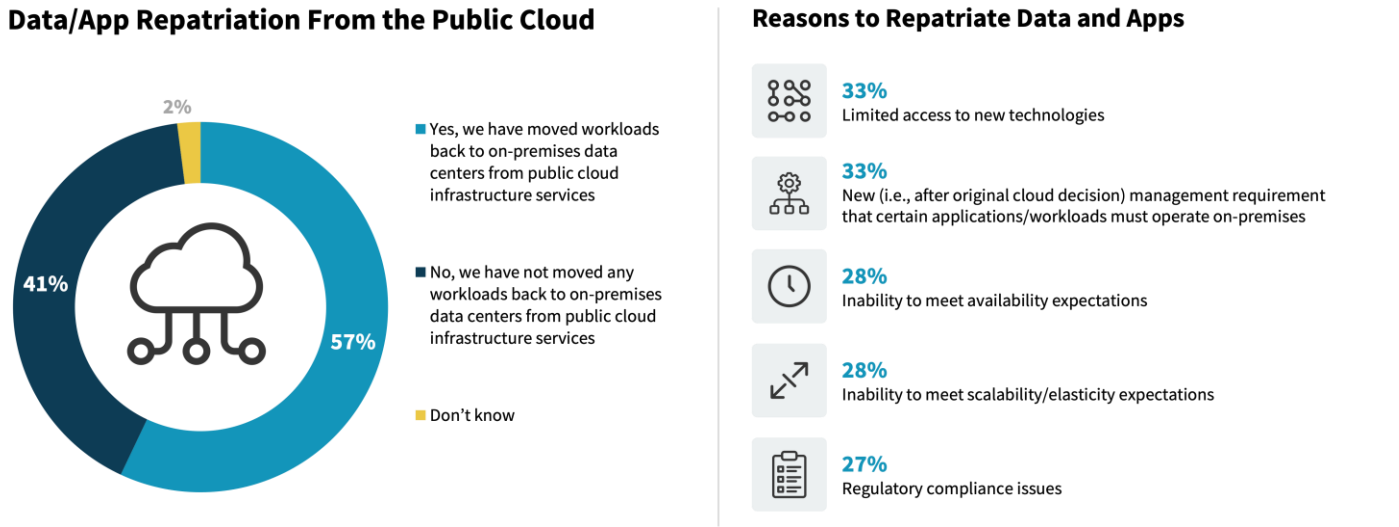
This report by TechTarget's Enterprise Strategy Group (ESG) documents the validation of Hitachi Content Platform (HCP) performance testing and confirms its consistent performance at scale for object storage. The testing was designed to measure performance with small and large objects, including throughput and time to first byte, to demonstrate the cost-effective performance that customers can achieve in their data centers for modern applications and analytics workloads.

The Challenges

According to 2022 DataOps research by Enterprise Strategy Group (ESG), organizations are prioritizing agility, transparency, and speed across their hybrid-cloud data ecosystems in search of rapid insight to support data-driven initiatives. In this context, agility means running workloads wherever they can drive the most value the fastest. 403 organizations across multiple industries were asked what business objectives are driving their data strategies. The top three responses were improving the quality of their products and services (53%), improving operational efficiency (53%), and improving business decisions and strategy (52%).[1]

In another survey, 57% of organizations indicated that they had repatriated their workloads (i.e., moved workloads back to on-premises data centers leveraging multi-, hybrid, or private clouds) for diverse reasons, including the need to access new technologies (33%), changing data management requirements for apps to run on-premises (33%), the need for higher levels of resilience (28%), the inability to meet scalability/elasticity expectations (28%), and compliance issues (27%) (see Figure 1).[2]

Figure 1. Workload Repatriation Rate and Drivers



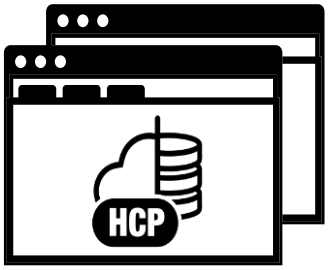
Source: Enterprise Strategy Group, a division of TechTarget, Inc.

[1] Source: Enterprise Strategy Group Research Report, *The State of DataOps*, August 2022.
[2] Source: Enterprise Strategy Group Survey Results, *2021 Data Infrastructure Trends*, September 2021.

Organizations face technology, operations, security, and financial challenges in order to be ready for new and evolving workloads and applications.

Hitachi Content Platform

HCP is object storage software that is designed to provide an on-premises cloud storage experience for storing, protecting, and scaling unstructured data. Its hallmarks are ease of use, efficiency, and massive scalability. HCP has the horsepower to support application modernization with Hadoop data lakes and next-generation analytics associated with modern S3-based applications and workflows. HCP also integrates with Hitachi Content Software for File (HCSF), a high-performance parallel file system that supports modern GPU workloads, as well as high-performance data analytics (HPDA) and other workloads that are highly concurrent and dissimilar. Previously, Enterprise Strategy Group (ESG) validated Hitachi Content Software for File performance for high-performance computing (HPC), artificial intelligence (AI), machine learning (ML), and real-time analytics.[3]



✓ Cyber Resilience

✓ Performance at Scale

✓ Intelligent Data Services

✓ Highly Multi-tenant

HCP supports private (on-premises), hybrid, and multi-cloud delivery models, third-party ISV applications across horizontal and vertical industry-specific use cases, public cloud APIs, and an impressive variety of deployment options. HCP deployment options are extremely flexible, with multiple on-premises and cloud-hosted instance options. Data can be stored on-premises—protected via erasure coding—or in the cloud. HCP uses customizable metadata management capabilities and numerous operational tools, including automation, policy-driven management, monitoring, reporting, and compliance and security capabilities to store and manage unstructured data at scale. The HCP portfolio has received FIPS-140-3 level 1 certifications, including Open SSL Provider. A single cluster can scale to more than an exabyte of data under management, supporting hundreds of billions of objects. All-flash configurations are supported, along with policy-driven tiered storage. Hitachi Vantara offers a full range of consumption models, including financial, OpEx, -as-a-service, and managed services.

Key features of the HCP portfolio include:

- Consistent Price-performance—HCP's flexible architecture is highly performant, with fine granularity for a variety of workload performance requirements and predictable cost—comparable to public cloud storage pricing—as performance scales from petabytes to exabytes. Organizations can scale up or out using hardware, software, protocols, and microservices. All-flash object storage offers consistent performance, scaling from thousands to billions of files. Consistent performance is critical for analytical workloads or IoT data where there is a huge volume of data to be processed and success is measured by how quickly the ingest/analysis can be completed.

- Advanced S3 Capabilities—HCP is compatible with the modern S3 APIs made popular by Amazon. HCP supports everything from basic create, read, update, and delete (CRUD) capabilities to advanced analytics functions provided by S3Select. HCP allows for modern applications to utilize the S3 APIs without modification while enabling applications to take advantage of the richer capabilities being defined by S3.

- Compliance—HCP provides functionality that supports organizations' efforts to meet stringent regulatory compliance requirements, including enforceable data retention policies and defensible disposal with the ability to apply multiple legal holds to each object, which is especially important in highly regulated industries.

[3] Source: Enterprise Strategy Group Technical Review, *Hitachi Content Software for File*, April 2022.

- **Cyber Resilience**—HCP allows applications to use the AWS S3 API with strong fidelity and a broad set of object version retention, encryption, and worm capabilities, strengthening its object storage against cyber and ransomware attacks. HCP has tight integrations with Veritas, Commvault, and Veeam through S3 Object Lock capability. HCP's product compatibility guide page provides a complete list of certified independent software vendors (ISVs) with version and interface compatibility for each supported application.

- **Intelligent Data Services**—Organizations can use Hitachi Content Intelligence in conjunction with the intelligent, policy-based tiering inherent in HCP to perform real-time data classification as data is ingested to identify sensitive information inside of unstructured data based on user-defined criteria and execute actions based on user-designed workflows for business, governance, or other objectives. For more complex data pipelines, HCP integrates with message brokers like Apache Kafka, RabbitMQ, and AWS SNS. HCP addresses the challenges of higher-than-expected costs due to a lack of funding by simultaneously leveraging multiple storage targets and cloud services to move data to the most appropriate location automatically based on value, policies, and workflow. HCP is designed to overcome the difficulty of meeting internal service level agreements caused by increasing IT requests and new workloads with deep knowledge of what is being stored, enabling the object store to perform complex functions.

Accelerating Productivity with Data Services

- **Optimize queries, workflow, and search to find the right subset of data.**
- **Eliminate unneeded data from large datasets to optimize performance.**
- **Connect and collect data from as many sources as needed, informing better decisions.**
- **Streamline IT ops, backup, legal discovery, and classification while improving accuracy with data enrichment.**

- **Multi-cloud Workflows for Multiple Tenants**—HCP enables organizations to consume services from different cloud providers natively as though they are operating in the cloud from local, on-premises infrastructure. In addition, HCP can send data to any S3-compatible endpoint (Sync to) for applications like transcription or image recognition and create workflows that simplify and optimize the ability to operate on data. HCP can also monitor AWS S3 buckets and bring new data back to HCP (Sync from) with S3 event notification through a GUI or Bucket_notification configuration to enable workflow automation. These capabilities can enable organizations to transform data more efficiently by creating it on-premises and then leveraging cloud services (like AWS transcoding) to operate on the data, transform it in the cloud, and return the processed data on-premises to be indexed and analyzed.

- **Highly Multitenant—**Multitenancy enables an organization to share one HCP cluster across multiple sub-units (departments for enterprise/government or subscribers for a service provider). Each sub-entity can customize its part of the cluster according to its requirements. HCP's compatibility with AWS S3 notification, S3 Select, and ObjectLock enables organizations and service providers to provide their services to multiple business divisions or tenant organizations securely.

- **Visibility and Notifications**—HCP uses Grafana, an open observability platform to provide customers with better visibility into their HCP environments with six dashboards, including system overview, system health, and troubleshooting dashboards. In addition, the system health dashboard provides alert notifications to customers via email when thresholds are exceeded or other criteria are met.

Enterprise Strategy Group (ESG) Tested

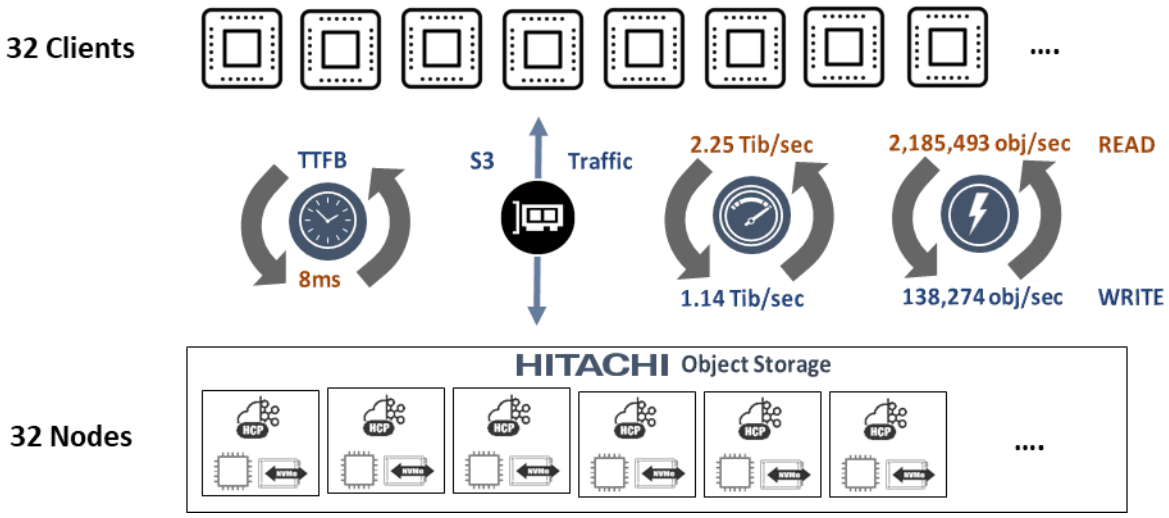
ESG audited Hitachi Vantara HCP software performance testing using a test bed located in AWS. Testing was designed to demonstrate read and write performance and time to first byte.

The test bed was based on a real customer deployment and used a 32-node HCP system built on standard AWS EC2 with local NVMe drives and 100GbE networking that was shared with other workloads to better simulate real-world customer environments. The network bandwidth available to each node was between 45 and 60 Gbps and was completely saturated. With more available bandwidth, higher throughput could reasonably be expected. It's important to note that in customer environments, the solution architecture will vary and can be scaled independently to meet the requirements of specific workloads.

Testing was focused on cloud-native applications using the S3 API for a wide variety of use cases. The mix of test workloads included 100% read, 100% write, and a blend of both. For each test, two object sizes were used in three configurations, 100% small (1KB), 100% large (64MB), and a mix of the two.

Retrieving data from object storage is done using the GET (e.g., read) task, while transferring data to the object store uses the PUT (e.g., write) task. Time to first byte metrics were gathered in real time from the application.

Figure 2. HCP Performance Test Bed in AWS



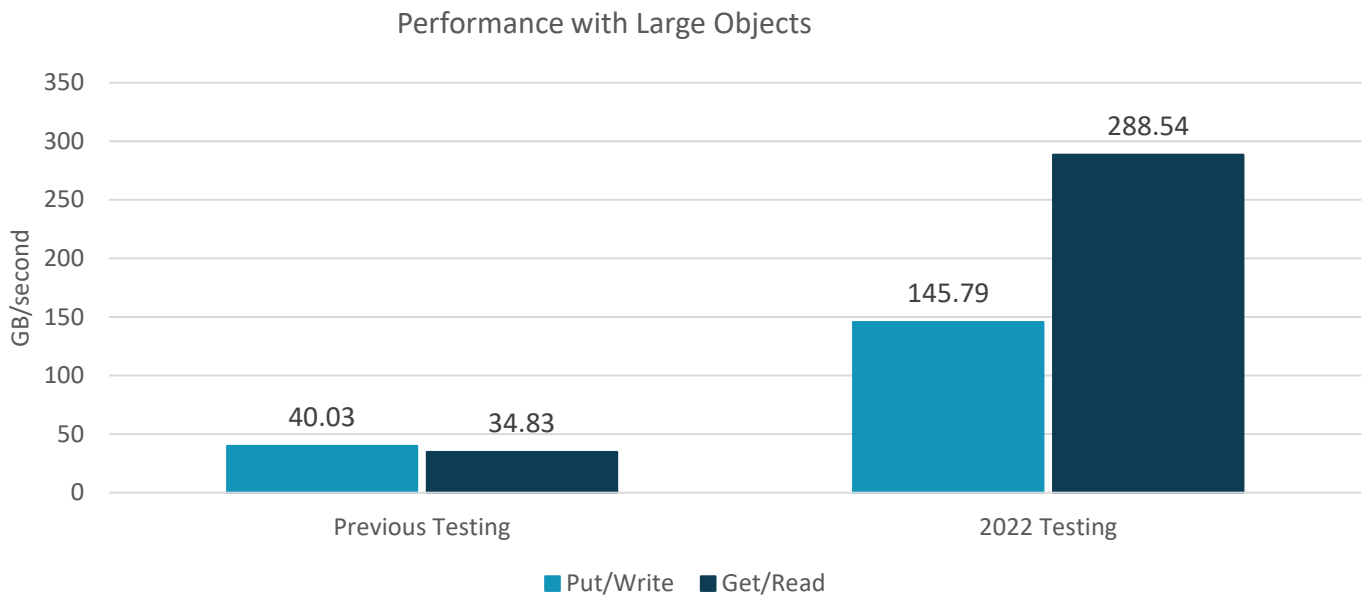
Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Large Object Testing

The key metric for large objects is bandwidth. Previous testing used 10MiB objects, similar to real-world objects, such as image files or backup data. Current testing used 64MiB objects, designed to emulate more modern real-world application needs, such as AI/ML and IoT data. Performance and scalability were tested using a 32-node HCP configuration.

Figure 3 shows the average performance for GETs and PUTs that HCP delivered for large objects. Large-block GET and PUT average performance is approximately four times the performance measured in 2020.

Figure 3. HCP Performance for Large Objects



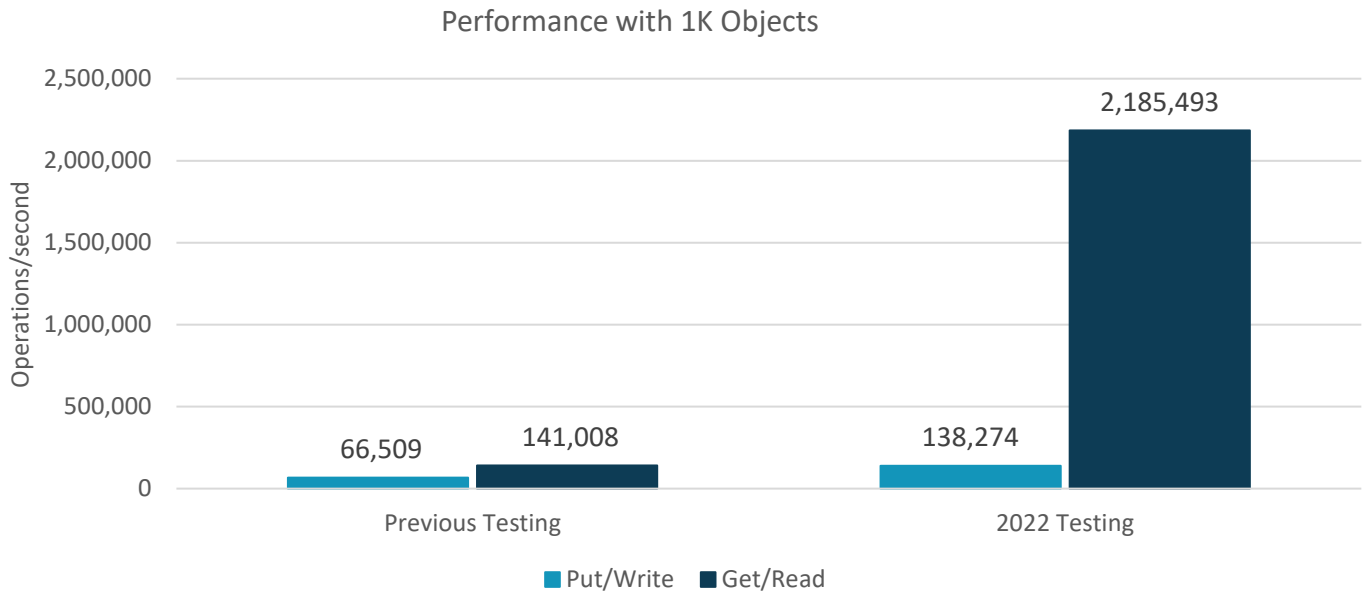
Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Small Object Testing

Next, we reviewed the throughput of 1KiB objects measured in operations/second. A real-world example would be small data files or metadata. Metadata performance is critical to maintaining high-performance data access for object storage.

As Figure 4 shows, small-block GET and PUT average performance has increased by approximately 10 times the performance measured in 2020.

Figure 4. HCP Performance for Small Objects

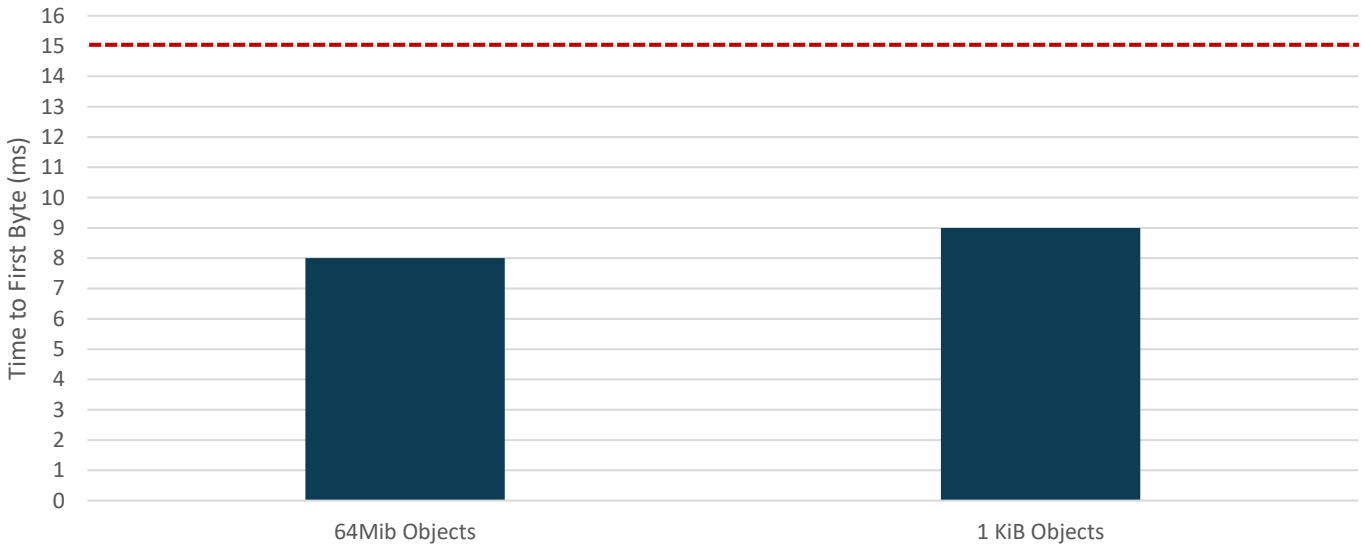


Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Time to First Byte (TTFB)

Time to first byte measures the responsiveness of the object store to requests made for data—that is, how long it takes for the first byte of data to reach the requester, enabling the user to begin using the data. This provides insight into how object storage affects productivity.

Figure 5. Time to First Byte



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

In previous testing, Hitachi Vantara established a high-performance time to first byte threshold of 15 milliseconds (ms). Note that 20 ms is considered an excellent response time that would not be noticeable to users. During both the large and small object testing, this metric was collected in real time, and the current HCP deployment yielded a best TTFB response time of 8ms for 64MiB objects and 9ms for 1KiB objects, both well below the 15ms threshold.

Why This Matters

Object storage is no longer relegated to archival usage focused solely on scale and cost per gigabyte. Today, organizations are storing mountains of data in object storage platforms and are using it for strategic business objectives, leveraging AI/ML for business analytics and customer insights, internet of things (IoT) and industrial internet of things (IIoT) analysis, and other modern production workloads. As a result, performance matters more than ever in object storage.

Enterprise Strategy Group (ESG) validated that HCP offers object storage performance optimized for modern workloads. Testing demonstrated average throughput of 145.79 GiB/sec for PUTs and 288.54 GiB/sec for GETs, more than eight times the performance measured in previous testing. Small object GET performance showed a more dramatic increase, to more than 2.1 million objects per second, more than 15 times the performance measured in 2020. These performance levels mean that organizations can count on HCP for production workloads, including modern analytics and AI/ML applications. In addition, HCP delivered a best TTFB response time of 8ms for large objects and 9ms for small objects, indicating excellent responsiveness and enabling organizations to entrust HCP with even more performance-sensitive workloads.

Customer Examples of HCP Performance

Enterprise Strategy Group (ESG) also reviewed several customer examples that demonstrate the real-world performance benefits of HCP object storage.

- 2 TB/minute throughput maintained for a 22 PB Hadoop data lake: A government agency used HCP to address a requirement to run massively scalable analytics on a Hadoop data lake. This customer was collecting huge amounts of streaming data, including voice, data logs, machine logs, and security event logs for analysis. Data lake seeding was being driven from Hadoop-based applications, normalized, and pushed to HCP using native Hadoop tools. The original performance requirement was to deliver 1 TB of data per minute from object storage. HCP was able to support twice the performance requirement, delivering 2 TB/minute throughput for the 22 PB Hadoop data lake using the S3A protocol. All-flash ingest nodes were deployed as an archival and data lake seeding target from Cloudera, orchestrated by Apache S3a. This supported the customer's application modernization migration to Kubernetes-based Spark and S3 Select workloads writing directly to HCP. The HCP solution also enabled the client to reduce their Hadoop licensing costs by 70%. In short, the HCP solution delivered faster, more accurate data insights across multiple data sources with cost efficiency.

> *"Based on the budget we have to work with, HCP can help us to keep data longer-term, which provides a deeper analysis of our users and data. Previously, we could retain just one to two months of data. With HCP, we can keep more than a year of data on the same budget, and scale far beyond that in the future. Flexibility is a top priority for us and we're very happy with this architecture because HCP allows us to have data-dependent scalability without compromising on performance requirements."*
> **•IT Services Director, Government Agency**

- CapEx and OpEx savings of $100M: A customer in the financial services industry had a global initiative to merge 51 archive solutions with separate technology to a single archive technology, distributed globally across four main regions. They selected Hitachi Vantara to provide HCP technology so they could aggregate data that supports the extraction, transformation, and load (ETL) processes between the existing archives and HCP. As they were displacing siloed repositories, they needed to validate and optimize data quality, aggregate and centralize data, and legally retain relevant data. Their goal was to create a single source of truth that simplified discovery and reporting. With HCP, this customer was able to transform their environment to a cloud services model and optimize their IT spending, reporting CapEx and OpEx savings of $100M USD while simultaneously mitigating future risk.

- Continuous availability: A US-based firm in the financial sector is leveraging HCP to meet stringent SEC document retention compliance requirements and an internal cloud service that enables their developers to leverage S3 object storage in-house. SEC regulations require them to enforce data retention policies, provide a searchable index, and ensure that data is retrievable and viewable, with backups stored offsite. There are four sites with 15 PB of total capacity in HCP. They are leveraging multiple high availability and disaster recovery capabilities, like active-active replication to provide business resilience. In the event of an outage, whether of an individual system or an entire data center, the client can redirect their applications and users without any downtime while services are restored.

- Higher performance with lower costs: A banking and investment firm based in the United States used HCP to consolidate legacy object and Fibre Channel SAN storage to support an in-house dedicated system with local storage for data processing and analytics. The short-term storage requirement for online customers, with sensitivity to the responsiveness of the application, drove the need for consistent, low-latency performance. The HCP environment used HCP all-flash ingest nodes to provide storage for their REST and S3-based applications. With their legacy system, they could only retain data on the in-house platform for about a week. The public cloud was used as an archive for long-term retention. HCP enabled them to use the application as they intended. Using a single pool of high-performance S3 storage, they eliminated the need to send sensitive data out to the public cloud. They were able to retain all of their data on-site for years rather than days. In summary, the client was able to implement a high-performance object storage solution using HCP that not only replaced slow legacy object stores but also allowed them to consolidate premium SAN storage and improve compliance while lowering costs.

- 15 ms or lower time to first byte: A large IT services provider used HCP to build a custom cloud-native application to provide digital services for their client. The primary requirement was for a solution capable of massive scale without compromising performance across multiple unpredictable workloads. The previous platform had the ability to scale capacity but not performance, so the user experience suffered. Another key requirement was the ability to locate files quickly and efficiently. Using custom metadata with HCP allowed them to tag the information inside the object store so users could find exactly the information they needed without having to crawl the entire data set. Time to first byte was an important performance metric for the client, and the HCP all-flash cluster demonstrated that it could achieve 15 ms or lower time to first byte, which was more than sufficient for the client's needs and gave them confidence that the solution could handle peak loads and spikes.

The Bigger Truth

The need for rapid insight is forcing organizations to prioritize agility, transparency, and speed across their data ecosystems with the goal of improving operational efficiency, improving collaboration, and accelerating time to value from investments in support of data-driven initiatives.

Traditionally, organizations have turned to object storage for highly scalable, cost-efficient, long-term storage with easy data retrieval—but not for performance. Hitachi Vantara has been raising the bar for performance, enabling organizations to use all their data to produce business insights and speed the development of new and innovative products. So much insight can be extracted from organizations' huge volumes of archived and new data that organizations can't afford to throw it away, but extracting value from it requires usable performance.

Hitachi Content Platform offers object storage for modern applications and workloads, including the latest AI/ML and analytics applications. HCP delivers the high performance and scalability that enterprises demand of their business-driving workloads.

Enterprise Strategy Group (ESG) validated:

- Average large object GET performance of 288.54 GiB/sec throughput and PUT performance of 145.79 GiB/sec.
- Average small object GET performance of more than 2.1 million objects per second and PUT performance of 140,008 objects per second.
- Time to first byte performance of 8 ms for large objects and 9ms for small objects.

These kinds of performance results continue to shift the role of object storage. Now, the object storage advantages of massive scalability, fast data retrieval, and cost-efficiency can be used with tier-1 production workloads. HCP customers shared some of the results they are experiencing:

- 2 TB/minute throughput maintained for a 22 PB Hadoop data lake.
- CapEx and OpEx savings of $100M USD while simultaneously mitigating risk.

This is a long way from traditional object storage. 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 their reliability, security, availability, and enterprise-class features. This HCP performance validation adds to the company's strong resume.

Of course, organizations' mileage may vary, as these tests were run in a controlled environment, 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 optimization, ESG recommends that you take a serious look at Hitachi Content Platform.

Enterprise Strategy Group by TechTarget

Enterprise Strategy Group is an integrated technology analysis, research, and strategy firm that provides market intelligence, actionable insight, and go-to-market content services to the global IT community.

www.esg-global.com contact@esg-global.com 508.482.0188