Western Digital.

CLOUDIAN®

REFERENCE ARCHITECTURE AUGUST 2022

Reference Architecture of Cloudian® on Ultrastar® Data60

Puspanjali Panda, Saravanakumar Pandian, Pavan Gururaj

Western Digital Corporation

Abstract

This reference architecture describes a Highly available cloud storage solution using Cloudian HyperStore® deployed on Western Digital's Ultrastar® Data60 storage platform. The reference architecture explains the design, deployment and performance summary.

Contents

Executive Summary

In the era of digital explosion, organizations that generate terabytes or even petabytes of data face many challenges. In recent times, data growth has been fueled by mostly online models for business, education, and entertainment. The unprecedented growth of data, and how to manage and monetize it, is a common challenge of modern enterprises. Legacy storage systems such as file and block struggle to manage the volume and velocity of unstructured data – despite their rich feature sets.

Another challenge IT leaders are finding is that it is difficult to keep pace with the explosive growth of data, which is driving up operational costs related to powering, cooling, and managing storage systems. Organizations started the trend towards implementing a software-defined storage (SDS) environment, combining S3-compliant object storage solution with industry-standard server hardware to address these challenges.

This document demonstrates the Cloudian HyperStore, a fully S3 API compliant hybrid cloud storage software deployed on Western Digital's Ultrastar Data60, which is the next-generation high-density storage platform for software-defined storage. It provides a low cost, resilient, high- availability, manageable, and scalable storage solution.

Problem Statement

In the midst of the digital evolution era, the growth of data is happening at a much more rapid rate than expected. That data is produced not only in data centers but increasingly at the edges as well. Because of this data growth, storage volumes and pools continue to expand, which leads to the scaling issues seen in many current deployments. Analysts from IDC expect data generation will grow to 175 zettabytes by 2025[1]. Almost 80% of that data will be unstructured.

Inflexible legacy solutions can't keep pace with this data growth, resulting in silos of data that are complex, expensive, and costly to manage. IT struggles to deliver against the massive scale required for modern workloads. This has given rise to a new, increasingly dominant class of storage – object storage. Object storage has become the de-facto standard for this class of data, just as this class of data has become the dominant data type. Cloud object storage architectures have emerged to manage and cater to this relentless growth of unstructured data. Using modern object storage approaches, organizations can collect and manage unstructured data, which may arrive in a variety of formats, far more easily.

The need of the hour is an object storage solution with ease of management, low TCO, linear scalability, and high availability to meet enterprise's growing data storage needs. In this document we demonstrate a solution to address the growing demand of unstructured data by deploying Cloudian HyperStore software on Western Digital's Ultrastar Data60 storage platform.

Solution Highlights

Western Digital, a pioneer in reliable, high-density industry-standard hardware for software-defined storage projects, is partnering with Cloudian, a file and object storage company specializing in S3-compatible object storage systems. The following sections of this paper provide an overview of this S3-compatible object storage solution – built on Western Digital, powered by Cloudian.

The solution combines:

- **Western Digital Ultrastar Data60:** The Ultrastar Data60 storage platform is a key element of next-generation disaggregated storage and software-defined storage (SDS) systems. It addresses the demanding storage needs of large enterprise customers, storage OEMs, cloud service providers, as well as resellers/integrators requiring dense, shared HDD.
- **Cloudian HyperStore:** Cloudian HyperStore is a scale-out object storage system designed to manage massive amounts of data. It is an SDS platform which runs on any standard Linux server. This dramatically reduces the cost of data center storage while still providing extreme availability and reliability.

[1] IDC Data Age 2025

By combining Cloudian HyperStore software with Western Digital Ultrastar Data60, organizations can achieve:

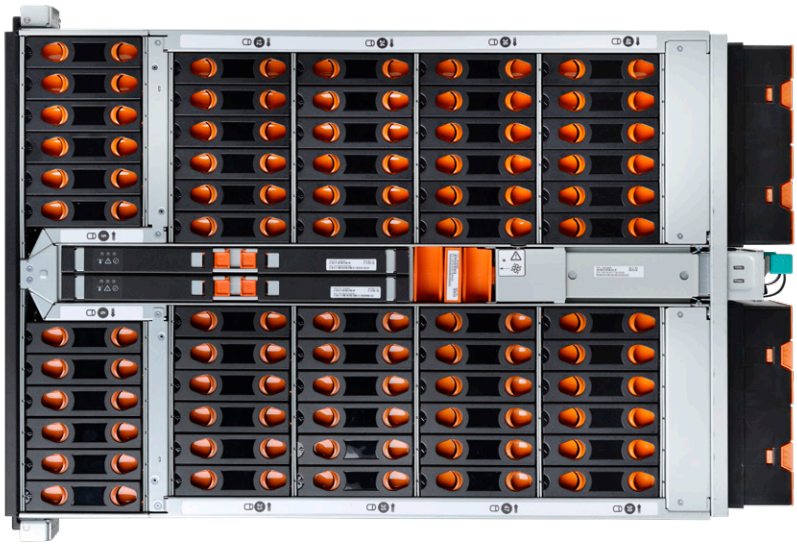
- Resilience and high availability
- Linear scalability
- Low TCO
- 100% native S3 API compliant object storage
- Ease of deployment
- Easy data management
- Integrated billing, management, and monitoring

Technology Overview

Ultrastar Data60 Storage Platform

Delivers High Density and Flexibility

The Ultrastar Data60 is a key element of next-generation disaggregated storage and software-defined storage (SDS) systems, delivering high density and the flexibility to balance performance with cost. The Ultrastar Data60 provides up to 1.2PB[2] of raw storage using our 20TB HDDs in a compact and efficient form factor. Western Digital HelioSeal® drives ensure cool running, quiet operation, and high reliability.



An Industry Leader in JBOD Storage

Direct-attached storage platforms don't have to be complicated, but you wouldn't know it looking at other solutions on the market. Many high-capacity storage enclosures include features and performance that enterprises rarely need, with a price tag to match. Others offer simple JBOD storage—but keep costs down by skimping on components, manufacturing, and warranties. Western Digital's Ultrastar storage platforms include unique technologies not found in any other storage platform: patented IsoVibe™ and innovative ArcticFlow™. IsoVibe reduces vibration-induced performance degradation, while ArcticFlow overcomes the cooling issues by introducing cool air into the middle of the platform. Combining these technologies with HelioSeal hard drives provides a solution designed for long-term reliability and reduced drive returns, enabling the safekeeping of all the digital content stored on the platform.

[2] One terabyte (TB) is equal to one trillion bytes and one petabyte (PB) is equal to one thousand TB. Actual user capacity may be less due to operating environment.

IsoVibe

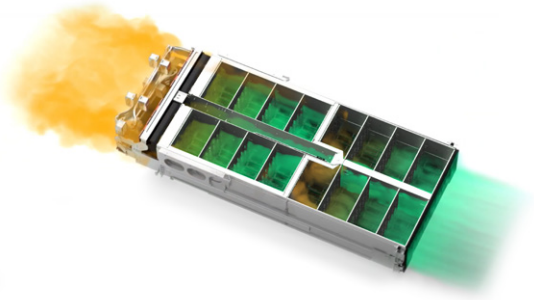
IsoVibe Vibration Isolation Technology

Precise cuts in the baseboard provide a suspension for the drives in the chassis, isolating them from transmitted vibration. The result is that consistent performance is maintained, even when all the drives are working hard.

ArcticFlow

ArcticFlow Thermal Zone Cooling Technology

Dividing its enclosure into multiple zones created an innovative air-ducting design that moves cool air directly into the center of the enclosure. The setup more effectively cools the controller modules that collect data from all the drives. This results in lower fan speeds, reduced vibration, lower power consumption, quieter operation, and ultimately higher reliability.

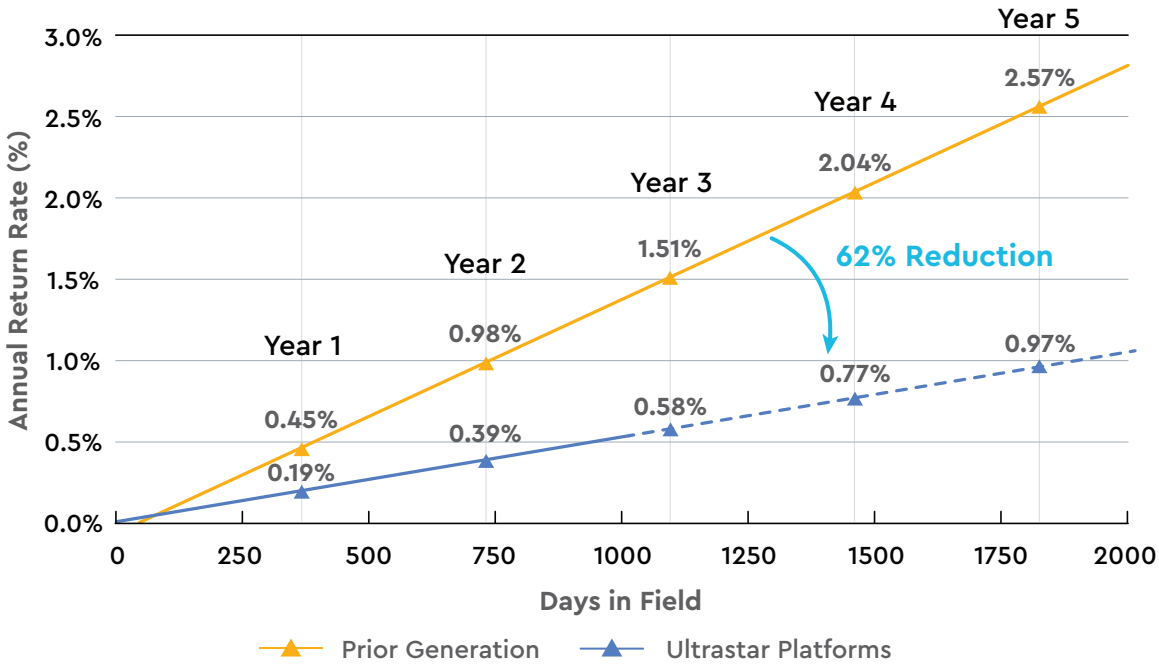


Resource Manager

Western Digital Resource Manager

A GUI-based tool, the Western Digital Resource Manager enables real-time monitoring and management of the platform and provides a consolidated dashboard displaying the most critical information. Other views allow platform configuration, health monitoring, and maintenance.

After implementing both ArcticFlow and IsoVibe in product enclosures, Western Digital product management engineers compared the field-return rates for the two product sets – which had been sold into the same types of customer environments and were supporting the same types of workloads with the same kind of disk drives. In the span of about three years, field-return rates have dropped by 62% (see Figure below). These findings are based on a very large sample size – in the range of hundreds of thousands of drives.



Designed for the Enterprise and the Cloud

This platform addresses the demanding storage needs of large enterprise customers, storage OEMs, cloud service providers, and resellers/integrators that require dense, shared HDD storage. The Ultrastar Data60 provides the flexibility to specify the HDD combinations to balance capacity, performance, and cost.

Platform Specification Summary

Specification	Value
Maximum Drives	• 60 × 3.5" drive bays
Drive Interface	• 12Gb/s SAS • 6Gb/s SATA
Available Drive Capacities	• HDD up to 20TB CMR or up to 20TB SMR
Fabric Adapter(s)	• Dual redundant I/O Modules (IOM) • 6 Mini-SAS HD ports per IOM
Power	• Dual 1600W or 1800W, 80+ Platinum • 200–240V AC input, auto ranging, 50–60Hz
Cooling	• 4 main enclosure fans, front-to-rear system cooling with zero-loss backflow prevention • 1 IO module fan • Dual PSUs with built-in fans
Serviceability	• Cable-free, hot-swappable IOM, power supply, fans, and drives

Cloudian Overview

Cloudian HyperStore is a scale-out object storage system designed to manage massive amounts of unstructured data, both objects and files, to a single, limitlessly scalable storage pool. It ensures unlimited scale, multi-datacenter storage, fully automated data tiering, and support for all S3 applications. It enables organizations to deploy private and hybrid clouds. Hybrid cloud storage deployed using Cloudian lets organizations manage storage resources on-prem and in the public cloud within a single namespace. At costs down to 1⁄2 cent per GB per month, Cloudian HyperStore's efficient architecture and industry-standard components can reduce Capex costs up to 70%.

Cloudian Features/Benefits

- Limitlessly scalable capacity
- Start with just three nodes and grow
- Non-disruptive capacity expansion
- 100% native S3 API for guaranteed compatibility
- Replication and erasure coding
- Bucket-level granularity for all storage policies
- Encryption for secure data storage
- Comprehensive multi-tenant services
- Quality of Service (QoS) controls
- Integrated billing, management, and software
- Geographically distributed replication

Cloudian HyperStore Components

The Cloudian HyperStore is a fully distributed architecture that provides no single point of failure, data protection options (replication or erasure coding), data recovery upon a node failure, dynamic re-balancing upon node addition, multi-datacenter, and multi-region support. The diagram below shows all of the service components that comprise a Cloudian HyperStore system.

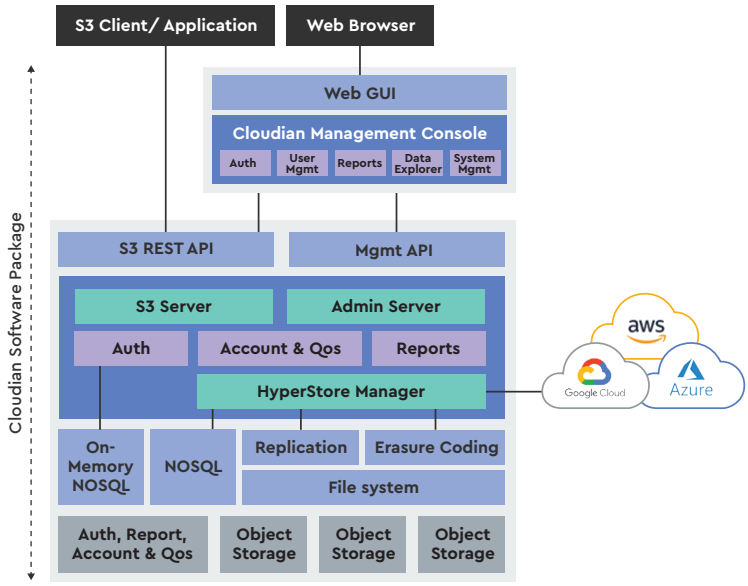
The Cloudian HyperStore system supports several types of services each of which plays a role in implementing the overall Cloudian HyperStore S3 object storage service:

S3 Service

- High-performance S3 proxy service processes S3 REST requests incoming from client applications.

Cloudian HyperStore Service and the HSFS

- Cassandra object store can optionally be used for small S3 data objects.
- Cloudian HyperStore File System (HSFS) is the area of Linux® file system on Apache Cassandra® nodes, where large S3 objects are stored.
- Cassandra capabilities are used to determine the distributed data management information such as the nodes that a specific key should be written to and replicated on and HSFS is used as the storage layer to store S3 object data.
- Within the HSFS, objects can be stored and protected in either of two ways: Replicated storage or Erasure coded storage.



Cassandra® DB Services

- The Cloudian HyperStore system uses the open source storage platform Cassandra to store several types of data. The Cloudian HyperStore system creates and uses several distinct "key spaces" within Cassandra.

Redis® DB Services

- Redis Key Value Data Store is a lightweight, open-source database that supports a variety of data and also supports Cloudian HyperStore S3 service features.
- Redis Credentials DB stores user credentials and additional S3 operation supporting data such as multi-part upload session information and public URL access counters.
- Redis QoS DB stores user-level and group-level QoS settings that have been established by system administrators. The DB is also used to keep count of user requests, so that QoS limits can be enforced by the system.

Administrative Service

- HyperStore Administrative Service leverages a RESTful HTTP API to enable system administrators to easily perform administrative operations.

Cloudian Management Console (CMC)

- The CMC is a web-based user interface for Cloudian HyperStore system administrators, group administrators, and end users.

Supporting Services

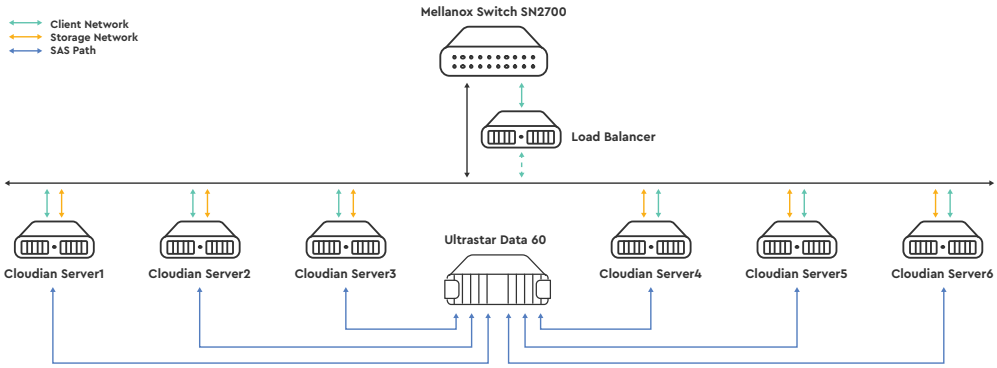
- Cloudian Monitoring Agent – The Cloudian Monitoring Agent runs on each Cloudian HyperStore node and monitors node health and performance statistics.
- Cloudian Monitoring Data Collector – The Cloudian Monitoring Data Collector runs on one node in each service region, and regularly collects data from the Monitoring Agents.
- Puppet – The HyperStore installer uses this to implement initial configuration and for ongoing configuration management.
- Dnsmasq – Dnsmasq is a lightweight domain resolution utility suitable for small networks. This utility is bundled with Cloudian HyperStore software and the Installer installs and configures Dnsmasq to resolve Cloudian HyperStore service domains.

Cloudian Deployment on Ultrastar Data60

Cloudian HyperStore is an S3-compatible object storage platform designed to manage limitless amounts of data in a cloud-native data format. It helps to consolidate your unstructured data to a single, dynamically-scalable storage environment. It delivers a fully S3 API compliant, multi-tenant, and multi-datacenter hybrid cloud storage solution. Cloud service providers use Cloudian HyperStore software to deploy public clouds and manage private clouds. Enterprises use HyperStore software to deploy private and hybrid clouds.

We have verified a functional deployment of Cloudian HyperStore with a single, high-density Ultrastar Data60 storage platform. The Ultrastar Data60 Storage Platform is acting as the object storage backend, which is hosting all the Cloudian data. Ultrastar Data60 with six servers are used in this deployment.

Cloudian Deployment Design



Configuration Details

1. HDDs are part of Ultrastar Data60 JBOD. Each storage server has 10 × 18TB HDDs which are part of the JBOD. 10HDDs are exposed to each server using zoning configuration. Total of 60 HDDs in 6 servers.
2. There are two additional SSD disks connected per server which are plugged in the backplane of each server to store the metadata information and have no connectivity to the JBOD(UD60).

Cloudian Server Configuration Details

Cloudian Storage Server Details	
Storage Product	Ultrastar Data60
Hyperstore Version	7.3.1.1 and 7.4
Storage Interface	6 × 12Gb/s SAS-3 host connections
Server Details	Dell PowerEdge R550
OS	Cent OS 7.9
Kernel Version	3.10.0–1160.el7.x86_64
NIC	1 x CX5 – (100Gbps)
CX5 OFED Package Version	5.4–1.0.3.0
CPU	Intel® Xeon® Gold 5318Y
CPU Core Details	Dual socket server with 24 core CPU each. 96 logical cores in total with HT enabled
Memory	256GiB
Switch	Mellanox® SN2700
Network Speed	100G
Cable	QSFP Cables used for client connectivity
Number of HDDs on each Storage Server	10 * 18TB each & Model: WUH721818AL5204
Number of SSDs on each Storage Server	2 * 3.2TB each connected directly to Dell Server back plane for meta data storage

High Availability

The Western Digital Ultrastar Data60 solution with Cloudian HyperStore contains high availability attributes. By default, every node includes the core services installed and running, such as Cassandra, Redis, HyperStore, S3, and the Cloudian Management Console. Redis services are installed in a master and slave configuration and can fail over to another node if an error is detected. For example, a deployment of six nodes can tolerate up to two node failures if the default read and write consistency level of quorum is met.

Cloudian Management Console

The Cloudian Management Console (CMC) can be installed independently on a separate node in the cluster that is running core services, such as Cassandra, S3, HyperStore, and Redis. The CMC also can work in a load balancer configuration for high-availability and to scale performance. Alternatively, the CMC can be run on a less powerful server separate from the data nodes.

Load Balancers

The Cloudian HyperStore software works with most load balancers that are available on the market. Services that are available for load balancing are: S3, Admin, and the CMC.

Backup

By using third-party applications, data that is stored in Cloudian HyperStore can be backed up to other storage mechanisms, such as tape, another Cloudian HyperStore deployment, or other cloud storage providers.

Tiering

Cloudian HyperStore allows users to transition objects on a per-bucket level basis to Amazon S3™, Amazon Glacier™, or another Cloudian HyperStore cluster.

Multi-Tenancy

Multi-tenant support enables service providers to offer storage to multiple customers. System administrators can isolate different customers or departments of an organization into separate groups on a single HyperStore cluster. Each tenant can have group administrators to manage administration tasks such as adding or removing authorized users, maintaining, and implementing quality of service settings, and creating usage reports.

Cloudian HyperStore Management

One Simple Web Based GUI

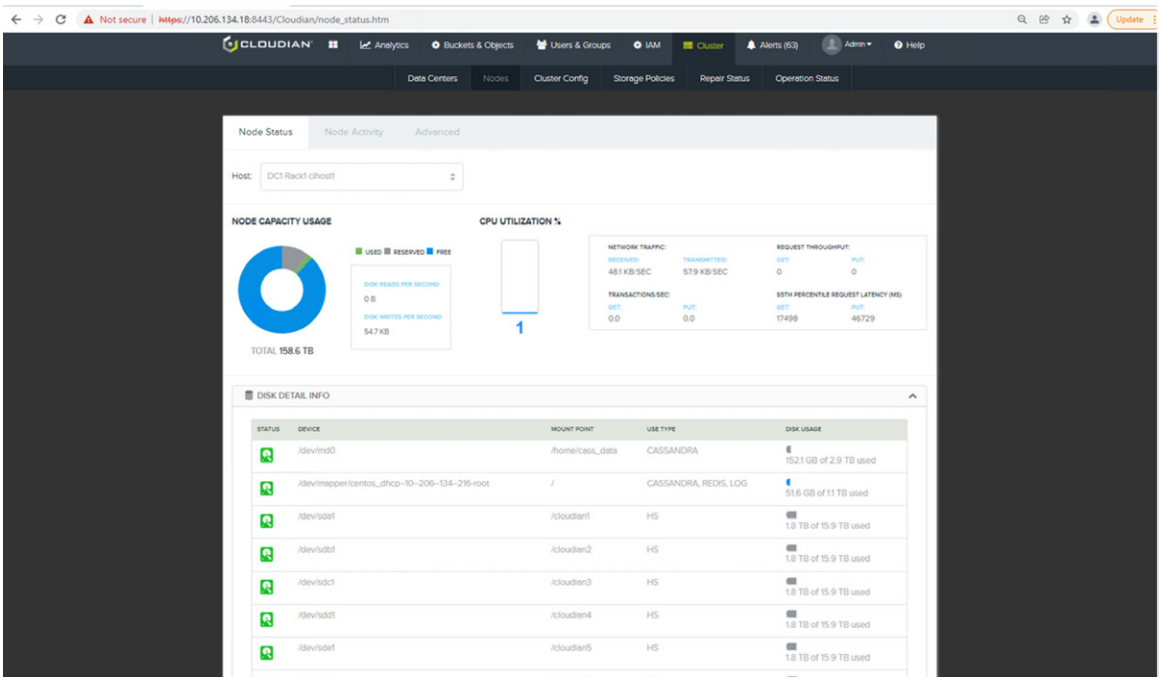
The Cloudian Management Console (CMC) is a web-based user interface for Cloudian HyperStore system administrators, group administrators, and end users. The functionality available through the CMC depends on the user type associated with a user's login ID (system administrative, group administrative, or regular user).

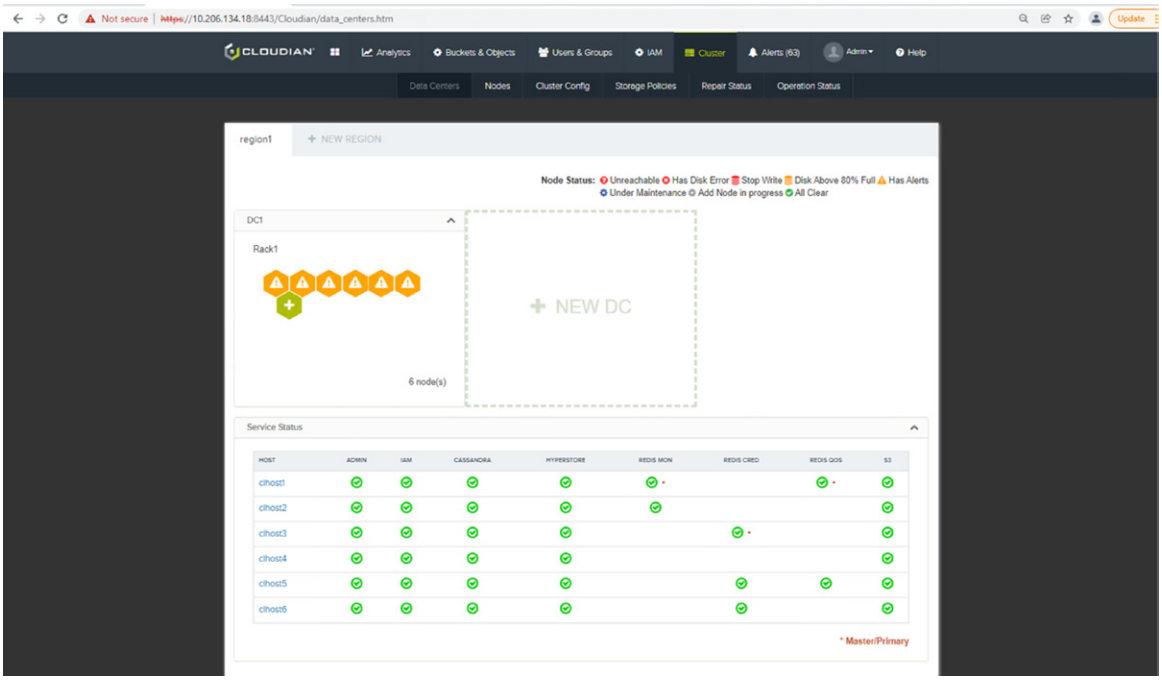
As a Cloudian HyperStore system administrator, you can use the CMC to perform tasks such as:

- Provisioning groups and users
- Managing quality of service (QoS) controls
- Creating and managing rating plans
- Generating usage data reports
- Generating bills
- Viewing and managing users' stored data objects
- Setting access control rights on users' buckets and stored objects

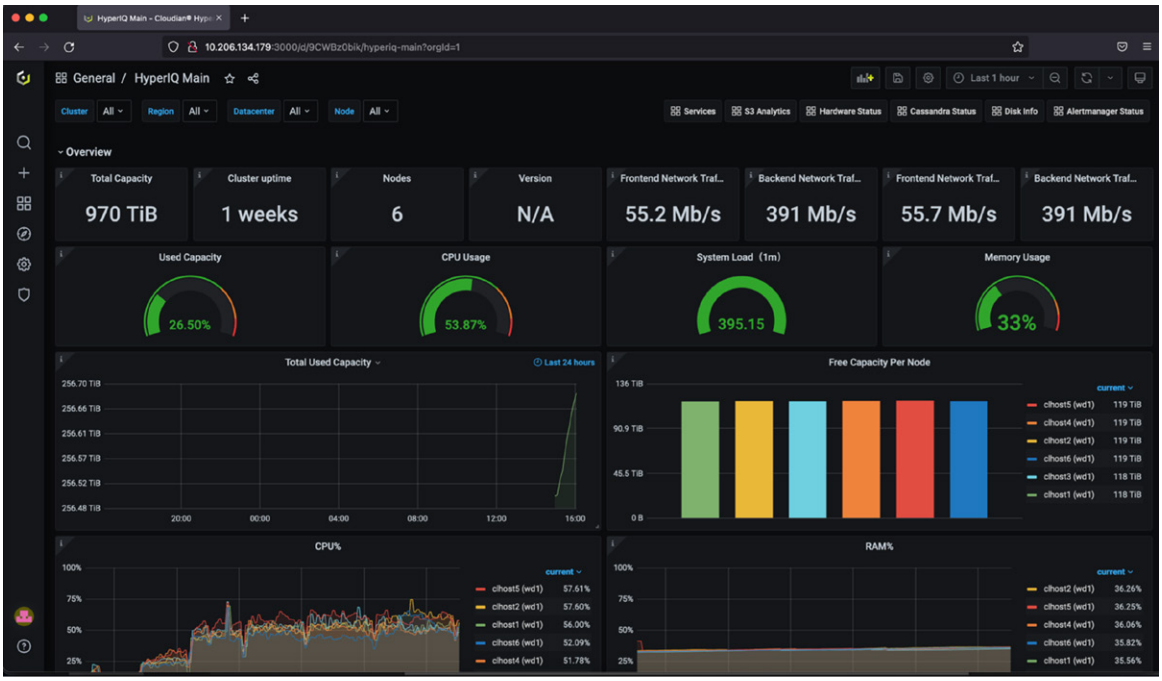
Group administrators can perform a more limited range of administrative tasks pertaining to their own group. Regular users can perform S3 operations such as uploading and downloading S3 objects. The CMC acts as a client to the Administrative Service and the S3 Service.

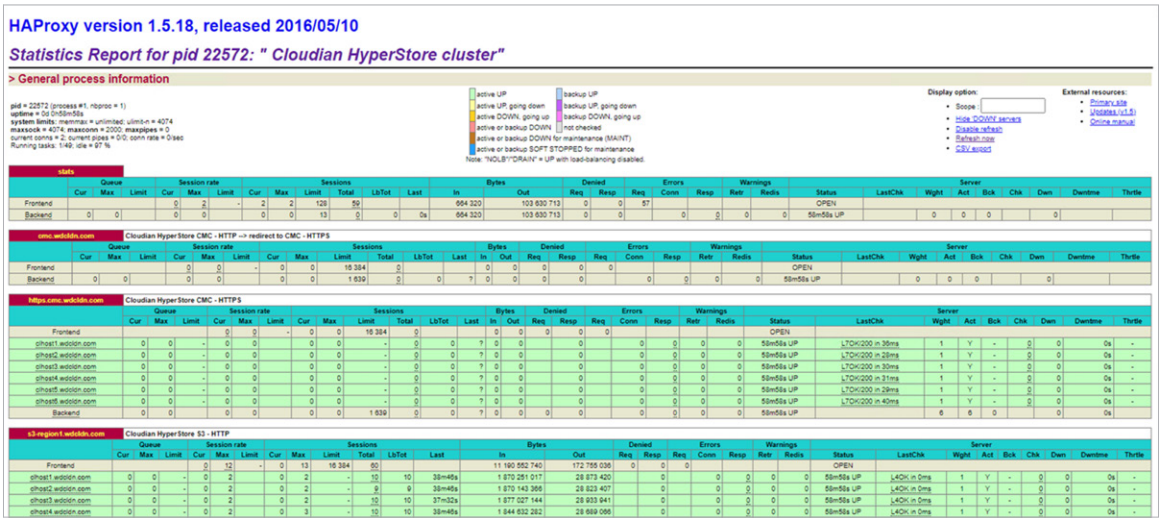
The CMC dashboard provides a high-level view of the status of your Cloudian HyperStore object storage service. If you have multiple service regions, there is a separate dashboard view for each region.





Cloudian HyperStore's scale-out architecture provides many benefits for availability and performance. A single HyperStore cluster provides a pooled resource of storage capacity, IOPS, and bandwidth, based on the number and type of server used. Applications share this pool of resources, and with proper planning applications can leverage HyperStore to provide predictable and stable performance.





This document describes the implementation and design of Cloudian HyperStore platform on Ultrastar Data60. This document will explain the data bandwidth, throughput, and latency charts for varying IO workloads (different number of threads and object sizes) to understand the expected performance deliverables from your Cloudian HyperStore cluster. When assessing data storage performance requirements, the three key metrics to be considered for an application to store and recall data from a storage system are as follows.

Bandwidth Throughput Measured in Megabytes/Second (MB/s)

This is the amount of data that can be processed by a storage system within a second. Maximum throughput for a disk could be for example 100 MB/s, which allows 100MB in a second. A disk will have different bandwidth numbers for reading and writing data as the mechanisms for those different IO operations are different. It is much easier to read data than to write data.

IO Operations Per second (IOPs)

IOPs is the number of read or write operations that can be achieved in one second by a disk. A disk will have different IOPs ratings for reading and writing data as the mechanisms for those different IO operations are different. It is much easier to read data than to write data.

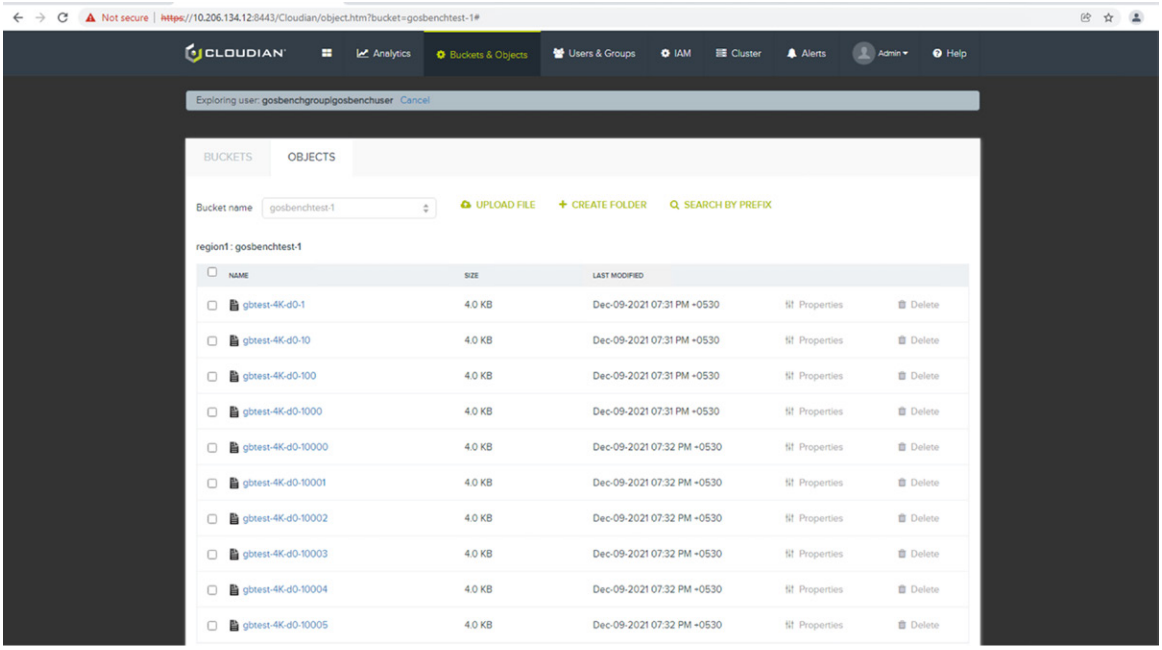
Response Time Requirements (Latency in Milliseconds)

The time taken from the application to send the IO request to the storage system and for the storage system to process the request and acknowledge back to the application that the request has completed. In simple terms it is the wait period between two consecutive sequential IO requests. This is most important for workloads that are sending frequent small data requests to a storage system, as the time between each request may result in IO being queued and acknowledgments received too late that could cause a bottleneck. Latency is impacted by all aspects of the client to storage connectivity stack. For larger data object sizes, this is typically a less important factor. There is a direct relationship between IOPs and Throughput (MB/s). A certain amount of IO operations will also give a certain throughput of megabytes each second, so these two are related.

Testing Methodology Overview

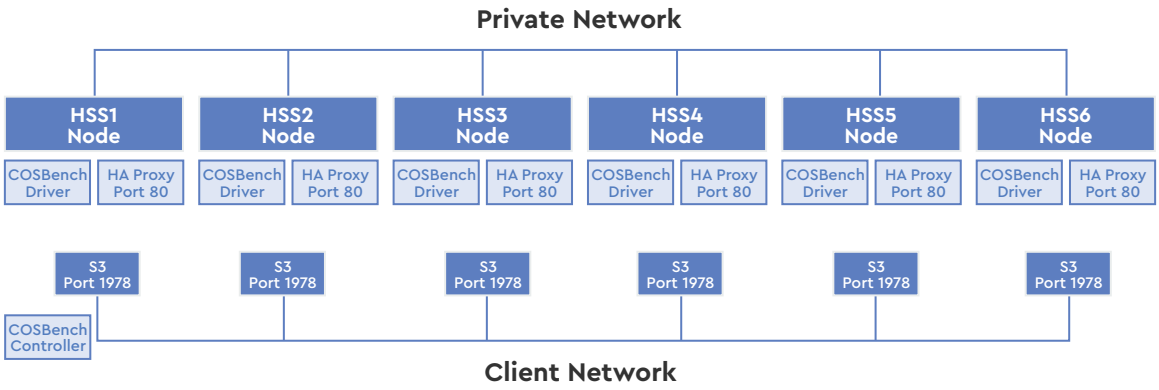
Gosbench was configured to run on each HyperStore cluster node to avoid impact on production client network and to ensure there were sufficient Gosbench worker nodes to stress the cluster capabilities.

Each node also had haproxy load-balancing software installed and was configured to ensure that every transaction generated by Gosbench was sent over the network, rather than taking advantage of local caching and skewing the results. This testing methodology has been compared with using standard external Gosbench clients and the difference in results are negligible compared to the benefit of flexible testing with no production impact.



As shown in the below figure, the test environment included the following configuration settings.

- HyperStore configured to run S3 API on port 1978 and haproxy on port 80.
- All HyperStore nodes were used to generate Gosbench workloads and send the workloads to the local running haproxy address.
- Each node running the haproxy service was configured to forward traffic to all cluster nodes except for the local node (itself) to simulate real client workload, preventing local cache hits and limiting bandwidth to actual physical connectivity available to the nodes.



Different applications will exhibit different IO characteristics depending on their function. To provide a starting point for the assessment of application IO requirements, the following table shows the most important IO metrics to optimize based on some common application use cases. This at least provides an initial assessment of the type of storage to use to support the application IO requirements.

Most important requirement

To be considered but less important then other factors

Least important factors

Applications	File size	Bandwidth reqs	IOPs reqs	Latency reqs
Virtual Machines	Small, Mixed			
OLTP Databases	Small			
Email Systems	Small, Mixed			
File Sharing	Mixed			
Backup	Large			
Archive	Large			
Media (Video, Images, Audio)	Large			
Healthcare PACs (X-Ray, MRI, Ultrasound)	Large			
Security (Images, Video, Voice Recordings)	Large			

Once you understand the application requirements for IOPs/bandwidth/latency you must start to consider how the application generates that workload. The key variables to consider from the application perspective include the following considerations.

Data Object Size

The application/service that needs storage access will issue a request to read or write a certain amount of data at the same time. This is called the IO size and could be for example 4KB, 8KB, 32KB up into the TBs range. The minimum amount of data to read/write to a single physical disk drive is the size of one sector, which was 512 bytes only, but now 4KB is the norm. The data object size read or written by the application to the storage platform has a significant impact on storage performance. Object storage platforms are designed for larger data objects which allows for greater throughput performance scalability as opposed to smaller objects which are typically representative of transactional workloads.

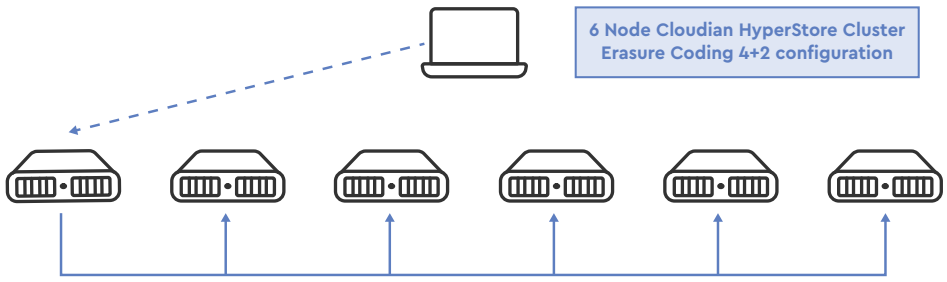
(IOPs x Object size = Throughput)

Read\Write Ratio

IO write operations are more expensive on resources than read operations. Every time a new object is written to the storage platform, data protection parity calculations must be performed that impact IO. Typically, storage systems deliver greater performance for read IO overwrite with the same configuration. Understanding the mix is important to accurately size the storage system.

Number of Client IO Threads

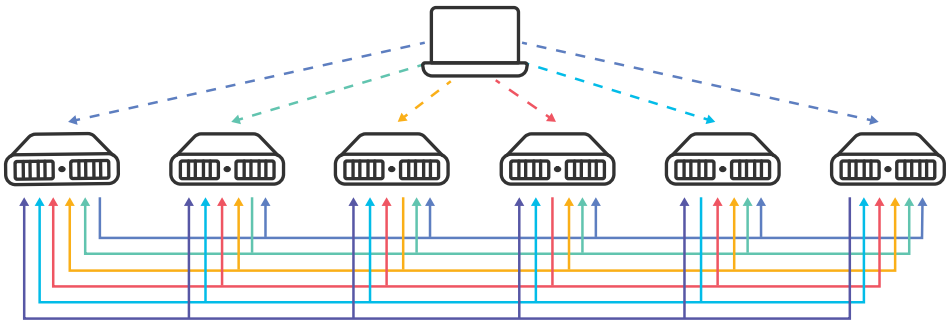
Cloudian HyperStore is a scale-out storage architecture which pools multiple independent nodes into a single storage namespace. To utilize all available hardware to deliver the required performance, the system will round robin client requests across each node in the cluster so that all nodes are working in parallel.



Single Threaded Write
Node 1 acts as coordinator node for the IO request
Calculates the parity and breaks the data object into EC fragments
6 fragments created (EC4+2) and 1 fragment distributed to each node over the backend storage network

As shown in the above figure, a single stream from an application will provide limited performance scalability as it is using only one node for each IO request.

To maximize the advantages of this capability, multiple "worker" threads allow an application to send multiple data streams to the storage system at once as can be seen in the below figure.



Multi Threaded Write (6 threads)
All nodes acts as coordinator nodes for each IO request
Each co-ordinator node calculates the parity and breaks the data object into EC fragments
6 fragments created (EC4+2) for each IO thread and 1 fragment distributed to each node over the backend storage network

There is also an upper limit on the number of connections per HyperStore cluster node to consider. The number of connections a node can simultaneously process is a function of the hardware capability and the workload, so mileage can vary, but Cloudian system engineers take 100 concurrent connections per node as a rule of thumb when sizing HyperStore clusters.

Support S3 Multi Part Upload (MPU)

If the client application is using S3, the ability to use the MPU feature of the S3 API can enhance the performance of even a single threaded application as MPU will split a larger file into smaller parts and distribute them evenly across all cluster nodes. Each part is treated as an individual object, therefore multi parts can use multi nodes.

Random\Sequential IO Ratio

Seek time for data on individual disks has always been a performance consideration, having to consider the time it takes for the drive head to find the right data block and position the head to locate with the block. Every time a reposition occurs then performance is impacted. Traditionally, having a larger file written contiguously on the same platter allows a single to reposition to transfer more data faster without having to move to another location on the platter to find the data. This is usually referred to as a sequential data pattern.

Random patterns usually occur when you have smaller data block sizes which get intermingled with each other and the drive must keep moving the head to find the next block. Although this is still an underlying consideration for all magnetic disks (not SSD) most enterprise storage solutions compensate for this by virtualizing physical disk resources and distributing data across multiple drives (and now nodes) to parallelize operations.

HyperStore cluster throughput, latency, and transactional capacity can vary based on the number of nodes in a cluster, as well as the CPU core to disk ratio, networking, and application implementation. In general, adding nodes would increase throughput and lower latency. Refer to the below details for the maximum performance achieved with the Cloudian HyperStore Cluster with Ultrastar Data60.

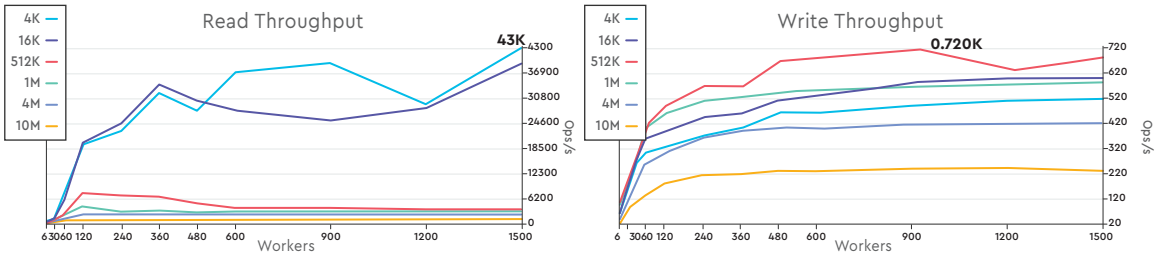
Performance Results

Cloudian Performance Test Results (Proserv)

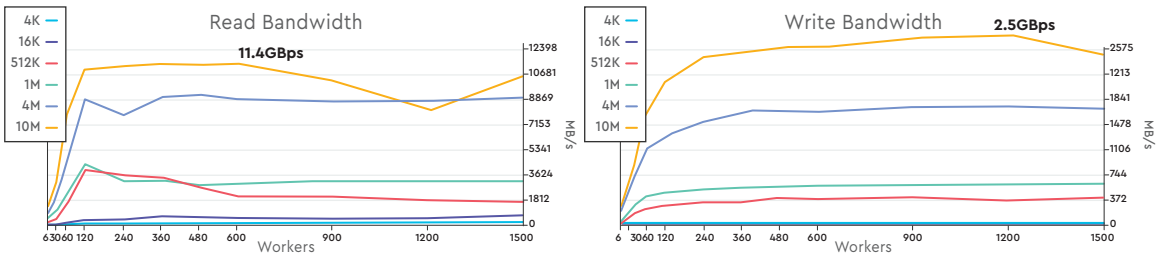
	Read BW		Write BW	
Storage Policy	BW	IOps	BW	IOps
EC4+2 (256GiB Memory)	11.4 GBps	43K	2.5 GBps	0.720K
RF3(256GiB Memory)	11.4 GBps	38K	2.8 GBps	1.1K

HyperStore EC (4+2) Policy Performance Results

Hyperstore IOPs Results (EC4+2 Policy):

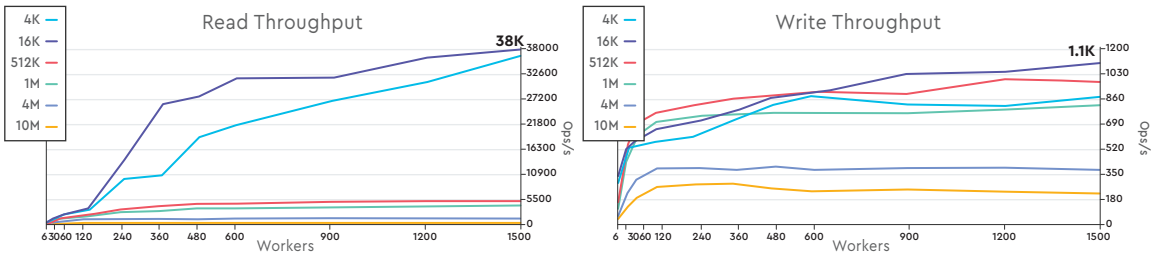


Hyperstore Bandwidth Result (EC4+2 Policy):

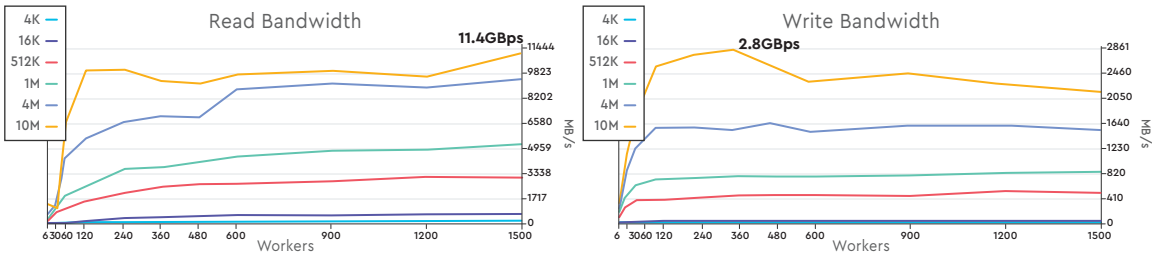


HyperStore RF3 Policy Performance Results

Hyperstore IOPs Results (RF3 Policy):



Hyperstore Bandwidth Result (RF3 Policy):



Features & Benefits Review

- Limitlessly scalable capacity
- Start with just three nodes and grow
- Non-disruptive capacity expansion
- 100% native S3 API for guaranteed compatibility
- Replication and erasure coding
- Bucket-level granularity for all storage policies
- Encryption for secure data storage
- Comprehensive multi-tenant services
- Quality of Service (QoS) controls

Use Cases

- Backup and disaster recovery
- File consolidation
- Media active archive
- Collaboration
- File lifecycle management
- AI / Machine Learning
- Data analytics
- Storage-as-a-Service (STaaS)

Conclusion

As a global leader of data storage technology, Western Digital consistently monitors the changing technological, strategic, and organizational requirements in the data center. Our observations show that as the complexities of IT environments have increased in general, so have the complexities of data protection environments in particular. Efficient data protection and data availability in all scenarios call for infrastructure and management consolidation, with the main goal of ensuring high data availability and fast disaster recoveries. To meet these requirements, we are offering a consistent data protection strategy with our solution.

By combining Cloudian with Western Digital's Ultrastar Data60 platform, organizations can achieve:

- Unlimited scale
- Multidata-center storage
- Fully automated data tiering
- Resilient, high-availability storage
- Robust high availability

References

https://documents.westerndigital.com/content/dam/doc-library/en_us/assets/public/western-digital/product/platforms/ultrastar-data60-hybrid-platform/product-brief-ultrastar-data60-hybrid-storage-platform-gov.pdf

https://documents.westerndigital.com/content/dam/doc-library/en_us/assetsv/public/western-digital/collateral/tech-brief/tech-brief-arcticflow-and-isovibe.pdf

https://cloudian.com/products/hyperstore/

https://cloudian.com/resource/

For Cloudian software, license and documents, please contact @ Cloudian_Support and Cloudian_Resource.

5601 Great Oaks Parkway
San Jose, CA 95119, USA
www.westerndigital.com