Cloudian HyperStore

Overview

Cloudian HyperStore is offered as either an appliance or software only using commodity servers and storage devices as a complete system or with node instances in public clouds – AWS, Azure, and Google. The appliance-based systems can scale with multiple nodes that have devices for metadata and data with differing numbers and type of processors with multiple cores.

The Cloudian HyperStore software includes functionality expected of object storage systems to compress and encrypt data, protect data using forward error correction with erasure coding, and seamless scaling with addition of nodes and dynamic rebalancing. Additionally, HyperStore includes the ability to auto-tier data using a policy engine and data mover to public clouds, multi-tenancy support with assignable permissions, and billing/chargeback and quota support for APIs.

The HyperStore system supports access using the S3 protocol for objects but does not have native support for file access. Objects are stored in an underlying EXT4 filesystem. Cloudian also has software that can be installed on a server or in a Virtual Machine to provide file access and controls. Cloudian HyperFile is a NAS controller that provides file to object access. The HyperFile offering provides:

- NFSv4, CIFS/SMB 3.0 (SAMBA), and FTP access
- AD and LDAP support
- Global namespace access (user HyperStore)
- Object versioning (using HyperStore)
- Snapshots and WORM setting
- Scale out with multiple NAS controllers with global namespace

The HyperStore software also has a plugin for Hadoop access. Both OpenStack and CloudStack integrations are supported as well.

Highlights

- Scaling object storage system – appliance or software only
- Data stored as files
- S3 APIs
- Geo-distribution for Forward Error Correction
- Auto-tiering to public clouds
- Additional NAS controllers for file access
- File migration engine with NAS controllers
- Encryption
- Compression
- Multi-tenancy
- Billing/chargeback & quota APIs
- Object versioning
- Can have node instances in clouds – AWS, Azure, and Google
- Kubernetes S3 operator for container support

Cloudian has additionally positioned HyperStore as a persistent storage solution for containers and Kubernetes environments. HyperStore is deployable as a container and features a Kubernetes S3 operator that allows object storage to be dynamically provisioned for containers.

Usage and Deployment

Cloudian positions the HyperStore for use as on-premises or public cloud object storage for a number of purposes: content repository including retained backups and online archives, analytics data, file services, hybrid store with public clouds, container storage, and multi-cloud support. HyperStore software is for those building their own object store and larger service provider environments.

- Characteristics
  - Scale – Scaling with multiple nodes. Capacity scales in device increments to a virtually unlimited capacity in a cluster.
  - Protection / Durability / Resiliency – Forward Error Correction with Erasure Coding is used to protect from data element or node failure. Distribution of data across geographies is controllable based on configuration data. Asynchronous remote replication to multiple sites is supported.
  - Index and Search – Index and search capability is supported through integration with Elastic Search.
  - Performance – Performance information has not been published. Does have QoS controls and flash for metadata.
  - Access Methods – Amazon S3 API over HTTP/REST for object access. HDFS plugin for Hadoop access. File access requires HyperFile controllers to get NFSv4, CIFS/SMB 3.0, and FTP access. Support for S3 Select.
  - Geographic Access – Geo-Distribution will distribute data across geographic areas with configuration settings.
  - Security and Compliance – Multi-tenancy is supported with multiple administrative and user access roles. Communications are encrypted with HTTPS. Data at rest would be encrypted with internal key management. Does have Common Criteria certification but not SEC 17a-4.
  - Metadata – System and user metadata are added to the object storage.
  - Billing and Chargeback – Chargeback and billing capabilities are available with usage data.
- Applications
  - Content repositories, cloud-based object storage, and Hadoop storage

  - Service providers or private cloud storage.
  - Big data analytics storage
  - Cloud archiving
  - Large scale web or cloud applications
  - Persistent storage for container environments
- System environments
  - Any environment with web access – either through private cloud or cloud service providers
- Deployment and Administration
  - HyperStore has a management GUI (Cloudian Management Console) with self-service dashboard.
  - HyperStore runs as a container

Key Capabilities

Architecture and Deployment

As with most object storage systems, Cloudian HyperStore uses individual nodes that are connected over Ethernet to hosts and other nodes to store and retrieve data. The data is in the form of objects accessed by the S3 protocol. Multiple nodes can be configured with a minimum of two required. The nodes are connected in a ring structure with data protected across the nodes either by storing three copies on three separate nodes or mathematically distributing the data using erasure error correcting codes. With the addition of nodes to the ring, data is automatically redistributed as all resources – processors and devices – are used in a common pool.

Key to the Cloudian implementation is the use of a Cassandra no-SQL database. The database stores information about the configuration, where the data elements are distributed, and metadata for the objects.

In addition to the nodes, HyperStore extends the abstraction with use of virtual nodes or vNodes. The storage resources within a node are sub-divided into smaller I/O groups and called vNodes. This allows for greater parallelism for data access and for improved recovery times in case of a device failure.

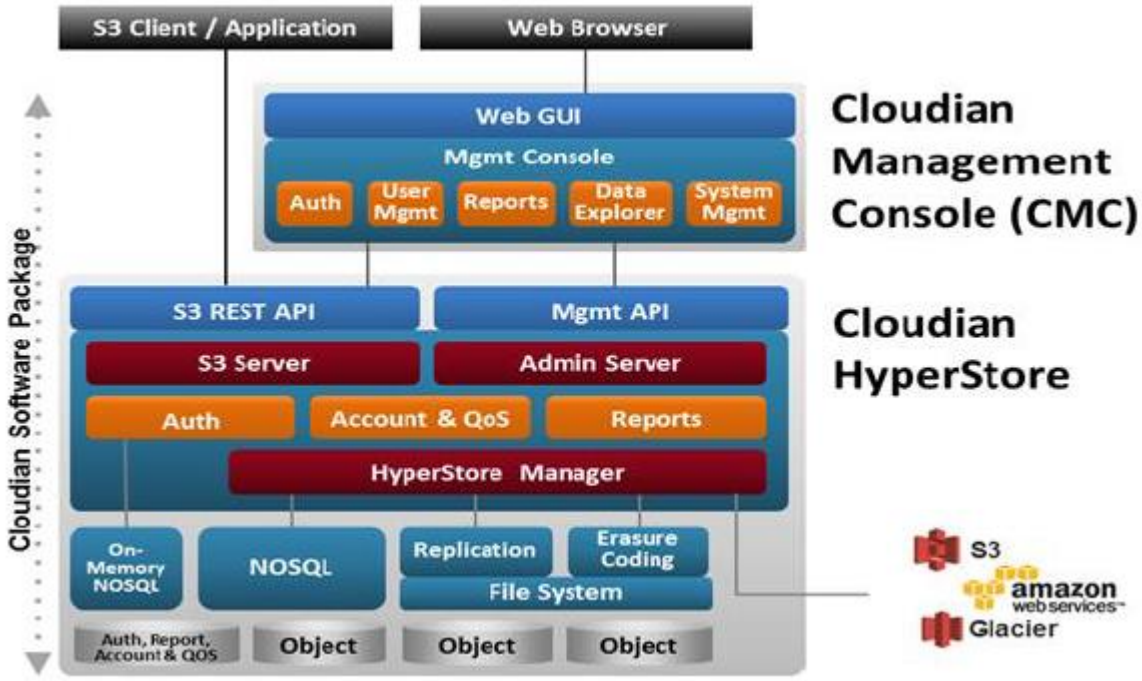

Figure 1: Cloudian HyperStore Architecture (source: Cloudian)

Access to data over the Ethernet interface is using a REST API with the S3 protocol. File access is through NAS appliances called Cloudian HyperFile that uses the Cloudian HyperStore as the backend storage. Cloudian features a full implementation of the S3 protocol in use for AWS. The following diagram illustrates S3 operations.

Operations on the Service	Operations on the Buckets	Operations on the Objects	Additional Features
Get	Get Bucket (List Objects)	Get Object	Initiate Multipart Upload
	Put Bucket	Put Object	Upload Part
	Delete Bucket	Delete Object	Upload Part - Copy
	Head Bucket	Head Object	Complete Mulitpart Upload
	Get Bucket Website	Get Object ACL	Abort Multipart Upload
	Put Bucket ACL	Put Object ACL	List Part
	Put Bucket Website	Put Object (Copy)	
	Delete Bucket Website	Delete Multiple Objects	
	List Mulipart Uploads	Options Object	
	Get Bucket ACL	Post Object	
	Get Bucket Lifecycle	Get Object Torrent	
	Get Bucket Policy	Post Object Restore	
	Get Bucket Tagging		
	Put Bucket Lifecycle		
	Delete Bucket Lifecycle		
	Delete Bucket Policy		
	Get Bucket CORS		
	Get Bucket Location		
	Get Bucket Logging		
	Get Bucket Notification		
	Get Bucket requestPayment		
	Get Bucket Versioning		
	Put Bucket Policy		
	Get Bucket Object Versions		
	Put Bucket CORS		
	Put Bucket Versioning		
	Put Bucket Logging		
	Put Bucket Notification		
	Put Bucket Tagging		
	Put Bucket requestPayment		
	Delete Bucket CORS		
	Delete Bucket Tagging		

Figure 2: S3 Protocol Operations Example

Software Architecture

The Cloudian HyperStore software executes on a server with a Linux OS. Object data is stored in EXT4 files on the nodes. Services executing in the nodes perform the primary functions:

- S3 Service – Processes S3 requests from hosts according to the protocol operations.
- HyperStore Service and HyperStore FileSystem – The HyperStore service exploits Cassandra software to perform the internal data partitioning, automatic replication of data, data redistribution in a cluster on expansion, quorum calculations, selection of nodes for data storage, and other functions. The Cassandra database also is used to store small S3 data objects. Larger objects are stored in the Linux EXT4 filesystem, which is called the HyperStore FileSystem.
- Cassandra Database Services – Information in the database not only includes the "map" of where data is distributed and small S3 objects, user metadata and system defined metadata is also stored for every object. System metadata includes:
  - Creation time
  - Last modified time
  - Last accessed time
  - Size of object
  - Access Control List
  - Version
  - Public URL if assigned
  - Compression type
  - Encryption key
  - Auto-tiering state
- Redis Database Services – Two Redis databases are used, one to store credentials for access to objects and another for QoS (Quality of Service) settings at the user or group level that have been set by an administrator.
- Administrative Service – Administrative actions are accessed through an API to the administrative service. All control setting and management functions through the administrative service. The Cloudian Management Console uses the administrative service for all of the management functions it provides.
- Support Services – Other support services:
  - Cloudian Monitoring Agent – Monitors node health and performance and has threshold notifications.
  - Cloudian Monitoring Data Collector – Collects data from Monitoring Agents on nodes within a defined region and stores in Cassandra database for reporting.

Consistency is configurable with the Cloudian system. The default is strong consistency where data must be written to a quorum where a failure will not result in loss of data before

acknowledging complete. Eventual consistency may be set where the quorum does not have to be completed before acknowledging, resulting in improved performance.

Data Protection and Security

Device and Node Failure Protection

As explained earlier, administrators can select device and node protection by choosing on a per bucket basis to either replicate data to three different nodes or use error correction with erasure codes to distribute data across different nodes. These are termed a 'storage policy.' There are three Information Dispersal Algorithm (IDA) erasure-coding configurations supported:

- 2+1 – For each object, there are two data fragments with one parity fragment
- 4+2 – For each object, there are four data fragments with two parity fragments
- 9+3 – For each object, there are nine data fragments with three parity fragments.

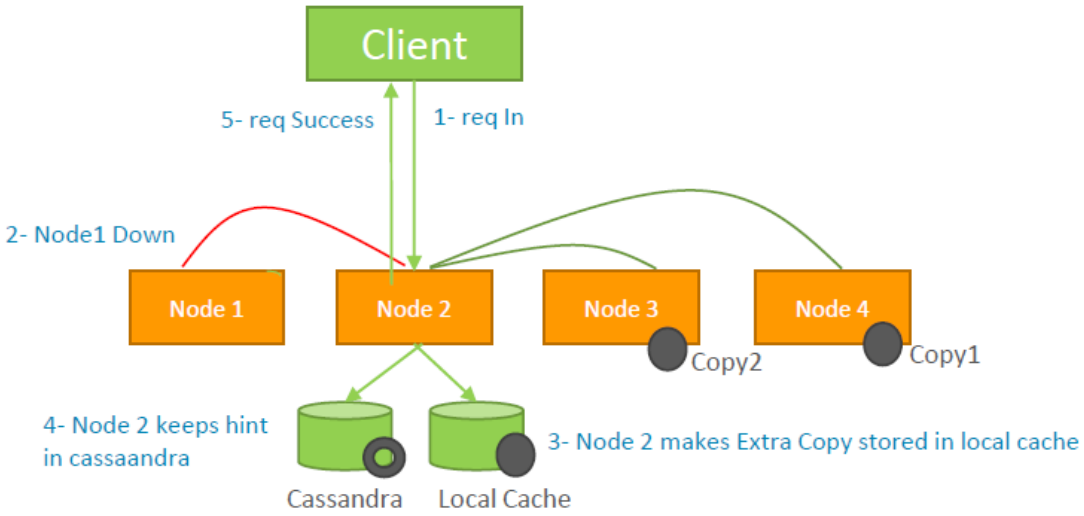

Figure 3: Cloudian Data Protection (source: Cloudian)

Remote Protection

Geographic dispersion is supported as well as asynchronous remote replication.

Data Integrity

Data integrity is assured with two methods. One is 'repair on read,' which is checking the added integrity check data on a read operation on the reconstructed data or on all replicas depending on the storage policy. Repairs are made with the protection data whenever an error is detected. The other method is for a background scan of all data where the integrity check data is evaluated against the data and corrections made where necessary. The frequency of the scan is a settable parameter.

Multi-Tenancy and Regions

Identity and Access Management allow administrators to manage by users and groups, providing effective isolation. The same isolation by use of credentials is used for Quality of Service and Billing controls. Multiple regions may be supported with shared multi-tenancy for geographic optimization. Data can be placed in specific regions for security or other reasons.

Encryption

Encryption is handled at either the bucket or object level using transparent key management done at the node level. AES-256 server-side encryption is used to encrypt data at rest. SSL is used to encrypt data in transit.

Advanced Features

The advanced features of Cloudian HyperStore add value that may be critical for certain uses. The advanced features for performance are included in the performance section of this document.

Versioning

Objects can be updated without the prior version being deleted or overwritten with the new one. Multiple versions are supported with access allowed to the prior versions as a subset to the object ID. Version information is kept in the per-object metadata.

Quotas

Per user quotas can be set on either capacity or number of objects or both.

WORM

WORM mode can be set on a per-object basis by restricting deletion. WORM mode is the default when versioning is used. WORM can also be set for filesystems with HyperFile.

QoS Controls

User and group level Quality of Service settings are supported. User QoS places upper limits on service usage by individual users. Group QoS setting places upper limits on the aggregate

service usage of users within the group. Enforcement of QoS is accomplished by rejecting S3 requests that would exceed the limits. Service usage can be defined by:

- Storage quota by number of kilobytes or by number of objects
- Peak HTTP request rate – number of requests per minute
- Peak data upload or download in kilobytes per minute

Integrated Billing Management and Monitoring

Usage reporting is done on a system-wide, group-wide, or individual user basis. Billing can be based on capacity used or usage rates – transfers in or out. Reporting and monitoring information is available through the GUI or accessible through the API.

Self-Service Portal

Role-based access is supported through the GUI, allowing users controls over provisioning, reporting, billing, QoS, and other settings allowed by the administrator.

Auto-Tiering

Cloudian Auto-Tiering is automatic movement of objects from HyperStore to another S3 system on a predefined schedule. Supported targets are Amazon S3, Amazon Glacier, or another HyperStore system. Auto-tiering is configured on a per-bucket basis by an administrator or by a user if granted permission.

Data Reduction

Cloudian supports three types of data compression: LZ4 (Lempel-Ziv), snapp and zlib. The different methods can have slightly different effectiveness. LZ4 is the most commonly used compression algorithm. Compression is performed as the data is written to the storage nodes and decompressed upon access.

Container Support

HyperStore supports a Kubernetes S3 operator to provide persistent storage for container environments. The Kubernetes S3 operator allows dynamic provisioning of object storage for container applications.

HyperFile NAS – File Access

For file access using NFSv3, FTP, and CIFS/SMB 2.1 (SAMBA) protocol, Cloudian provides a NAS dual controller system that uses the HyperStore for storage of file data. The HyperFile NAS controllers are in an active/passive configuration for high availability. HyperFile supports scale out with multiple controller pairs and support a global namespace across the HyperFile systems.

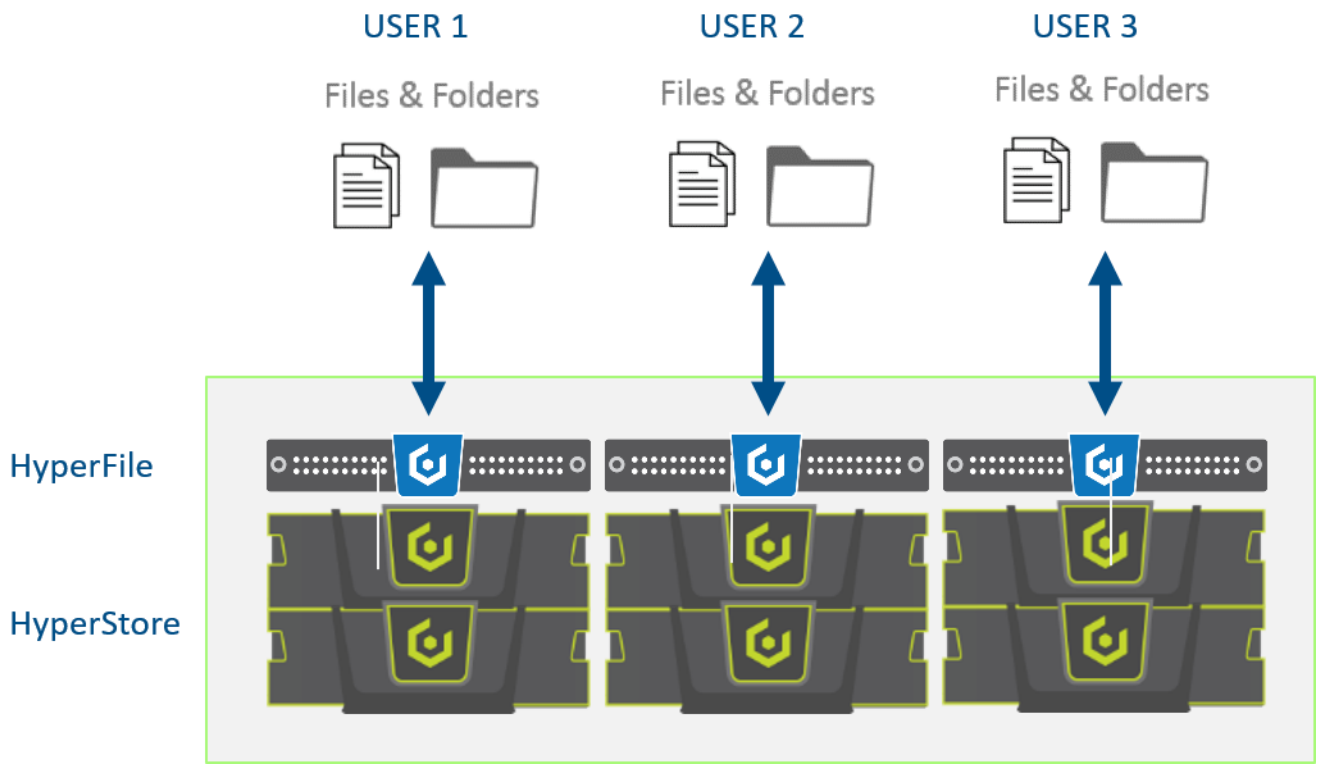

Figure 4: Cloudian HyperFile (Source: Cloudian)

There are two modules, one being Access Point which acts a NAS controller and the other optional component, Global View Manager.

File Migration Engine

HyperFile includes software for migrating data from existing NAS systems. The file migration software will index the mounted file system on the donor NAS system and then assume the IP addresses. Writes will be sent through to the donor NAS system while pull the data from the donor systems. The following diagram illustrates the operation.

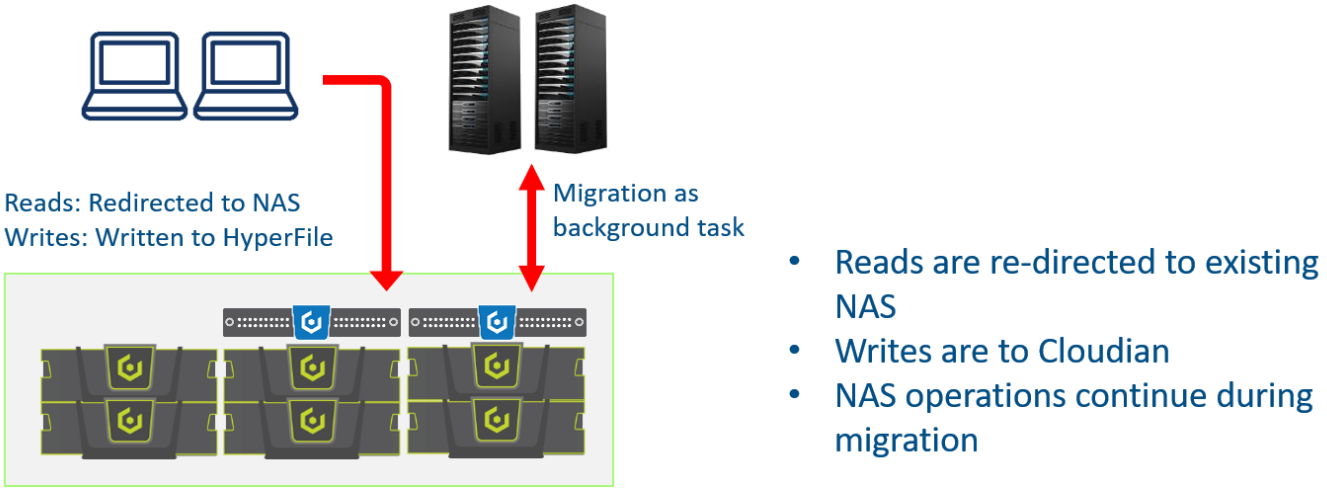

Figure 5: HyperFile File Migration (source: Cloudian)

Features of Connect for HyperFile

A number of advanced features are offered with Cloudian HyperFile:

- WORM
- AD and LDAP support
- Global namespace
- HyperStore data protection
- HyperFile GUI and CLI
- POSIX compliant

Significant Announcements

- Dec 2023 – start with announcement coverage

Futurum Group EvaluScale – Object Storage

The Futurum Group product review methodology "EvaluScale" assesses each product within a specific technology area. The evaluation of each product is based on its capabilities, with capabilities for each technology segment grouped into distinct categories. The products are evaluated based on the following 4 criteria categories:

- Performance / Capacity
- Basic Functionality
- Advanced Capabilities

- Ability to Execute

The full Object Storage EvaluScale can be found here.

The Futurum Group Opinion and Outlook for Cloudian HyperStore

Cloudian HyperStore has the opportunity to be an inexpensive entry-level object storage platform with the ability scale large either with a software only implementation or to the maximum of the appliances sold. There are several appliance models offered. Build your own implementations can license the software for customer implementations. The ability to mix on-premises nodes with nodes configured on any public cloud is unique and adds flexibility. Nodes can be only in cloud locations as well with a minimum of three nodes.

Cloudian has many advanced features with the exception of defined regulatory compliance such as SEC 17a-4. Unique for Cloudian is their file system software that is deployed on physical or virtual appliances that utilizes the S3 object storage. The file system software is very advanced with features such as file migration, global namespace as part of the global management console, object versioning, and WORM.