Deploying Hybrid Storage with Cloudian and Google Cloud Storage

By Jon Toor, CMO at Cloudian

This article discusses integrating Cloudian and Google Cloud Platform (GCP) to store data locally on Cloudian and tier data to GCP using configurable tiering settings.

Cloudian HyperStore is petabyte-scalable, on-premises object storage for unstructured data. Cloudian storage resides as an appliance in the data center and provides an on-premises storage solution for data center applications that require local access to information.

A hybrid configuration of Cloudian with GCP combines the fast performance of local disk-based storage with the capacity scalability and cost of cloud storage.

Typical use-cases for using Cloudian with Cloud Storage include:

- **Data tiering for performance**: Provide local access for frequently accessed media assets to maximize performance, while tiering older, less-frequently accessed data to GCP.

- **Data tiering for archive**: Use integrated, policy-based migration tools to migrate data from Cloudian to GCP for long-term archive.

- **Capacity expansion**: Migrate data from local storage to GCP to facilitate capacity expansion; both on-premises and GCP storage can be managed within a single namespace.

- **Data distribution**: Store data locally (for fast data access), then migrate to GCP for content distribution using cloud-based applications.

    Examples of solutions built with Cloudian include:

- **Video / surveillance:** Retention of security or monitoring video with rich metadata tagging to facilitate rapid search.

- **Backup/archive:** Maintain recent backups on-premises for rapid recovery, while tiering older data to the cloud for disaster recovery (DR) purposes. Compatible with most popular backup solutions, including Veritas, Commvault, Veeam, Arcserve, and Rubrik.

- **Media and Entertainment:** Use Cloudian storage on-premises for active archive as part of the media workflow, migrate data to GCP for long-term archive or content distribution.
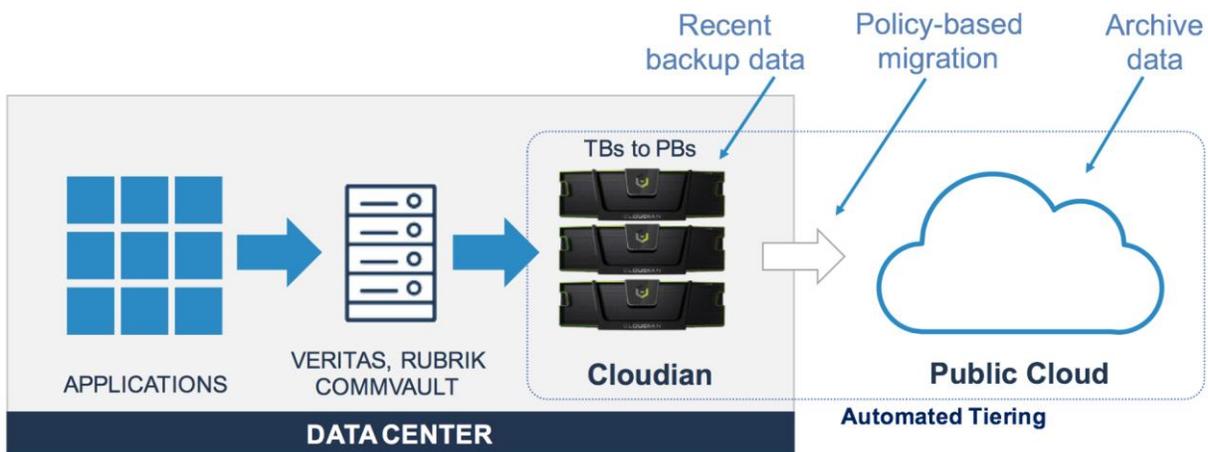
Cloudian provides on-premises storage, scalable from TBs to hundreds of PBs. Integrated tools enable policy-based data migration to GCP for backup, archive, or capacity expansion. When data is migrated, metadata is retained on premises for rapid search.

Cloudian Hyperstore Features

Cloudian HyperStore incorporates a suite of tools including system management, monitoring, and reporting. Key features include:

- **Cloud Storage API**: Full API compliance for investment protection and application interoperability.
- **Multi-tenancy**: Accommodate multiple users with isolated storage domains.
- **QoS**: Manage SLAs with Cloudian's granular QoS capability.
- **Billing**: Manage chargebacks using configurable metrics.
- **Erasure coding**: Configurable data protection.
- **Compression**: Integrated data compression.
- **Encryption**: Integrated encryption helps ensure security for data at rest.
- **Works with objects and files**: Consolidates object storage and backup/archival file storage in one pool.
- **No single point of failure**: Fully distributed, peer-to-peer architecture
- **Scale on demand**: Scales from TBs to 100s of PBs and/or multiple data centers.
- **Deployment options**: Deploy as appliances or software on industry standard servers.

Distributed Data

Collaboration or remote backup are simplified with Cloudian's distributed data capabilities. Each HyperStore software implementation starts with three or more distributed nodes. Objects are then replicated or erasure-coded across the available nodes for data durability and availability.

Administrators can configure the number of replicas or define the erasure-code strategy required to meet SLA and cost objectives. In addition, administrators can add policies for data to be tiered to Cloud Storage. Reads and writes are always performed at the local data center with remote replication performed in the background to avoid latency of remote writes.

Security

HyperStore simplifies the data encryption process by providing transparent key management, using AES-256 server-side encryption for data stored at rest, and supporting SSL encryption for data in transit. And with Cloud Storage API-compatible object-level ACLs, system administrators can secure buckets and objects with either no access, read-only, or read-write permissions for everyone or for named users and groups.

Compression

HyperStore seamlessly uses data compression to reduce storage and network consumption significantly while accelerating data replication speeds and reducing network bandwidth requirements. With less data to store on disk and less data to move over the network, businesses can get more life out of their existing storage and network investments, further improving their ROI and lowering their TCO. HyperStore offers three different types of data compression technology: lz4, snappy, and zlib.

Multi-tenancy

HyperStore supports multi-tenancy, where each account is logically segmented and data for accounts is only accessible by account users and the group administrator. Advanced identity and access management features allow system administrators to provision and manage groups and users in each account, to define specific classes of service for groups and users, and to configure billing and charge-back policies. Both administrators and users benefit from reporting options and account and data management capabilities.

File Access Support

Cloudian HyperStore Connect for Files allows enterprises to offer scalable file services on top of Cloudian HyperStore object storage using industry-standard protocols such as NFS, CIFS, and FTP. HyperStore's file system integration supports user access control with Active Directory and LDAP, version control for deleted files, and multi-threaded parallel data access.

Use Cases

Cloudian customers are building a number of solutions using HyperStore, including the following.

Backup and Archive

Cloudian object storage offers an alternative to conventional disk and tape. This provides significant TCO savings compared to other disk-based solutions, and with far greater speed and convenience than tape, HyperStore makes the ideal backup target for large capacity environments. For details, see the Cloudian Backup Solution PDF.

APPLICATIONS   BACKUP SERVERS   ONSITE BACKUP   ARCHIVE

Veritas, Rubrik Commvault

Compatible with backup solutions such as Veritas, Commvault, Veeam, Arcserve, and Rubrik, Cloudian HyperStore offers petabyte-scalable, on-premises storage that can be implemented as a "bolt on" solution, eliminating the need to overhaul the existing storage architecture. The ROI of Cloudian object storage as a backup target is discussed in this report.

Multiple backup options are available to meet specific data protection objectives. You can configure a hybrid cloud and tier a portion of data to Cloud Storage for archive or data protection. Or you can locate Cloudian nodes at different sites to automatically provide off site backup and recovery capabilities.

Hybrid Cloud

The following illustration shows policy-based migration to GCP. When data is migrated, metadata is retained on premises for rapid search:
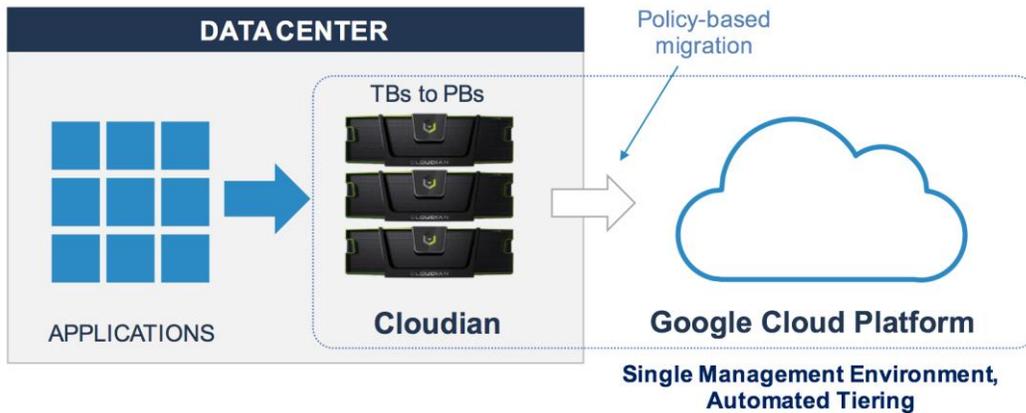

A hybrid cloud allows both both on-premises and cloud-based storage pools to be managed as one. With automated tiering, information stored to Cloudian can be selectively migrated to GCP based on data policies such as file type, frequency of access, file size, or specific metadata parameters. Both the on-premises and cloud storage can be managed as a single, limitlessly scalable storage pool.

To enhance data accessibility, Cloudian maintains a copy of all metadata on-premises, allowing the search all data in an instant. Retrieve only the information required.

Big Data Analytics

Hadoop analytics can be run directly on HyperStore software and appliances. This in-place analytics capability lets you derive meaningful business intelligence from data. Automated tiering allows data to be migrated to GCP for backup, archive, or capacity expansion.
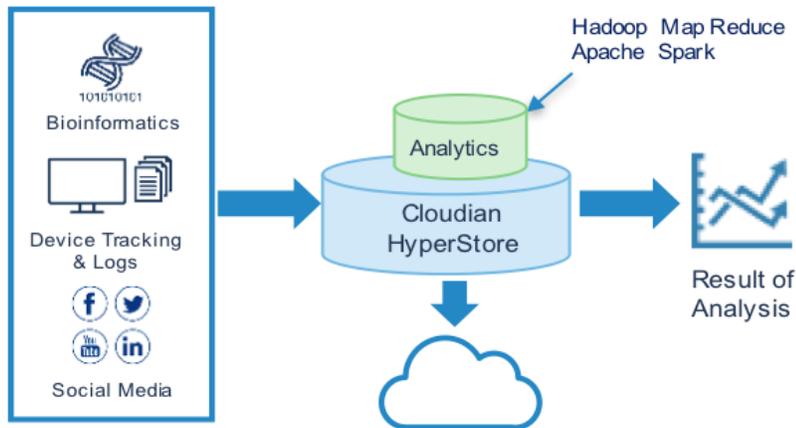
Cloudian HyperStore can emulate HDFS storage for Hadoop and Spark workloads, which allows compute and storage to scale independently in large environments. With Cloudian, you can efficiently store blocks of any size from 4 KB to multiple TB, and you can reduce storage footprint with integrated erasure coding and compression.

Features such as encryption protect data at rest, while TLS can help secure data in flight. For details, see this reference architecture PDF.

Storage-as-a-Service

Enterprises can offer storage as a service by combining on-site storage with GCP services for effectively limitless storage capacity and a superior cost structure. Integrated data tiering allows selective data migration to GCP for backup, archiving, or capacity expansion.
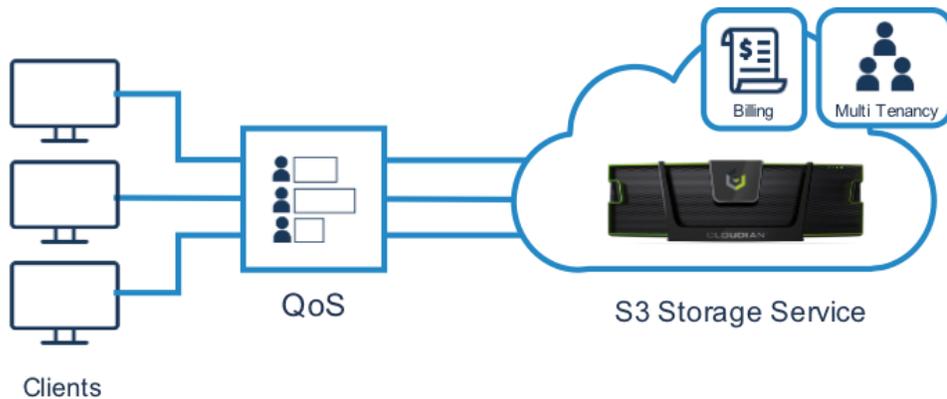
Cloudian employs a shared-nothing, peer-to-peer architecture, so adding capacity is non-disruptive. Performance scales linearly with added nodes and failure domains are limited to a single node. Choose from data protection options and configure specific users accounts for the data durability they choose including the ability to lose a node or even an entire site without data loss or service disruption.

Multi-tenancy is built-in and enforced with robust QoS tools. Self-service provisioning is made simple through an end-user GUI, while policy generation and billing are automated based on administrator settings.

Deploying Cloudian Hyperstore

Cloudian is deployed on premises in your data center in one of two configurations:

- **Deploy as an appliance**: Cloudian is available as a fully-configured hardware/software appliance. Available in a range of capacities beginning at 24 TB, each appliance acts as a single node, with a typical starting configuration of three nodes. Additional nodes can be added to the cluster to grow both capacity and storage bandwidth. Performance scales linearly with added nodes.

- **Deploy as software on your own servers**: Cloudian is also available as software that runs on your own servers and storage. The software can be deployed on an X86 server or on a VM. Standard hard drives provide the storage capacity.

Data is stored to the Cloudian object storage environment through any application that can store data to GCP. Examples include applications from Adobe, Computer Associates, Commvault, Citrix, HortonWorks, IBM, Veritas, Red Hat, and others.

Cloudian can also be used as a repository for file-based data. Standard SMB, NFS, and FTP file types can be stored using available software or hardware solutions.

Data Tiering to GCP

In a hybrid cloud configuration, Cloudian provides an onsite data repository with policy-based data migration to GCP. The combined on-premises-plus-cloud environment employs a single namespace, effectively delivering a unified management environment. Migration is policy-driven, based on rules such as data age, frequency of use, file size, and file type. When data is migrated to the cloud, a copy of the metadata is retained on-premises, enabling rapid search of all data, both on-premises and in the cloud.

For any data storage system, granularity of control and management is extremely important, because different data sets have varying management requirements. It is often necessary to apply different Service Level Agreements (SLAs) as appropriate to the value of the data to an organisation.

Cloudian HyperStore manages data tiering through lifecycle policies, as shown in this screenshot:

Bucket Lifecycle policies specify rules and destinations for tiering.

Auto-tiering is configurable on a per-bucket basis, with each bucket allowed different lifecycle policies based upon:

- Which data objects the lifecycle rules apply to, for example:
- All objects in the bucket
- Objects for which the name starts with a specific prefix (such as prefix "Meetings/2015/")
- The tiering schedule, which can be specified using one of three methods:
- Move objects X number of days after they're created
- Move objects if they go X number of days without being accessed
- Move objects on a fixed date — such as December 31, 2016

When a data object becomes a candidate for tiering, a small stub object is retained on the Cloudian cluster acting as a pointer to the actual data object, so the data object still appears that it's stored in the local cluster. To the end user there is no change to the action of accessing data, but the object does display a special icon denoting the fact that the data object has been moved.

For auto-tiering to GCP, a GCP account is required along with associated GCP account-access credentials. After they've been auto-tiered to GCP, the objects can be accessed either directly through GCP (using the applicable GCP credentials) or through the local Cloudian system.

Next Steps

Here are some resources to get you started on your journey to object storage. Peruse these for information, or visit the [Cloudian website](#):

- [Free Cloudian Trial](#)
- [Google Cloud Platform Solution Brief](#)
- [Data Protection Solution Brief](#)
- [5 Reasons to Deploy a Hybrid Cloud](#)
- [White Paper Collection](#) Try out other Google Cloud Platform features for yourself. Have a look at our [tutorials](#).

Last updated March 27, 2017.