CLOUDIAN®

Object Storage Buyer's Guide

The technical and financial benefits of object storage and how to evaluate the technology for your environment.

Table of Contents

Introduction

The data revolution is upon us. Organizations in every industry generate exponentially more volumes of unstructured data than ever before. They also retain, re-use, and learn from that data to a far greater extent. Add to this the emergence of cloud and IoT, and it becomes apparent that unlimited, affordable storage scalability is an increasingly important component of a company's long-term success.

Today's massive amounts of data are now generated and stored on-premise, at the edge, and in the cloud, greatly increasing storage network complexity. To meet the needs of this transformation, the way we store and manage unstructured data must change.

That's where hyperscale object storage comes in.

In this ebook you'll learn about today's storage challenges and how object storage provides unique capabilities to address them. You'll acquire a better sense of how object storage works, and how it can fundamentally redefine data management at scale. Ultimately, you'll gain peace of mind from knowing the value of object storage for your organization.

Armed with this information, you'll be much better equipped to take advantage of the data revolution – and reap maximum ROI from your storage investment.

1: The Storage Challenge

Each succeeding generation of computing creates greater and greater volumes of data. In 1986 the world's combined data totaled 2.6 exabytes (EB). In 2016, 16 zettabytes (ZB) of data was generated worldwide (1 ZB = 1,000 EB). By 2025 the amount of usable business data will soar to 163 ZB, according to market research firm IDC.

IDC

**2025
163 ZB**

160 ZB

100 ZB

**1986
2.6 EB**

**2016
16.1 ZB**

20 ZB

1986   1996   2006   2016   2026
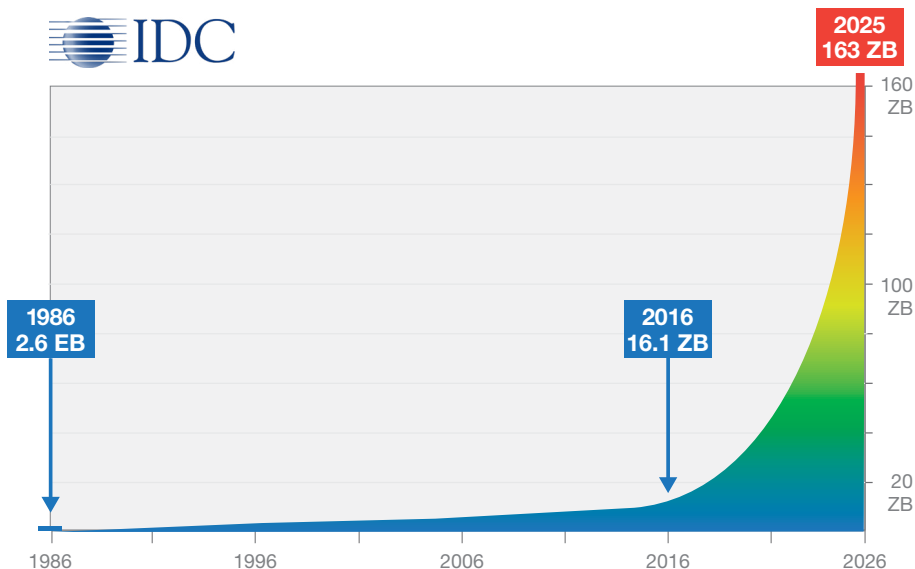
IDC predicts the total amount of data generated worldwide will skyrocket to more than 163 zettabytes by 2025.

Unstructured data represents the vast majority of this information – 80%, by most accounts.

Scale Capacity Without Complexity
As a result, capacity requirements for unstructured data are growing more than 50% year over year. No wonder that **CAPACITY GROWTH is the No. 1 challenge** facing most organizations that generate and use vast amounts of data.

Rapid Data Search
Users across every industry increasingly demand that data be easily searchable and instantly accessible. This desire for speed is a direct consequence of our growing dependence on the internet. When the world's data is instantly available on the web, users expect their own data to be just as searchable and accessible.

Instant Data Access
Businesses today recognize the value of data across all phases of their operation, including archived information. In diverse use cases such as healthcare,

security, and media and entertainment, unstructured data has enduring value as we find new ways to analyze and re-purpose it. For instance:

- Broadcasters now re-use media files multiple times as their content is formatted for distribution in broadcast, mobile streaming, and web versions.
- Medical firms re-analyze archived records, looking for predictive insights.
- Manufacturers examine machine sensor data, trying to anticipate failures before they occur.
- Vehicle makers now collect millions of miles of driving data, endeavoring to train autonomous cars and trucks.

All of this data has tremendous value, but only if it can be economically stored, efficiently analyzed, and instantly searched. Traditional storage lacks the needed economics and scale to achieve these tasks, and tape media lacks the required access time and search capability. Meeting these objectives is attainable only with object storage.

FASTFACT Unstructured data, also referred to as "file data," is any information not managed by a database application, including high-resolution digital media, healthcare records, and engineering files. It accounts for about 80% of all capacity generated today.

Hyperscale object storage provides the most scalable, most economical solution for meeting storage challenges

Object storage:

- Strips complexity from the data center
- Addresses an organization's demands for global data search
- Ensures data durability
- Accommodates new requirements (such as GDPR) regulating where data is stored and how it is managed.

2: Solving the Challenge with Object Storage

Data center managers, content creators, broadcasters, researchers, healthcare providers, and software developers all require storage solutions that help them contend with the explosive growth in unstructured data.

Object storage offers unique capabilities to meet these needs. Far from being an incremental improvement over earlier technologies, it is a fundamentally different way of storing data. Here's why:

Exabyte Scalable

Modular design for non-disruptive growth
Traditional storage systems were designed with an upper limit on capacity. Object storage solves the scale problem with an architecture that eliminates this limitation. But this exabyte-plus scale is only part of the solution. How easily the system grows is just as critical as how big it can get. With object storage, scaling is non-disruptive, so you can add capacity when needed.

Files and Objects Together

Consolidated storage for unstructured data
Files remain the most common means of managing unstructured data. Object storage lets you combine file and object data in a single pool. Unlike NAS, in which the file hierarchy limits growth, object storage stores files in a flat file system that can expand forever.

Geo-Distribution

One storage pool that spans the globe
With the advent of IoT, remote sensing technologies, and low-cost 4K cameras, high-capacity data is now created everywhere. This paradigm shift places new demands on networking and storage technologies. Object storage addresses this challenge with a distributed system in which nodes may be deployed wherever needed. Low-cost, remote storage lets analysis happen where the data is collected, rather than having to load the network with raw information. Whether providing local storage for remote applications or disaster recover (DR) capability across sites, object storage possesses the needed data management capability to perform these tasks simply and economically.

Cloud Integration

Hybrid cloud and multi-cloud ready
Most organizations today plan to use both cloud and on-prem storage. Analysts predict continued rapid growth for both storage types. Object storage employs the S3 API, so it speaks the language of the cloud. It also incorporates data management features that simplify data placement, cloud and on-prem storage become two parts of a single global namespace.

The Benefits of Object Storage

| Scalability | Searchability | Geo-Distribution | Rapid Access | Simplicity | Data Durability | Data Protection | Compliance |

3: How Does Object Storage Work?

Object storage combines multiple technologies in a single, integrated system to solve the capacity data management challenge. Here's a quick look at the key elements.

Flat File System

Every storage system needs a way to index, or locate, information so it can be managed and retrieved. With traditional SAN and NAS technologies, these indexing schemes have built-in limits. NAS file systems use a hierarchy, like branches on a tree. SAN uses direct addressing, like postal addresses. Both have scale limits. Furthermore, with these systems, performance degradation may slow growth before the theoretical limits are reached.

Object storage, on the other hand, employs a flat file structure that has no limits. An "object" includes:

- User data (usually a file
- A unique ID that is created from the data itself
- User-defined metadata, which can be used to describe that object's content

Objects are then organized in "buckets," which are analogous to folders for files. These containers can hold similar or related objects, allowing them to be managed as groups.

Clustered Nodes

Object storage is always a clustered system, never a single device. Any "node" in the cluster can see and retrieve any data. The cluster can be expanded simply by adding more nodes. The nodes themselves are stand-alone devices — typically industry-standard servers running a software-defined storage application — interconnected via either LAN or WAN. Each has a metadata catalog, knows where in the cluster data is stored, and can respond independently to data requests. For enhanced performance, nodes can also work in parallel to accelerate the delivery of large objects.

Retrieving Objects

To retrieve data from an object store, you simply ask for it by its object ID. Objects may be local or at other sites, but because they are in a flat address space, they are retrieved in exactly the same way.

How Does Object Storage Solve the Storage Capacity Problem?

To visualize how object storage works, first visualize traditional storage as a parking lot. That lot has a limited number of spaces. When it's full, you need to find a new parking lot. Furthermore, as it fills up, it will take you longer to find an empty space.

Now visualize object storage as valet parking. You leave your car with the attendant who gives you a ticket, a unique ID. He is then free to park your car anywhere space is available. He keeps a record of where your car is parked, and is therefore not limited to the space in a single lot.

Object vs File Storage: What's the Difference?

A file is information written in a specific format. That format is known to an application, and is identified by the file's suffix. A ".jpg" file format, for example, would be usable by an application that manages images.

An object, on the other hand, is a package that contains information (perhaps a file or a part of a file). There is also metadata that describes the user data. A unique ID identifies the object. The object can be divided into "slices" to be distributed across multiple nodes.

Unlike files, an object can include a large amount of metadata, which enhances search capabilities using sophisticated Google-like tools.

	OBJECT STORAGE	FILE STORAGE
PERFORMANCE	Performs best for big content and high stream throughput	Performs best for smaller files
GEOGRAPHY	Data can be stored across multiple regions	Data typically stored locally
SCALABILITY	Can scale infinitely to exabytes and beyond	Operational limits reached at a few PBs
ANALYTICS	Customizable metadata allows data to be easily organized and retrieved	Limited number of set metadata tags

8

How Does Object Storage Work? (continued)

Data Protection

To guard against component failure, such as a failed hard drive, all storage systems employ data protection. With object storage, the built-in failure protection can do much more. It can guard against whole-device failure, rack failure, or even site failure.

When data is written to object storage, it is protected with either **ERASURE CODING**, or **DATA REPLICATION**, or both. Erasure coding breaks the data into smaller segments ("slices") and writes them to multiple nodes within the cluster. It also writes additional data segments to other nodes for data durability purposes, ensuring that the object can be restored even if nodes fail.

Nodes may be located at one data center, or distributed across multiple data centers to protect from site failure. Configurations always allow for multiple node failures to ensure data durability.

The second option, replication, writes identical copies on multiple nodes to ensure availability. Both erasure coding and replication can operate across sites for DR purposes.
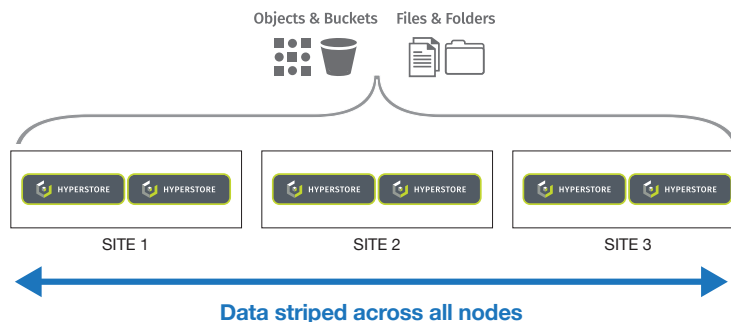
Data Replication

Writes identical copies on multiple nodes, and can operate across multiple sites.

REMOTE SITE

PUBLIC CLOUD
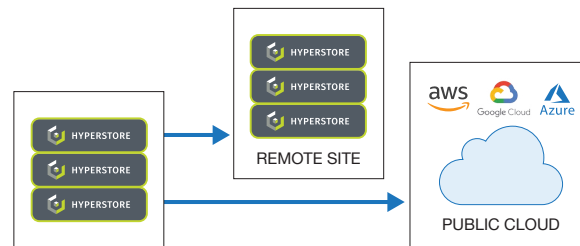
Erasure Coding

Writes to multiple nodes within the cluster; also writes to nodes outside the cluster to allow for multiple node failues.

Objects & Buckets   Files & Folders

SITE 1   SITE 2   SITE 3

Data striped across all nodes

The S3 API

The Amazon S3 Application Programming Interface, better known as the S3 API, is the most common way in which data is stored, managed, and retrieved by object stores. Originally created for the Amazon S3 Simple Storage Service, the widely adopted S3 API is now the de facto standard for object storage.

Not All S3 APIs Are Equal

Compared with established file protocols such as NFS, the S3 API is relatively new and rapidly evolving. Among object storage vendors, S3 API compliance varies from below 50% to over 90%. This difference becomes material when an application — or an updated version of that app—fails due to S3 API incompatibility.

Cloudian is the only object storage solution to exclusively support the S3 API. Launched in 2011, Cloudian's many years of S3 API development translate to the industry's highest level of compliance.

Employing the S3 API makes an object storage solution flexible and powerful for three reasons:

Standardization
With Cloudian, any object written using the S3 API can be used by other S3-enabled applications and object storage solutions; the existing code works out of the box.

Maturity
The S3 API provides a wide variety of features that meet virtually every need for an object store.

Simplicity
End users planning to deploy object stores can access the plentiful resources of the S3 community — both individuals and companies.

The Difference Maker: Metadata + Search

User-defined rich metadata tags are another key attribute of object storage. Each object includes one or more tags with system metadata (such as creation date), and user metadata that can be used to describe the object's contents. To find information, Google-like tools let you search the metadata for specific attributes.

User metadata can include information about the content, such as:

• The location where an asset was created
• The project the object was created for
• The specific subject of the data

Search tools like Elasticsearch and Kibana let you both search metadata and also create graphical views of data to categorize information and more easily spot trends.

4: Object Storage and the Cloud

Large-scale object storage adoption first occurred in the cloud, and today all major public clouds are built on the technology. Nearly all the web services you use every day – including Facebook, Netflix, and Google – rely on it. Now, the same technology is available for use in your data center: the same APIs and the same limitless scalability.

Cloud integration is built-in with object storage, so public and private clouds can be merged into a single storage pool where the public cloud acts as another storage tier, giving you new options for data management. Here are several ways that on-prem object storage and the public cloud work together to solve problems.

S3-Compatible Storage Services

With object storage, it's easy to launch S3-compatible storage services. Many managed service providers (MSPs) and enterprises today offer S3-compatible services in either public or private cloud service models. Cloudian incorporates features that simplify this with multi-tenancy, quality of service controls, billing, and access controls.

Hybrid Cloud: Ideal for DR

A hybrid cloud lets you manage both public and private storage pools as one, with data management tools that manage policy-based tiering or replication among environments. Migrate or replicate data based on file type, frequency of access, file size, or other parameters of your choice.

Hybrid cloud use cases vary. DR is easy to manage and cost-effective with policy-based replication to the cloud. For capacity expansion, data tiering can effectively provide unlimited capacity. Policies maintain the most commonly used data on-prem, while infrequently used data is moved to the cloud.

Two other use cases, data analysis in the cloud and content distribution, let you capitalize on the compute capability and geographic reach of the cloud, while also letting you access data locally for performance.
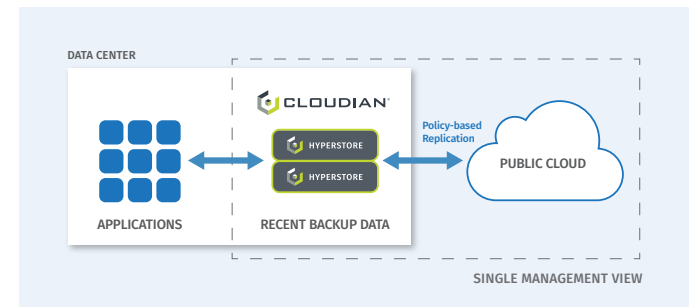
Multi-Cloud: Multiple Cloud Vendors in One Storage Pool

Multi-cloud lets you merge clouds from multiple vendors in a single management environment. Combine storage from Amazon, Google, and Microsoft, plus private cloud storage, to a single pool with one set of management APIs.

With multi-cloud, your organization can use different clouds for different reasons, letting you take advantage of various features or data center locations. It also allows you to use clouod resources selectivity for cloudbursting or other events that call for a temporary use of more storage resources than usual.
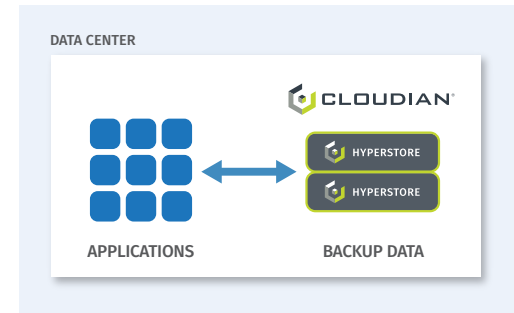
S3-Compatible Storage Services

Hybrid Cloud: Ideal for DR
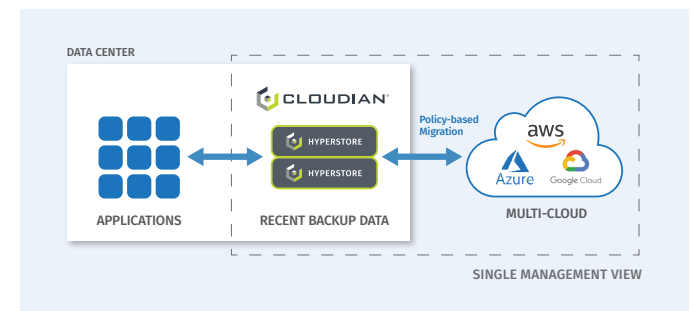
Multi-Cloud: Multiple Vendors in One Storage Pool

5: The Financial Benefits of Object Storage

Cost-effective acquisition and operation is designed into object storage.

Built specifically for large-scale data management, object storage delivers the lowest CAPEX and OPEX of any enterprise storage system. And unlike traditional enterprise storage where acquisition costs — measured in cost per terabyte — tend to increase with scale, object storage systems become more efficient and less costly with scale.
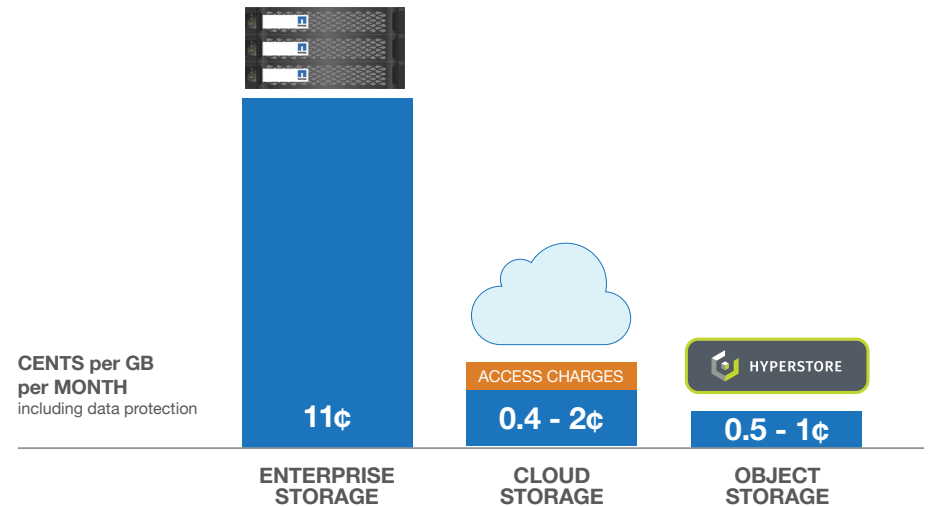
CAPEX Savings

Object storage CAPEX benefits often reach 70% savings vs. traditional tier 1 storage. One reason is that it functions on industry-standard hardware. This attribute eliminates the need for proprietary platforms, which keeps both acquisition and maintenance costs low. As your system grows, the open systems model ensures that your costs always remain in line with the industry's best pricing.

Deployment Flexibility
With the choice to deploy as a VM, as bare-metal on the server of your choice, or as vendor-supplied hardware appliances, object storage can be installed in whatever manner makes the most sense for your organization.

Modular Scaling
An object storage cluster expands in a modular fashion, with nodes being added to the cluster as needed. No longer do you need to plan capacity increases months in advance — or pay for storage you're not using.

CENTS per GB
per MONTH
including data protection

ENTERPRISE STORAGE	CLOUD STORAGE	OBJECT STORAGE
11¢	ACCESS CHARGES / 0.4 - 2¢	HYPERSTORE / 0.5 - 1¢

OPEX Savings

Traditional storage becomes complex to manage as the number of systems and associated middleware tools grows. Object storage saves by consolidating data to a single system and leveraging built-in management tools such as automated DR between sites. Support costs are reduced as well.

Because object storage runs on standard hardware, it eliminates your need to deploy and maintain proprietary storage servers. In fact, the Cloudian object storage solution can reduce your enterprise storage costs with up to 95% less management overhead, 30% less power/space/cooling, and a highly robust design that ensures maximum productivity with up to 14 nines of data durability.

6: The Technical Benefits of Object Storage

Object storage is not an architectural enhancement, but rather a fundamentally different approach to storage with benefits tailored for large-capacity use cases.

Limitless Scalability

Object storage's flat file system eliminates the scaling limitations of traditional storage, which in turn allows an object storage system to grow much larger than traditional storage.

Driven by Metadata

Metadata identifies properties of a storage object, and can be customized and updated over time. Google-like search and data visualization tools make information easy to find and analyze.
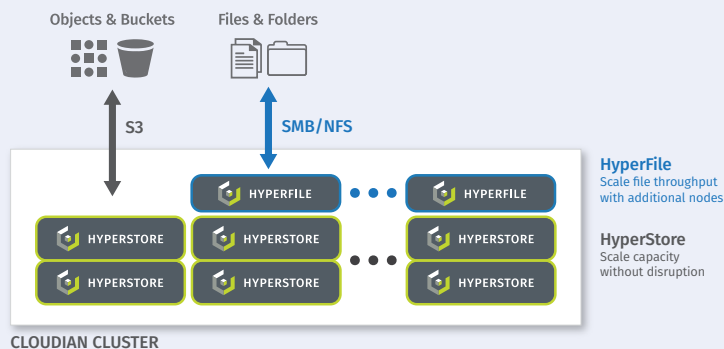
Cloud Native

Object storage is the storage technology of the cloud. On-prem object storage employs the same language (the S3 API), so the two storage environments can be managed as one for use cases such as DR and on-demand capacity expansion.

Files and Objects

File storage systems face limitations in capacity and cost. Object storage offers limitless scale at lower cost. While not intended for transactional data where operations-per-second is critical, object storage presents an economic option for large pools of less latency-sensitive files.

Meet Cloudian HyperStore

Consolidate your unstructured data — both objects and files — to a single, limitlessly scalable storage pool with Cloudian HyperStore. Available as either standalone software or fully integrated appliances, HyperStore enterprise object storage provides unlimited capacity scalability, intuitive management tools, uncompromising data protection, and the industry's most compatible S3 API implementation.

Objects & Buckets **Files & Folders**

S3   SMB/NFS

HYPERFILE ••• HYPERFILE
HYPERSTORE HYPERSTORE HYPERSTORE
HYPERSTORE HYPERSTORE HYPERSTORE

HyperFile
Scale file throughput with additional nodes

HyperStore
Scale capacity without disruption

CLOUDIAN CLUSTER

Enterprise NAS File Services

Cloudian HyperFile®, a scale-out NAS controller for Cloudian HyperStore, delivers limitlessly scalable enterprise file services, on-prem. Together with Cloudian HyperStore, HyperFile provides a cost-effective solution for your capacity-intensive, less frequently used files. Seamlessly grow from terabytes to petabytes. Independently scale capacity and performance. And get the features of enterprise NAS at one-third the cost.

7: Object Storage Use Cases

Data Protection

Object storage makes an ideal target for data protection systems from Rubrik, Commvault, Veritas, and Veeam. All are proven compatible with Cloudian and are simple to integrate. DR planning is simple, too, with built-in replication to the cloud or to a remote site.

Media and Entertainment

To accommodate high-resolution formats and rapidly expanding content libraries, object storage provides studios and post-production teams with unlimited capacity and modular growth. Rich metadata dramatically increases the ability to find media assets. Most asset management software supports the S3 API, making object storage a simple plug-in alternative to tape-based archives.

Storage-as-a-Service

For service providers and organizations looking to offer S3-compatible storage services, object storage is the ideal solution. Cloudian provides features such as quality of service controls, multitenancy, and billing to make services easy to manage.
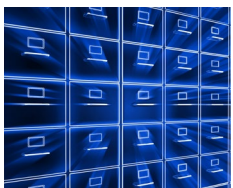
File Services

Files and objects are easily combined in a single pool, delivering scale and cost savings benefits vs. traditional file platforms. Faster than cloud and more cost-effective than enterprise NAS, object storage is the advanced file services option.

Healthcare

Object storage integrates seamlessly with PACS systems and vendor neutral archives (VNA) to provide healthcare professionals with a unified view of all patient data. Object storage is about one-third the cost of the proprietary storage systems commonly employed for healthcare storage.

Video Surveillance

With 4K cameras, higher frame rates, and longer retention periods, security personnel need exponentially more storage than in the past. Object storage is proven with video management software and provides a scalable, affordable solution for video surveillance.

Artificial Intelligence

For rapidly expanding learning databases, you need a storage system that can start small, and grow limitlessly and affordably. Object storage provides the ability to scale on-site and to use the cloud to flex with your needs.

8: Case Study: WGBH

WGBH, the PBS television station based in Boston, is well known for its slate of award-winning programming such as "Nova," "Frontline," "Masterpiece," and "Antiques Roadshow."

Fifty years of content creation has resulted in an enormous archive, most of which was stored in a tape library or on external hard drives located within in a large vault at the WGBH studios . Meanwhile, an ongoing transition to 4K and 8K media created ballooning capacity demand, while the growing rate at which media was created added to the sheer volume of material to be archived. Media retrieval time and data protection were challenges as well.

In the search for a new solution, the WGBH team decided that a hybrid cloud approach would best meet its needs. A hybrid cloud environment combines on-prem and public cloud storage: on-prem for the working copy, public cloud for a DR copy. This solution features the rapid data access of an on-prem system with simplicity of cloud-based DR.

Eventually, WGBH's research led them to object storage — specifically, Cloudian's HyperStore.

HyperStore provides WGBH with fast access to data, limitless capacity, modular and easily managed growth, high density, metadata tagging to facilitate search, and low cost. The initial deployment consisted of a 3PB cluster, housed in three 4U-high appliances. Consuming just 21 inches of rack height, this cluster consumes less than 1/10 the space of the equivalent tapes and library facilities.

VIEW VIDEO
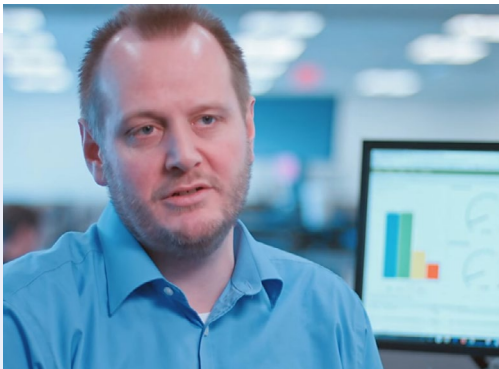
"With Cloudian, DR became automatic. We store data to the archive and it's automatically replicated to the cloud. That's a lot simpler and more reliable than managing tapes."

Shane Miner
WGBH Senior Director of Technical Services

AMERICAN EXPERIENCE   Masterpiece   NOVA   FRONTLINE.   ANTIQUES AR ROADSHOW

9: Conclusion

Hyperscale object storage addresses the capacity storage challenge in ways that traditional NAS and SAN architectures cannot. Object storage is a different kind of storage, not a middleware-enhanced spin on an older technology. As the only storage type to come of age during the cloud era, it provides unique capabilities that let you reshape your storage strategy to meet the needs of a geographically dispersed, cloud-connected enterprise.

Think Big

Consolidate data, organization-wide, to a single, exabyte-scalable data fabric. Cloudian's modular design makes it easy to grow. Expand capacity and geographic reach simply by adding nodes anywhere you need capacity. Performance scales, too, thanks to the peer-to-peer, shared-nothing architecture.

Cloud Connected

On-prem storage, private cloud, hybrid-cloud, or multi-cloud: It's your choice. Connect seamlessly with the public cloud, then use our integrated tools to replicate or fluidly migrate information. Whether your goal is DR, capacity expansion, or a data archive in the cloud, we make it effortless to integrate on-prem and cloud storage.

Capacity Where You Need It

Built to be distributed, Cloudian can be deployed across sites or across the globe. Place nodes within your data center, at your DR site, or at remote offices — then control them in a single data fabric. All fabric-connected devices work as one, so you can store, find and protect information wherever it resides.

By bringing the flexibility and simplicity of public cloud storage into your data center, you can simplify management costs and reduce TCO by 70% versus conventional storage systems.

THINK
BIG

▸ **Exabyte Scalable**   ▸ **Geo Distributed**   ▸ **Cloud Integration**   ▸ **Native S3**