# The definitive guide to Cloud Object Storage System dispersed storage

*Modernize your data center storage for efficiency*

## Highlights

- Part one: What is dispersed storage?
- Part two: How dispersed storage works: Step-by-step
- Part three: Concentrated dispersal mode
- Part four: The benefits of dispersed storage
- Part five: Dispersed storage solutions
- Part six: Conclusion

According to IDC, object-based storage capacity is expected to grow at a CAGR of 30.7% from 2016 to 2020, reaching 293.7 EB in 2020. Likewise, object-based storage market revenue is expected to hit USD 19.8 billion that same year.[1]

Although individuals generate most of this data, IDC estimates that enterprises are responsible for 85% of the information in the digital universe at some point in its lifecycle.[2] That means organizations take on the responsibility for architecting, delivering and maintaining information technology systems and data storage systems to meet the demand. As Figure 1 illustrates, data continues to grow and companies need a new way to store and retrieve all this data more efficiently.

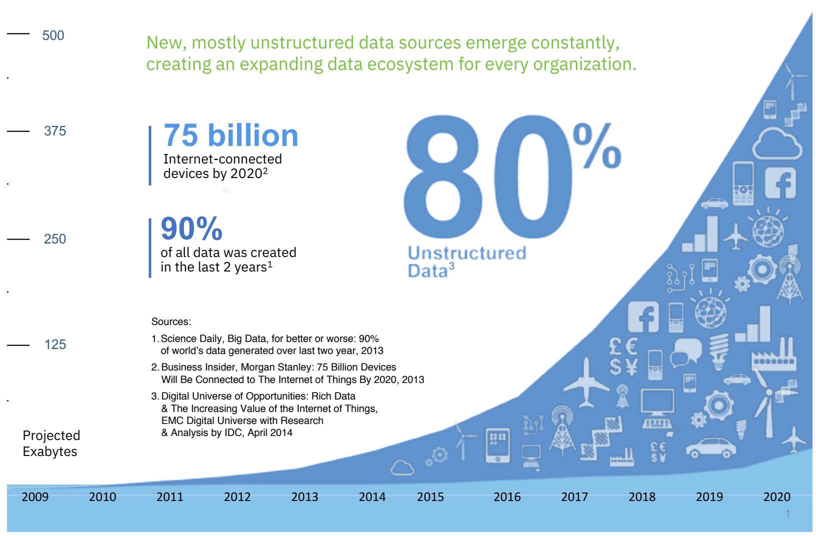


*Figure 1*: IBM Cloud Object Storage System Appliances Landscape

Many technological advances are helping with this data growth challenge to some degree. Computing is getting faster and cheaper. Virtualization is driving up efficiency and use. Storage devices are growing in terms of capacity while declining in price — more bits per device at a lower cost — and recently getting faster with the advent of solid-state technologies — although not currently at a suitable price point for all workloads. Delivery mechanisms such as cloud computing are also helping to lower costs and drive efficiencies. But in some cases, the advances in technology — specifically the capacity expansion of storage devices — are putting a strain on traditional methods of protecting and preserving digital information. Traditional storage protection technologies such as RAID are simply inadequate when it comes to protecting digital information from data loss at petabyte-scale and beyond. Traditional storage architectures are not designed to scale to the petabyte range. They're less secure. They're less reliable. And they're more expensive.

Consider the following challenges that traditional storage systems face once they reach petabyte-scale:

- Data integrity suffers when system size is 10 billion times larger than the bit error rate of a hard drive.

- Data availability suffers when hundreds of drives fail every day and require a week to rebuild.

- Data security suffers with millions of devices and multiple copies in multiple locations.

Enterprises that need to store large volumes of unstructured data must look beyond their current storage solutions and evaluate new approaches. This paper can help you understand how dispersed storage works and how its benefits have helped other organizations achieve high levels of scalability, availability and security while controlling storage costs.

## Part one: What is dispersed storage?

### Dispersed storage defined

The IBM® Cloud Object Storage System™ uses an innovative approach for cost-effectively storing large volumes of unstructured data while helping ensure security, availability and reliability. Cloud Object Storage technology uses Information Dispersal Algorithms (IDAs) to separate data into unrecognizable "slices" that are distributed through network connections to storage nodes locally or across the world. The collection of distributed storage appliances creates what is called the Cloud Object Storage System. With Cloud Object Storage dispersed storage technology, transmission and storage of data are inherently private and secure. No complete copy of the data resides in any single storage node, and only a subset of nodes needs to be available to fully retrieve the data on the network.

### Background

By taking the methods that the Internet used for data networking and applying them to data storage, dispersed storage is designed to allow companies to store massive amounts of content — video, audio, photo, text — securely and reliably.

Much like the Internet used an open protocol (TCP/IP) based on the improved design of packet switching in comparison to the established telephony protocols used in older circuit-switched networks, dispersed storage is a commercial-grade implementation of an IDA.

IDA technology transforms data into slices by using equations such that a subset of the slices can be used to re-create the original data. These slices, which are like packets but are for data storage, are then stored across multiple storage appliances, also referred to as storage nodes. Slices are created using a combination of erasure coding, encryption and sophisticated dispersal algorithms.

Dispersed storage systems are well-suited for storing unstructured data like digital media of various types and sizes, including small size documents produced by desktop productivity applications, and server log files, which are typically larger files. Currently available industry-standard hardware, software and networking technologies are not cost effective for dispersal of structured data in latency sensitive, high IOPS workloads like transaction-oriented databases because of the overhead in processing associated with slicing and dispersing.

## What is information dispersal?

At the foundation of the Cloud Object Storage System is a technology called information dispersal. Information dispersal is the practice of using erasure codes to create redundancy for transferring and storing data.

An erasure code is a Forward Error Correction (FEC) code that transforms a message of k symbols into a longer message with n symbols such that the original message can be recovered from a subset of the n symbols (k symbols).

*Simply speaking, erasure codes use advanced deterministic math to insert extra data in the original data that allows a user to need only a subset of the coded data to re-create the original data.*

An IDA can be made from any FEC code. The additional step of the IDA is to split the coded data into multiple segments, which can then be stored on different devices or media to attain a high degree of failure independence. For example, using forward FEC alone on files on your computer is less likely to help if your hard drive fails, but if you use an IDA to separate pieces across machines, you can now tolerate multiple failures without losing the ability to reassemble that data.

## The math behind information dispersal

How do these IDAs work? In algebra, when you have a system of equations with five variables, you can solve for those variables when you have at least five outputs from different equations using those variables.

As shown in Figure 2, there are five variables (a through e) and eight different equations that use these variables, with each yielding a different output. To understand how information dispersal works, imagine the five variables are bytes. Following the eight equations, we can compute eight results, each of which is a byte. To solve for the original five bytes, we may use any five of the resulting eight bytes.

This is how information dispersal can support any value for k and n. k is the number of variables, and n is the number of equations.

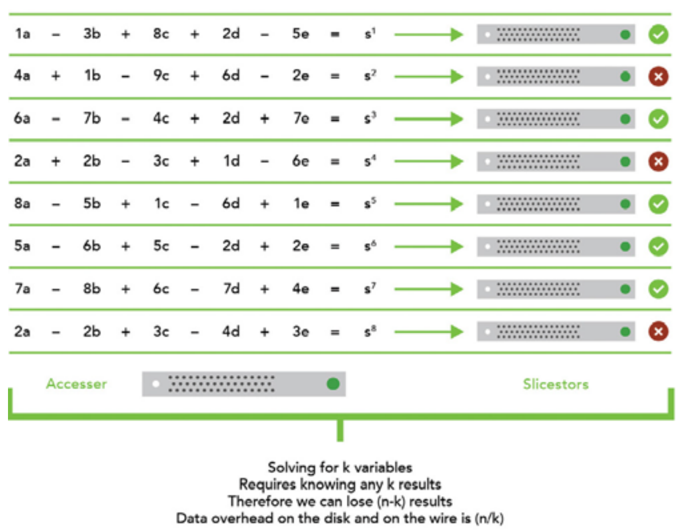


*Figure 2*: Mathematical equations behind information dispersal and retrieval algorithms

## Part two: How dispersed storage works: Step-by-step

At a basic level, the Cloud Object Storage System uses three steps for slicing, dispersing and retrieving data.

1. Data is virtualized, transformed, sliced and dispersed using IDAs. In the example in Figure 3, the data is separated into 12 slices. Therefor the “width” (n) of the system is 12.
2. Slices are distributed to separate disks, storage nodes and geographic locations. In this example, the slices are distributed to three different sites. With Concentrated Dispersal mode, described later in this paper, more than one slice can occur on a single storage node.
3. The data is retrieved from a subset of slices. In this example, the number of slices needed to retrieve the data is 7. Therefor the “threshold” (k) of the system is 7.

Given a width of 12 and a threshold of 7, we can refer to this example as a “7 of 12” (k of n) configuration.

The configuration of a system is determined by the level of reliability required. In a “7 of 12” configuration,” five slices can be lost or unavailable and the data can still be retrieved because the threshold of seven slices has been met. With a “5 of 8” configuration, only three slices can be lost, so the level of reliability is lower. Conversely, with a “20 of 32” configuration, 12 slices can be lost, so the level of reliability is higher.

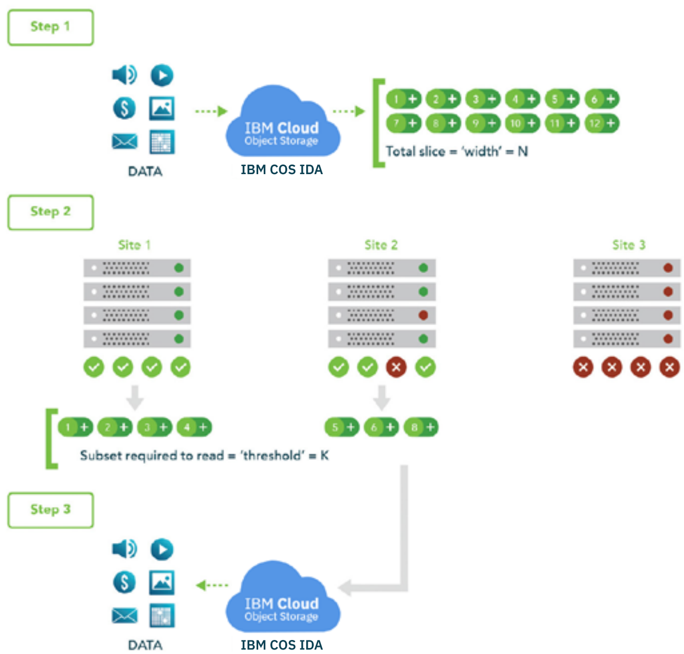


*Figure 3*: Step-by-step data slicing, dispersal and retrieval in a Cloud Object Storage System

## Multi-site failure example

With dispersed storage, only a subset of slices is required to retrieve the data. This allows a dispersed storage system to tolerate appliance failures both within a single site and across multiple sites.

1. Data is virtualized, transformed, sliced and dispersed using IDAs. The “width” (n) of the system in this example is 12.
2. Slices are distributed to separate disks, storage nodes and geographic locations. In the example shown in Figure 4, the slices are distributed to four geographically dispersed sites.
3. The data is retrieved from a subset of slices. In this example, the number of slices needed to retrieve the data is seven. So even though failures are occurring across all three sites, the data is still available to be retrieved because the “threshold” of seven available slices has been reached.

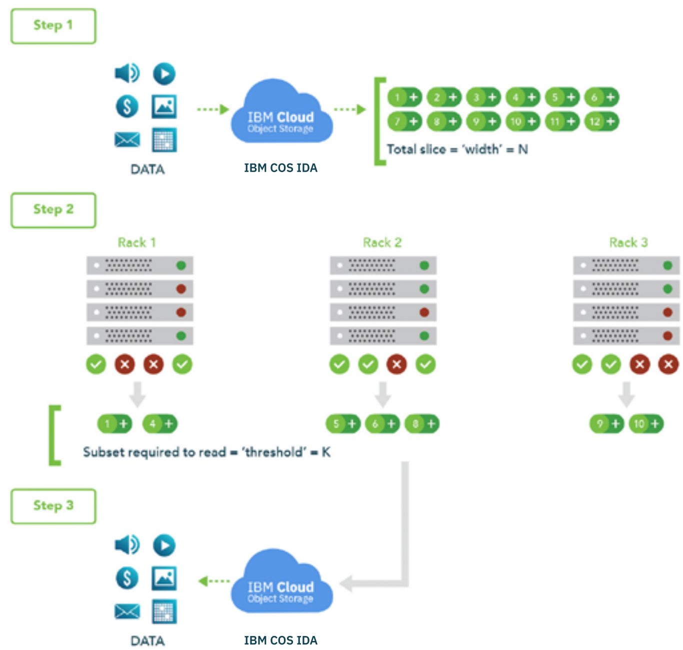


*Figure 4*: Data slicing, dispersal and retrieval in a Cloud Object Storage system to withstand failures within a single site or across multiple sites

## Single-site or multi-device failure example

A dispersed storage system can also be deployed in a single site with the ability to tolerate the failure of multiple appliances within that site.

1. Data is virtualized, transformed, sliced and dispersed using IDAs. The “width” (n) of the system in the example in Figure 5 is 12.
2. Slices are distributed to separate disks, storage nodes and geographic locations. In this example, the slices are distributed to four different racks within a single site.
3. The data is retrieved from a subset of slices. In this example, the number of slices needed to retrieve the data is seven. So even though each rack has experienced one or more device failures, the data is able to be retrieved because the “threshold” of seven slices has been met. Even with five slices unavailable, the data can be bit-perfectly re-created.

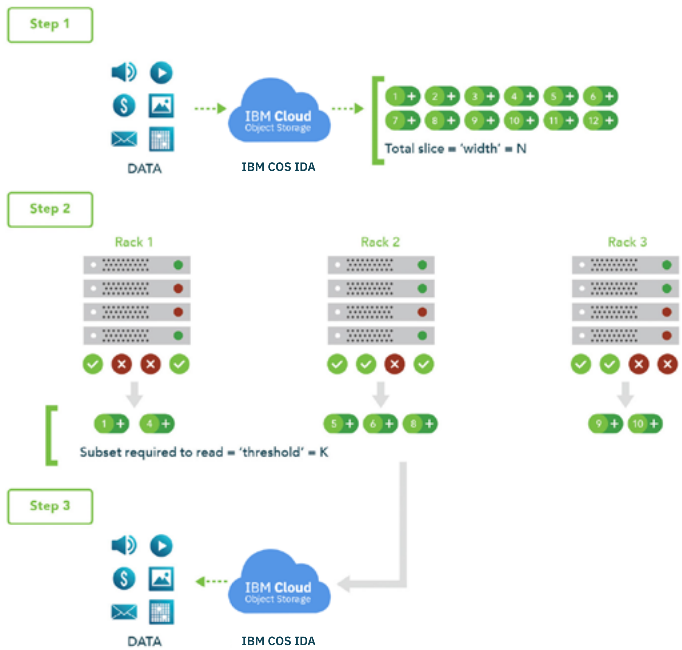


*Figure 5*: Data slicing, dispersal and retrieval in a Cloud Object Storage system to withstand multiple failures in a single site

## Components of a Cloud Object Storage System

The Cloud Object Storage System enables the creation of storage systems using three software components – the IBM Cloud Object Storage Manager, IBM Cloud Object Storage Accesser® and IBM Cloud Object Storage Slicestor®. These software components can be deployed on a wide range of compatible industry-standard hardware platforms, as virtual machines. In the case of the Accesser, as an application running on a Linux-OS. Physical and virtual deployment can be combined in a single system. For example, virtual machines for the Cloud Object Storage Manager and the Accesser and physical servers for the Slicestor.

### Software

Each of the three software components serves a specific function as a part of Cloud Object Storage:

- The Manager is responsible for monitoring the health and performance of the system, configuring the system and provisioning storage, managing faults, and other administrative and operational functions.
- The Accesser is responsible for encrypting or encoding data on ingest and decoding or decrypting data when read. This is in addition to managing the dispersal of data slices resulting from this process across a set of Slicestor nodes.
- The Slicestor is responsible for the storage of data slices.
- When the Manager, Accesser, and Slicestor software are deployed on IBM-certified industry-standard hardware platforms, there are a number of benefits:
- Reduced time to production on initial deployment because hardware and software compatibility and configurations are predefined and validated by Cloud Object Storage.
- Optimized hardware configuration to help maximize the value of Cloud Object Storage.
- Increased system reliability due to monitoring and management of hardware health at a lower component level.
- Access to Cloud Object Storage support staff that is familiar with both the hardware and software components of the system

### Object storage foundations

Cloud Object Storage is based on a simple object storage approach that efficiently stores billions of data objects in a single flat namespace and exposes the data through a REST interface using the HTTP-based protocol.

### The old way: File-based

Traditional storage systems organize data in a hierarchical file system and expose the data through NAS-based protocols like NFS and SMB. The file system approach tends to be ideal for human users storing small amounts of data.

### The enhanced way: Object-based

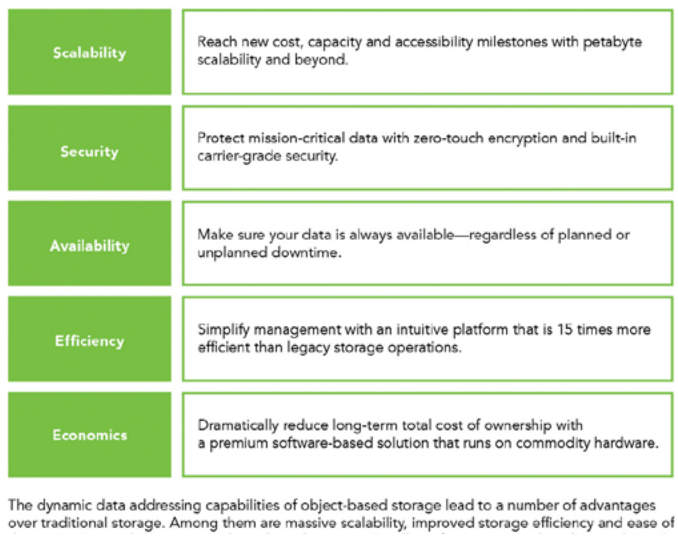


*Figure 6*: Advantages of object-based storage over traditional storage in a petabyte-ready storage system

With file system storage, data is closely tied to its location. Object-based storage overcomes this limitation by decoupling data from its physical location in the storage system. Analyst firm Forrester Research cites valet parking as an apt analogy for object-based storage . When you valet park your car, the attendant gives you a claim ticket that allows you to retrieve your car when needed. While the attendant has your car, he or she might move it around as needed to optimize space in the parking lot or garage. The claim ticket identifies your car, not a particular parking space. With object-based storage, an object ID identifies a particular piece of data, but not its specific location in the storage system. Data can be moved around in the system as needed, and the object ID is the "claim ticket" needed to retrieve the data, wherever it resides.

## Access methods

### Object-based access methods

The underlying storage pool of a dispersed storage system can be shared and is jointly accessible by the compatible S3 REST API. Simple PUT, GET, DELETE and LIST commands allow applications to access digital content. The resulting object ID is stored directly within the application. Using the standard S3 interface makes IBM Cloud Object Storage compatible with many other S3-compatible applications and storage systems. Accesser, which accepts all the access to the storage, can be deployed as a virtual machine, as an embedded Accesser on a storage server or as an appliance.

### REST API access to storage

REST is a style of software architecture for distributed hypermedia information retrieval systems such as the web. REST-style architectures consist of clients and servers. Clients initiate requests to servers. Servers process requests and return associated responses. Requests and responses are built around the transfer of various representations of the resources.

The REST API works in way similar to retrieving a universal resource locator (URL). But instead of requesting a web page, the application is referencing an object.

REST API access to storage offers several advantages:

- Tolerates Internet latency
- Provides for "programmable" storage
- Provides efficient global access to large amounts of data

### File-based access methods

Dispersed storage can also support the traditional NAS protocols — SMB/CIFS and NFS — through integration with third-party gateway appliances. Users and storage administrators can transfer, access and preserve data assets over standard file protocols.

## Security features

### Data security

IBM SecureSlice™ technology is used to help ensure confidentiality, integrity and availability of data stored on a Cloud Object Storage System. SecureSlice combines two algorithms: an IDA and an All-or-Nothing Transform (AONT). AONT is a mode of encryption in which the information can only be deciphered if all the information is known. The diagrams shown in Figures 7 and 8 illustrate basic write and read operations using SecureSlice.

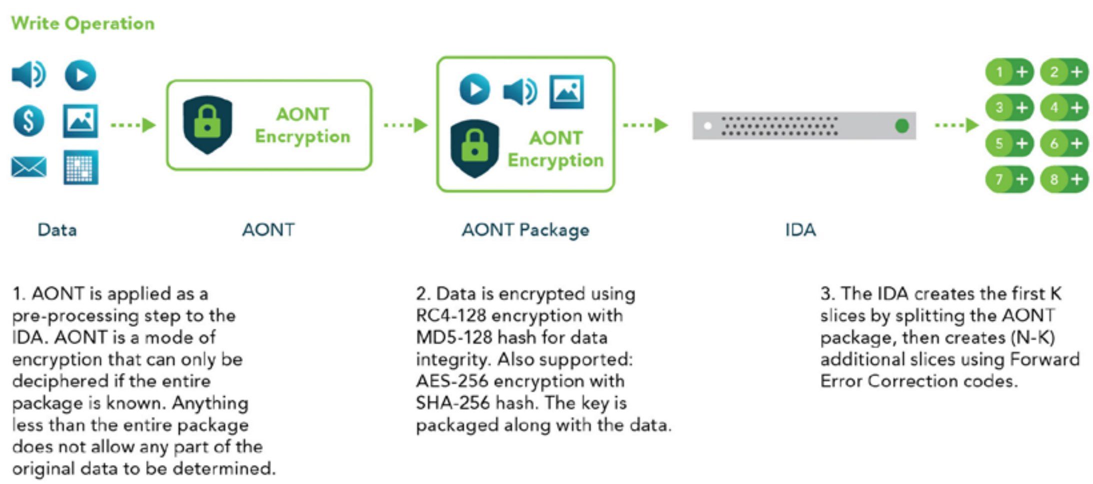


*Figure 7*: Basic write operation in a Cloud Object Storage system

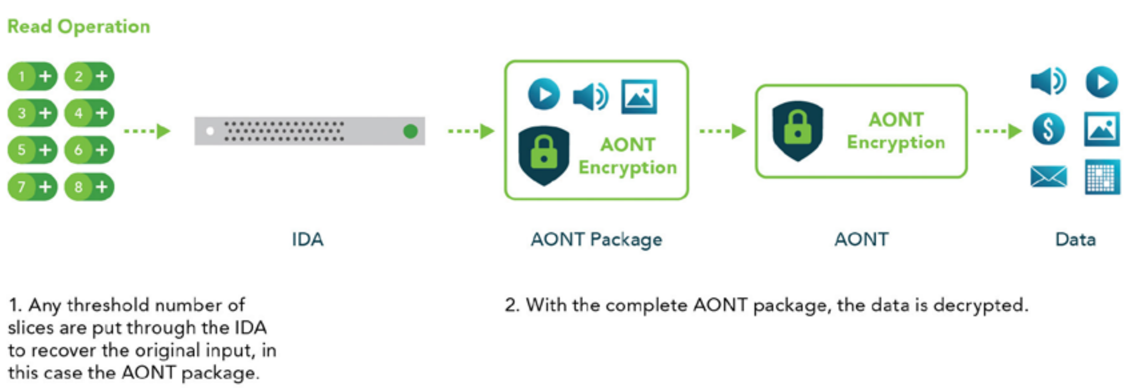


*Figure 8*: Basic read operation in a Cloud Object Storage system

**Internal of AONT Encoding**

When a segment of data is to be stored in a dispersed storage system, an integrity check value is first appended to the data. The integrity check value can be any well-known constant value, so long as its length is sufficient. This value will be checked after decoding, to help ensure that no corruption has occurred.

If any slice used in the reassembly of the data segment has been corrupted, there is a high probability that the integrity check value will also be corrupted. The dispersed storage invalid data from reaching the user. Should the integrity check value be corrupt on a given slice, the dispersed storage system will attempt to find a valid combination of slices to retrieve the complete data segment.

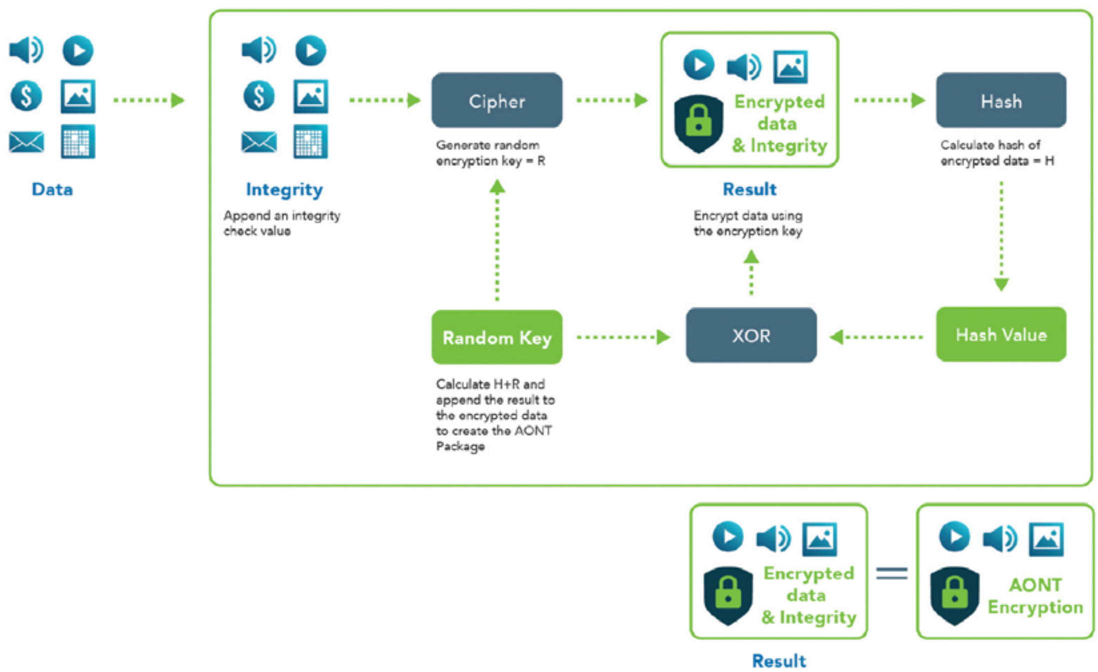


*Figure 9*: Internals of AONT operation in a Cloud Object Storage system

**Network security**

All network traffic flowing into or out of appliances in a dispersed storage system is encrypted using TLS, SSL or SNMPv3 with AES. Storage nodes may be placed virtually anywhere without complex firewall or VPN setup. See Figure 10.

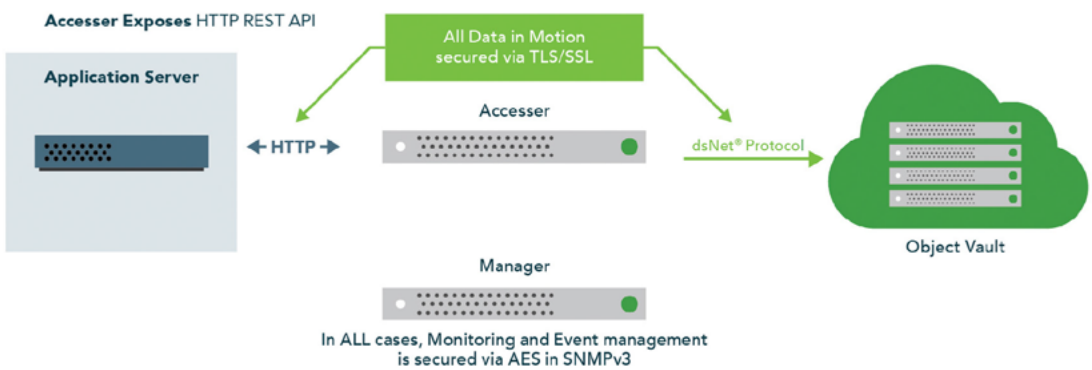


*Figure 10*: Network security in a Cloud Object Storage System

**Device registration — Certificate Authority trust**
It is not enough to simply say devices use TLS or SSL. These protocols do not prevent "man in the middle" strikes unless the connections are authenticated. Authentication with TLS/SSL requires the use of digital certificates, and these certificates must be verifiable as belonging to a valid node in the storage network. To accomplish this, nodes are given a signed digital certificate at the time they are approved into the storage network.

Such approval requires an administrator to log in to the management interface, view the request, and authorize it. The administrator can see the IP address, MAC address and fingerprint of the device making the request, and verify that each is valid before accepting the device into the system. Once approved, the node will be granted a certificate signed by the certificate authority (CA) for the storage network. All devices in the storage network trust this certificate authority, and by extension, any node that owns a valid certificate signed by this CA. Appliances may, at some future time, be retired or become compromised. At this point in time, the Manager may revoke the device's certificate by adding it to the Certificate Revocation List (CRL), which is periodically polled by every node in the system. See Figure 11.

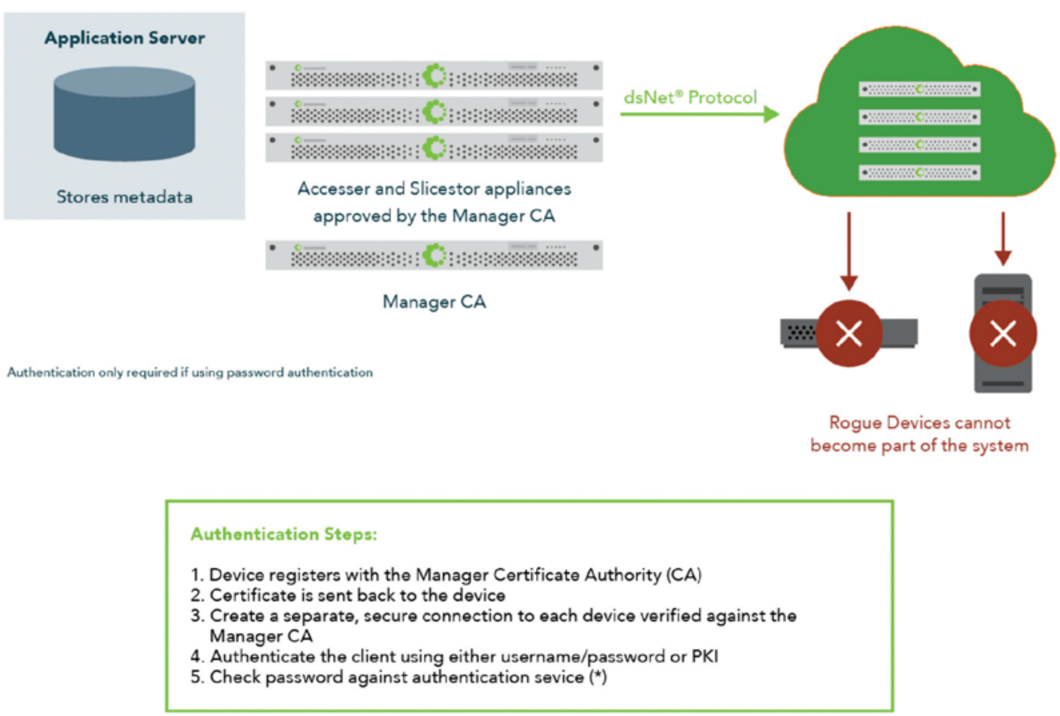


*Figure 11*: Device registration and authentication in a Cloud Object Storage system

## Availability features

The availability features of a dispersed storage system provide continuous error detection and correction, helping ensure bit-perfect data availability.

### Integrity check on all slices and files

A dispersed storage system checks for data integrity through an intelligent background process that proactively scans and corrects errors, scans data slices for integrity, rebuilds any corrupted slices, and checks for both slice integrity and file data integrity prior to delivery. This helps ensure bit-perfect data delivery through proactive correction of bit errors as well as correction of latent soft errors that may occur during normal read/write operations. It also helps ensure that data cannot be modified without authorization and that malicious threats are detected. See Figure 12.

### Continuous error correction

If a slice is determined to be corrupted—meaning the integrity check value is invalid—the Slicestor starts the distributed rebuilder technology to replace the slice with a valid slice. If the slice is missing, the distributed rebuilder technology re-creates a valid slice. Continuous error correction increases system availability because it is not waiting for data to be read to detect errors, as shown in Figure 13. This is crucial with long-term archives and massive digital stores where information isn't as frequently read. The distributed rebuilder model allows for predictability—the rebuilder is "always on" at a moderated rate, making I/O performance much more predictable—as well as scalable, as the rebuilder grows with storage.

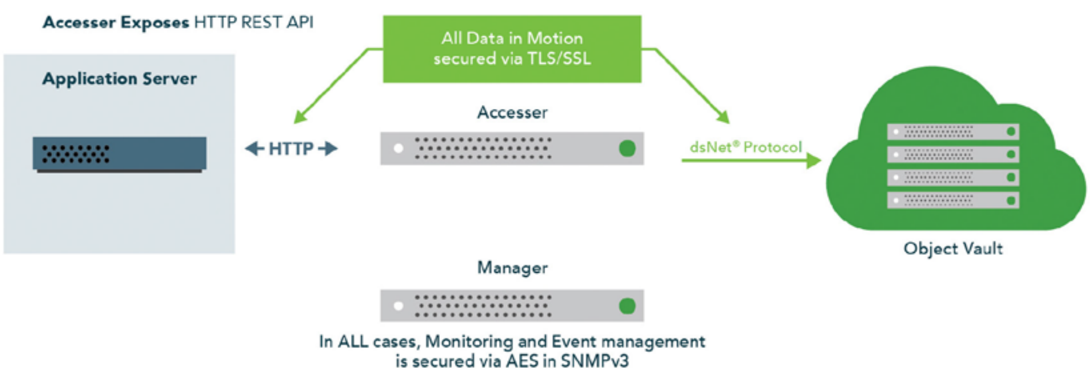


*Figure 12*: Data integrity checks in a Cloud Object Storage system for a bit perfect data availability

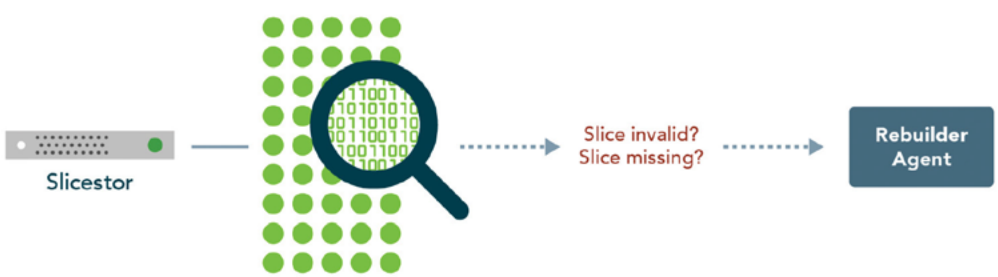


*Figure 13*: Continuous error correction by the Slicestor in a Cloud Object Storage system

## Performance optimization features

Dispersed storage uses IBM Cloud Object Storage SmartWrite and IBM Cloud Object Storage SmartRead technology to optimize writes and reads of slices, resulting in improved throughput and efficiency.

### Improved performance of k of n writes

SmartWrite enables a successful write operation even if the full width of slices can't be written—for example, if there is a failure condition at a node or within the network. See Figure 14. SmartWrite attempts to write all slices. Once the required write threshold of slices is achieved, the write is considered successful. The remaining slices continue to attempt to write asynchronously. If a slice write operation times out, it will be detected and rebuilt.

### Improved performance of k of n reads

SmartRead predicts the optimal network routes and storage nodes to more efficiently retrieve data. Data is reassembled in segments, and for each segment, thousands—if not millions—of combinations of slices are examined to help determine the best delivery path, as shown in Figure 15. SmartRead ranks storage nodes by their on- demand performance and requests the optimal combination of slices to recreate the data. If a slice request is not performing, SmartRead requests a slice from another node.

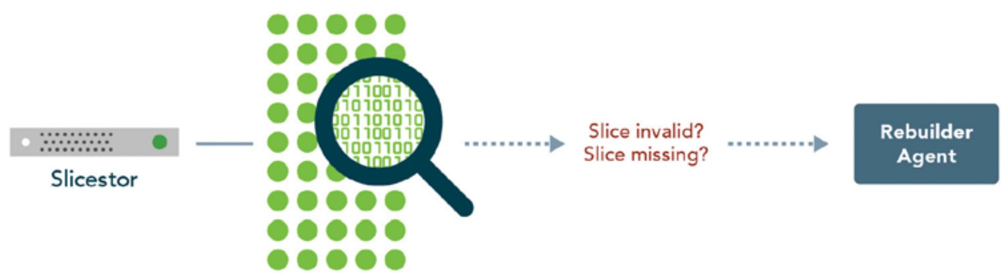


*Figure 14*: Continuous error correction by the Slicestor in a Cloud Object Storage system

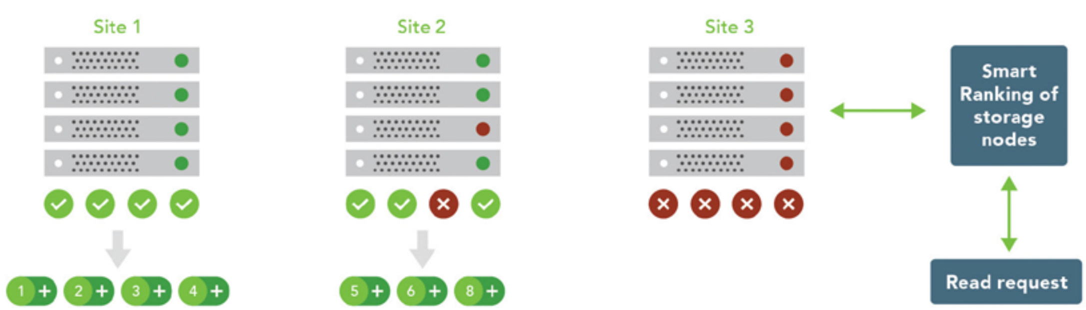


*Figure 15*: Improved read performance due to SmartRead technology in a Cloud Object Storage system

## Part three: Concentrated dispersal mode

Beginning with Cloud Object Storage System release 3.12, a new operational mode was introduced that reduces the number of storage nodes required for a reliable and efficient small Cloud Object Storage system. This new mode, called Concentrated Dispersal (CD) mode, allocates multiple slices of an object to the same storage node (but no more than one slice per disk) to optimize the reliability, availability and efficiency of the Cloud Object Storage system. The system will automatically detect and configure the storage pool with CD mode or standard dispersal depending on the number of storage nodes assigned to that pool. No other features or functionality is changed except for the number of slices on a storage node. Figure 16 illustrates how slices are more compact and allowing for smaller configurations with CD mode.

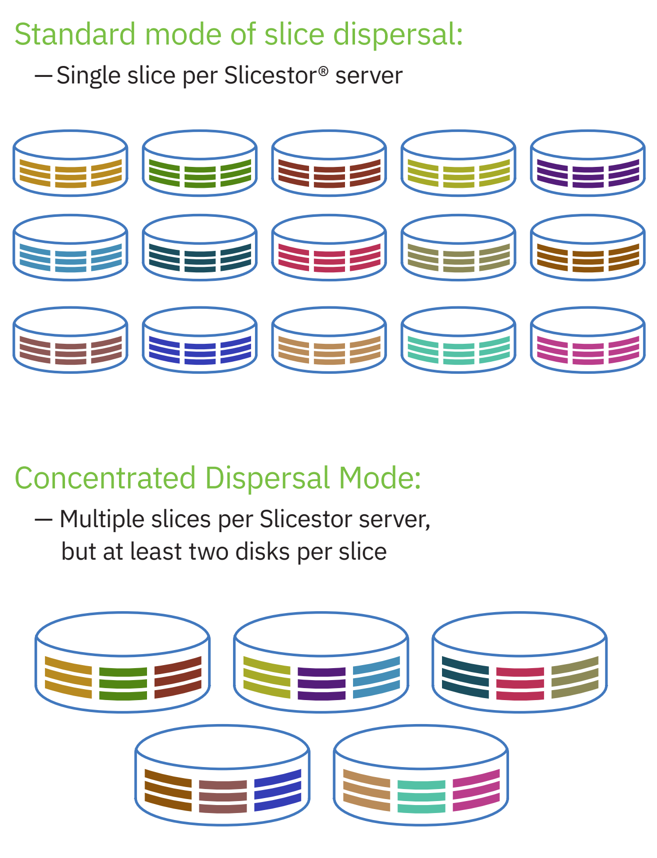


*Figure 16*: Concentrated Dispersal mode comparison

### CD mode benefits

The key benefit of CD mode is that it provides the ability for customers to start building their object storage system with much smaller configurations than were previously possible. New systems can start with up to 75% less space or capacity than previous initial configurations. A new system can now start with as little as 72 TB useable capacity and grow to a massive scale. Customers who start small lose no features or functionality. The same system software that runs on the petabyte scale systems runs on the smaller systems. As with any Cloud Object Storage configuration the system can be expanded with zero downtime by adding additional storage nodes. The additional storage nodes can operate either in CD mode or standard dispersal mode. The expansions can be accomplished on line without ever losing access to data. The system will move slices in the background to the newly added storage nodes while maintaining full access to the data. In this way, customers can start with terabytes and grow to petabytes and exabytes all online without stranding any investment.

## Part four: The benefits of dispersed storage

### Scalability benefits

Dispersed storage provides massive scalability with significantly reduced administrative overhead. Systems are permitted to grow from terabytes to petabytes to exabytes.

#### Multiple drivers of scalability

To store and manage data at the petabyte, exabyte and beyond level, an architecture that can scale is crucial. With no centralized servers, capacity and performance for a dispersed storage system can be scaled independently. The object storage foundation of dispersed storage enables data mobility, scalability and storage efficiency crucial for limitless scale storage.

Dispersed storage delivers a single addressable global namespace that virtualizes all individual storage nodes—providing a single point of management. Additional benefits of utilizing a global namespace approach include the ability to open up more storage pools for larger working pools of disks, migrate data transparently and reduce the number of mount points and/or shares in an environment.

The dispersed storage protocol can be used within application servers or devices, each independently accessing storage nodes. This helps enable massive parallel writes and content distribution to be achieved. And it avoids the choke points of a gateway, helping improve performance in a distributed environment.

### Availability benefits

Dispersed storage maintains nearly 100 percent data integrity even as millions of physical bit errors occur or as multiple drives, servers, containers or locations change or are replaced. The information that goes in is the same that comes out, completely authenticated, and the information is accessible from virtually anywhere, anytime. Data is always available with an architecture that can tolerate simultaneous failures.

#### Configurable availability and zero-downtime upgrades

Dispersed technology provides exceptional data protection and availability. Dispersal is significantly better than many other storage solutions, because it does not replicate data to overcome the shortfalls of other implementations and does not suffer the significant risk of data loss that other storage solutions can experience, during the rebuild process, which may take many hours—or even days—for even a single hard drive. It is configurable to provide higher levels of fault tolerance (k of n) when compared to RAID 5 (1 of n) and RAID 6 (2 of n) used in many other storage solutions. By using IDAs and storing the resulting slices on independent hardware that can be either in a single site or geographically dispersed, Cloud Object Storage helps drive reliability and availability without replication.

The Cloud Object Storage system allows enterprises to tolerate entire site failures and still have seamless access to data without expensive copies. Taking a data center offline for routine maintenance does not change availability. With dispersed storage, zero-downtime upgrades are possible. Rolling upgrades enable the system to remain operable with data accessible throughout the process. No scheduled maintenance window is required.

## Security benefits

Dispersal helps ensure data confidentiality even when multiple drives, servers, containers or locations are compromised. Data in motion and data at rest is encrypted to help make it completely unrecognizable and inherently secure to eliminate opportunities for security breaches.

### Exceptional security for data at rest and data in motion

Dispersed storage technology uses encryption, AONT, an integrity check value and IDAs to split data into inherently secure slices. Each slice is encrypted, but no external key management is required. All data is computationally secure unless a "threshold" of slices is available to decrypt it. The result is slices that do not contain any representation of the data and that require a threshold number to re-create the data bit-perfectly.

These slices are stored on independent hardware. No full copy of the data exists on any storage volume. Individual servers containing slices are useless without possessing a threshold number of them, which must be taken from many different physical locations. This means the likelihood of a security breach is significantly lower and in some cases eliminated. Further, since the slices a re created prior to traversing the network, the slices are protected against on-the-wire security breaches. The result is exceptional data security for both data over the network as well as data at rest.

## Economic benefits

The Cloud Object Storage System delivers significantly lower total cost of ownership for storage systems at the petabyte level and beyond by significantly reducing and in many cases eliminating expensive replication and associated incremental costs. Hardware, electricity, floor space, support and management costs are also reduced.

### No copies, lower costs

Dispersed technology eliminates the need for costly replication. The Cloud Object Storage System delivers the equivalent availability of up to four replicated copies of data while reducing storage requirements by up to five times when compared to traditional approaches.

When compared to a popular storage file service, Cloud Object Storage requires:

- Less than one third of the raw storage for data dispersal with a single copy of data and no replication required
- Almost one third of the power and cooling costs with less resources required.
- Almost one fourth of the floor space due to capacity savings.
- No additional costs for hardware, software and tape copy

## Part five: Dispersed storage solutions

### Use cases

The Cloud Object Storage System is an ideal solution for enterprises who need to securely store large volumes of unstructured data with high availability and where latency is not a primary consideration.

### Unstructured data delivery

With the unprecedented growth in new digital information, use cases have emerged that enable organizations to store and distribute limitless data. A distributed and decentralized storage architecture along with an object storage interface helps enable enterprises to deliver data to their users across the globe. These use cases include content repository, storage-as-a-service, enterprise collaboration, backup, and archive.

### Content repository storage

The Cloud Object Storage System provides one of the most reliable, scalable platforms for your business-critical data.

### Effectively store and protect valuable content

Consumers access content from different locations worldwide, making it a business priority to protect irreplaceable originals at scale. Cloud Object Storage's content repository solutions deliver data availability at petabyte and beyond scalability. The scalable Cloud Object Storage System delivers carrier-grade security for a single copy of original content before dispersing it geographically. Cloud Object Storage technology helps ensure data integrity from start to finish.

Organizations need content storage that can be distributed across their infrastructure to effectively store and distribute content. Cloud Object Storage provides a high-availability environment, long-term file integrity and access, and authentication enforcement. Whether organizations have less active, fixed or frequently accessed content that users are collaborating on, Cloud Object Storage offers a security-rich, reliable and cost-effective approach. A shared content storage repository can be accessed in a safeguarded manner by people inside or outside of an organization, enabling collaboration across geographies.

### Case study: A major league baseball team

**Problem:** For Major League baseball teams, video means practically everything. Video photographers capture every pitch and every hit of every player at every game. As a result, the team collects a large amount of unstructured video data to store, secure and make available to their coaches and employees.

**Solution:** The Cloud Object Storage benefits include massive scalability, geographic independence, multi-tenant features and the ability to use proprietary, off-the-shelf technology, which provides additional cost savings. A baseball team placed Cloud Object Storage software in the back end of their system, with a controller as the local interface to the global file system for seamless data access. They promptly noticed the difference it made to the IT staff.

**Results:** Cloud Object Storage technology gave this major league baseball team the ability to have cross-site access to data from each of its data sites without sacrificing security. The Cloud Object Storage solution eliminated up to 30% of IT professionals' time to serve as a backup administrator and coordinate all the replication routines. The IT security chief can keep all data up to date, backed up and replicated across all of their sites. The coaches, scouts and trainers have access to the content through the cloud by logging in and finding what they need to retrieve. In addition to enhanced data security and seamless data collocation, they value the system's transparency to the user. Based on the successes the team experienced with Cloud Object Storage technology, they are expanding the cloud storage system into their affiliates in the minor leagues.

## Storage-as-a-service

Cloud Object Storage software helps deliver new levels of storage capacity and availability with carrier-grade data security to a company's user base.

**Delivering storage capacity and availability**

Being able to sell capacity to customers on a centralized infrastructure is a must for service providers and large enterprises. Cloud Object Storage helps these organizations implement storage-as-a-service solutions that consolidate users and customers onto a single platform. Cloud Object Storage helps streamline management and efficiently scale storage to meet their demands. With secure multi-tenancy, zero-touch encryption and robust management APIs, IT can build a storage offering that is as scalable and reliable to manage and cost-efficient.

### Case study: A leading data storage integrator

**Problem:** A leading data storage integrator decided to expand their managed and hosted data storage services. Aware that many of their potential customers were operating in hyper-growth environments, they realized that the platform they were reselling fell short. To serve these customers and grow their business, they needed a new storage-as-a-service offering.

**Solution:** Cloud Object Storage reliably stores a large amount of data cost-effectively, making it well suited for the cloud. After a thorough evaluation, this data storage integrator selected the Cloud Object Storage System to provide a competitive data storage-as-a-service solution that is more cost efficient and easier to deploy than public cloud providers.

**Results:** The Cloud Object Storage System has allowed this data storage integrator to provide storage services with data reliability that is greater than 9 nines while still competing with public cloud pricing. The simple pricing model is particularly attractive to potential customers in hyper-growth environments. The Cloud Object Storage System also enables this data storage integrator to provide customers with rich data and insights such as performance reports, capacity consumption reports and technology updates that are not typically available from public cloud offerings.

## Enterprise collaboration

Cloud Object Storage provides security-rich, distributed access to valuable content, making it easier to enable workplace productivity across the globe.

### Collaboration and productivity

Today's workforce is constantly on the move, with businesses reaching across the globe. To be successful, employees need seamless access to mission-critical data from virtually anywhere, at any time. The Cloud Object Storage solution delivers a data hub that allows business to provide global access to data. By simultaneously protecting it on-premises with zero-touch encryption, they provide security-rich, distributed data access that enables enterprise collaboration and improves productivity.

---

### Case study: A leading global marketing and implementation agency

**Problem:** A leading global marketing and implementation agency produces advertising and marketing communications for clients across all media and languages. The company was growing rapidly, with each campaign consuming a lot of storage space in the production environment. Their IT team was looking for a storage solution that would cost-effectively address their expanding amount of unstructured data. They also needed a solution that would enable their international workforce to collaborate globally and without interruption.

**Solution:** Optimized for storing high volumes of data-driven content, the Cloud Object Storage solution met this agency's requirements and was implemented in all of their main studios. With a single addressable global namespace, Cloud Object Storage delivers a unified, single point of management and access that can scale beyond the limits of traditional centralized metadata servers. Their employees can now write to it and any of the offices can pull the data back up to their systems as needed.

**Results:** Cloud Object Storage helps their staff to archive much more aggressively and limits the amount of expensive production storage they use. Cloud Object Storage also allows their staff to collaborate between offices and gives their clients easier, safer, and more reliable access to the assets they need to help ensure regulatory compliance. With the majority of their production studios connected, productivity has increased across the company.

---

## Backup

The Cloud Object Storage System provides scalable backup and always- on data availability for dependable recovery and security up to 80% lower infrastructure cost.

**Cost-effective, security-rich and accessible storage**

It is a challenging task for IT to collect and back up data from diverse application servers and user machines. Storing this data long-term and at scale is even more difficult. Cloud Object Storage backup solutions provide a cost-effective, accessible storage platform for long-term data protection. SecureSlice zero-touch encryption protects data before cost-effectively distributing it across multiple sites, helping ensure long-term, bit-perfect protection at scale. The Cloud Object Storage System enables faster access to data once it's backed up, speeding business recovery time in the event of a disaster.

### Case Study: Major retailer

**Problem:** For a major retailer, the steady production of unstructured data — including videos, photos and more — was increasing at a phenomenal rate. At the same time, IT storage platforms were struggling to scale without dramatically increasing in price and decreasing in reliability.

**Solution:** Cloud Object Storage enables business to efficiently store, manage and access data at petabyte scale and beyond. Using erasure coding — a type of forward error correction — the Cloud Object Storage solution offers far higher data resiliency than other storage solutions and requires far less storage capacity than standard object storage solutions. Cloud Object Storage demonstrated the scalability, efficiency, security and simplicity of their object software storage and the retailer's upper management was quickly on board.

**Results:** The retailer is well past a petabyte of data. The Cloud Object Storage Backup Solution economically stores mountains of video and visuals. The major driver for the move to IBM was cost-efficiency. The savings were quick. The retailer realized approximately a 50% savings per byte compared to its previous storage vendor. Cloud Object Storage's method of slicing and structuring data makes it virtually impossible to damage or steal critical information, helping ensure the data is protected.

## Active archive

The Cloud Object Storage System keeps content accessible with a scalable, reliable and security-rich long-term data archive.

**Scalable, reliable and security-rich archive storage**

Many organizations are seeking an archival solution that provides their users with rapid access to their data. IBM offers an archive storage solution that combines virtually limitless availability with one of the highest levels of data integrity and confidentiality.

Proactive error correction is crucial to keeping a long-term archive healthy, since information isn't as frequently read. Cloud Object Storage employs an intelligent background process that scans storage nodes, checking for and correcting errors.

Deployments of the Cloud Object Storage solution can span multiple data centers. An archive distributed across multiple offsite locations helps protect data against a potential single location failure or catastrophic disaster, making it more securely accessible for long-term retention. It is not tied to a specific server or storage device, and the data is automatically reconstituted as new storage nodes are installed in the system. The Cloud Object Storage archive solution is designed to enable organizations to meet their compliance requirements and long-term preservation goals.

### Case study: An Internet-based photo publishing service

**Problem:** With an active and growing archive of billions of photos in constant motion, this company faced significant challenges to make sure it could keep pace with its customer's needs and maintain the same levels of performance, availability and reliability. At petabytes of raw storage and a double-digit growth rate, the cost to store this data was growing rapidly too. They had to find ways to reduce the cost of storage and make it easier to manage.

**Solution:** The Cloud Object Storage approach to ingesting and storing data solved their most critical criteria — to eliminate single points of failure and deliver high levels of fault tolerance. Advanced erasure coding techniques disassociated the performance and reliability of individual components from application level performance and reliability. This allowed the company to have continuous availability of its data, making it far less susceptible to potential hardware and software problems in its storage tier.

**Results:** This company has over 150 PB of storage in production and is growing rapidly with a limitless capacity to scale. The company is seeing significant power consumption and management cost savings across the board. They are now able to manage the entire storage platform, containing billions of objects and over 150 PB of capacity, with only three part-time storage administrators.

## Part six: Conclusion

Enterprises that need to store large volumes of unstructured data must look beyond their current storage solutions and evaluate new approaches. Dispersed storage is one such innovative approach for cost-effectively storing large volumes of unstructured data while helping ensure security, availability and reliability.

This paper described the features and benefits of dispersed storage in five critical areas:

- Availability: Data is always available whether or not there is planned or unplanned downtime.
- Scalability: Systems can grow from terabytes to petabytes to exabytes.
- Security: Data confidentiality is maintained even when multiple drives, servers, containers or locations are compromised.
- Economic: The need for costly replication is eliminated, significantly lowering the total cost of ownership for storage systems at the petabyte level and beyond.
- Efficiency: Manage petabytes of storage per administrator.

For these reasons, dispersed storage is an ideal solution for enterprises who need to store large volumes of unstructured data and where latency is not a primary consideration. Enterprises with content storage, active archive or content distribution needs should evaluate dispersed storage as a different technology option.

### Is dispersed storage right for your organization?

Use the following checklist to determine whether your organization could benefit from dispersed storage:

- Do you have applications that require long-term retention of data?
- Does the data consist of large, unstructured objects?
- Do you have 500 usable terabytes or more of this data?
- Do you have requirements for data security, availability, scalability, and cost-effectiveness?
- Do you have the infrastructure required to support dispersed storage, including:
  - Network connectivity?
  - High-quality bandwidth (if geo-dispersed)?

## About Cloud Object Storage

Cloud Object Storage provides organizations the flexibility, scale and simplicity required to store, manage and access today's rapidly growing unstructured data in a hybrid cloud environment. Relied upon by some of the world's largest repositories, these solutions turn storage challenges into business advantages by reducing storage costs while reliably supporting both traditional and emerging cloud-born workloads.

To learn more, contact your IBM Business Partner:

**System Analysis Services, INC.**
**6176316946 | carl@sysgbs.com**
**http://systemsanalysisservices.com/**

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
October 2017



1 Worldwide File- and Object-Based Storage Forecast, 2016-2020 (IDC #US41685816, September 2016)

2 "Objectively better-leverage object-based storage to store petabytes of data efficiently," DMD Data Systems, Inc., Blog, 4/2016, http://bit.ly/2flq1SK