IBM Cloud **Object Storage**

Primary and Secondary Storage Solutions

IBM Cloud Object Storage Use Cases and Solutions Guide

January 2020

Copyright

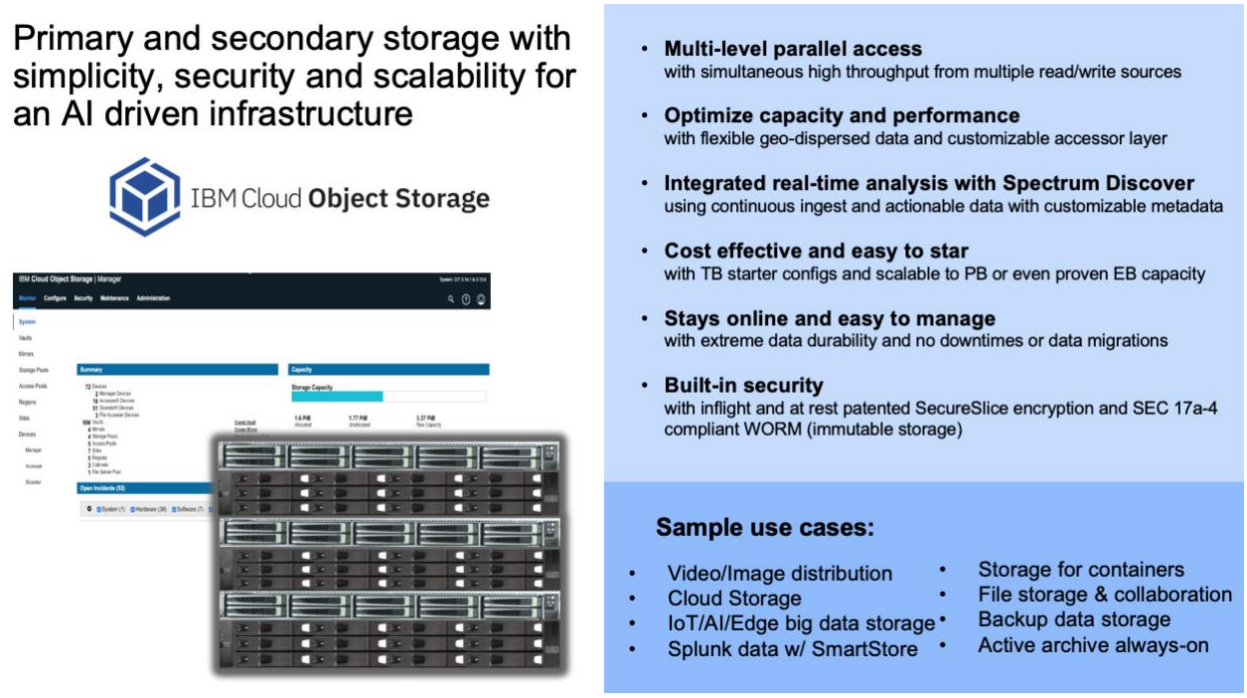
Revision History

Version	Product	Changes	Date
v1	IBM Cloud Object Storage – data center	Initial release	January 2020

Contents

Chapter 1. Storage innovation for AI and big data solutions

IBM Cloud Object Storage (COS) is an industry leading and cost effective storage platform for AI, Big Data primary workloads and secondary storage solutions. The system is designed for extreme scalability and always on availability with a share nothing multilayer architecture and distributed multi-site efficiency. IBM COS is also the perfect solution for traditional secondary storage requirements such as remote file storage consolidation and collaboration, active archive and easy to access backup repositories. With the ability to deploy in the data center or use as a service in the public cloud, this system is designed for hybrid flexibility. The system is easy to manage and deploy at scale and is used by customers who store many of the world's largest object storage systems. IBM COS customers have lowered the cost of their storage and according to an independent analysis, IBM COS customers find that the system is 76% more efficient to manage and these same customers have received an average 255% ROI and a 8 month payback on their investment. Storage solutions can start with as little as 72 TBs of useable capacity (24TB per node) and grow online to exabyte plus scalability with mixed capacity and multiple generations and configurations in a cluster and no forklift upgrades required. The solution enables investment protection that can save customers with short term and long term cost savings. Applications can access storage from any location using geo-dispersed protection and efficiency and concurrent and secure cloud native accessibility. Security is built in with a patented technology that can encrypt data in flight and at rest with automated or customer provided keys and a RESTful S3 API that provides an air gapped solution to protect against physical breaches. Data remains available with and protected from failure without replication, snapshots, complex upgrades or complex expansions. Leveraging the integration of IBM Spectrum Discover users can quickly and easily search, tag and analyze data with ease. With over 800 technology patents driving innovation and new fully configured options, our architecture is designed for customers demanding workloads today and a foundation for the growing requirements of modern applications in the future. **IBM Cloud Object Storage is turning storage challenges into business value.**



IBM Cloud Object Storage Summary

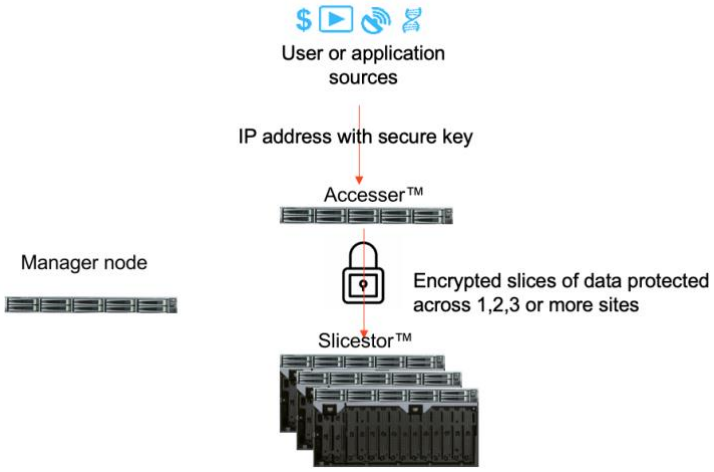
The Cloud Object Storage System can be deployed in multiple configurations in the data center with multiple levels of optimization. Customers can also utilize our technology as a service in the IBM Cloud. The system can be deployed for a single use case such as secondary storage for an online access to archive and backup data or due to its unique architecture it can be deployed for multiple use cases simultaneously including AI and Big Data workloads. The system writes all the data securely and with full data protection including the metadata to ensure the best performance at any scale from terabytes to exabytes. Our customers continue to increase their capacity and some have actually doubled their capacity over the past few years showing the confidence they have and the value they have found in our solution. We currently have multiple installations of our software with over 1EB of capacity and we have an installation of 1EB in a single namespace cluster that is continuing to expand. Customers appreciate our storage technology and scalability and they also highlight the stability and support that IBM offers to store the growing data repositories and critical data for their organization.

Chapter 2. The IBM Cloud Object Storage Advantage

IBM Cloud Object Storage Architecture

IBM Cloud Object Storage consists of three types of node. The three node types are:

- IBM Cloud Object Storage Manager
- IBM Cloud Object Storage Accesser®
- IBM Cloud Object Storage Slicestor®

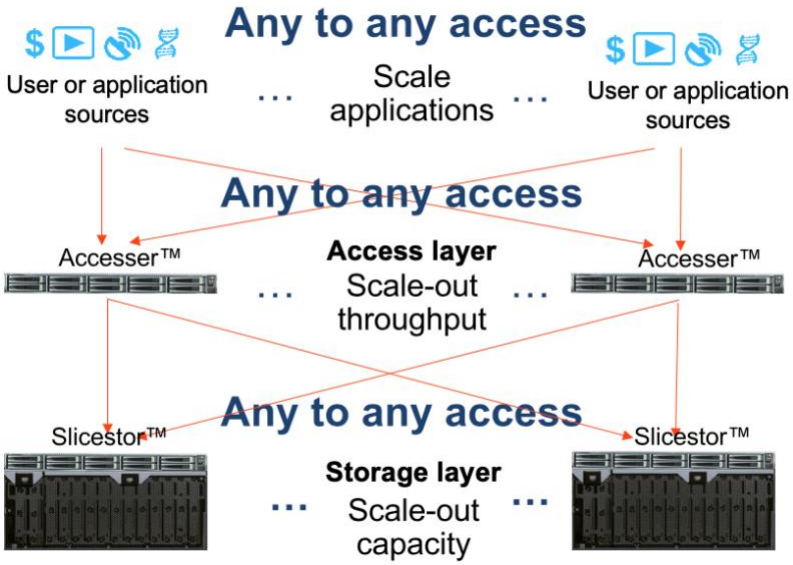


A simple but flexible architecture

Each Cloud Object Storage System has at least one Manager node, which provides out-of-band configuration, administration and monitoring capabilities. There is also one or more Accesser nodes, which provide the storage system endpoint for applications to store and retrieve data. Finally there are three or more Slicestor nodes, which provide the data storage capacity for the Cloud Object Storage System. The Accesser is a stateless node that presents the storage interface of the Cloud Object Storage System to client applications (via an IP address) and transforms data using an information dispersal algorithm (IDA). Slicestor nodes receive data to be stored from Accesser nodes on ingest and return data to Accesser nodes as required by reads.

The Manager and Accesser can be virtual (VM or containers) or can be physical devices. Slicestor nodes are currently only supported as physical devices and can deployed on optimized IBM hardware or most any server vendor such as Cisco, Dell, HPE, Lenovo, Seagate, Supermicro or others. Each Cloud Object Storage System or cluster has a single or dual (clustered for high availability) Manager node, which provides out-of-band configuration, administration and monitoring capabilities. There is also one or more Accesser nodes, (physical or virtual) which provide the storage system endpoint for applications to store and retrieve data. Accesser nodes can scale out for increased performance and throughput without incurring the cost of additional storage. There is no data or metadata stored in the accessor but it performs the processing or performance layer for the system. The final node type is the capacity layer or Slicestor nodes. There are three or more Slicestor nodes in each system. These nodes provide the data storage capacity for the Cloud Object Storage System.

The IBM Cloud Object Storage software is compatible with a wide range of servers from many sources, including a physical or virtual appliance. In addition, IBM conducts certification of specific servers that customers want to use in their environment to help insure quick initial installation, long-term reliability and predictable performance.



A shared nothing any to any to any architecture

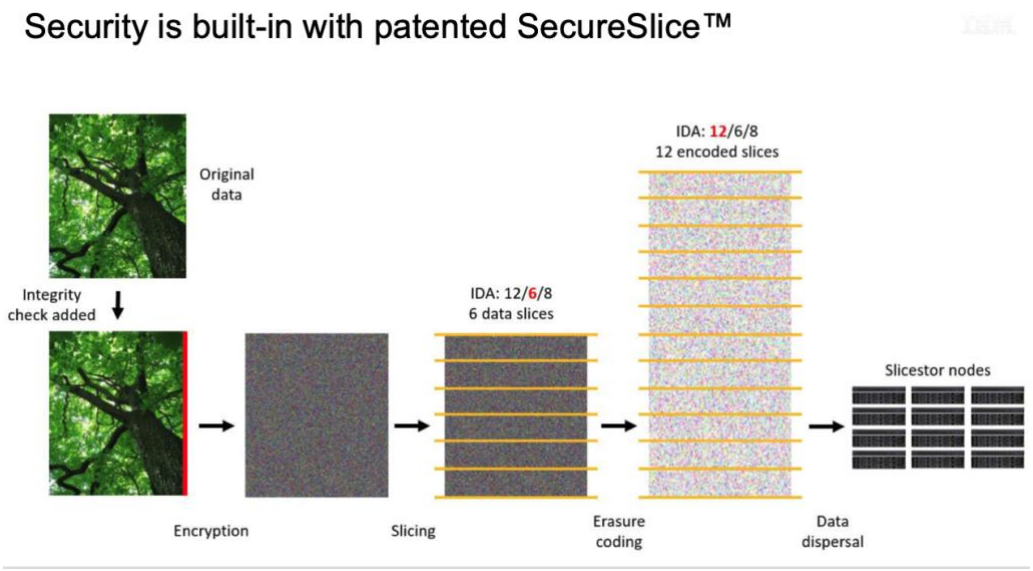
The Accesser is a stateless node that presents the storage interface of the Cloud Object Storage System to client applications and transforms data using an information dispersal algorithm (IDA). Slicestor nodes receive data to be stored from Accesser nodes on ingest and return data to Accesser nodes as required by reads.

The IDA transforms each object written to the system into a number of slices, such that the object can be read bit-perfectly using a subset of those slices. The number of slices created is called the IDA width. The number required to read the data is called the IDA read threshold. The difference between the width and the read threshold is the maximum number of slices that can be lost or temporarily unbailable while still maintaining the ability to the read the object. For example, if a system with a width of 12 and a read threshold of seven, data can be read even if five of the 12 stored slices cannot be read.

IBM Cloud Object Storage Built-in Security

Transmission and storage of data is inherently private and is designed to withstand attacks from the outside and within. No copy of your data resides in any single disk, node or location. Data in motion is encrypted

using total layer security (TLS). Data at rest is encrypted using SecureSlice encryption. Data can further be encrypted from the application or user to ensure security is never compromised.



Patented SecureSlice™ with customizable configurations

SecureSlice encryption provides high-level confidentiality for data at rest on Slicestor storage nodes as long as no more than N Slicestor nodes have their data exposed, where N = IDA Read Threshold-1. In typical Cloud Object Storage System deployments, N ranges from four to 25 depending on scale and configuration. SecureSlice is a standard product feature, with no additional license fee. It can be configured to use any of the following combinations of encryption and data integrity algorithms:

- RC4-128 encryption with MD5-128 Hash for data integrity
- AES-128 encryption with MD5-128 Hash for data integrity
- AES-256 encryption with SHA-256 Hash for data integrity

TLS is supported on network connections within the Cloud Object Storage System for data-in-motion protection. TLS is supported on Client-to-Accesser networks for data-in-motion protection. Multiple authentication methods are supported for data and management access:

- Internally managed username and password
- Active directory or Open LDAP server
- S3 secret access key
- Public key infrastructure (PKI) certificate and private key
- One user may authenticate using:
- Username and password
- Certificate and private key

Critical configuration information is communicated in a security-enhanced manner or digitally signed to prevent a potential outsider from assuming an administrator's role. For virtually any bucket, a user may be granted owner, read/write or read-only privileges. Buckets may also be configured with classless inter-domain routing (CIDR)-schemed IP Access restrictions. When bucket security is not desired for a bucket, it may be configured as anonymous read or anonymous read/write. This feature enables access to content in a bucket without authentication.

Object-level Access Control List (ACL) support in the S3-compatible Cloud Storage Object API, enables the association of an ACL with each individual object.

Role-based Access Control in the Cloud Object Storage Manager provides managed restriction of access to functionality by role for the following six roles:

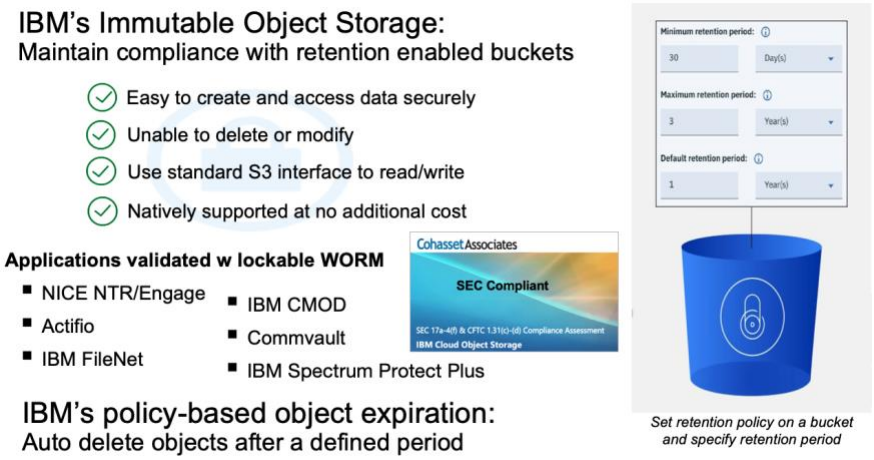
- Super user
- System administrator
- Security officer
- Operator
- Vault provisioner
- Vault user

Lockable Read-only Retention Buckets

For companies that require the ability to store data with specific policy based retention rules, retention periods can now be set. See Figure 5.This capability allows for the creation of data vaults designed per government mandated compliance SEC Rule 17a-4(f) to help meet the requirement that “electronic records must be preserved exclusively in a non-rewritable and non-erasable format.” Once set, the data cannot be overwritten or deleted. IBM will enforce the controls and will protect data per the controls (including a predefined retention period) by an application or bucket creator. Setting retention is done with a simple click on a screen or using an S3-compatible API.

One cannot change the data after it is set. However, you can still alter and configure the access controls that are not related to compliance by using a separate read only access policy. For example, you can grant read access to business partners or designated third parties (as sometimes required by regulations). In most cases, a vault should be created, a protection level applied, and then data is uploaded to the vault where it will be governed by the designed policy.

Locking down data (no deletion or changes) for a specific period of time (retention period) starts by creating a vault either with the GUI or through a REST API. A default retention period along with a minimum and maximum retention period can be initially set. Retention periods can also be set at the object level through the REST API. Access to objects either creating or reading objects is still using the S3-compatible API.



. IBM’s Immutable WORM Object Storage

Availability and Reliability for Always-on Long-term Data Storage

In an Cloud Object Storage System, the reliability and availability characteristics of the system are configurable. For extremely demanding applications, reliability of 15 nines or more and availability of 8 nines can be provided. Customers can configure for more typical levels of reliability and availability and potentially achieve economic savings as a result. The configurability of the Cloud Object Storage System allows you to choose the combination of reliability, availability and economic efficiency that suits your requirements. Data durability is designed to be maintained over time by built-in integrity checking and self-repair capabilities.

The Cloud Object Storage IDA is designed to enable durable storage, helping provide reliability and availability without storing multiple copies of the data. Availability and reliability are maintained regardless of a potential failure of hard drives and other components, complete failure of Cloud Object Storage System nodes and site outage or destruction.

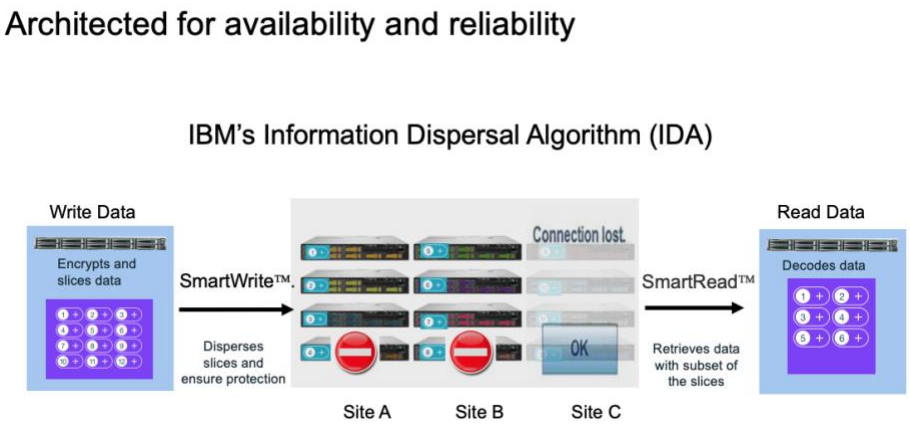
The IDA can be configured to provide high levels of reliability (15+ nines) or availability (8 nines), or to provide a lower level of reliability or availability with less physical storage capacity needed for the same usable capacity. The system allows the flexibility to use different IDA configurations for different vaults.

Distributed Rebuilder which is initiated on reads uses all Slicestor nodes in the system to identify slices that are missing or corrupt and perform the necessary repair, such as:

- Slice data lost due to failure
- Slice data corrupted by a disk-level
- Unrecoverable Read Error

Disk lifecycle management — low-level monitoring of disk health in Slicestor nodes — allows data to be moved from a failing drive to a healthy drive before the drive fails.

Multi-level data integrity incorporates checksums to handle physical media errors that often occur in large-scale storage systems. Integrity is checked at both the slice and object levels. Corrupted slices are not used and are repaired by the Distributed Rebuilder.



Information Dispersal Algorithm for availability and reliability

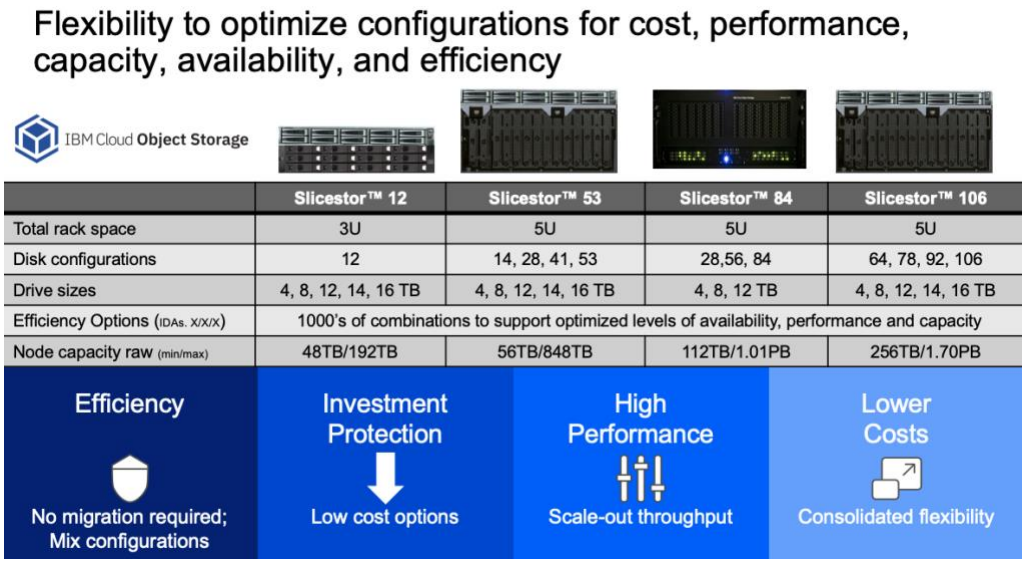
Always-on storage with high availability

For the example in the graphic, 12 slices of an object are distributed to 12 Slicestors using a patented SmartWrite™ process. Note that using our geographical dispersion the data is then written to one or more

sites and in this example we write data to 3 sites, site A, site B and site C. Now data is stored and ready for access from any of the sites with concurrent access from any and all sites. In this example we also set a read threshold of 6 which means we only need 6 slices of data to be able to read the entire object. This leads to the magic of Information Dispersal. An application or user can access any object even if an entire site goes offline, or an entire site plus 2 more nodes. ANY 6 nodes can be down and we still maintain access to our data.

IBM Cloud Object Storage Components

Since the beginning, the Cloud Object Storage System has been designed to run on x86 server hardware as software defined storage, whether it be IBM Gen2 branded appliances, or any selection from the dozens of x86 commodity hardware including two Cisco certified validated designs (CVDs) that have been tested and qualified by IBM and Cisco. An instance of IBM COS software (i.e. Manager, Accesser, or Slicestor) runs on each node or Virtual machine in the cluster. The Manager nodes consist of a CPU, RAM, and OS disks which may also consist of a virtual machine. The Accesser nodes consist of a CPU, RAM and OS disks which may also consist of a virtual machine or containerized application. The Slicestor nodes consist of a CPU, RAM, OS disks, and 6 or more data disks. Each Slicestor manages its data disks, consisting of anywhere from 6 to 106 drives depending on the configuration, as stand-alone devices having no built RAID controller. All data protection and storage control is handled in software by the IDA as well as a DLM (Disk Lifecycle Manager) process.



IBM Cloud Object Storage Slicestor nodes

The IBM Cloud Object Storage provides cost efficiencies for current and next generation applications. Our current IBM branded Slicestor nodes are built using four models of storage capacity nodes, Slicestor 12, Slicestor 53, Slicestor 84 and Slicestor 106. Each of these capacity nodes provides investment protection with no migrations or forklift upgrades required. We have optimized our efficiencies for much greater consolidation and options to scale out different capacities vs. previous models and competitors.

A 3 rack configuration providing over 40PB capacity in 1 or 3 sites

The benefit of our nodes is found in the ability to attach more, or less, storage to each controller in response to the client requirements for more or less density. The system is designed for investment protection and extreme efficiency as more options will be available to scale up by offering different size storage nodes and multiple disk enclosures per controller that will increase the efficiency of storage node for the growth of hybrid cloud and next generation applications. COS also supports multiple types of configurations within a cluster allowing it to scale-out in multiple directions.

A mixed configuration

IBM COS Expansion for Multiple or Growing Use Cases

Storage capacity is provided by a storage pool using multiple Slicestor nodes. Three or more Slicestor nodes can be grouped to create a device set which are configured to be part of a storage pool. A single Cloud Object Storage System may have one or multiple storage pools and each of the storage pools may have one or more storage device sets.

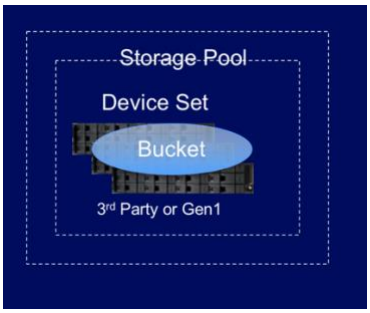
A storage bucket is not part of the physical architecture, but is an important concept in an Cloud Object Storage System. A bucket is a logical container or a virtual storage space, upon which reliability, data transformation options — for example, IBM Cloud Object Storage SecureSlice and IDA algorithm — and

access control policies may be defined. Multiple buckets can be provisioned on the same storage pool. The system can support millions of buckets and billions of objects in those buckets.

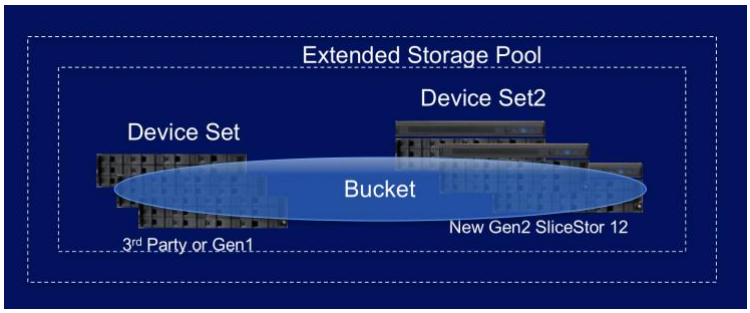
The IDA combines encryption and erasure-coding techniques to transform data to highly reliable and available storage without making copies of the data, as would be required by traditional storage architectures. By enabling reliability and availability without storing multiple copies of the data, the Cloud Object Storage System can offer significant TCO savings.

IBM COS scales up by offering different size storage nodes and soon multi-disk enclosures that will increase the efficiency of storage node for next generation applications. IBM COS also supports multiple types of configurations within a cluster allowing it to scale-out in multiple directions. A Gen2 system can start out as small as 3 nodes and grow online to an almost infinite # of nodes. The system is designed to be very easy to expand without a forklift upgrade for long term investment protection and always keeping full access to all the data (always on and available).

One can start with 3 nodes from a previous generation IBM appliance configuration or a software defined configuration using any one of our 3rd party reference options and then add 3 more nodes of the latest generation and grow the capacity online with full investment protection and without any forklift upgrades. This is very important for big data volumes that can grow to multiple PBs as systems can be upgraded or maintained on a schedule that best matches budgets or unknown variable growth rates. This can be a huge cost savings and means resources can be maintained for longer periods of time and this increases the efficiency and lowers the costs of the overall system saving thousands or even millions of dollars over time.

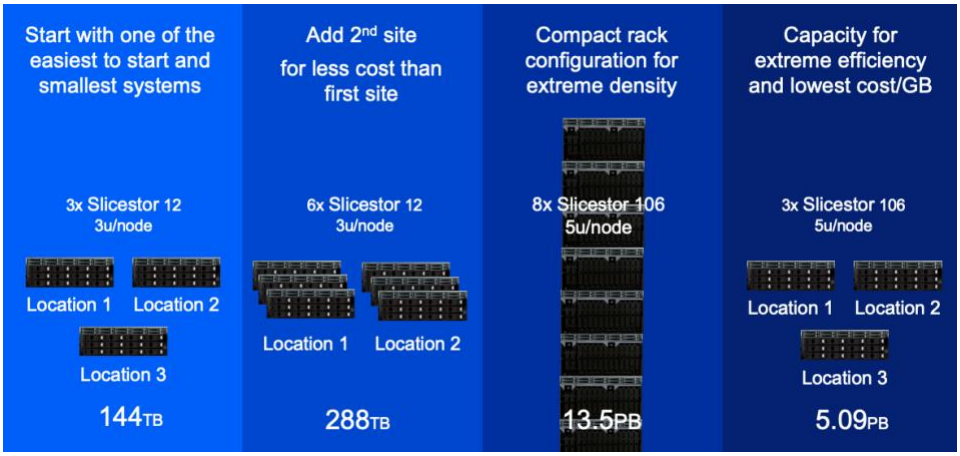


Easy to start with device set



Grow a bucket by adding device sets or create another bucket

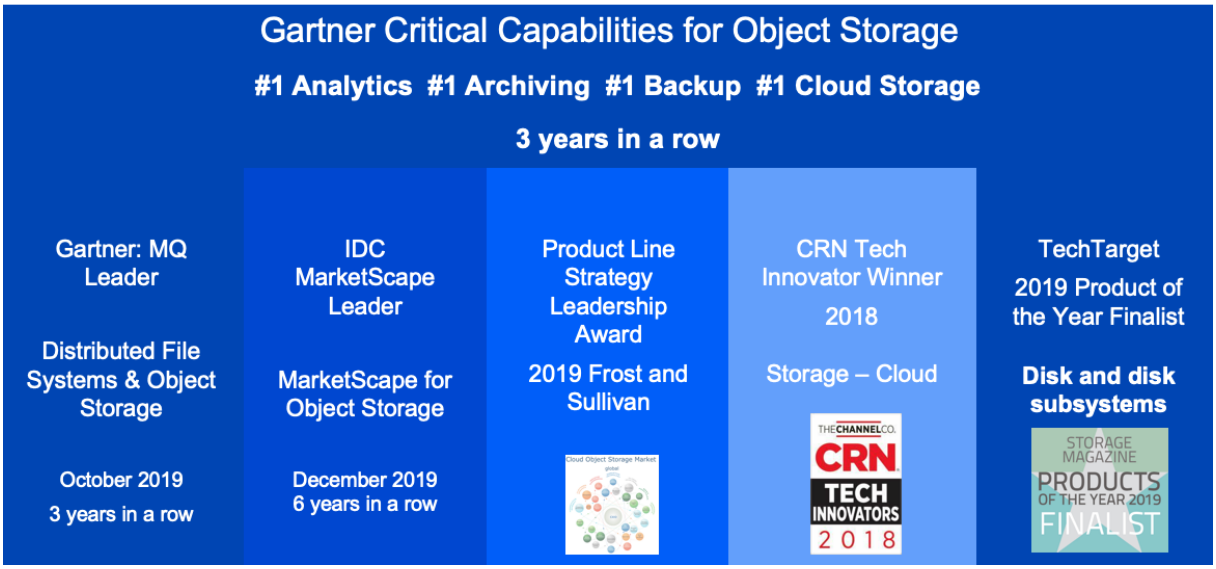
IBM Cloud Object Storage has customers that have started with 144TB and have grown to PBs and even EB of data. IBM COS also has a single cluster of 1EB of capacity deployed in production and has multiple customers with large mission critical systems over 1EB of capacity each.



Lower initial costs. Low cost growth.

IBM Cloud Object Storage is not just about lowering the cost of the storage or bringing faster and more consolidated storage to the enterprise. Our storage is also about enabling enhanced efficiencies for new and existing workloads. Offering a design for scale-up and scale-out capacity with customized expansion factors for workload optimization, our solutions drive more and more customers to rethink their storage for multiple workloads and for large scale deployments.

IBM Cloud Object Storage Industry leadership

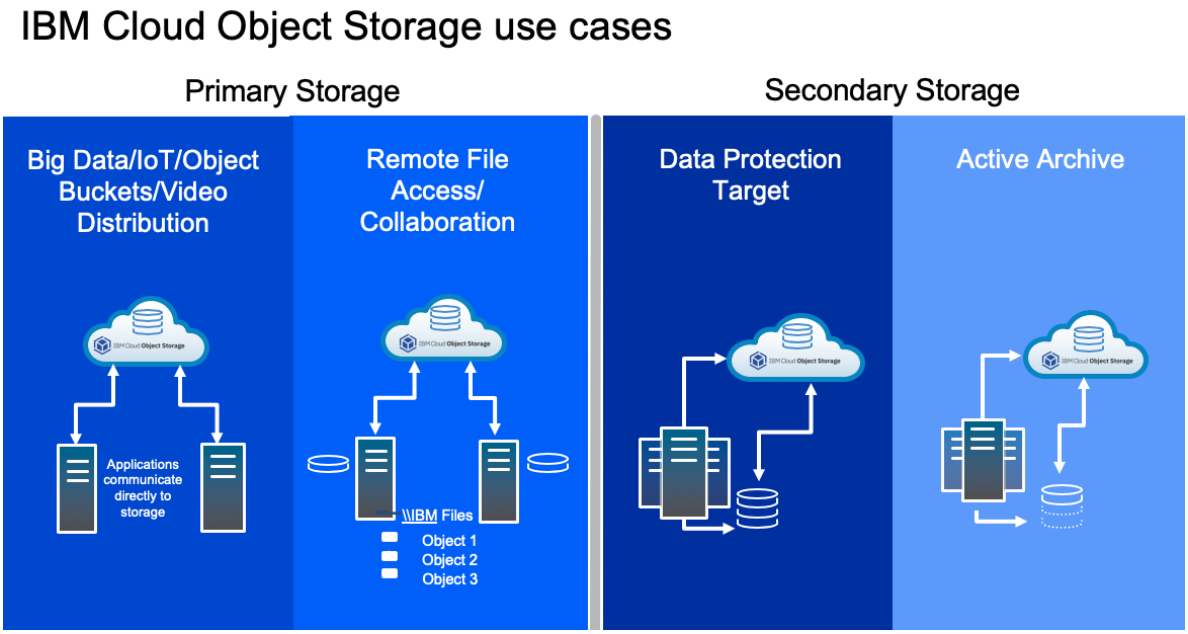


Optimized storage for multiple concurrent use cases

Chapter 3. The IBM Cloud Object Storage Use Cases and Solutions

IBM Cloud Object Storage Solution Categories

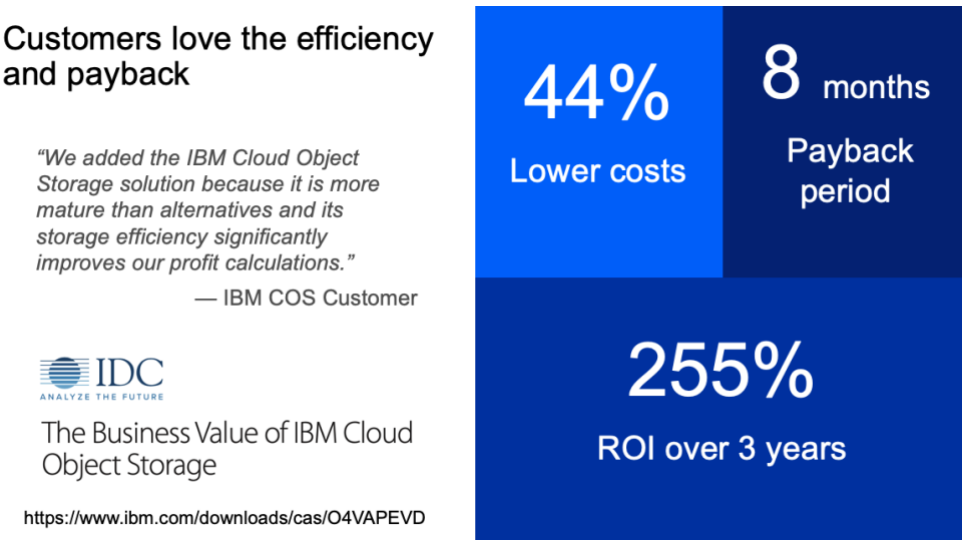
The two main solution categories for IBM COS include primary and secondary storage. The largest use case and biggest growth area for IBM COS is primary storage. This use case presents buckets to users or applications leveraging the native S3 REST API for primary persistent storage for areas such as Ai and analytics, IoT applications, video/image distribution or content storage or even for native S3 buckets for service providers. As container applications grow with the momentum of containerized environments such as OpenShift, IBM COS becomes a great solution for persistent storage with the easy to leverage S3 Rest API. The second primary storage use case is remote “file” access or file collaboration. When using file storage solutions with IBm COS, some of the data may actually be cached or treated as primary storage by the user as in the case with Panzura but the majority of the files which are less frequently access are treated more as secondary storage and do not provide high performance access.



The use cases of IBM Cloud Object Storage

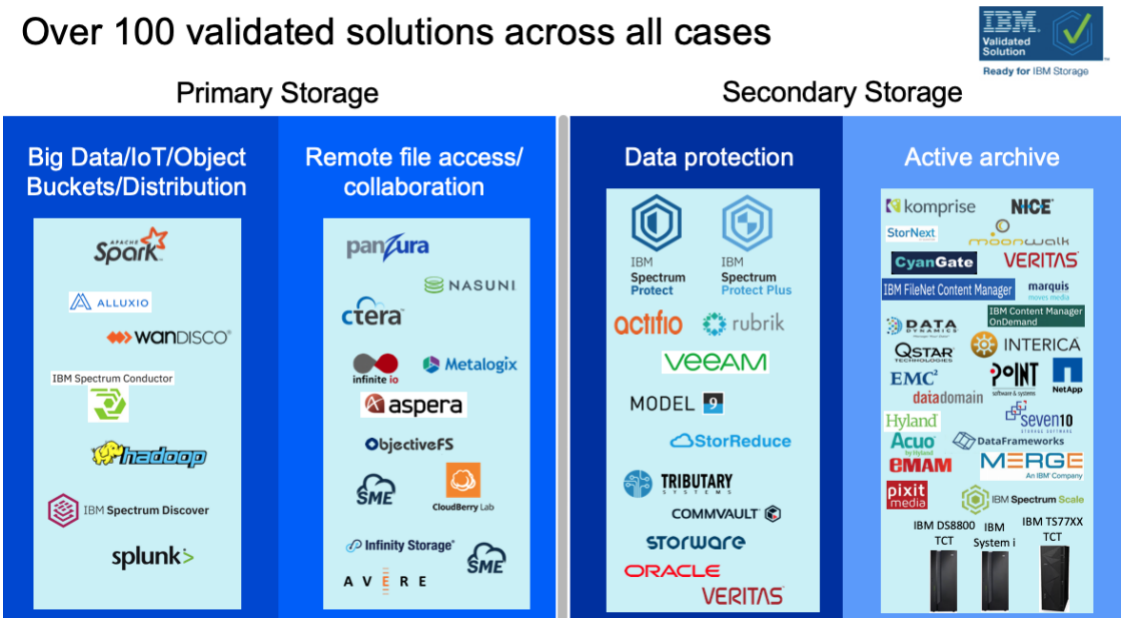
The other solution category is secondary storage. Secondary storage has been the typical use case when customers think of object storage in the data center or even in the public cloud. Secondary storage consists of using IBM COS as a data protection target for backup applications such as IBM Spectrum Protect, Commvault, Veritas NetBackup or other backup application. Another use case for secondary storage is creating an active archive to move data off of primary storage and onto a warm storage layer but one that is still easily accessible and online when needed. Some of the applications that have been tested and integrated with IBM COS for active archive include, IBM FileNet, Komprise, and Nice NTR.

Based on actual data from IBM Cloud Object Storage customers, IDC recently completed a report that analyzed the cost savings of using IBM COS in their environment. The results highlighted the efficiency and productivity gains by deploying an IBM Cloud Object Storage solution. IDC found the IBM Cloud Object Storage customers have lowered their cost an average of 44% and received payback for investment in only 8 months. The IBM COS solution also received and ROI over 3 years of 255%.



IBM Cloud Object Storage for AI enabled workloads

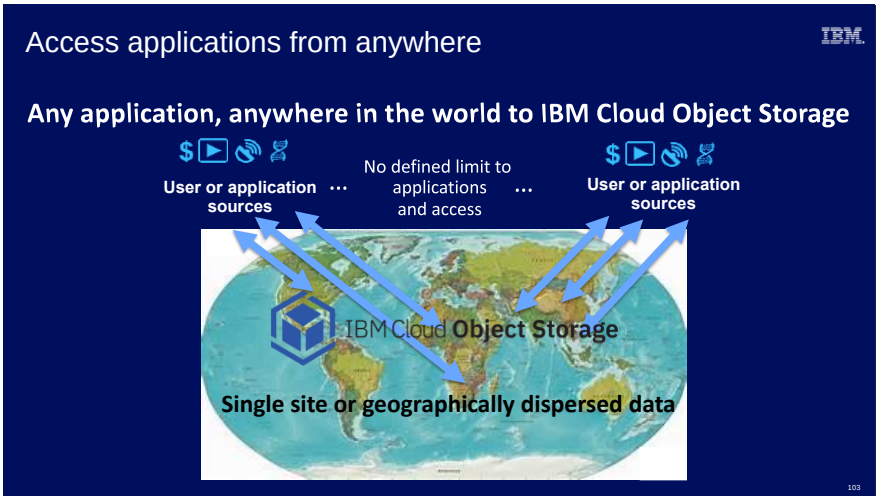
IBM Cloud Object Storage has over 100 validated solutions and we are constantly evaluating and testing more and more applications and solutions. All our validated solutions have been tested to work with our solution and we are dedicated to ensuring that our platform provides the best experience in the industry. We have a dedicated engineering team that validates application interfaces, performs performance characterization, and provides documents that can help customers in the integration with IBM COS and the IBM or 3rd party solutions.



Over 100 certified and validated solutions

IBM Cloud Object Storage Provides a Cloud Experience from Anywhere

The advantage of IBM Cloud Object Storage is data can be secured safely in your data center(s) but accessed from any location in the world as required by the business or organization. This access is controlled with a set of secure keys and can be limited or made available based on your requirements.



Object Storage is about access from anywhere

The IBM Cloud Object Storage System not only can be accessed from anywhere in the world but the system can also be access from most any application environment. Since there is no need to export or mount a file system or storage device file the system is easy to access from any operating system that can access an IP address and even new container environments like docker, kubernetes or even OpenShift.

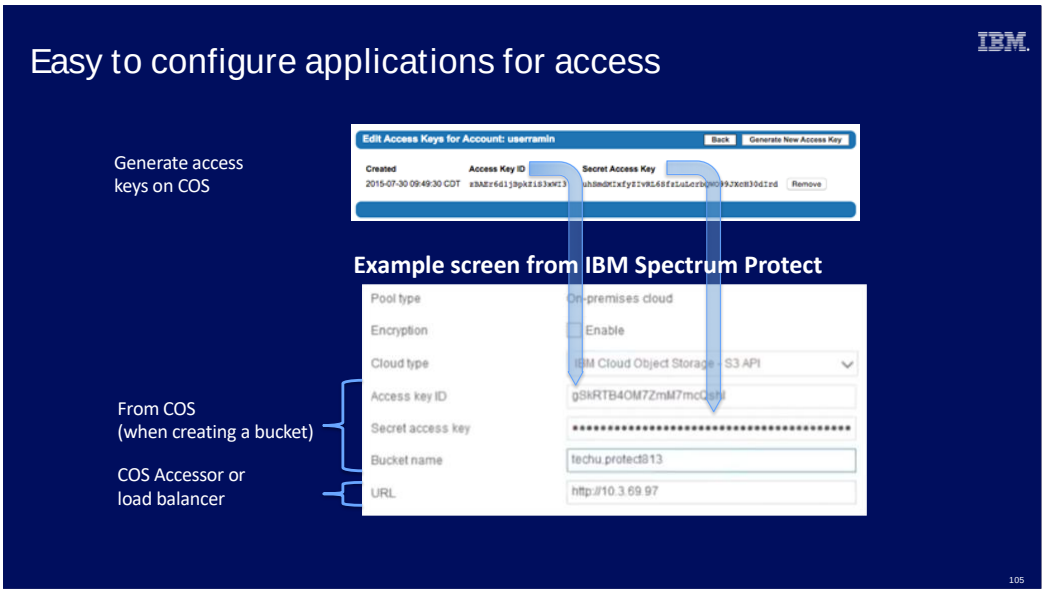
Access from any environment

IBM Cloud Object Storage is more than just storage efficiency, always on storage or the ability to storage massive amounts of data that is easy to scale. IBM COS provides cloud storage and access to data from any location with the security, control and cost savings of storage in a data center. The advantage of IBM COS data located in your data center is there are no surprises with access fees or hidden costs to access data from any applications in any location including the public cloud. Customers pay one price and access from anywhere.

IBM Cloud Object Storage secure access

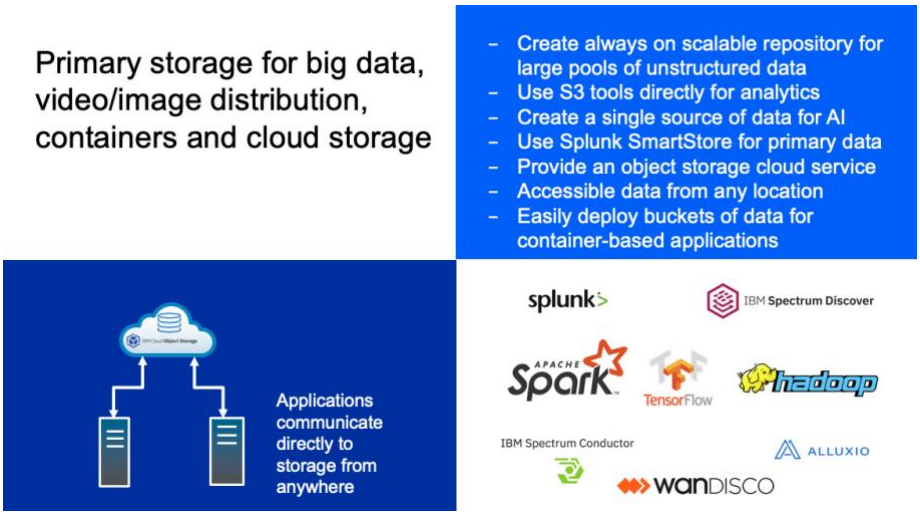
IBM Cloud Object Storage leverages the S3 Restful API which is used by the largest cloud providers and object storage systems. Using the S3 Restful API means that access to storage is not limited to an operating system like block and file storage. The advantage is that an application can access data easily without the restriction of the storage attached to an operating system. This makes is very easy for any application including applications that leverage new environments like containers that leverage Docker or

OpenShift or new application architectures like microservices. To access data, all that is required is the IP address of the accessor node or end point, the bucket name and the access key and secure key which provides the secure access to specific objects. When a bucket is created in IBM COS the system provides the access key ID and the secret key which is used by the applications to access the data. To configure an application to access an IBM COS bucket to read or write data, the IP address and bucket name along with the keys are configured for secure access. The example below show the screen from IBM COS on the top which outputs the keys when a bucket is created. On the bottom is IBM Spectrum Protect which is configured using the keys from IBM COS to store backup data.



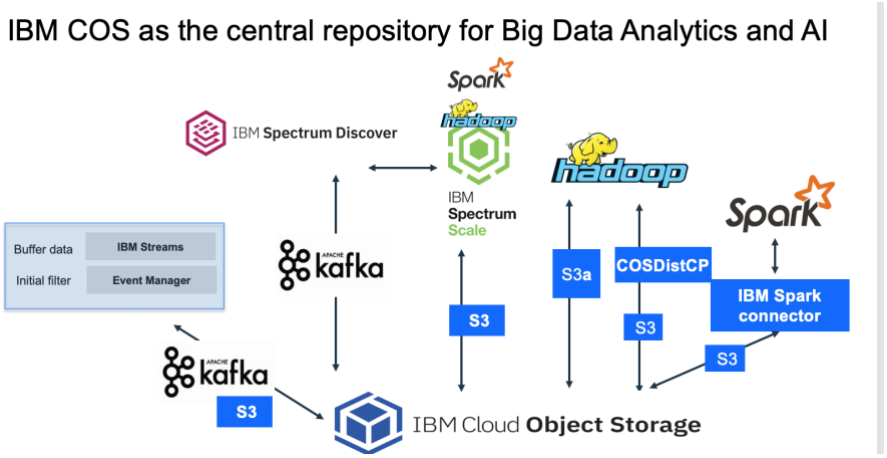
Easy to configure application access to object storage

IBM Cloud Object Storage for AI and big data primary storage solutions



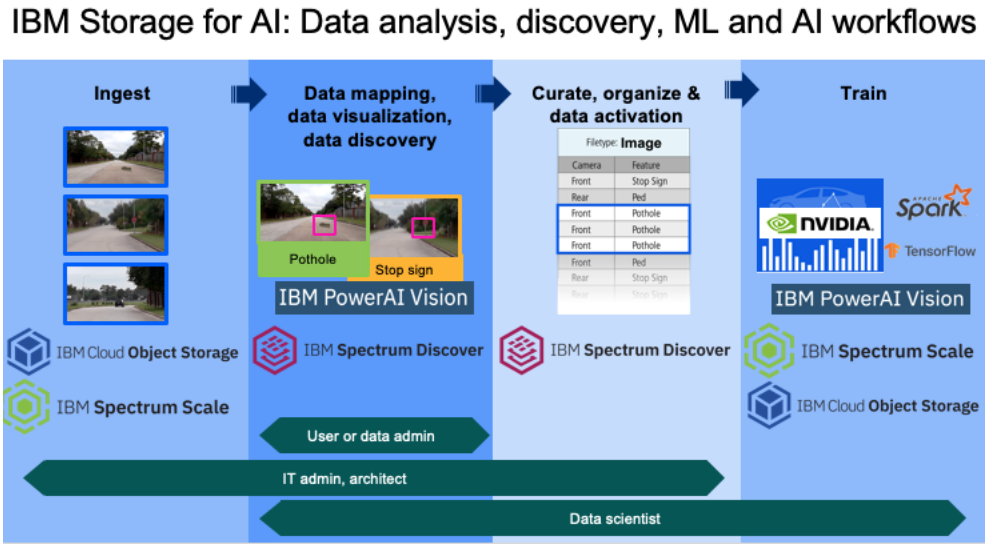
IBM Cloud Object Storage for primary AI and big data workloads

Using IBM Cloud Object Storage, organizations can build a centralized data repository, leveraging cost-effective and scalable storage that makes it possible to collect and store nearly unlimited amounts of data of any type, from any source. Data remains in its native format and doesn't need to be moved in and out of IBM Cloud Object Storage; rather, the IBM Cloud Object Storage-based data lake is the persistent data store for the analytics.



IBM Cloud Object Storage for Analytics

IBM Cloud Object Storage can be integrated directly to applications, analysis tools or even to a high speed file storage like IBM Spectrum Scale. IBM Spectrum Discover can be integrated to bring powerful insights in data and together with Spectrum Scale and IBM Cloud Object Storage create a complete infrastructure framework for transformational AI environment.



Using container platforms such as Red Hat OpenShift with IBM Cloud Object Storage is easier than block or file storage. Applications can manage the storage without a connector and just the S3 compatible REST API to access an always on persistent storage layer. If one prefers a file interface then the opensource S3FS FUSE connector is available to provide a file system interface to IBM Cloud Object Storage. Using IBM Cloud Object Storage with containers or container environments like Red Hat OpenShift is very easy. Developers creating containerized applications can use IBM COS directly using the S3 REST API without any storage connector or gateway as object storage is persistent and easy to access through the S3 REST

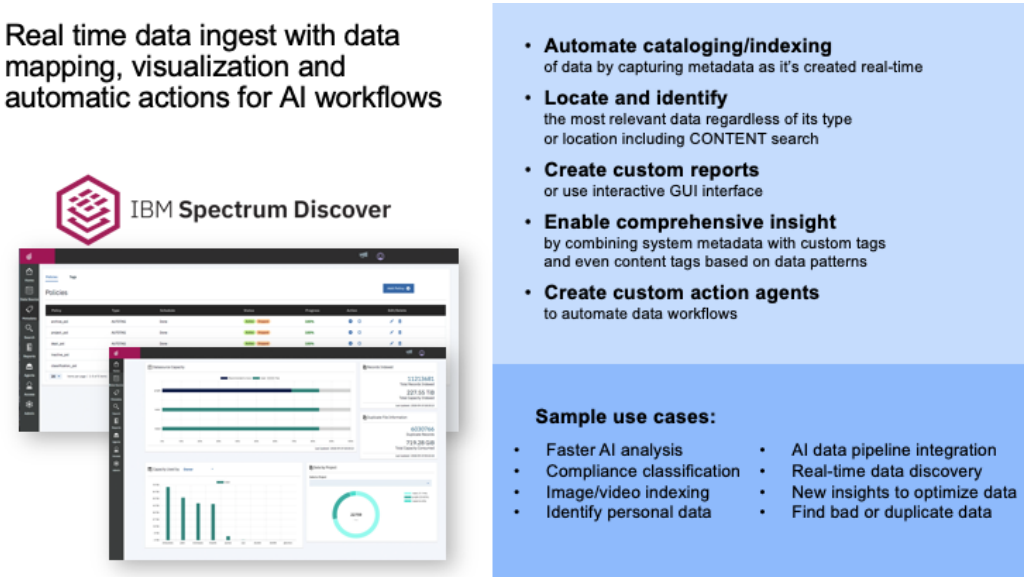
API. For traditional applications that require a file system a file access connector is available for IBM COS using the opensource S3FS Fuse module. See: https://github.com/s3fs-fuse/s3fs-fuse

As analytics and AI play an increasing role for organizations as they modernize and transform for the future, the access and efficiency of the data becomes something that brings more and more value to the organization. When integrating with Hadoop applications such as Apache Spark or MapReduce an opensource connector is available to allow an object storage system like IBM Cloud Object Storage utilize these analytics tools through a standard HDFS (Hadoop File System) interface. This interface is shipped with Hadoop distributions as an s3a connector. IBM has created an opensource connector that is optimized further for object storage called the Stocator connector and is available also as a supported interface from IBM. When using this interface Spark application have seen up to 18x improvements on some standard operations. See: https://github.com/CODAIT/stocator

When migrating large amounts of data from HDFS or Hadoop data storage an opensource tool has been created to optimize the transfer of that data to IBM COS. This is a useful tool for migrating large amounts of data to object storage. See: https://github.com/IBM/cos-distcp

Spectrum Discover – Deeper insight and analysis to IBM Cloud Object Storage

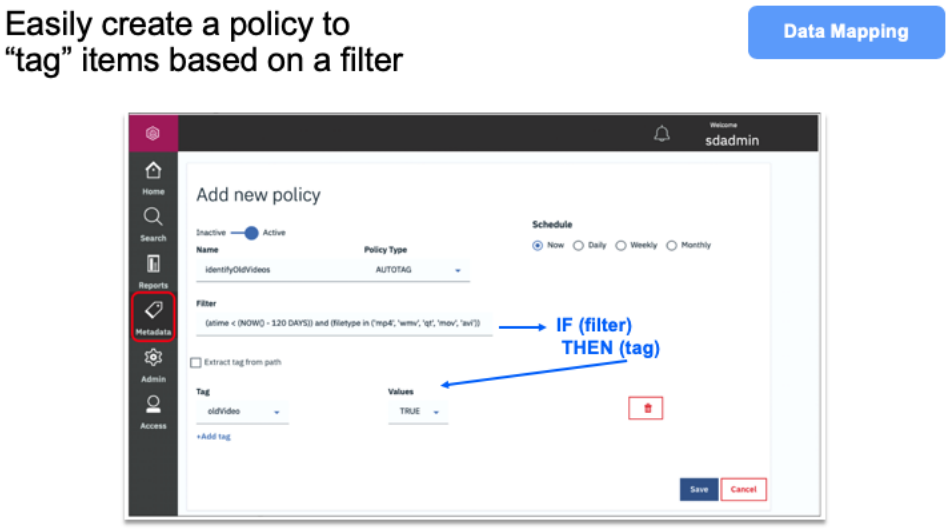
One of the most powerful AI and analytics tools to integrate with IBM Cloud Object Storage is Spectrum Discover. Spectrum Discover is a user friendly tool that automatically catalogs all the objects from IBM Cloud Object Storage organizing the metadata into a powerful and customizable database that can be analyzed and additionally be labeled for powerful insights into big data repositories. Users can locate and identify the most relevant data regardless of its type or location. Using either a simple SQL query command or actionable API scripts or commands, users can enable comprehensive insight into the data in a fast and efficient way.



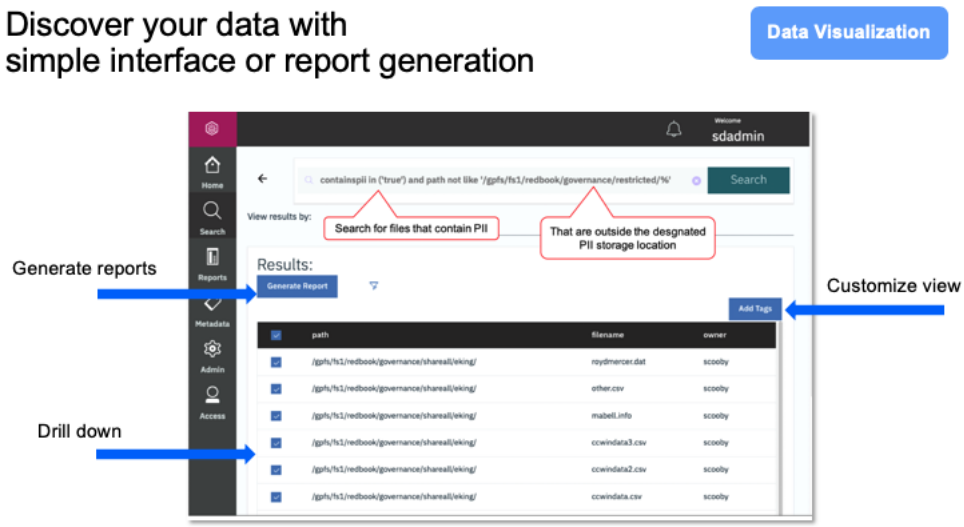
IBM Spectrum Discover

Spectrum Discover can also be used to create custom tags and policy-based workflows to orchestrate content inspection and activate data in AI, ML and analytics workflows. One key feature of Spectrum Discover is the ability to add new tags as policies and to map filters to a key or a tag. This ability allows users to create custom metadata on objects. To create a policy and tag data, a user creates a filter and

then based on the filter a tag is create on any items that match the filter and a value is added for the new tag.



Users can easily visualize data and organize/sort through millions or billions of object with ease. With Spectrum Discover, a user can use SQL like search strings and visually see the results. Results can also be customized to show different fields or information about the data. We can also generate reports that can be exported or sent to other individuals.

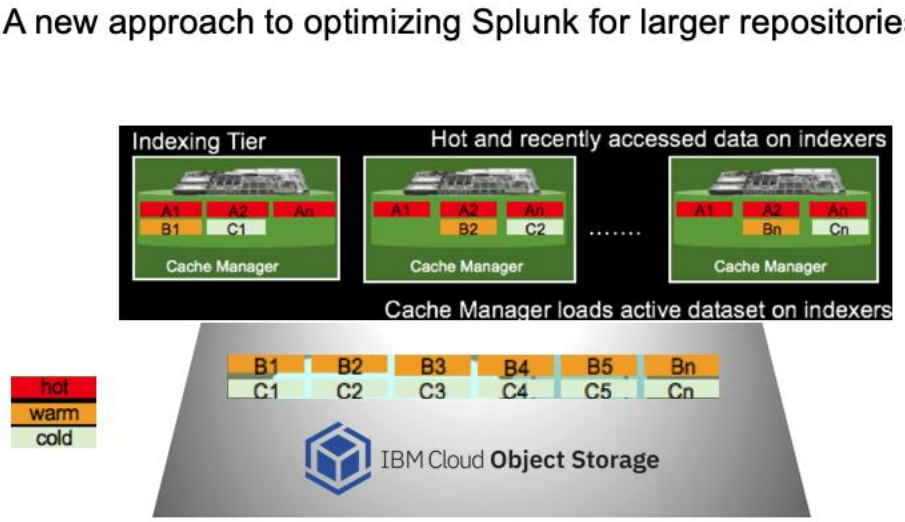


Splunk SmartStore – An AI solution that leverage IBM COS

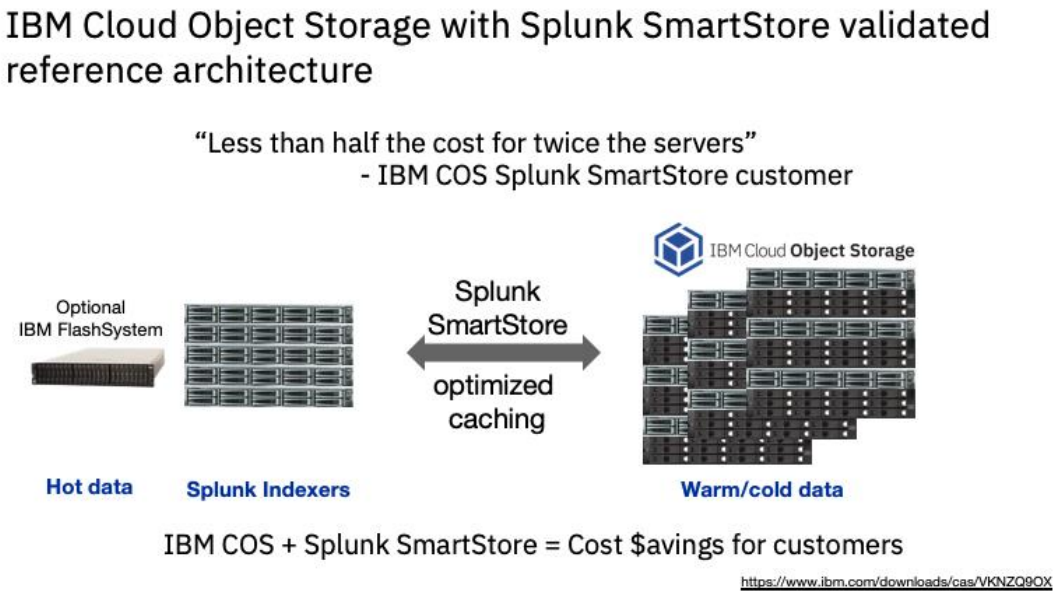
Simplify growth and lower costs

Splunk SmartStore is a new Splunk architecture that uses object storage for warm and cold data. It is a distributed scale out model that brings data closer to the compute nodes (Indexer) on-demand and provides a high degree of compute/storage elasticity and makes it incredibly cost efficient to achieve longer data retention at scale. The Computer nodes for Splunk are called Indexers and hot data or data that is recently

used is now the only data kept on compute nodes. While the indexers no longer keeps the majority of warm data on the Indexer, Splunk cache manager can quickly recall the data from the new capacity storage tier on IBM Cloud Object Storage.



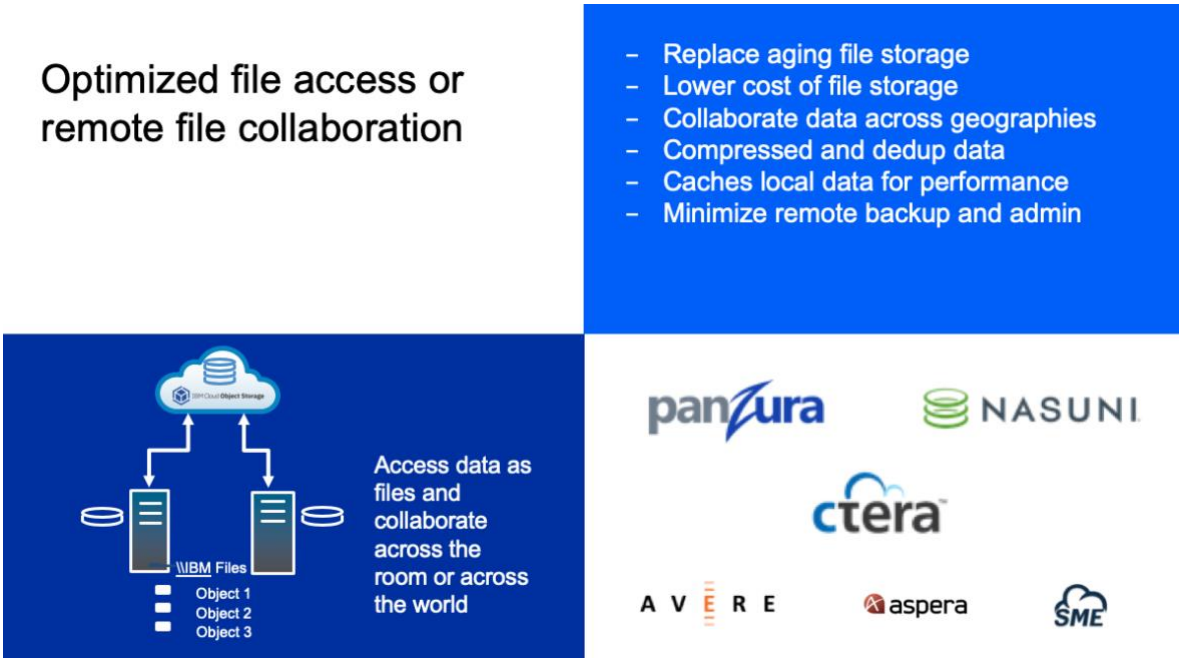
The architecture using IBM Storage with Splunk SmartStore is focused on the Indexers or the compute nodes. The compute nodes retain the active data or the hot data in a Splunk system and then mange the access to warm or cold data efficiently on an object storage system.. The Active data (hot data) on the indexers could use direct attached storage on the compute nodes or they can optionally use IBM high performance flash SAN storage like IBM FlashSystem storage. Using IBM FlashSystems provides more flexibility and control with a consolidated and extremely high-performance approach to hot data. Smart Store then leverages an object storage system like IBM Cloud Object Storage for the warm and cold data. The master copies are kept on IBM COS which lowers the amount of capacity on the expensive hot tier. Offering a combined solution with hot and warm/cold data means IBM can drive performance and efficiency in multiple ways to provide the most cost-effective solution to the customer.



IBM Cloud Object Storage for Modern File Storage

modernize file storage for new efficiency and cloud scalability

Traditional file storage is being transformed very quickly as more modern file options are being optimized for larger repositories of data and more efficient and scalable file storage solutions. If you have ever managed or tried to expand a growing traditional file-based NAS, then you can relate to the problems that traditional scale-up single attached NAS can create. Current market trends show that changes are occurring and traditional NAS has seen better days. IDC estimates that scale-up file storage is shrinking and expected to decline 5.0 percent CAGR for 2017-20221. File storage can start out very cheap and economical, but unstructured file storage is still growing very fast. The problem is that managing duplicate copies of files and balancing backend storage and tuning controller nodes becomes a challenge and the NAS can become complex and expensive very fast. Customers need a solution built for scale for modern applications and workloads with secure remote access to data.



IBM Cloud Object Storage for remote file access and collaboration

IBM is focused on helping customers modernize file-based environments and help customers prepare for and utilize more modern workloads like AI and analytics and remote collaboration and file access.

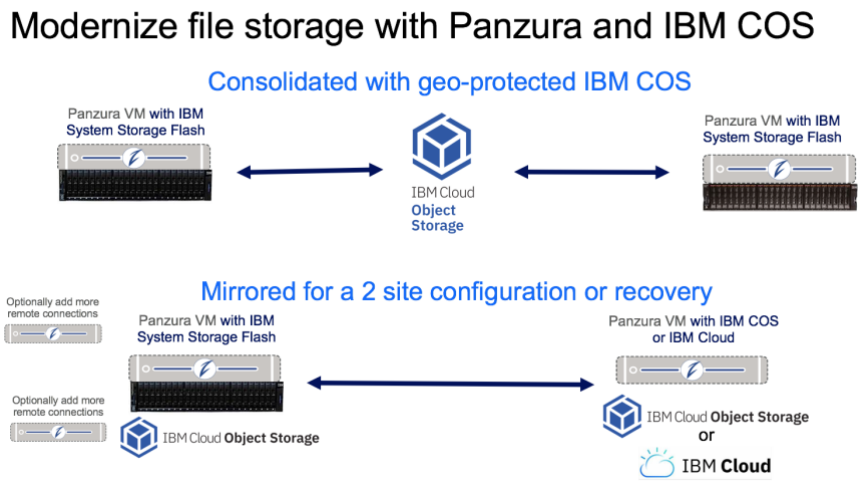
IBM has a number of key partnerships and solutions for file service that are integrated and validated with IBM Cloud Object Storage. IBM Spectrum Scale provides a high performance file system that can easily move data that is less frequently used to a warm capacity tier using IBM Cloud Object Storage. IBM has two partner solutions that provide a unique integrated file interface using IBM COS. Panzura and their Panzura Freedom™ NAS software provides a local high performance file tier that enables global file optimization and CTERA file services with their secure large scale collaboration, backup and security services. These solutions have been validated with IBM Cloud Object Storage and can now be sold directly by IBM and partners to create a solution that is optimized for file access with the scalability and cost efficiency of an IBM COS solution. Other file services solutions include Nasuni, SME, Cloudberry and for faster transfer of file data across the network IBM COS supports the integration of IBM Aspera with their high speed data transfer.

Why IBM Cloud Object Storage is best for modern file services?

- **Fast data access for local data and optimized storage for less frequently used data**
  Keep active data on flash and dedup/compress file data for increased efficiency on IBM COS with Panzura
- **Single global namespace across multiple sites leveraging the simplicity of IBM COS**
  With geo-dispersed IBM COS data can be concurrently accessed across multiple sites
- **Consolidating multiple data silos for lower operational and capital costs**
  Leverage IBM COS any to any to any architecture with share nothing access nodes and online upgrades
- **Secure data for extended periods**
  Data encryption with SecureSlice and optionally manage your own keys for large volumes of data
- **Scale on demand when needed**
  Easy to add more capacity to EB+ without forklift upgrades or downtime
- **Durability to last a lifetime**
  Objects/files can last millions of years with 99.9999999999999 of durability
- **Resiliency and always on availability of data**
  Data can withstand multiple outages or even a site disaster w/ up to 99.999999 availability
- **Cost efficient for on line data**
  Save cost vs keeping data on expensive primary storage and be ready to use your data when needed

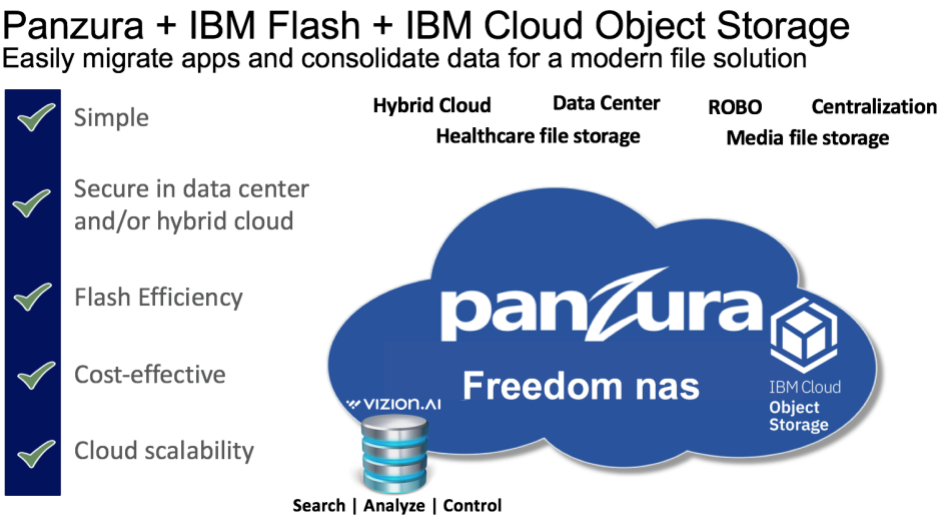
Replace Multi-site File Storage (Netapp Dell/EMC etc.) with Panzura, IBM Flash and IBM COS

Panzura Freedom™ hybrid cloud storage helps you rapidly deploy cloud-based file services with IBM Cloud Object Storage. PCFS provides global NFS and SMB/CIFS access with Cloud Object Storage that you can use to store, manage and protect all unstructured data. The Panzura Freedom family provides a secure end to end encrypted hybrid cloud storage solution with intelligent caching, deduplication, compression, snapshots and global byte range file locking that is trusted by government agencies.



Modernize file storage with IBM COS and Panzura

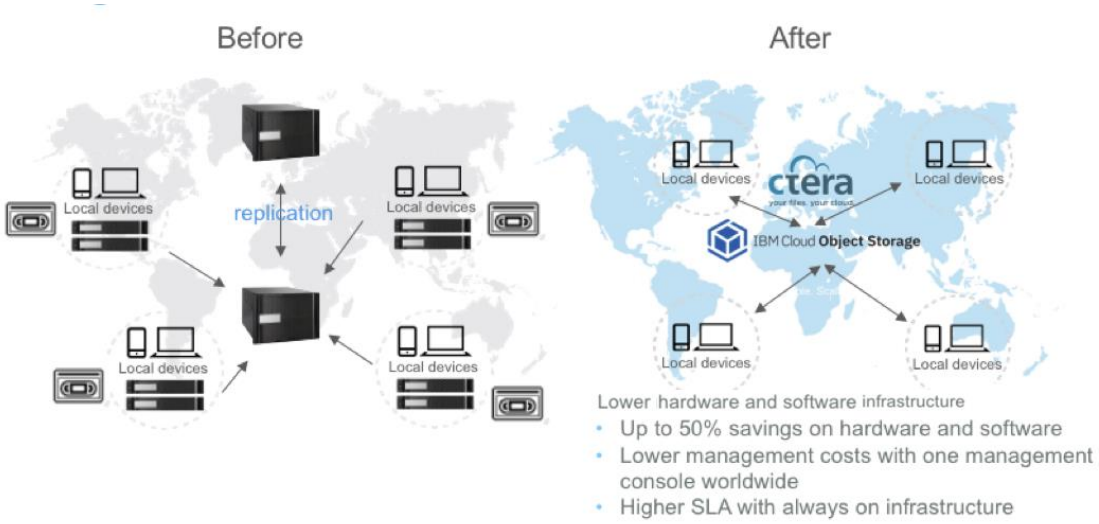
IBM and Panzura offer all the features of an enterprise file system with the economics scalability and simplicity of cloud storage. Combining Panzura Freedom NASTM with IBM Cloud Object Storage provides improved cloud economics and local performance for data in the cloud and also assures your data is secured and protected. Panzura builds an intelligent hybrid cloud storage solution with Cloud Object Storage. It consolidates your unstructured data on cloud storage while keeping active data cached close to the users. IT staff will find it behaves like local file servers and NAS, but is based in a cloud data repository and can support hundreds of offices.



IBM COS and Panzura Freedom NAS

Push the Easy Button for File Collaboration using CTERA with IBM COS to Maximize Security and Minimize Costs

CTERA's Enterprise File Services Platform is the only cloud-enabled, managed-storage delivery solution that combines enterprise endpoint backup software, file sync and share software, and cloud storage gateway appliances to address enterprise user file access. CTERA helps customers to securely sync, share, protect and govern data everywhere using IBM Cloud Object Storage.



IBM COS brings simplicity with CTERA for collaboration

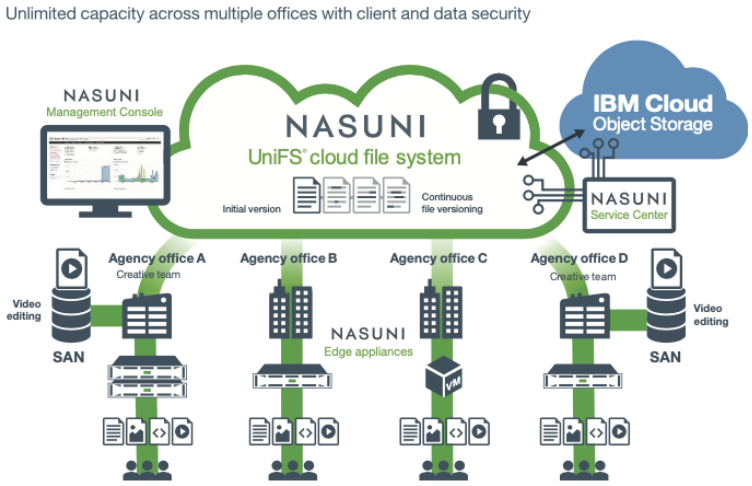
IBM Cloud Object Storage with the CTERA Enterprise File Services application provides two primary use cases for the enterprise:

- Enterprise file sync and share (EFSS) and file collaboration
- Remote data protection and backup for endpoints and servers

IBM and CTERA provide easier scalability and management designed to supply the convenience of always-on storage while also supporting enhanced data security and regulatory compliance. The solution also has management controls to help limit costs. Compared with traditional data-center storage solutions, this solution provides more business agility by empowering line-of-business operations to deploy storage more quickly and simply on their own instead of depending on a storage administrator. The CTERA Enterprise File Services platform helps you preserve data security and better manage compliance and regulatory issues—requirements that can be difficult to meet when using other public cloud services.

Speed Collaboration with Faster, more Flexible Storage and File Sharing using NASUNI

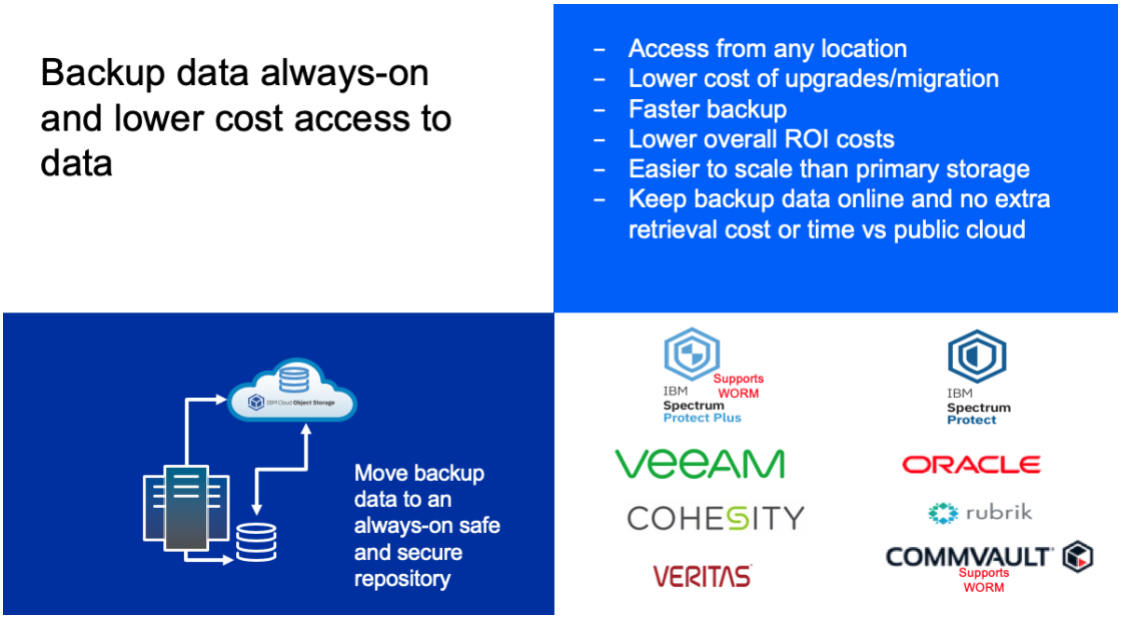
Nasuni (NAS Unified) is a leading provider of cloud-scale enterprise file services. Powered by UniFS, the innovative cloud-native file system, Nasuni capitalizes on IBM Cloud Object Storage to integrate and surpass the silo capabilities of traditional NAS, distributed file systems, data protection software, file synchronization and share solutions and disaster recovery infrastructure. With Nasuni's virtually unlimited primary storage capacity, industry-leading recovery points and recovery times, global file access and unmatched scalability, enterprises can meet "cloud first" and business growth objectives, extract more value from their file data and transform expensive storage infrastructure into a more affordable, as-a-service asset.



IBM Cloud Object Storage for Modern Data Protection

Warm storage for data management and back up data

Data protection is an important part of an organizations data strategy. Storing that data has typically been on offline tape storage or expensive primary storage. Storing data to public cloud resources means that data must leave your data center(s) and can create security concerns or issues meeting SLA requirements. One of the fastest growing traditional use cases for IBM Cloud Object Storage is using IBM COS as a warm secondary storage for a modern data protection solutions. As most data management and backup applications now leverage the S3 object interface to IBM COS, customers are now able to create a more efficient and simplified approach to data protection and maintain control and leverage the efficiencies of a modern storage solution.



IBM Cloud Object Storage has been validated for all the major leading backup applications

One of the easiest and fastest ways to start with object storage is to use IBM COS as a backup target for data protection applications like Spectrum Protect, Spectrum Protect Plus, Commvault, Veeam, Rubrik, etc. Typically this replaces keeping backup data on primary storage or a backup appliance like Dell/EMC Data Domain or in some cases replacing tape when the customer has made a decision to move off tape. Note that IBM COS is not intended to provide a lower cost solution than tape but can provide a solution that does not require any migration cost and provide faster access to data and access from any location.

IBM Cloud Object Storage has been validated by our dedicated solutions engineering team for the leading backup and data protection applications and vendors. Validation makes the transition to IBM COS easier and with much less issues in the transition to an always online access to your data. We have customers who are backing up data from 100s of TB to 100s of PB and our experience has helped make the easy transition to storing backup data on IBM COS. Many times its as easy as setting the IP address of the Accesser node or load balancer along with the bucket name and secret key. Once set up the backup applications store data on IBM COS which can be deduplicated or compressed at the source for even greater storage efficiency.

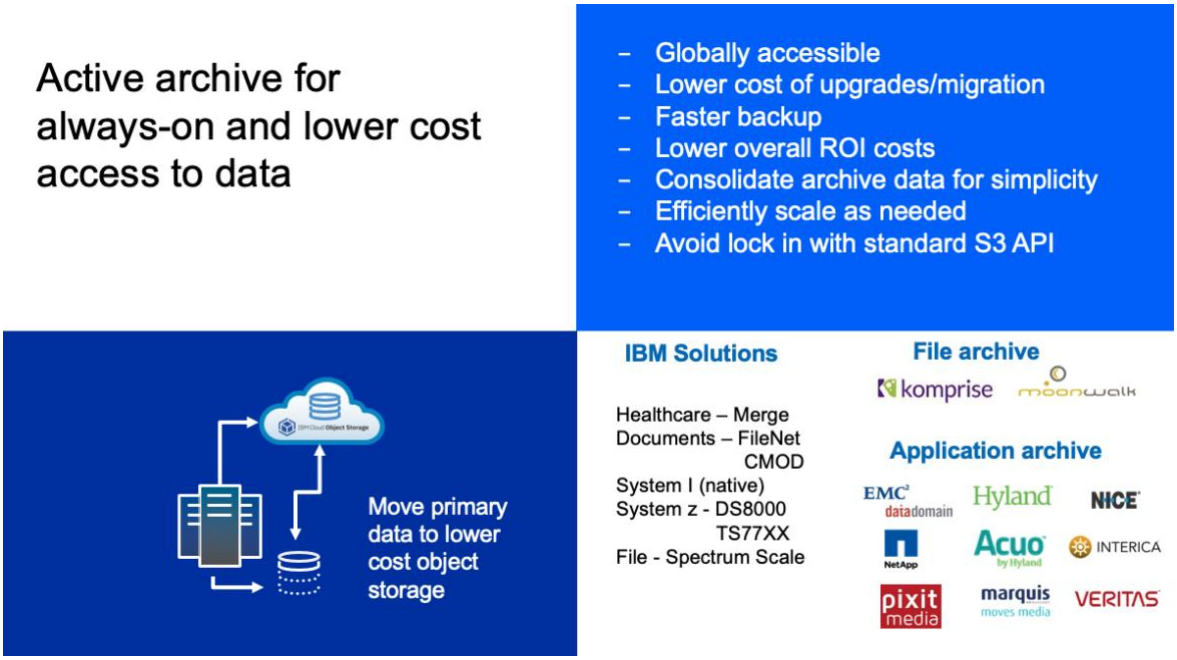
Why IBM Cloud Object Storage is best for data protection?

- **Fast data recovery**
  Data is easily available from any location and is always on line
- **Secure data for extended periods**
  Data encryption with SecureSlice and optionally manage your own keys
- **Scale on demand when needed**
  Easy to add more capacity to EB+ without forklift upgrades or downtime
- **Durability to last a lifetime**
  Objects can last millions of years with 99.9999999999999 of durability
- **Resiliency and always on availability of data**
  Data can withstand multiple outages or even a site disaster w/ up to 99.999999 availability
- **Cost efficient for on line data**
  Save cost vs keeping online data on primary storage and be ready to use your data when needed

IBM Cloud Object Storage for Online Archive of Primary Data

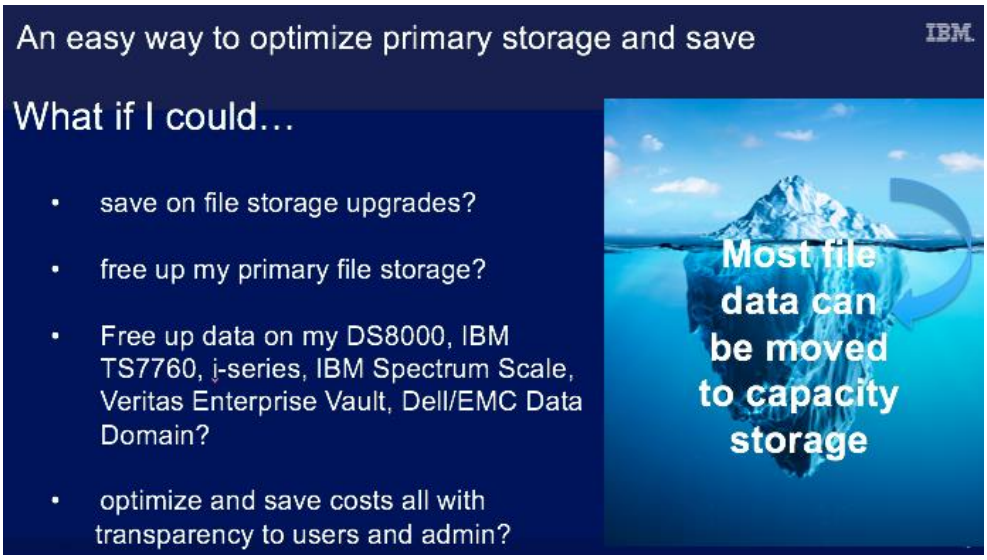
Warm storage for data that does not have to be kept on expensive primary storage

Archiving data to IBM Cloud Object Storage is all about improving efficiency and lowing cost and at the same time creating an always on access to the expanding amount of enterprise and user data. Enterprises are replacing traditional archives with IBM Cloud Object Storage to reduce cost and complexity while keeping data available. With low-cost warm and cold storage options, and with a broad ecosystem of integrated partner solutions, it's easier than ever to move and store archive data in IBM Cloud Object Storage. You can realize cost savings directly related to reducing IT infrastructure and operational costs and make data that was once out of reach accessible for your organization. Keeping data in a centralized big data system can also allow leveraging that data for AI and Analytics to bring additional value from your data to the business or organization.



IBM Cloud Object Storage for Active Archive

Since IBM COS provides a storage solution that can be geo-dispersed across data centers and is easier to scale and easier to manage than primary storage it provides a great solution for less frequently used data. There are a growing list of IBM solutions that support using IBM COS as an active archive including most recently added IBM Systems i. One easy solution that can be easily justified is moving data off of primary file systems like Netapp or Isilon or even a windows server using solutions such as Moonwalk that uses native file interfaces or Komprise that uses symbolic links to archive files off of primary file systems.



IBM Cloud Object Storage for active archive

Active archives take advantage of metadata in order to keep track of where primary, secondary, or even tertiary copies of data reside within the system. This allows for optimizing storage resources and placing data in the best location for performance, costs or other efficiency options. With the growing amount of unstructed data in the data center and even in the cloud the need for archive applications is growing each year.

IBM COS active archives provide online access, searchability and retrieval of long-term data and enables virtually unlimited scalability to accommodate future growth. In addition, active archives enhance the business value of the data by enabling users and applications including AI enabled workloads to directly access the data online, search the data and use it for business purposes.

Why IBM Cloud Object Storage is best for active archive?

- **Fast data access for archives**
  Data is always on line and ready to access from anywhere
- **Consolidating multiple data silos for lower operational and capital costs**
  Leverage IBM COS any to any to any architecture with share nothing access nodes and online upgrades
- **Leverage online accessibility to archives for new ways to create value from data including AI**
  Data is available from any location and is always on line
- **Cyber secure repository**
  Data is encrypted when received and not attached to an operating system and accessible only from API
- **Secure data for extended periods**
  Data encryption with SecureSlice and optionally manage your own keys
- **Scale on demand when needed**
  Easy to add more capacity to EB+ without forklift upgrades or downtime
- **Durability to last a lifetime**
  Objects can last millions of years with 99.9999999999999 of durability
- **Resiliency and always on availability of data**
  Data can withstand multiple outages or even a site disaster w/ up to 99.999999 availability (always on)
- **Cost efficient for on line data**
  Save cost vs keeping data on expensive primary storage and be ready to use your data when needed

Archiving files from Netapp, Isilon and other NAS to IBM COS

As unstructured data and especially file data continues to put strain and added costs for many enterprise companies and organizations the need find more efficient ways to lower costs but maintain access becomes important. For file data located on Netapp NAS systems or Dell/EMC Isilon systems or other file storage, IBM has partnered with Komprise to create a seamless way to transparently move and archive less used data to IBM Cloud Object Storage.

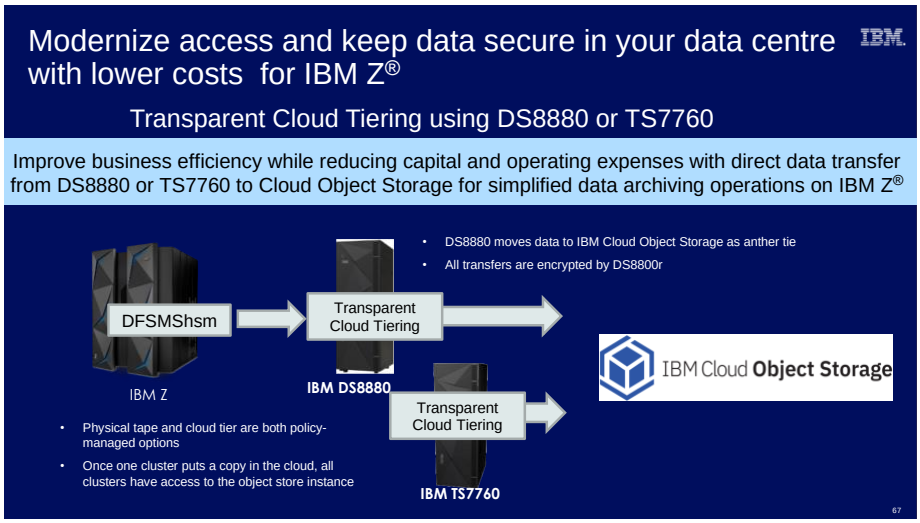
Komprise is a software only solution that manages the files on a file system and transparently moves those files after a period of time that is defined by the administrator. The Komprise system then creates a link that is transparent to the users and applications and moves the data off of primary storage to IBM Cloud Object Storage as a more cost effective and long term storage for less frequently used but always accessible data. Two other products that archive file data to IBM COS are DefendX(formally NTP) and Moonwalk.

The main Komprise dashboard

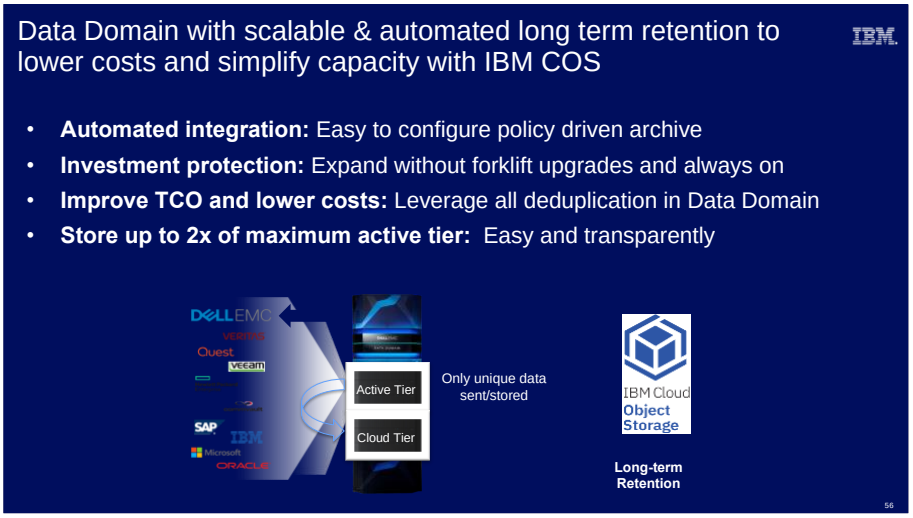
Archiving from Traditional Applications

Many applications have now built in a way to transparently archive data to an S3 storage system like IBM Cloud Object Storage. IBM had validated and tested the following applications and have customers who are successfully using the ease of scalability, low cost management and storage efficiency to lower their cost and consolidate their capacity for both opex and capex savings.

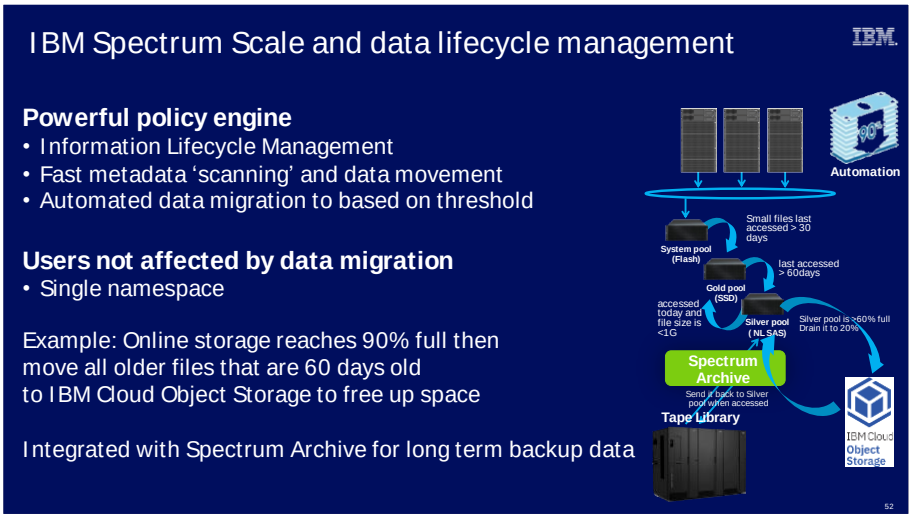
Veritas Enterprise Vault
IBM Spectrum Scale
Hyland/Acuo VNA
eMAM
IBM DS88XX and TS77XX
IBM FileNet
IBM Merge VNA
IBM Spectrum Protect Plus
Dell/EMC Data Domain
Nice/NTR
StoreNext



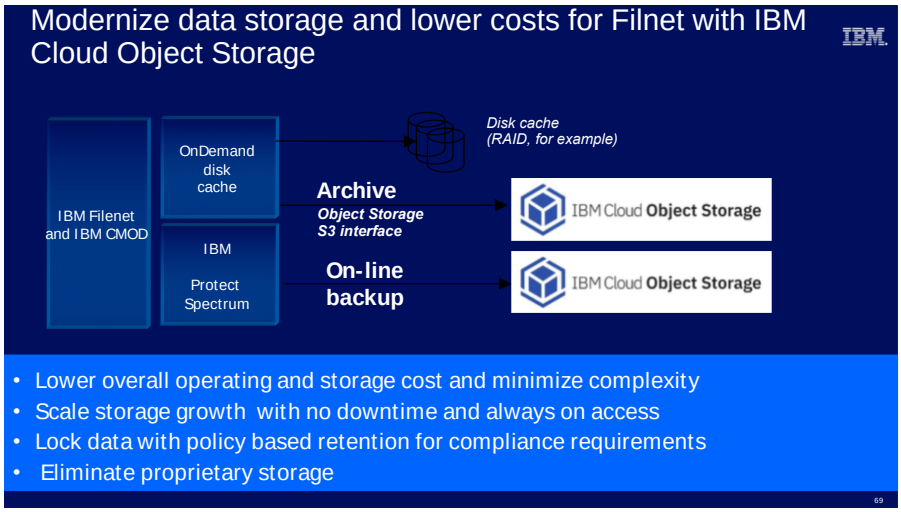
Leverage IBM COS with IBM mainframe and System z



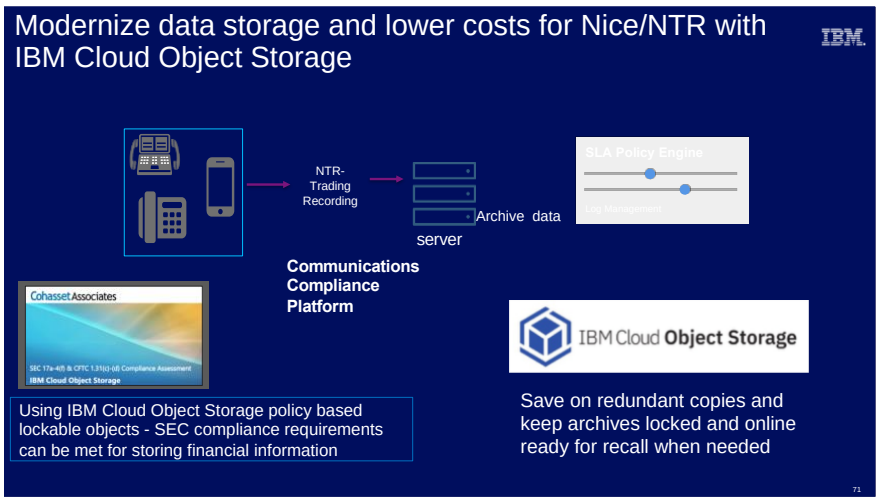
Leverage IBM COS with Dell/EMC DataDomain



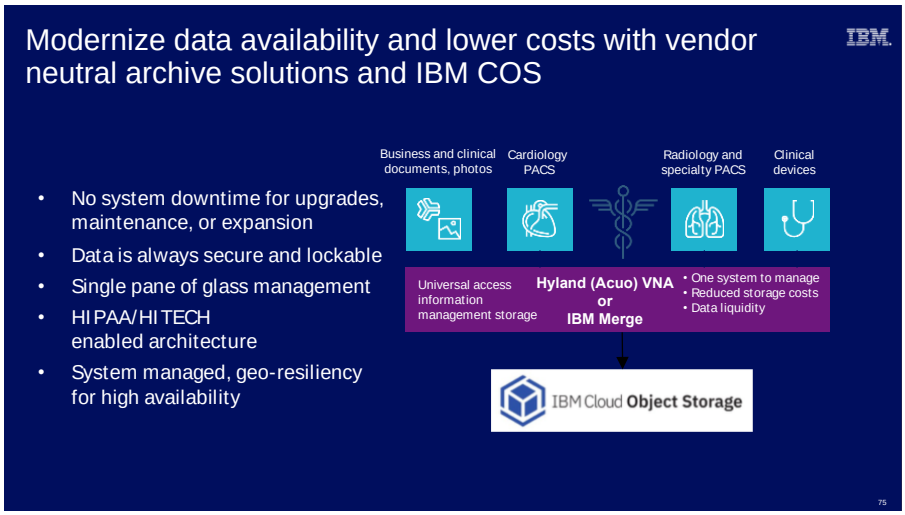
Leverage IBM with IBM Spectrum Scale lifecycle management



Leverage IBM COS with IBM FileNet and CMOD



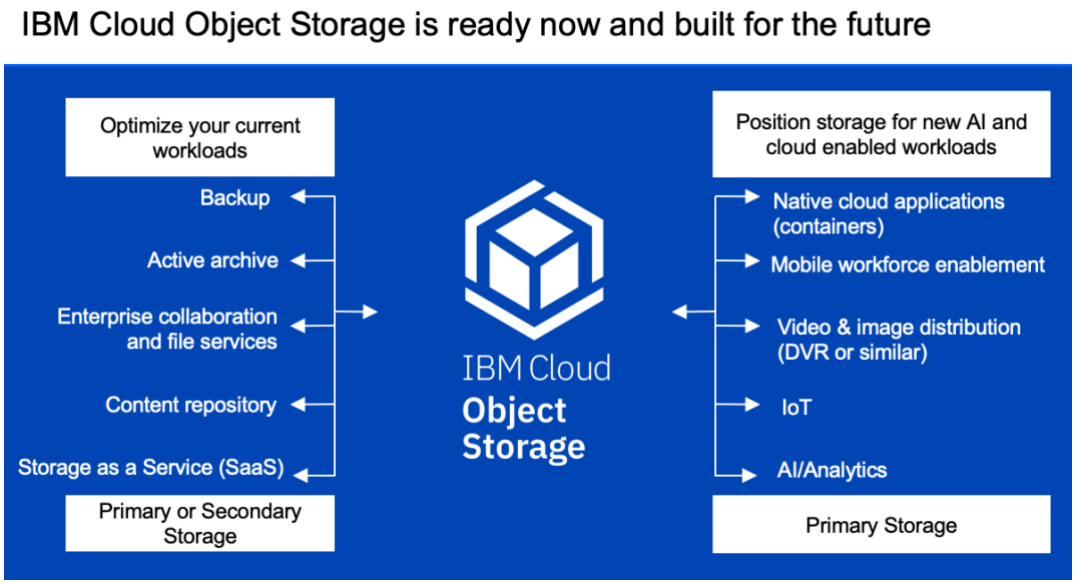
Leverage IBM COS with Nice NTR



Leverage IBM COS with Healthcare VNAs (IBM Merge and Hyland (Acuo))

Chapter 4. Conclusion

IBM Cloud Object Storage is bringing transformational and modern efficiencies to lower cost and global access to data in a highly resilient and easy to manage solution. It's easy to start with a small configuration or one that fits your budget and capacity requirements. A customer can start with as little as 72TB useable and grow into a second site or scale up capacity without any downtime. Our architecture is proven at EB scale and the dependability of our solutions means that customers can have confidence that their data is always available and safe without many of the issues found in other storage platforms. Our technology is constantly improving with over 800 technology patents and we are helping customers turn data in to expanding business opportunities.



IBM Cloud Object Storage for Primary of Secondary Storage

Related Documentation

IBM Cloud Object Storage Concepts and Architecture: System Edition, Copyright © 2019 IBM. This IBM Redpaper™ publication explains the architecture of IBM Cloud Object Storage on-premises offering and the technology behind the product. You can find a detailed discussion about the IBM Cloud Object Storage use case scenarios and deployment options in the IBM Redbooks publication "IBM Cloud Object Storage System Product Guide, SG248439".

URL: https://www.redbooks.ibm.com/redpieces/abstracts/redp5537.html

IBM Cloud Object Storage System Product Guide, Copyright © 2019 IBM. This IBM Redbooks® This publication describes the major features, use case scenarios, deployment options, configuration details, initial customization, performance, and scalability considerations of IBM Cloud Object Storage on-premises offering. You can find a detailed description of the IBM Cloud Object Storage architecture and technology behind the product in the IBM Redpaper™ IBM Cloud Object Storage System Product Guide, SG24-8439.

http://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248439.html?Open

References on the World Wide Web

IBM Support Knowledge Center for IBM Cloud Object Storage System: https://www.ibm.com/support/knowledgecenter/STXNRM

IBM Cloud Object Storage main web page: https://www.ibm.com/marketplace/cloud-object-storage-system

IBM Cloud Object Storage Datasheet: https://www.ibm.com/downloads/cas/QLZVPXKZ

Compliance with lockable WORM : https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WUL12394USEN&

Business Value of IBM Cloud Object Storage: https://www.ibm.com/downloads/cas/O4VAPEVD