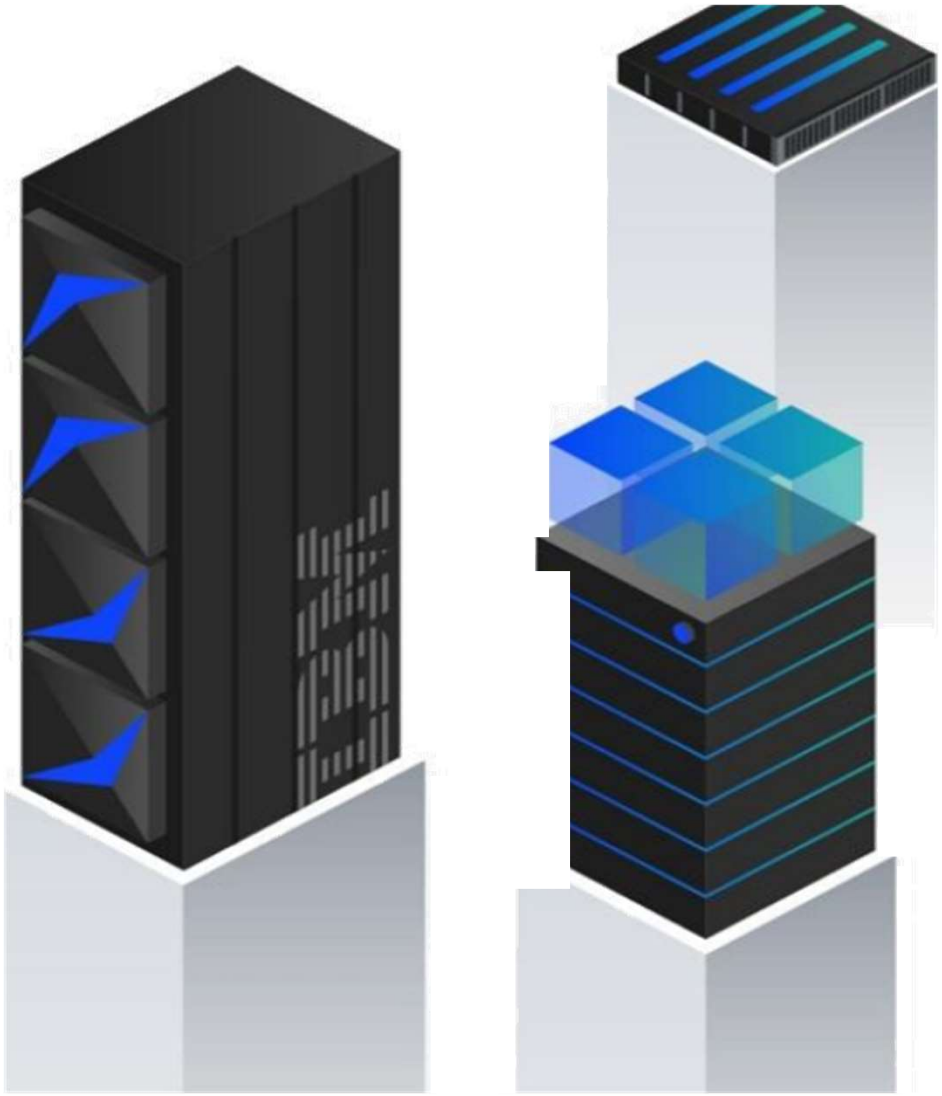
Archival Data Integration with SW tools and Tape for the Modern Workload

Sandro De Santis
Storage Technical Specialist, IBM Technology Sales, Italy
Sandro.de.santis2@ibm.com



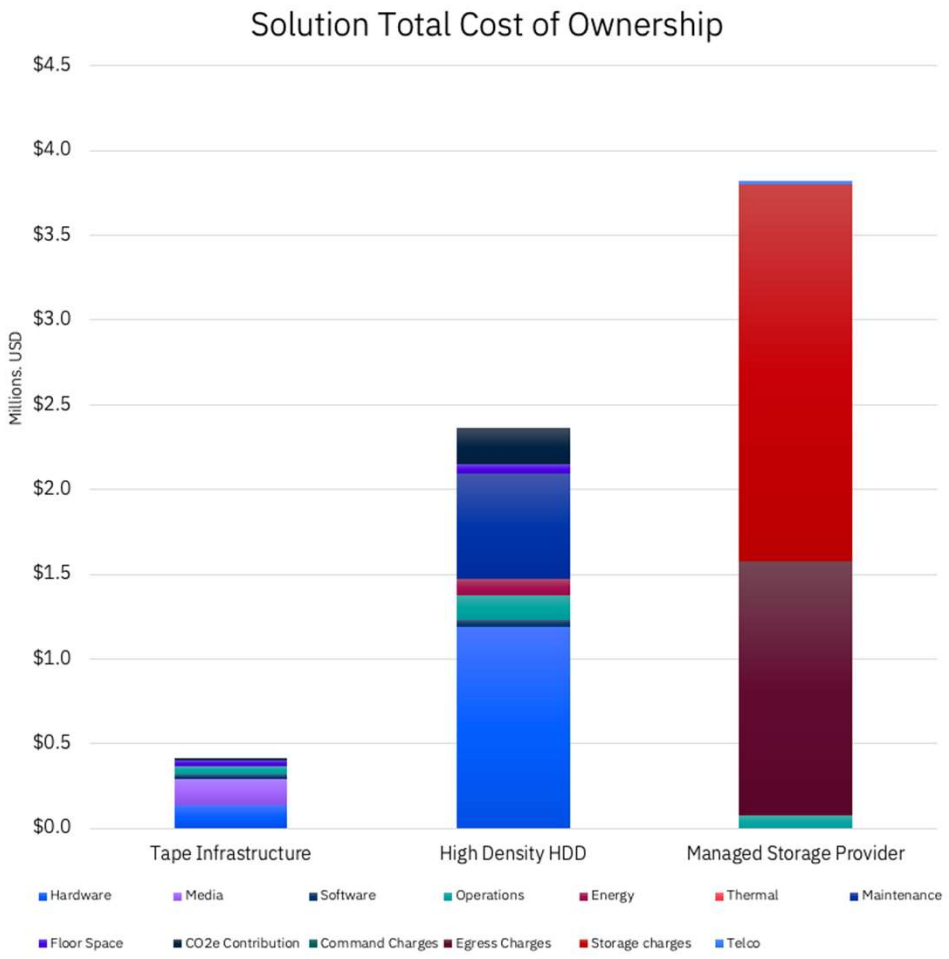
Tape TCO –reducing regulatory expenses

Premises cloud
is an average of

6X

more expensive than tape

- 20 PB Archive data
- 10% data CAGR
- 10 Years retention
- 1% data egress per month (MSP)
- Data transferred back to premises
- $200 per sq ft floor space cost
- 11 – Nines of durability



Easy-to-use, low-cost data archive storage

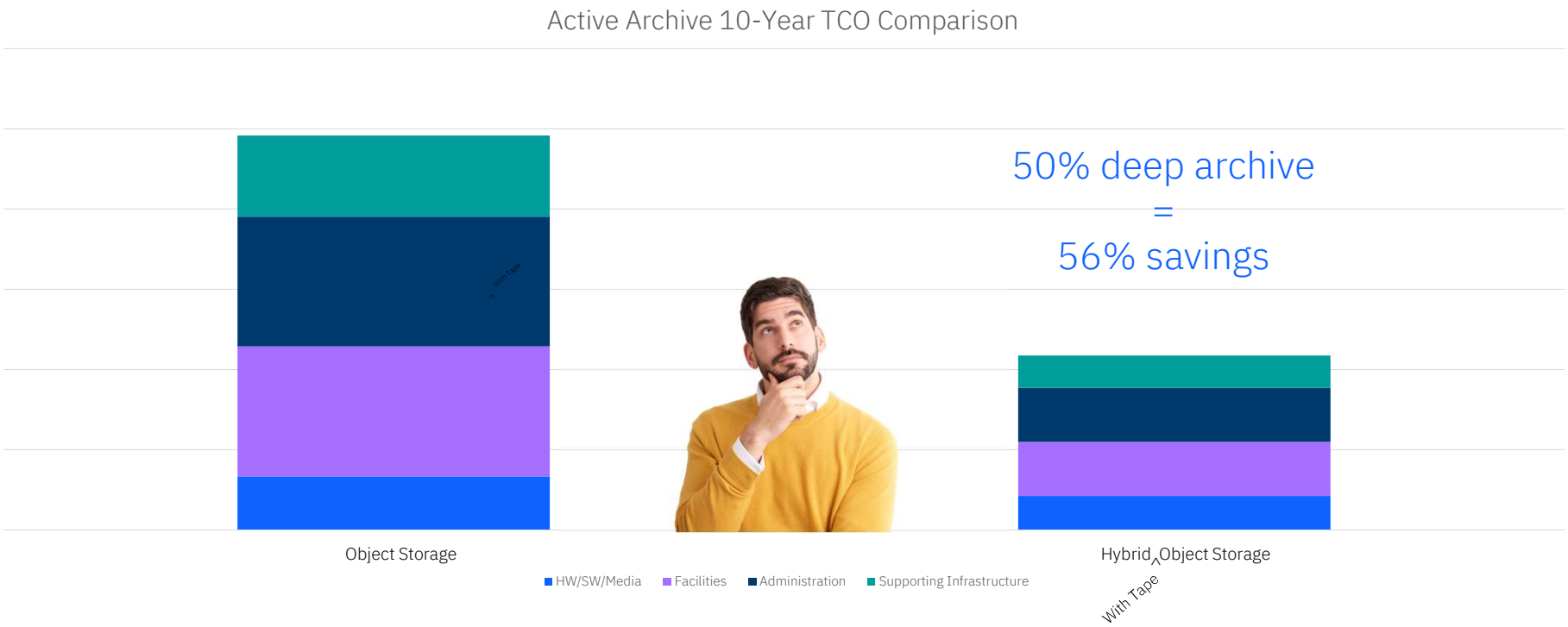
- Simplify tape usage
- No device-specific software
- Reduce TCO by up to 90%

"I had to do the calculations for the IBM solution twice, because I could not believe that the amount of savings could be true."

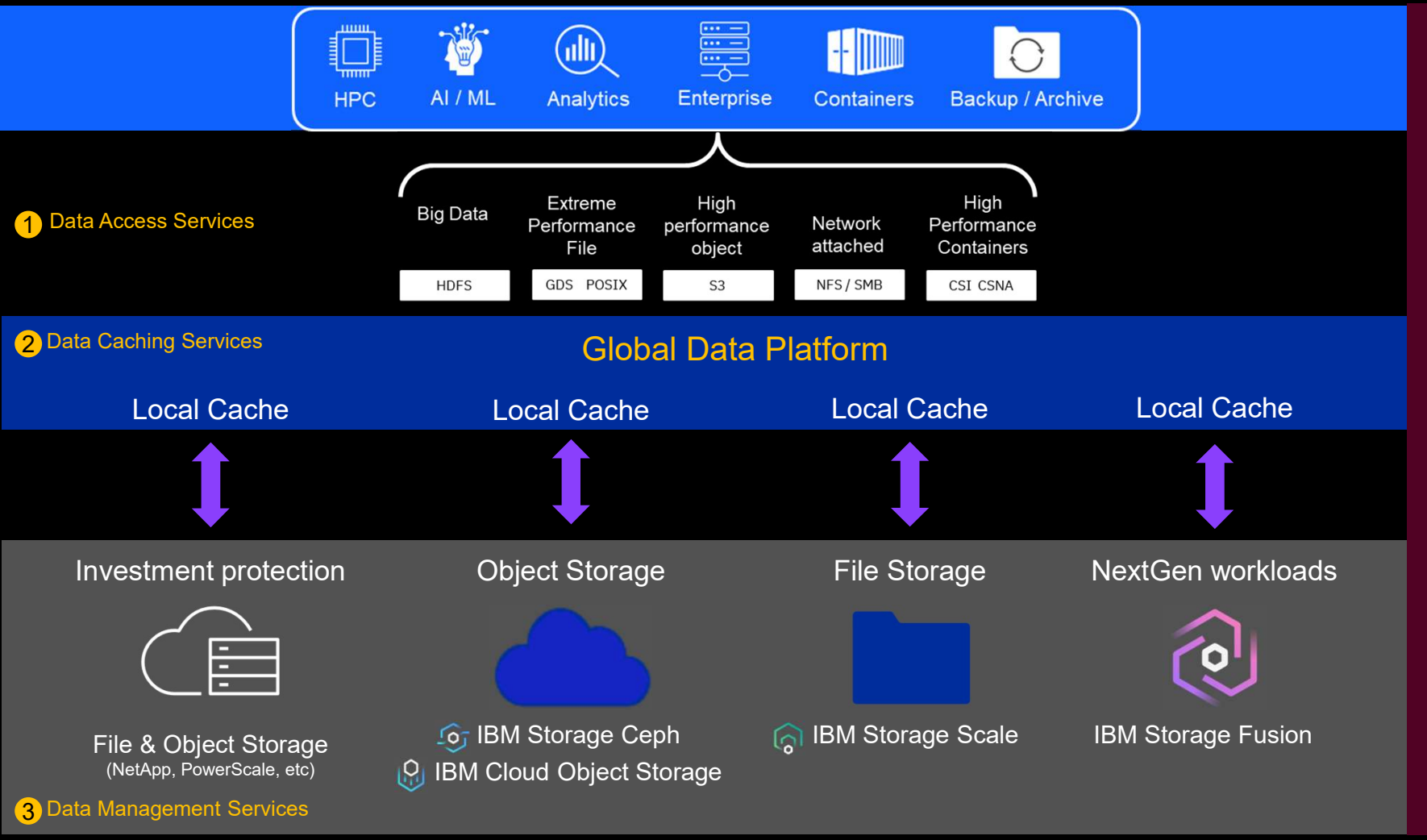
Constantinos Colombus,
Chief Technology Officer, AlphaTV

Economics of premises Active Archives – 20PB

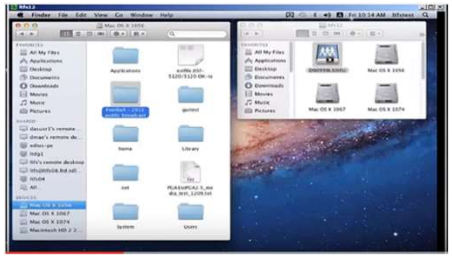
Savings nearly equal the archived data



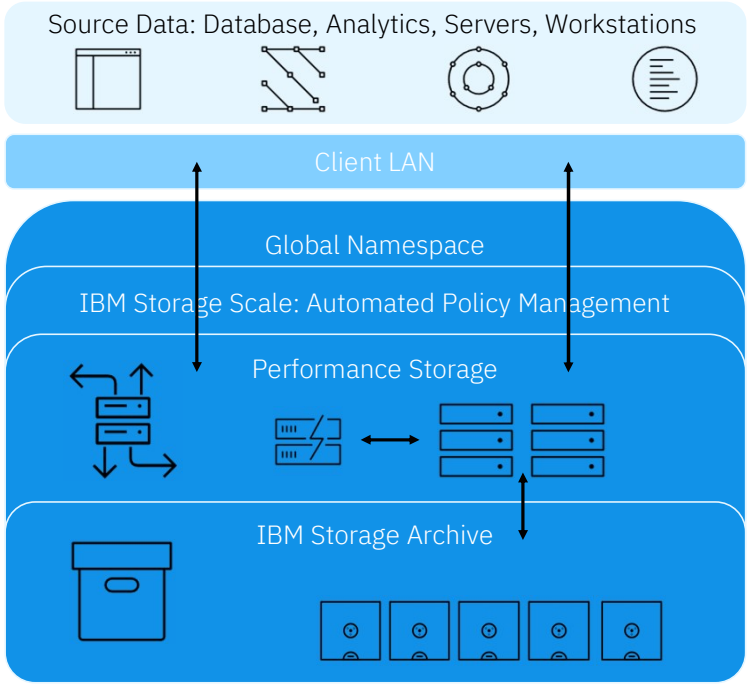
IBM's Global Data Platform for File & Object Data



IBM Storage Archive™ - Software defined storage, tape



- Graphical user interface
- Application interface utilization
- Policy-based migration
- Flexible data sharing
- Easy data access
- Data storage efficiency
- IBM Storage Scale™ and ESS integration

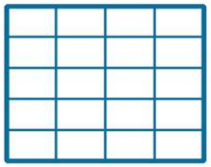


Is Moving Cloud Data to Tape Difficult?

File System

```
C:\folder\music.m4a
```

sysadmin required for integrity and scale



Database / Structured Data

```
SELECT * FROM table;
INSERT INTO table;
```

sysadmin and DBA required for scale, integrity and performance

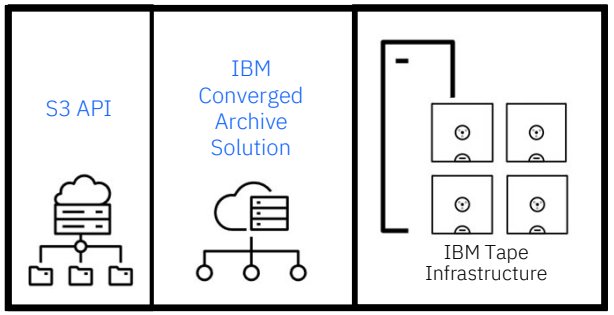
Object Storage

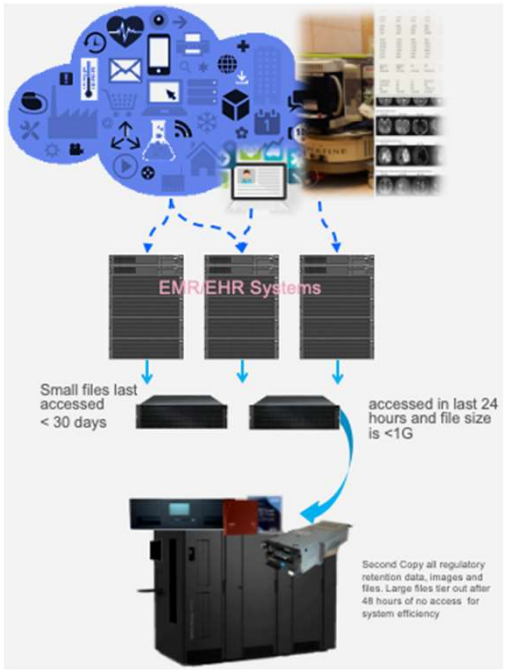
```
GET /object/KbglBn7qepo
PUT /object/KbglBn7qepo
```

sysadmin not required

Getting Cloud data to tape is easy

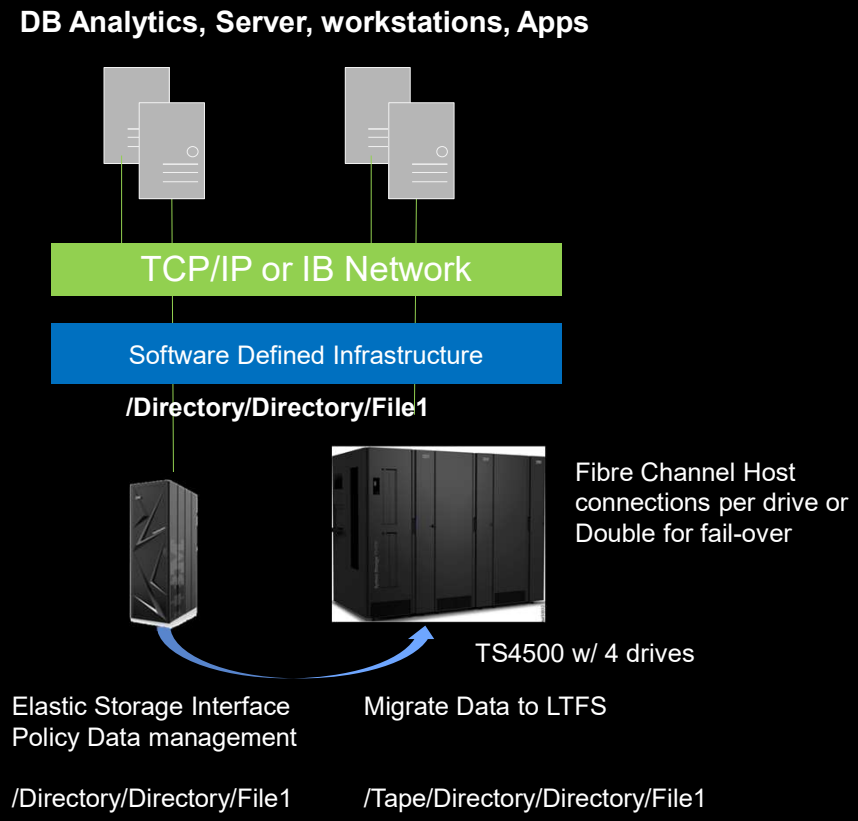
S3 API Data Mover





Sample Scale configuration – High Capacity or Growth

- **Data Capacity**
  - 3PB and Up
  - Can be expanded up to 417PB (native)
- **I/O**
  - Initial I/O 900 MBpS
  - Maximum I/O in largest I/O environment 51 GBpS
  - Support up to 128 tape drives
- **Use cases**
  - Near-line persistent Data storage
  - On or Off-Premise Storage Cloud (infrequent access)
  - NAS with tape for near-line availability of data
  - Production performance at a fraction of the cost
  - Data Exchange – Store and share large data, also good as secondary copy to deep storage
  - Any Industry - Medical, M&E, Education, Research, Government, Finance, Analytics, Start-ups



IBM Tape Portfolio – Sustainable Data Archiving

	TS2290	TS2900	TS4300	Diamondback	TS4500	TS7700 Tape Attach TS1150/1160
Max capacity*	18 TB	162 TB	5.04PB	27.8 PB*	417 PB LTO 877 PB** Enterprise	100 PB Enterprise**
Max # drives	1	1	21	14	128	16
Max # cartridges	1	9	280	1548 LTO	23,170 LTO 17,550 Enterprise	5,000 Enterprise
Withdraw from Market	Per drive Gen	Per Drive Gen	8Y+	N/A	10Y+	10Y+

- Actual capacity will vary depending on reserve and swap slot configuration

** TS1170 capacity with JF media

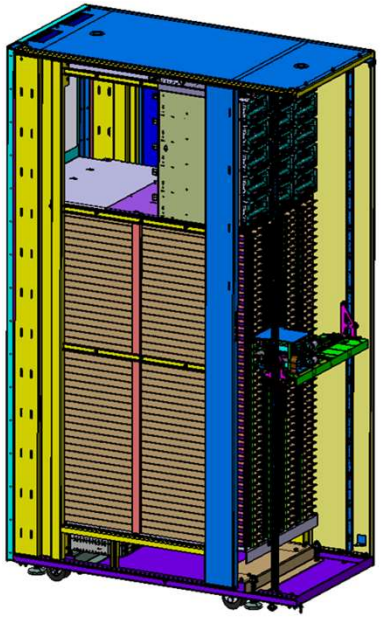
Diamondback – Storage Transformation

- Single Frame design – No expansion frames
- Up to 14 LTO-9 Tape drives
- Up to 1548 Tape cartridges
  - 27.8 PB uncompressed capacity*
- 1 cartridge service magazine with 10 I/O slots
- Single robot, high availability grippers
- Pre-loaded media option
- Self-service design
- GKLM library managed Encryption
- Open systems attachment including AIX®, Linux, MS Windows

- Support for LTO 9 tape drives
  - Native data transfer of up to 400 MBps
- LTO generation 9 media specification tape cartridge capacity of up to **18 TB**
  - (up to 45 TB compressed capacity)
  - Read and Write compatibility with LTO-8
- Ultra-high Density single frame capacity up to 27.8 PB* native
  - (up to 69 PB with 2.5 to 1 compression)
- 12 Gb SAS and 8 Gb Fibre Channel interfaces
- Open Recommended Access Order (oRAO)
- AES-256 Data-at-rest encryption
- IBM Archive™ support
  - Incorporates Linear Tape File System™ format
- Data security and regulatory compliance with encryption and WORM media

*Actual capacity will vary depending on swap availability and number of operational cartridges

Overview – Single frame / single robot tape library

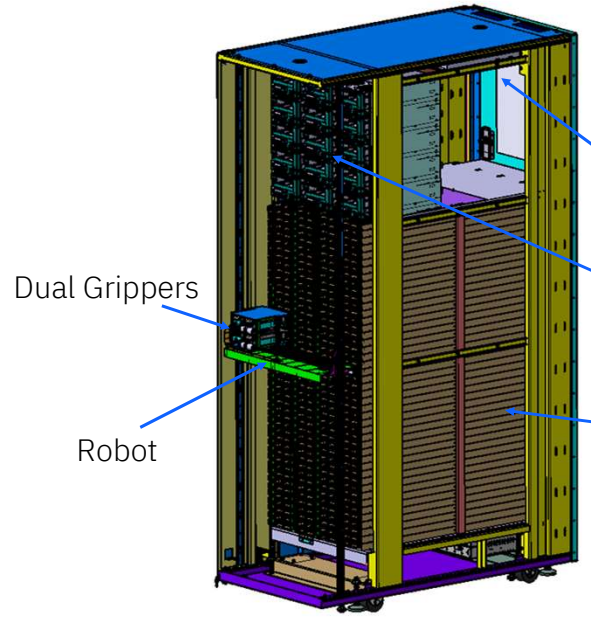


14 - LTO drive slots
1 - Cartridge Service Magazine
3 Columns, 5 Rows

1584 LTO cartridge slots

9 Tiers for LTO

4 Columns
44 Rows for LTO



Specifications Comparison

Feature	Diamondback	TS4500
Base Frames	1	1
Expansion Frames	Not Applicable	Up to 17
Tape Drives	Up to 14	Up to 128
Data Rate per hour	Up to 20.2 TB	Up to 184.3 TB
Supported Tape Drives	LTO-9	TS1150, TS1155, TS1160, TS1170* LTO-5, LTO-6, LTO-7, LTO-8, LTO-9
Supported Media	All drive supported Media	
Storage Slot Configuration	Up to 1584	Up to 23,170
Native Capacity	Up to 27.8 PB	Up to 877 PB**
I/O Stations	1	2 min, 8 Max
I/O Slots	10	Up to 144
High Availability Robots	1	2
Service and Support	Self-service CRU or IBM Service Support	IBM Service Support
Distributed Systems Attach	AIX®, MS Windows®, Linux®, Unix®	

- Upon availability of TS1170

** TS1170 capacity with JF media

IBM Deep Archive on Diamondback – Use Cases

A low-cost, on-premises cloud providing secure and durable storage for data archiving and online backup.

- Standard S3 Glacier commands
- No additional tape software
- Premise data control
- Data encryption
- No cost for retrieval priority
- Air gap storage
- Integrated system monitoring

NAS Cold Data offload

komprise

Cloud pools * with orchestration

VERITAS Enterprise Vault

Object Archive/Repatriation

Orchestra

MoonwalkSoft

<API data mover/orchestration>

rclone move with –min-age

Back-up Retention Archive

veeam

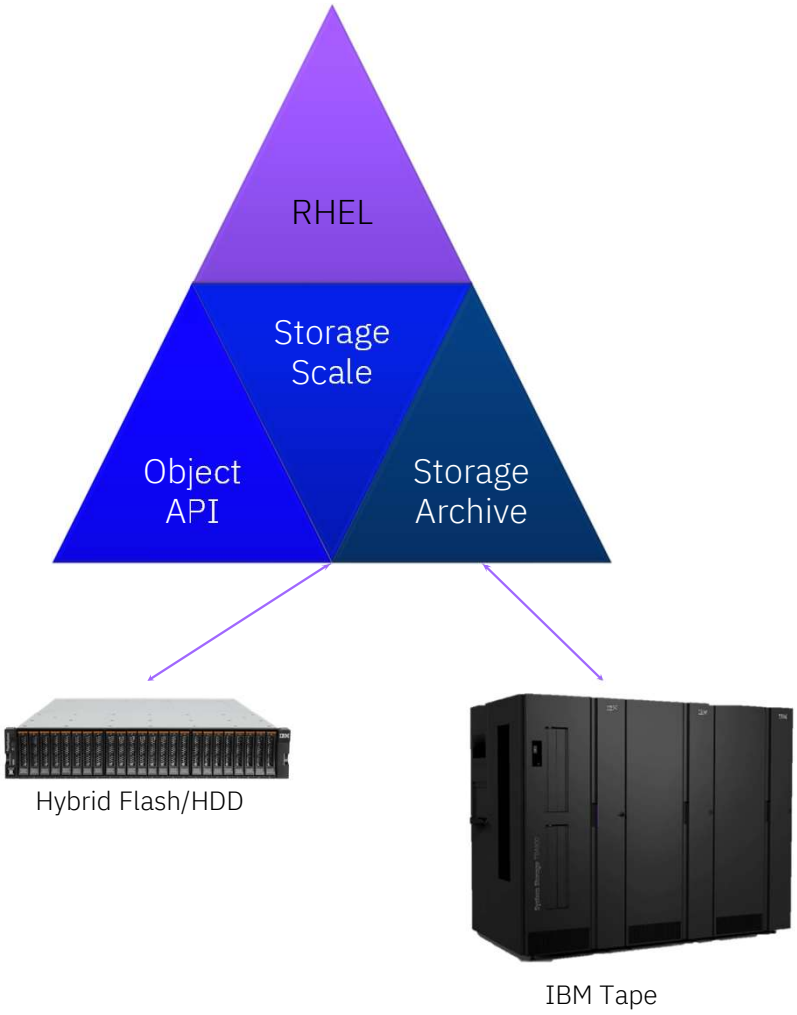
COHESITY

rubrik

https://www.ibm.com/docs/en/s3-deep-archive

IBM Deep Archive

- S3 Glacier compliant interface
- Install-and-go deployment
- Simple administration
- Up to 27PB object storage
- Up to 16.1TB per hour data rate



Key Components

- Server
  - RHEL
  - 128GB RAM Memory
  - HDD RAID-6 storage
  - RJ-45 10GbE integrated
  - 16GbFC tape connectivity*
  - Available HBA expansion slots
- Storage Scale and Storage Archive
  - Capacity License applied for Storage Scale
  - Server License for Storage Archive
- Open Object Gateway / API
- Attach to any IBM tape

*FC connectivity is not reserved and maybe split to support other external storage. Not included with iCAS LBS configuration services.

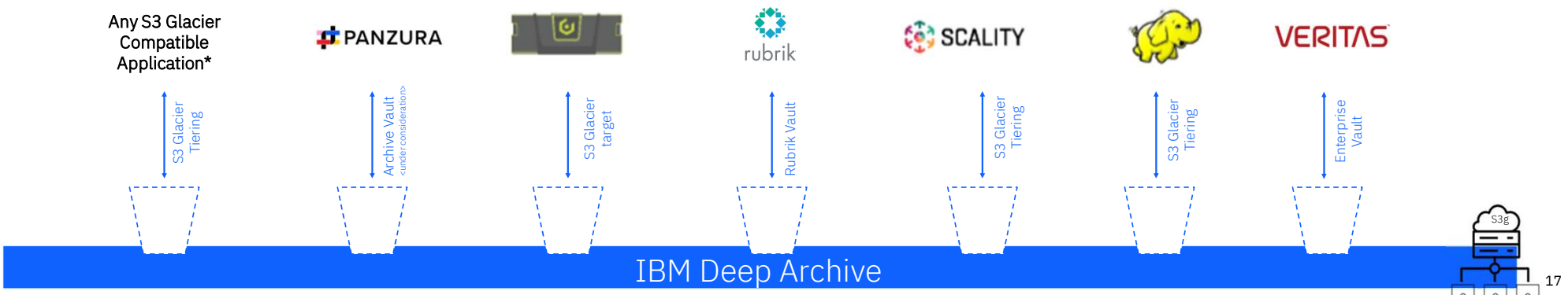
New Possibilities for Deep Archive Data

Any Application that supports the S3 Glacier Flexible Retrieval storage class may utilize IBM Deep Archive

Any existing S3 Object Storage using rclone move command
IBM Defender Data Protect *(1H2025)**
CEPH *(1H2025)**
IBM HPSS
IBM Storage Scale *(1H2025)**
Cohesity
Starfish
IBM Cloud Object Storage *(4Q2025)**
Commvault
Rubrik, rubrik Vault
Komprise
Veritas, Enterprse Vault
Moonwalksoftware/Panzura Symphony

Atempo
Duplicacy
CloudHQ
Nodeum NAS Cloud Gateway
Medium, AWS Storage Gateway
ARQ Backup
Synology NAS, Backup
Hitachi Vantara
Druva
Scality
Dell Data Protector
Orchestra.io
IBM TS7780 with Cloud Off-load *(2025)*

This is not a list of validated integrations; this list is based on published documentation. For IBM validated support visit the IBM ISV support page.



*Future support (not a commitment) for the Glacier Flexible Retrieval Storage class

IBM ISV Support Matrix: https://www.ibm.com/downloads/cas/ORLZ7EMG

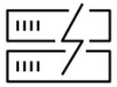
IBM Deep Archive: What it is and is not

IBM Deep Archive is:

- a deep archive storage using S3 API with glacier flexible retrieval storage classes.
- a user/application accessible data archive.
- a low-cost, high-capacity storage target.
- an easily deployed and managed archive.
- an automated archive requiring no additional software or specialized tape experience.
- S3 Glacier on-premises

IBM Deep Archive is not:

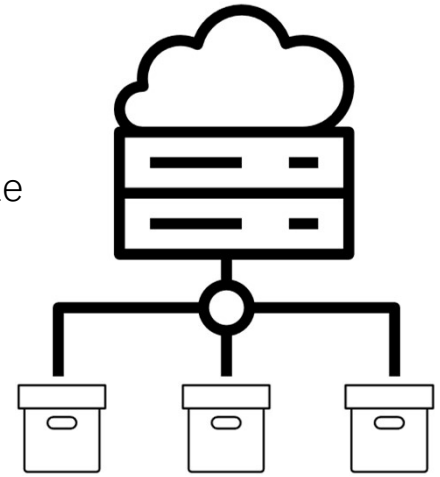
- a transactional storage.
- a primary high-performance high-availability fully redundant storage.
- a tiered solution for user object/file frequent accessibility.
- a full feature information lifecycle management platform.
- a fully automated data durability platform
- unified Object and File
- IBM Storage Scale + CES

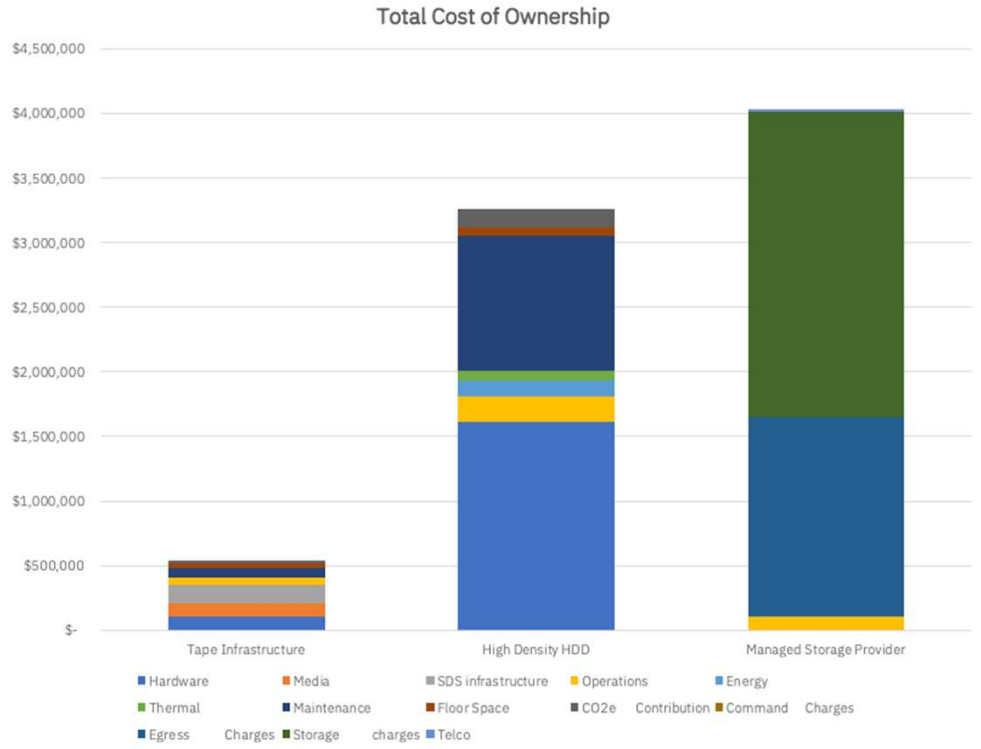


- a transactional storage.
- cloud-native mobile application Webservice.
- a tiered solution for user object frequent accessibility.
- AI and data analytics active storage.
- a rich S3 API compatible storage infrastructure.
- unified Object/File/Block
- IBM Storage Ceph Object

IBM Deep Archive on Diamondback

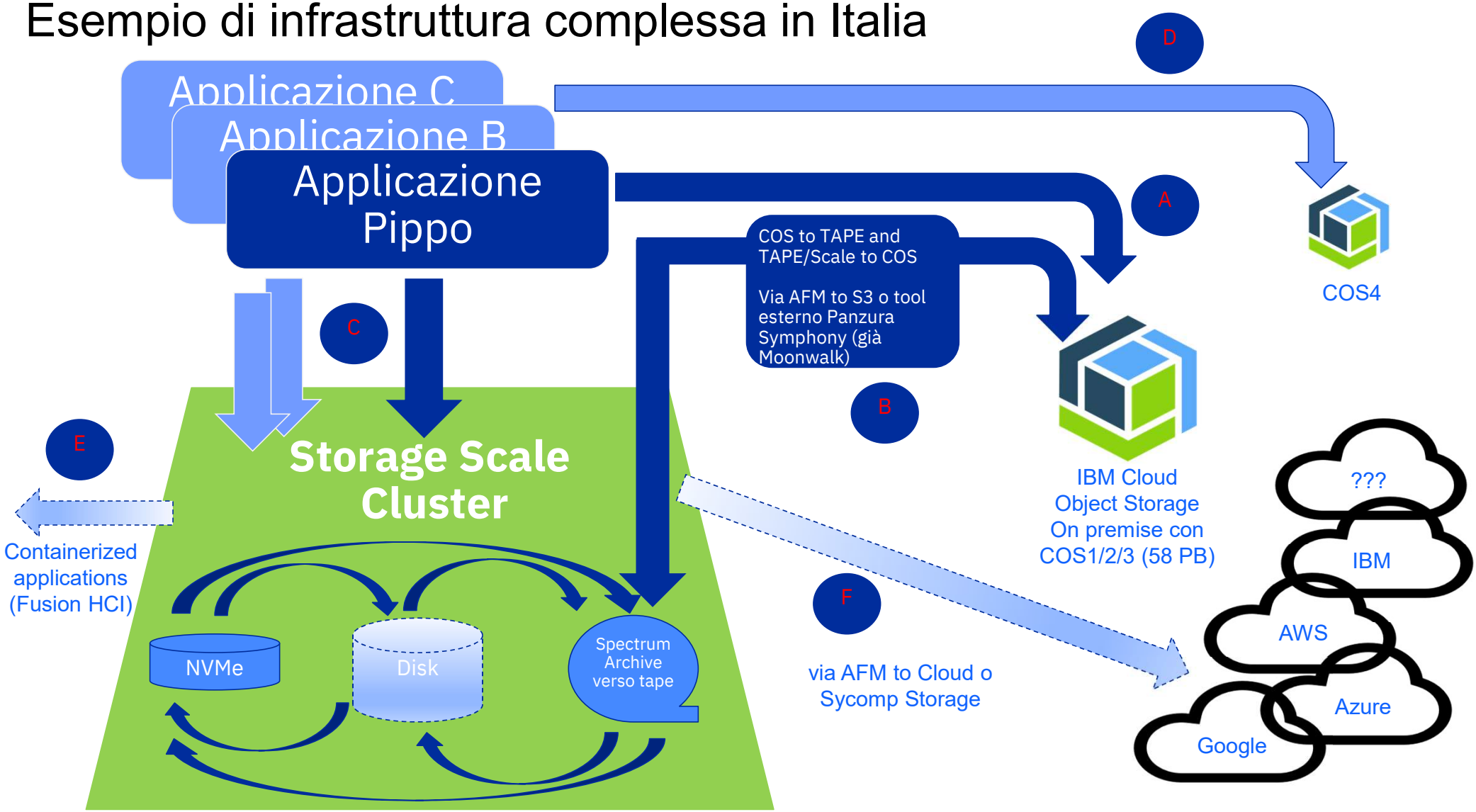
- Hybrid cloud archive
- S3 Glacier (FR) compatible
- Low-cost
- Secure
- Durable





IBM Cost Compare Tool: 10-year retention, 27PB data, 16TB HDD commercial S3, AWS Glacier Deep Archive, 1% recall/month

Esempio di infrastruttura complessa in Italia



Backup slides

IBM Tape Portfolio – Sustainable Data Retention

IBM Storage Archive

Single Drive Edition		Library Edition	Enterprise Edition		
				IBM S3 Deep Archive	
TS2290	7226	TS2900	TS4300	IBM Diamondback	TS4500
					
”Bridge Box” Stand-Alone tape drive Single Cartridge AC Power	Rackable tape drive Up to 2 independent drives Single Cartridge AC Power	“Media Changer” Rackable tape drive 1 tape drive Up to 162 TB	“Stackable Expansion Up to 48 drives Up to 48.38TB per hour 3U modular expansion Up to 11.52PB	“Scale-out” Up to 14 drives Up to 20.1TB per hour Simplified Service Up to 27.8PB	“Scale-up” Up to 128 drives Up to 184TB per hour Intelligent optimization Up to 877PB

	LTO-7	LTO-8	LTO-9
LTO Tape Drives Industry Interoperable Tape Format	6 TB Up to 300 MBpS 6Gb SAS Host Interface 8Gb FC Host Interface Downward R^{1x}/W^{1x} Compatible WORM and Encryption LTFS	12 TB Up to 360 MBpS 12Gb SAS Host Interface 8Gb FC Host Interface Downward R^{1x}/W^{1x} Compatible WORM and Encryption LTFS	18 TB Up to 400 MBpS 12Gb SAS Host Interface 8Gb FC Host Interface Downward R^{1x}/W^{1x} Compatible WORM and Encryption LTFS

IBM Tape Enterprise Portfolio – High-density Storage

IBM Storage Archive Enterprise Edition

TS4500

"Scale-up"
Up to 128 drives
Up to 20.1TB per hour
Intelligent optimization
Up to 877PB Native

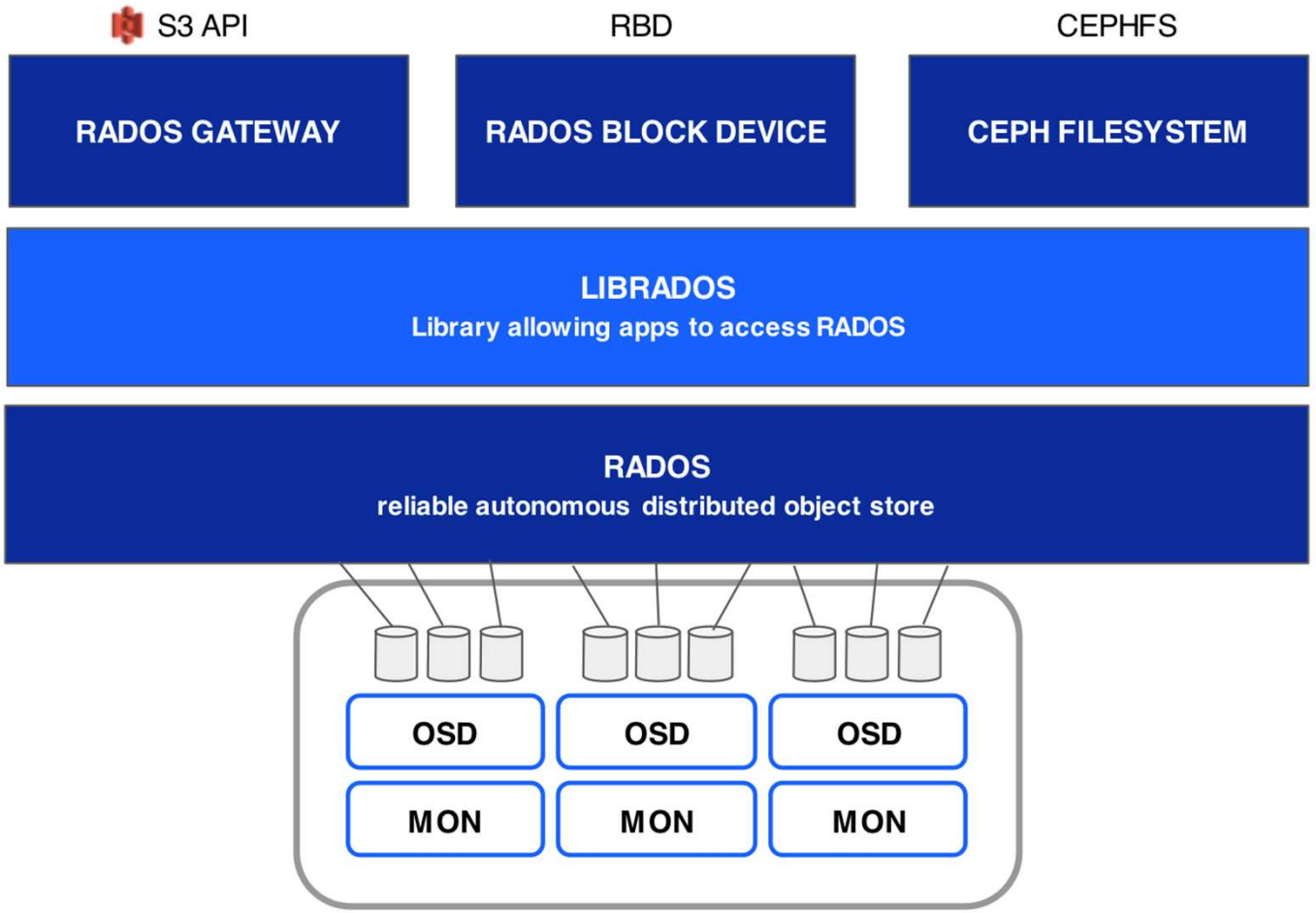
	TS1150	TS1160	TS1170
Enterprise Tape Drives Industry Interoperable Tape Format	10 TB 360 MBpS 12Gb SAS Host Interface 8Gb FC Host Interface Downward R^{2x}/W^{1x} compatible JD, JC Media WORM and Encryption Field Upgrade available only	20 TB 400 MBpS 12Gb SAS Host Interface 16Gb FC Host Interface Downward R^{1x}/W^{1x} Compatible JD, JE Media WORM and Encryption	50 TB 500 MBpS 12Gb SAS Host Interface 16Gb FC Host Interface Capacity Format Only JF Media Only Physical WORM and Encryption

IBM Storage Ceph Basic architecture

IBM Storage Ceph solution is centered around RADOS, which means:

Reliable Autonomous Distributed Object Store

also known as the “base engine part” of IBM Storage Ceph.



IBM Storage Ceph
Use cases

1

Object storage as-a-service

Scalability and performance for both small and large object stores.

2

Data lake for data analytics/machine learning applications

Cloud native data lake, with massive scalability and high availability to support demanding workloads

3

Cloud native (S3) Data pipelines

A unique new way of architecting the data flow in applications which is through event driven architectures

4

Active archive and nearline storage

Backup & recover into and from an object storage will help to improve RTO and RPO

IBM Storage Ceph as backup target

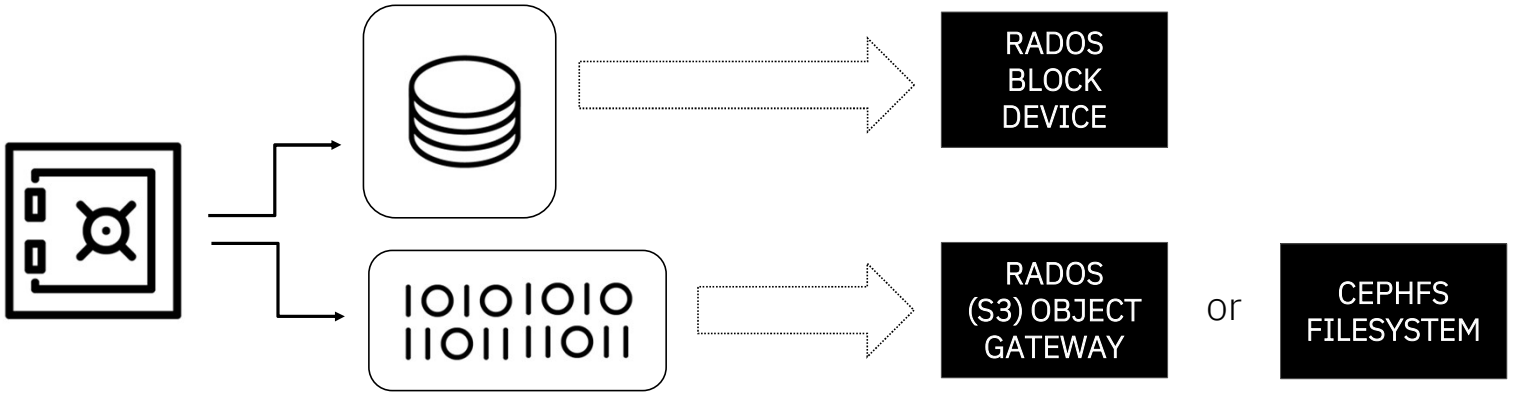
Disk library for backup data

IBM Storage Ceph offers a scalable storage solution to store backup data.

Accessible by AWS compatible S3 API for object or CephFS for filesystem shared access.

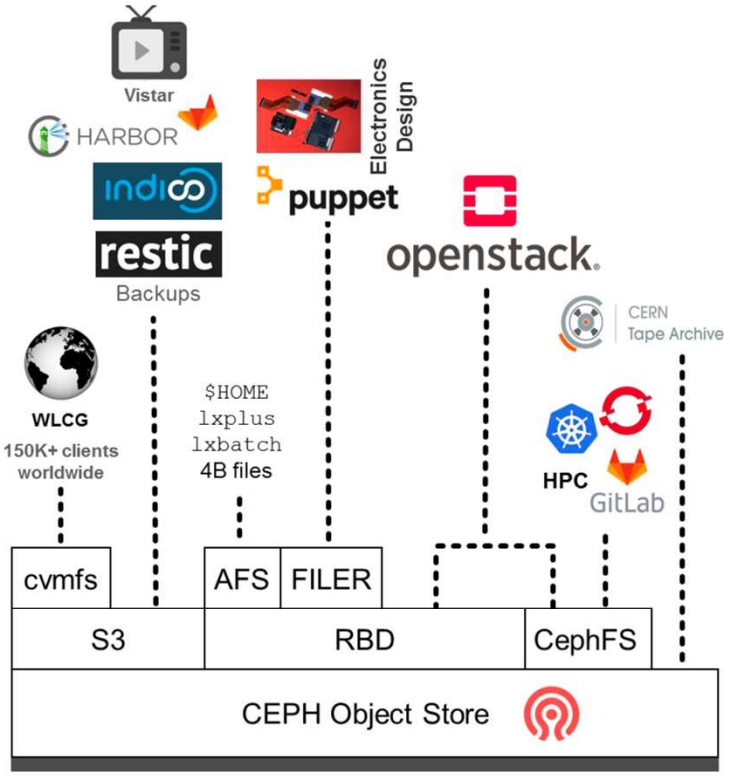
Database store

Data protection application’s database can be stored on IBM Storage Ceph Rados Block Device, providing resilient performant block storage.



Ceph at CERN

- **IT Services:**
  - Cloud Infrastructure: OpenStack, K8s, OpenShift
  - Code repositories, Container Registries, GitOps, Agile Infra
  - Monitoring: Open Search, Kafka, Gafana, InfluxDB, Kibana
  - Document Repositories // Web: Indico, Drupal, WordPress
  - Analytics: HTCondor, Slurm, Jupyter Notebooks, Apache Spark
- **Other Storage:**
  - NFS Filers, AFS, CVMFS, CERN Tape Archive, …
- **Physics Experiments and End-Users:**
  - Accelerator Complex Monitoring
  - Microelectronics Design
  - Engineering and Beams



Ceph at CERN

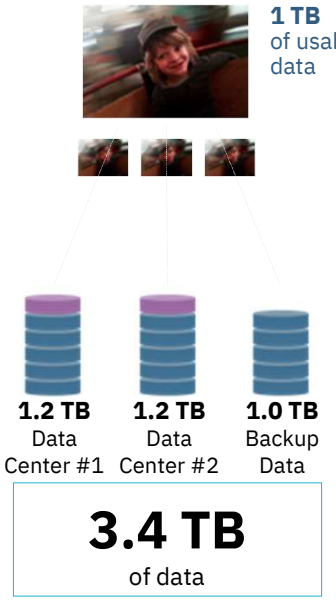
Application		Size (raw)	Version
RBD (OpenStack Cinder/Glance, `krbd`)	*Production, HDDs*	24.5 PiB	Pacific
	Production, full-flash EC 4+2	643 TiB	Pacific
CephFS (OpenStack Manila – K8s/OKD PVs, HPC)	*Production, HDDs*	7.9 PiB	Pacific
	Production, full-flash	782 TiB	Pacific
	Hyperconverged (HVs with flash storage)	892 TiB	Octopus
CERN Tape Archive (CTA)	*Tape DB and Disk Buffer*	235 TiB	Octopus
RGW (S3 + SWIFT)	*Production (4+2 EC)*	4.1 PiB	Octopus
S3, RBD: Backup to 2nd Location	*Production (4+2 EC, 3 replicas)*	25 PiB	Octopus

108858501

IBM Cloud Object Storage information dispersal

Redefining availability and economics of data storage

Traditional storage



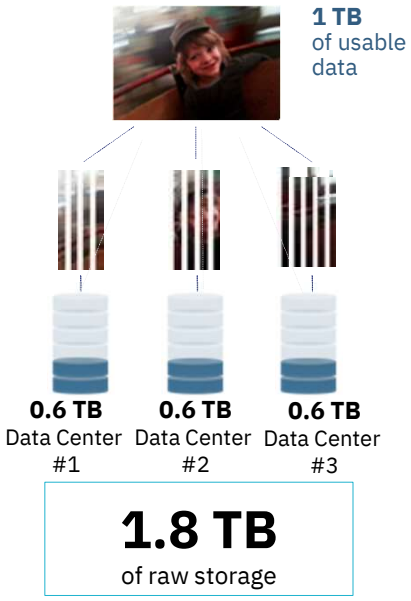
Traditional storage requires 3.4 TBs raw storage capacity for 1 TB of usable storage.

IBM Cloud Object Storage requires less than half the storage and 70% lower TCO*.

You can lose a disk, a server or even a whole site due to failure or disaster, and still quickly recover 100% of your data.

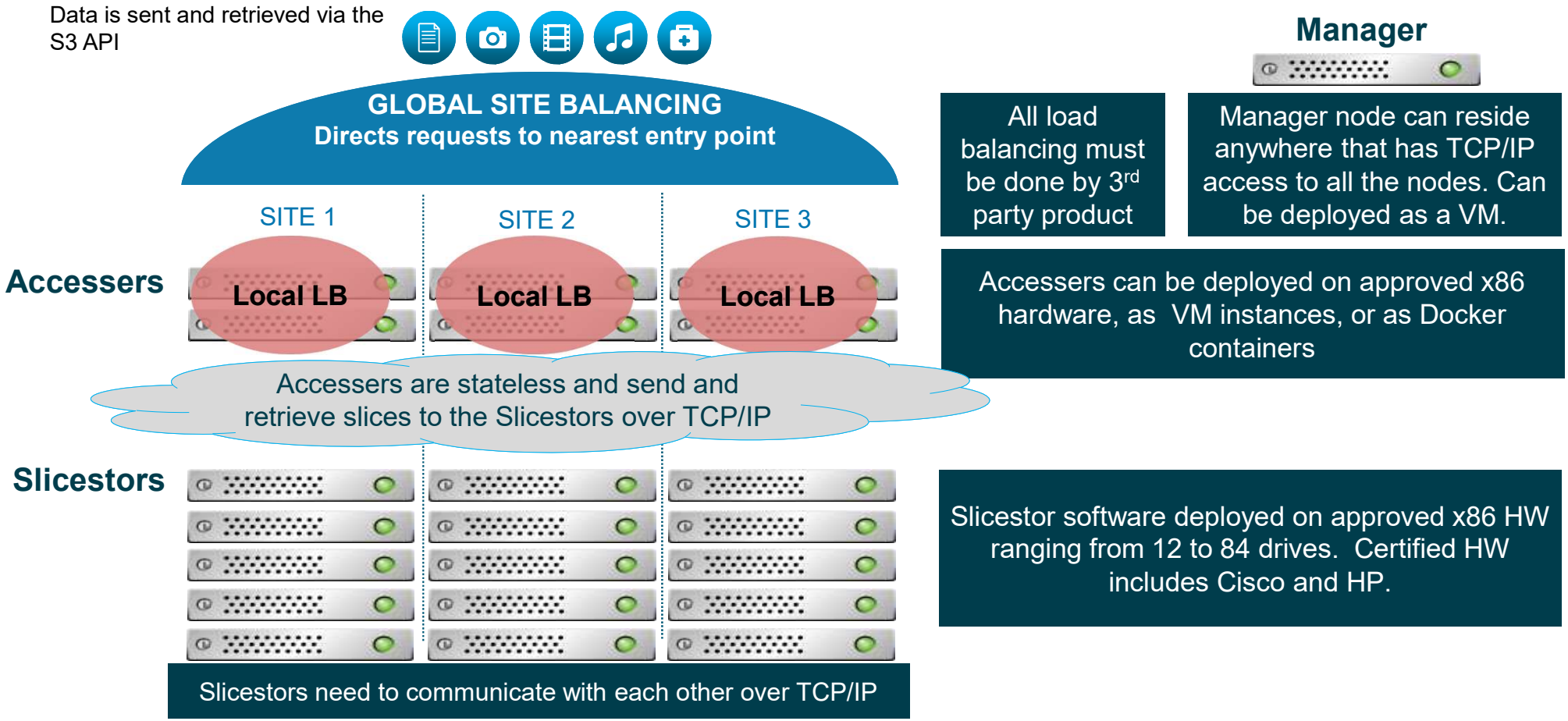
Slices are distributed geographically for durability and availability.

IBM Cloud Object Storage



Our object storage requires only 1.8 TBs raw storage capacity for 1 TB of usable storage.

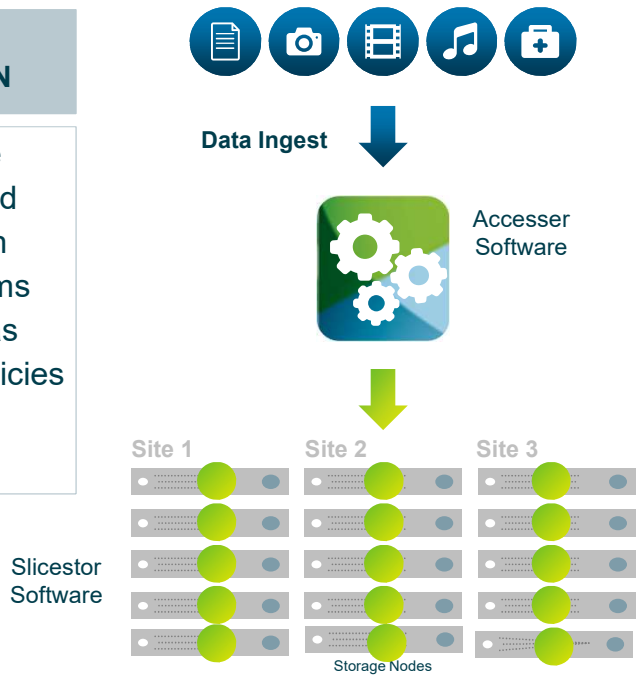
COS Architecture



How it Works

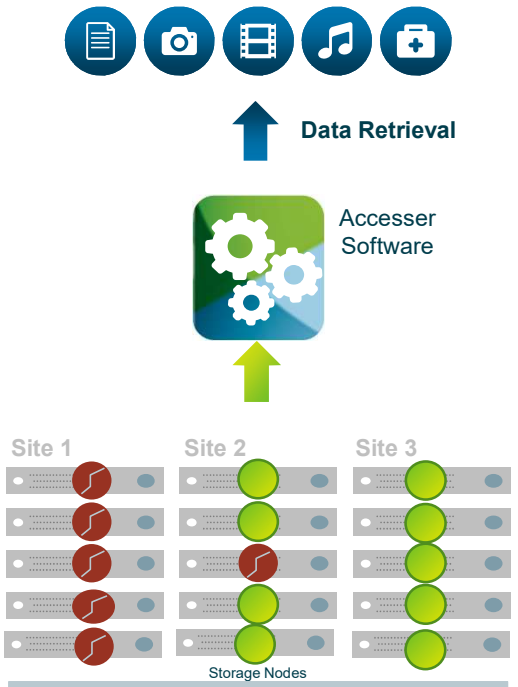
CONTENT TRANSFORMATION

IBM COS software encrypts, slices and applies Information Dispersal Algorithms otherwise known as erasure coding policies to the data.



Physical Distribution

Slices are distributed to separate disks and industry standard x86 hardware across geographic locations.



Reliable Retrieval

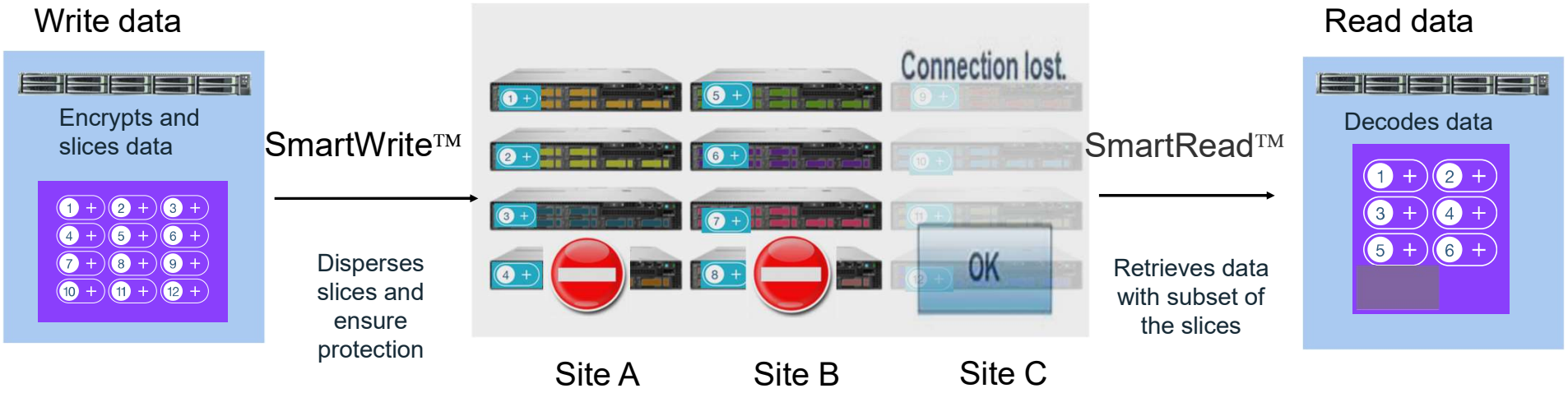
An operator defined subset of slices is needed to retrieve data bit perfectly in real time.

BENEFITS

The level of resiliency is fully customizable resulting in a massively reliable and efficient way to store data at scale as opposed to RAID and replication techniques.

IBM differentiator: Architected for availability and reliability

IBM's Information Dispersal Algorithm (IDA)



SecureSlice

AONT Encoding

Data

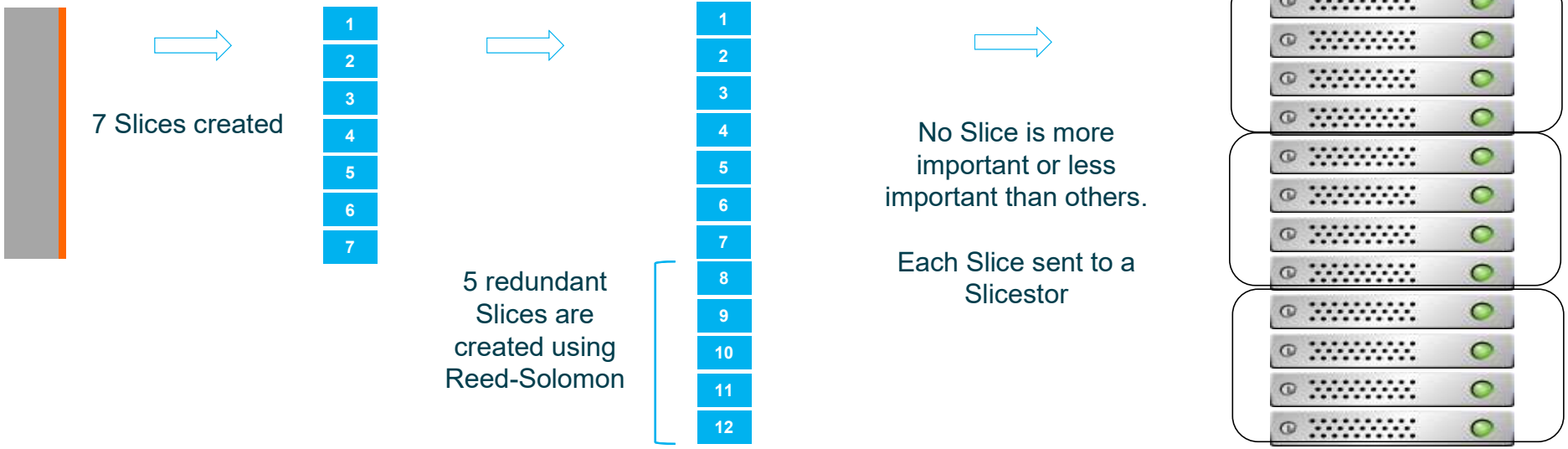
4MB Segments

Append Integrity Check Value

Encrypt using randomly generated key

Result appended to create AONT package

Information Dispersal



7 Slices created

5 redundant Slices are created using Reed-Solomon

No Slice is more important or less important than others.

Each Slice sent to a Slicestor

IBM COS Differentiators

PROVEN AT MASSIVE SCALE	ERASURE CODING IS INLINE	NATIVE COMPLIANCE WORM
Clients with well over 200PB in production scale performance & capacity concurrently	Not a post process. Data is fully protected upon write commit	3rd party validated SEC 17a-4f compliant WORM, Secure AES-256 encryption, S3 API interface
DISK LIFECYCLE MANAGEMENT	**ALL FEATURES INCLUDED**	**PERPETUAL LICENSING MODEL**
Predict drive failures and take appropriate action	Base price includes all interfaces	No need to repurchase licenses when refreshing hardware
CHOICE OF DEPLOYMENT	**ZERO-TOUCH ENCRYPTION**	**SOFTWARE PACKAGED w/ OS**
On-Premise, Hybrid, Off-Premise Software Defined, Hardware Aware	No need for 3rd party encryption solution or key management	No need to install and manage a separate OS