CLOUDIAN®

Next Gen S3 Compatible Object Storage

Adam Dagnall
Senior Sales Engineering Director - EMEA
adagnall@cloudian.com

Halvor Evenson
Senior Sales Director - Nordics
hevensen@cloudian.com

Welcome

Adam Dagnall

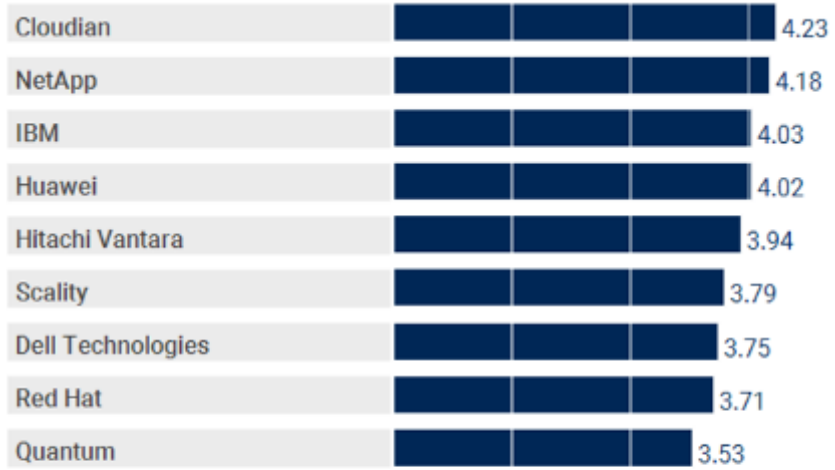
Senior Sales Engineering Director - EMEA

Halvor Evensen

Senior Sales Director - Nordics

Meet Cloudian

- Built to the AWS S3 Standard
- Software Defined Storage on Commodity x86 HW (On-Prem)
- Enables multiple use cases for enterprises
- The most security certified
- Start small with Exabyte scale
- Feature rich for shared cloud platforms
- VMware, Rubrik, Veeam v10+ (v12), Veritas, Commvault integration
- AWS Outpost & Local Zone ready & launch partner



Ranked #1 on Gartner Critical Capabilities Report

Trusted by leading organizations

Global 1000

verizon · NTT · etisalat · Mercedes-Benz · GE · Pfizer · SONY · MERCURIA ENERGY GROUP LTD · RICOH · McKESSON

Media/Imaging

N Major Network · WGBH · CBS

Retail/Consumer

Kroger · TESCO · POLO RALPH LAUREN · Schreiber · KOHL'S · petco · BB Major Retailer · I Global Furniture

Higher Ed

Johns Hopkins University · Princeton University · University of Leicester · Hamburg University · Berkeley University of California · University of Exeter · University of Oslo · Big Data Institute, University of Oxford · Osaka University

Tech

VERITAS · MOTOROLA · R Games · NUANCE

Healthcare

NIH U.S. National Library of Medicine · FUJIFILM Medical Systems · CDC · Public Health England · NEO Genomics · DELTA DENTAL · National Cancer Institute · MERCY Medical Center · eHealth Queensland

Service Provider

T-Mobile · CCL · www.switch.ch · vodafone · SaskTel · gtt · Telia · telenor · switch · TELMEX · NAVER Cloud

Finance

CBRE · UniCredit · S Payment Processing · A Credit Card · Bloomberg PolarLake · Rabobank · PICTET · PostFinance · State Street · CITADEL

Government

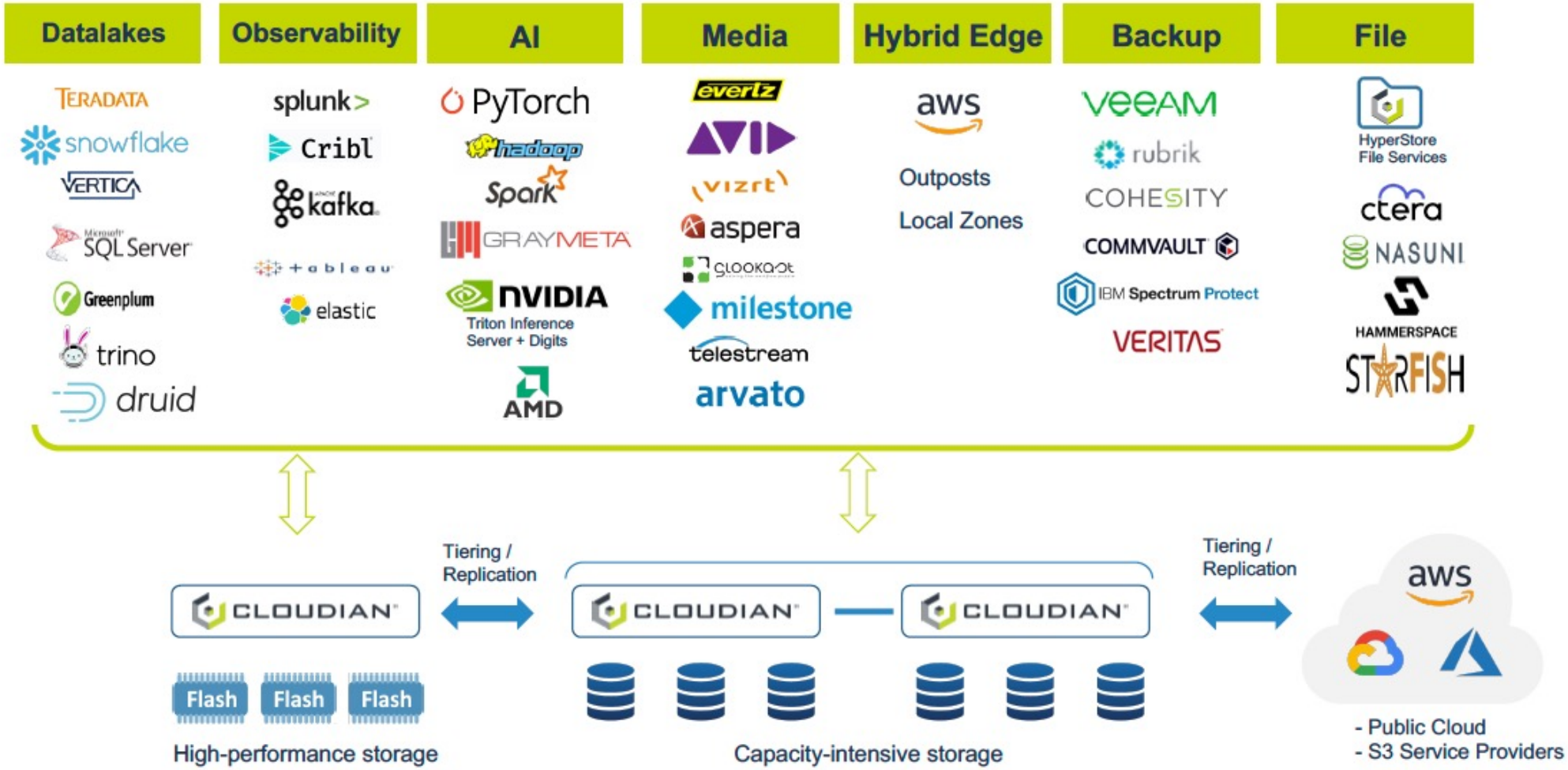
California Republic · DC.gov · U.S. Department of Defense · NIH National Institutes of Health · Federal Reserve Bank of Kansas City · Sheriff · U.S. Air Force · European Commission

Industrial & Engineering

DAIMLER · MITSUBISHI · Rockwell Automation · WILLIAMS F1 TEAM · RENAULT

CLOUDIAN

Cloudian App Store



CLOUDIAN

3 Key Tenets

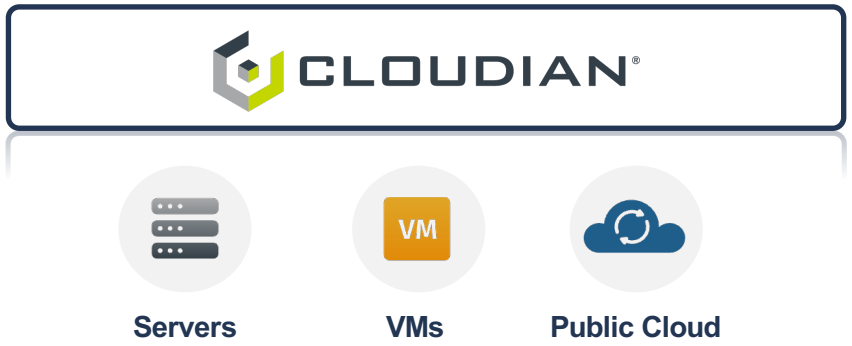
- Scalable
  - Modular Heterogenous node support
  - Scale performance and capacity to 100’s of PB’s
  - Introduce additional sites and protection schemes
- Secure
  - Object Lock for Ransomware protection
  - Hardened Shell (HSH) for secure access
  - Certified mode for compliance
- S3 Compatible
  - Native S3, runs on every node
  - Highest level of S3 compatibility outside of AWS

Deployment options

Software-defined

Deploy on the platform you prefer

- Flexible software licensing models
- Run on servers, VMs, cloud
- Industry-standard hardware



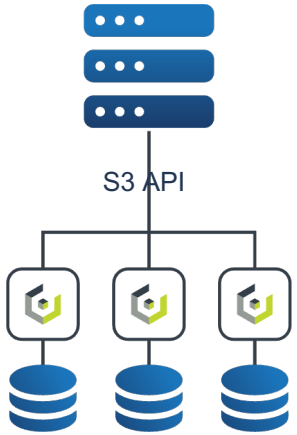
Cloudian appliance

Pre-configured solution

- High-density, storage-optimized servers
- 1U and 4U form factors
- HDD and all-flash configurations

Product Family

HyperStore
Object Storage

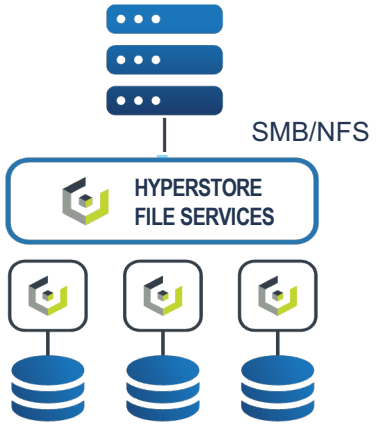


Most secure immutability

Best S3 compatibility

Multi-tenant

HyperStore
File Services

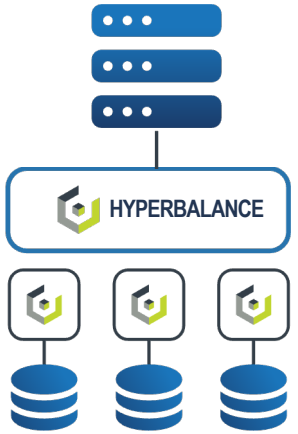


NFS/SMB protocols

Bimodal file/S3 object access

Scale-out

HyperBalance
Load Balancing



Load balancing

Performance optimization

Enhanced data availability

HyperIQ
Observability

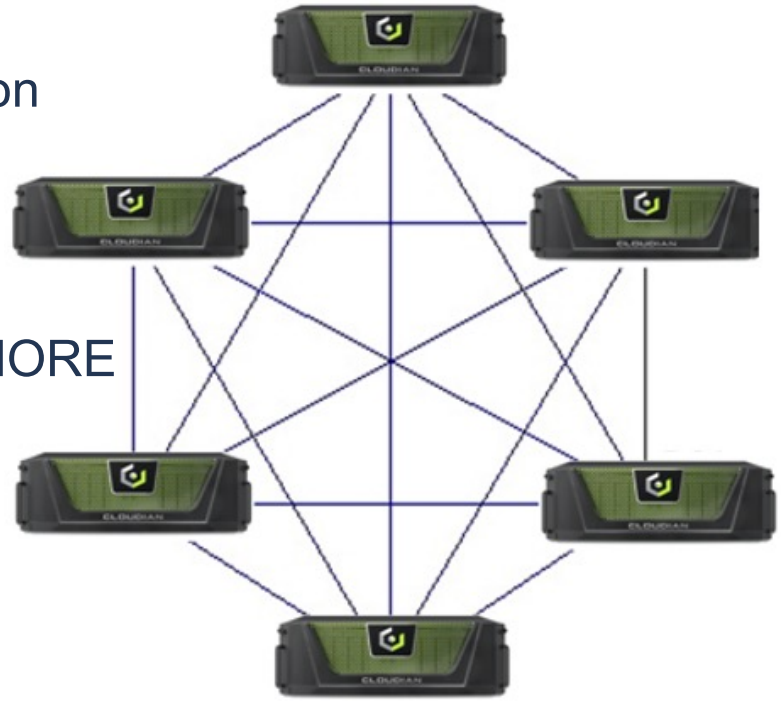
Time series database

Server, storage, and networking

Single-screen view

Cloudian HyperStore Design Principles

- **Flexibility**
  - Multiple Hardware and Software deployment options – including cloud
  - Various appliance (HW) options including high density and high-performance appliance options
  - Cluster design & architecture based on workload
  - Fully fledged Admin API for management & custom integration
  - True ‘Peer to peer mesh’ architecture
  - All core services run on all nodes
- **Security**
  - Object Lock, HyperStore Secure Shell (HSH), Encryption +MORE
- **Multitenancy**
  - RBAC, IAM, QoS, LDAP / SSO & Chargeback built in
- **Multicloud flexibility**
  - Tier to, replicate to, or run in the public cloud



Exabyte capacity Scalability

200 nodes/region

10 DCs/region

20 regions

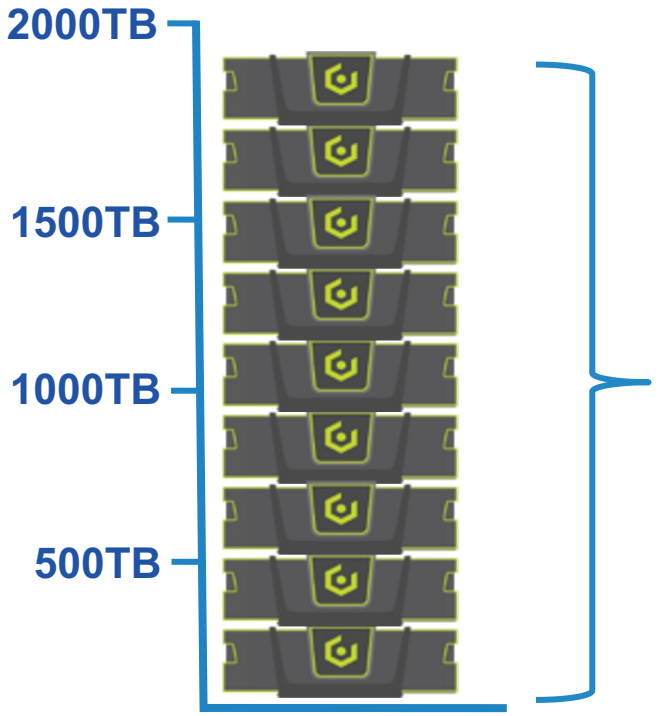
20 x

200 x

Node with 48 x 16TB disks

Region: 153PB

System: 3EB



Simple
Manage as one device

Scalable
Requests serviced by all nodes

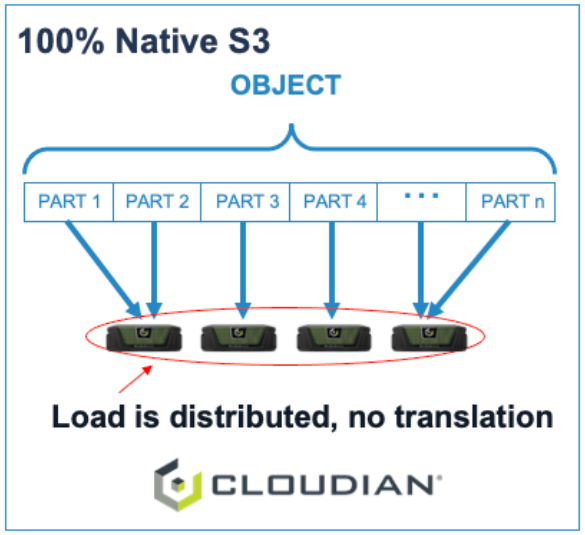
CLOUDIAN®

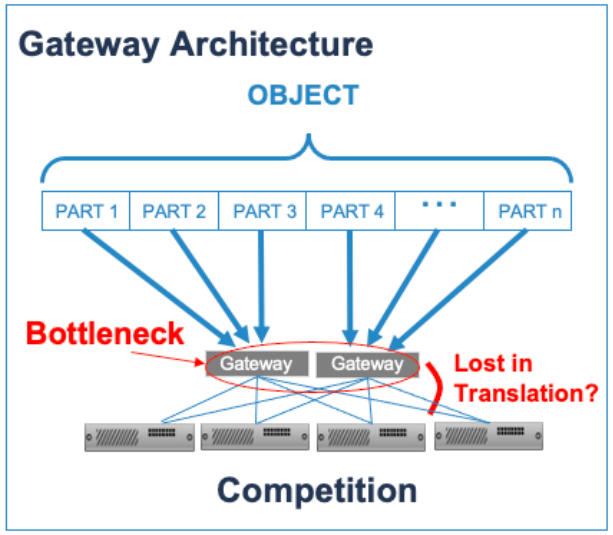
Cloudian - S3

- De-facto standard for object
- Complex API ~100 API calls
- Error level compatibility
  - S3 applications work as they should with HyperStore
- Native
  - Runs on every node

S3 API Topic	# Items	%Compliance
Amazon S3 - Common request headers	10	100.0%
Amazon S3 - Common response headers	10	100.0%
Amazon S3 - Signature Versions	2	100.0%
Amazon S3 - API	105	84.8%
Amazon S3 - Canned ACLs	7	100.0%
Amazon S3 - ACLs	5	100.0%
Amazon S3 - Error codes	83	94.1%
Totals	**222**	**90.6%**

Given some AWS S3 API are irrelevant, our actual API compliant % would be closer to around 96%





CLOUDIAN

Security +

Data immutability

S3 Object Lock for ransomware protection

Data cannot be changed or deleted

Tamper-proof protection

Security certifications

The most complete set of certifications in object storage:

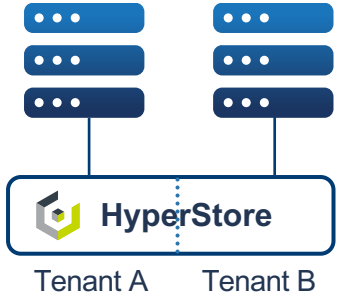
Common Criteria, FIPS 140-2 SEC Rule 17a-4(f), CFTC 17 C.F.R. § 1.31, FINRA 4511c, IDW PS 880 (German), OR §§ 957ff (Swiss) NIST 800-88

Secure multi-tenancy

Securely share single environment

Unique namespaces & management

QoS controls ensure performance

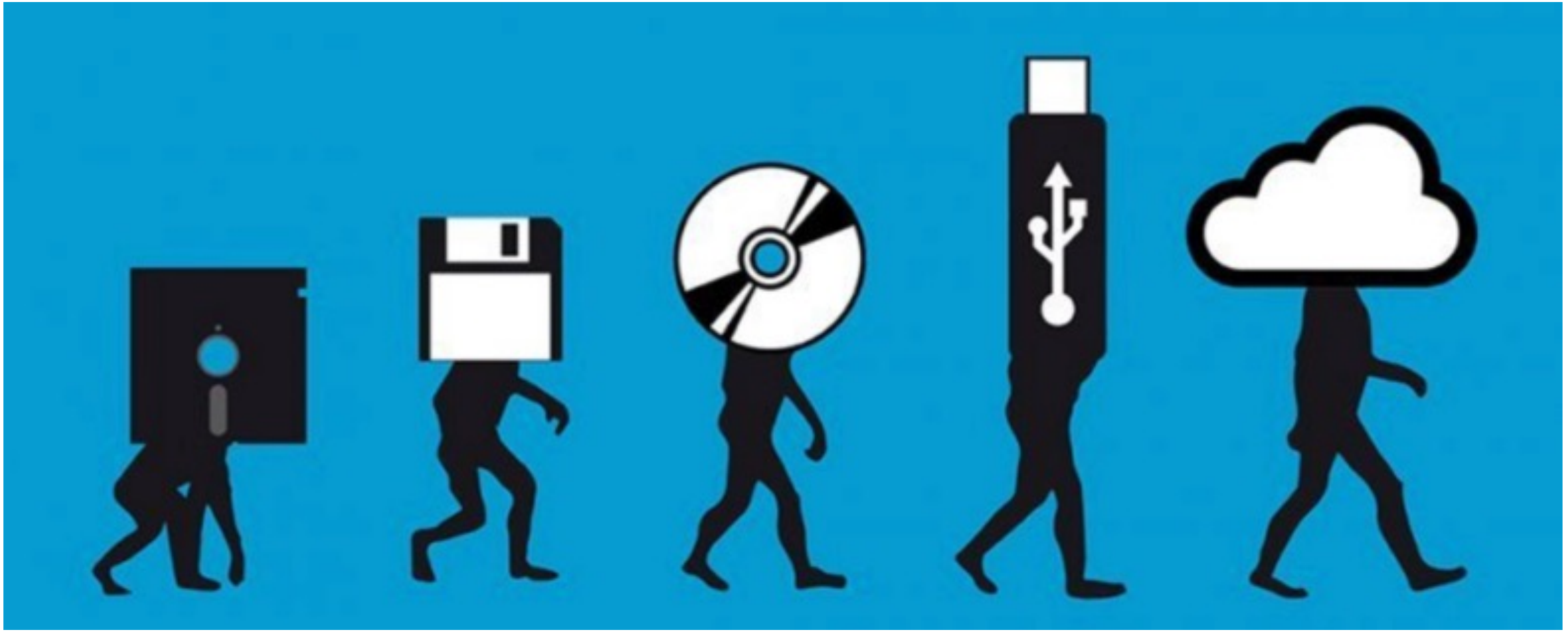


Data encryption

Data at rest: AES-256 encryption (NIST)

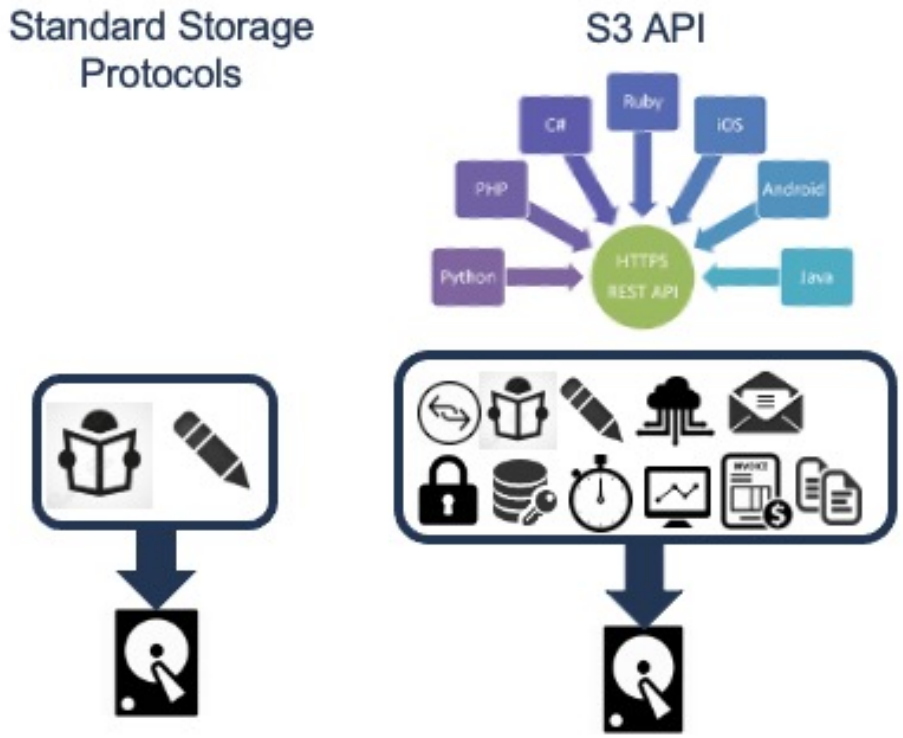
Data in flight: TLS 1.2 and 1.3 protocols (Internet Engineering Task Force)

Enterprise Storage Evolution



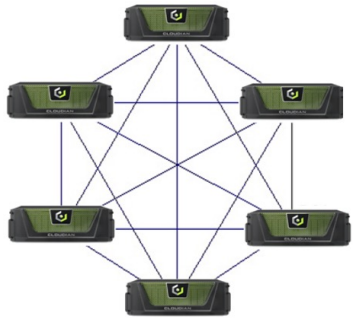
S3 is a Modern Access Protocol

- Providing more than just IO

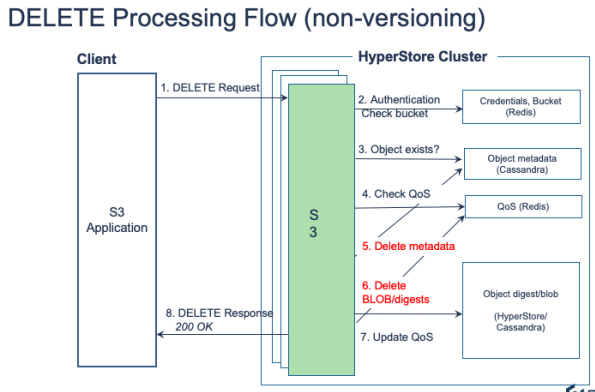


- Rich functionality including metadata functions IAM & SQS

- Running in a distributed environment with a single namespace



- Distributed operations are thereforce more complex i.e. deletes



How is Object Storage Used?

60% of share of wallet is spent on storage

Fast Data

Qumulo

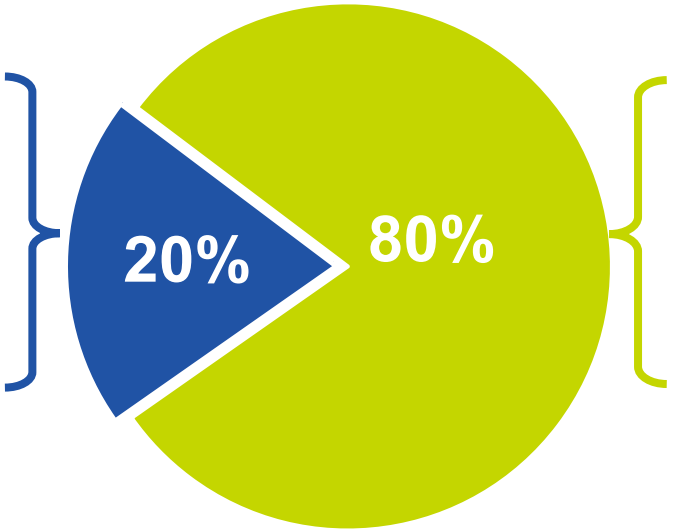
PURESTORAGE®

NUTANIX

VAST

Block & File

Data



Capacity Data

CLOUDIAN

Object Storage

Typical Object Storage Workloads

- Large data sets
- Backup workloads
- Archive workloads
- Secondary storage
- Tertiary storage
- WORM
- Ransomware protection
- Compliance & Regulatory
- Long Term Retention
- Cost optimised capacity

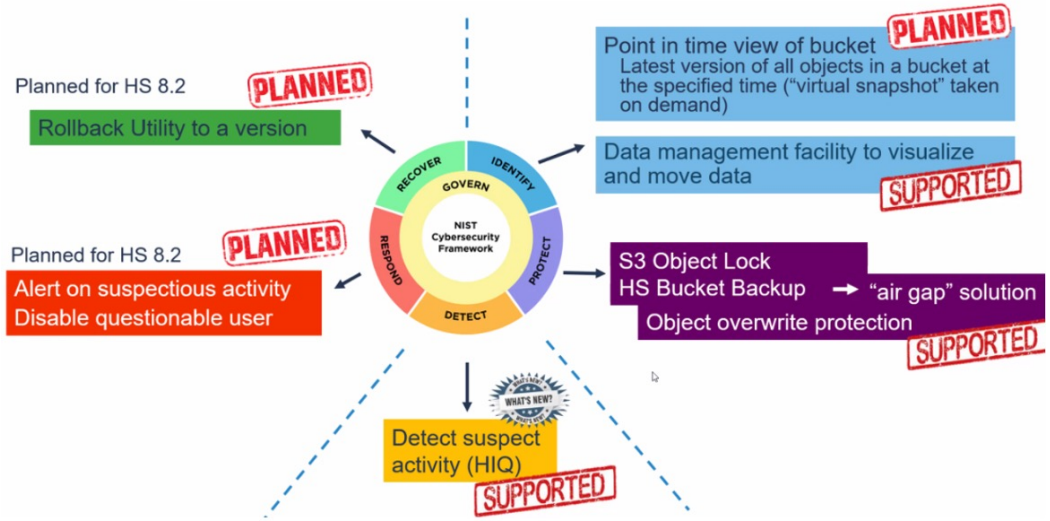
10 use cases where object storage stands out

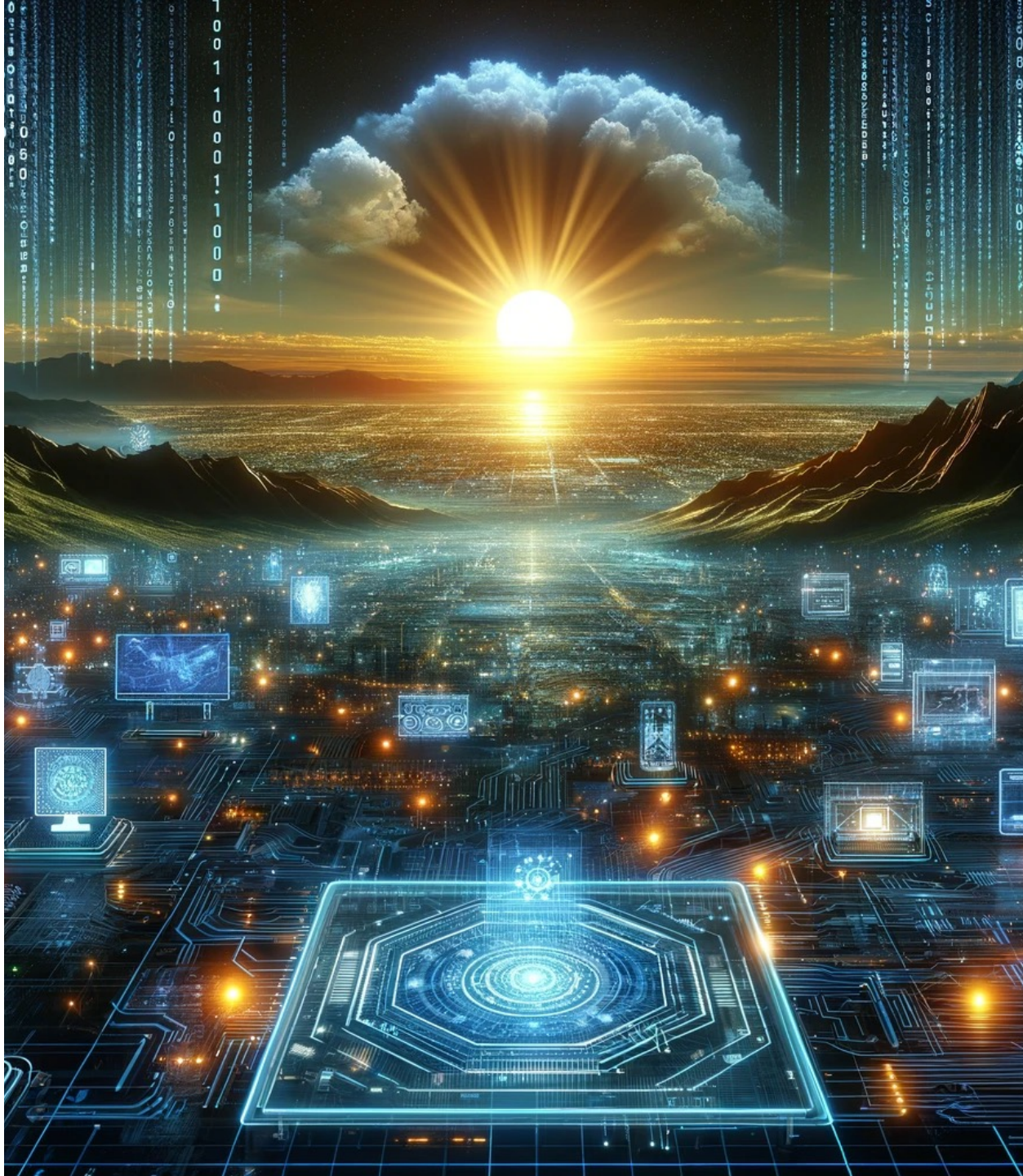
- Archiving
- IoT Data Storage
- Backup and DR
- Content distribution and CDN
- Cloud Storage Services
- Media and Entertainment
- Data Lakes
- Log and event Data Storage
- Data Sharing and Collaboration
- Genomic Data Storage

CLOUDIAN®

The Dawn of CyberStorage

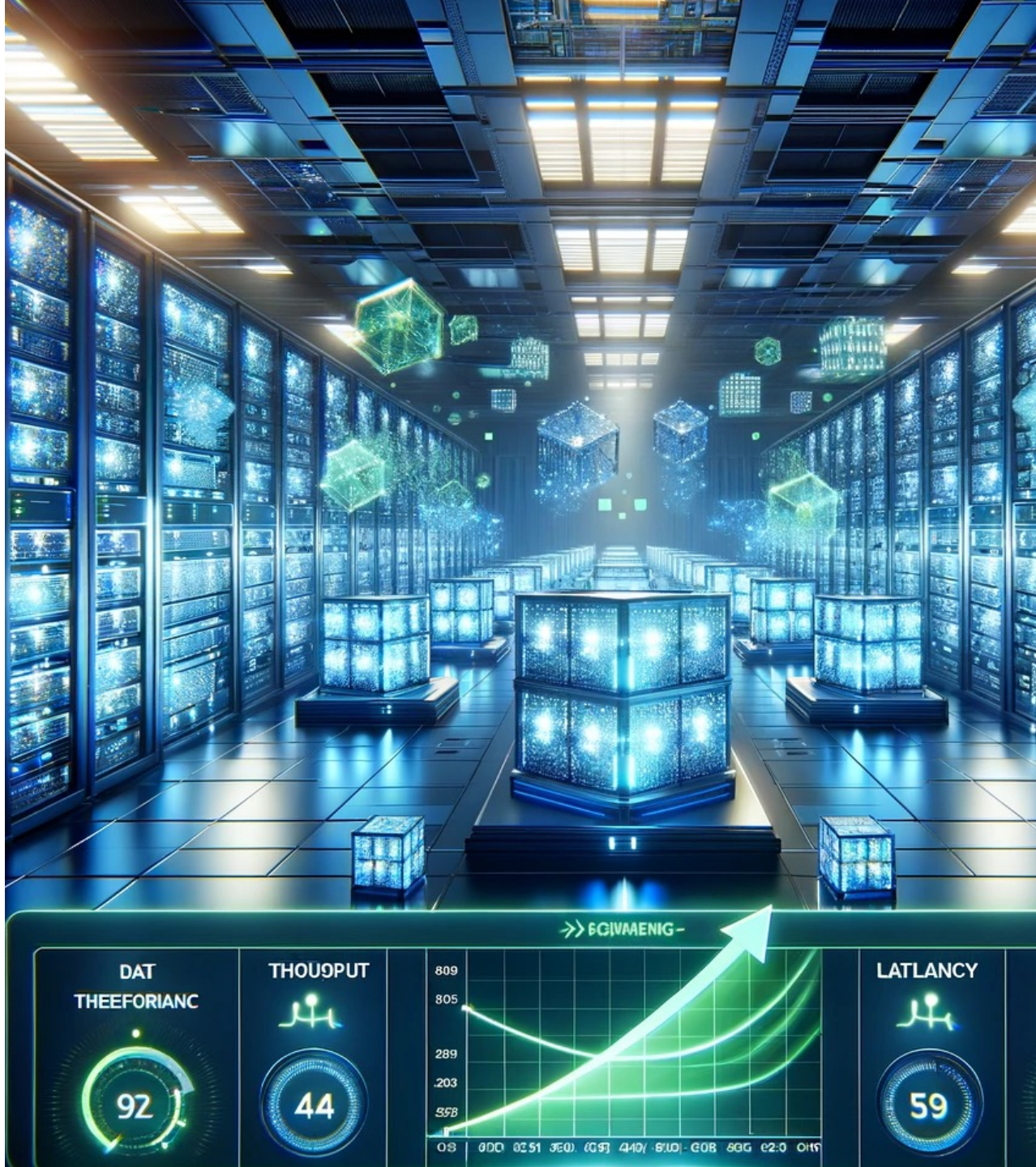
- CyberStorage refers to the use of online platforms and servers to save, manage, and access data over the internet.
- This can include everything from **personal files** and photos to **business data** and **applications**.
- It leverages cloud computing technology, providing users with the ability to store data remotely and **access from any device with internet connectivity**, enhancing flexibility, scalability, and often security through advanced encryption methods and redundancy systems





Object as Primary

- Primary dataset held in object
- Different IO profile to backup & archive
- Store primary data sets on highly available, highly durable platforms
- Application driven
  - Data Lakehouse
  - Analytics
  - Machine Learning (ML)
  - Artificial Intelligence (AI)
- HyperStore lends itself to more performant workloads through
  - Flash nodes & advanced tiering options
  - Hybrid Storage Policies
  - Adaptive bucket rules*
- Eliminate Redundant copies of data
- Eliminate Backup… maybe



The DataLake

- A data lake is a centralized repository that allows you to store all your structured and unstructured data at any scale.
- It serves as a foundation for analytics and allows for the storage of raw, unprocessed data as well as processed data.
- HyperStore provides flexible deployment & architecture options
  - Topology
  - Architecture options
  - Configurable protection schemes

DataLake Ecosystem

Applications with native S3 support

Arrow

- Apache Arrow is a language-agnostic software framework for developing data analytics applications that process columnar data.

Dremio

- Based on open-source technologies, including Apache Iceberg and Apache Arrow, Dremio provides an open lakehouse architecture

Druid

- Apache Druid is a real-time analytics database designed for fast slice-and-dice analytics ("**OLAP**" queries) on large data sets. Most often, Druid powers use cases where real-time ingestion, fast query performance, and high uptime are important.

Teradata

- Teradata can perform powerful OLAP(Online Analytical Programming) functions and process the data parallelly. Teradata provides better performance and linear database scalability than other DBMS data structures.

Cribl

- Cribl Stream helps you process machine data – logs, instrumentation data, application data, metrics, etc. – in real time, and deliver them to your analysis platform of choice. It allows you to: Add context to your data, by enriching it with information from external data sources.

Splunk

- Splunk is a big data platform that simplifies the task of collecting and managing massive volumes of machine-generated data and searching for information within it. The technology is used for business and web analytics, application management, compliance, and security.

Veritca

- Vertica is the unified analytics platform, based on a massively scalable architecture with the broadest set of analytical functions spanning event and time series, pattern matching, geospatial and end-to-end in-database machine learning.

Imply

- Imply develops and provides commercial support for the open-source Apache Druid, a real-time database designed to power analytics applications.

Greenplum

- Greenplum Database is a massively parallel processing (MPP) SQL database built on PostgreSQL. It can scale to a multi-petabyte-level data workloads and allows access to a cluster of powerful servers that will work together within a single SQL interface where you can view all the data.

MS SQL

- SQL Server 2022 introduces new object storage integration to the data platform, enabling you to integrate SQL Server with S3-compatible object storage, in addition to Azure Storage. The first is backup to URL and the second is Data Lake Virtualization.

How it works: Machine Learning Example

1) Archive Text & Image Files are digitised, then ingested into an S3 bucket

2) A containerised OCR Process runs against the assets (typically images and PDFs)

3) This process generates XML Files are then stored in the bucket along with the assets

4) The XML Files are then read by Elastic and indexed to make searchable

IO PATTERN: WRITE > READ > WRITE > READ > KEEP

Object Storage

Features that benefit AI workloads

Scalability

Cost

Accessibility

Versioning

Pre-processing

Model serving

Compliance

S3 can be leveraged for various deep learning applications

Here are some specific applications that can utilize S3 storage for deep learning:

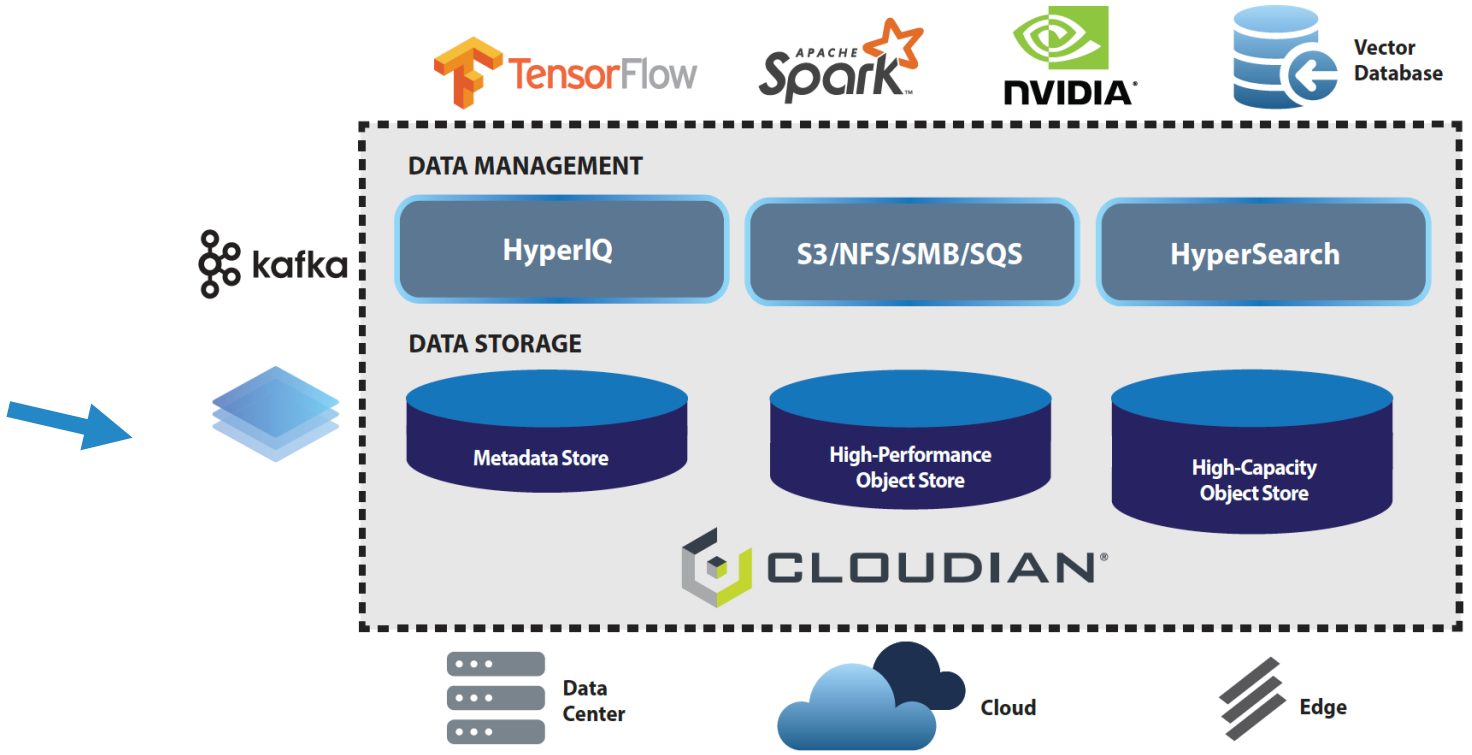
1. **Image Classification and Recognition**: Deep learning models for image classification and recognition often require large datasets of labelled images.
2. **Natural Language Processing (NLP):** NLP tasks such as text classification, sentiment analysis, and language translation require large corpora of text data.
3. **Object Detection:** Models for object detection typically require annotated datasets containing images & corresponding bounding boxes or segmentation masks.
4. **Recommendation Systems:** Deep learning-based recommendation systems often rely on large-scale datasets of user interactions (e.g., clicks, purchases, ratings).
5. **Time Series Forecasting:** For time series forecasting tasks such as stock price prediction, demand forecasting, or anomaly detection,
6. **Medical Imaging Analysis:** Models for medical imaging analysis, such as diagnosing diseases from medical images or segmenting anatomical structures
7. **Autonomous Vehicles:** Deep learning models used in autonomous vehicles and robotics applications often require large datasets of sensor data

The Data Platform is the foundation of AI

Data Platform

- Integrated set of technologies
- Meets end-to-end data needs

- Exabyte Scalable
- Secure
- Fault Tolerant
- IAM / Tags / Multi Tenancy
- Hybrid Cloud
- Geo Distributed



Cloudian meets AI needs

AI Requirement	CLOUDIAN®
1) **Scalable infrastructure** - Meet the massive AI storage demands - Seamlessly grow as data volumes expand	 - Limitlessly scalable - Modular
2) **Data governance and security** - Protect data from unauthorized access - Defend against ransomware - Comply with data sovereignty requirements	 - Security certifications - Data immutability
3) **Performance** - Rapid storage and retrieval - Real-time access	 - Flash-compatible - Parparallel processing
4) Cloud integration - Compatible with cloud tools - Data mobility to/from cloud	 - Supports cloud data protocols - Full cloud-integration