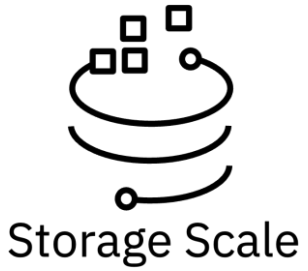
IBM Storage Scale

Heiko Lehmann
Tech Sales Leader Storage4AI & BigData DACH

Disclaimer

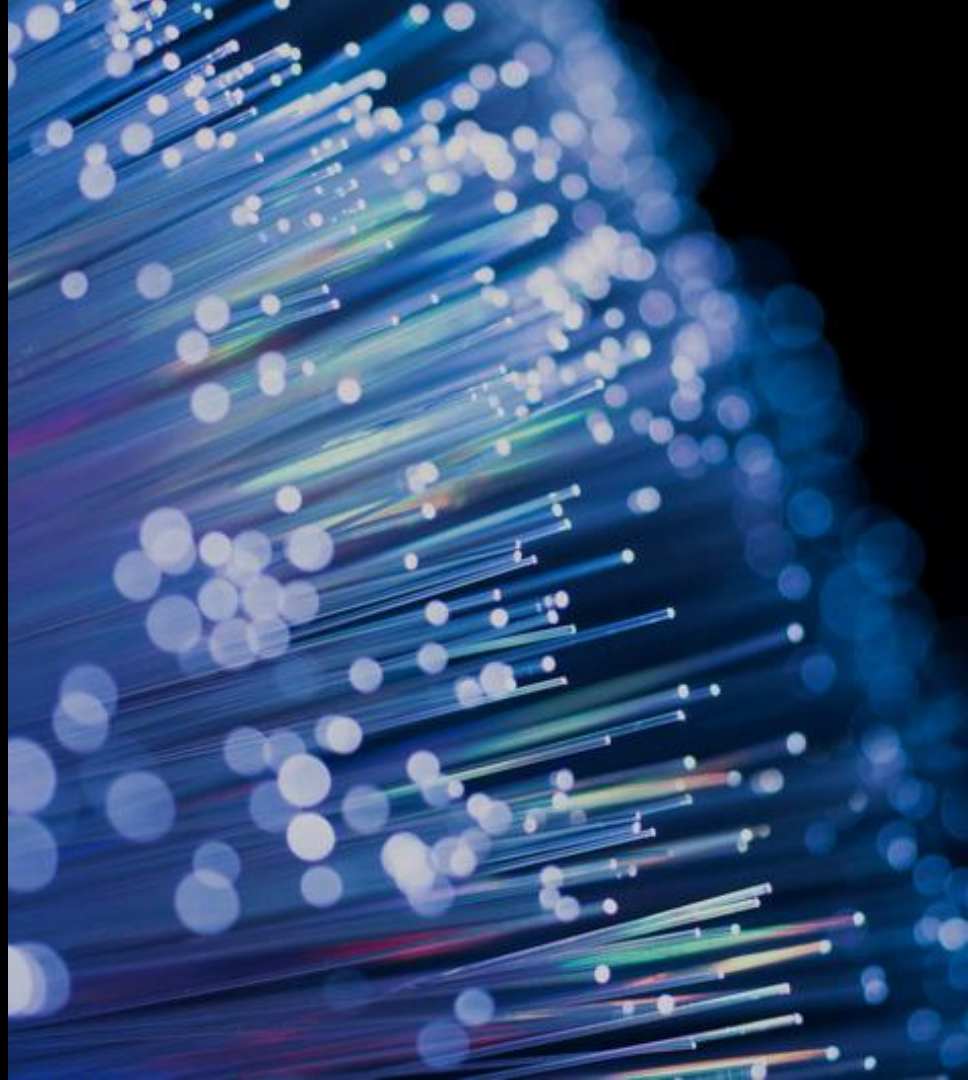


IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

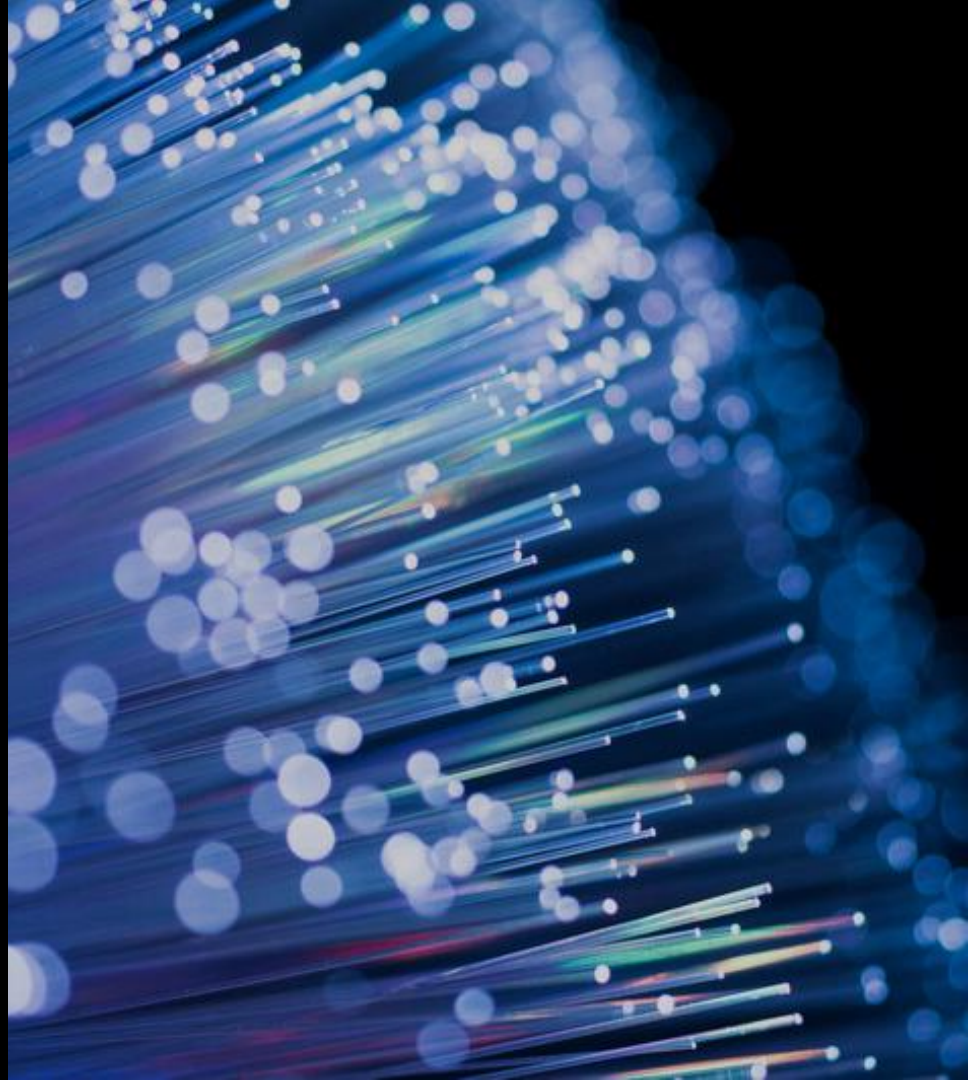
IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

GPFS 1.1
released 1998

Thank you for
25 years!!!



Market dynamics and challenges



pain points

- New generations of application, such as AI and analytics, bring new challenges of storage infrastructure. Distinctive qualities required include:
  - Ability to match performance requirements
  - Ability to scale anywhere from small to extremely large deployments
  - Ability to share data to different applications
  - Ability to minimize data movement, data copies
  - Ability to manage the data complexity
- Cyber resiliency requirements mandate securing the data storage layer

- Lack of expertise to build and deploy, for example hardware, networking and so on
- Time consuming to design, implement, optimize, and test the solution before deployment
- Risks in unproven architecture

Storage for Data and AI - POV

Workloads

Collect: Video, medical images, cloud storage, IoT, log files, genomics, big data

Organize: AI analysis, governance, operations

Analyze: HPC, analytics, AI, ML, DL

Problem

The best AI is built on a foundation of data that is collected and organized as carefully as it is infused into the business.

But infrastructure challenges impede progress.

Data silos and complex storage solutions that are not easily connected, make it difficult to get a holistic and timely view of the growing amounts of information.

Infrastructure that was not built for AI is not flexible enough to respond to new demands and provide the performance for fast and global insights.

Value

IBM Storage for Data and AI is infrastructure simplified for AI workloads.

It provides an optimized foundation for each stage of the AI ladder.

It eliminates and consolidates data silos driving faster and lower cost results.

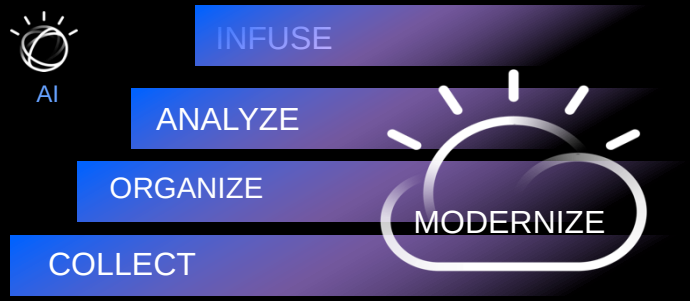
IBM Storage for Data and AI solutions are massively scalable and globally available. They provide faster insights and allow your infrastructure to grow with your business and your data needs.

Only IBM

- Provides 3X better performance and up to 1000X better scalability, saving OPEX and CAPEX
- Supports seamless archive to tape lowering cost of storage by up to 75%*
- Supports active cloud storage acceleration boosting performance up to 40GB/s and lowering cost of storage by up to 50% **
- Supports a heterogenous data catalog and policy engine that can scan a billion records in less than ½ second
- Offers time saving and improved data quality, speeding time to insights with continuous time real-time ingest***
- Supports a data catalog on Red Hat OpenShift for deployment on multi-cloud platform
- Offers suite licensing for IBM Storage AI software for simplicity and up to 40% lower cost
- Leverage file and object data to IBM Watson solutions and IBM Cloud Paks for Data

Comprehensive Storage for the AI Journey

- **Complexity**
- **Costs**
- **Integration**



Collect

(Retain)

Data lake for files

IBM ESS 5000

IBM Storage Scale

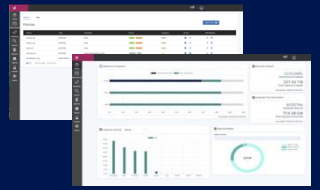
Red Hat OpenShift Integrated

Data lake for objects

IBM COS

IBM Cloud Object Storage

Organize



AI policy engine and

Data Catalog

IBM Storage Discover

Red Hat OpenShift Integrated

Analyze

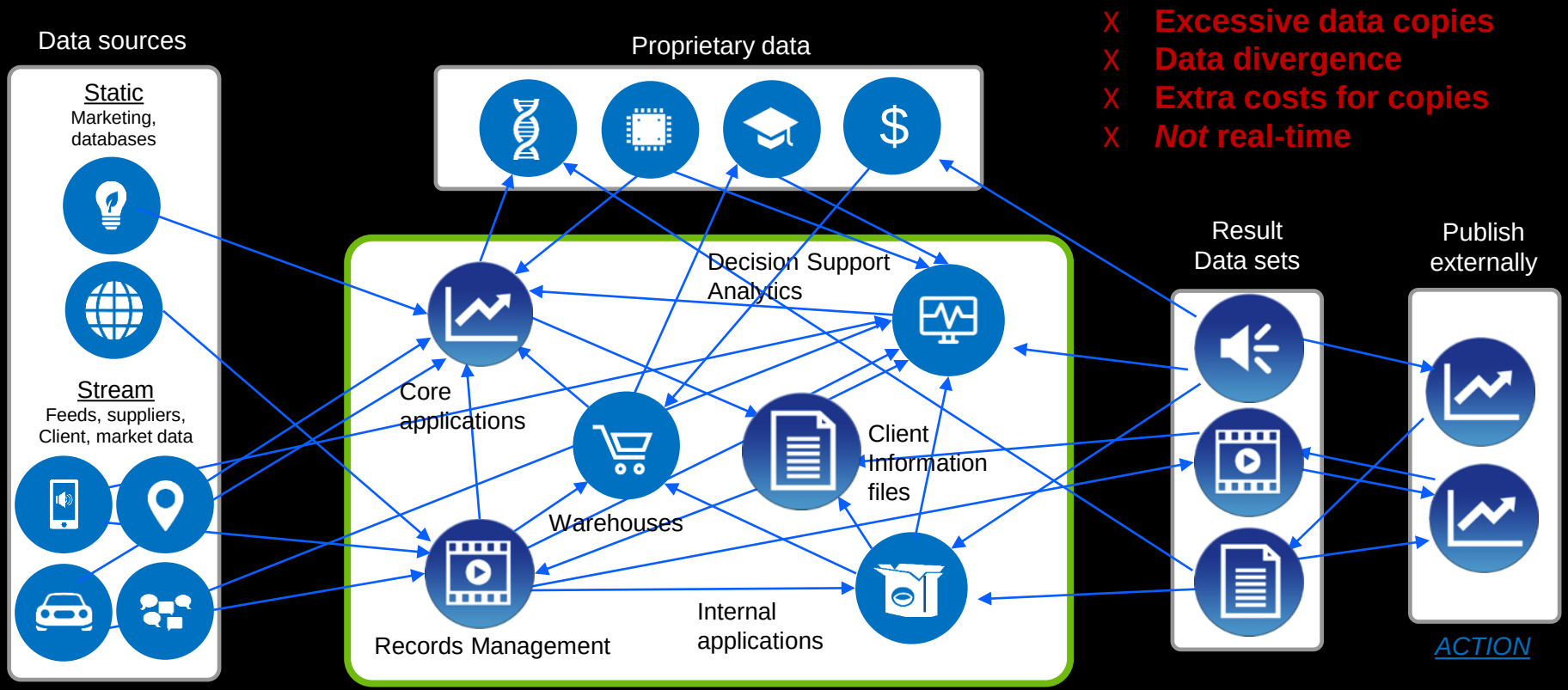
NVMe flash nodes

IBM ESS 3500

IBM Storage Scale

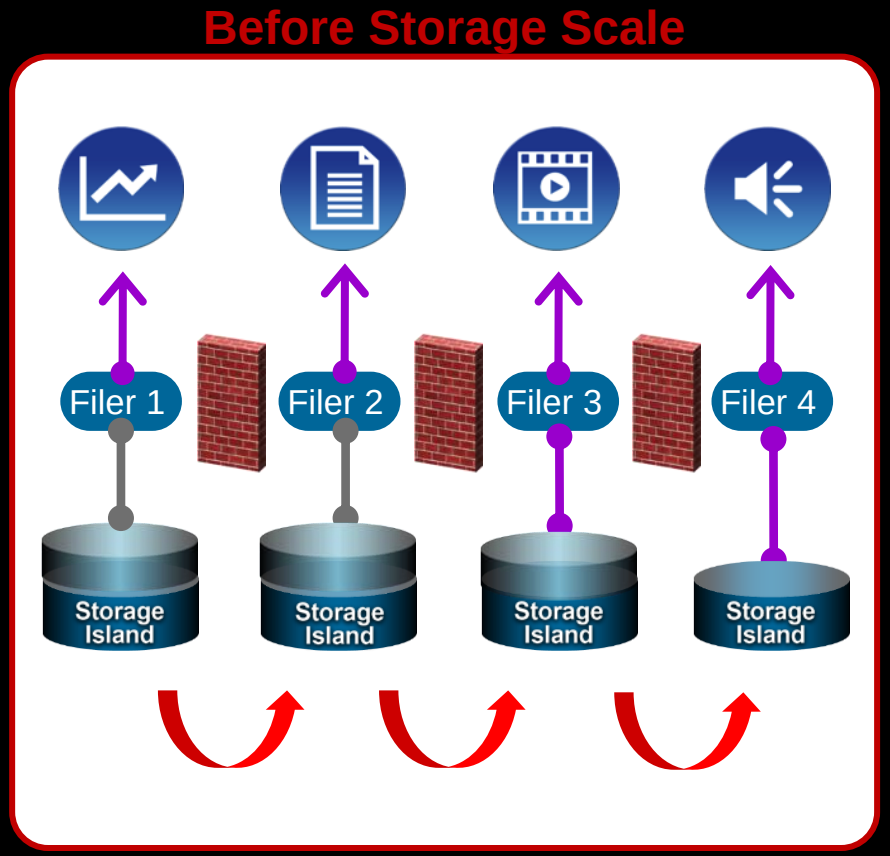
Red Hat OpenShift Integrated

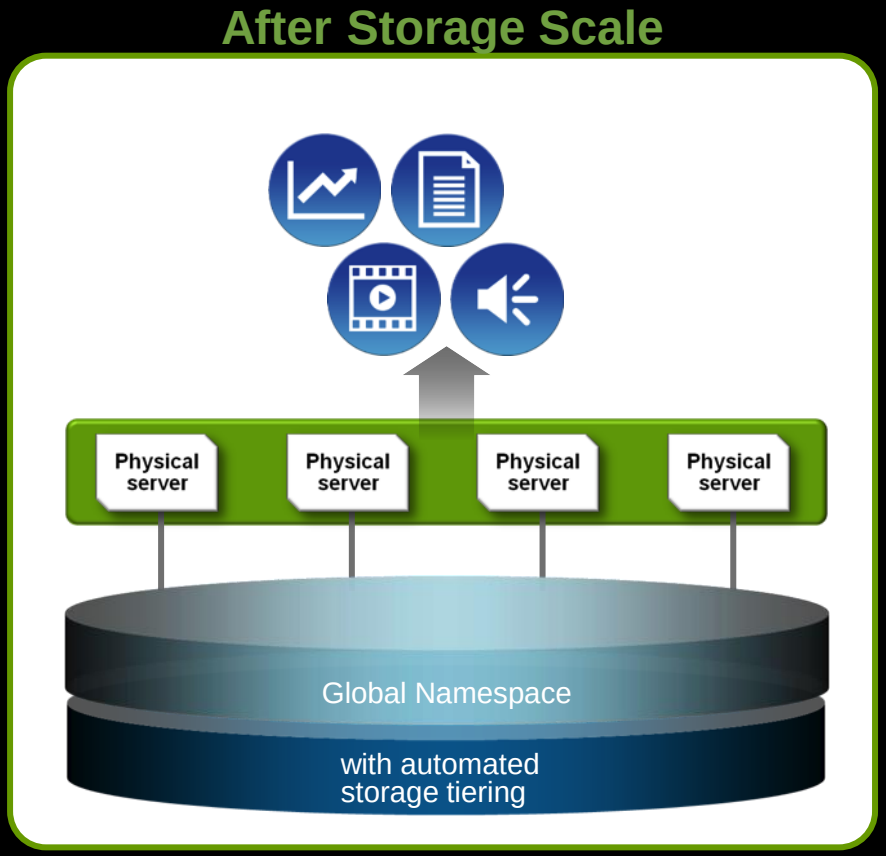
Enterprise data problem: complicated, massive, siloed, and costly



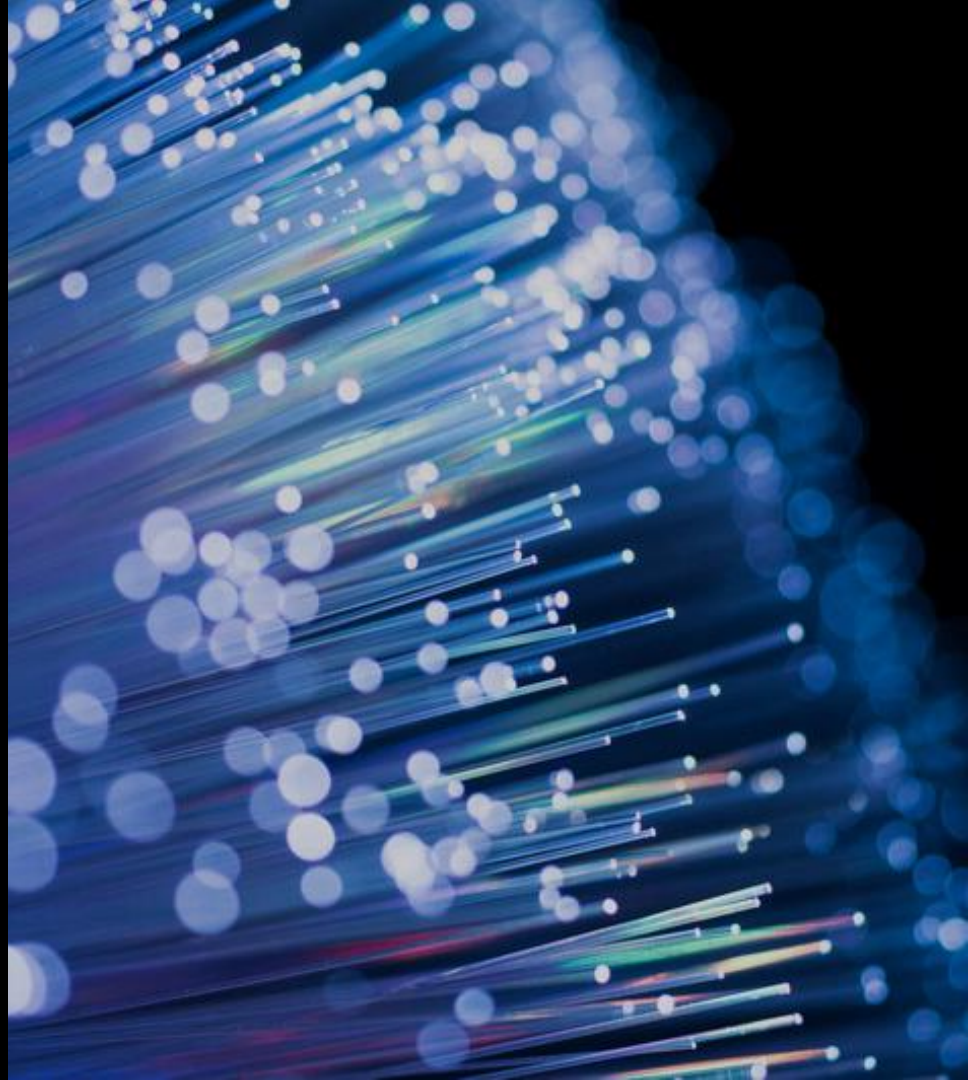
- ✗ **Excessive data copies**
- ✗ **Data divergence**
- ✗ **Extra costs for copies**
- ✗ ***Not* real-time**

High performance file storage data architecture





Storage Scale overview



IBM Storage Scale

Highly scalable high-performance unified storage for files and objects with integrated analytics

Performance: remove data-related bottlenecks
- with a parallel, scale-out solution
- 2.5TB/s demonstrated throughput

Ease of management: enable global collaboration
- with unified storage and global namespace
- Data Lake serving HDFS, files and object across sites

Economics: optimize cost and performance
- with automated data placement
- thin-provisioning preview and TRIM support, QOS on project preview

Robust: ensure data availability, integrity and security
- with erasure coding, replication, snapshots, and encryption
- end-to-end checksum, Storage Scale RAID, NIST/FIPS certification

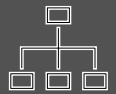
IBM Global Data Platform for Unstructured File & Object Data

Unstructured Data Services Framework

Applications and Workloads

Data Access Services

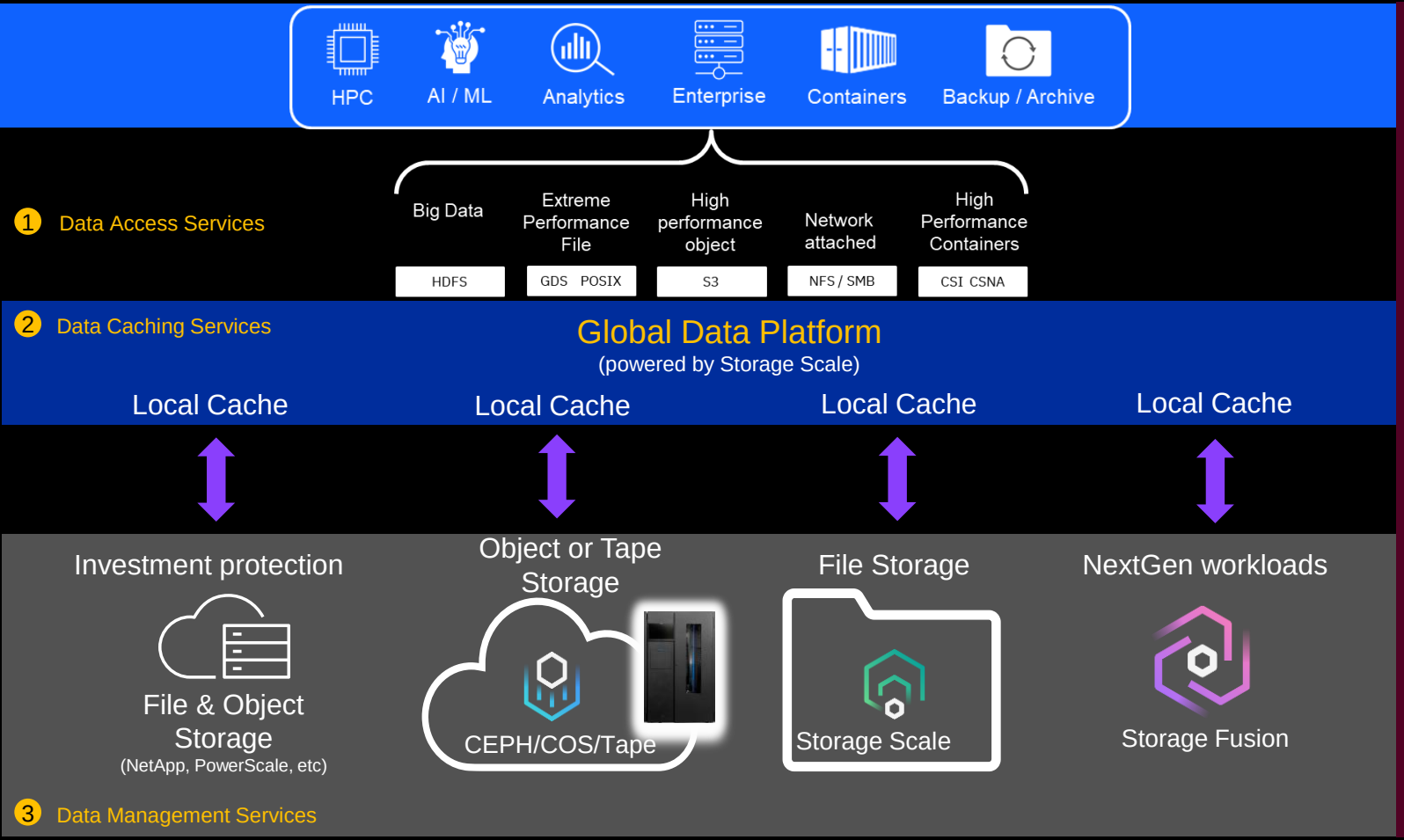
Data Caching Services



Data Management Services

Data Security Services

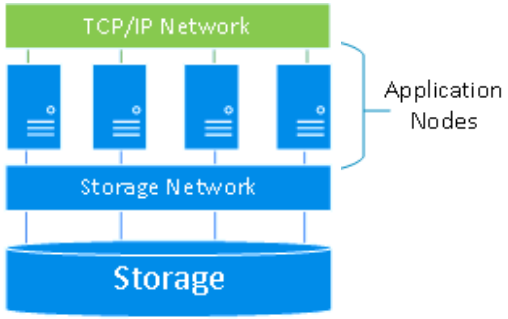
IBM’s Global Data Platform for File & Object Data



Storage Scale deployment models

Enterprise Integrated Model

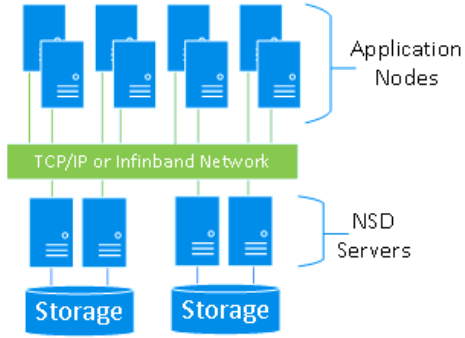
(SAN, NVMeoF, iSCSI)



Unify and parallelize storage silos

Network Shared Disk (NSD) Model

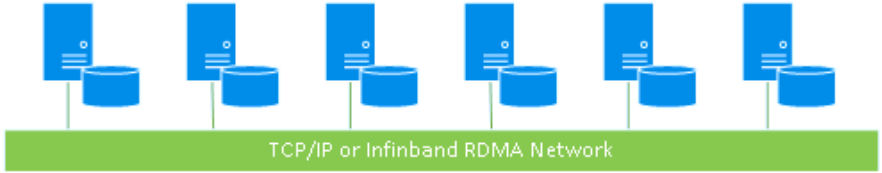
(twin tailed storage, erasure code)



Modular High-Performance Scaling

Shared Nothing Cluster (SNC) Model

(Storage Rich Servers (replication, erasure code))



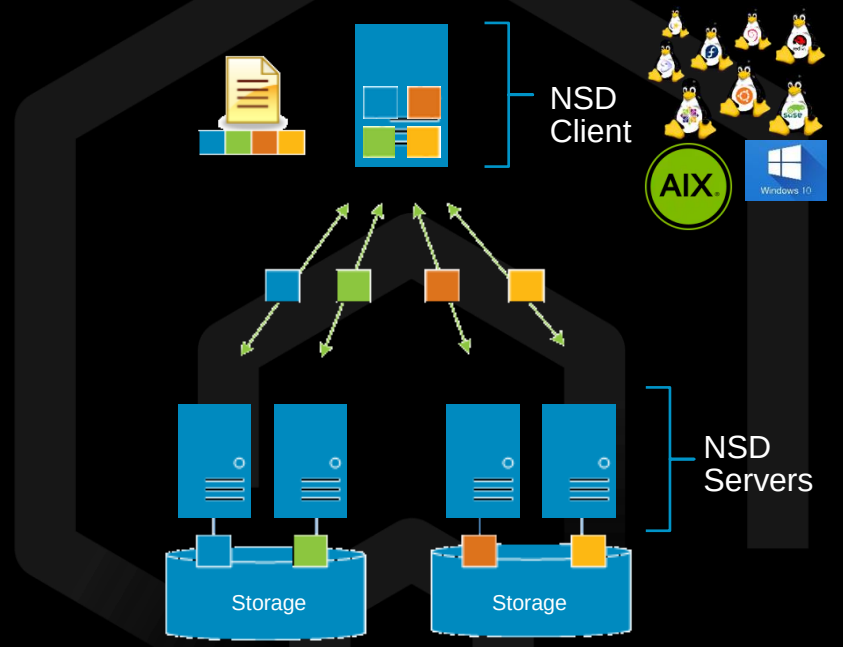
Span storage rich servers for converged architecture or HDFS deployment

High performance parallel architecture

High performance shared storage platform
for end-to-end collaborative common enterprise
data platform, analytics, and AI workflows

Parallelism eliminating bottlenecks / hot spots

- All NSD servers export to all clients in active-active mode
- Storage Scale stripes files across NSD servers and NSDs in units of file-system block-size
- File-system load is spread evenly
- Easy to scale file-system capacity and performance while keeping the architecture balanced



NSD client does real-time parallel I/O
to all the NSD servers and storage volumes/NSDs

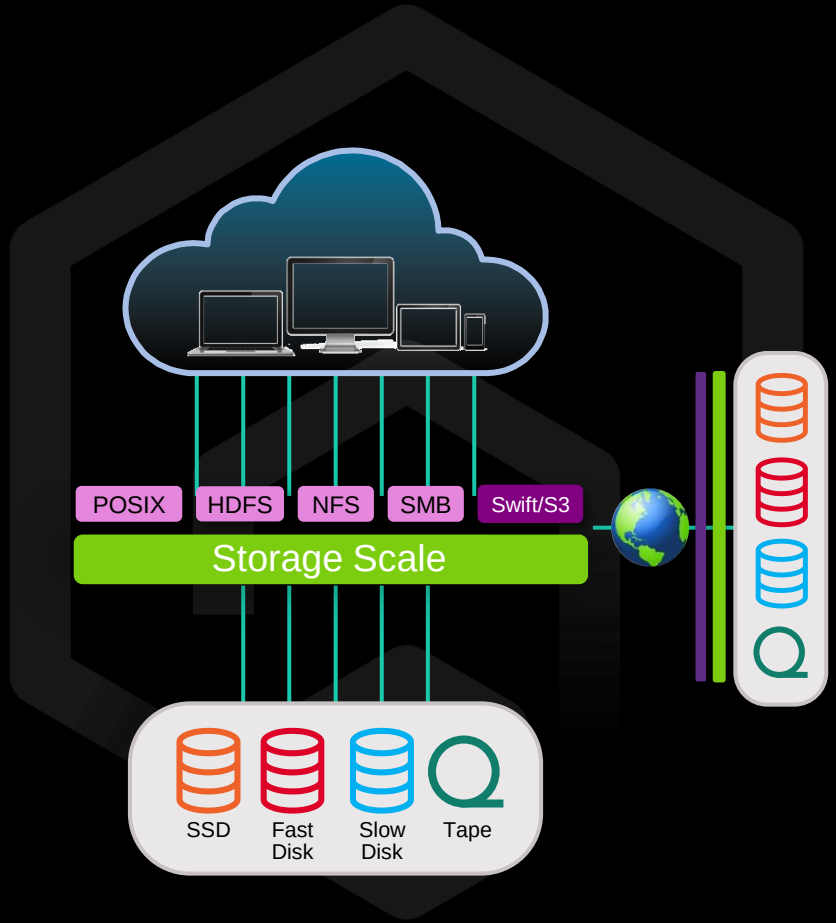
Storage management at scale

- Simplifies management by combining workflows on a single common enterprise data platform
  - Single global namespace can share data across the enterprise
  - Eliminate redundant copies of data
- Extensive monitoring and analytics
- Intuitive GUI supporting many capabilities
  - Performance, capacity, network monitoring
  - Active File Management
  - Transparent Cloud Tiering
  - Enhanced maintenance and support, including interaction with IBM remote support via CallHome
- Billions of files and yottabytes of data

  (yottabyte = 1.000.000.000 petabyte)
- Faster and simpler out-of-the-box experience
  - Easier setup and improved performance with less manual configuration

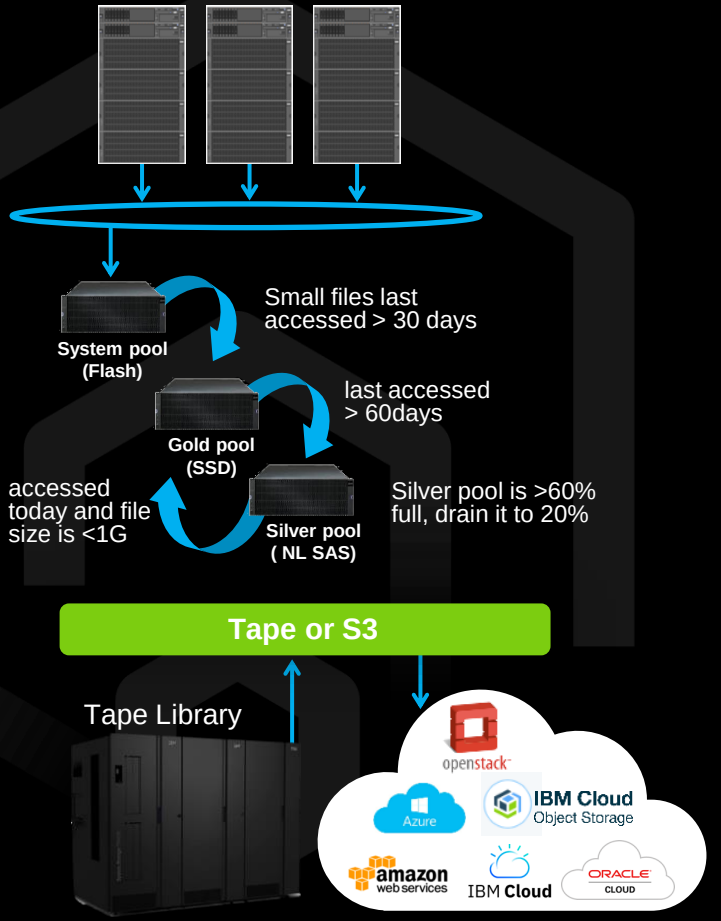
Store everywhere. Run anywhere.

- Unified Scale-out Data Lake
  - Access data using multiple protocols
  - High-performance concurrent access with integrity
  - Analytics on demand
  - Single management plane
  - Cluster replication and global namespace
  - Enterprise storage features across file, object, and HDFS
- Global collaboration with Advanced File Management
  - Filesystem caching and single namespace view across multiple geographically distributed remote sites
  - Extend collaborative workflows
  - Mitigate network bottleneck with advanced routing
  - Flexible configuration with writer and read-only sites
  - Disaster recovery for enterprise resiliency



Improve data economics

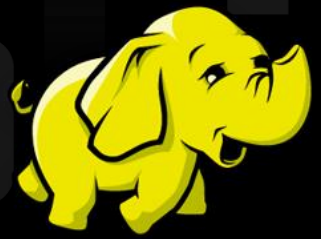
- Information Lifecycle Management with intelligent automatic tiering of data
  - Policy engine for fast metadata scans
  - Automated data movement based on policy
  - Movement among multiple types of storage: Flash, SSD, HDD, external tape, object storage, and cloud
  - Reserves high speed storage for work in progress, moves everything else transparently to lower cost storage tiers
  - Visibility / access to content regardless of storage tiers
- Enterprise scale
  - Seamless expansion and upgrades
- Policy driven compression increases effective capacity



Software-defined open platform

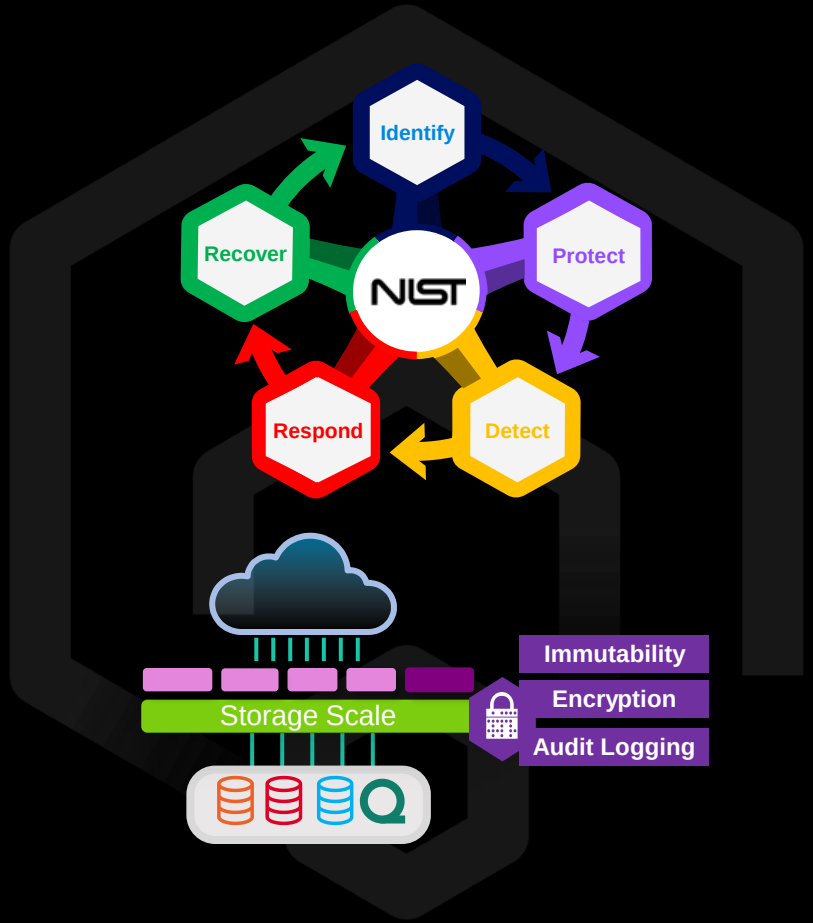
- Hardware agnostic multi-vendor
  - Can be built entirely from enterprise commodity components
  - Mix NVMe, Flash SSD, SAS, NL-SAS, tape, cloud
  - Free from hardware vendor lock-in
- Software defined
  - Software defined storage approach reduces vendor dependency, placing control with the clients
  - Can eliminate expensive proprietary hardware
- Integration with Cloud
  - Transparent Cloud Tiering
- Software, integrated solution or Cloud deployment
- Multi-protocol access to data
  - Industry standard, network file sharing protocols, including SMB, NFS, S3 Object, Hadoop/HDFS

- Transparent integration with Hadoop data
  - Hadoop clusters directly read and write to Storage Scale
  - No waiting for data transfer between storage systems
  - Faster time to results
  - Single Data Lake for all applications
  - Enterprise data management on Hadoop data
  - Hortonworks / Cloudera certification



Security and cyber resiliency

- Centralized authentication and access control
- Data encryption and cryptographically secure erase
- Immutability
- Audit logging
- Data protection through snapshots, replication, backup, and/or disaster recovery
- Data dispersal and erasure code for faster rebuild times
- End-to-end checksum to catch errors
- Integration with IBM Storage family
- NIST/FIPS certification



Cyber Resiliency solution with Storage Scale

The **Storage Scale** existing feature integration with **Storage Protect, Storage Archive,** and **Tape Storage** enables organizations to implement an effective, easy-to-manage, and automated Cyber Resiliency solution.

Download the full Blueprint document here

Cyber resiliency is an organization's **ability to continue** delivering the intended outcomes despite adverse cyber incidents

Robust protection against external cyber events

(For example, malware, ransomware attacks, and human elements)

NIST framework + IBM Storage

Safety of business-critical data

Storage Scale deployment methods

Software

Storage Scale System
Integrated Solution

Cloud service

Storage Scale on the Cloud

- Storage Scale is fully supported by IBM Service/Support for production usage in specific clouds
- Customers can exploit their existing cloud provider skill sets, expertise, cloud accounts
- Deployment characteristics:
  - IBM Cloud: IaaS bare metal servers
  - AWS and Oracle: PaaS fully automated deployment templates
- Licensing model: Bring your own license

IBM Cloud

ORACLE CLOUD

amazon web services

High performance storage
with cloud economics

Fast procurement, provisioning,
and deployment
Minutes to hours, not days

Agile scalability
for HPC requirements
Off premises bursting; Hybrid multi-cloud

Storage Scale supports OpenShift and Container Storage Interface

- The IBM Storage Scale Container Storage Interface (CSI) driver enables container orchestrators, such as Kubernetes and Red Hat OpenShift, to manage the life-cycle of persistent storage for containers
- Storage Scale supports new CSI and Red Hat OpenShift releases
- Open source CSI driver available at: https://github.com/IBM/ibm-spectrum-scale-csi

Storage Scale editions and licensing at a glance

Editions have various function levels:

- Data Access edition (DAE) - often used for HPC
- Data Management edition (DME) - adds functions valuable in commercial environments
  - Free Developer edition (DE)
- Erasure Code edition (ECE) - aimed at hyperscale, web-scale service providers

Capacity licensing: built for simplicity

- Easy to purchase, expand, budget, renew
- Entitled to unlimited number of IBM Storage Scale client and server licenses

Existing IBM Storage Scale socket-licensed customers

- Can stay on existing sockets-based licensing for as long as they wish
- Passport Advantage site ID defines boundary

Feature	Data Access or Standard Edition	Data Management, Advanced or Developer Edition	Erasure Code Edition
Multi-protocol scalable file service with simultaneous access to a common set of data	Yes	Yes	Yes
Facilitate data access with a global namespace, massively scalable file system, quotas and snapshots, data integrity and availability and filesets	Yes	Yes	Yes
Simplify management with GUI	Yes	Yes	Yes
Improved efficiency with QoS and compression	Yes	Yes	Yes
Create optimized tiered storage pools based on performance, locality, or cost	Yes	Yes	Yes
Simplify data management with Information Lifecycle Management (ILM) tools that include policy-based data placement and migration	Yes	Yes	Yes
Enable worldwide data access using AFM asynchronous replication	Yes	Yes	Yes
Immutability (WORM / Write Once Read Many)	Yes	Yes	Yes
Asynchronous multi-site Disaster Recovery		Yes	Yes
Hybrid cloud (Transparent Cloud Tiering)		Yes	Yes
Protect data with native software encryption and secure erase, NIST compliant and FIPS certified		Yes	Yes
File audit logging		Yes	Yes
Watch folder		Yes	Yes
Erasure coding	Scale System only	Scale System only	Yes

Positioning and integration within the IBM Storage portfolio

PowerAI

Genomics
Medicine
Workload

NVIDIA
IBM SpectrumAI

Blueprint and
Reference architecture

Consumed as services on Public Clouds

amazon web services
IBM Cloud
ORACLE CLOUD

Cloud Data Services

IBM Storage Scale

IBM Storage solutions

IBM Storage Suite for IBM Cloud Paks

IBM Storage Scale
IBM Cloud Object Storage
IBM Spectrum Virtualize for Public Cloud
IBM Storage Discover
Red Hat OpenShift Container Storage
Red Hat Ceph Storage

Additional IBM Storage offerings it works with

IBM Spectrum Virtualize

IBM Storage for Data and AI

IBM Storage Scale

ESS 3500 and ESS 5000

IBM Cloud Object Storage

IBM Storage Discover

IBM Storage for Data Resilience

IBM Storage Scale

IBM Storage Discover

DS8000

IBM Tape

IBM Storage Protect Plus

IBM Storage CDM

IBM Storage Protect

IBM Storage Archive


IBM