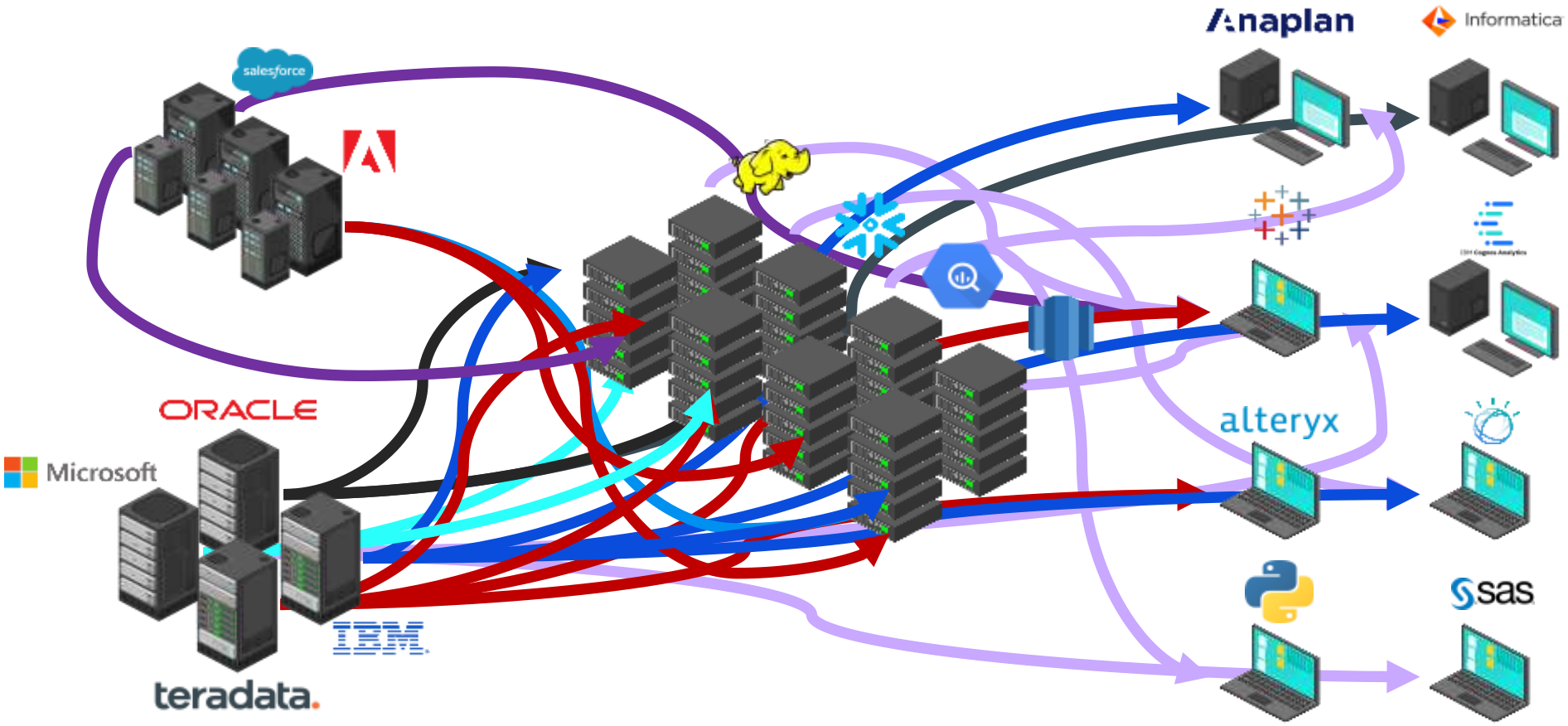
IBM Vendor Update - Storage

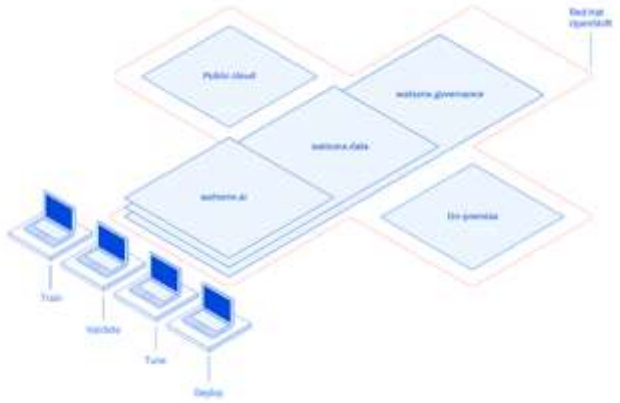
Chris Maestas
Chief Architect for Storage File and Object Systems
IBM Data and AI Storage Solutions
https://www.linkedin.com/in/cdmaestas

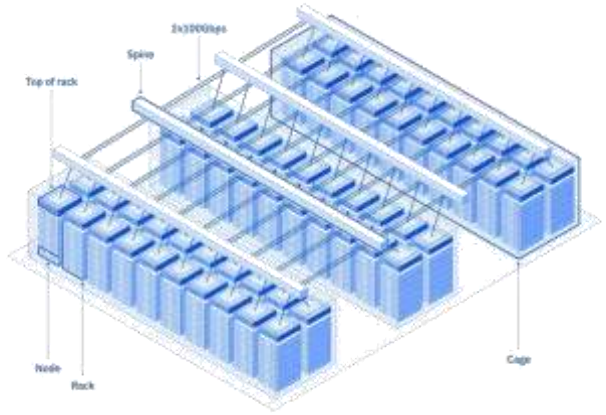
Data is everywhere in a hybrid and multi-cloud world and the

compute (GPU or CPU or DPU) wants from remote to local!



Storage requirements for AI and HPC





Tuning/Inferencing

watsonx .ai	Storage Acceleration Storage Abstraction	Efficient GPU support	Rapid deployment HA/DR/Backup Simplified Day-2 operations
watsonx .data		Metadata catalog integration	
watsonx .governance			

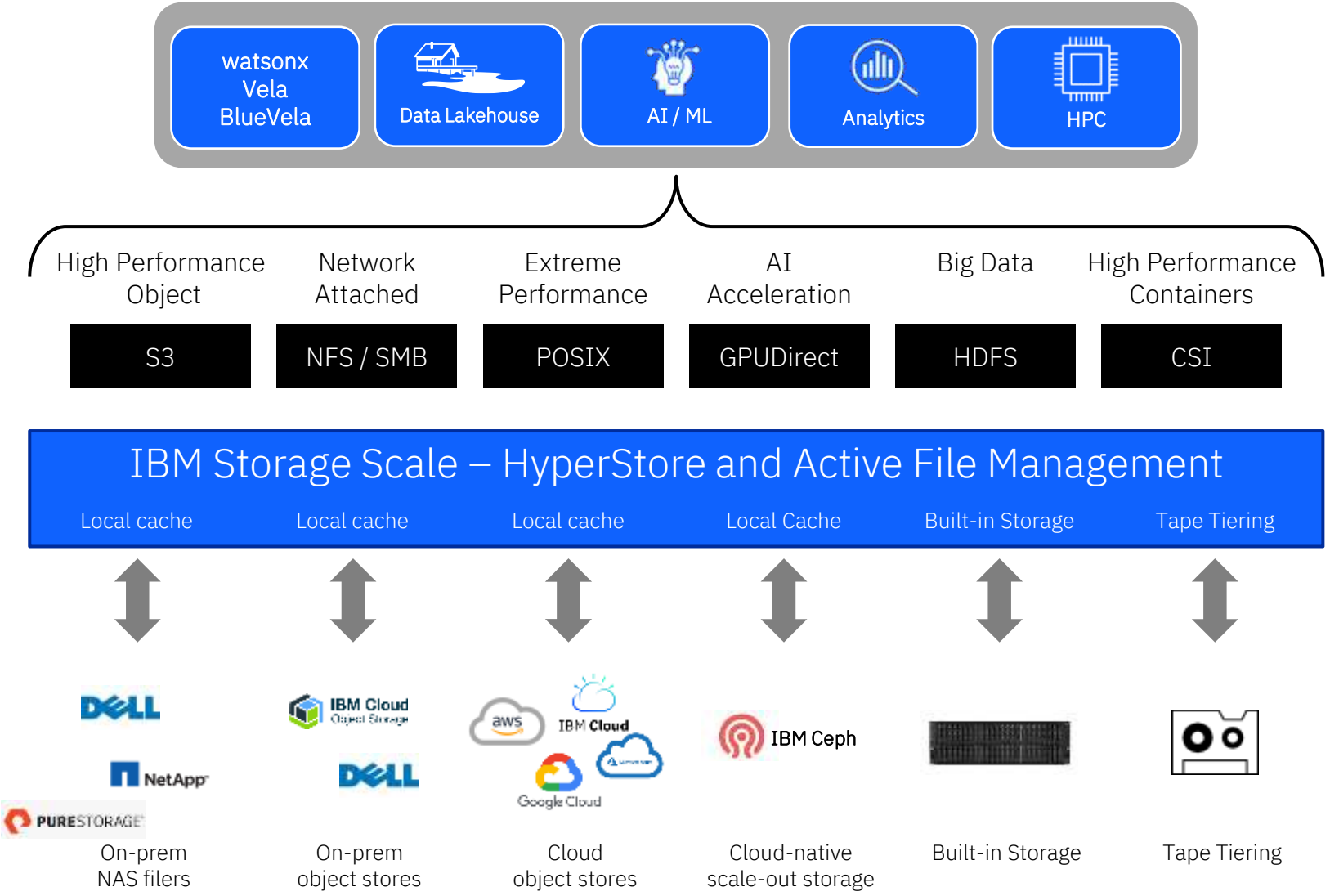
Training and Modeling

Maximum Performance	Efficient GPU support High bandwidth Low latency
Scalability	Linear capacity scaling High density

IBM Storage Scale – Capability Highlights

Storage Access, Abstraction and Acceleration to maximize CapEx investment

Storage Scale



Multi-Protocol Support Access

Simultaneous multi-protocol access including GPUDirect support

Outcome: Enable globally dispersed teams to collaborate on data regardless of protocol, location or format

Storage Acceleration

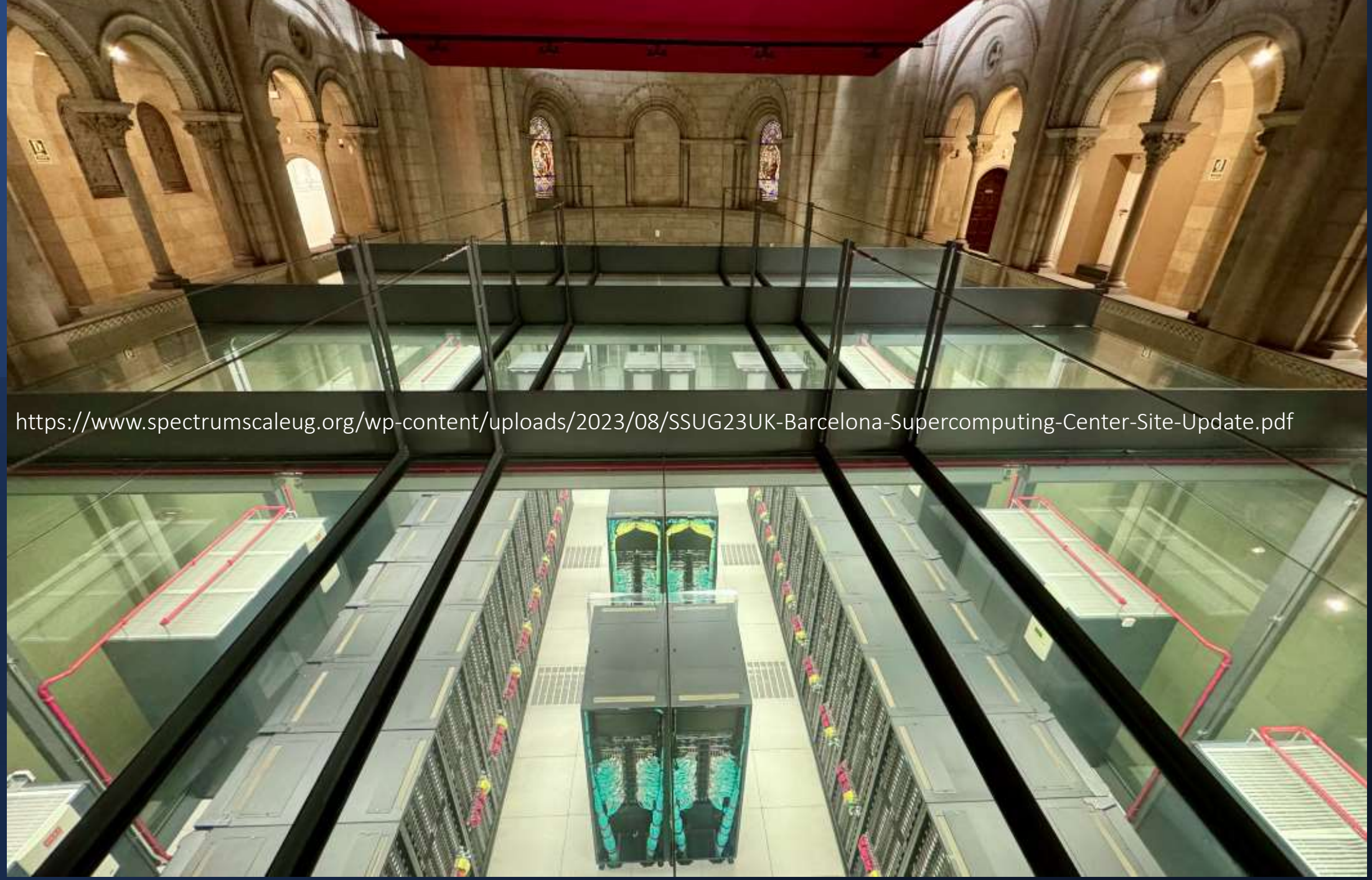
Automatic, transparent caching of back-end storage systems

Outcome: Accelerates data queries and improves economics by fronting lower performance storage

Storage Abstraction

Single global namespace delivers a consistent, seamless experience for new or existing storage

Outcome: Reduce unnecessary data copies and improve efficiency, security and governance


https://www.spectrumscaleug.org/wp-content/uploads/2023/08/SSUG23UK-Barcelona-Supercomputing-Center-Site-Update.pdf

MN5: IO Partition

https://www.spectrumscaleug.org/wp-content/uploads/2023/08/SSUG23UK-Barcelona-Supercomputing-Center-Site-Update.pdf

ESS model	#ESS	Drive Capacity	Total # drives	Raw capacity	Net capacity	Read perf	Write perf
ESS 3500 Capacity model	50	NL-SAS 18TB	20400	367PB	248 PB (8+3P)	1.6TB/s (IOR 100%read)	1.2TB/s (IOR 100%read)
ESS 3500 Performance model	13	NVMe 15.36TB	312	4.79PB	2.81PB (8+2P)	600GB/s 1Mio iops 4KB	600GB/s 500K iops 4KB

Total net storage capacity: **650** PB

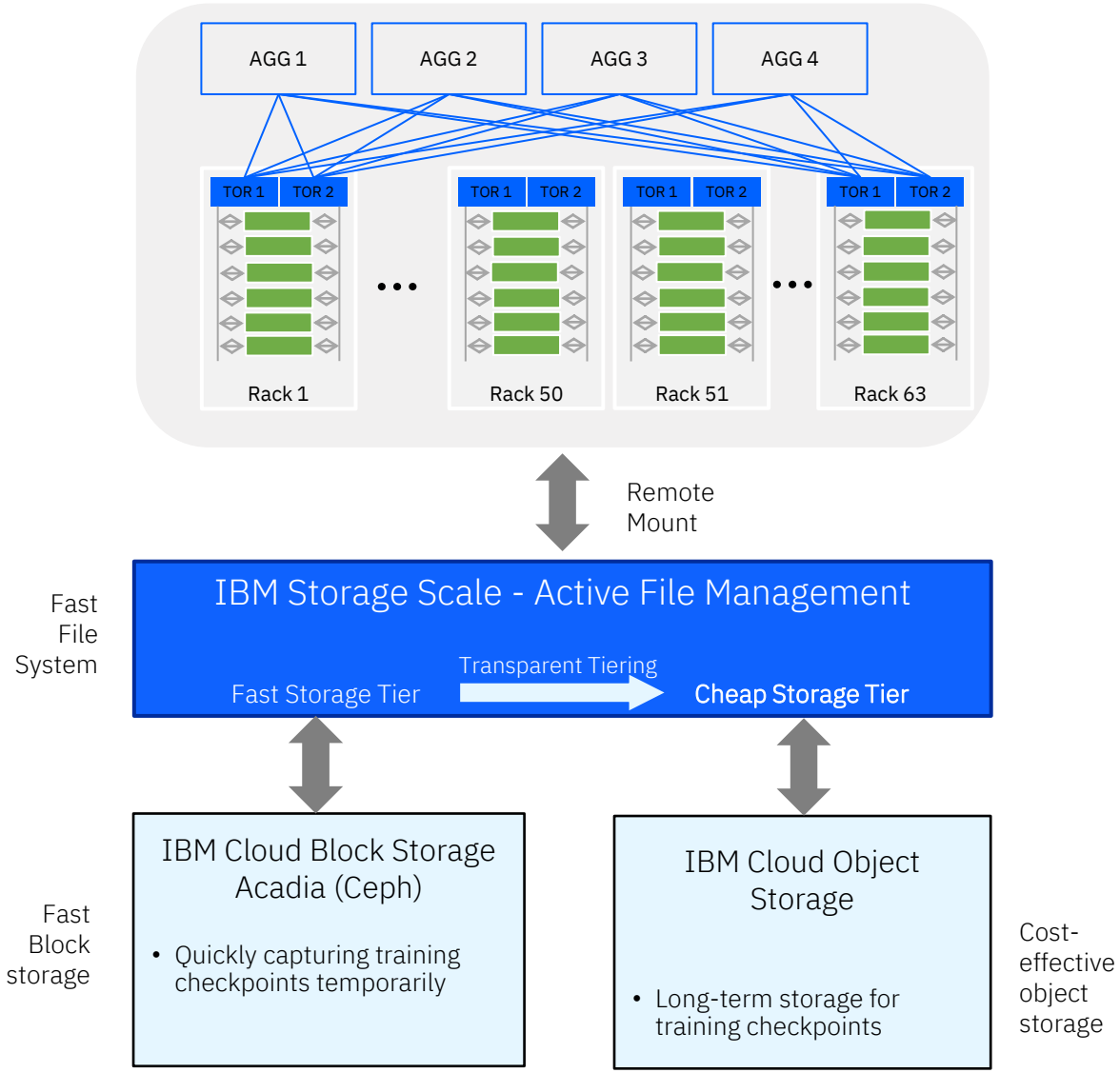
Element	Element	Size
IBM TS4500	2	
Tape Enterprise	20100	400 PB
Drives	64	

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Julich Lab Jupiter Exascale HPC with IBM Storage

IBM Storage Scale

An integral part of the Vela architecture



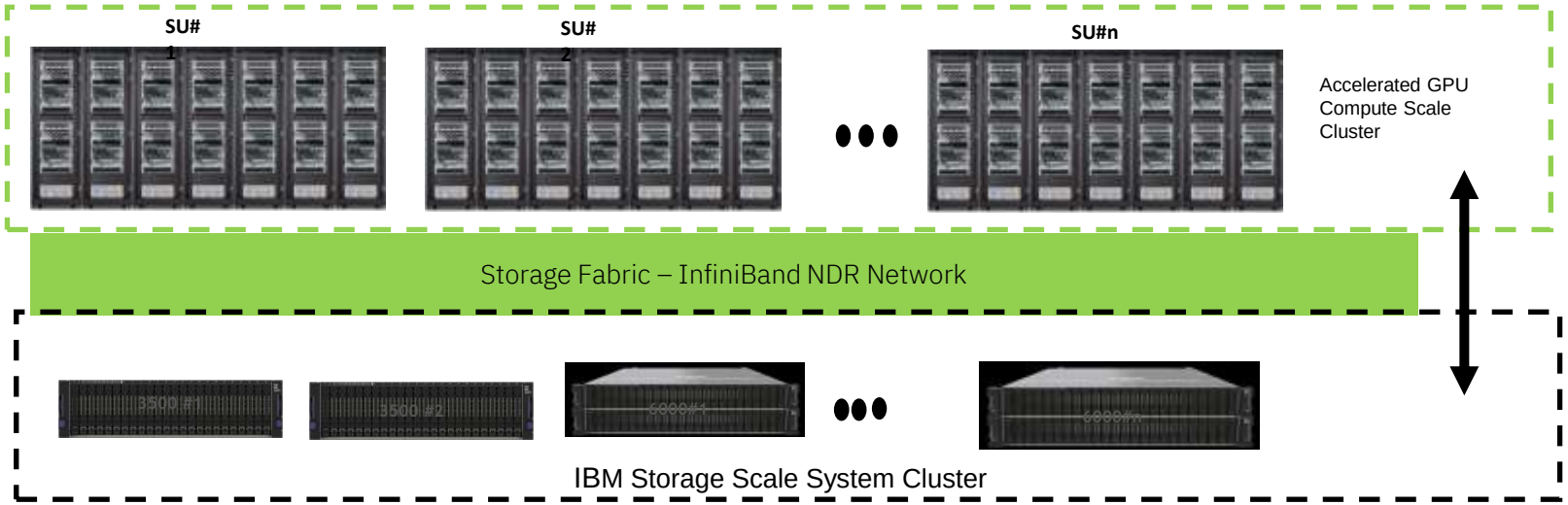
- Built completely on **IBM Cloud** infrastructure
- Dedicated **IBM Storage Scale cluster** on IBM Cloud instances
  - Cloud-Native Scale Access (CNSA) on GPU compute cluster
  - Shared POSIX file system semantics
  - One volume for training data
    - Fit complete training dataset
  - One volume for checkpointing
    - Can accumulate ~10 days of checkpointing
- Fast storage tier: **Acadia (Ceph) block storage** in IBM Cloud
- Large cost-effective data repository using **IBM Cloud Object Storage**
  - Two-tier architecture where AFM transparently moves data between the object storage and file system

Training performance improvements:

- Storage Scale improved training step time variation by 5X

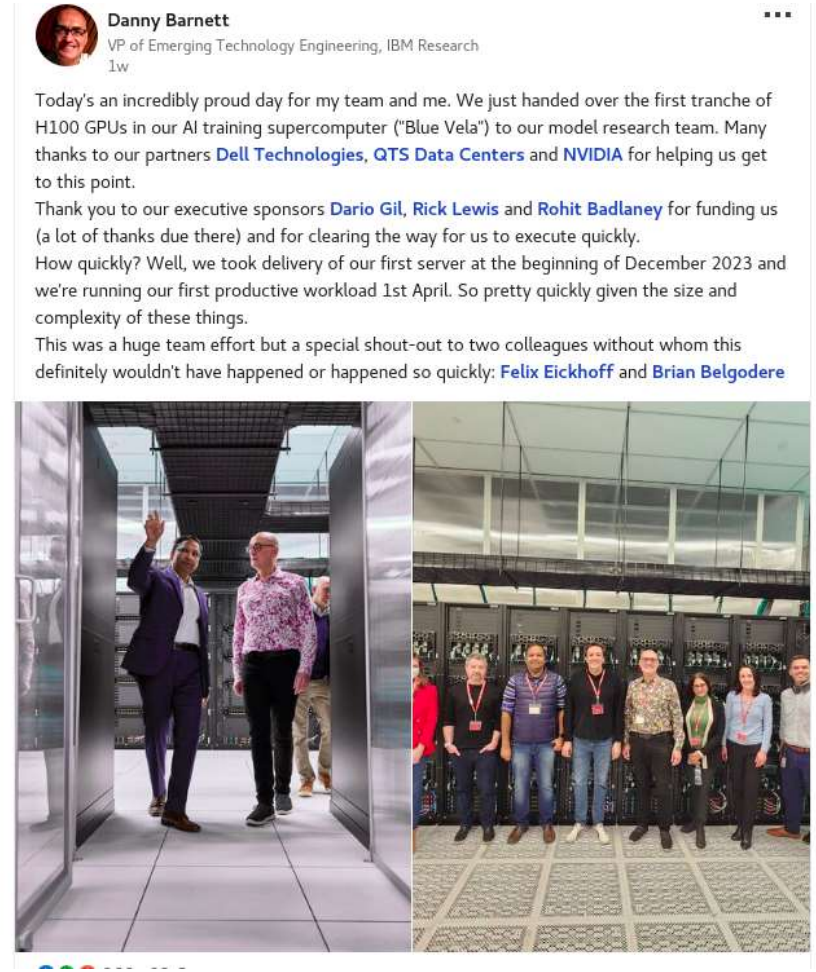
https://research.ibm.com/blog/AI-supercomputer-Vela-GPU-cluster
https://research.ibm.com/blog/vela-ai-supercomputer-updates

IBM Blue Vela- HGX “SuperPOD” Storage Fabric (IBM Cloud/ IBM Research)



- IBM Cloud and Infrastructure
- AI Supercomputer Scalable with many H100 HGX Systems
- 1st Phase 1000s of HGX GPUs

- AI and Data platform to deliver enterprise AI service
- Training LLM models with 100B+ parameters
- Faster results – quality & speed of the training models.



https://www.linkedin.com/posts/dannybarnett_todays-an-incredibly-proud-day-for-my-team-activity-7180654456361910274-4fvU?amp;utm_medium=member_ios

IBM Storage Scale

Active Protection for Cyber Resiliency built on the NIST framework

Storage Scale

IDENTIFY

- Cyber Resiliency Assessment Tool, Probes 100s of different controls and best practices

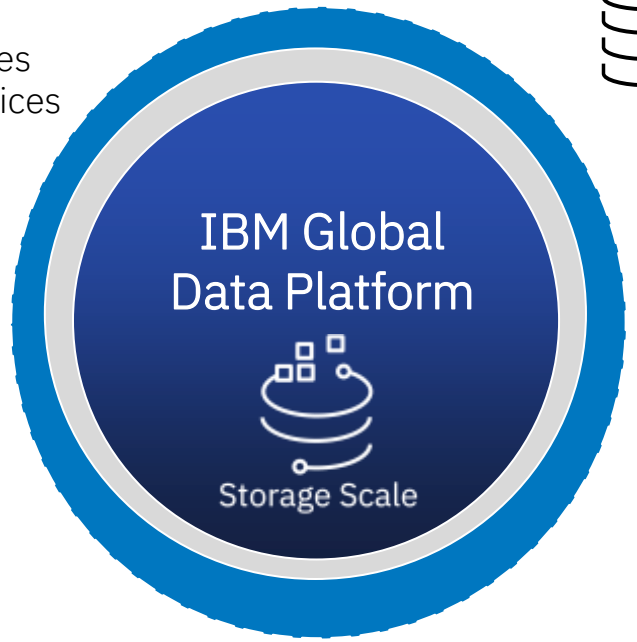
Governance

- Data Catalog allowing for data orchestration and data migration control and accountability
- Watson Knowledge catalog

RECOVER

Recover Operations and Data Quickly

- Instant Restore with Storage Scale AFM
- Storage Scale and Storage Protect – recover multi-petabyte filesystems in hours
- QRadar Incident Forensics



PROTECT

Active Protection against cyber attacks

- Multifactor Auth, RBAC, Privileged Access Monitoring (IBM Security Verify)
- Safeguarded Copies via immutable snapshots, logical air gap
- Scan snapshots for signs of ransomware
- Log all Admin & user actions

DETECT

Detect Suspicious Behavior

- QRadar and Splunk SIEM integration
- File Audit Logging, Watch Folders
- Analyze backup data for signs of ransomware (Spectrum Protect)
- Reporting: QRadar User behavior analytics
- IBM Flash Core Modules entropy detection

RESPOND

Alert and take action

- Automated action upon threat detection (QRadar)
  - Snapshot, Block Session , Etc..
- Alerts automatically prioritized based severity of the threat and criticality of the assets involved

Multiple ways to deploy IBM Storage Scale

IBM Storage Scale software

ARM, x86, IBM Power, IBM System z, Kubernetes or Virtual Machines

IBM Storage Scale System 6000 or 3500

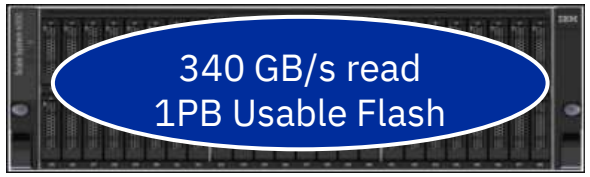
IBM Storage Scale in Public Cloud

AWS Azure IBM Cloud Alibaba Oracle Cloud Google Cloud

The easiest way to deploy for HPC and AI Training

IBM Storage Scale System

All NVMe Flash 6000



- 48 TB to 1+ PB flash
- Up to 15+ PB capacity per 3500
- Scales from 1 to 10000s+ of clients

3500 Hybrid NVMe Flash + HDD

Performance Optimized

340GB/s per 6000

10M+ IOPS per 6000

Parallel access to data

Locally cached global data with dynamic memory pools

Cost Savings

Compression-enabled NVMe QLC storage 2Q24

Integrate existing non-IBM storage and cloud

Turn off/turn on unused storage w/ tape

Mix and match old/new

Cyber-Secure

End-to-end encryption with customer keys

Lower RTO with Safeguarded Copy quick recover option

Ensure protection with CyberVault service

Global Connectivity

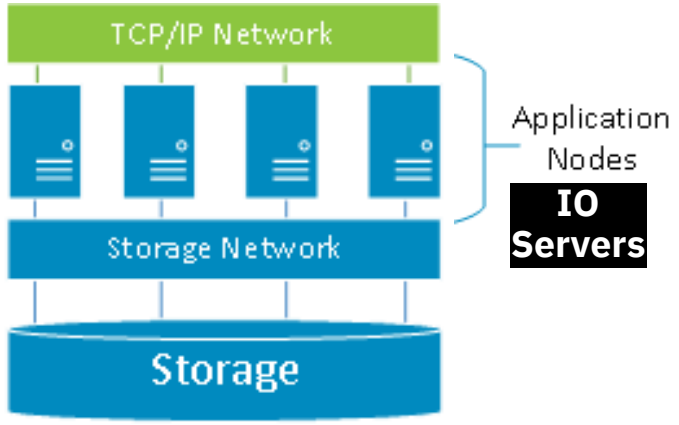
Break down silos by connecting remote systems, cloud data and non-IBM storage

Connect any node to a global data platform online when needed

1 Up to 2PB+ can be obtained with 3:1 compression using FCM disks

Deployment models to run the data with the compute

Enterprise Integrated Model (SAN, NVMeoF, iSCSI)



Block based

SAN based filesystem

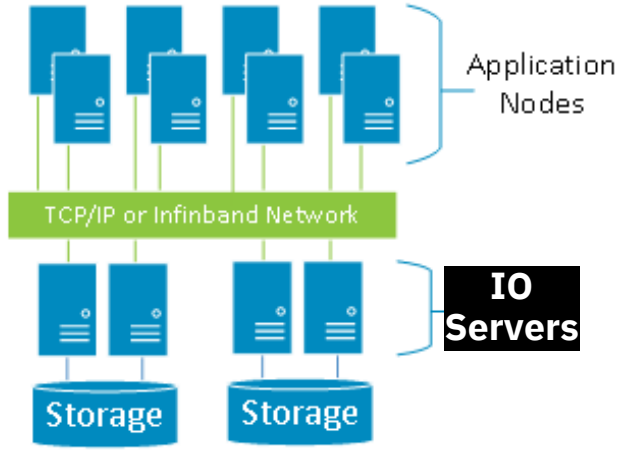
PureScale

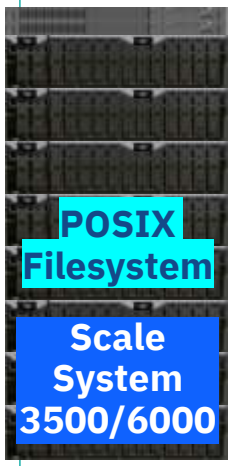
Sailfish

DS8k SDS Scale

Unify and parallelize storage silos

Twin tailed storage with erasure coding





POSIX Filesystem

Scale System 3500/6000

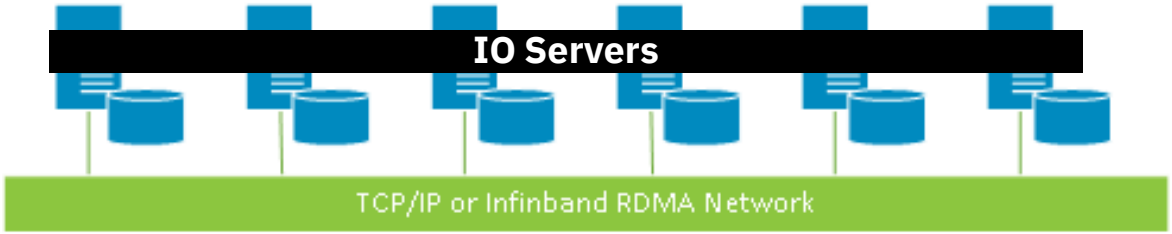
Modular High-Performance Scaling

Shared Nothing Cluster (SNC) Model (Storage Rich Servers (replication, erasure code))

Scale (FPO)

Erasure Code Edition (ECE)

Storage Fusion HCI or SDS



Hadoop

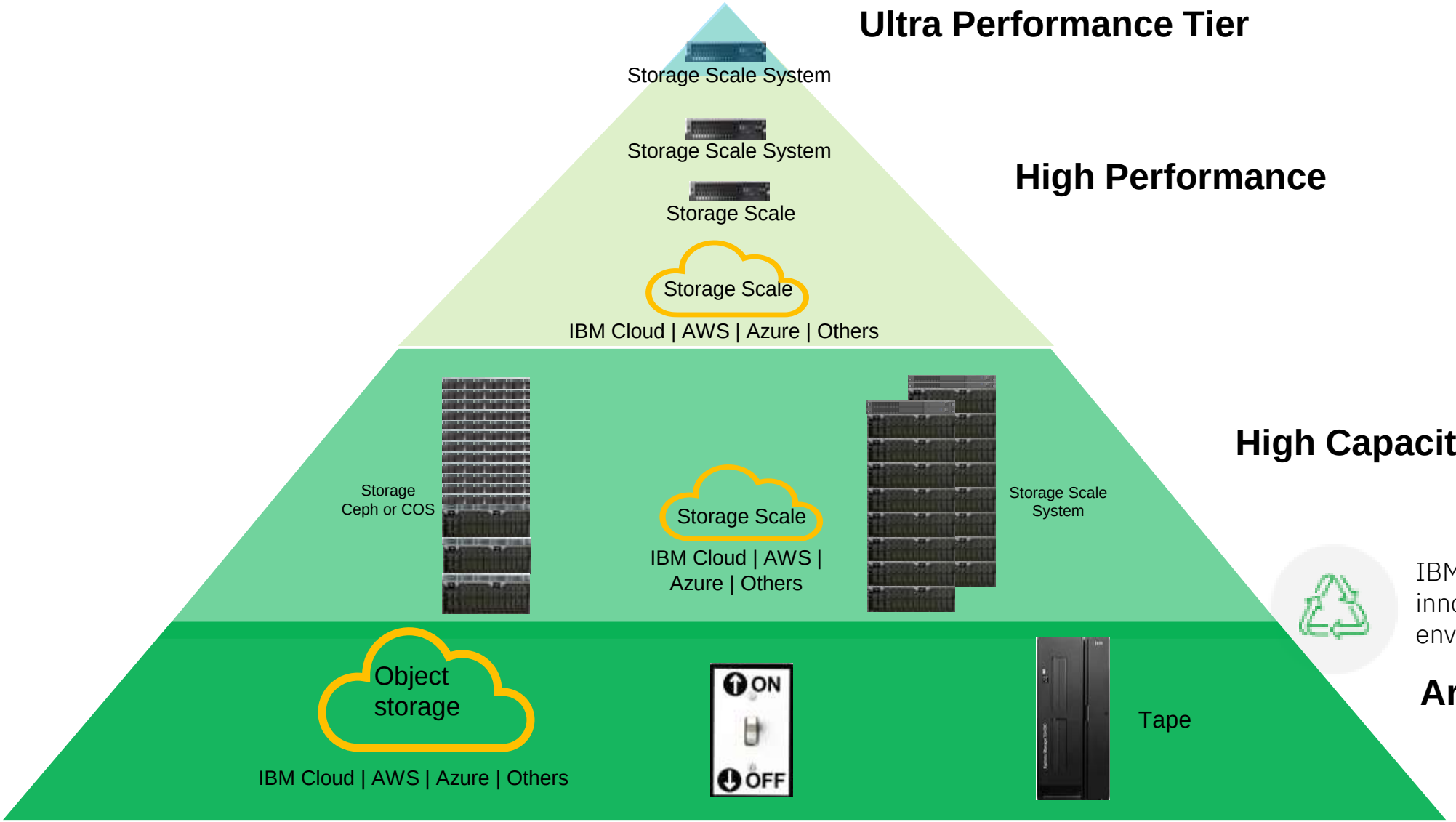
Object Storage

SDS POSIX

Span storage rich servers for converged architecture or HDFS deployment

Tiering models to move the data to and from the compute

Lowers Energy Consumption and Costs with Data Lifecycle Management



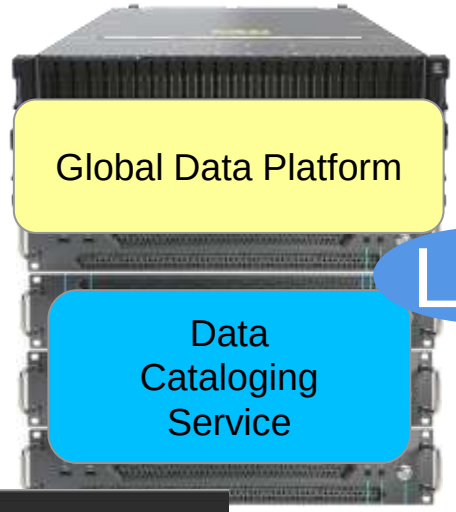
Data Orchestration to the compute

Data Orchestration providing a future scalable architecture with Tape

50TB

TS4500 or Diamondback Racks

Access Services

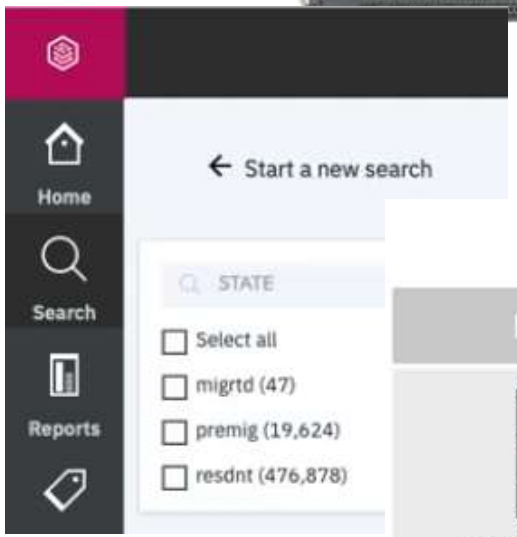


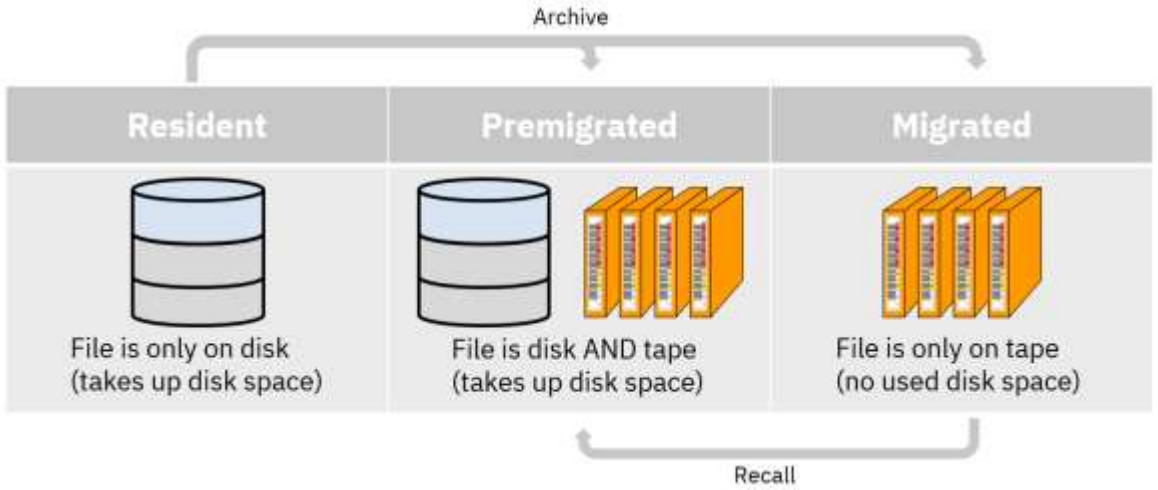
LSF

Architecture to ingest and recall data to tape libraries

Data Management Services

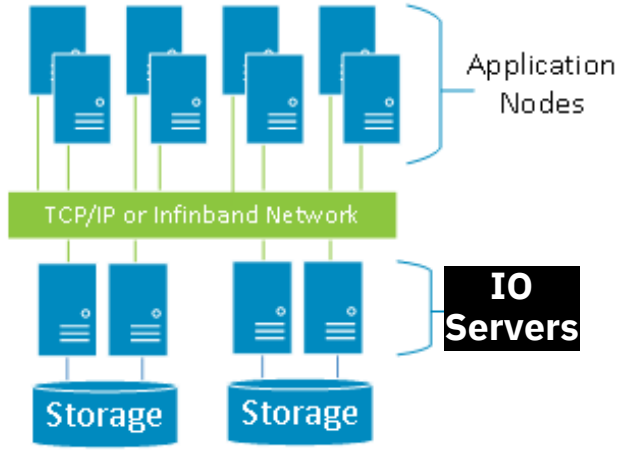
Orchestration

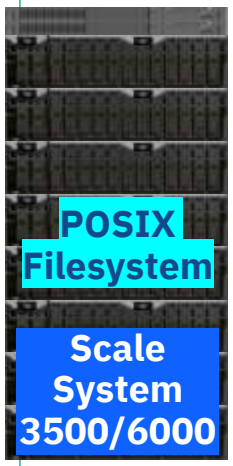




Run the data closer to the compute with local storage

Twin tailed storage with erasure coding

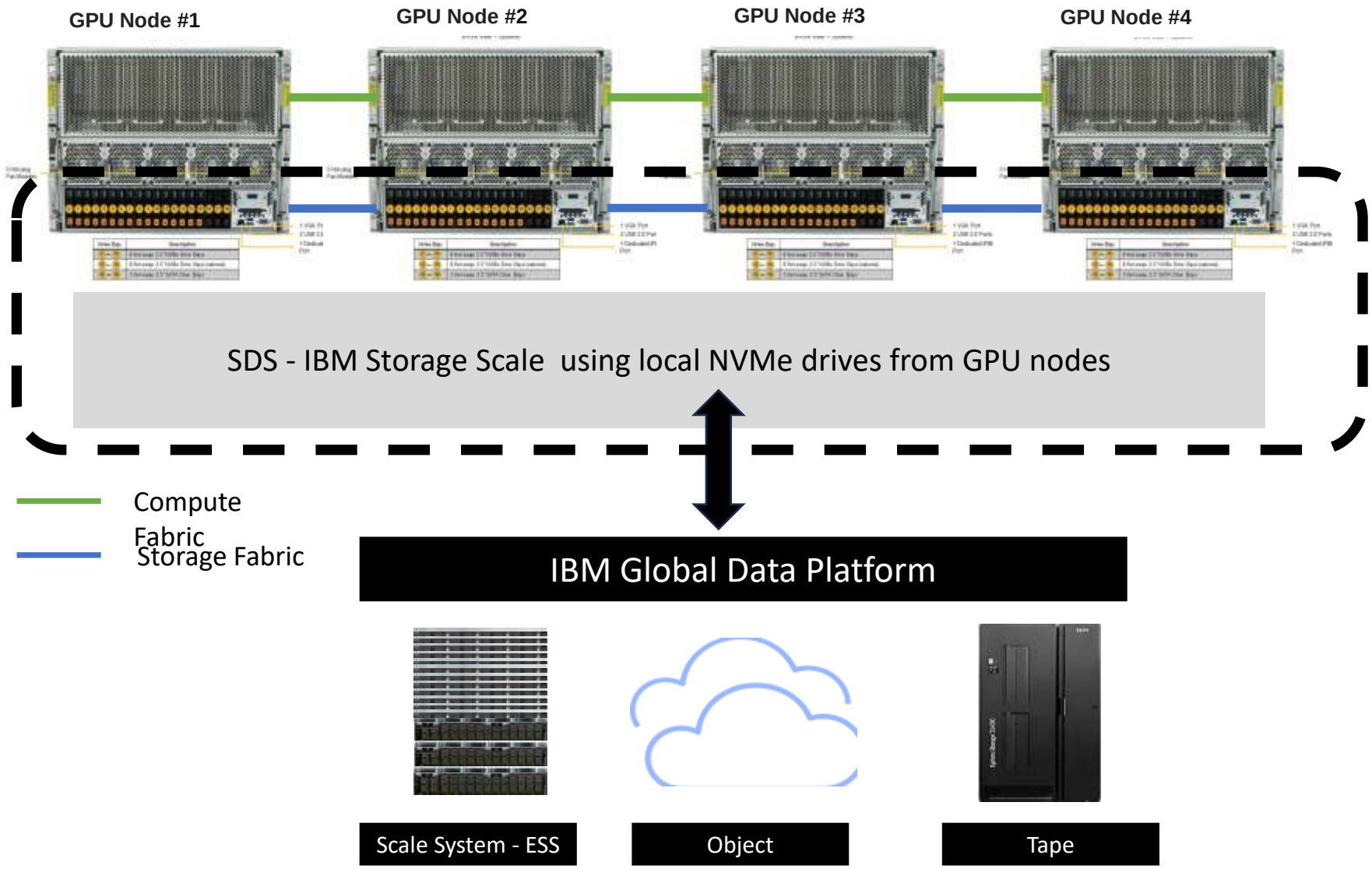




Modular High-Performance Scaling

Shared Nothing Cluster (SNC) Model (Storage Rich Servers (replication, erasure code))

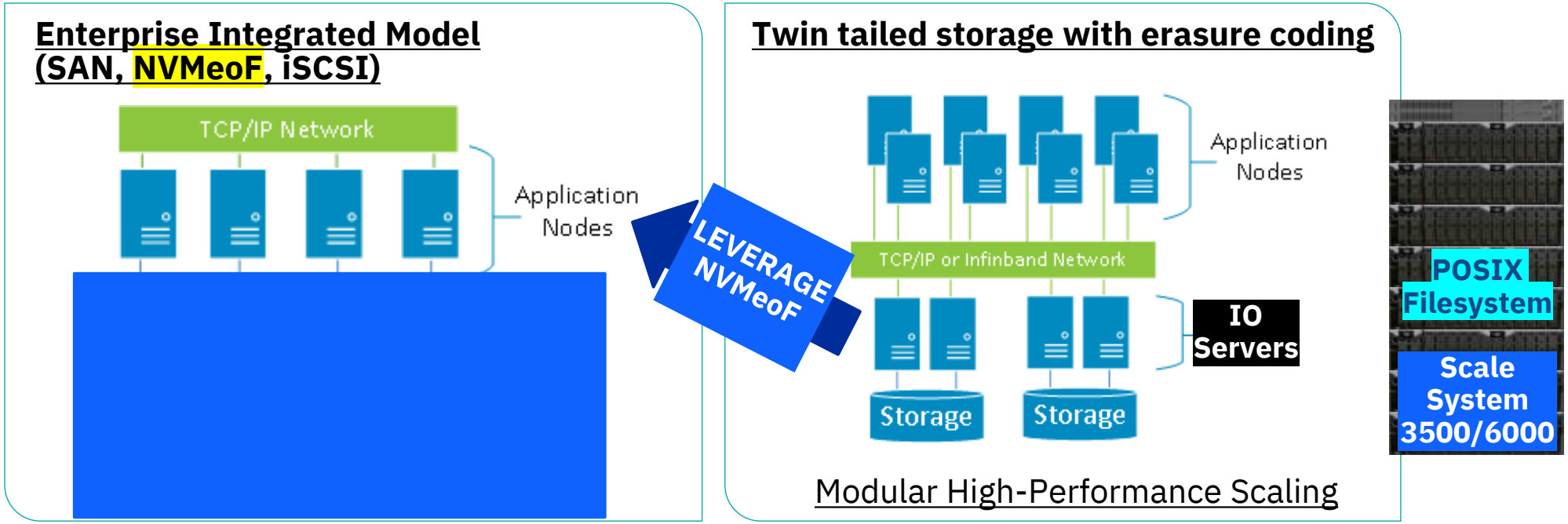
A new approach to Data at the compute – History - Converged Solution Architecture



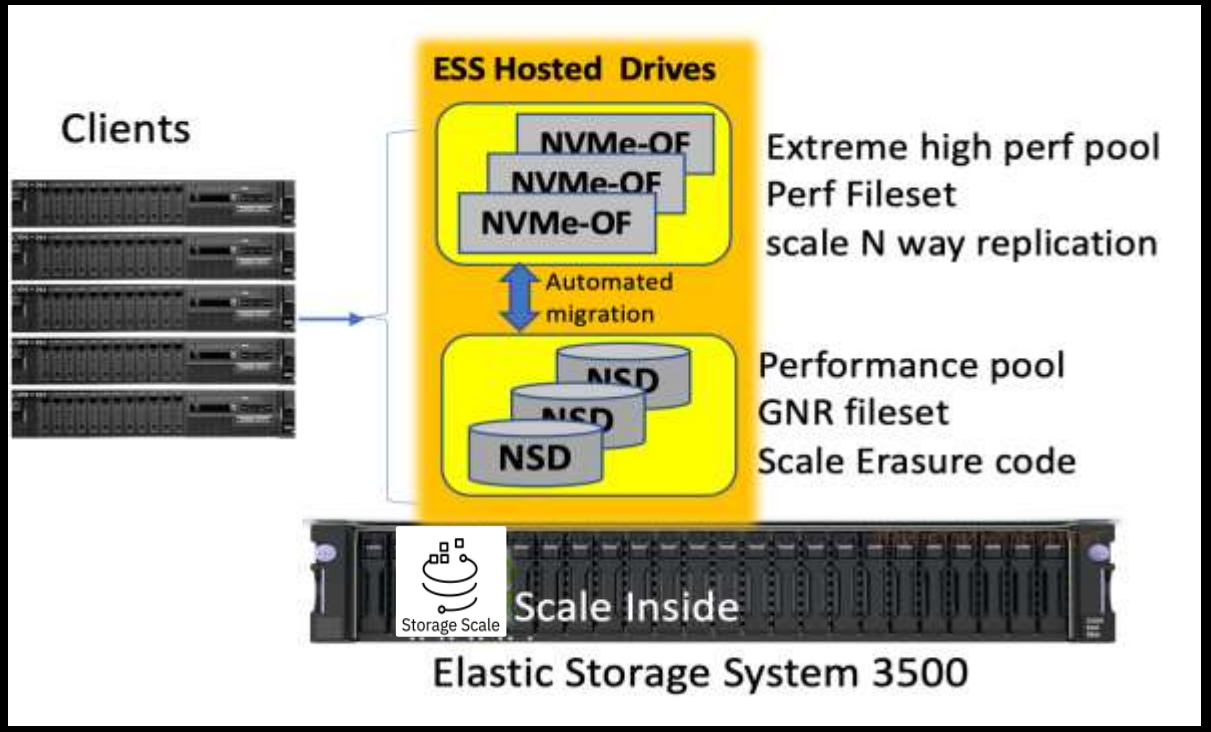
IBM Storage Scale configuration

- Converged GPU and Storage solution for AI training workloads
- High Performance parallel storage using NVMe local drives from GPU nodes
  - Minimum 2 drives per node; 8 drives across 4 node cluster
  - Max 16 drives per node; 64 drivers across 4 node cluster
- 2 x 200 Gbps storage network per node

Run the data closer to the compute emulating local storage with NVMeoF



At SC22 demonstrated: Integrated NVMeoF for an extreme performance tier to the compute



Measured over 10+ Million IOPs and 100s GB/s

Use Case

Data analytics (AI/ML) needing very high rand IOPS with high throughput

High performance Scratch / Shuffle space

System Config

3.84 TB, 7.68 TB, 15.36 TB or 30.74 TB

4x CX6-VPI Adapters / canister

Performance and Features

- **Integrated extreme high IOPs storage Pool**
- **Dedicated performance pool (12x drives)**
- **Easy configuration and setup**
- **Automatic data migration between pools**
- **Integrated RAS support**

Introducing HyperStore! (uStore codename) Caching and Acceleration to the compute

Accelerate AI and Analytics by storing the data as close to the compute as possible

Leverage both shared storage (e.g. NVMeoF) and storage inside of the compute node

Support **Asymmetric Replication** with one erasure-encoded copy of the data and one performance copy for high-speed access

Creates a **Shared Co-operative Cache** across all compute nodes. Any node can access all cached data, regardless of physical location.

The first release, writes update all copies. In a follow-on release, allow writes to performance copy only with **Eventual Reliability** (e.g. Burst Buffer)