

Breaking data silos to support mixed workloads



Kevin Tubbs, Ph.D

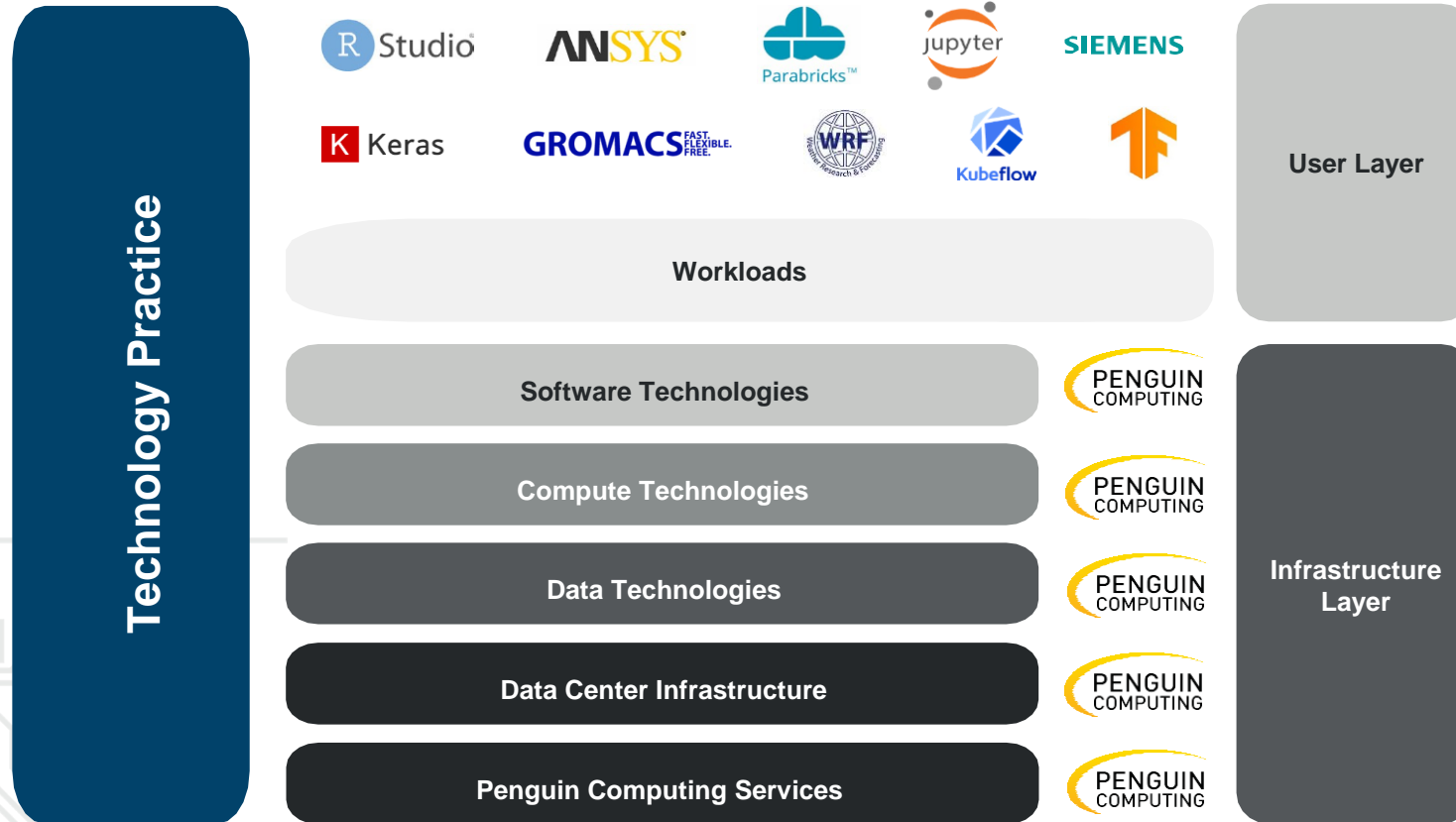
SVP, Strategic Solutions Group

Penguin Computing

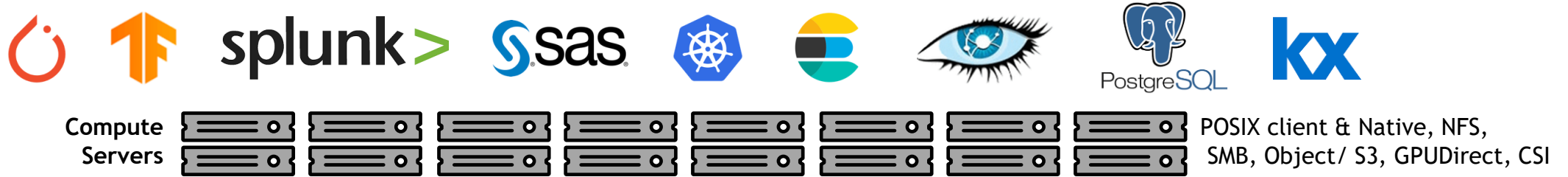
Presented at

Hyperion User Forum April 29th 2021

Architectural Model



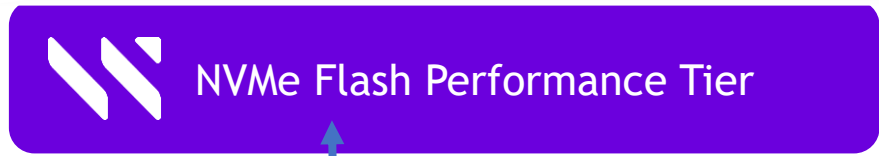
THE LIMITLESS DATA PLATFORM



Ethernet and/or Infiniband

UNIFIED NAMESPACE

WekaFS on Industry Standard Servers



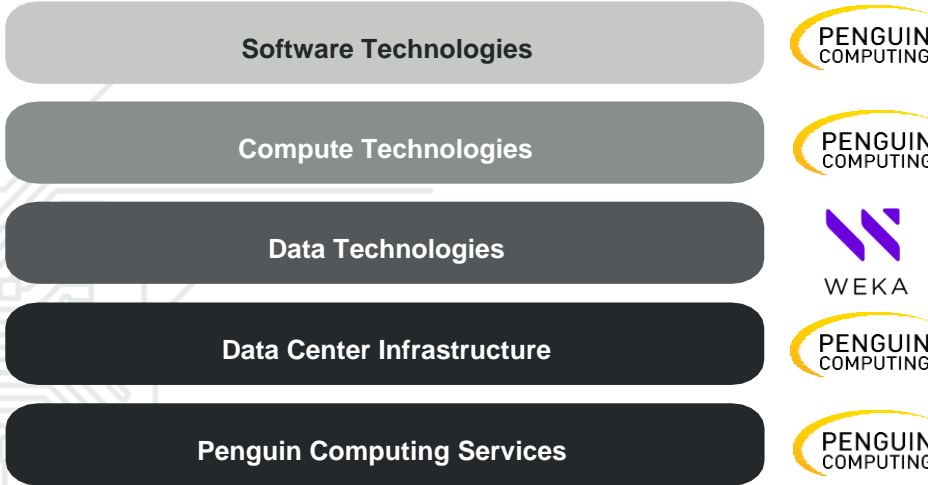
ActiveData - WekaFS

Penguin Computing ActiveData Components

Data Practice



HPC Workloads Analytics/ML Workloads AI/DL Workloads



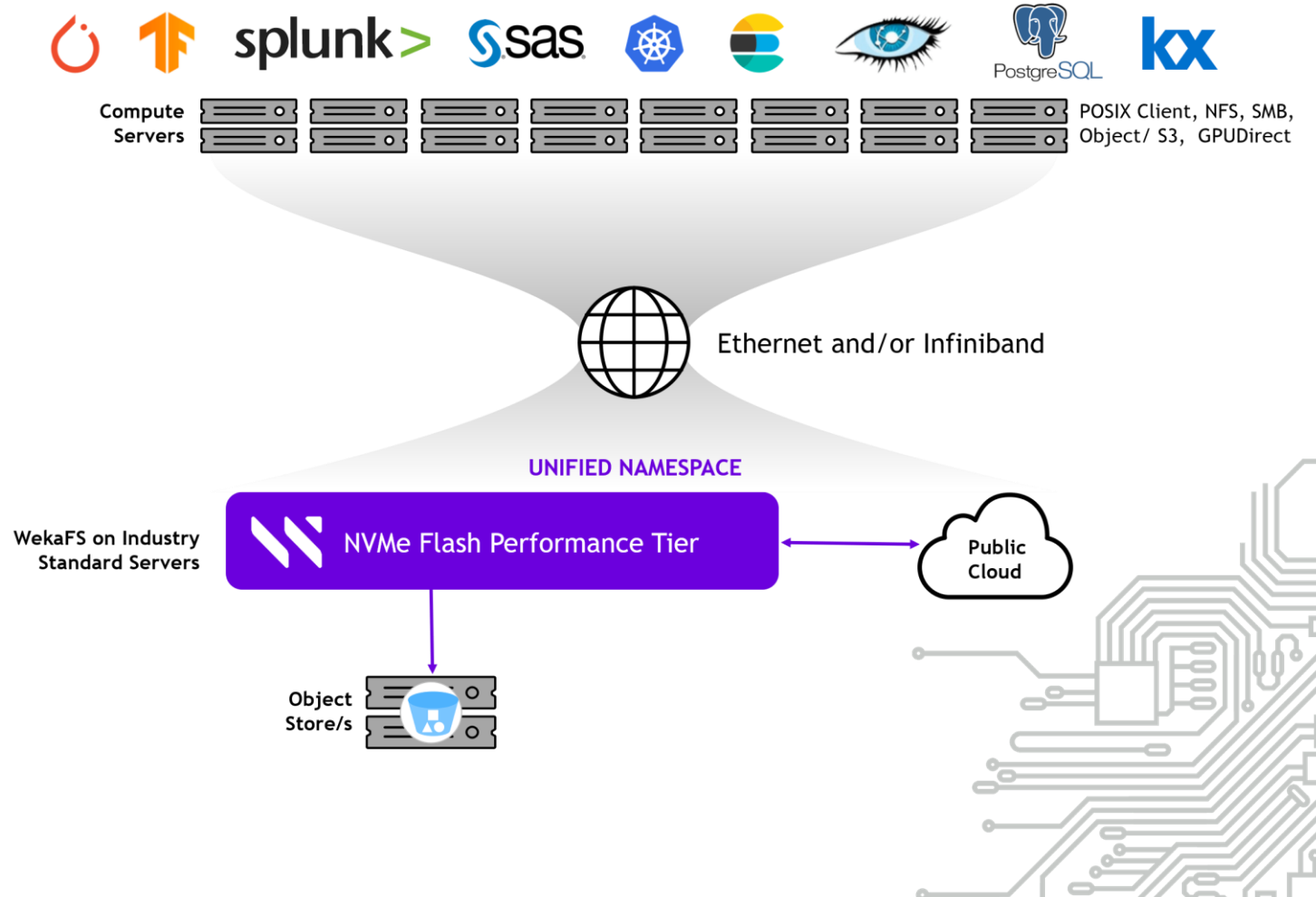
User Layer

Infrastructure Layer



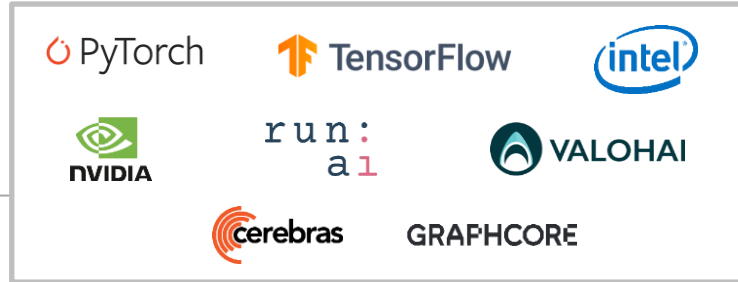
IDEAL PLATFORM FOR CONSOLIDATION

- Multi-protocol ready
 - SMB
 - NFS
 - S3/Object
 - GPUDirect
 - CSI
- Diverse workloads
 - File and Object
 - Small and Large file workloads
 - IOPS/Throughput/Latency/Metadata
- Built In User & Data Management
- Active Directory integration
- Authenticated hosts (PKI)



WORKS WITH YOUR APPLICATIONS

ARTIFICIAL INTELLIGENCE



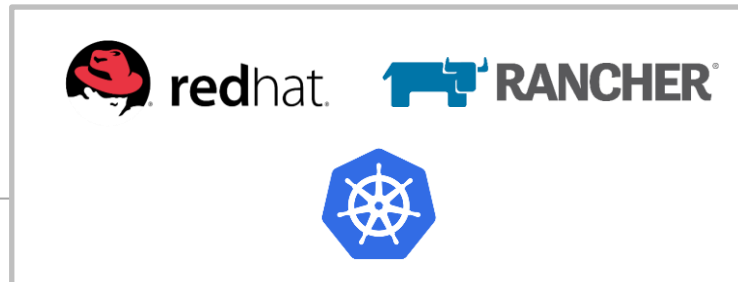
DATA ANALYTICS



LIFE SCIENCES



CONTAINERS & ORCHESTRATION



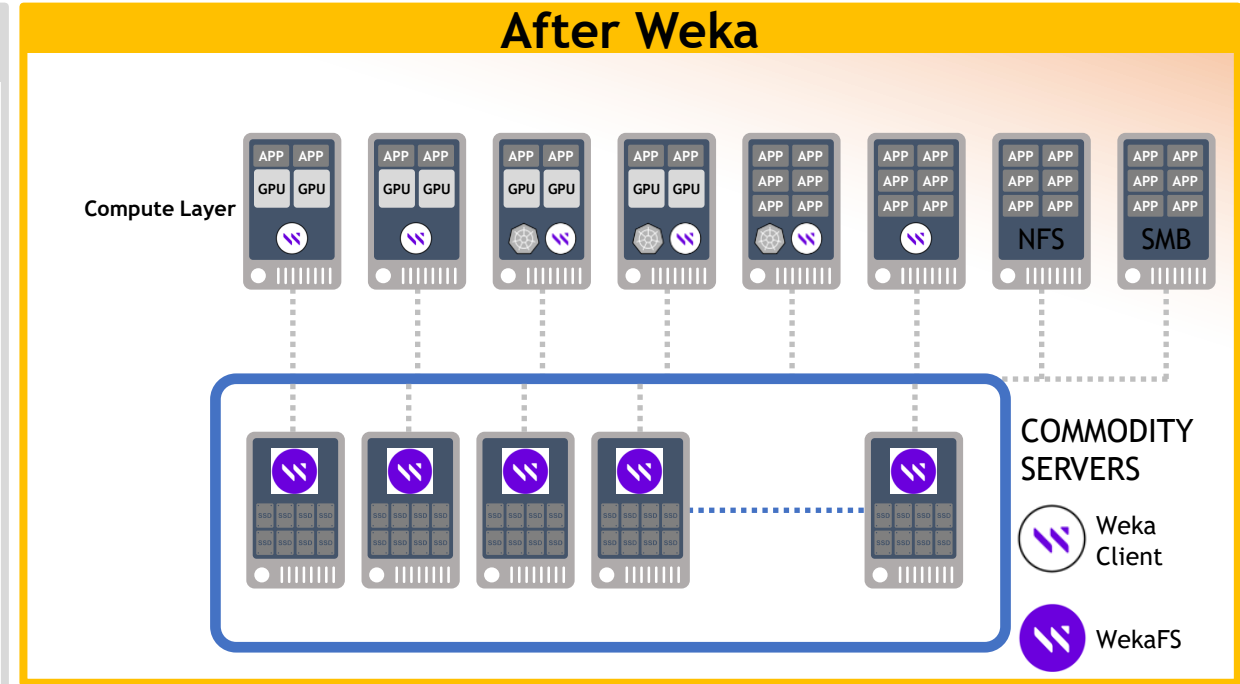
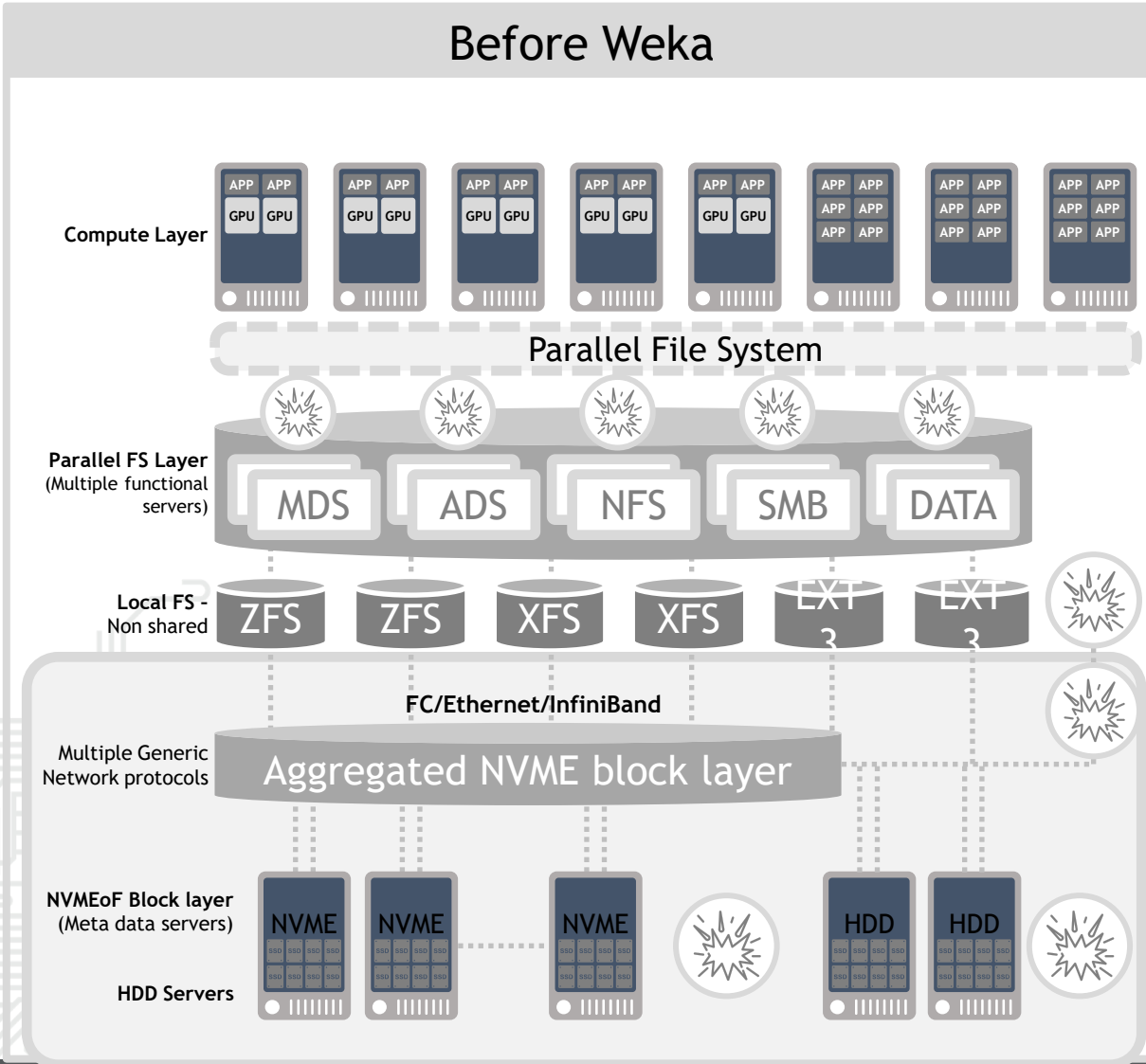
OBJECT STORAGE



DATABASES



BREAKING DOWN SILOS



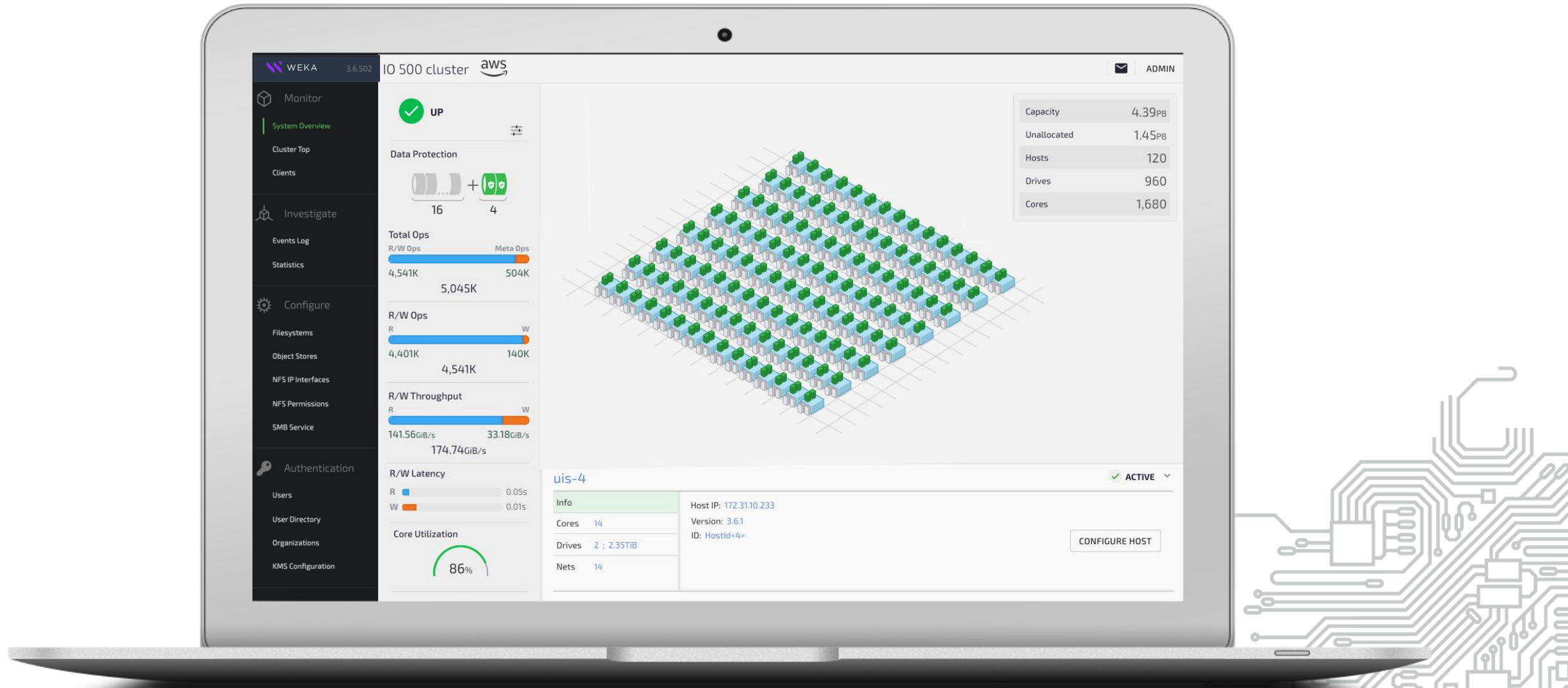
Before

- Complex to scale
- Requires high levels of expertise to optimize per specific workload
- Several layers that must be optimized and configured independently
- Monolithic MD structures

After

- ✓ Linear scalability on all dimensions: Throughput, IOPS, Metadata maintaining low latency
- ✓ Dynamically adapts itself for varying workload
- ✓ Simple management and maintenance
- ✓ Fully distributed MD across all the cluster nodes
- ✓ Multi-protocol access
- ✓ Superior performance

INTUITIVE USER INTERFACE

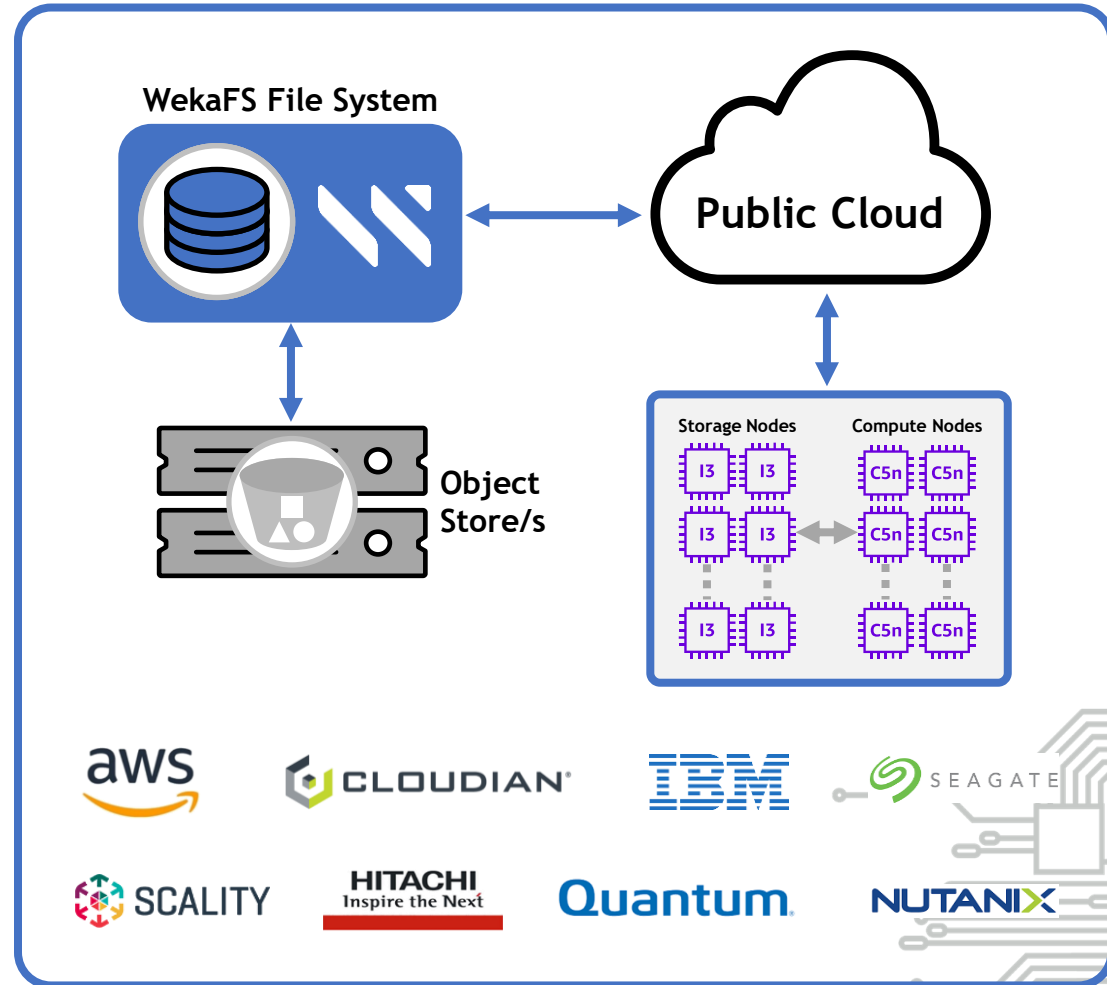


SIMPLIFIED DATA PROTECTION

UNIFIED NAMESPACE



- Transparent backup, archive, data migration
- Snap-to-object allows burst to cloud
- Disaster recovery and business continuity baked into solution
- Tiering to more than one object storage solution increases durability



ENTERPRISE-GRADE SECURITY

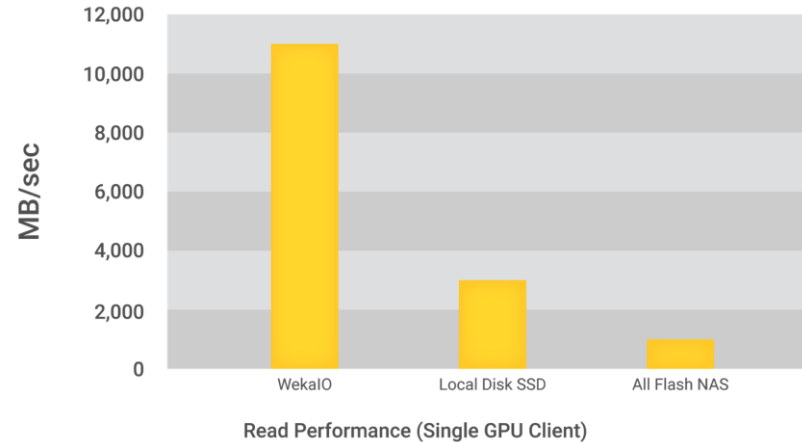
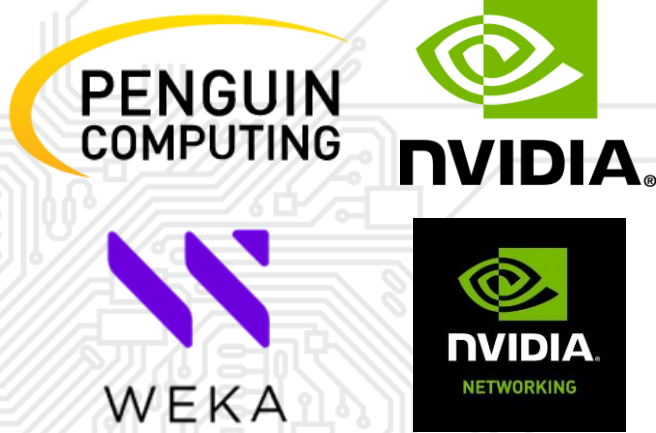
- Encryption for data at rest and in flight
- Negligible performance impact
- Encryption strength
 - 512-bit XTS-AES, AES-256-GCM
- KMS integration
 - Hashicorp Vault , KMIPS (v3.7x)
- ACL synchronization across Windows and Linux
- Multi-tenancy
 - Authenticated host mounts
 - Temporal and mount rules controlled internally
 - Org. level provisioning of security controls including admin access



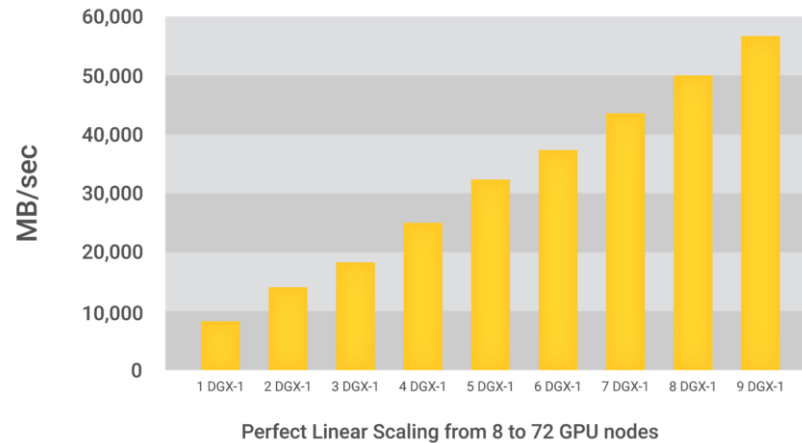
ActiveData - WekaFS Performance

162GB/Sec
& 45M IOPs
Performance to a
single GPU client

MASSIVE SINGLE CLIENT
PERFORMANCE

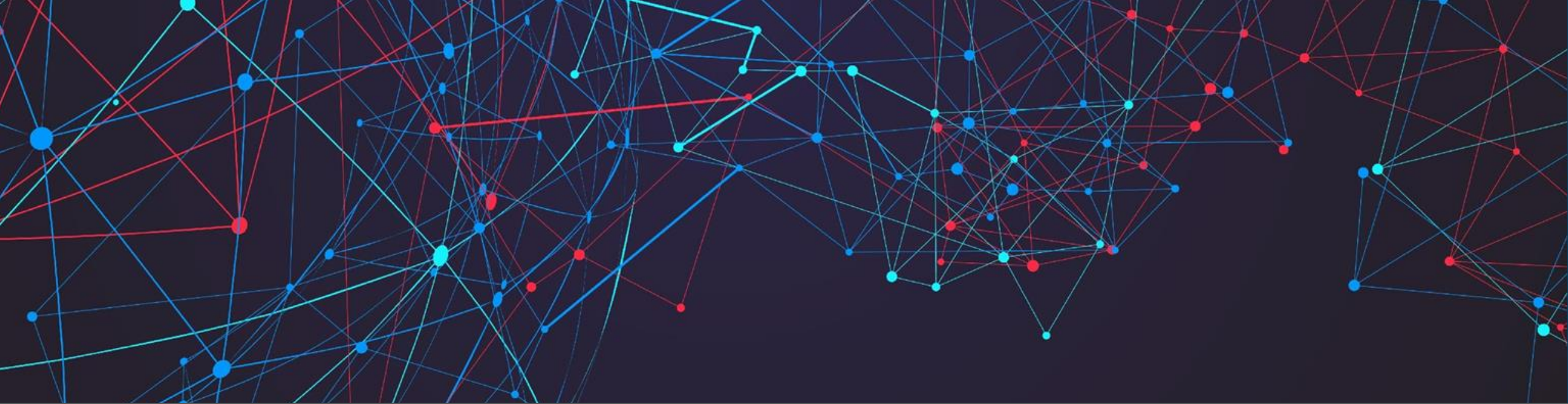


HIGH-SPEED NETWORK SATURATION



PERFORMANCE SCALES LINEARLY TO MULTIPLE CLIENTS





Solution Case Study

The Challenge

- **Traditional heritage as a Penguin HPC User**
 - Has growing AI expertise developed and delivered in cloud
 - Demanding need to expand and scale corporate AI capability
 - Focus AI hires on data scientist and engineers or AI Infrastructure IT talent?
- **Transition to Unified Enterprise HPC and AI Strategy**
 - AI projects and customers driving enhanced and accelerated capabilities
 - Need to devise a strategy abstracts platform complexity – Cloud, On-Premises, or both?
- **Design, Validate, Procure, and Integrate Emerging Technologies for Organization**
 - Accelerate bringing AI capabilities to their business

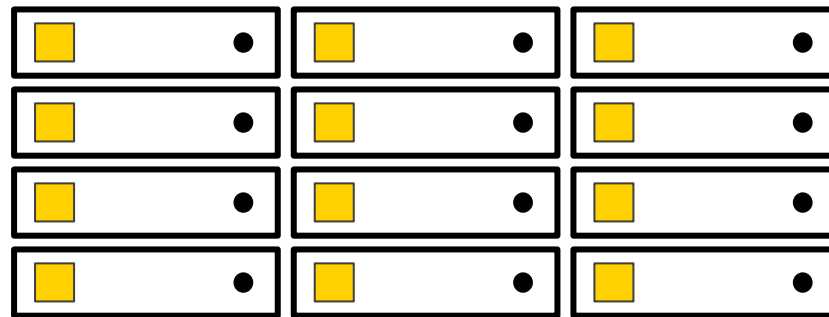
Solution Goals

- **Provide Capability to Support AI/ML and HPC Workloads**
 - Growth enabled in all directions – HPC, AI, Data, Networking
- **Improve User Experience**
 - Personalized environment
 - Sessions tied to persistent VM rather than compute servers – No timing out
 - Single remote access interface for all applications – Down from 5
 - Enable a single user experience that supports HPC and AI
- **Improve Security**
 - Using current employee identities
 - Balance technical computing best practices with corporate security policies
 - Enable Data-At-Rest Storage Encryption

HPC Architecture



TrueHPC
(Turnkey Enterprise HPC)



Add AI Architecture

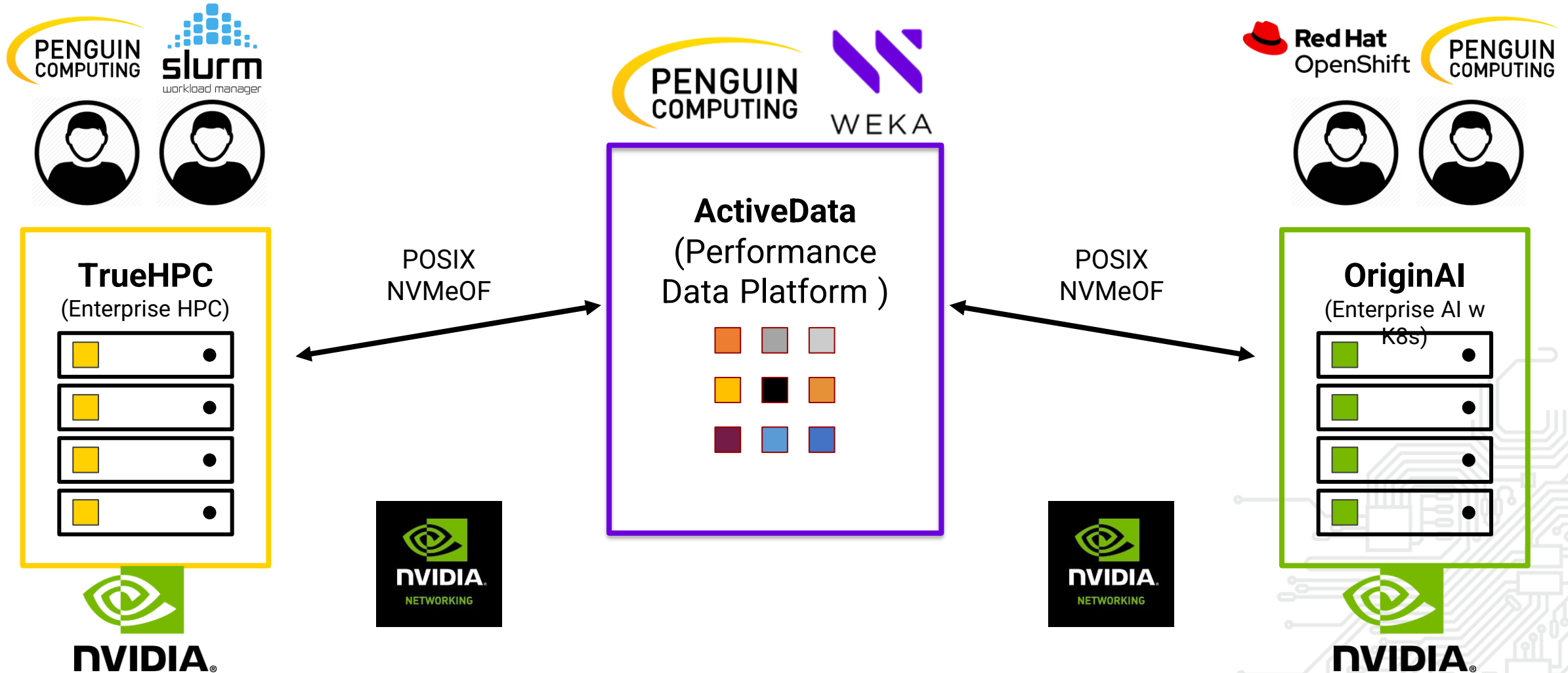


TrueHPC
(Turnkey Enterprise HPC)

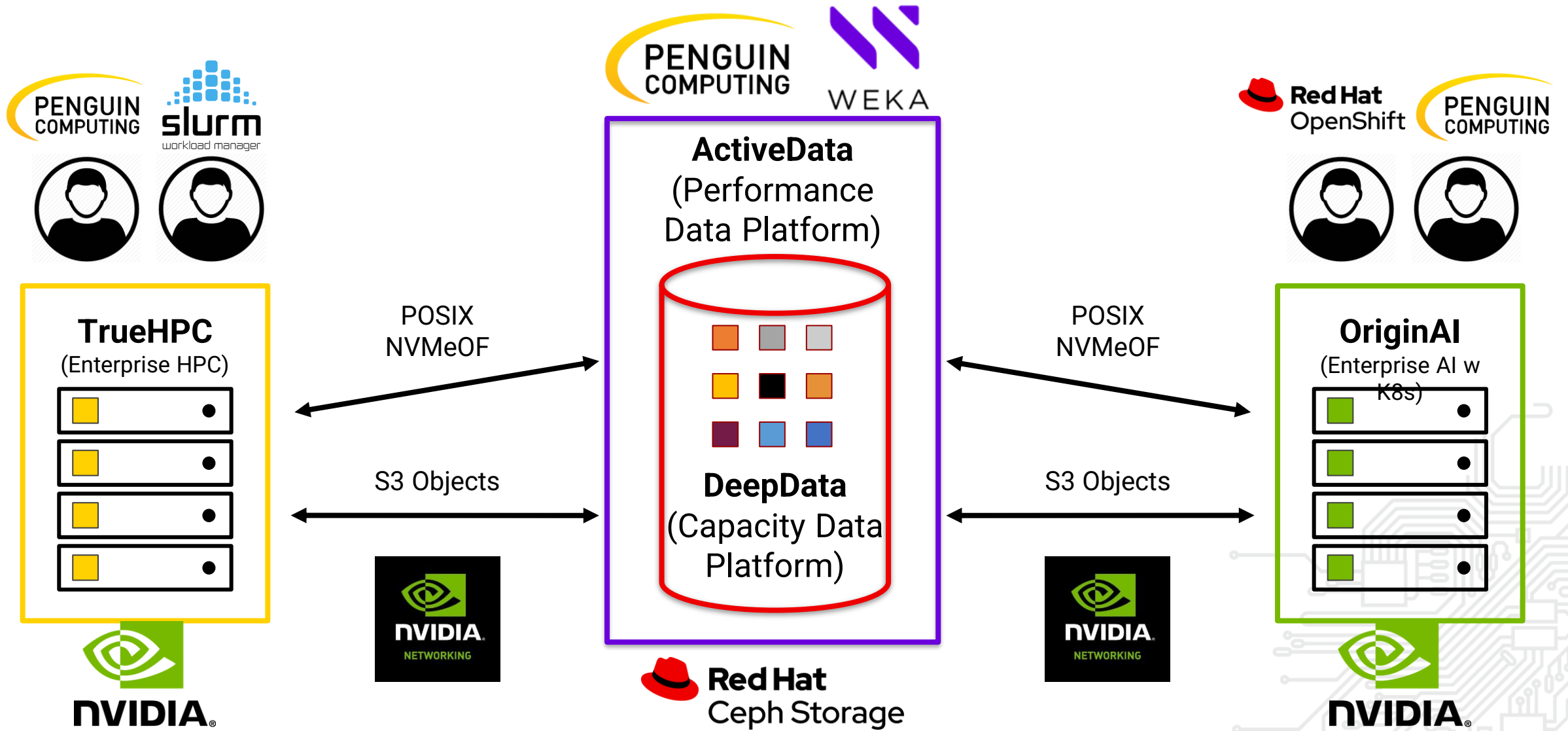


OriginAI
(Turnkey Enterprise AI w/ K8s)

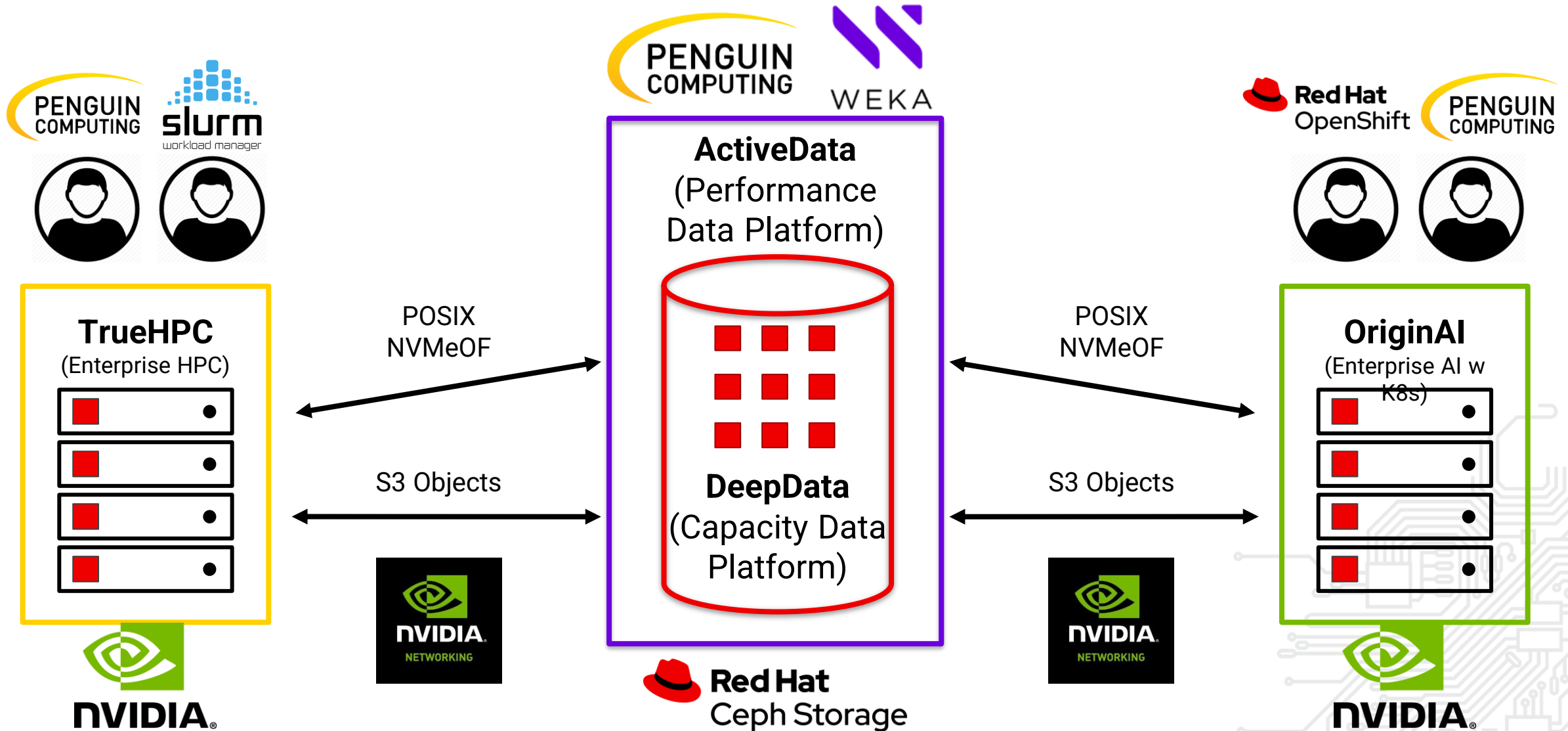
Add High Performance Data Store



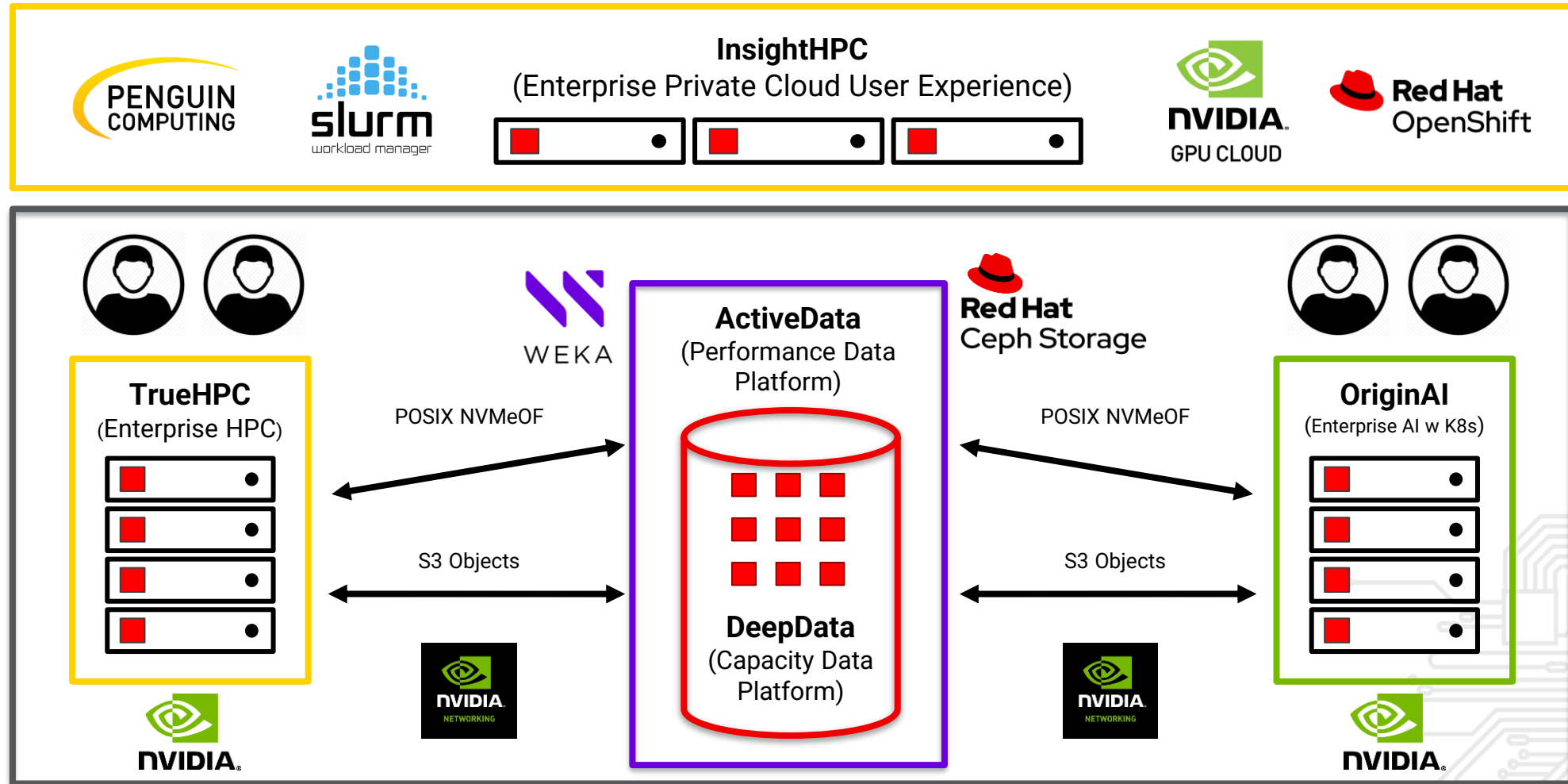
Add High Capacity Data Store



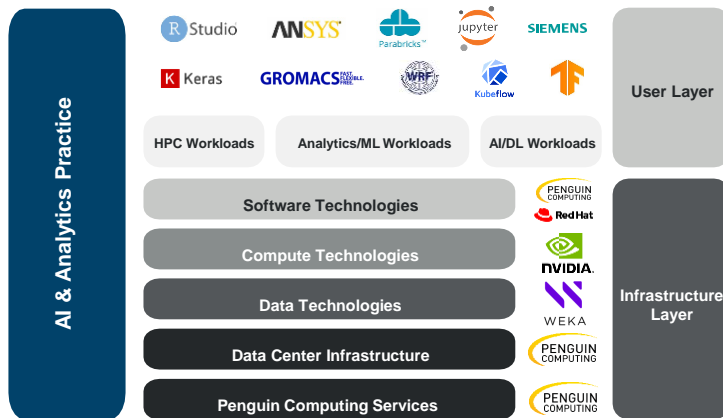
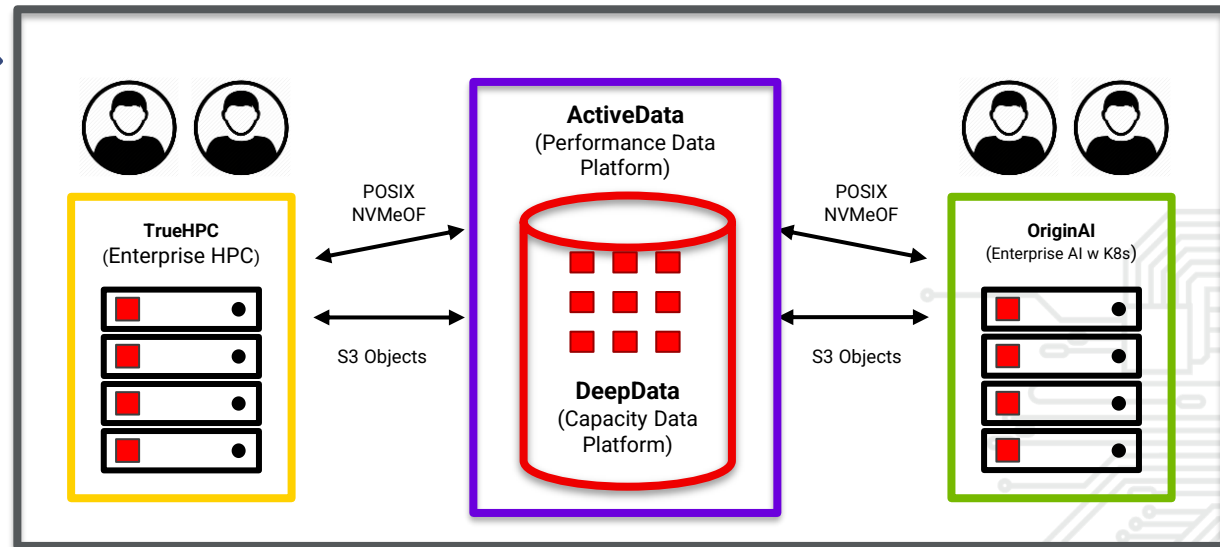
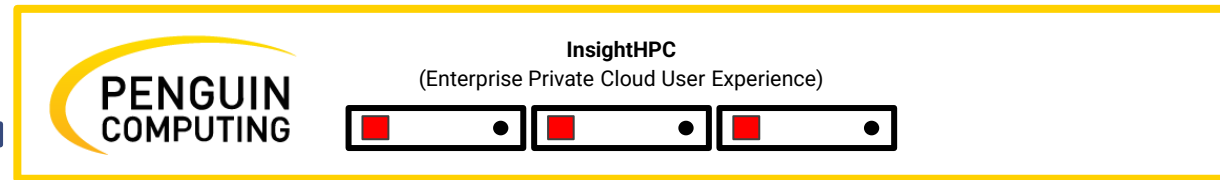
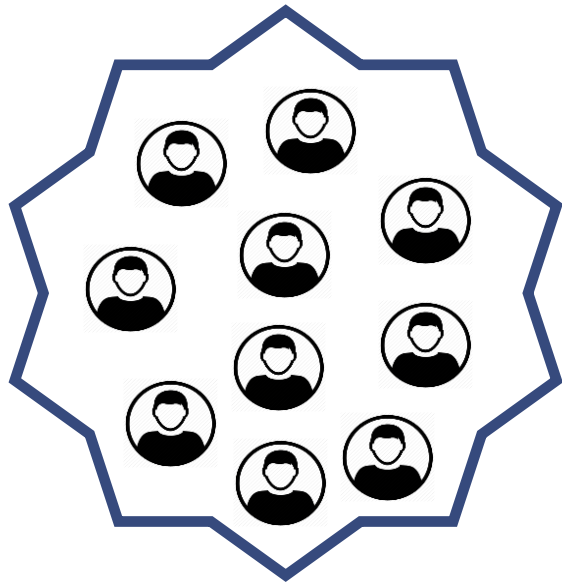
Security - Add Data-At-Rest Encryption



Diverse Community - Add Cloud User Experience



Security - Corporate Network Integration



Lessons Learned

- **Improved Productivity**
 - Reduced latency in total system: data, network and compute
- **Improved Resource Utilization**
 - Increased access via improved tools
 - Move to parallel and batch submissions and add K8s resource manager and scaling
- **Limitless Data Platform**
 - Ceph object S3 compliance
 - Weka native storage tiering
 - Weka bi-directional Object to File translation
- **Improved Workload Portability**
 - Cloud-focused technologies prepare users for cloud strategies



Questions?

OriginAI with OpenShift for
AI/ML Workloads