SNIA DEVELOPER CONFERENCE
SDC 25
By Developers FOR Developers

Hyatt Regency Santa Clara, CA
September 15-17, 2025

Activating Tier 0 Storage Within GPU- and CPU-based Compute Clusters

Floyd Christofferson
Hammerspace

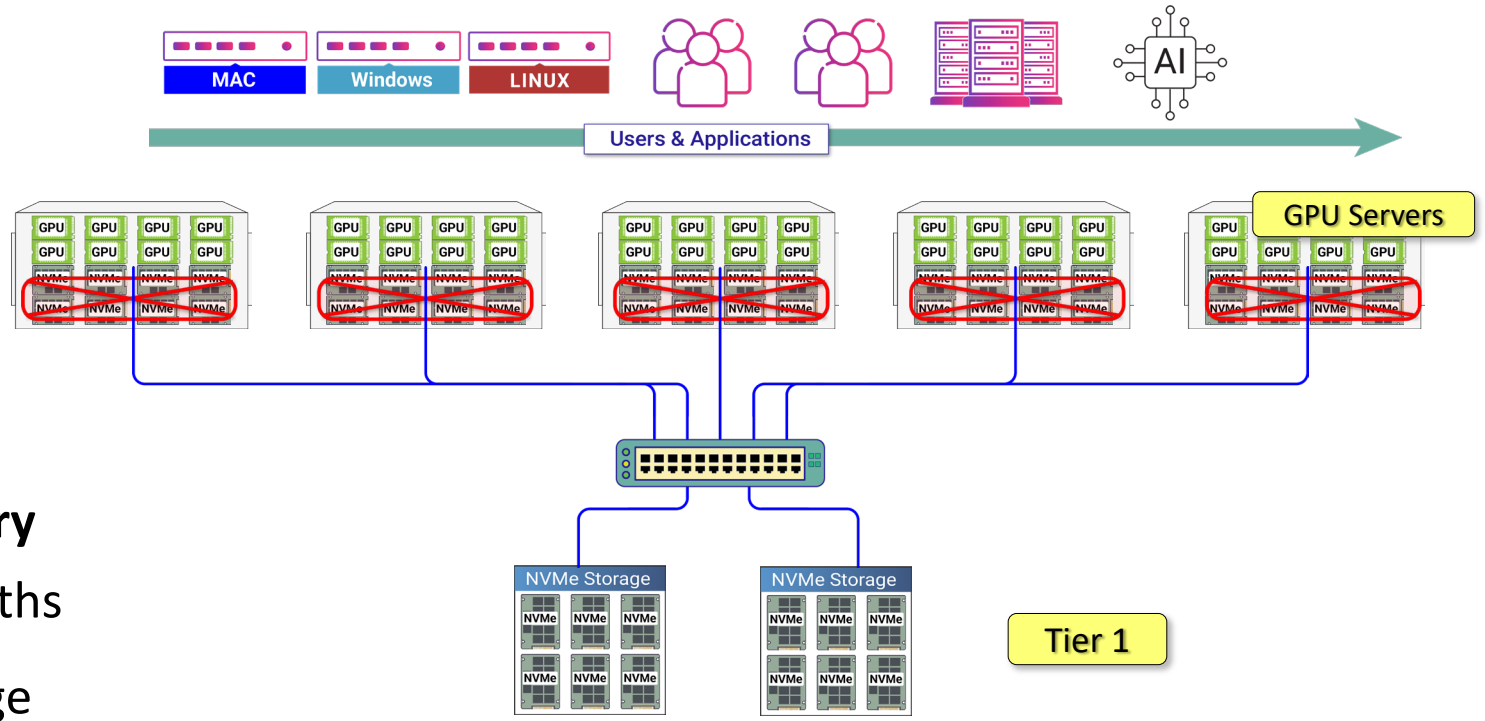
www.sniadeveloper.org

The Problem of Enough Storage Performance for GPUs

How To Activate Underutilized Local NVMe on GPU Servers

GPU Server Local NVMe Storage Underutilized

- Siloed, and largely unprotected
- Not shared across the cluster
- As NVMe densities grow, this equals large volume of stranded, under-used storage capacity

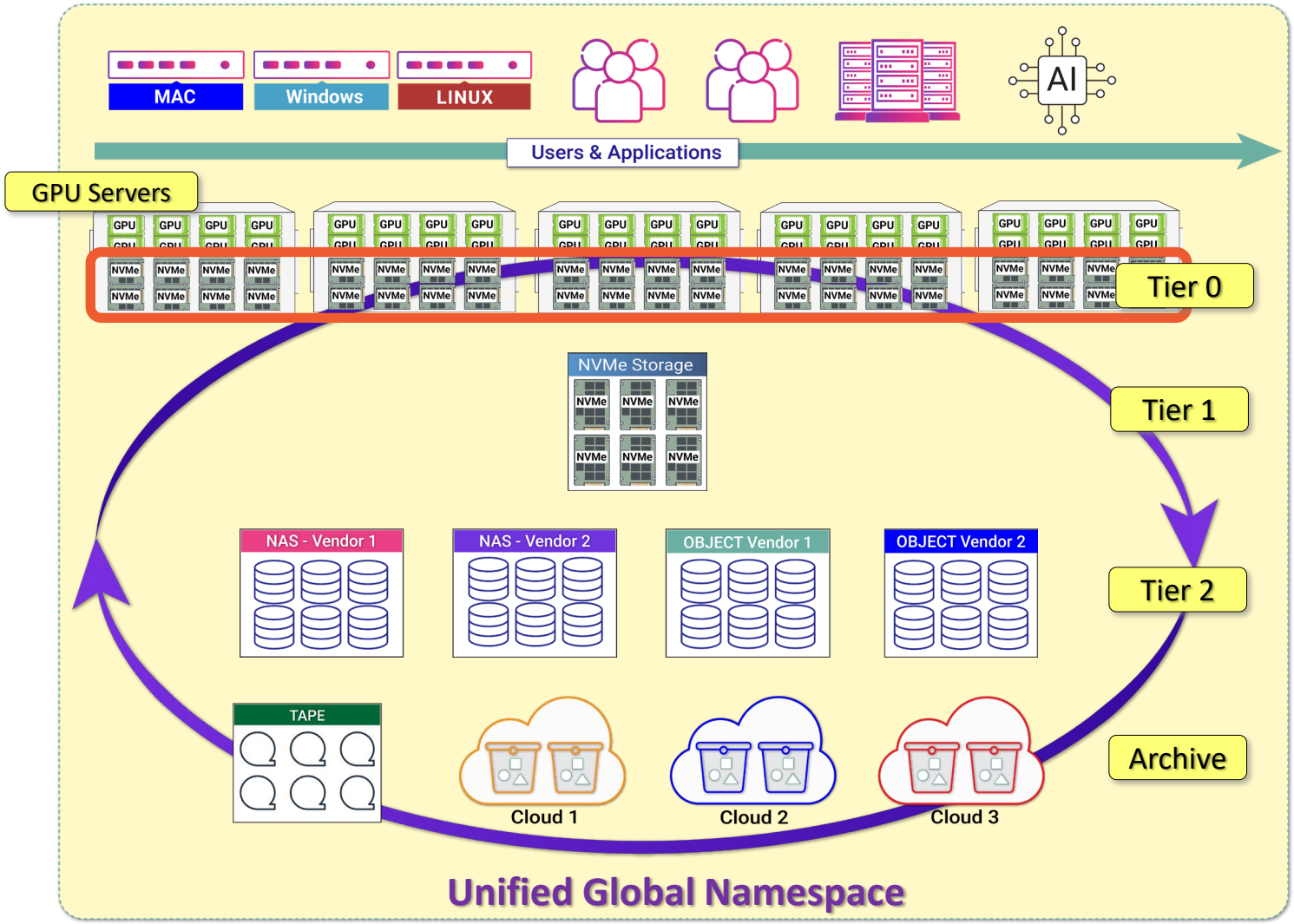


AI Workloads Surpassing GPU Server Memory

- More complex models, longer context lengths
- Forcing I/O over network to external storage
- Network bottleneck adds latency

Activate Tier 0 Within Existing Clusters

- Unifies local Server Storage with all other storage tiers
- Ensures all data is protected
- Reduces load on expensive Tier 1
- Eliminates vendor silos
- Automates AI workloads and data placement into & out of any other storage
- Spans multi-site & cloud



Comparing Tier 0 via Different Network Choices

On premises:

- Tier 0 is faster than any networked external storage
- Reduces need for expensive networks
- Reduces dependency on expensive Tier 1

1. Tier 0 vs. Tier 1 - On Premises (90% Network Efficiency)

(Assuming Checkpoint Size of 500GB)

	Storage Type	Optimal Throughput	Effective Throughput	Time to Write 500GB	
Tier 0	Local NVMe (8-drive GPU server)	112 GB/s	112 GB/s ✅	~4.5 sec ✅	
Tier 1	100 Gb/s Network	12.5 GB/s	11.25 GB/s	~44 sec	10x Slower
	200 Gb/s Network	25 GB/s	22.5 GB/s	~22 sec	5x Slower
	400 Gb/s Network	50 GB/s	45 GB/s	~11.1 sec	2.4x Slower

Note: *External Tier 1 storage throughput adjusted assuming 90% practical network efficiency.*

Cloud

- Tier 0 is dramatically faster
- Increases GPU utilization
- Reduces downtime
- Reduces Tier 1 costs

2. Tier 0 vs. Tier 1 - Cloud Storage (90% Network Efficiency)

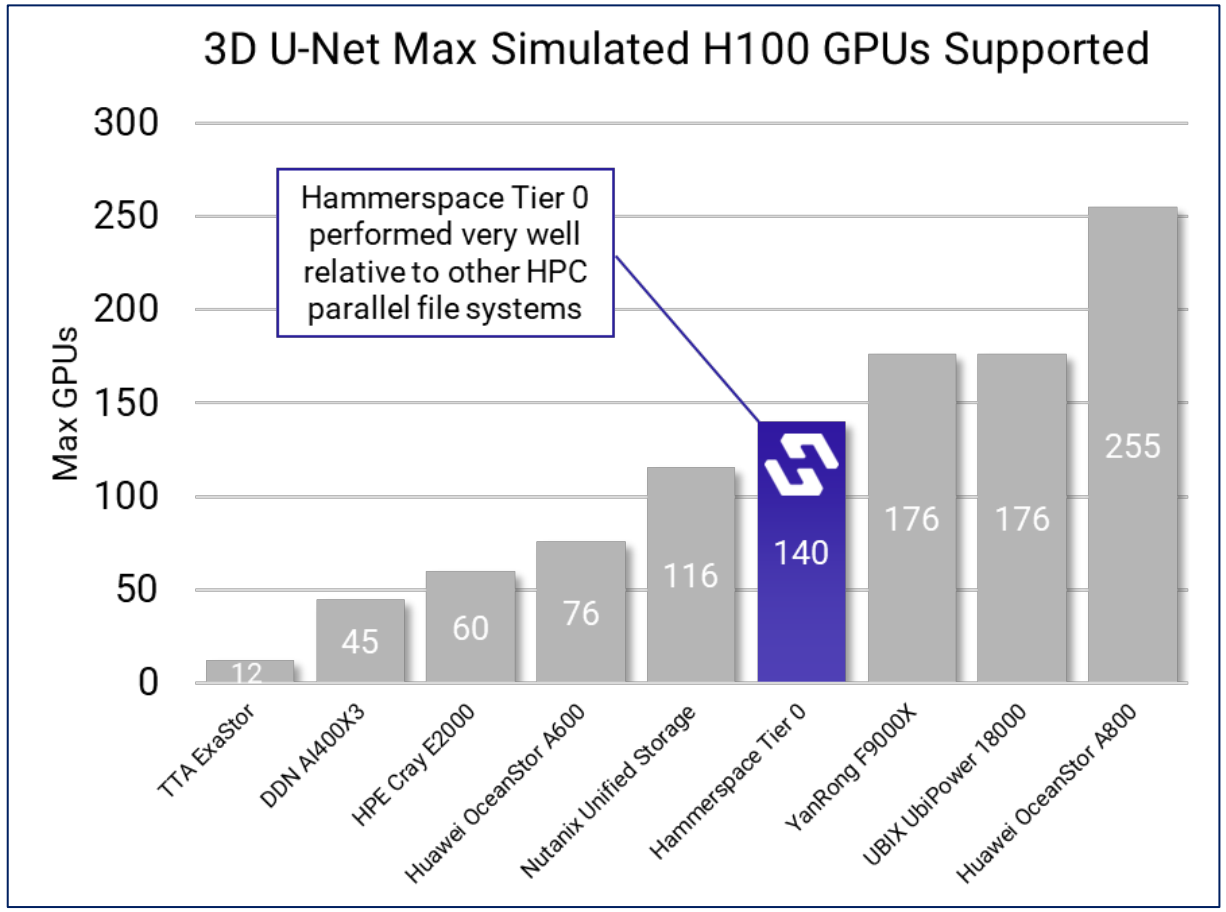
(Assuming Checkpoint Size of 500GB)

	Storage Type	Optimal Throughput	Effective Throughput	Time to Write 500GB	
Tier 0	Local NVMe (8-drive GPU server)	112 GB/s	112 GB/s ✅	~4.5 sec ✅	
Tier 1	AWS EBS (gp3)	1 GB/s	0.9 GB/s	~555.6 sec (~9.3 min)	123x Slower
	AWS EBS (io2)	4 GB/s	3.6 GB/s	~138.9 sec (~2.3 min)	30x Slower
	AWS S3 (via high-speed network)	1 GB/s	0.9 GB/s	~555.6 sec (~9.3 min)	123x Slower
	Microsoft Azure Ultra Disk	10 GB/s	9 GB/s	~55.6 sec	12x Slower
	Microsoft Azure Premium SSD v2	1.2 GB/s	1.08 GB/s	~462 sec (~7.7 min)	103x Slower
	Oracle - OCI Block Volume	1 GB/s	0.9 GB/s	~555.6 sec (~9.3 min)	123x Slower

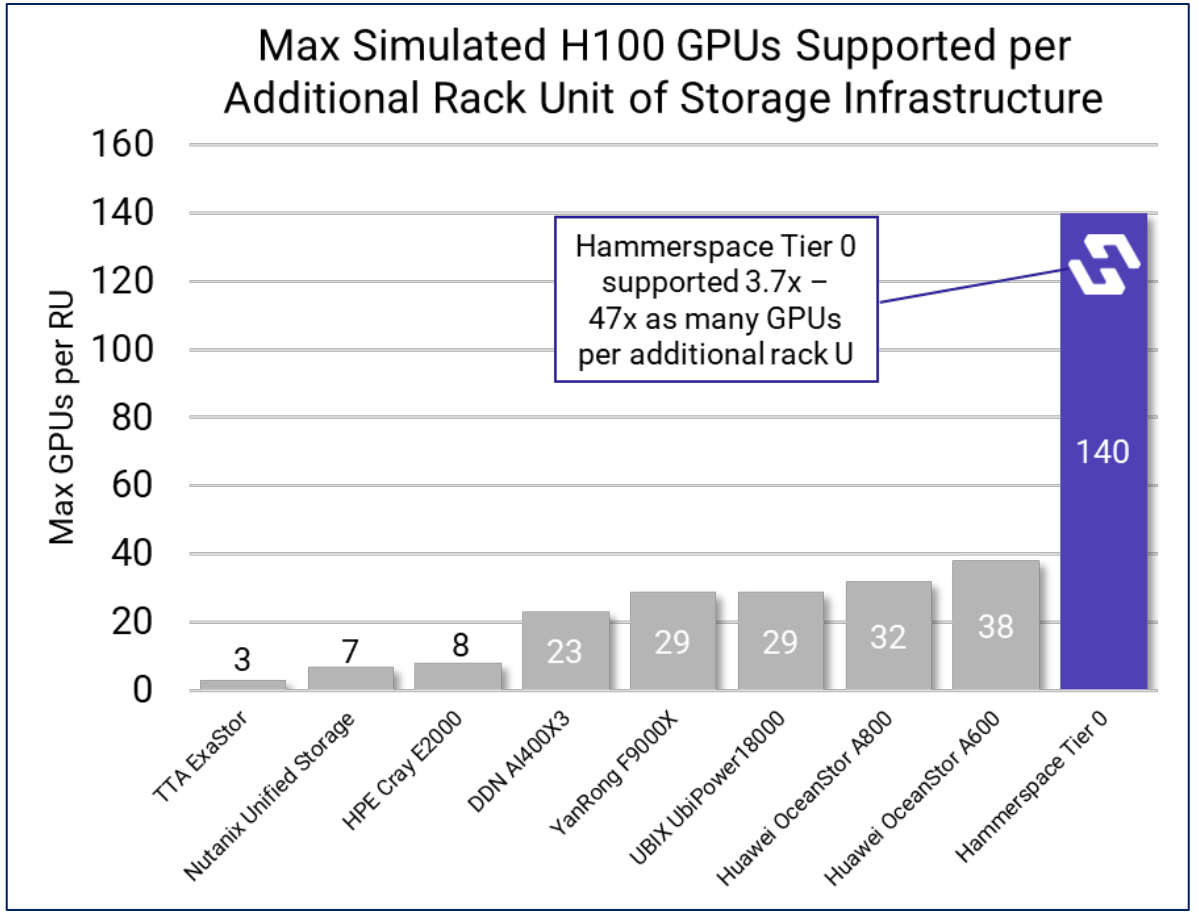
Note: *External Tier 1 storage throughput adjusted assuming 90% practical network efficiency.*

Tier 0 Performance Benchmarking: MLPerf v2.0 Results

Tier 0 Performed Well Relative to Other HPC File Systems

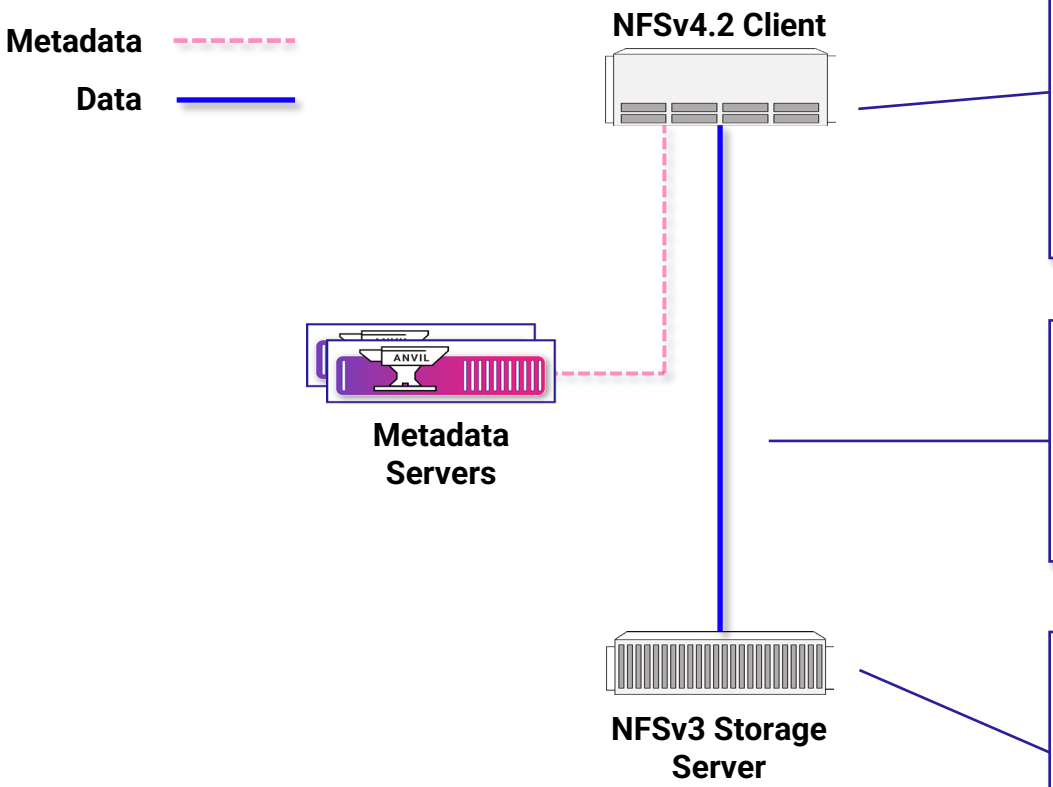


Tier 0 Excels When Measuring Efficiency (Performance per RU)



Standards-Based Parallel File System Architecture

-- Parallel NFS v4.2 with Flex files --



Hardening a Standards-Based Parallel File System

- 2018: Enhanced Parallel NFS spec with pNFSv4.2 with Flex Files
  - Eliminated NFS GETATTR chattiness
  - Added telemetry feedback
  - Added N-Connect
- Part of every modern Linux distribution

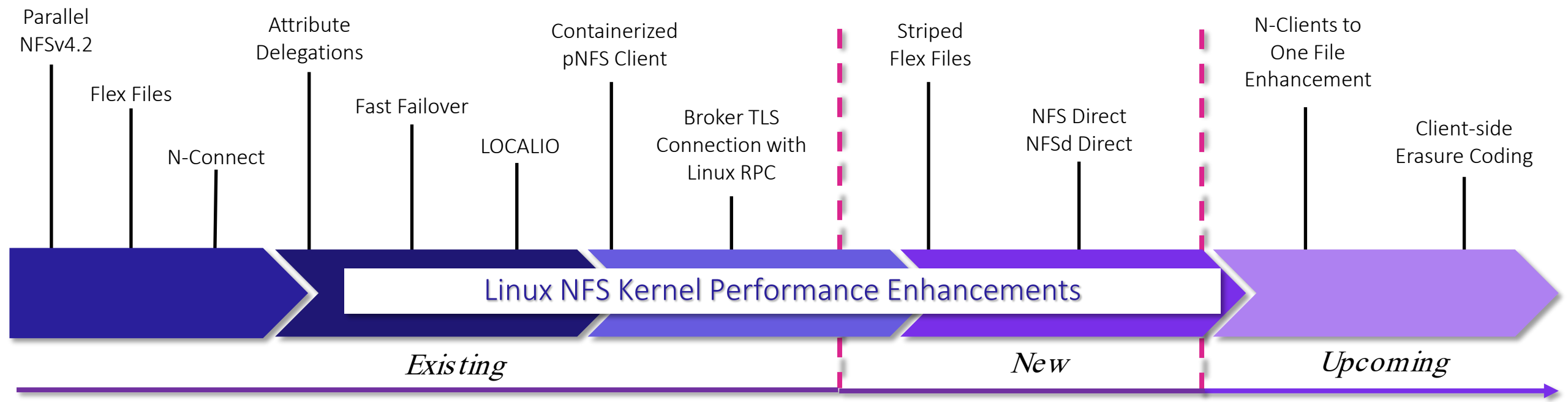
Separate Metadata and Data Paths

- Direct data path using TCP or RDMA
- Provides for multiple parallel connections
- Metadata layer acts as global control plane

Use Any NFSv3 Storage System

- Add any NFS storage volume
- Including GPU server local NVMe
- Automate data protection and placement

A History of Innovation and Commitment to Open Standards



Hammerspace is Engineered to Activate These Capabilities,
Which are Included in All Major Linux Distributions

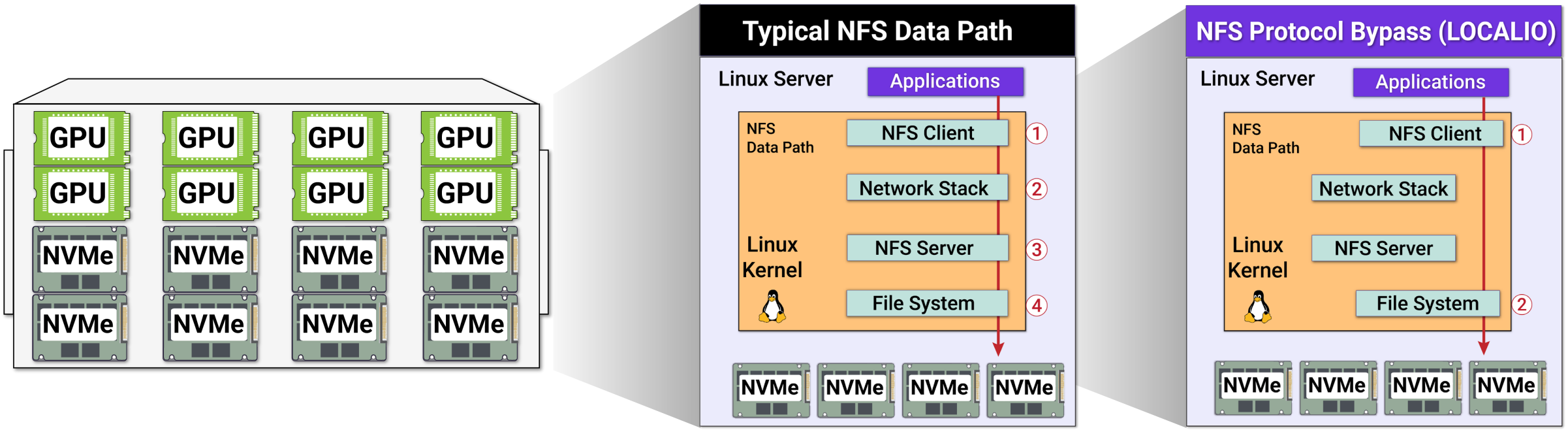
This Means Customers Never Have to Install Proprietary Client Software

Added Benefits Waiting in the Linux Kernel

Further Reduce Latencies and Maximize GPU Utilization Using LOCALIO (NFS Protocol Bypass)

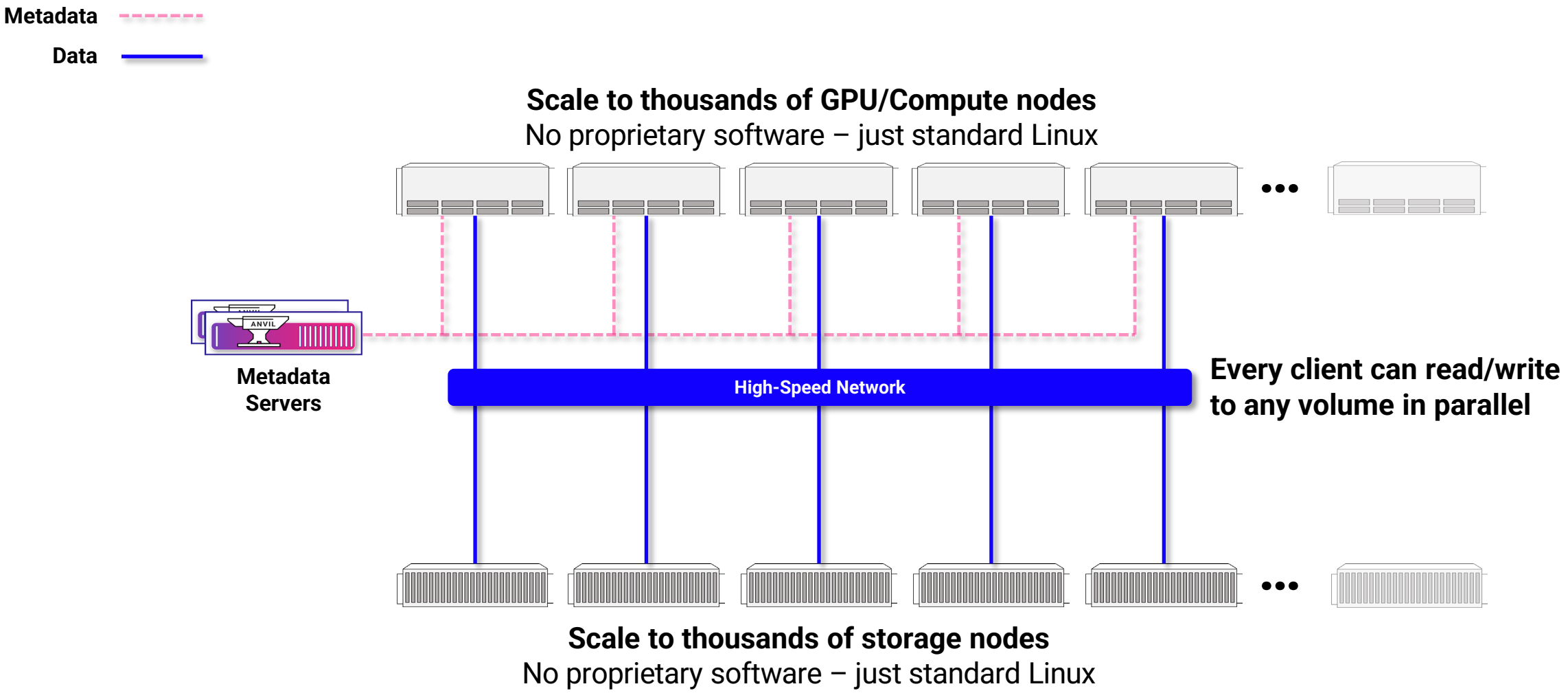
>3x faster read performance

Reduces CPU usage in GPU servers

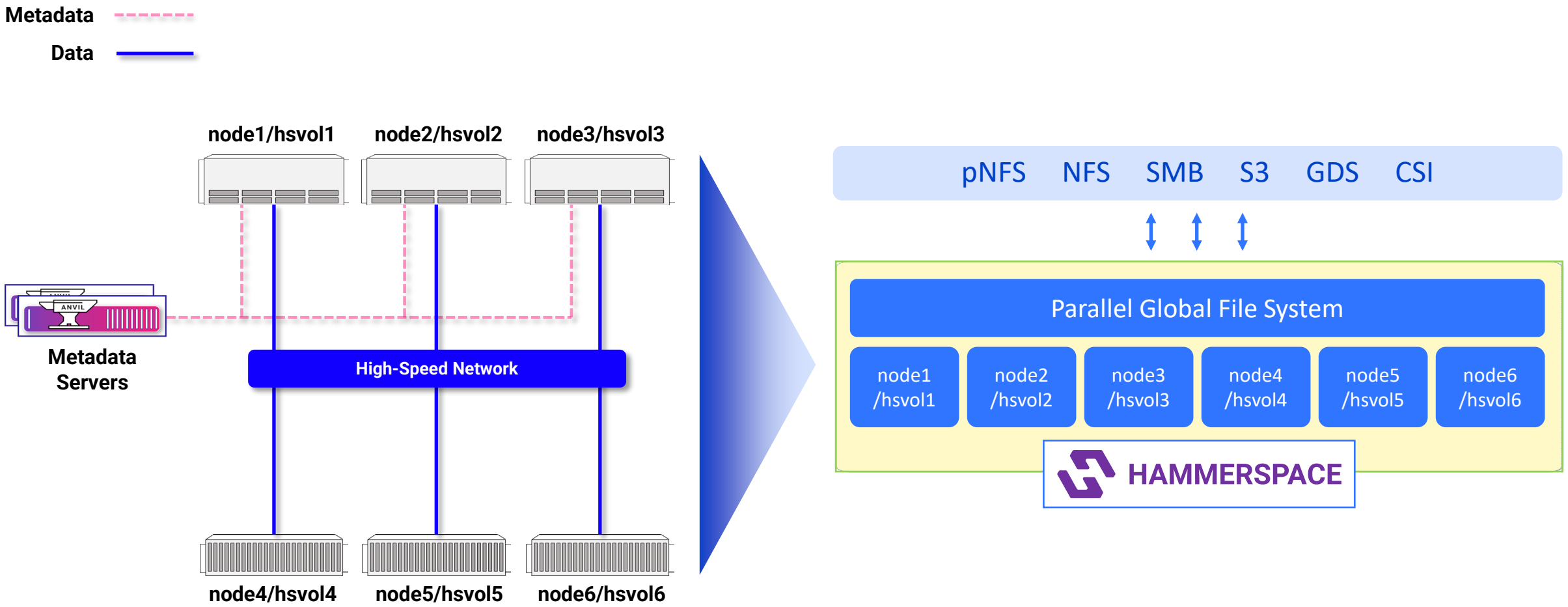


LOCALIO Was Released in Linux Long-Term Support kernel 6.12
Was included in RHEL10 in May 2025

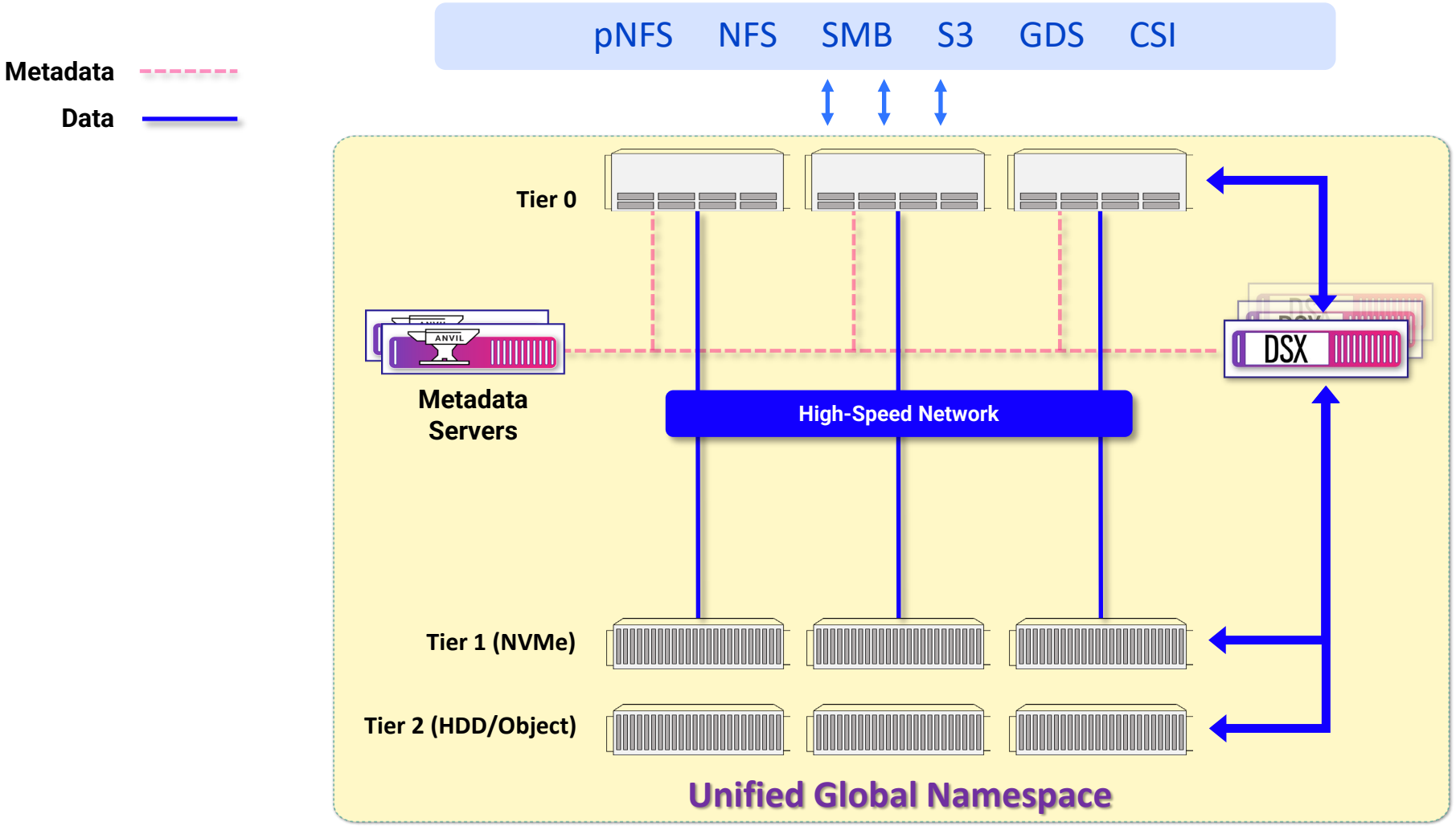
Standards-Based Parallel File System Architecture



Adding Tier 0 Nodes to the Global Namespace

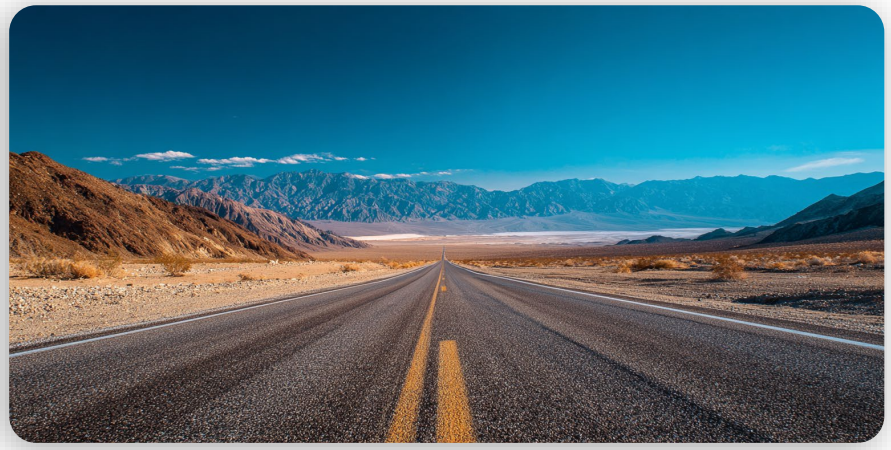


Orchestrating Data Within a Tier 0 Environment



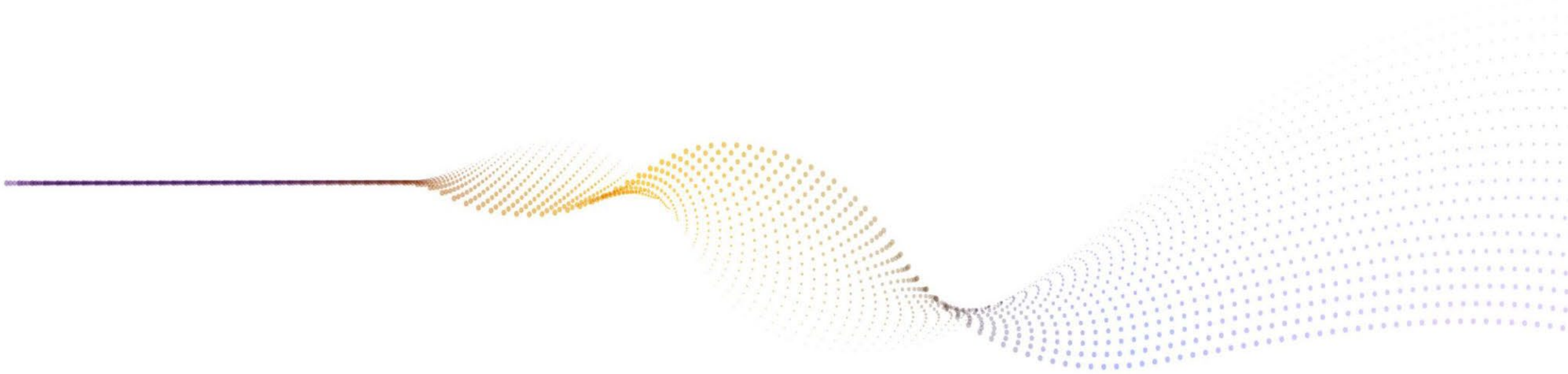
- DSX Nodes Act as Data Movers
  - NFS to NFS
  - NFS to Object
- Scale number of DSX nodes based on size of cluster, amount of data being moved

Open Highways vs Walled Gardens

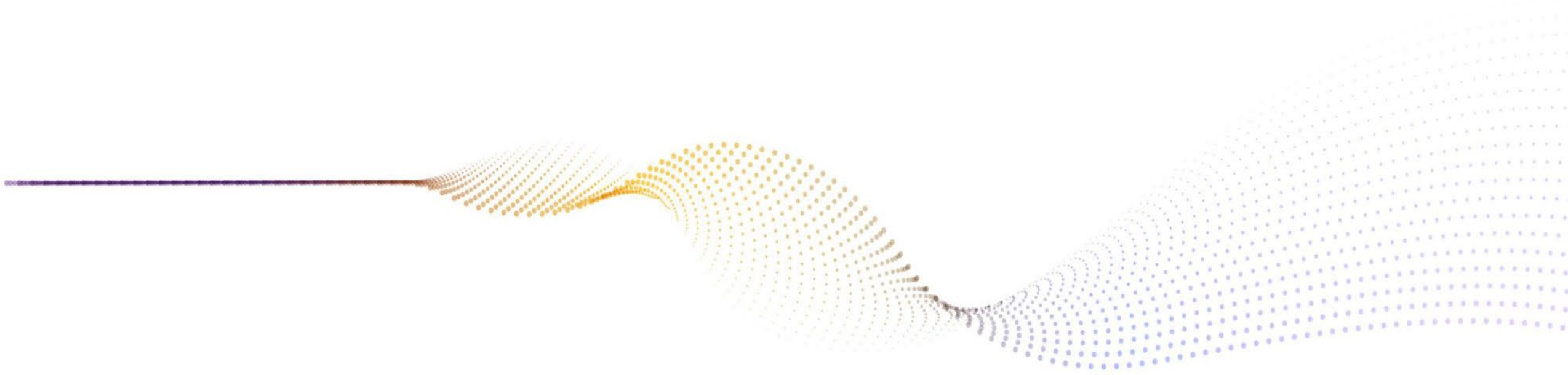


- **Standards-based:**
  No proprietary client software, commodity hardware/networking
- **Extreme Performance:**
  Tier 0 with existing storage GPU Server – Local NVMe speeds
- **Unified access:**
  Extends to unified namespace of any storage, on-prem, cloud

- **Proprietary:**
  Clients, agents, specialized hardware required
- **Bottlenecked Performance:**
  High-performance I/O bottlenecked across the network – Adds cost, latency
- **Siloed Access:**
  Data must be copied, adding risk, cost.



Questions?



Thank you for attending!

Please remember to rate this session. You get access the presentations at
http://sniadeveloper.org/conference

