SNIA DEVELOPER CONFERENCE

SDC 24

BY Developers FOR Developers

September 16-18, 2024
Santa Clara, CA

Standards-based Data Platforms for HPC and AI

David Flynn

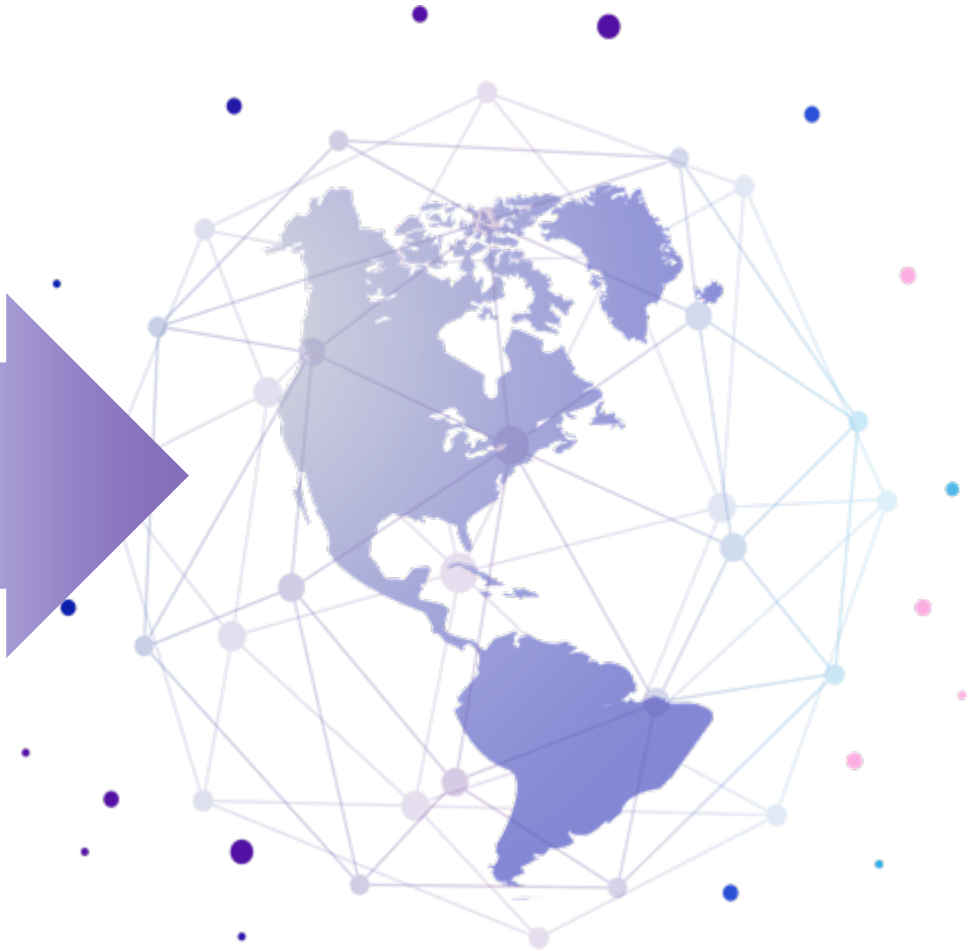
CEO & Founder, Hammerspace

Major Industry Trends Driving New Storage Requirements

LLM Training
Load and Iterate in GenAI

Large, Decentralized
Data Sets

Multi-Site
Multi-Cloud
Remote AI Researchers



SDC 24

Goal: Consolidate File Systems and Data Sources

Proprietary → **Standards-Based**

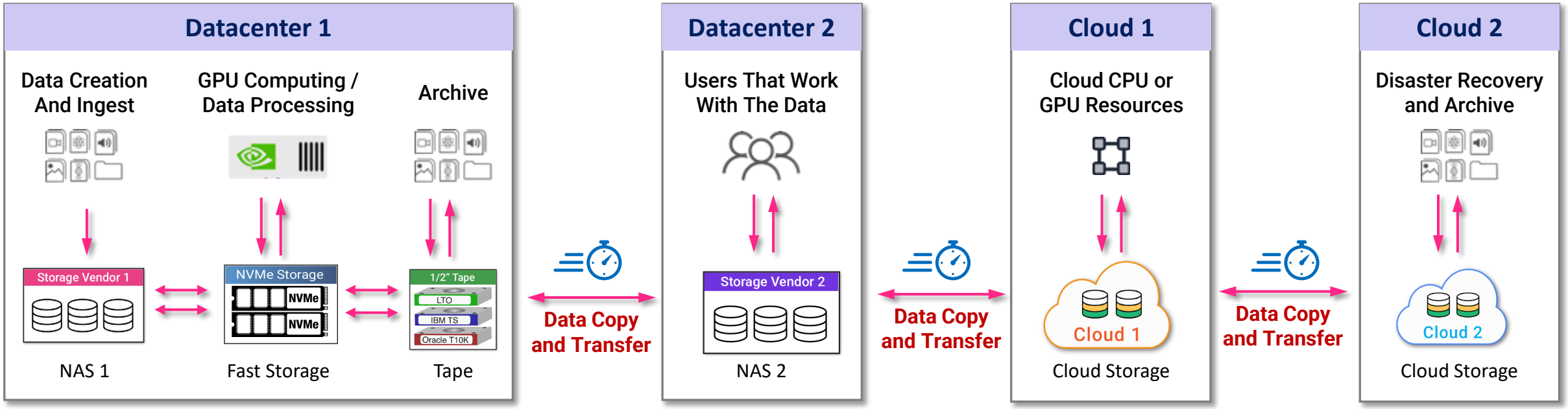
	Traditional HPC Parallel File Systems	Scale-out NAS w/ NVMeoF	Global Data Platform *(parallel file system + data orchestration)*
Large block, large file read/write	🟢	🟡	🟢
Read intensive	🟢	🟢	🟢
AI random read / write	🟢	🟡	🟢
Data distribution	🟡	🔴	🟢
Archive	🔴	🔴	🟢

Evolution of High-Performance File Systems

SDC 24

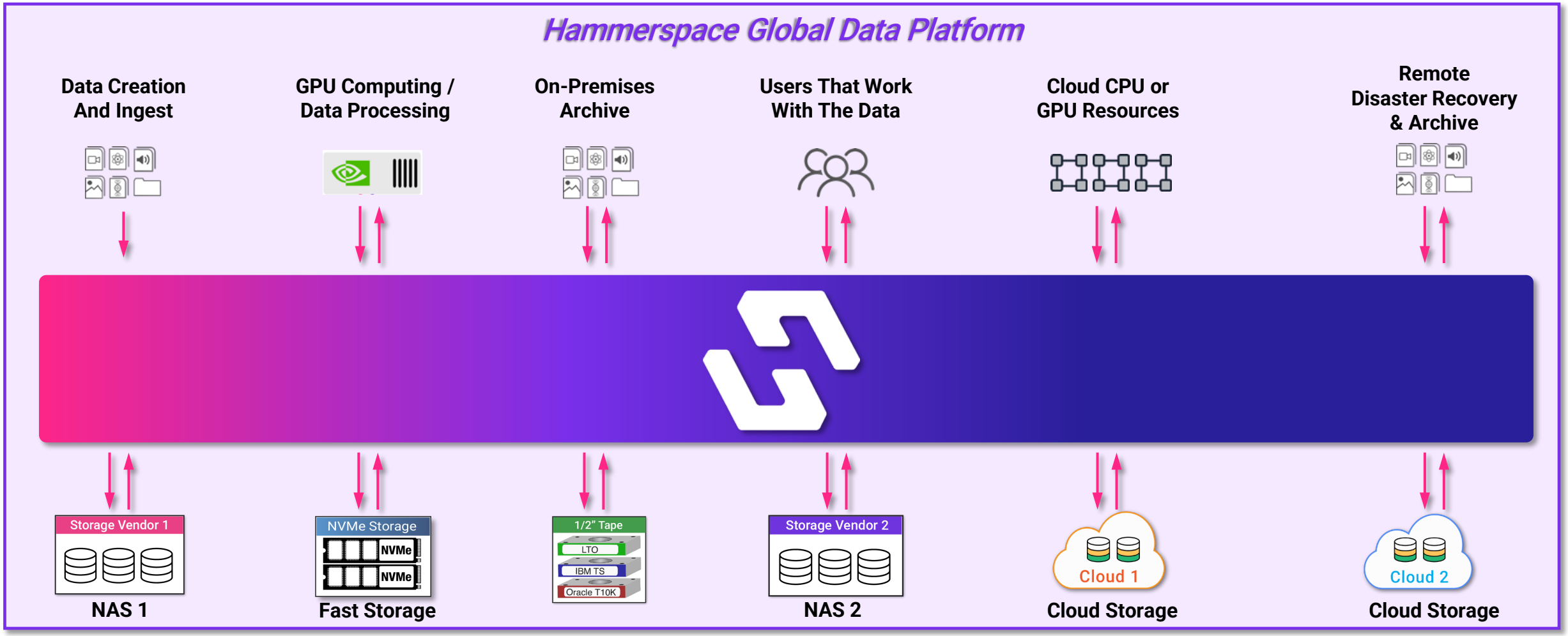
Why? Data Silos Inhibit Progress, Reduce Data Quality

Valuable data **trapped in silos**	Getting data to global users **takes too long**	Infrastructure is **not ready for AI**
Data copy sprawl **impacts cost, governance and security**	**Lack performance** to keep GPUs utilized	**Lack agility** to use elastic cloud resources



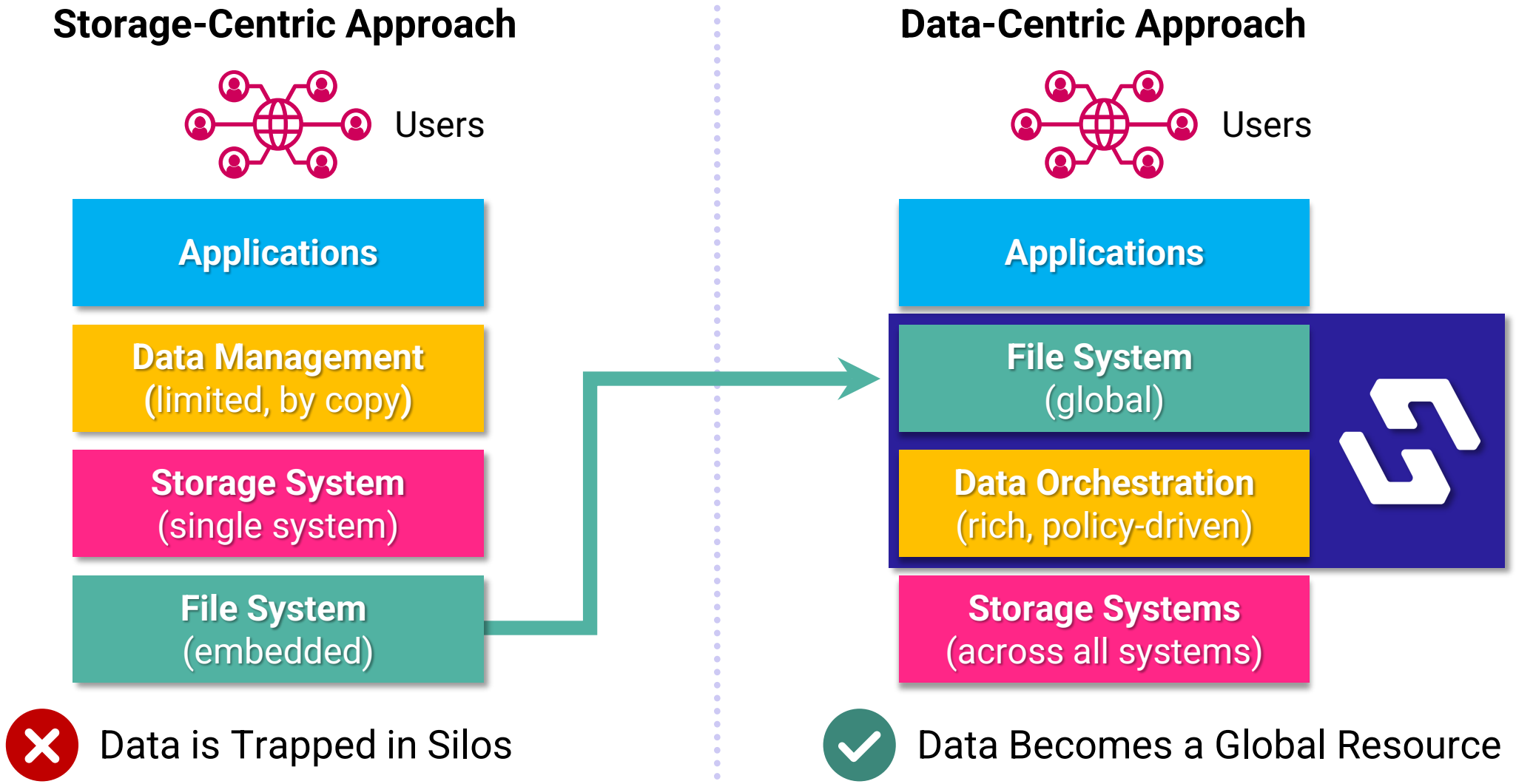
Standards Are A Must for Global Data Use

Unify and automate unstructured data across any data center, any cloud, anywhere



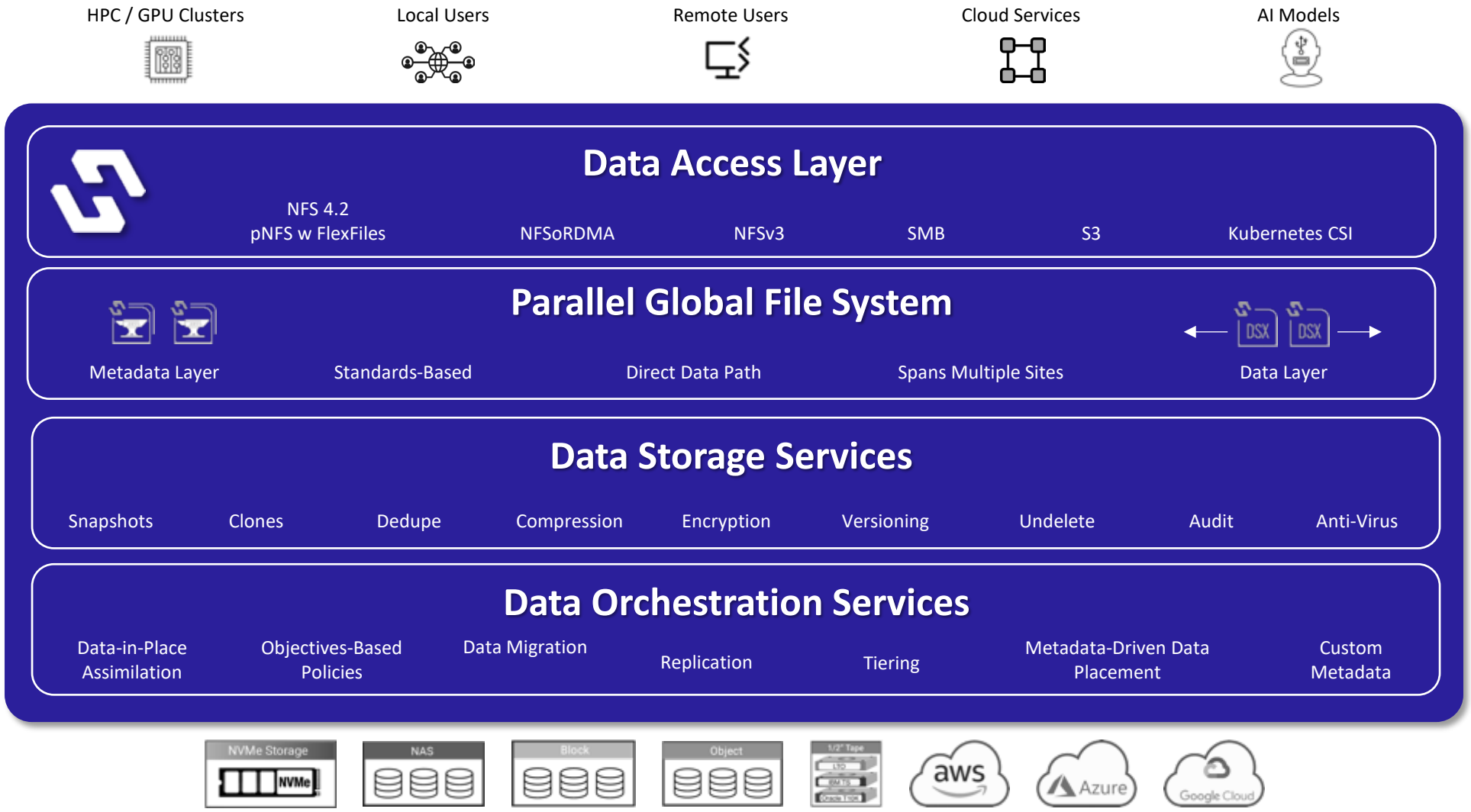
SDC 24

Distributed Workloads Require a New Architecture



SDC 24

High-Performance Global Data Platform Architecture

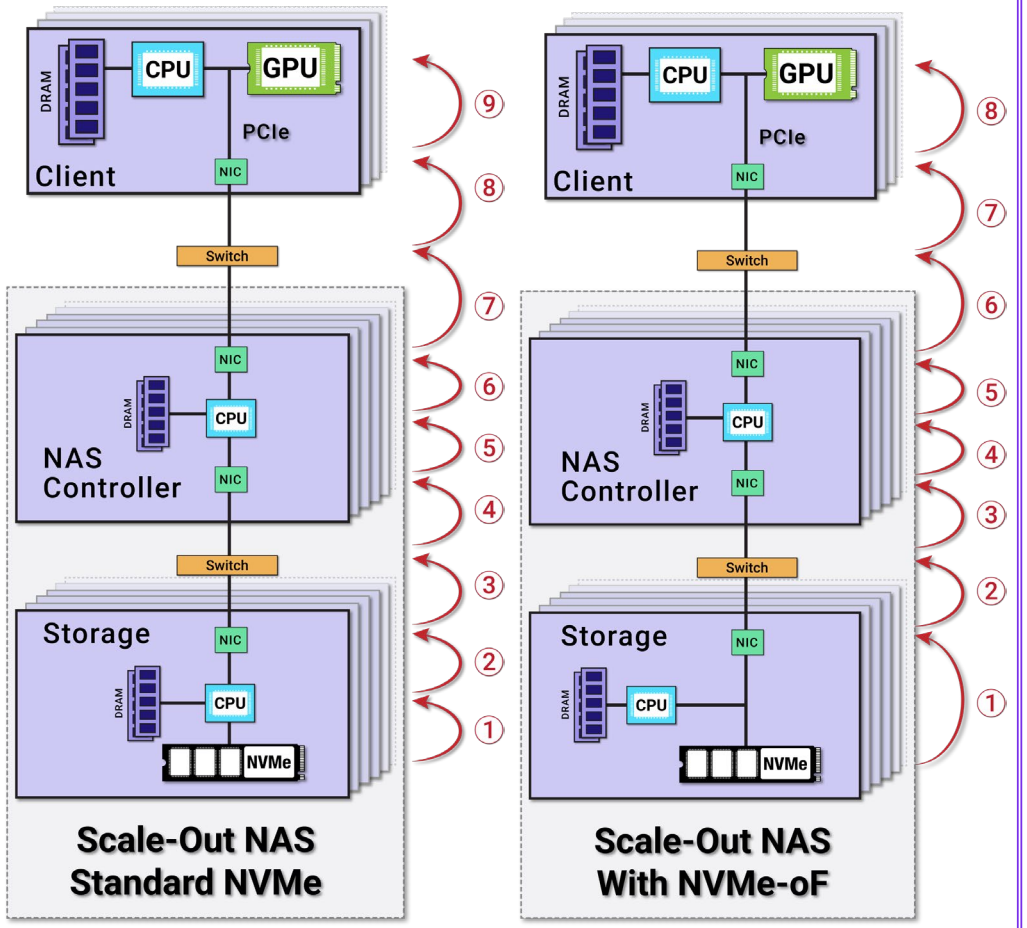


Runs on New or Existing Storage from All Leading Vendors

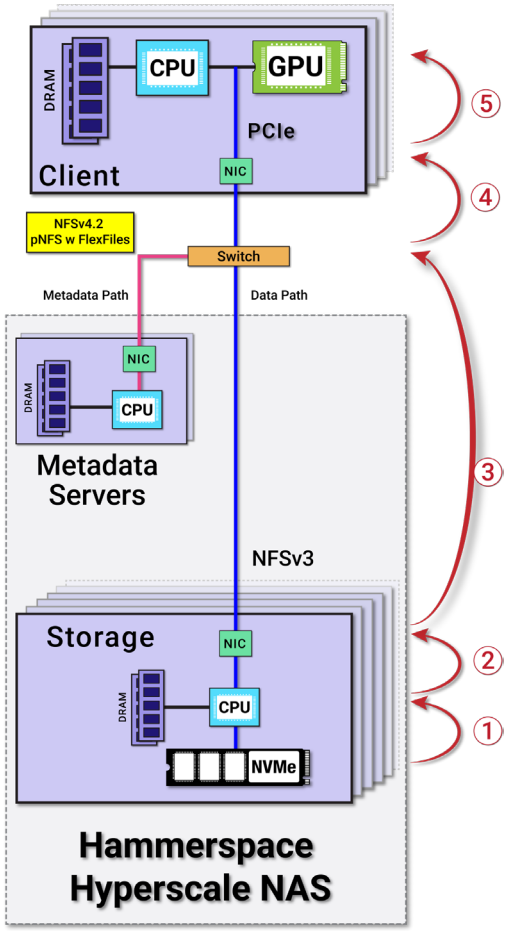
SDC 24

NFSv4.2 Enables Storage Efficiency & Performance

Scale-Out NAS Architectures



Hyperscale NAS Architectures

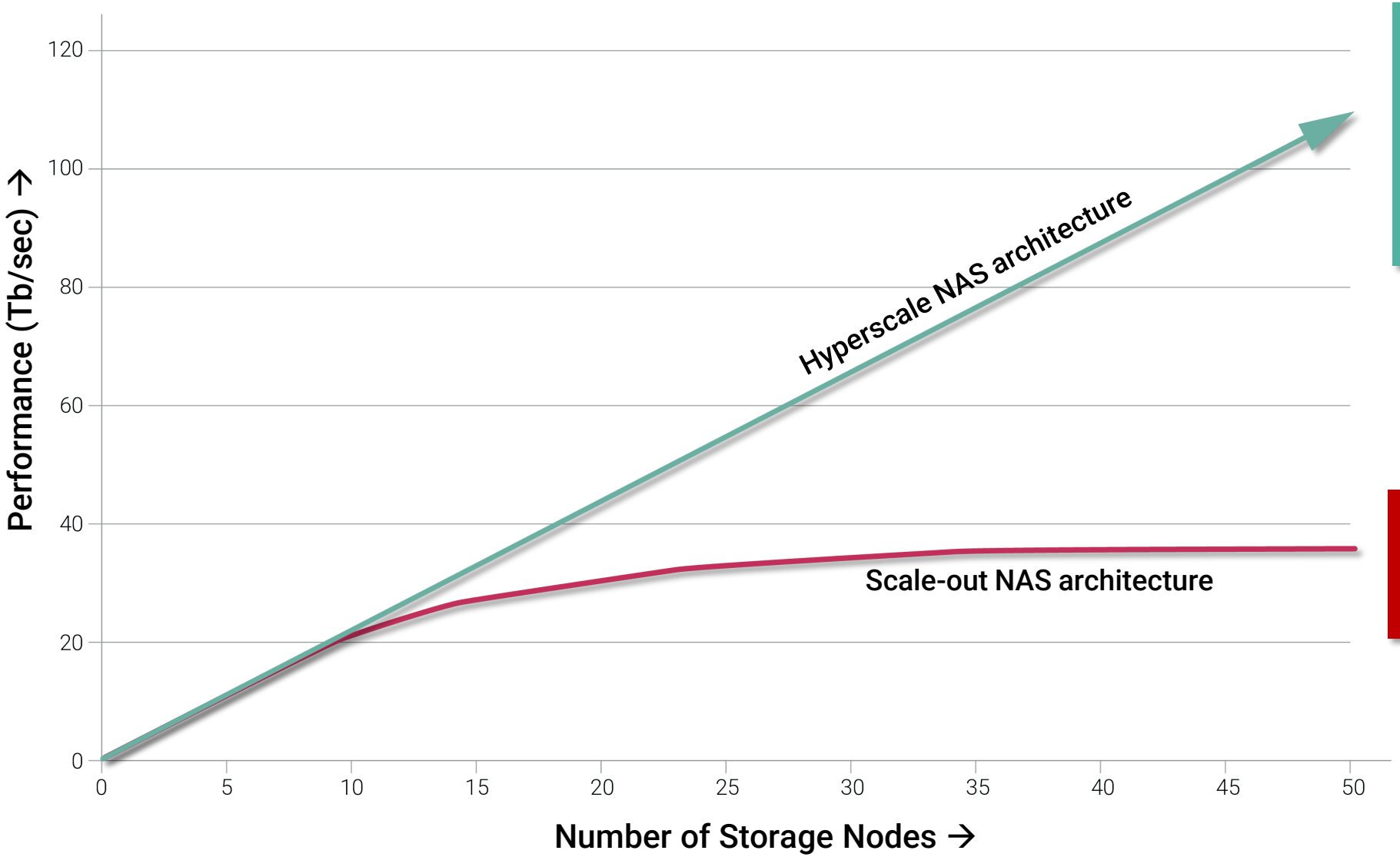


2x

Reduction in Servers, Networking, Latency, Watts and Rack U

SDC 24

NFSv4.2 + Parallel File System



Hyperscale NAS increases performance linearly as node count increases.

So far proven at over 1,000+ storage nodes.

Scale-Out NAS performance plateaus as storage nodes increase

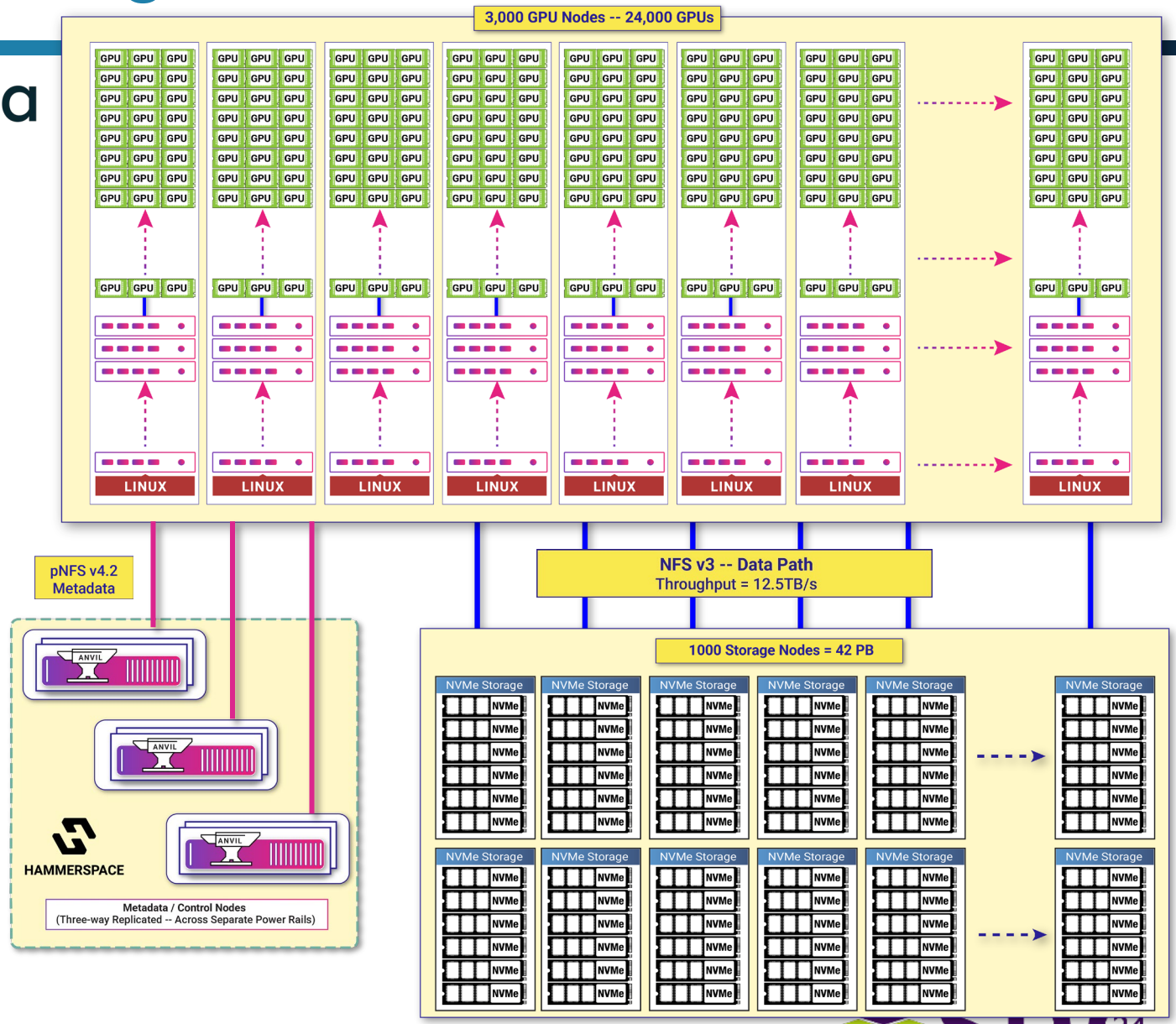
Customer Story: LLM Training

About the Customer

- Meta's AI Research Super Cluster
- Powering Llama 2 & 3 LLMs
- Massive performance and scale demands
- Plan to grow to 350,000 NVIDIA H100s by end of 2024
- Evaluated leading storage vendors as well as considered "build your own"

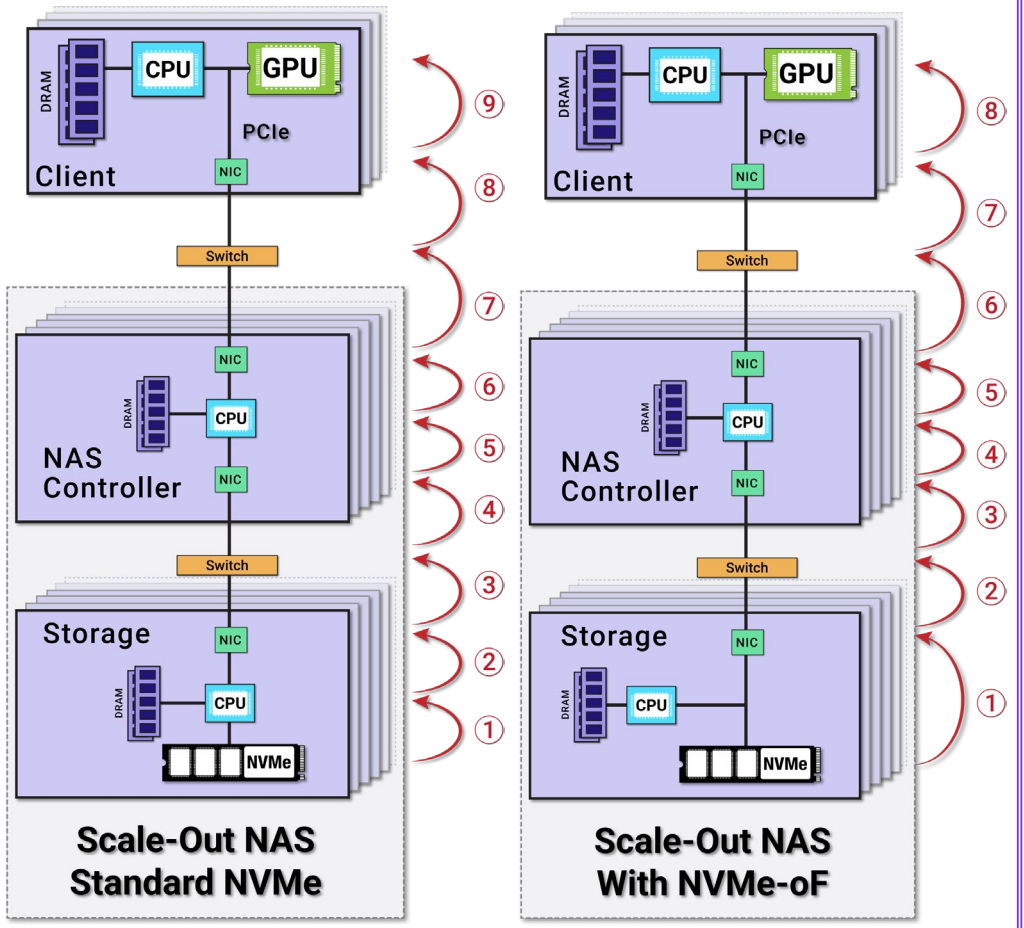
Hyperscale NAS Solution

- Triple redundancy on metadata nodes
- 42PB across existing **1,000+ node storage cluster**
- Feeding 24,000 GPUs, soon to be 350,000, then 1M
- Aggregate performance of **12.5TB/sec (100Tb/sec)**
- Everything is **standards-based** and **plug-n-play**
- Customer was able to use **existing OCP storage servers**
- Exceptional reliability (100's of storage nodes can fail)

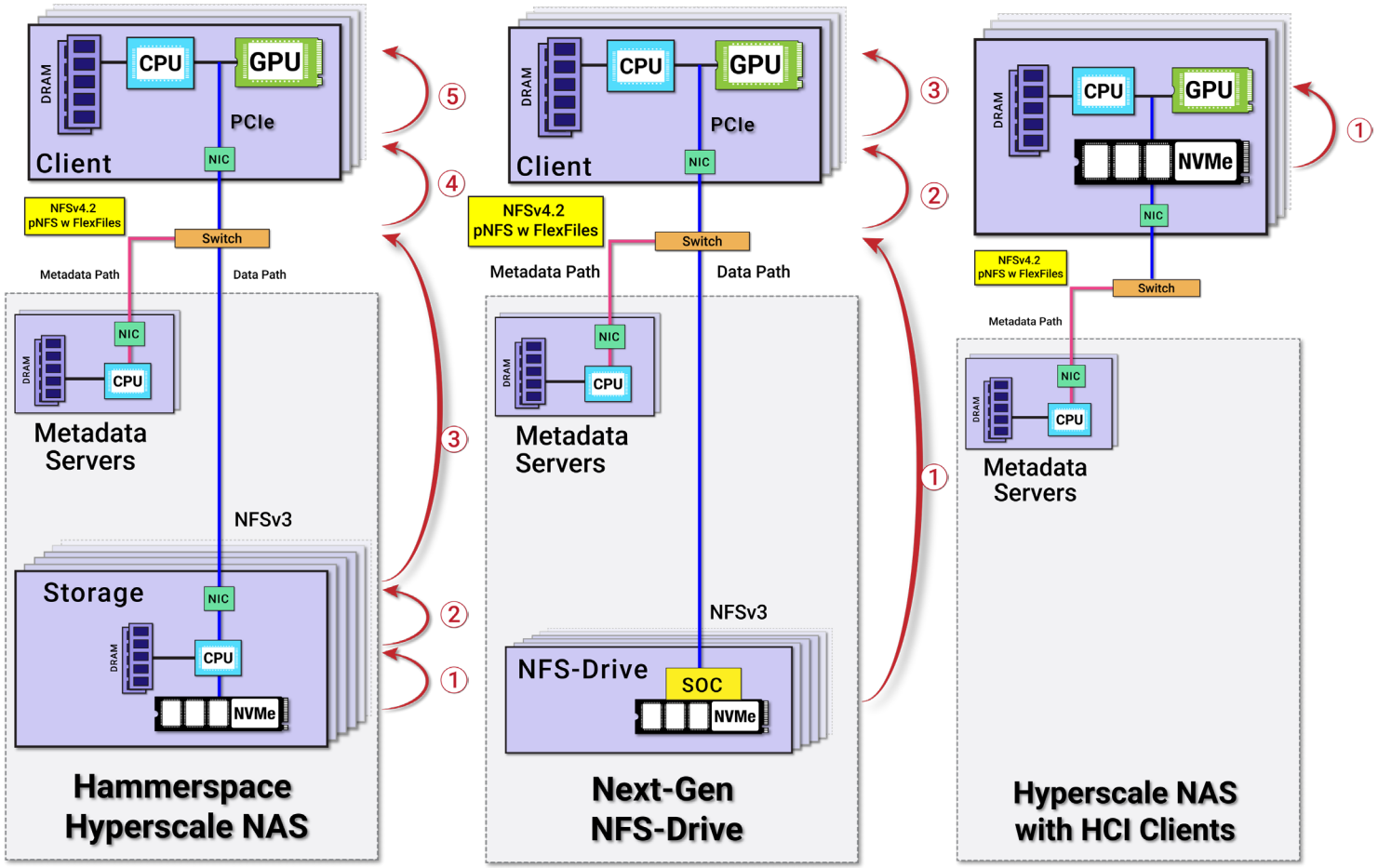


Future Advancements Will Continue to Drive File and Object Storage Infrastructure Efficiency and Performance

Traditional NAS Architectures



Hyperscale NAS Architectures

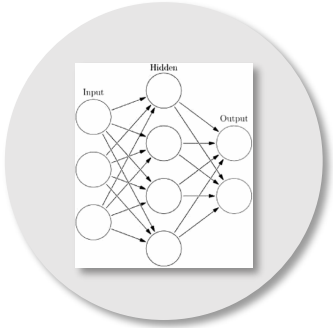


2x
Improvement in Latency, BW and IOPS per $, Watt and Rack U

5x
Improvement in Latency, BW and IOPS per $, Watt and Rack U

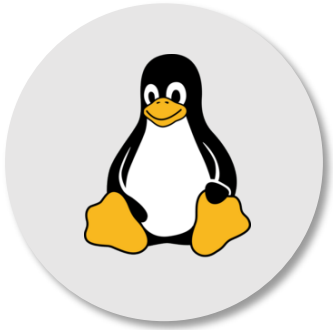
10x
Improvement in Latency, BW and IOPS per $, Watt and Rack U

Standards-Based Global Data Platforms



Why Needed Now

- AI and GPU Computing workloads are forcing HPC into the mainstream
- Performance and agility are needed now like never before



Why Possible Now

- Linux won the Unix wars. Now the significant investment to make a much
- Smarter (and more complex) NFS client needs only be done in Linux.

Why Hammerspace

- Unique talent: Technical CEO, CTO is Linux kernel maintainer, best-in-class team
- Unique technology: Addresses requirements of AI and GPU computing

Thank You

Please take a moment to rate this session.

Your feedback is important to us.

SDC 24