Data Platforms for AI & ML

Molly Presley – SVP Global Marketing
molly@hammerspace.com

FMS 2024

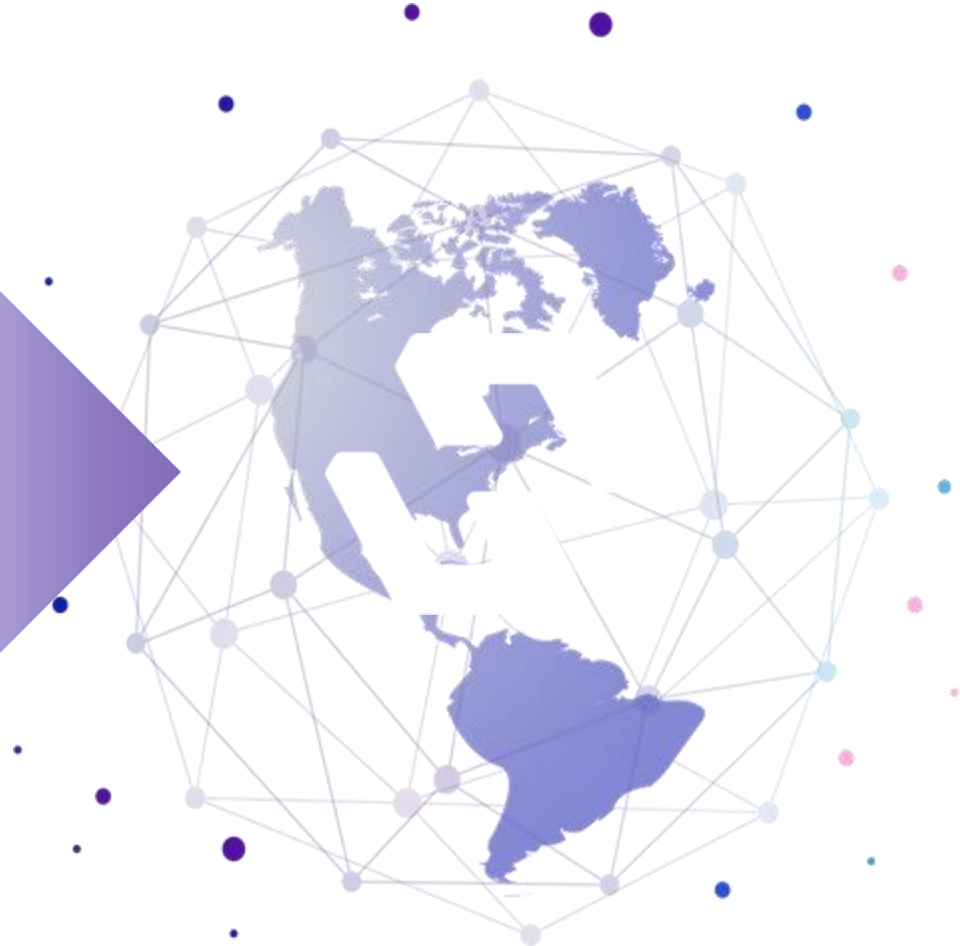
the Future of Memory and Storage

Major Industry Trends Driving AI Storage Requirements

LLM Training
Load and Iterate in GenAI

Large, Decentralized
Data Sets

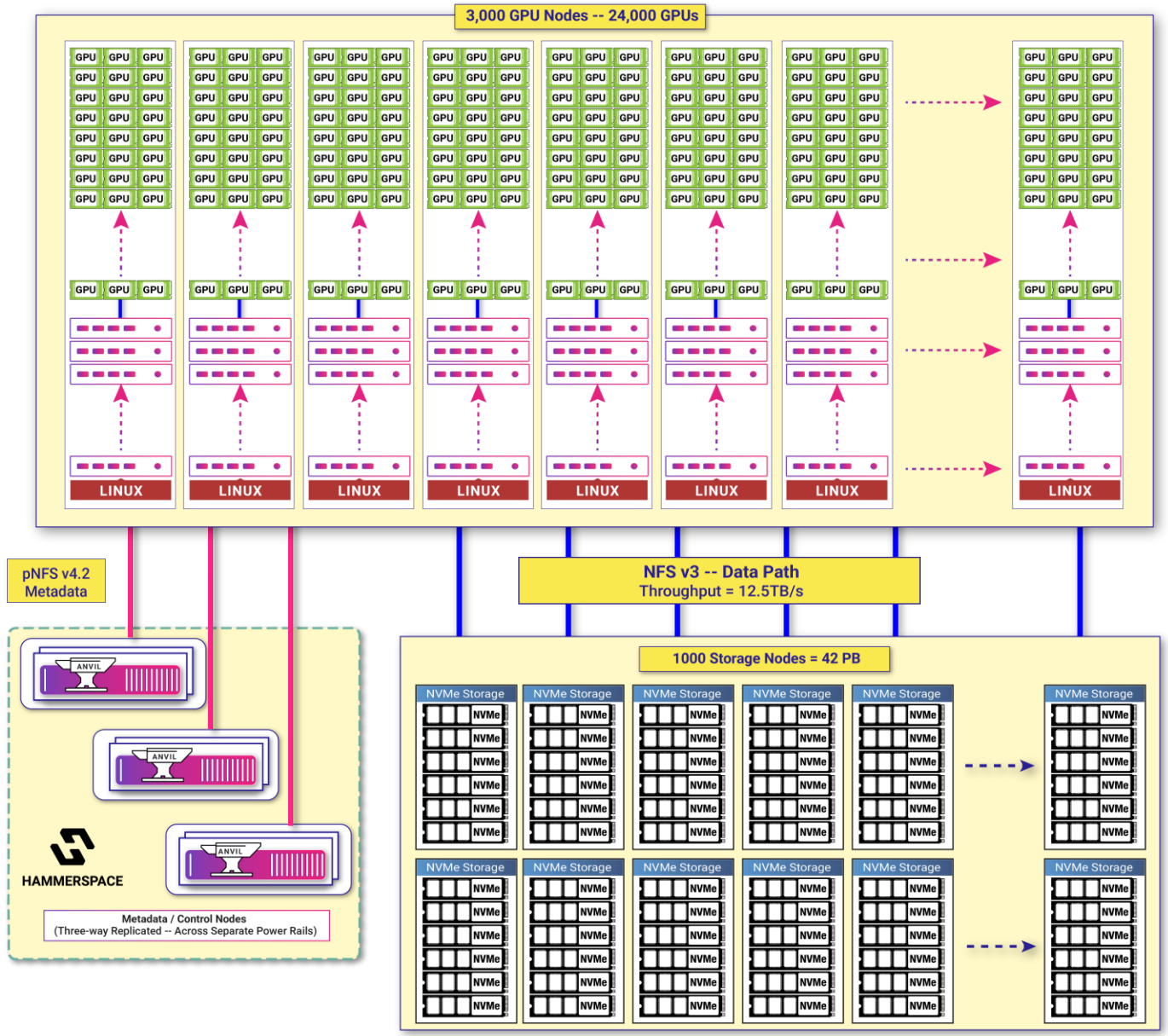
Multi-Site
Multi-Cloud
Remote AI Researchers



HAMMERSPACE

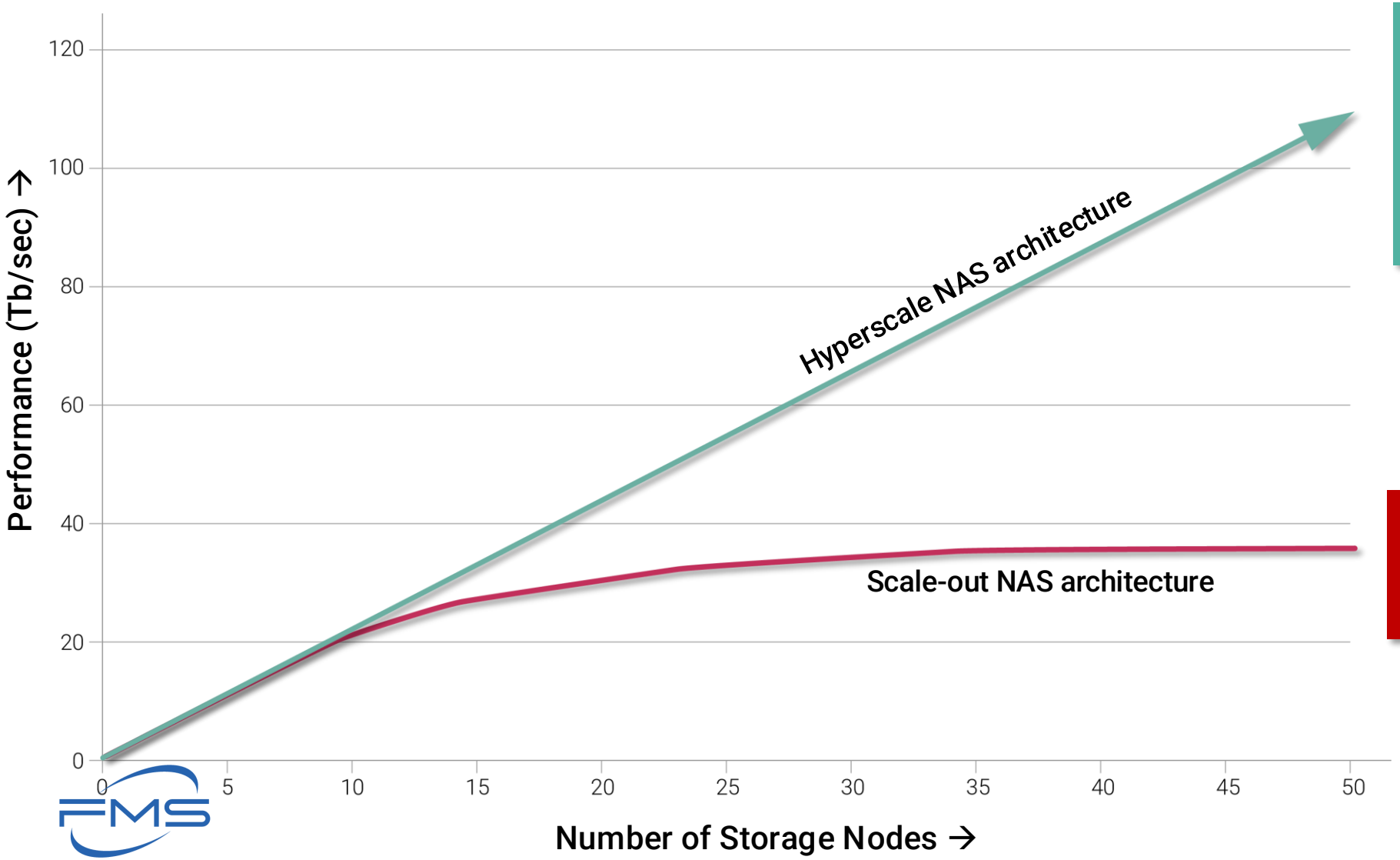
Meta Llama LLM Training Data Architecture

- Standards-based and plug-n-play
- 1,000+ node storage cluster
- 24,000 GPUs, soon to be 350,000, then 1M
- Highly reliable (100's of nodes can fail)
- Performance of 12.5TB/sec (100Tb/sec)
- Use existing storage servers (OCP)



FMS
the Future of Memory and Storage

Massive Linear Performance Scalability Required



Hyperscale NAS increases performance linearly as node count increases.

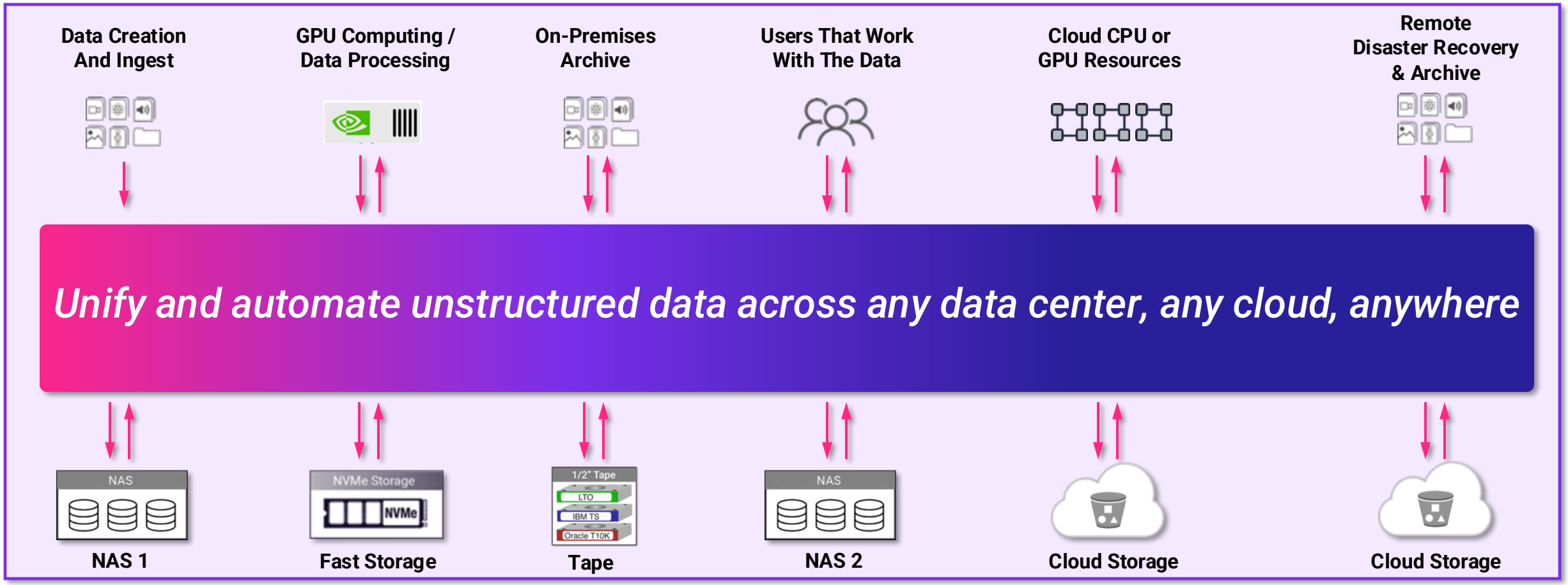
So far proven at over 1,000+ storage nodes.

Scale-Out NAS performance plateaus as storage nodes increase

Data Platform Requirements for Optimal AI Results

	Scale-Out NAS	HPC File Systems	Webscale Architectures	Hyperscale NAS	Hyperscale NAS Global Data Platform
SAFE Reliability, Availability, Serviceability	✓	✗	✓	✓	✓
EASY Standards-Based, Plug-N-Play	✓	✗	✗	✓	✓
FAST HPC-Class Performance	✗	✓	✗	✓	✓
AFFORDABLE Cost Effective at Scale	✗	✓	✓	✓	✓
DATA ASSIMILATION Instant, Data-In-Place, No Copy	✗	✗	✗	✗	✓
DATA ORCHESTRATION Non-Disruptive, Cross-System Data Movement	✗	✗	✗	✗	✓
GLOBAL FILE SYSTEM Multi-Site, N-Way Active/Active	✗	✗	✗	✗	✓
PROGRAMMABLE METADATA Custom Tags, Cross-Platform Metadata Inheritance	✗	✗	✗	✗	✓
ADVANCED DATA SERVICES Global Snaps, Clones, Versioning, WORM, etc.	✗	✗	✗	✗	✓

Global Data Platform's Power AI/ML



Thank you!

molly@hammerspace.com

FMS

the Future of Memory and Storage