

# Orchestrating the Future

Seamless Workload & Data Management  
in Distributed HPC

David Flynn  
CEO & Founder, Hammerspace

Matthew Shaxted  
CEO & Founder, Parallel Works



# “Hammerspace is Like Magic”

*Principal Engineer, Meta*



Hammerspace is an extra-dimensional space that is instantly accessible and infinite in size.



# The Dawn of a New Era





# Evolution of Data Permanence and Access





# Evolution of Data Permanence and Access



# Evolution of Data Permanence and Access





# Evolution of Data Permanence and Access



# Evolution of Data Permanence and Access





# Enterprise Data Architectures Are Evolving Fast to Meet Demands of AI



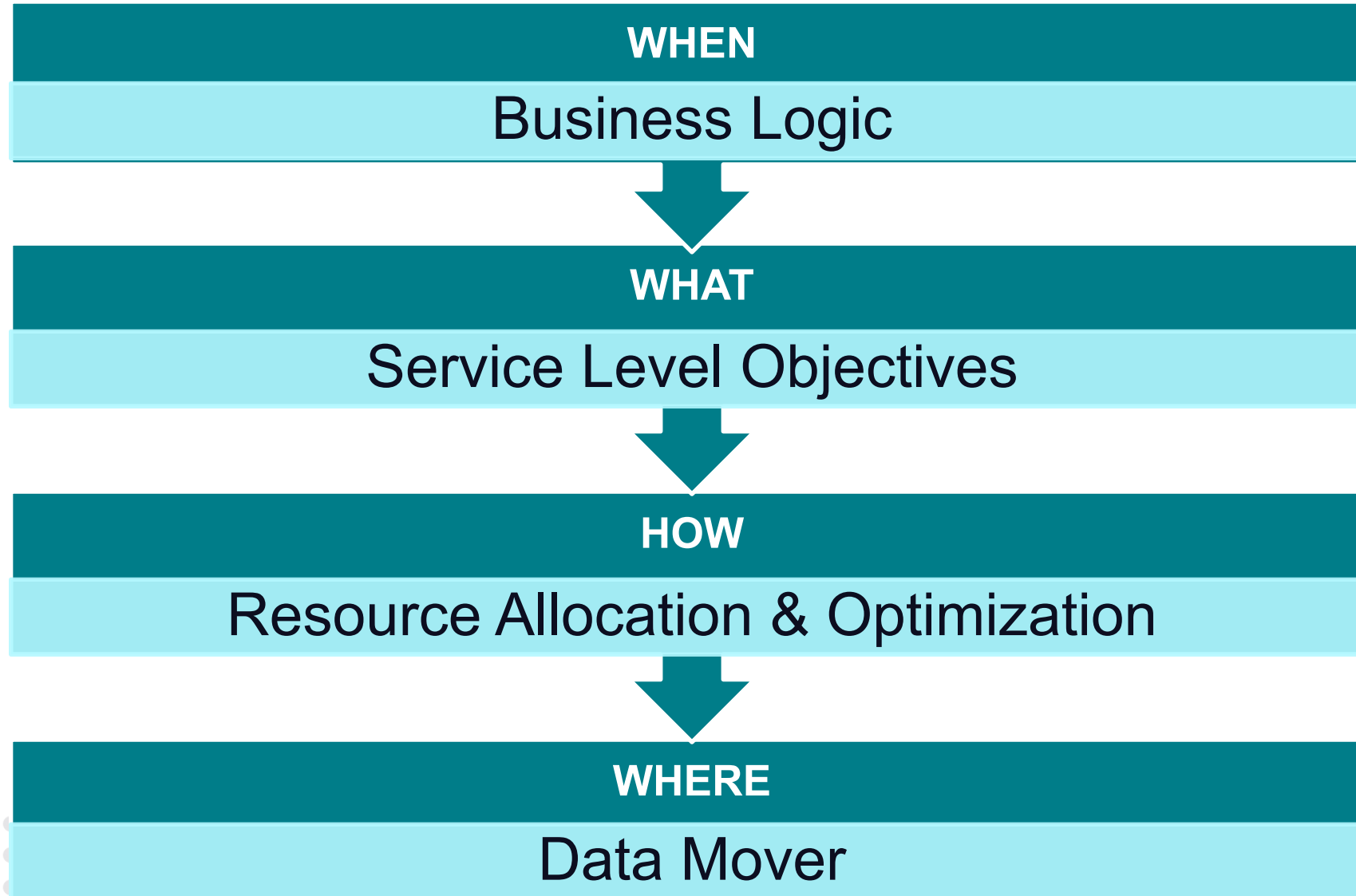
**FROM:**  
**Data at Rest**  
**Siloed in Storage**



**TO:**  
**Data in Motion**  
**to Compute or Storage**



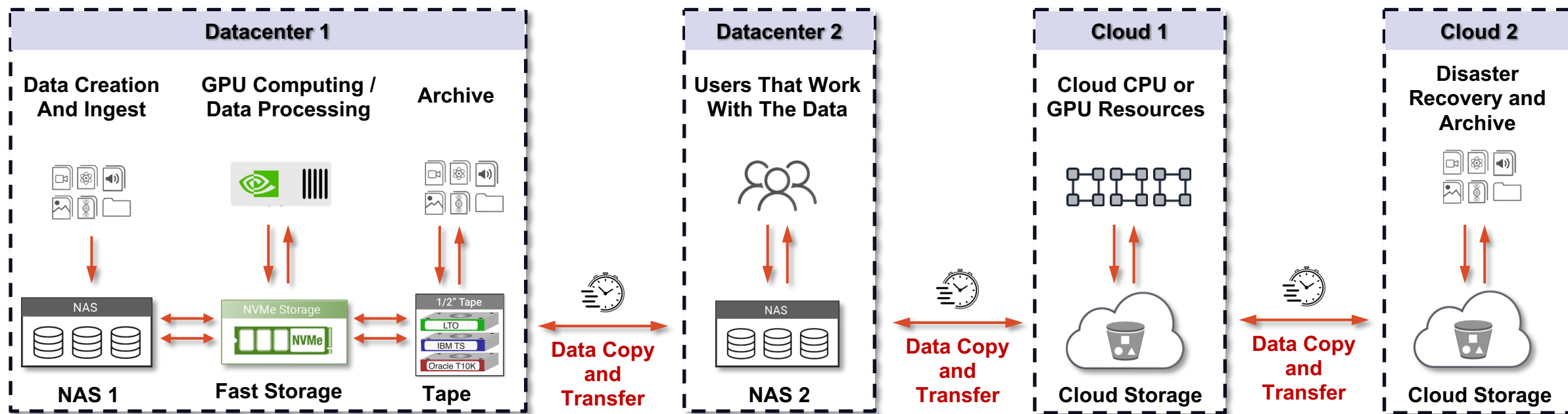
# Business Objectives Dictate Data Placement





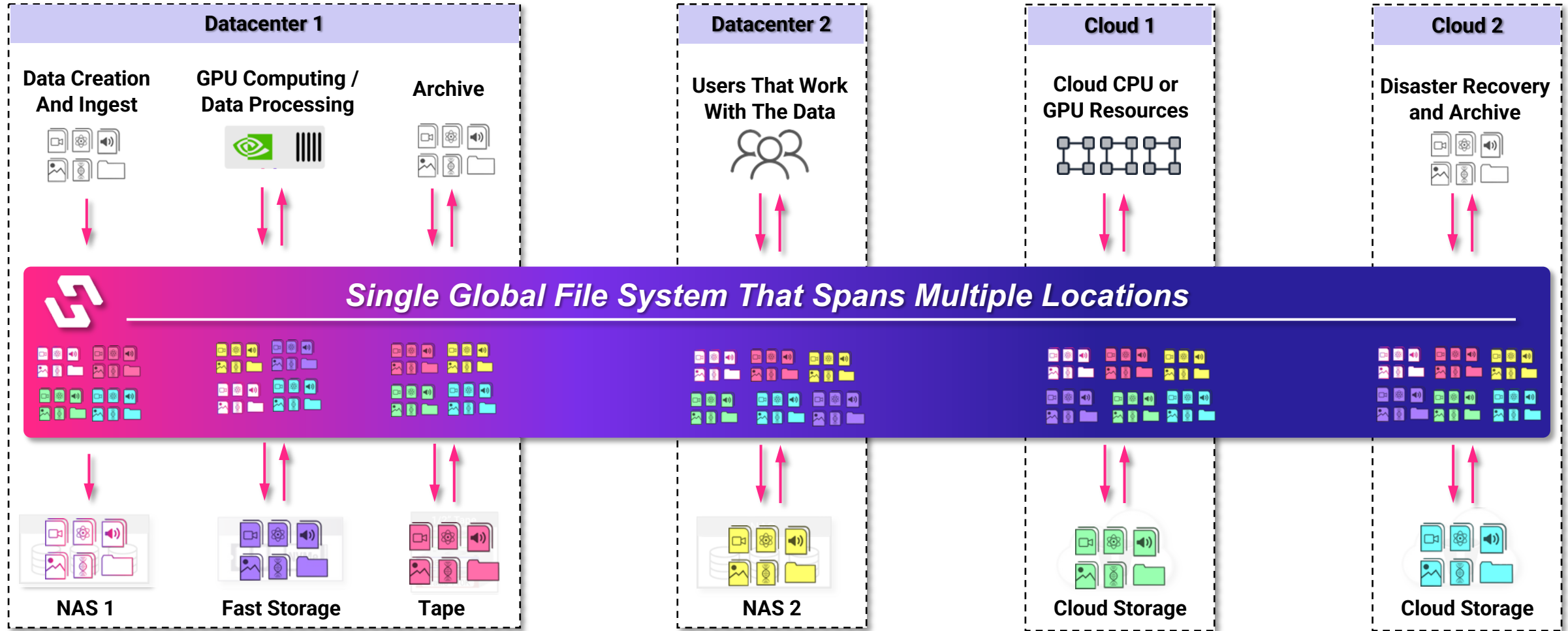
# Data Silos Inhibit Productivity

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



# Hammerspace

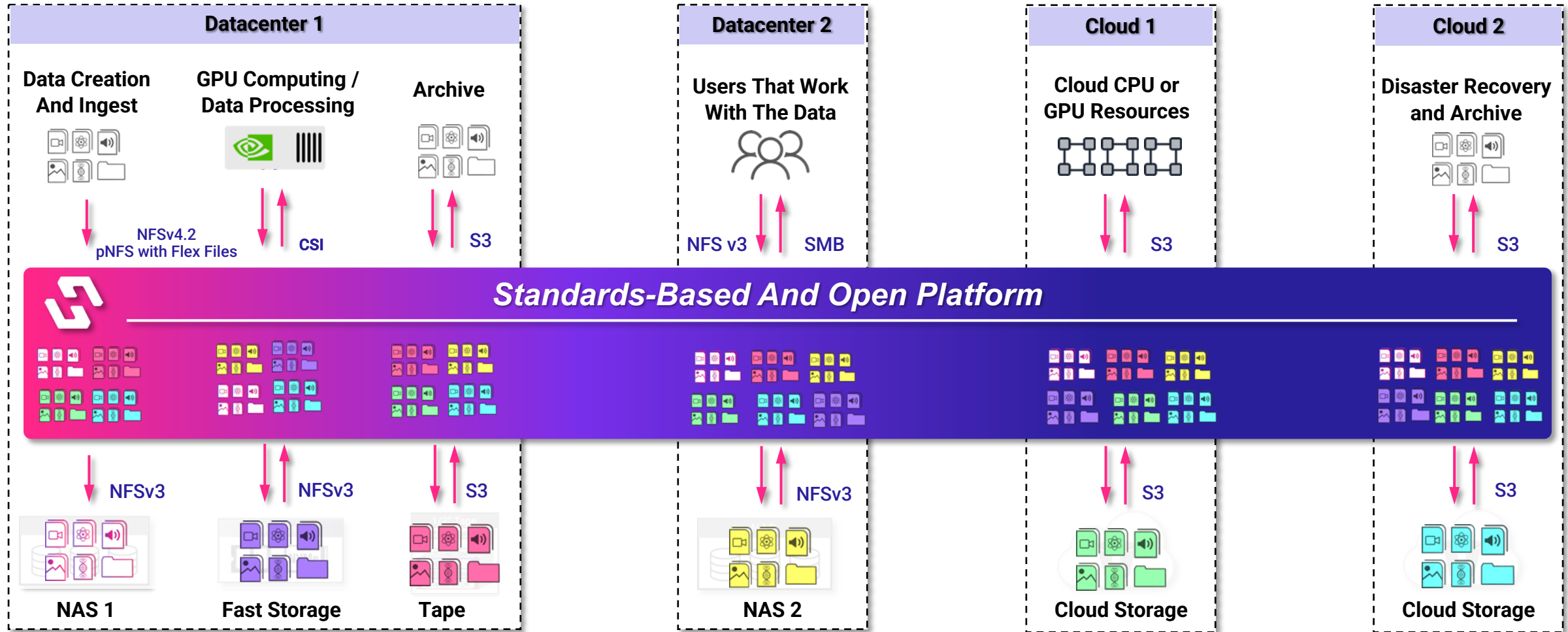
## *Breaks Down Data Silos*





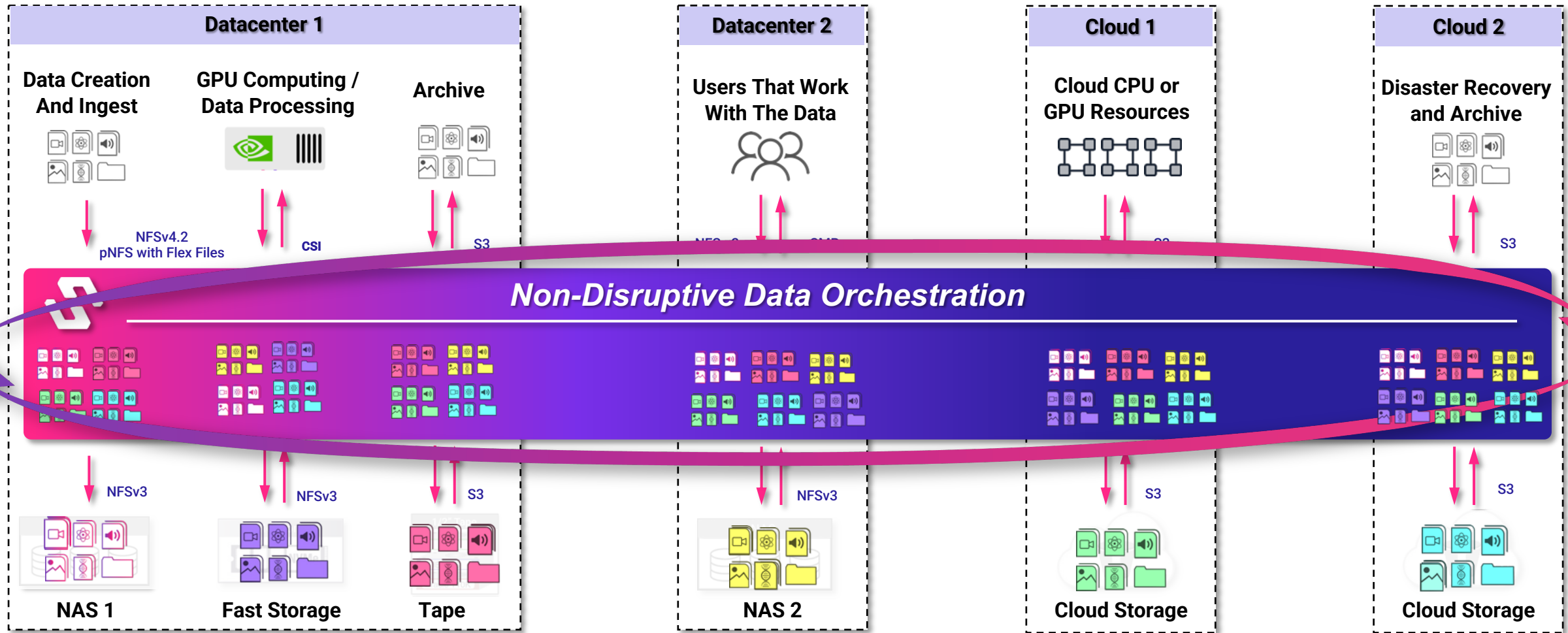
# Hammerspace

## *Eliminates Infrastructure Lock-in*



# Hammerspace

## *Orchestrates Dynamic Data Movement*





# Hammerspace

*Global Data Platform for File and Object Data*



## Hammerspace Global Data Platform

Data Creation  
And Ingest



GPU Computing /  
Data Processing



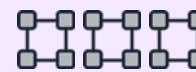
On-Premises  
Archive



Users That Work  
With The Data



Cloud CPU or  
GPU Resources



Remote  
Disaster Recovery  
& Archive



NAS 1



Fast Storage



NAS 2



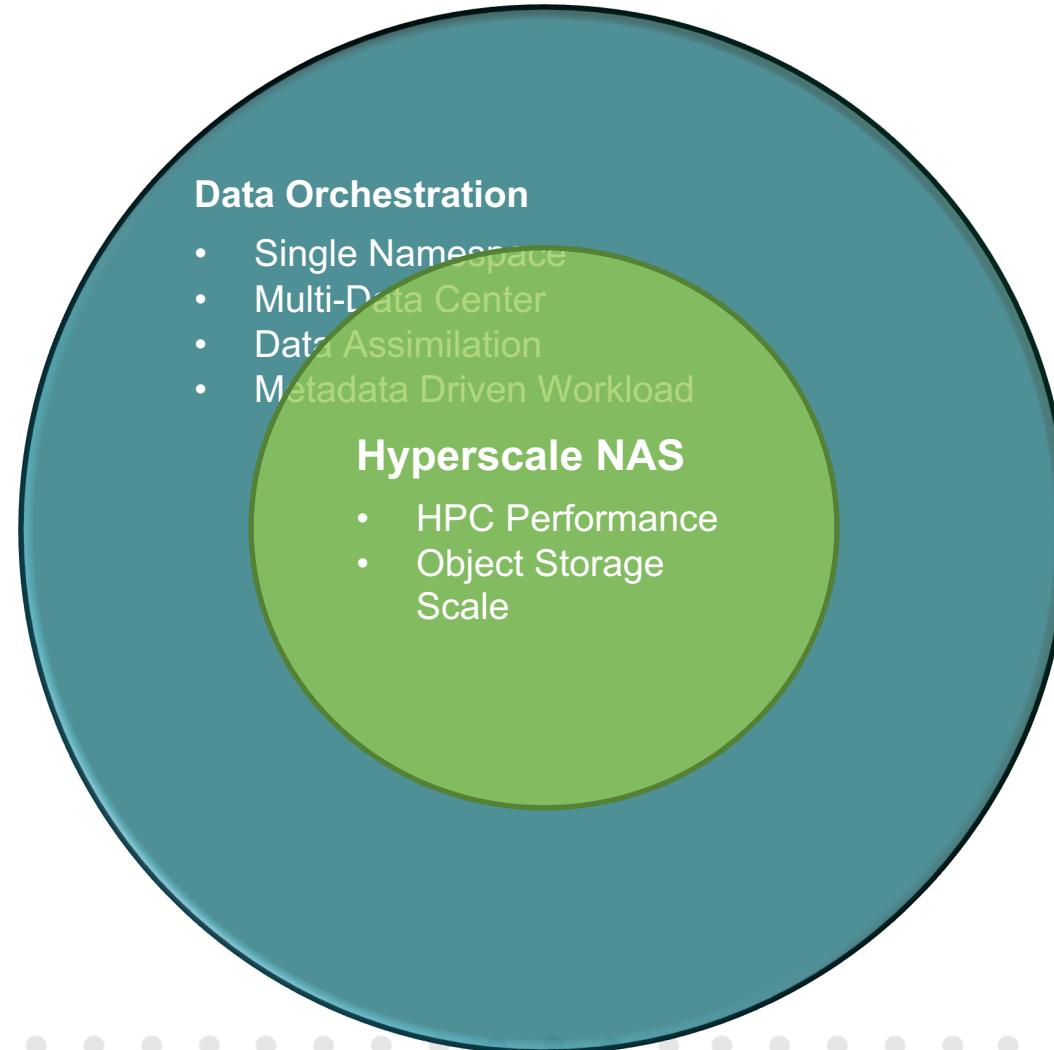
Cloud Storage



Cloud Storage

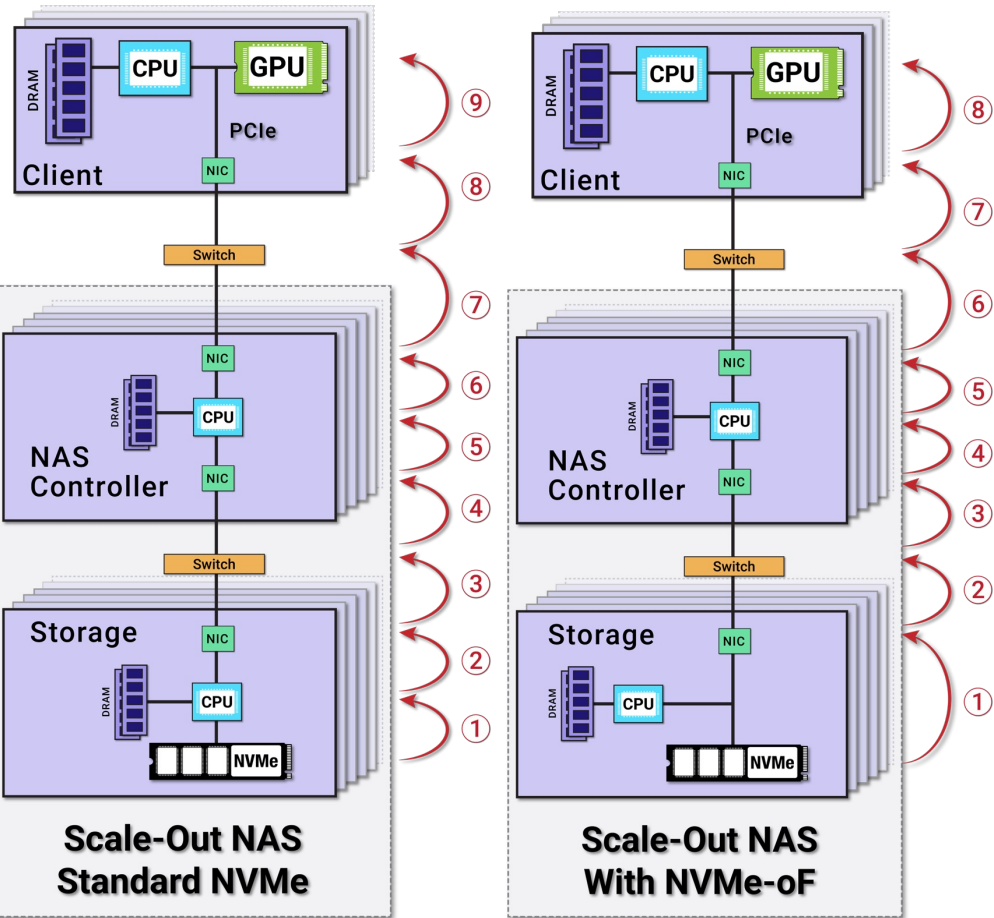
HPC QUANTUM DATA AI

# Hammerspace Superpowers

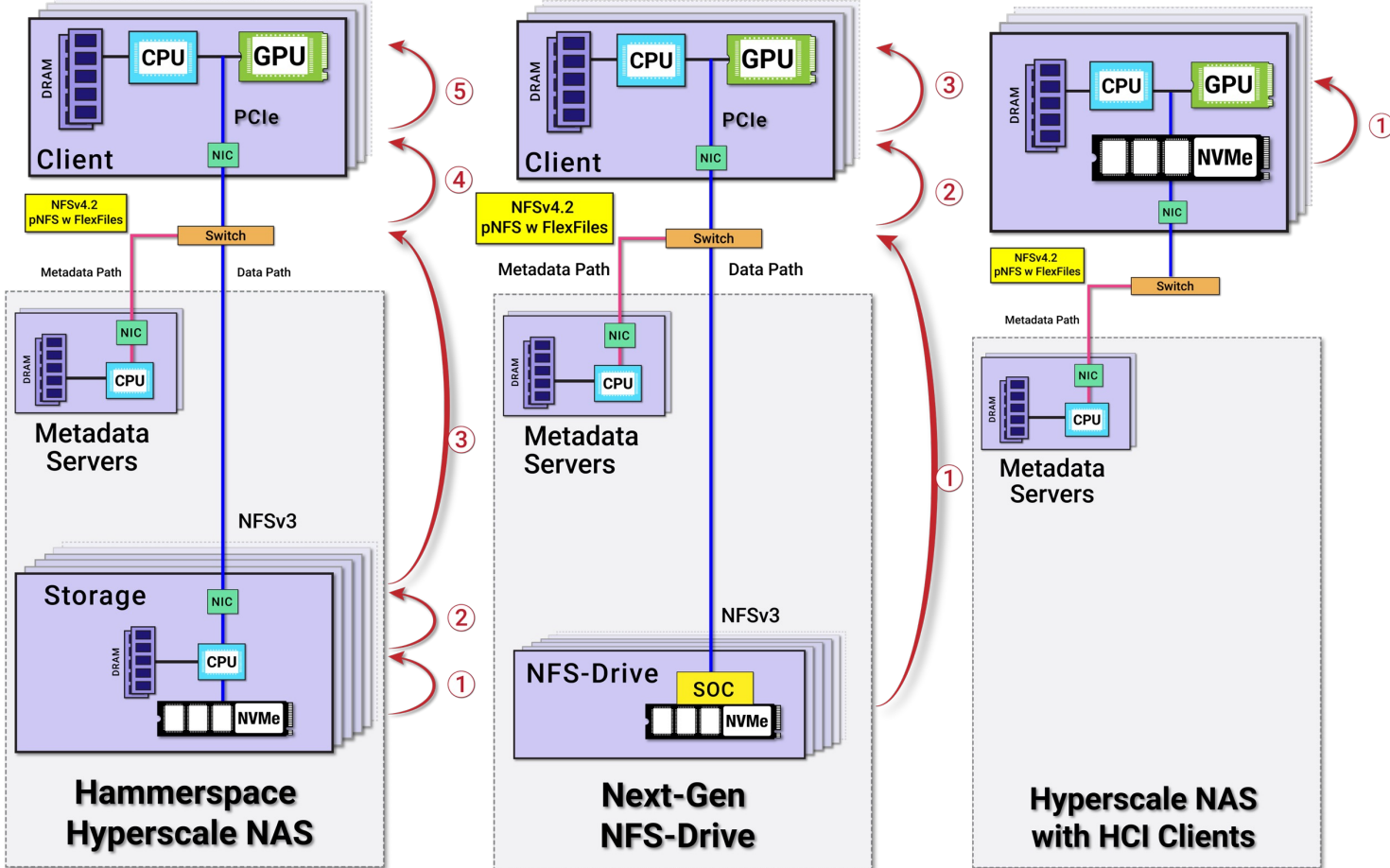


# Future Advancements Will Continue to Drive File and Object Storage Infrastructure Efficiency and Performance

## 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





**Contact**

**Matthew Shaxted, CEO**  
[shaxted@parallelworks.com](mailto:shaxted@parallelworks.com)

# Challenges for the Modern Computing Organization

Technology diversity & capacity constraints necessitate agile task placement



**Rapidly evolving technology** (cpu, gpu, fpga, neural, quantum) requires agility to get first mover advantage.

When on-premises resources are constrained or in maintenance, seamless **access to diverse computing sites is essential.**

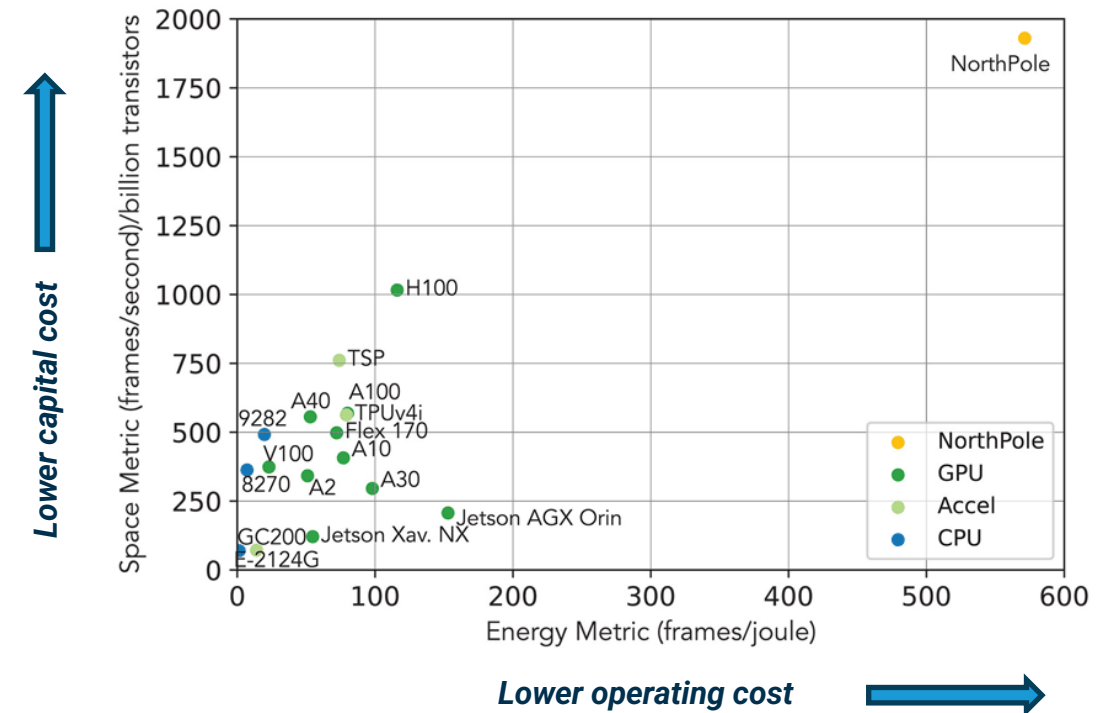
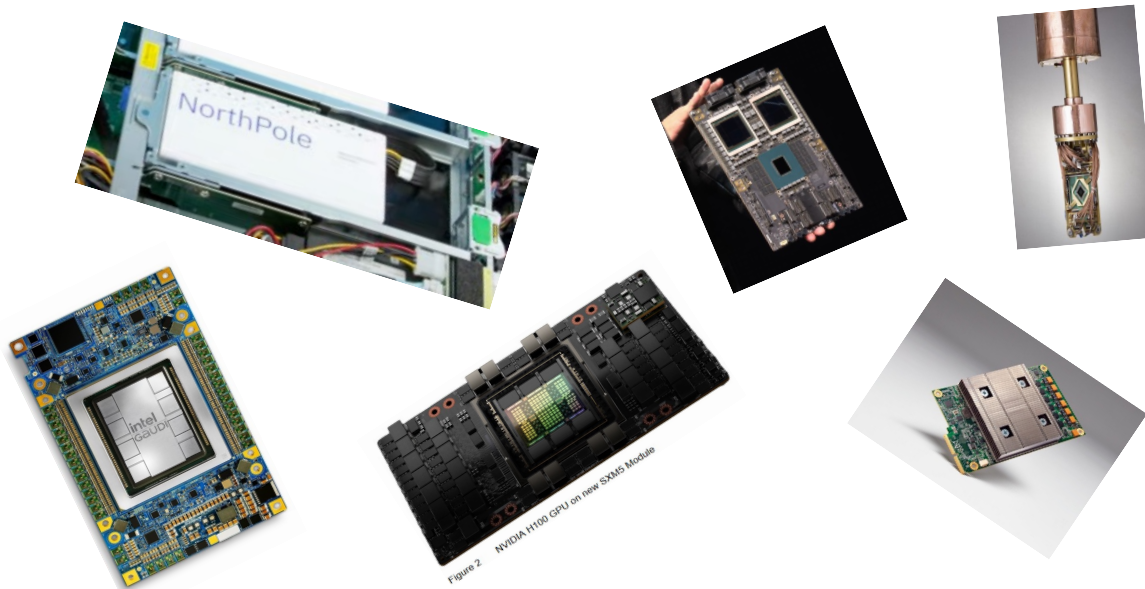


Chart from Carlos Ortega-Otero IBM NorthPole:  
[https://www.youtube.com/watch?v=7s1M09z\\_qI8](https://www.youtube.com/watch?v=7s1M09z_qI8)



# Parallel Works ACTIVATE



HPC & AI control plane for R&D teams who rely on compute-intensive applications and AI models. Easily and consistently provision, manage, and share on-premises, cloud and hybrid high-performance compute resources at scale for simulation, analytics and AI.

The image displays three overlapping screenshots of the Parallel Works web interface. The central screenshot shows the 'Home' dashboard with a sidebar menu and a 'My Compute Resources' section listing various cloud providers and instance types. The top-left screenshot shows the 'Access' section with a grid of application icons. The top-right screenshot shows the 'Workflow' section with a table of active jobs and their status. The bottom-left screenshot shows a 3D visualization of a simulation result. The bottom-right screenshot shows a configuration page for a workflow.

**Access**

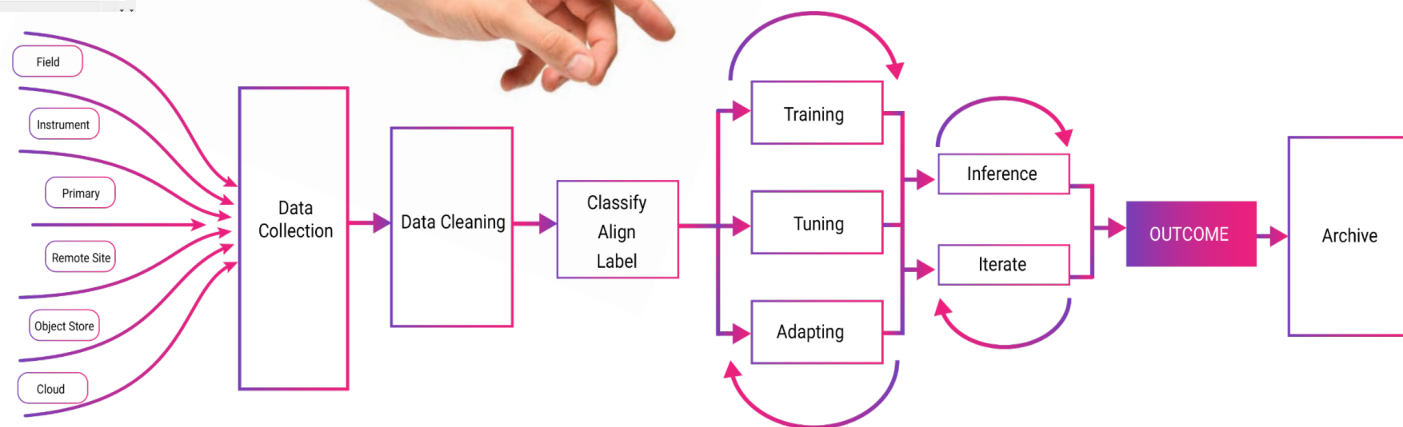
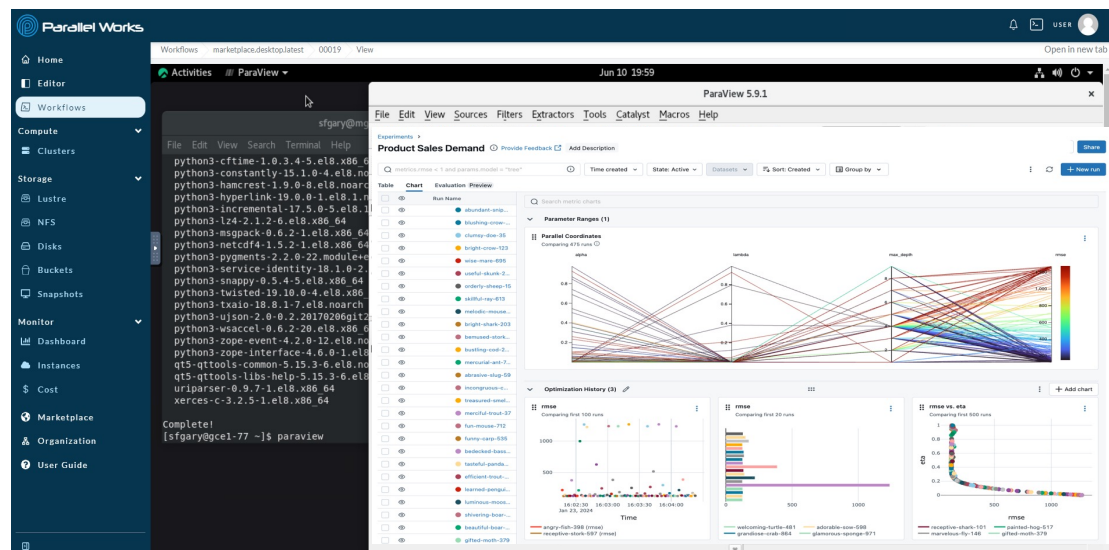
**Workflow**

**Compute Resources**

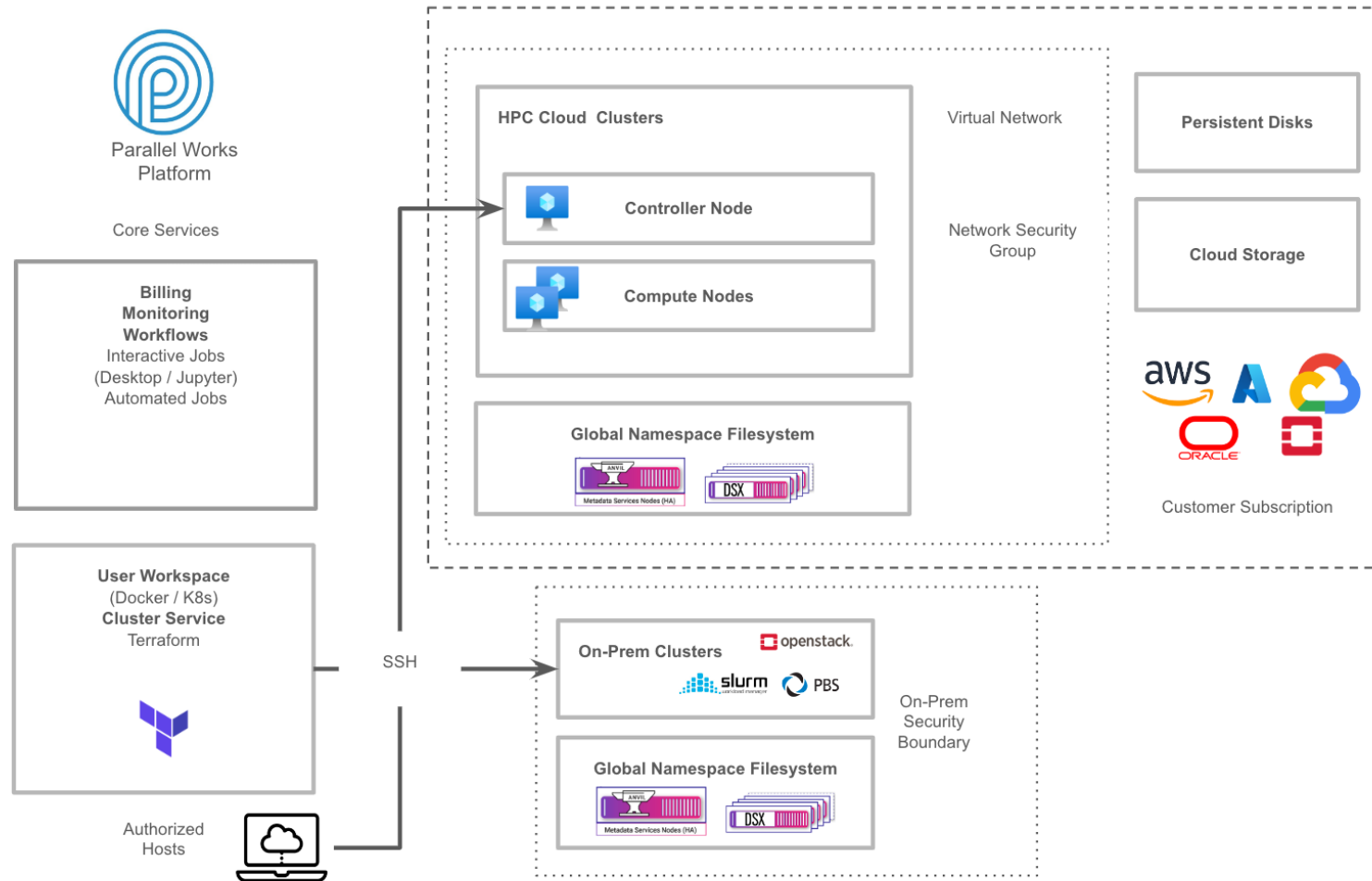
Provider	Group	Name	Description	Instance Type
AWS	awsdemo	group-aws-contrib	description: AWS SLURM cluster	
AWS	aws2	group-aws-contrib	description: AWS SLURM cluster	
AWS	aws2	group-aws-contrib	description: AWS SLURM cluster	
Azure	azure2	group-azure-contrib	description: Azure HPC Cluster	
Google	google2	group-google-contrib	description: GCP H2	
Intel	intel-medium	instance	description: Intel Developer Cloud - Medium Instance	
Intel	intel-small	instance	description: Intel Developer Cloud - Small Instance	
Junos	noaa	cluster	description: On-Premise Cluster	
Hinc	koeir	cluster	description: Koehr DRC Cluster	
Hinc	onyx	cluster	description: Onyx DRC Cluster	
Podmtz	cluster	description: Penguin MT2 Slurm Cluster		
Podmtz	cluster	description: Penguin MT2 Slurm Cluster		

© 2024 Parallel Works, Inc. - v4.104

# R&D, Task Steering & Control to Optimal (or Available) Environments

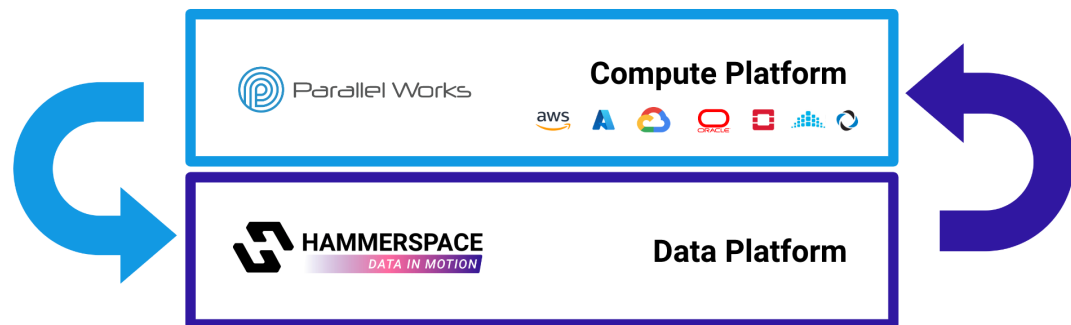


# Connection to / Assembly of Uniform Compute Clusters for Application Portability





# Fit Storage Requirements Precisely to Workloads



**Add Storage** Add Storage Cancel

Resource name  
New storage name; this must be unique, between 2-255 characters, and use only lowercase letters and numbers.

Description  
Limit 100 characters

Tags  
Make your resource easy to find. Separate tags with commas.  
e.g linux, ubuntu, windows, centos...

Select storage type

☐ AWS Elastic File System

☐ Google Filestore

☐ Azure Files

☒ HAMMERSPACE

☐ Load from marketplace: use an item from the marketplace to initialize this resource.

Parallel Works

Home  
Editor  
Workflows

Compute  
Clusters

Storage  
Lustre  
NFS  
Disks  
Buckets  
Snapshots

Monitor  
Marketplace  
Organization  
Admin  
User Guide

Storage / Hammerspace Azure Global Namespace *hammerspace*

Sessions Definition JSON Properties Sharing

General Settings

Cloud Infrastructure \* pw-canary-eastus

Group \* pw-azure

Hammerspace Options

Region \* eastus

Admin Username \* admin

Admin Password \* .....

Use High Availability \* No

Storage Type \* Standard\_LRS

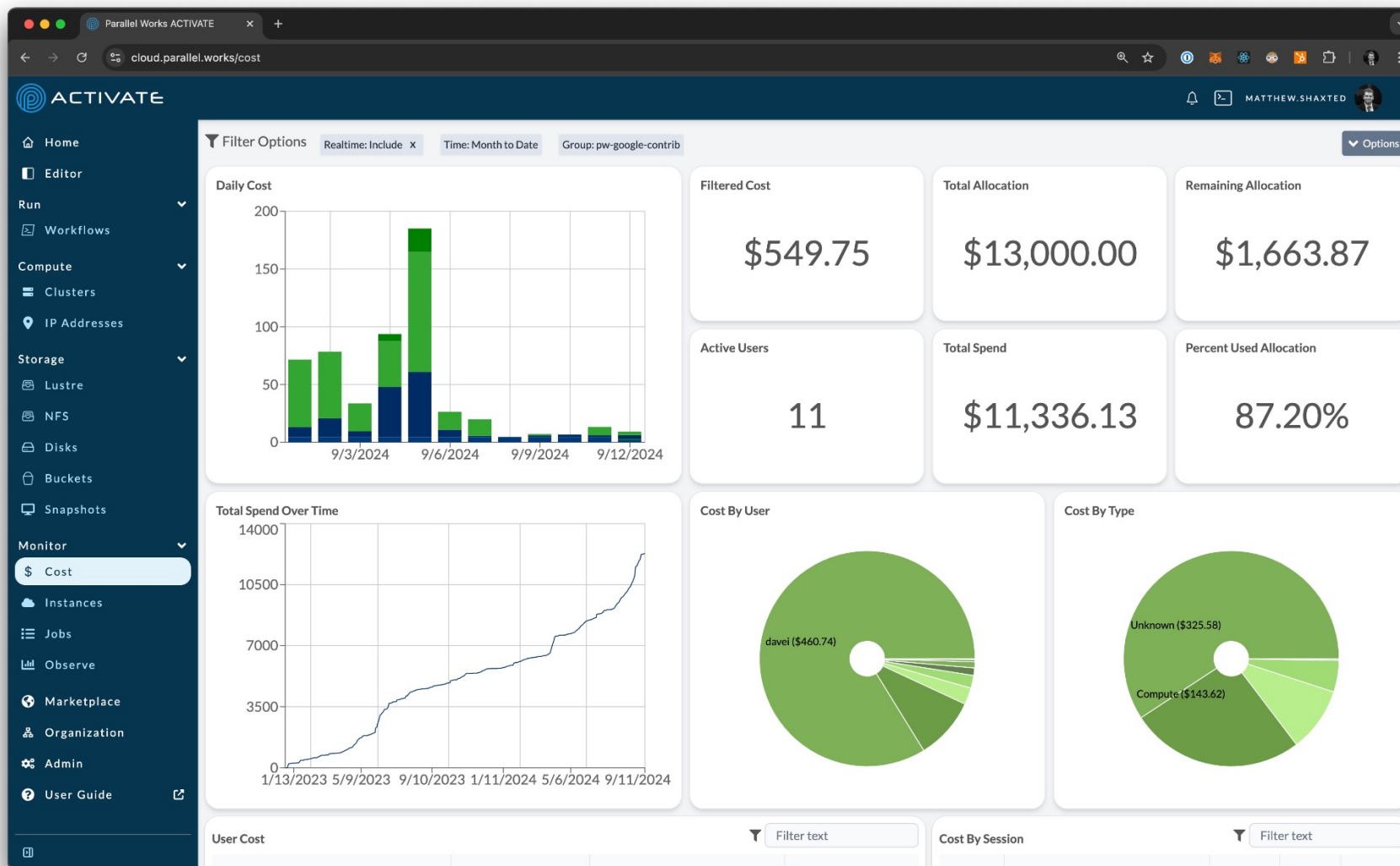
Anvil Instance Type \* Standard\_F8s\_v2 (8 vCPUs, 16 GB Memory, amd64, 64 GB Local Disk)  
[See all sizes](#)

Metadata Disk Size (GB) \* 256

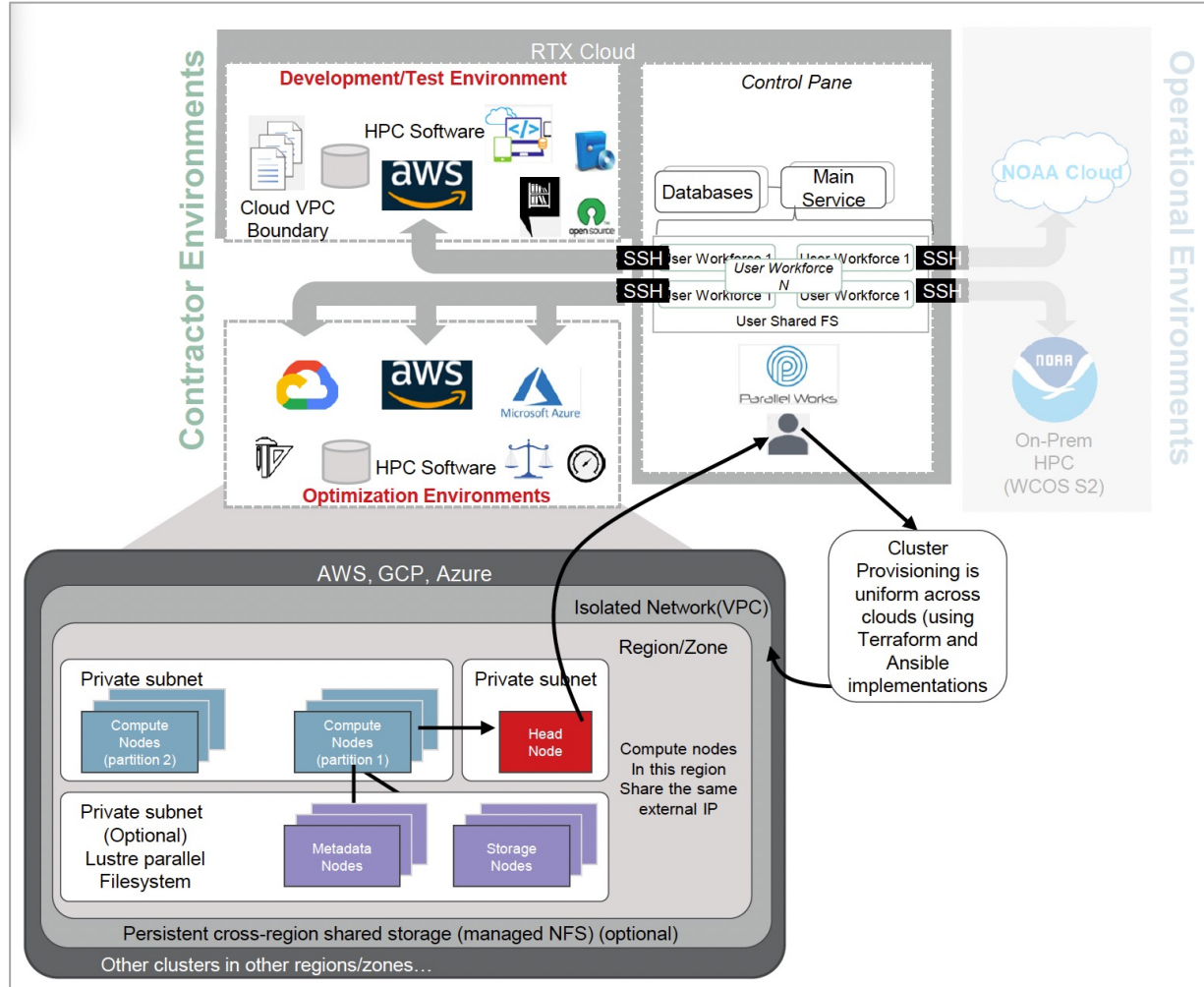
Dsx Instance Type \* Standard\_F8s\_v2 (8 vCPUs, 16 GB Memory, amd64, 64 GB Local Disk)  
[See all sizes](#)

Data Disk Size (GB) \* 256

# ~3 min Cost Control in Cloud Environments: FinOps for HPC



# Streamlining Research to Operations



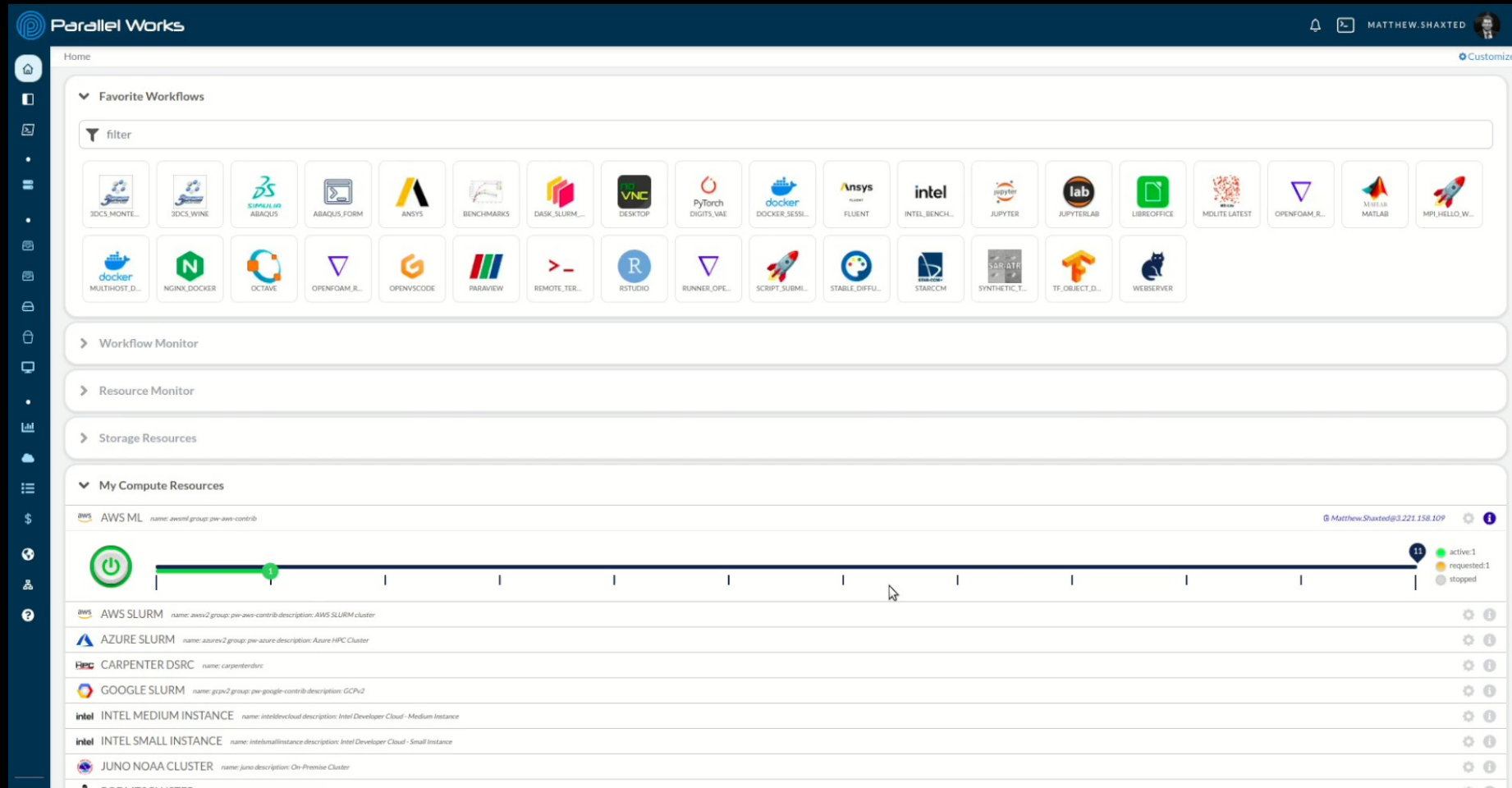
**Problem:** Customer experienced a high barrier when moving production workloads from local dev environments to operational systems.

**Solution:** Parallel Works helped customer to spin up a cloud environment that closely mirrored the current on-premise operational environment (WCOS S2) to act as a testbed for operational workloads.

**Key Results:** The cloned cloud environment empowered practitioners to migrate production workloads to operational systems 3x faster than a local dev method, streamlining research to operations.



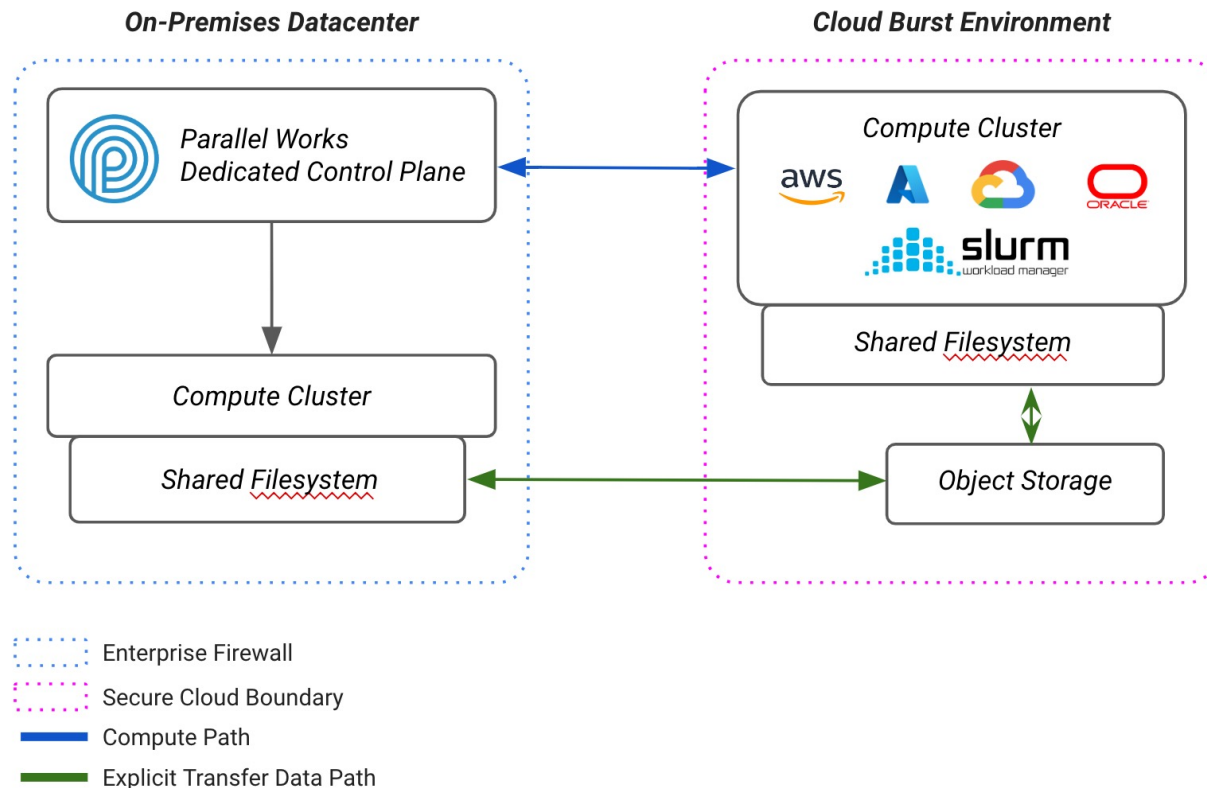
# Lowering Friction for End Users To Do Their Work



Execute complex ML/GenAI stacks via turnkey webforms on cloud or on-premise GPU resources (e.g., Stable Diffusion).

# Business-as-Usual for Burst Computing

## Explicit Multi-Hop Transfers to Object Storage



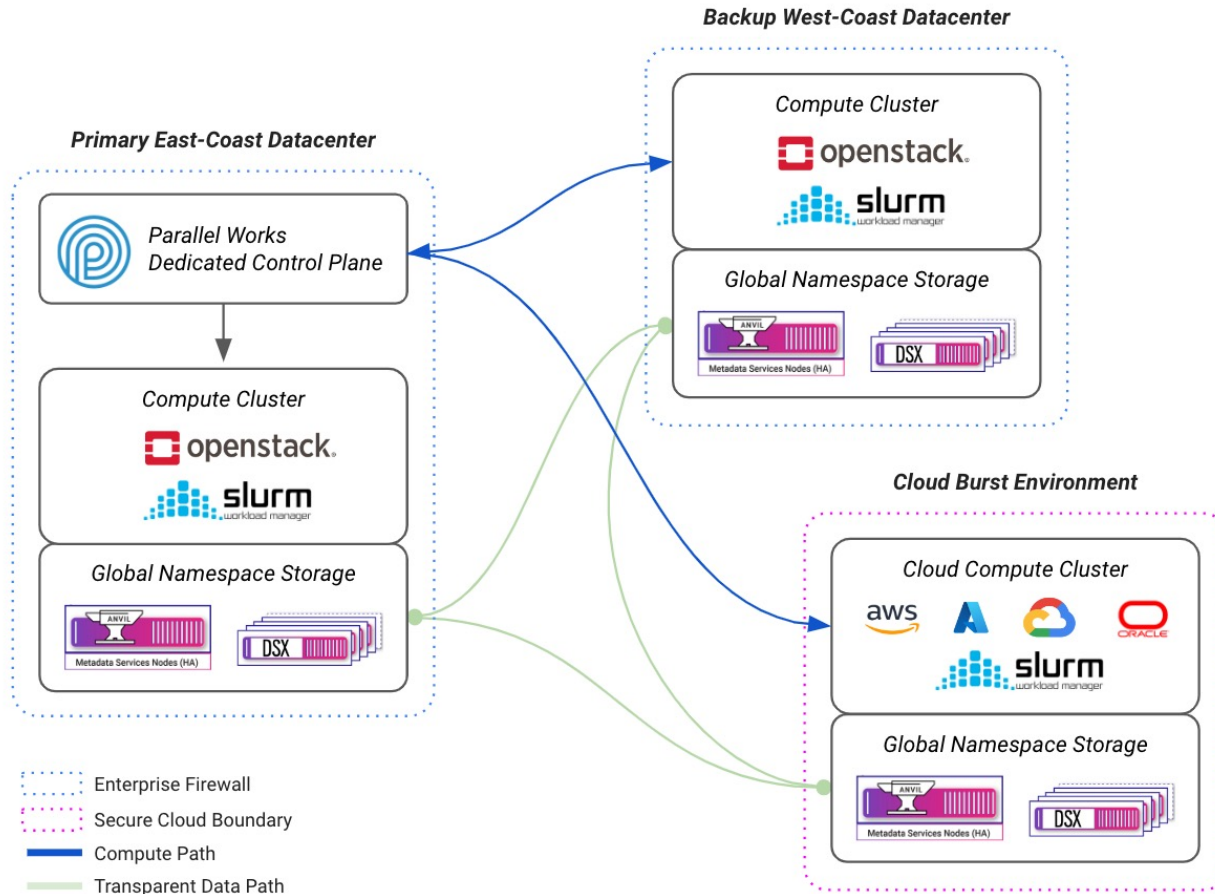
**Different:** running in cloud takes a different file availability approach as on-prem systems, adding to multi-site friction.

**Difficult:** requires knowledge and understanding of CSP-provided CLI tools.

**Slow:** multiple hops required to/from on-prem filesystem, object storage and cloud shared filesystem.

# Converged Burst Computing

## Transparent Data Availability at Any Site



**Familiar:** data presented exact same way as on-prem filesystems, lowering cognitive load for end users.

**Portable:** allows workloads to easily scale between all sites (cloud or on-prem).

**Scalable:** adding new sites or cloud keeps the same topology without having to learn new data movers.



# Right-Sizing of HPC & AI Infrastructure



- Right-sizing is the process of matching workloads to optimal resource shapes to reduce TCO.
- Recent customer study looked at a \$30m cloud program budget w/ multiple FS performance tiers in comparison to traditional full on-prem hardware refresh.
- Showed that using a cloud mix could deliver competitive TCO for a portion of workloads, but not a full 1:1 mapping.
- Cloud benefits such as supply chain agility for new large system and tech updates throughout the program lifecycle factor into the “more is better” tradeoff.
- Additionally global namespace filesystem as opposed to Managed Services resulted in >25% additional compute and storage.



**HPC + AI**  
**WALL STREET**

2024

**Thank you!**

[david.flynn@hammerspace.com](mailto:david.flynn@hammerspace.com)

[shaxted@parallelworks.com](mailto:shaxted@parallelworks.com)

**HPC** **QUANTUM** **DATA** **AI**



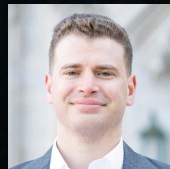
# Orchestrating the Future: Seamless Workload and Data Management in Distributed HPC Environments



**Moderator:**  
Alex Woodie  
*Managing Editor, Datanami*



David Flynn  
*Co-founder and CEO,  
Hammerspace*



Matthew Shaxted  
*CEO,  
Parallel Works*