Storage for a New Generation of AI/ML

Presented by

Somnath Roy – Principle Engineer – Samsung Memory Solutions Lab

Current State of AI/ML

- **Focus on Large-Scale AI/ML (at least >1PB storage for training data)**
  - Large-Scale Use cases:
    - Fraud prevention and risk analysis
    - Natural Language Processing
    - Real-time price optimization
    - Autonomous driving
- **Compute has evolved rapidly with new algorithms and GPUs**
  - In fact with the advent of GPU direct, NVDIA is claiming bottleneck is on storage
- **Can large-scale storage performance keep up with compute?**
  - High read BW requirement (>1TB/s per rack) for running AI training at scale with thousands of GPUs in parallel

DSS: Performant & Scalable Object Storage

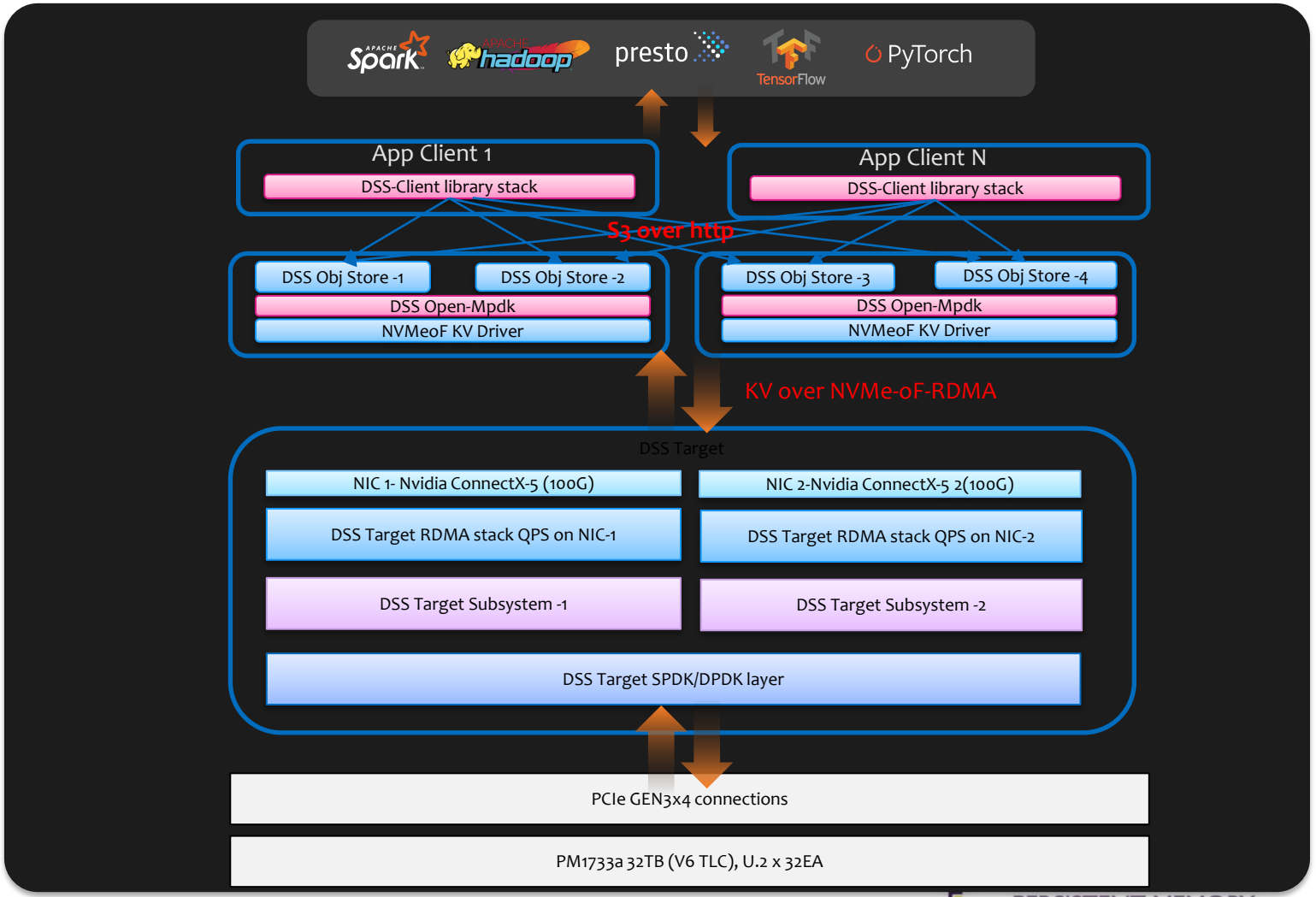
Disaggregated Storage Solution(DSS)

Services

- Samsung developed – open sourced
  https://github.com/OpenMPDK/DSS
- NVMeoF based S3 Service
- High Read Throughput Object Storage
- Disaggregated Storage and compute
- Shared everything architecture
- Zero copy key-value transfer
- Easy Scaling at Exabytes

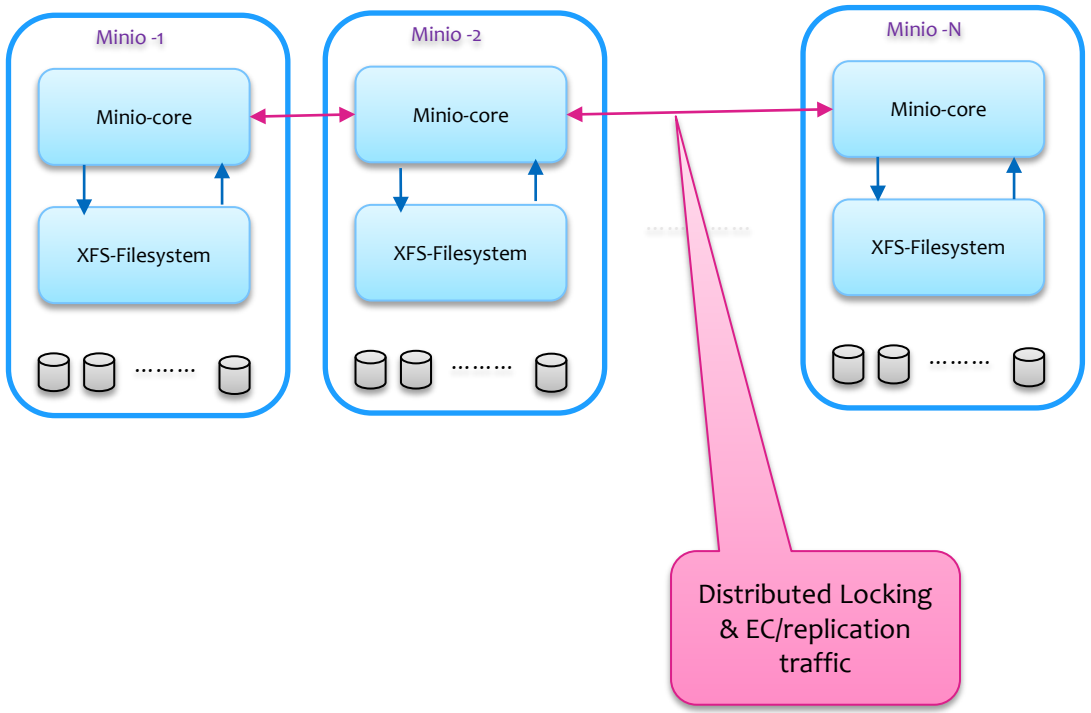
Use Cases

- Large scale high READ throughput AI training
- Image Analytics
- Audio/Video AI
- Metaverse

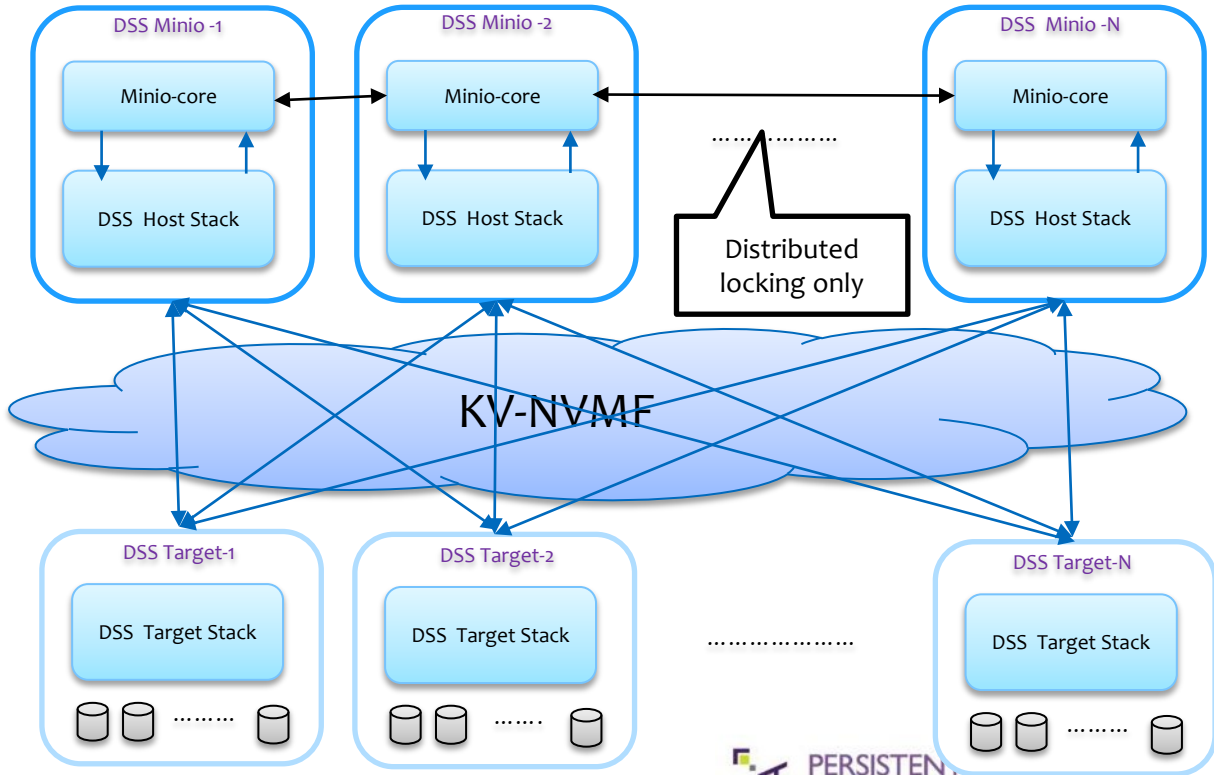


DSS Enhanced Minio Object-Store

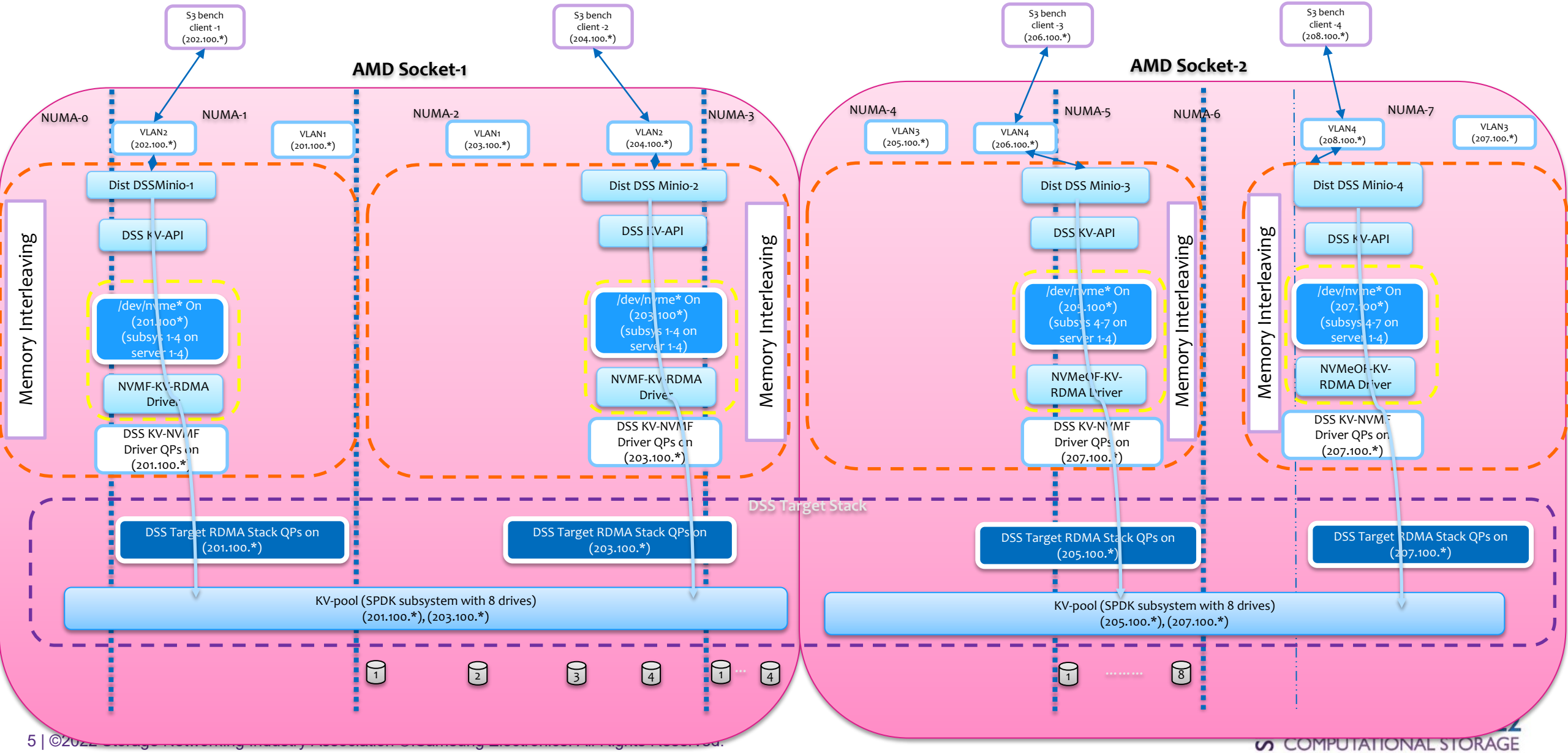
Stock Minio Shared-nothing architecture (Compute has to grow along with storage)



DSS Minio disaggregated, Shared-everything architecture (Compute and storage can grow independently)

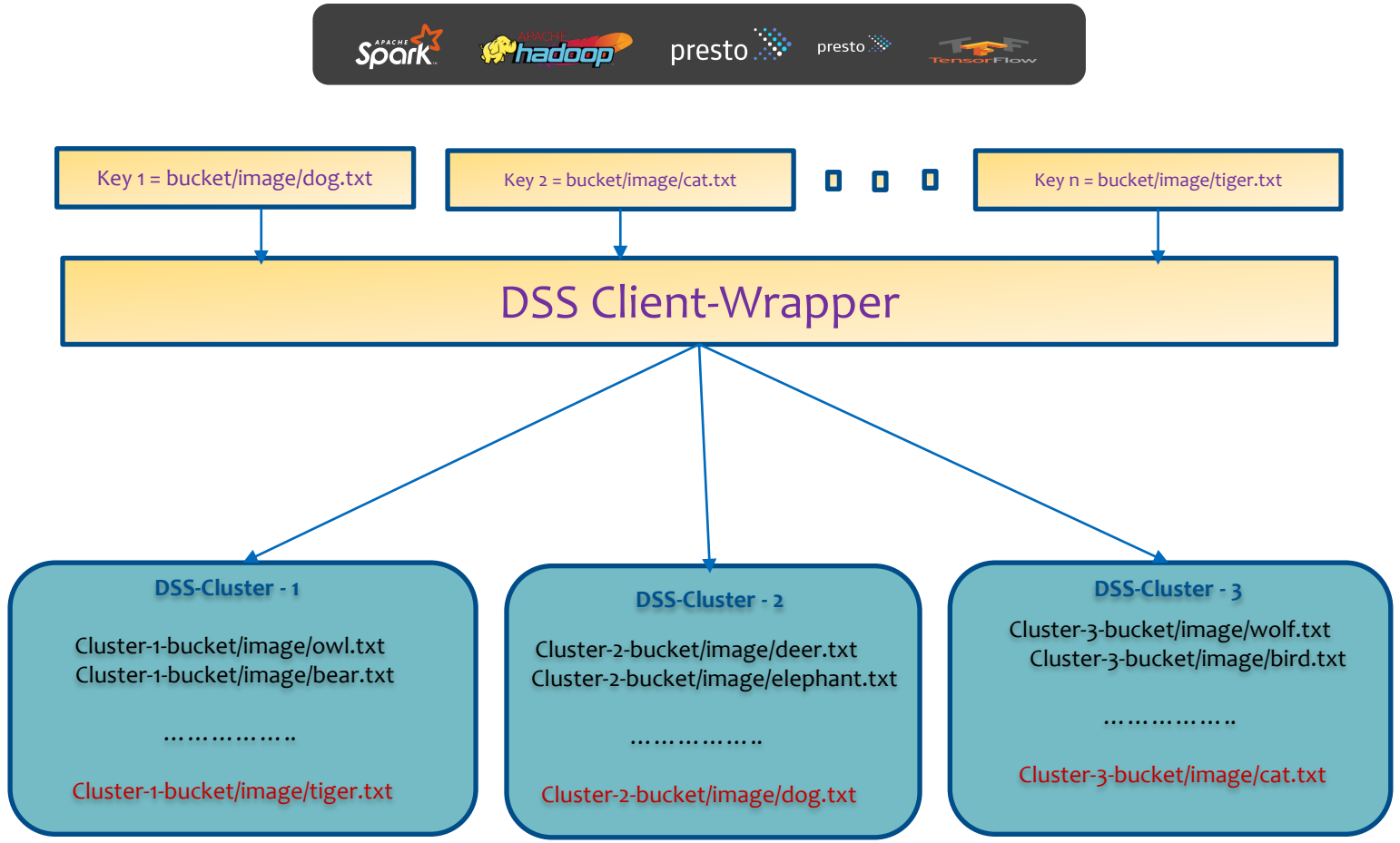


Reference Minio + DSS deployment model for AMD

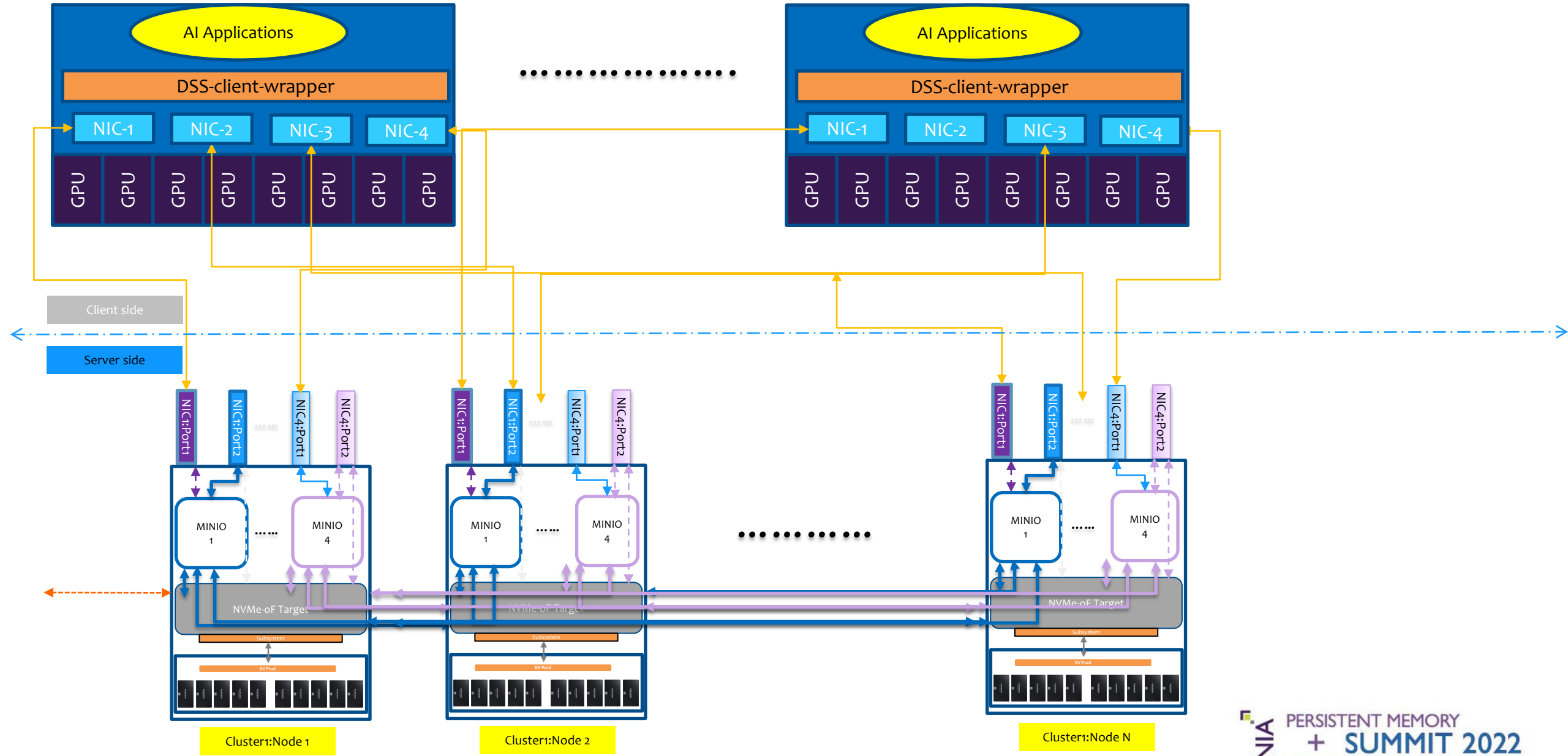


DSS Client Wrapper

- Bucket Abstraction
- Key Distribution
- Cluster Expansion
- Rebalance
- Standard S3 Operations (Get/Put etc.)



DSS Deployment View

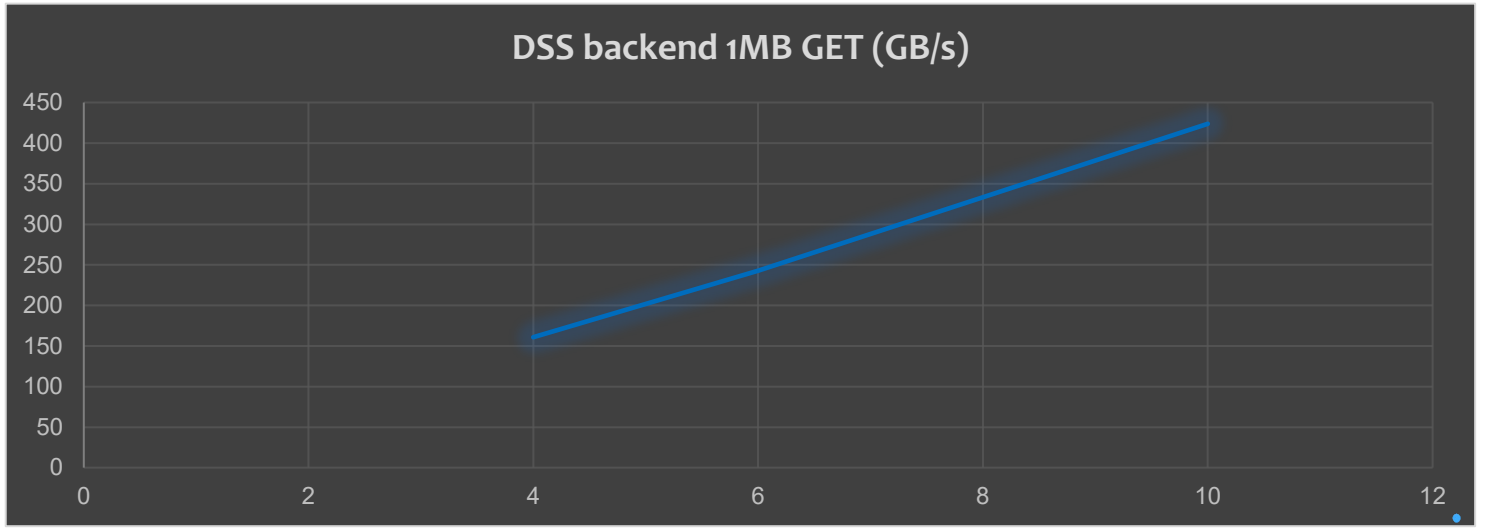


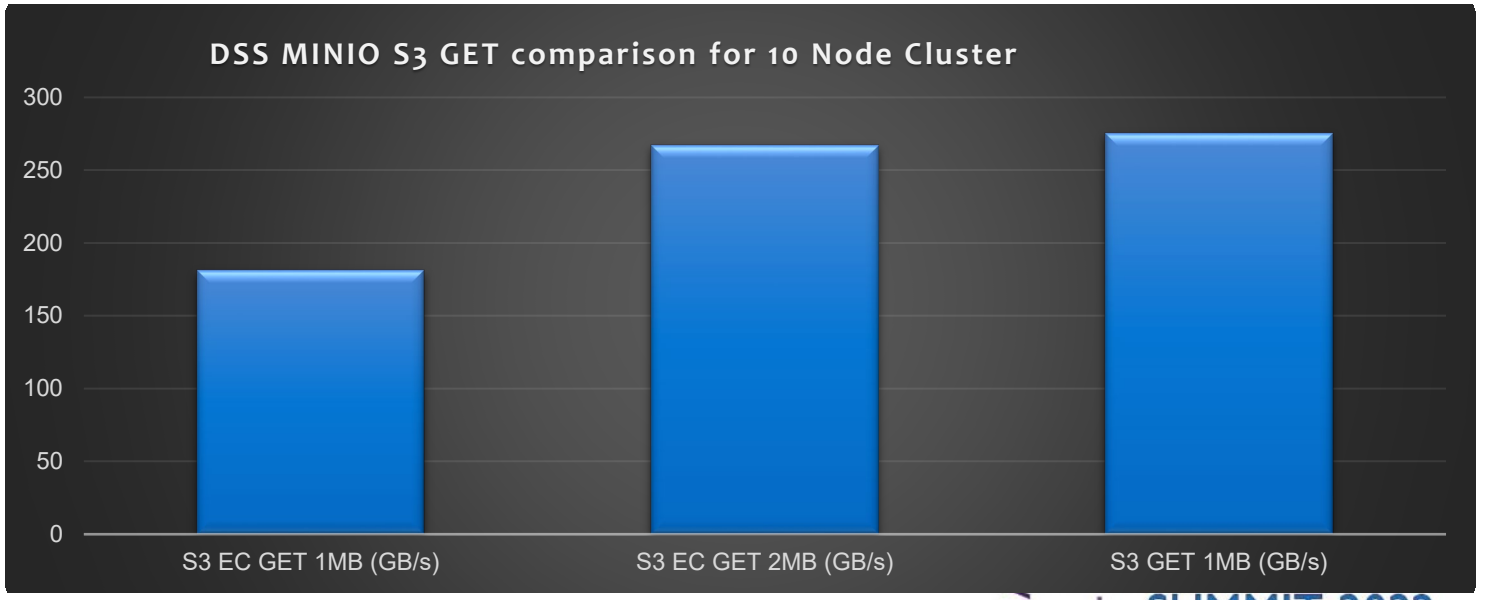
DSS GET Performance

Setup:

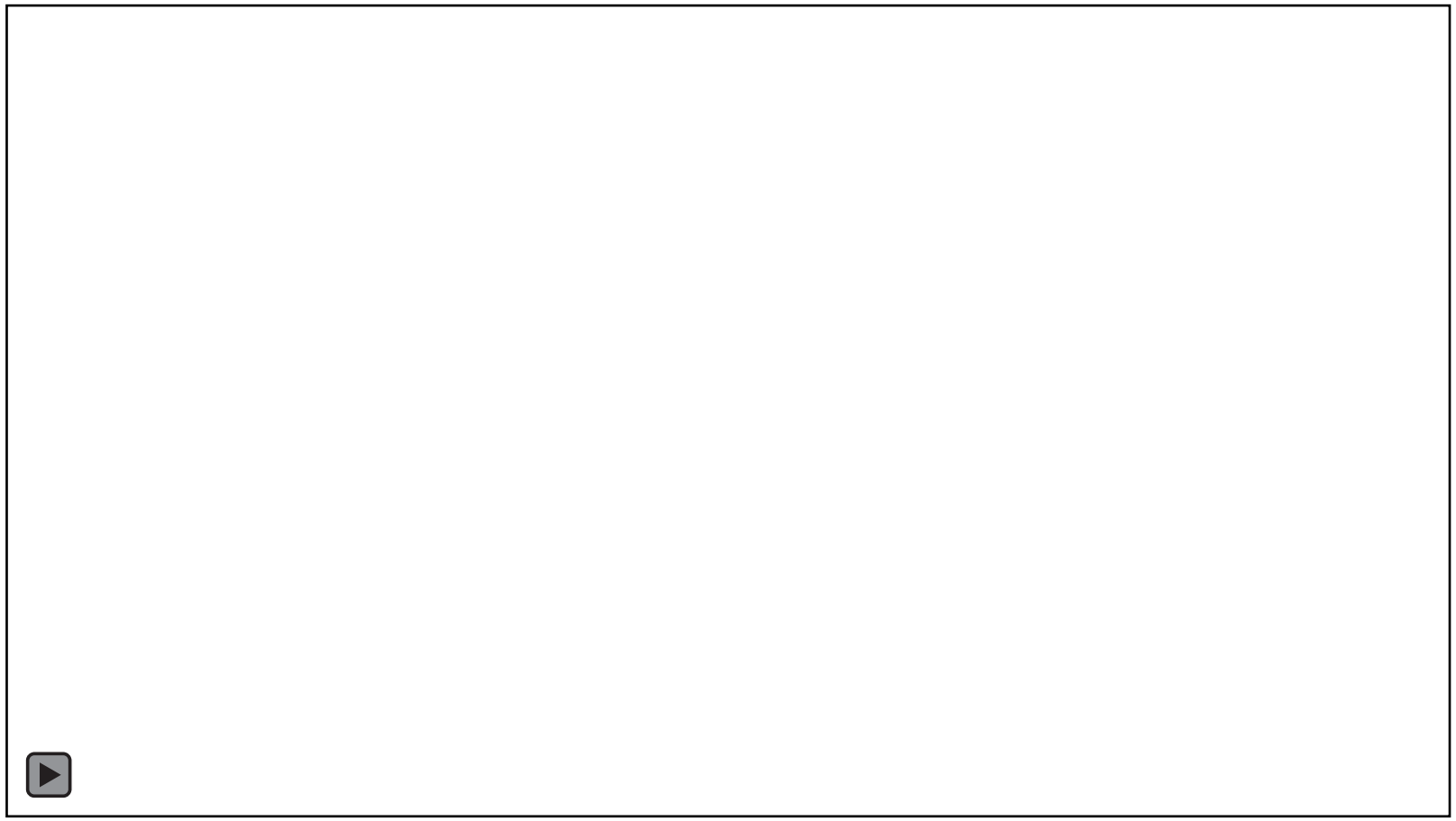
- Client - 16x Dell PowerEdge R6525, 2 x DGX A100
- DSS S3 Server
  - 10x Dell PowerEdge R7525 Gen4 servers
  - Dual socket AMD EPYC 7742 64-Core
  - 1TB physical memory
  - 4xMellanox Dual port 100/200Gb (ConnectX-6)

- SSD - 16x PM1733 4TB Gen4 NVMe SSD per DSS S3 server
- Total data set generated during test ~400TB
- Top chart is just DSS backend performance across 10 node, no S3 involved
- Tool used home grown dss test cli
- Bottom one with DSS optimized Minio
- Tool used standard S3-benchmark





AI Benchmarking Tool

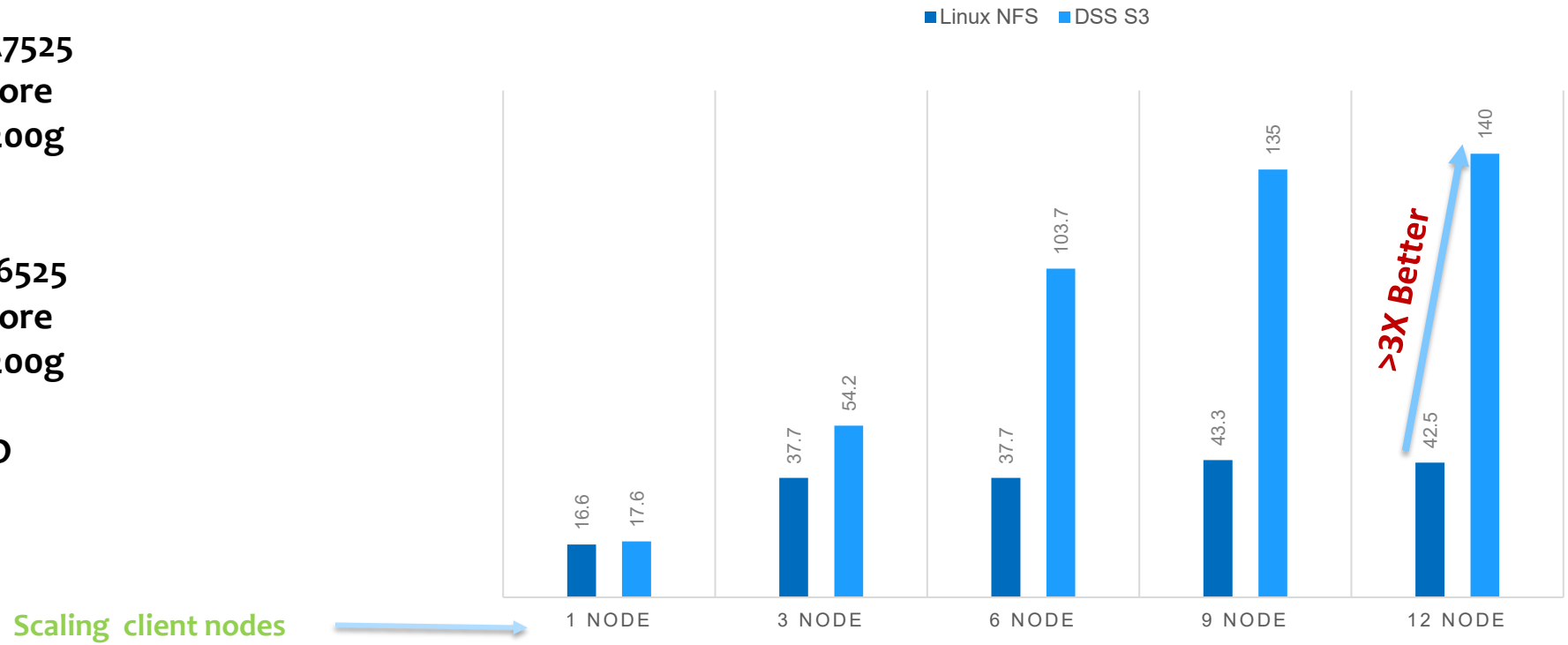


- Benchmarking various storage solution based on NFS, S3 at AI training level
- Platform where developers can add their ML framework, custom data set, training method, models and storage backend
- Demo is showing a custom training with a custom data set that is only capturing data load time and BW from storage servers on NFS/S3

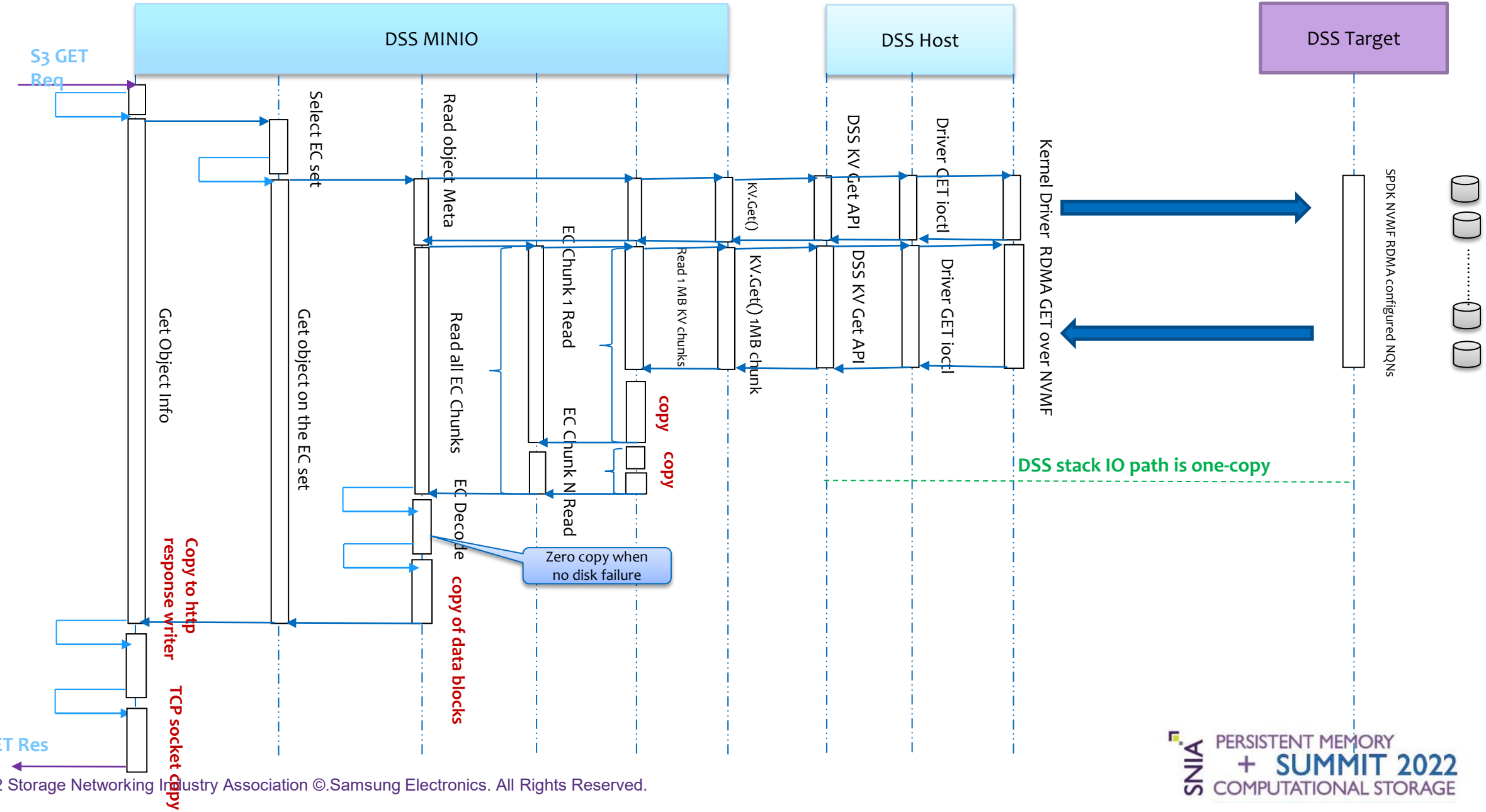
DSS S3 vs Standard NFS

Setup:

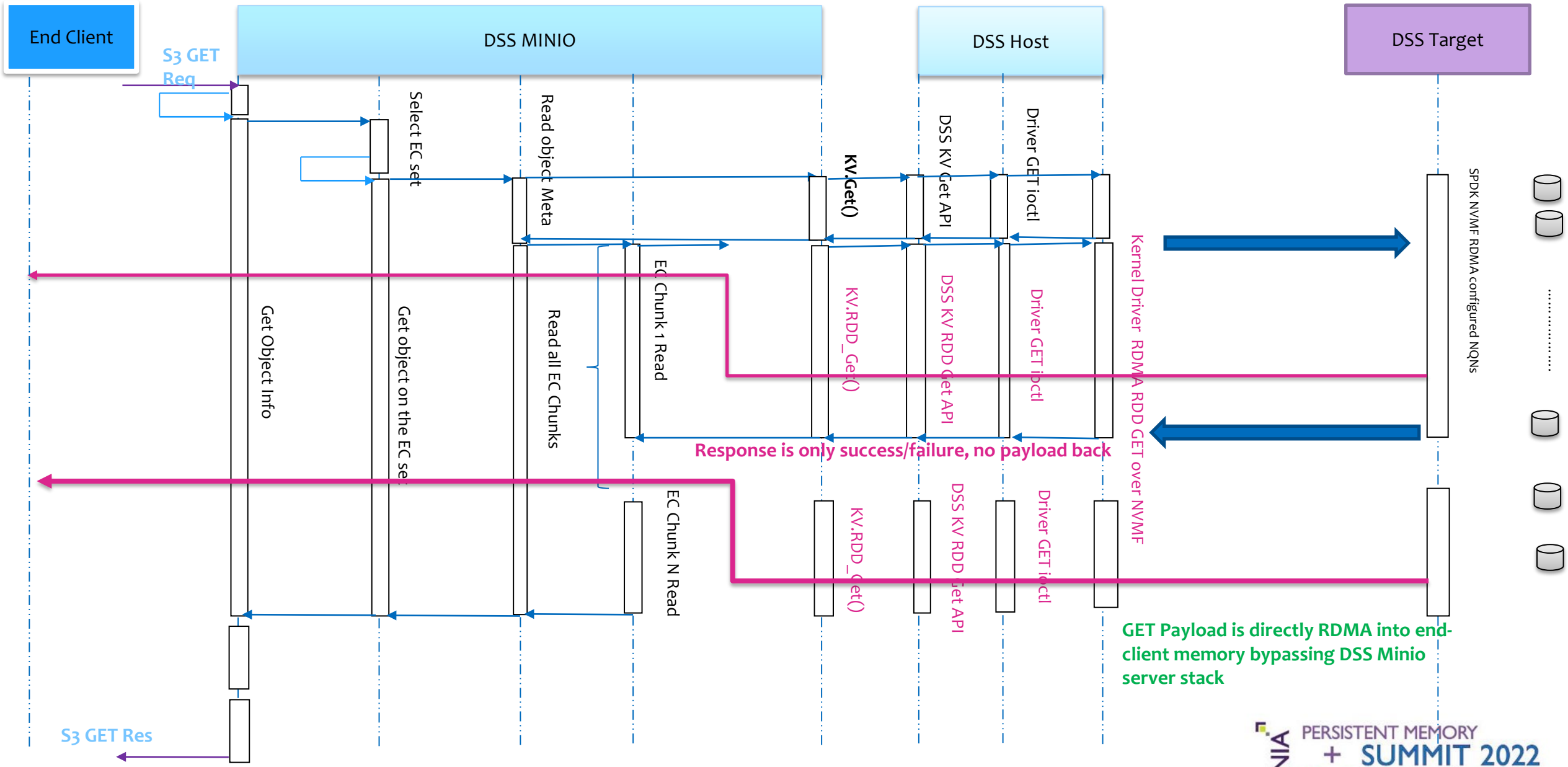
- Client – 12xDell PowerEdge 740xd
- DSS S3 Server
  - 6x Dell PowerEdge R7525
  - AMD EPYC 7742 64-Core
  - Mellanox Dual port 200g (ConnectX-6)
- NFS server –
  - 6xDell PowerEdge R6525
  - AMD EPYC 7742 64-Core
  - Mellanox Dual port 200g (ConnectX-6)
- SSD - PM1733 4TB NVMe SSD



IO flow during S3 GET request



IO flow during S3 GET request for next Gen DSS



DSS Availability

Open Source Announcement

- https://github.com/OpenMPDK/DSS
  - https://github.com/OpenMPDK/dss-sdk
  - https://github.com/OpenMPDK/dss-ansible
  - https://github.com/OpenMPDK/dss-minio
  - https://github.com/OpenMPDK/dss-ecosystem

Complete Ecosystem

- AI Benchmarking Framework supporting user preferred training and models
- Client Wrappers supporting Pytorch and Tensorflow
- Host and Target Stack

Thank You

(som.roy@samsung.com)

Please take a moment to rate this session.

Your feedback is important to us.

