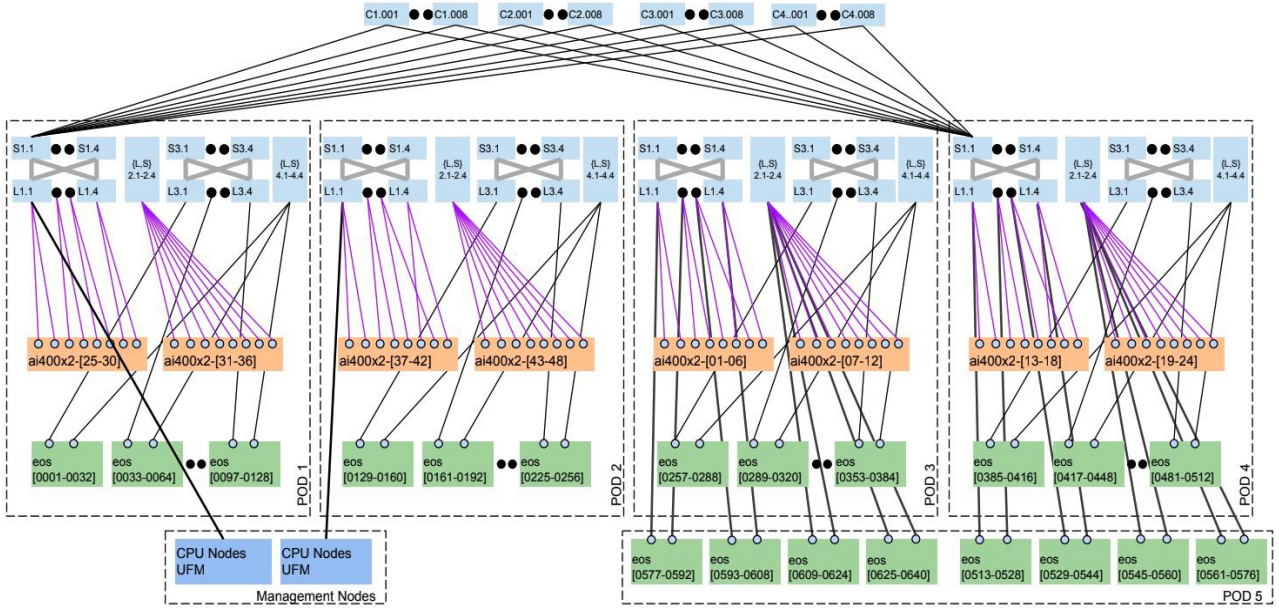
JLUG 2023

JLUG 2023: Lustre Today

- Continuing Leadership for High-End HPC Clusters (e.g. Top 100)
- Most Broadly User Solution for HPC Production Systems
- On-Demand Solutions Available from Most Public Cloud Providers
- Expansion in the Generative AI Infrastructure & LLM Training Market
- Rapid Adoption by Specialized GPU Cloud Service Providers
- Capacity *and* Performance for Specialized Inference Workloads

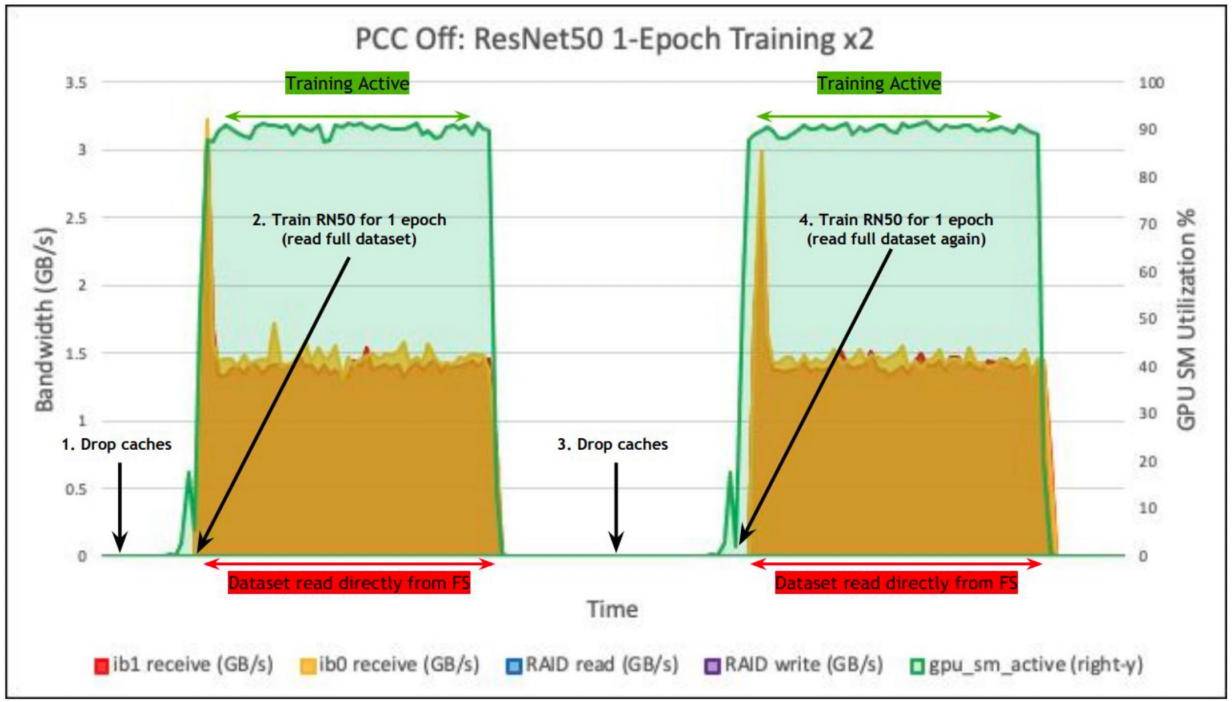
Example: NVIDIA EOS Supercomputer

- Announced 2022
- 4608 x H100 GPUs
- 1 x NDR 400 Gbps port per GPU
- Separate NDR Network for Storage
- Lustre File System with 192 OSSs
- 2 TB/sec Read *and* Write BW

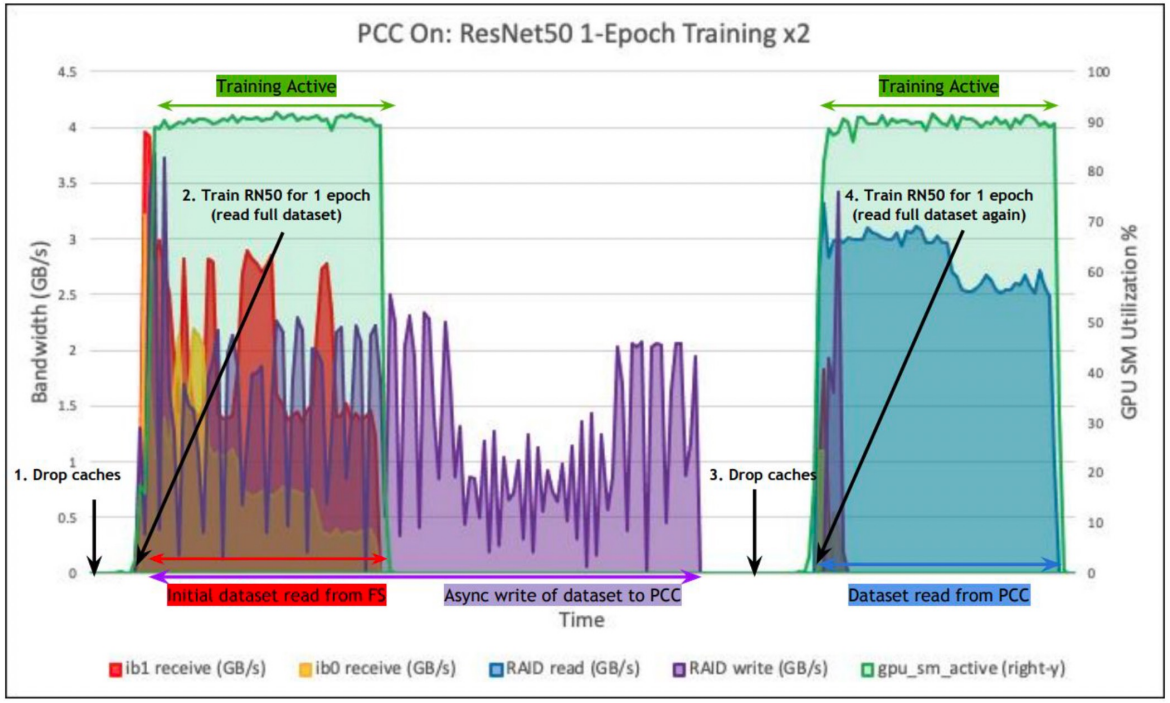


Persistent Client Cache for DL Training

Without PCC



With PCC



Future Focus Areas

- 10s of Millions of Metadata Operations
- High-performance Single-treated I/O
- Kerberos Support Across the Entire Datacenter
- Improver Resilience and Recovery
- Management and Automation for Large-scale Data Centers
- Continuing Broad Platform Support (e.g. RISC-5)

Thanks to our Sponsors!

NVIDIA

The NVIDIA DGX™ H100 system features eight NVIDIA GPUs and two Intel® Xeon® Scalable Processors

ddn

ScaleWorX

GIGABYTE™

SCSK

FUJITSU

Prunus-Solutions

株式会社 コスモ

Lenovo