PER3S 2024

AI FACTORIES: CONVERGENCE OR COMPLEMENT?

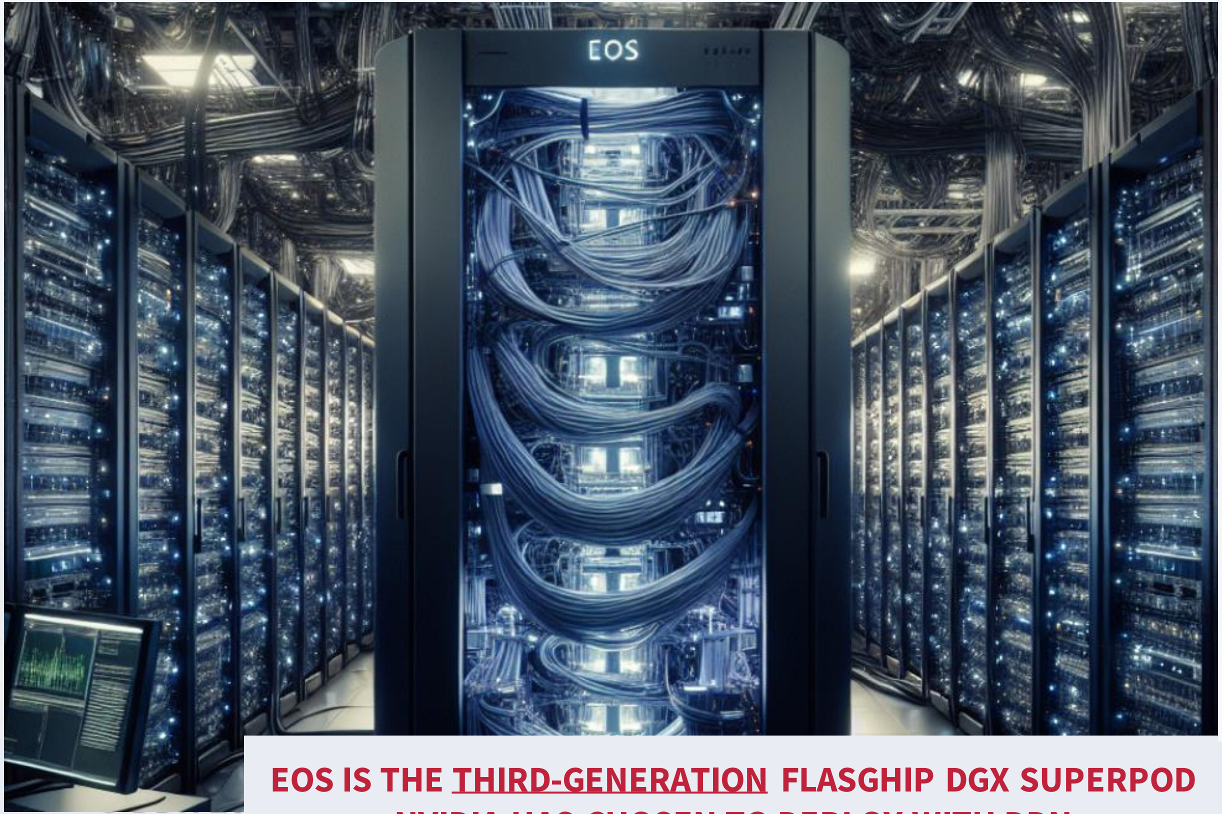
NCP and duality of use

Jtacquaviva@ddn.com

May 28th, 2024

ddn

EOS NVIDIA Flagship System



576 DXH: 4608 H100 GPU

NDR400 IB Compute and Storage

Storage:

48 AI400NVX2

EXAScaler 6

12 PB flash

4.3 TB/s Read

3.1 TB/s Write

DDN **Hot node** for Accelerated AI Training

EOS IS THE THIRD-GENERATION FLASGHIP DGX SUPERPOD NVIDIA HAS CHOSEN TO DEPLOY WITH DDN

EOS is part of a long list of DDN and NVIDIA SuperPODs

NVIDIA Eos
World’s Fastest SuperPOD

NEC AI Research
Japan

Berzelius
Linköping University

Cambridge-1
UK Life Sciences

PARAM Siddhi AI
India Research and R&D

NAVER AI Cloud
South Korea AI Services

HiPerGator
Uni of Florida

Lambda
US Cloud SuperPOD

NVIDIA Selene
World’s First SuperPOD

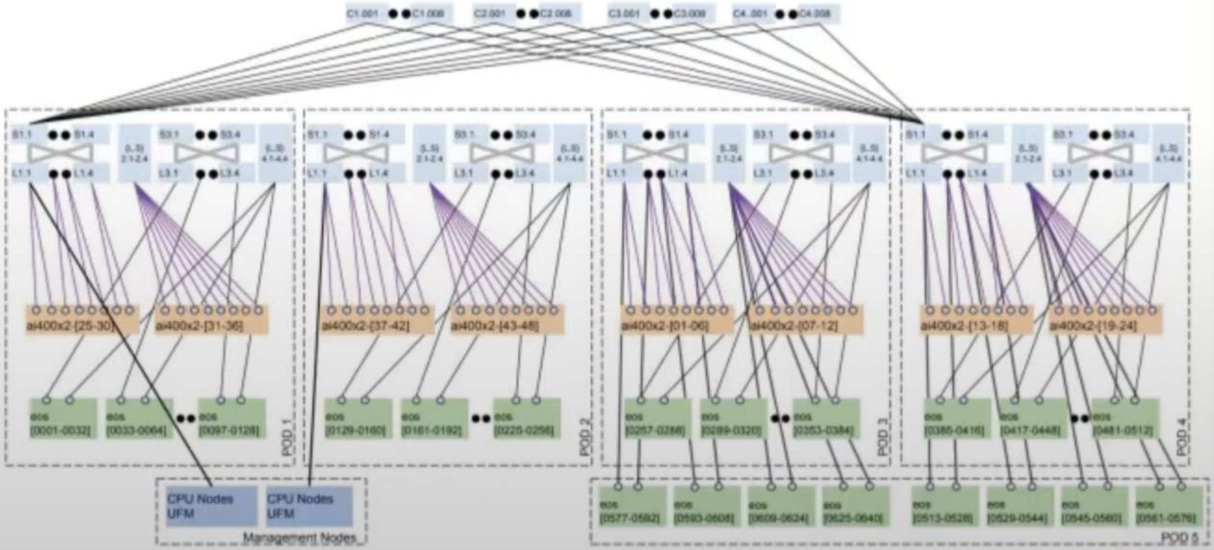
Scaleway
Europe Cloud SuperPOD

EOS looks very much like a fat-node fat-tree HPC system

Hierarchical Design: 32N Scalable units, 128N PODs

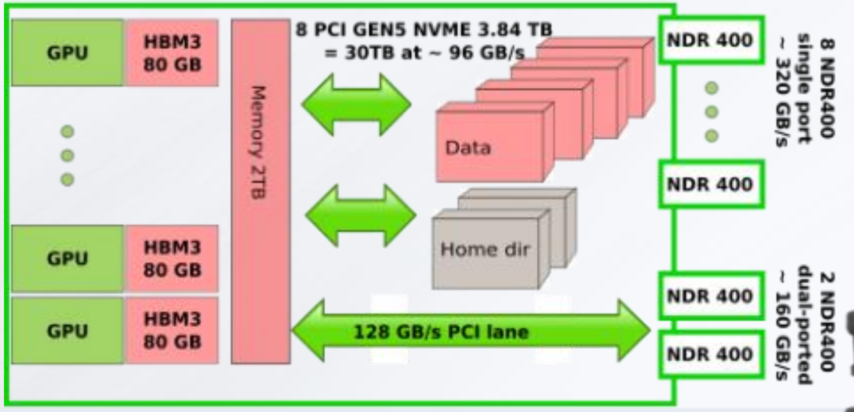
- DGX H100: 8xNDR400 ports for compute and 2x NDR400 port storage
- Separated, non-blocking fabrics for both compute and storage
- Three-level fat tree topologies

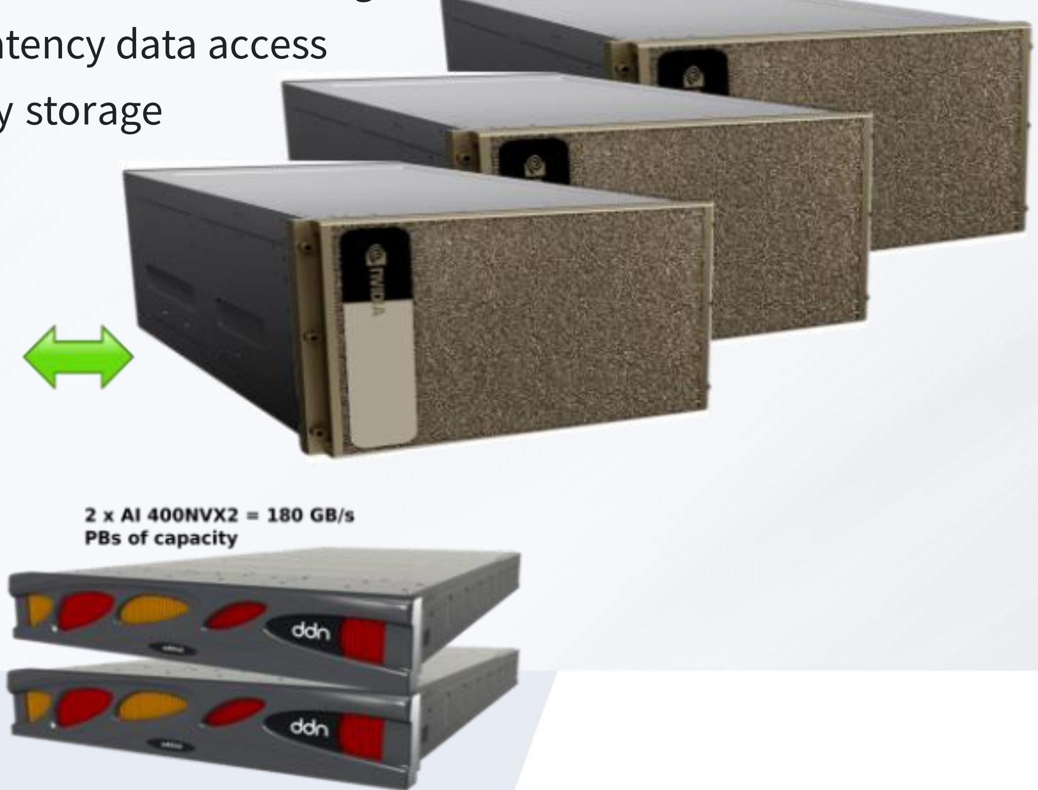
- 48 DDN AI400X2 storage appliances connected with HDR IB
- Appliance IB connections are interleaved across 4 PODs
- Target a minimum performance of 2 TB/s (read) to support DL training at scale



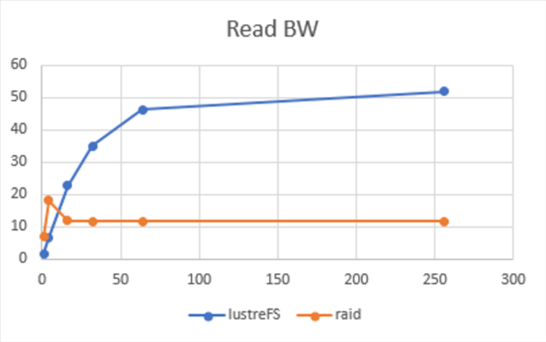
DGX Architecture Overview

- DGX H100: Dual Networks 8 NDR400 Compute / 2 NDR400 Storage
- Internal Storage for no-interference small latency data access
- 2 NDR400 for high bandwidth / high capacity storage

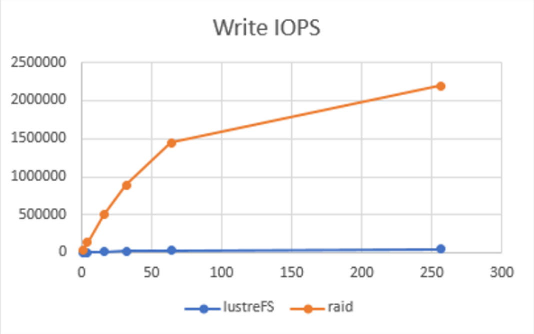




Storage micro-benchmarking



Comparative bandwidth measurements on a DGX platform. Using FIO with threads number ranging from 1 to 256 and large payload.
The Lustre delivers x5 the read performance and x2 the write performance of the local storage.



Comparative latency measurements on a DGX platform. Using FIO with a threads number ranging from 1 to 256 with a small payload. Local storage delivers x5 the IOPS (IO operations per second) than Lustre and x100 the IOPS of Lustre for write operations. Lustre version 2.12 used in this experiment does not support the most recent IOPS write optimizations

Next-Gen AI Data Caching with DDN Hot Nodes

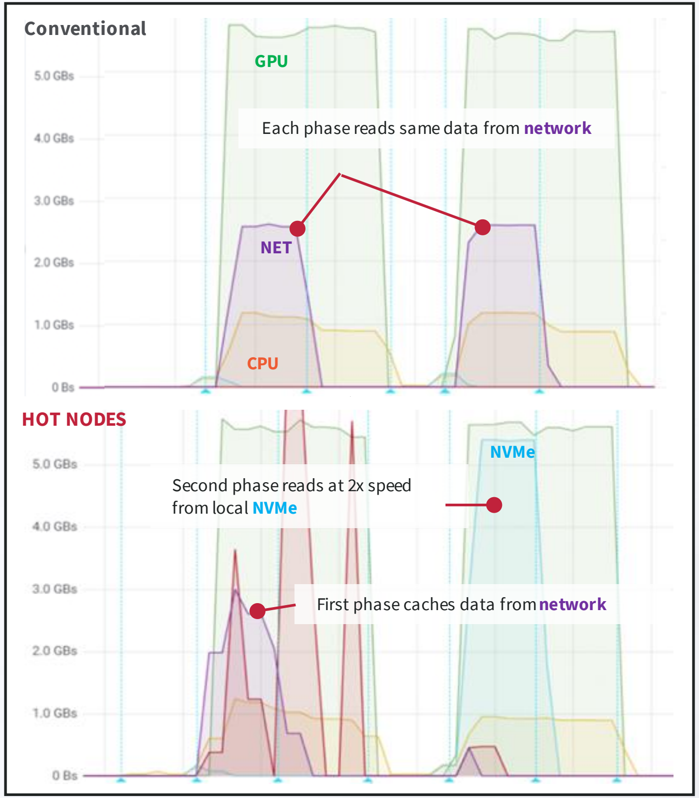
Leverage local flash to maximize benefits of unified, global shared namespace

- Achieve full AI application performance with data cached on local nvme devices in client, without any manual and risky data management overhead.
- Automated data movement from shared space to local node with intelligent policy-based cache management makes the process entirely transparent for users.
- Delivers significant efficiencies and AI workload improvements with large number of nodes engaged simultaneously for training, especially for at-scale NLP.

Cacheing Data in DGX Local NVMe for Multi-Epoch Training

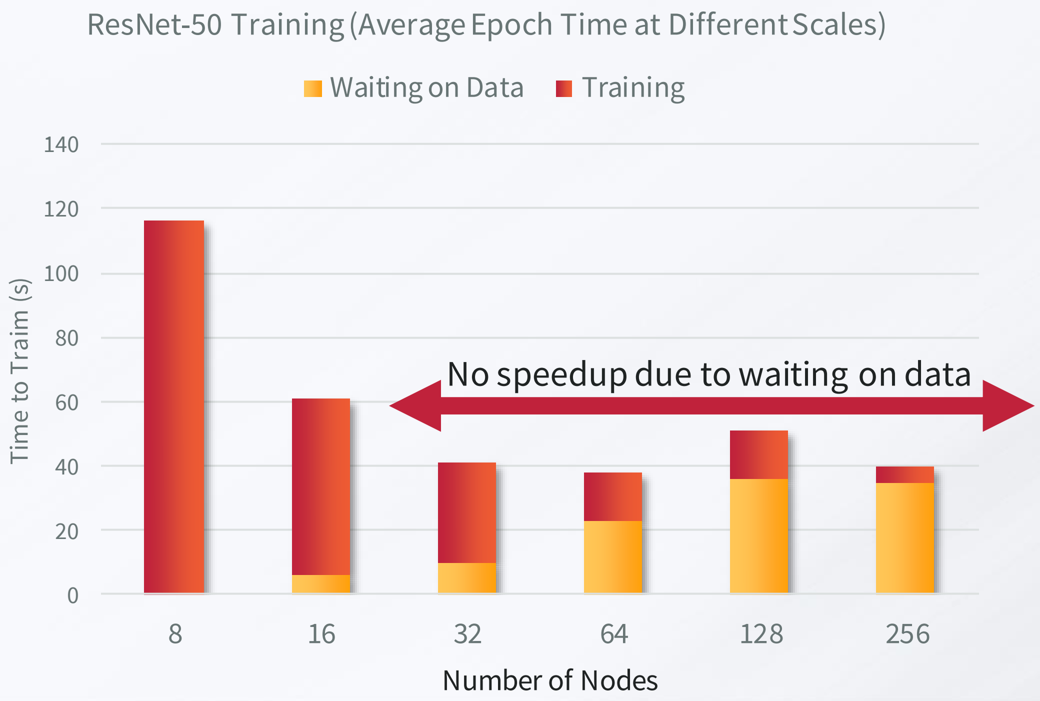
- ResNet50 benchmark on DGX-A100 + AI400 without caching
- Each phase reads same data from network (purple)
- Compute runs in parallel with IO (CPU orange, GPU green)

- ResNet50 with caching on internal NVMe devices
- First phase also reads from network (purple)
  - Total data read volume is similar, second read from RAM
- Computation also reads from network/RAM while files copied
- Write to cache storage on NVMe (red)
- Second phase reads from NVMe at double bandwidth (cyan)
- GPU usage (green) the same, CPU usage (orange) lower
  - No network/server load on second and later runs



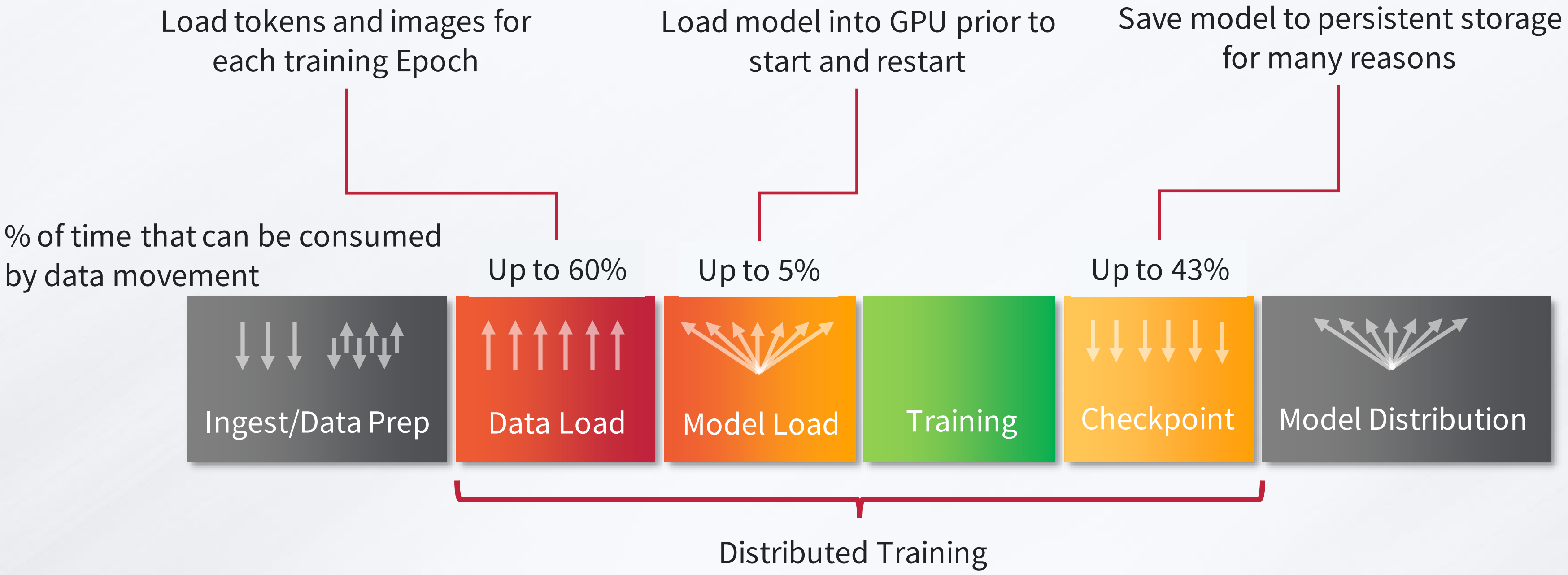
Waiting on Data can be a Critical Bottleneck

As models scale the Data Movement component for Multi-Epoch Training can become the dominant factor in Training Time



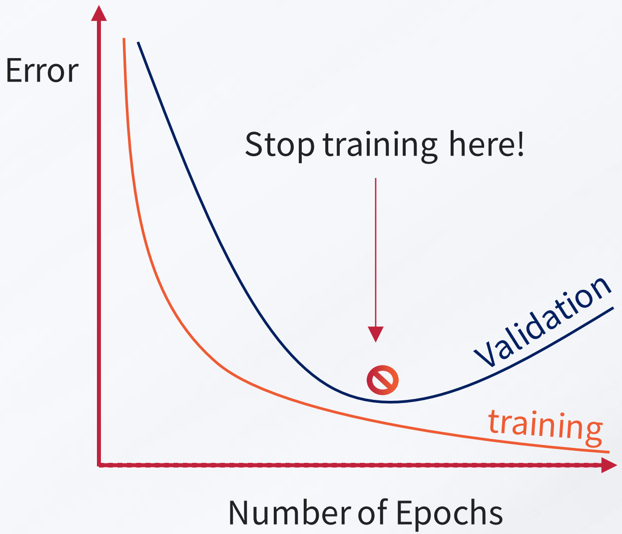
https://ar5iv.labs.arxiv.org/html/1910.01196

AI Training is Storage Intensive



Checkpoints are a Critical Step in Deep Learning Training

- **Prediction Accuracy** – improve accuracy by lowering learning rate from a checkpoint
- **Multi-System Training** - continue training model across different nodes or clusters/cloud
- **Transfer Learning** – if goals change, start afresh from a checkpoint
- **Better Fine Tuning** - pick out less trained states to restart new experiments
- **Early Stopping** - for large models, without sufficient regularization, the error on the evaluation dataset can start to increase



Checkpoints are a Critical Step in Deep Learning Training

Master AI Infrastructure: Learnings from the past 5 years

Reference Architecture ai400nvx2

Workload Guidelines Storage BW per GPU

- Small data set. LLM. Text processing: 1.4 GB/s Read 1 GB/s Write
- Medium dataset. Compressed video processing. Checkpointing Distributed training LLM: 2.8 GB/s Read 2 GB/s Write
- Large dataset. Uncompressed video processing. Checkpointing Distributed training LLM 5.6 GB/s Read 4 GB/s Write

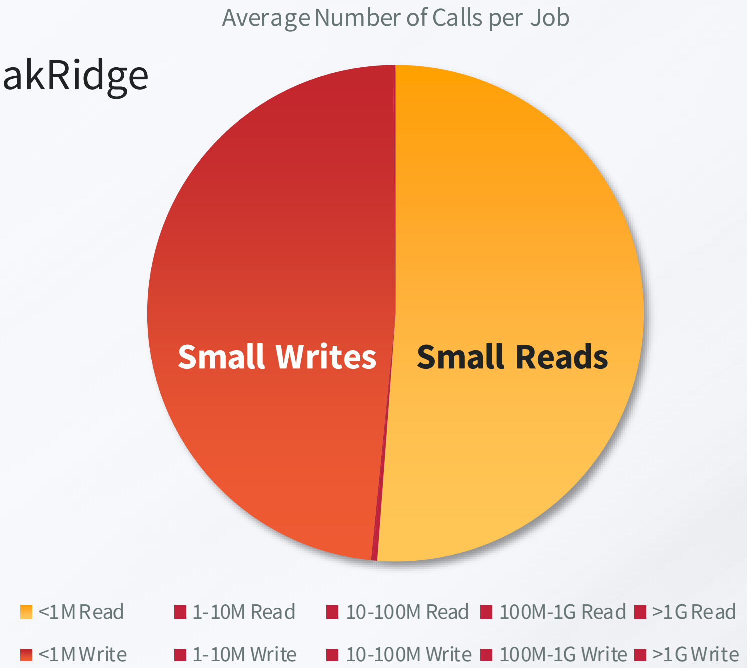
Operation Guidance: EXAScaler Management Facility

- Automated deployment and upgrade
- HA configuration
- Health Monitoring
- Prometheus Exporter

Machine Learning Needs Fast Reads and Writes

Analysis of over 23,000 Machine Learning Jobs at OakRidge

- "Most ML jobs are perceived to be read-intensive with a lot of small reads while a few ML jobs also perform small writes."
- "Our study showed that ML workloads generate a **large number of small file reads and writes**…"



~50% W/R ALSO VALIDATED DURING EOS PRESENTATION (SC23)

PAUL, Arnab K., KARIMI, Ahmad Maroof, et WANG, Feiyi.
Characterizing machine learning i/o workloads on leadership scale hpc systems.
In : *2021 29th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*.
IEEE, 2021. p. 1-8. *https://arnabkrpaul.github.io/publications/mascots21.pdf*

New DGX SuperPOD RAs: AI400X2-QLC and AI400X2-Turbo



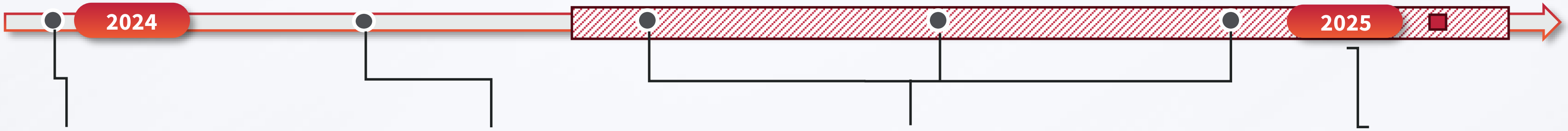
AI400X2-QLC	AI400X2	AI400X2 TURBO
More Useable Capacity	*Optimal Performance/Capacity*	*Extra Performance Boost*
90 GB/s (r), 70 GB/s (w), 3.5M IOPs	90 GB/s (r), 65 GB/s (w), 3M IOPs	115 GB/s (r), 75 GB/s (w)*
1 or 2 PB useable (QLC)	120, 250, 500 TB useable (TLC)	120, 250, 500 TB useable (TLC)
6 RU • 4.4 KW • 15K BTU/hr	2 RU • 2.2 KW • 7.5K BTU/hr	2 RU • 2.2 KW • 7.5K BTU/hr
HDR200/100GbE/200GbE QSFP 56 (8) or NDR200/200GbE QSFP 112 (8)	HDR200/100GbE/200GbE QSFP 56 (8) or NDR200/200GbE QSFP 112 (8)	NDR200/200GbE QSFP 112 (16)

FULLY INTEGRATED! NO EXTERNAL SWITCHES!

** AI400X2-Turbo DVT Numbers - Improvements expected with GA in Q1 2023*

EXAScaler Roadmap 2024

Updated 2024-02-15



EXA 6.3

Appliances
- QLC Appliance

Data Management
- Client-Side Compression

EMF
- HA Reliability Improvements

Data Services
- Data Services as software
- Per User BucketFS path

EXA 6.3.1

Appliances
- AI400X2T “Turbo” appliance

Core File System
- Reliability on unstable networks
- **EMF**
- Simplified Online Upgrades
- Enhanced SOS reports
- **Data Services**
- Prometheus Exporter

CSI
- OpenShift support

EXA 6.3.x

Data Management
- Client-Side Compression v2

EMF
- Prometheus Exporters GA
- EXA Health GA

Data Services
- S3 AD authentication support

EXA 7.0

Core File System
- Large ldiskfs OSTs
- IPv6

EMF
- Push-button Online Upgrades
- Configuration Management

Data Services
- EMF Deployment

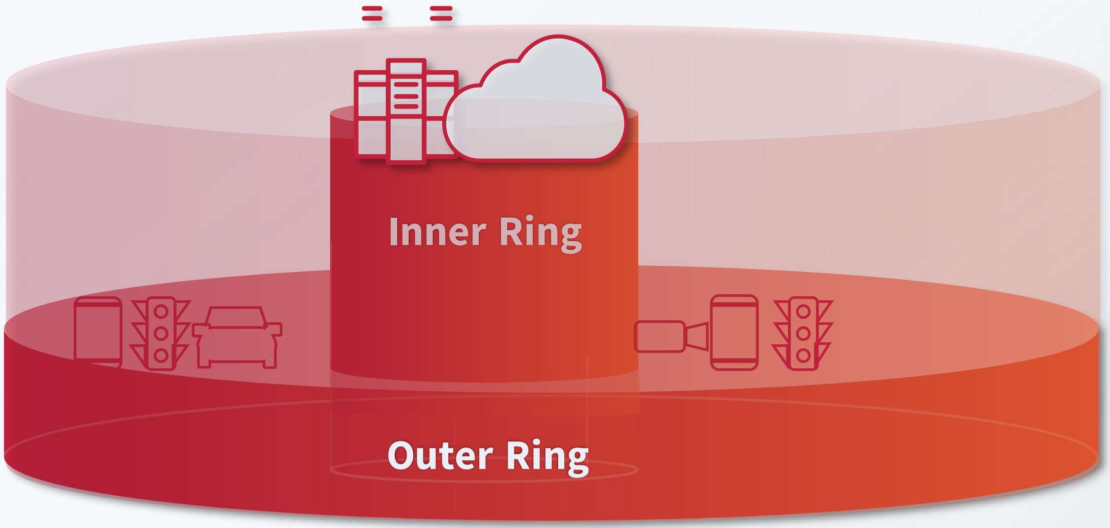
Subject to Change

From AI Training to AI Data Center

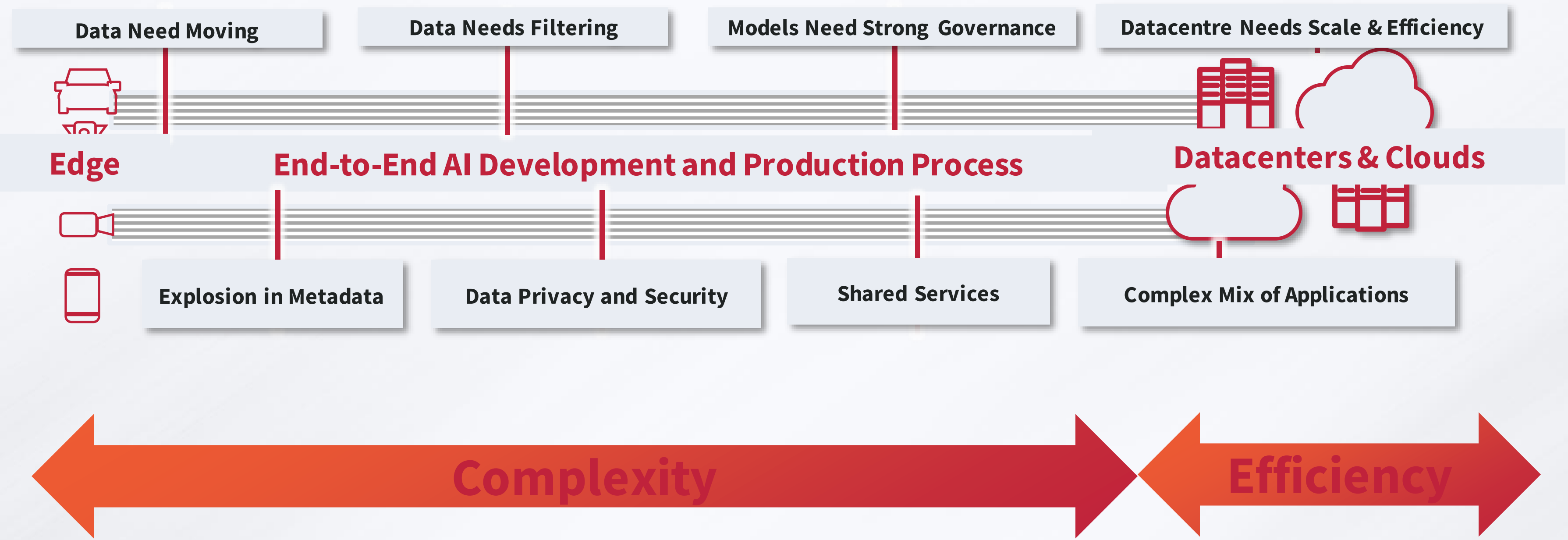
AI Data Centers Operates Different Rings of Data

DDN EXAScaler
Scale, Efficiency & Performance

DDN Infinia
Simplicity, Multi-Tenancy, SW-Defined



Data Journey from Acquisition to Training



Addressing Each Challenge with the Most Relevant Solution

DDN Infinia

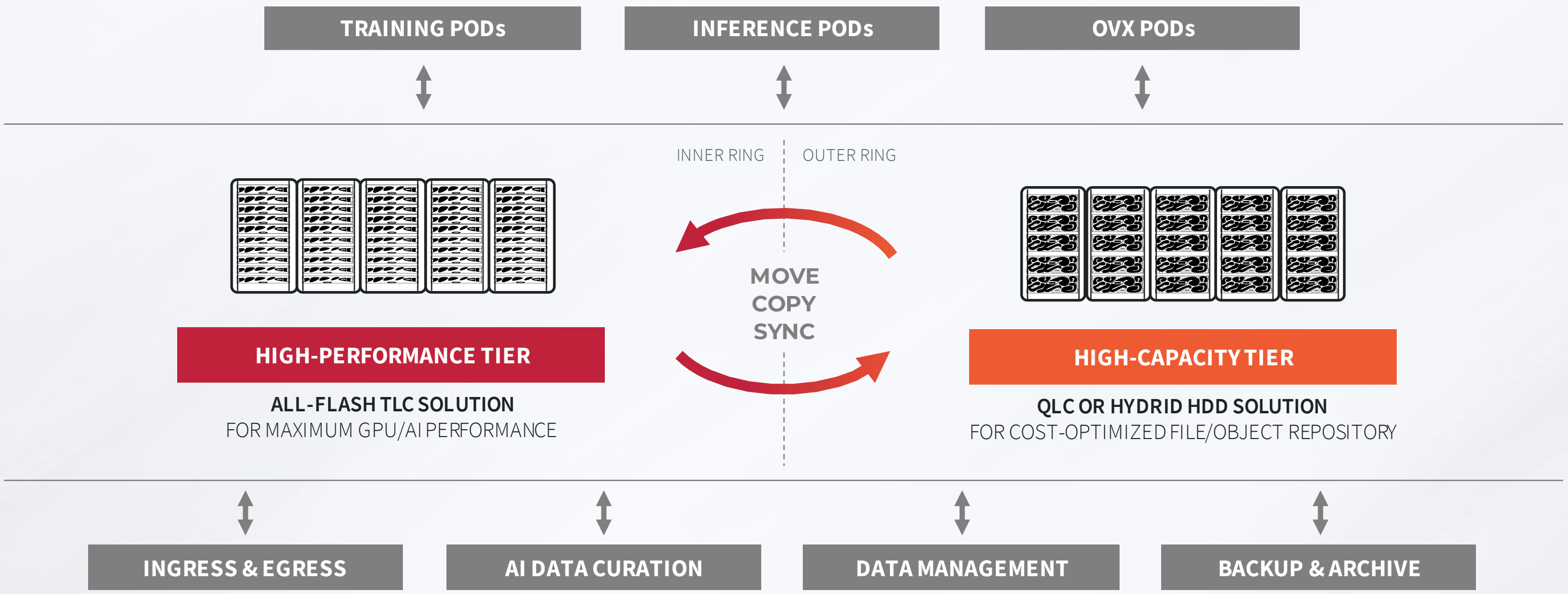
- Accelerate & Cost Reduce AI development
- Enable Strong Governance
- Simplify End-to-End

DDN EXAScaler

- Maximum Uptime
- Highest Efficiency; NCP & On-Prem
- HyperScale

AI Data Center: Inferences and Training

Data management for inference with versatile numerous inputs differs from high performance training on curated data



Storage Designed for Next Generation AI

Software Driven. Powerful Analytics. Strong Security. Highly Efficient for All Workloads

Scale Out S3 Object Storage
Start with 500TB and scale out just by adding servers

Zero Service Disruption
Expand, remove, upgrade, introduce new generations of hardware without service disruption

Cloud-Like Multi Tenancy
Manage SLAs for Tenants and Subtenants dynamically and easily

100% QLC Flash Simplicity and Stability
HDD-free Architecture, hyper simple management

DDN Infinia is 100% Fit Into Modern AI Development

- Fully Containerized Architecture → Elastic Scalability. Always On
- Multi-Tenancy that is 100% Dynamic → Easily change Tenant Allocation
- Massive Unstructured Data → Tag & Search without limits, collapse complexity

TAG, SEARCH, STORE

TENANTS & SUBTENANTS

DYNAMIC QoS

CLOUD NATIVE

SCALABLE

ELASTIC

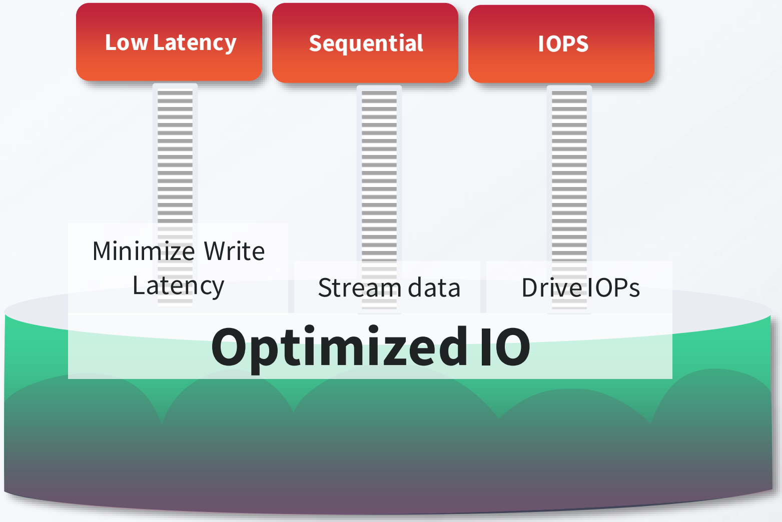
Infinia Implementation of the Bε-Tree Core Data Structure

- **Efficient Space Utilization**: good use of space within nodes
- **Reduced Disk I/O**: By optimizing the node fill factor, they can reduce disk operations, which is crucial for performance in large DB systems or metadata operations.
- **Scalability**: well-suited for large datasets and can scale.
- **Range Queries Efficiency:** efficient in performing range queries, making them useful in applications where such queries are common.
- **Flexibility in Node Size**: flexibility in setting node size, which can be tuned for specific application/workload needs.
- **Dynamic Nature**: dynamically adjust to insertions and deletions, maintaining balanced tree structure.

Keeping Every Application Optimized Regardless of Type

Hybrid IO Engines Handle Diverse Workload Needs with High Efficiency

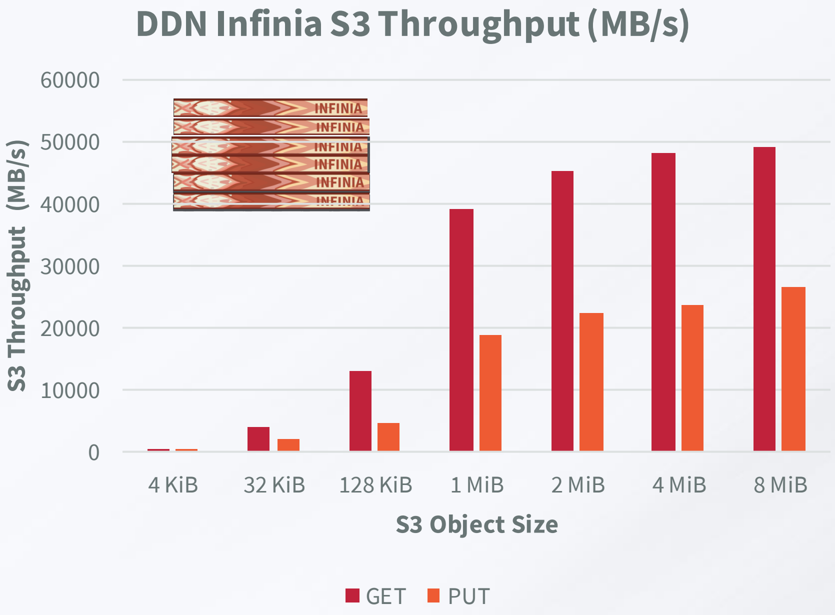
- Autonomous optimization for **low latency *and* IOPs *and* throughput and metadata intensive workloads**
- No read/modify/write overhead due to log structured operations
- Byte addressable with wide striping for large I/O blocks
- No Complexity for Administration - Elastic space for metadata and data



Infinia is Fast - Accelerates S3 workloads by 10-100X

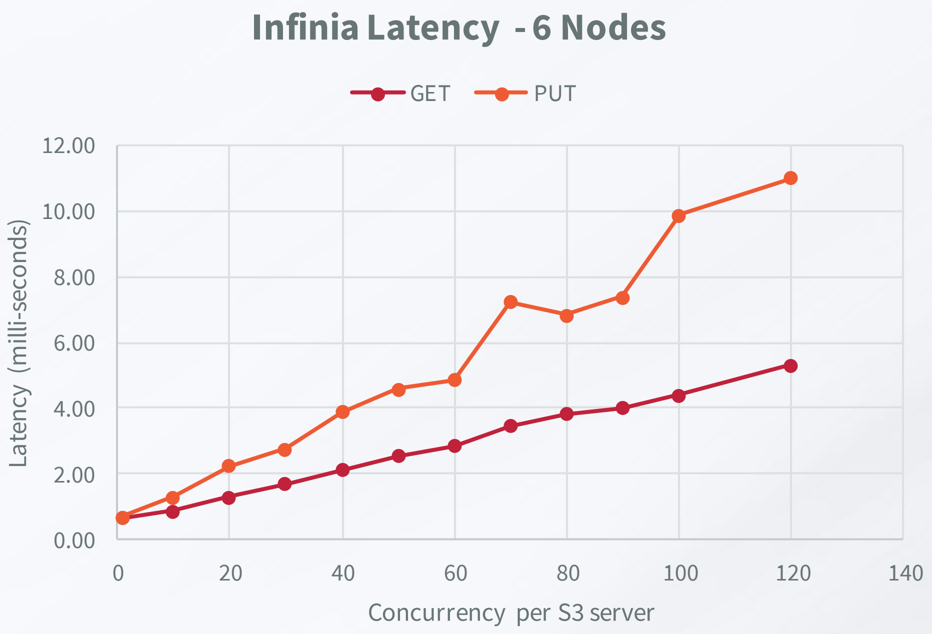
- DDN Infinia exceeds the claimed performance of all other object stores in Objects/sec and Throughput
- Accelerate Enterprise Analytics workloads like Apache Spark, Starburst Presto/Trino, Clickhouse,

Object Size	GET	PUT
4K	**147 K**	**67 K**
4M	**49 GB/s**	**26.8 GB/s**



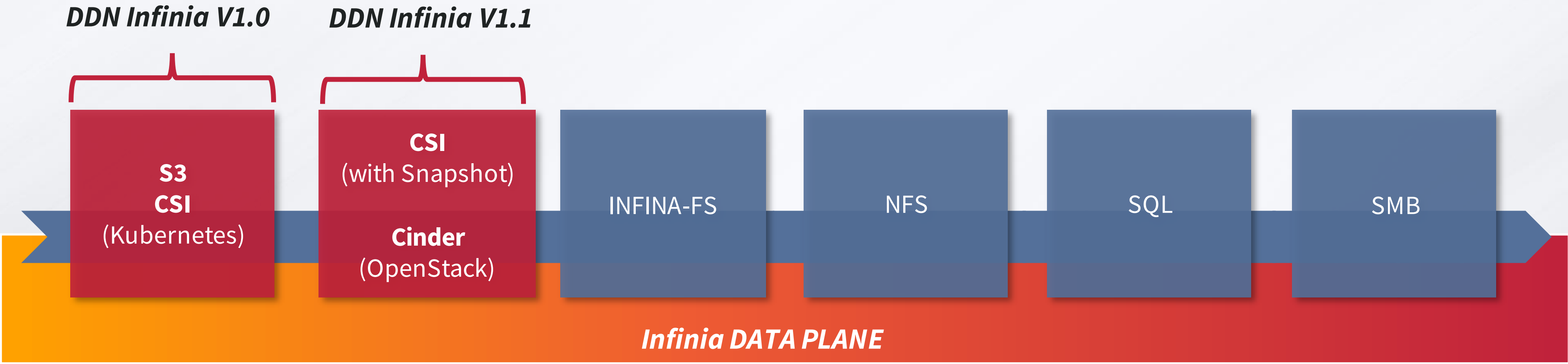
Infinia is Responsive - 100x Lower Latencies

- Infinia is the first Object Storage to deliver **sub-millisecond latencies** for PUTs and GETs
- 100x lower than AWS S3
- Improve website load times, database response times, accelerates Spark queries etc improving end user satisfaction



Infinia – Data Plane and Control Plane Architecture

- Infinia is a 100% software, dataplane designed to deliver storage services on a variety of protocols
- SW upgrades will introduce additional services including an Optimized Parallel Client, NFS and SQL.

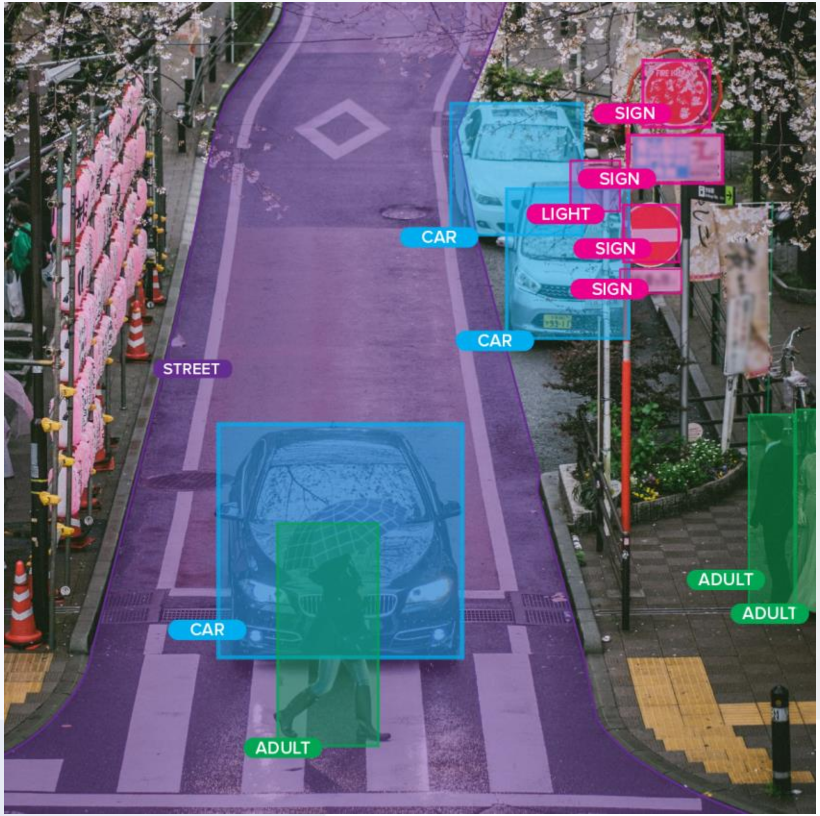


DDN Infinia Protects Service Levels Automatically

- DDN Infinia will automatically defend your Tenants ***and Subtenant*** Performance
- Full protection from Noisy Neighbours
- Complete Fair Share implementation in Software

Blurring Borders Between Metadata and Data

- AI Data tend to be metadata heavy
  - Every frame of an autonomous car is annotated by 100s of metadata
- Metadata allow to structure the Data-lake
  - Prevent Data-lake to turn in Data-Swamp
- Query-able Metadata: Data-LakeHouse
  - Data Lake + Data Warehouse



DDN Infinia SQL

- Create Tables, Insert Rows, Select a Subset
- Scale Out Performance
- Manage S3 Tags and File Extended Attributes like a Database

Create a 3 Column Table: Integer: Integer: Variable Character

```
redsql> CREATE TABLE redtest (col1 BIGINT, col2 BIGINT, col3 VARCHAR(48))
redsql> time milliseconds: 2.533413
```

Insert Values 1, 20 and "test"

```
redsql> INSERT INTO redtest VALUES (1, 20, 'test')
redsql> time milliseconds: 5.021347
```

```
redsql> SELECT * FROM redtest WHERE col1 > 70 AND col1 < 85
{"col1":83,"col2":20,"col3":"test83"}
{"col1":73,"col2":20,"col3":"test73"}
{"col1":80,"col2":20,"col3":"test80"}
{"col1":84,"col2":20,"col3":"test84"}
{"col1":78,"col2":20,"col3":"test78"}
{"col1":79,"col2":20,"col3":"test79"}
{"col1":71,"col2":20,"col3":"test71"}
{"col1":74,"col2":20,"col3":"test74"}
{"col1":75,"col2":20,"col3":"test75"}
{"col1":76,"col2":20,"col3":"test76"}
{"col1":82,"col2":20,"col3":"test82"}
{"col1":77,"col2":20,"col3":"test77"}
{"col1":72,"col2":20,"col3":"test72"}
{"col1":81,"col2":20,"col3":"test81"}
redsql> time milliseconds: 6.005832
redsql>
```

THE AI DATA COMPANY

Infinia Primary Dashboard

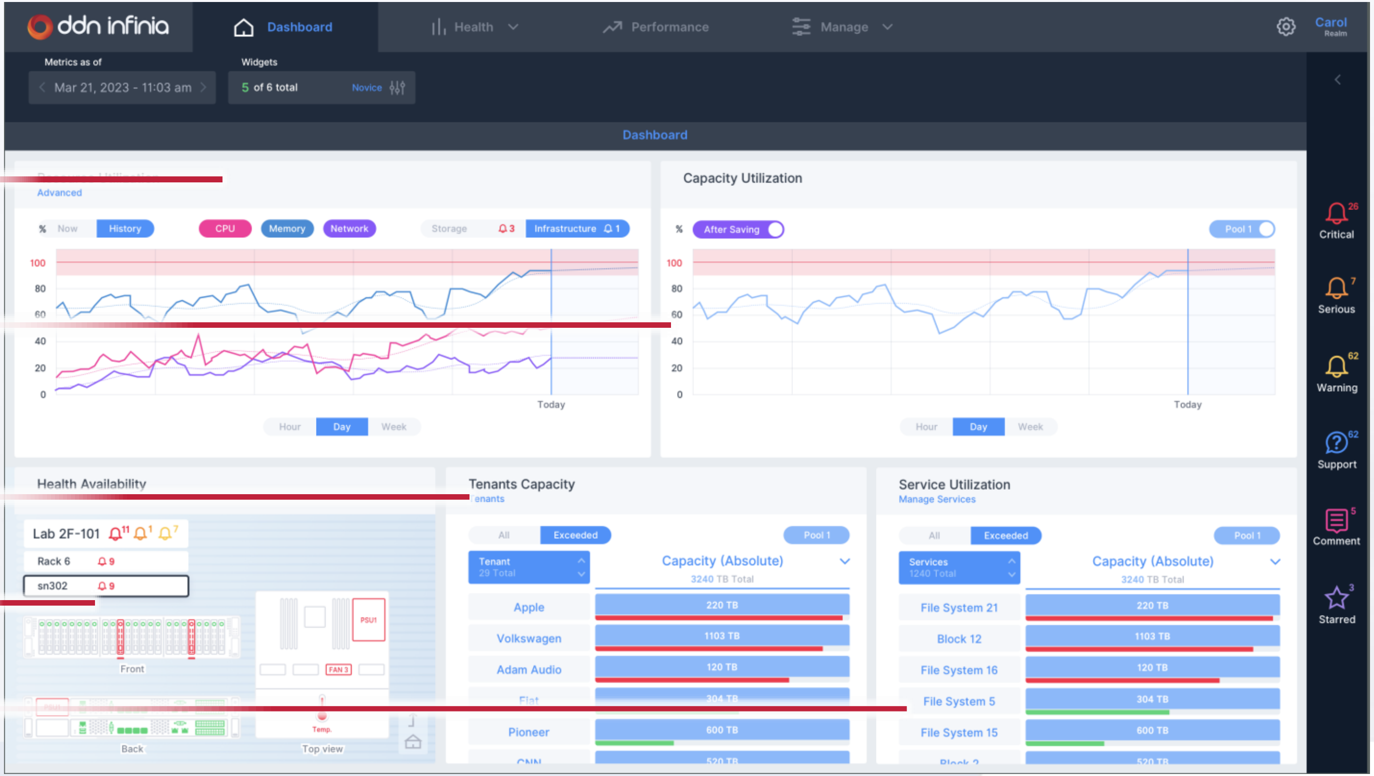
Resource Utilization

Capacity Utilization

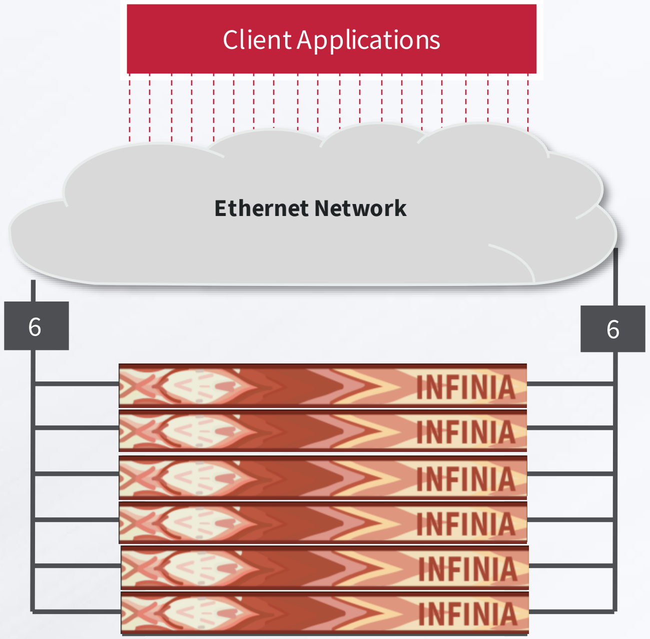
Capacity by Tenant

Physical Views of Health

Capacity by Data Service



Scale Out S3 in the Data Center and in the Cloud Up to 100’s of PBs

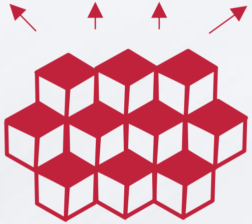


- S3 Object Storage
- Multi-Tenancy Namespace
- No Restrictions on Object Size
- No Restrictions on Metadata
- Full Multi-Part Upload Support
- Sub Millisecond Latency for PUTs and Gets

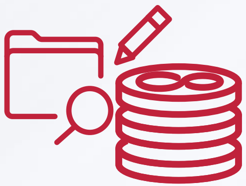
DDN Infinia will Enable the Next Generation of AI

The 4 big data challenges of the new Era of AI: Efficiency | Metadata | Sharing | Moving

Data services

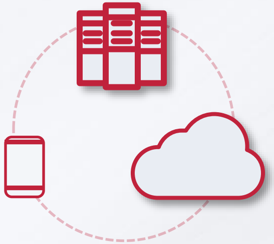


Containerized MicroServices



Scalable Data & Unlimited Metadata

Native Tenancy Architecture



Move Data; Edge, Cloud, Datacenter

And the Best Storage is?


Accelerate AI at Scale
Then the best storage is?


Accelerate AI at Scale
DDN!

ddn