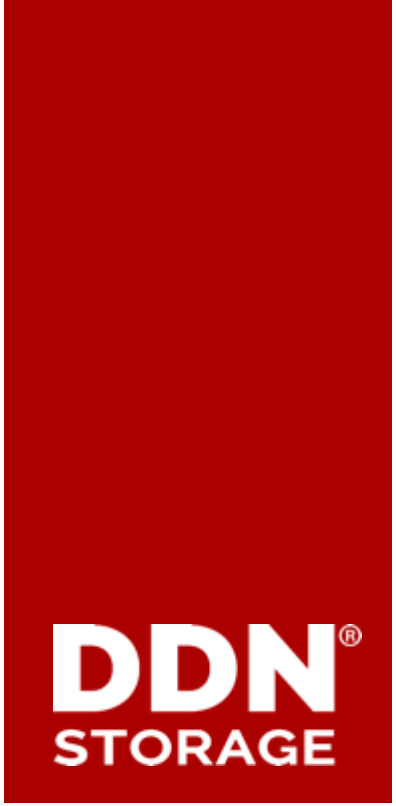


DDN Update 2019

Gfarm Workshop 2019

DataDirect Networks Japan, Inc.

2019/2/1

橋爪 信明 (nhashizume@ddn.com)

2018年主な出来事@DDN

- Declustered RAID(DCR)対応製品へラインナップ刷新
  - SFAOS11世代のHWへの切り替え元年
  - SFA200/400NV, SFA7990, SFA18K

- 6月 Intel社よりLustre部門を買収
  - DDNのWhamcloudディビジョンとして活動
  - Open Source Community Edition開発
  - Lustre L3サポート

- 9月 Tintri社を買収
  - DDN内の独立したTintriディビジョンとして活動
  - 仮想化・マルチクラウド自動最適化フラッシュストレージ

DDN SFA

ALL-FLASH AND HYBRID BLOCK STORAGE PLATFORMS

2019年より全グレードで新ラインナップを展開

200NV	400NV	7990	14KX	18K
				
20GB/s 1M IOP/s	40GB/s 2M IOP/s	20GB/s 1M IOP/s	60GB/s 4M IOP/s	90GB/s 3.2M IOP/s
24 NVME Slots	24 NVME Slots	Up to 450 SSD/HDD	48 NVMe Slots Up to 1872 SSD/HDD	48 NVMe Slots Up to 1872 SSD/HDD
EDR IB (4), OPA (2) FC32 (8), FC (8)	EDR IB (8), OPA (4)	EDR IB (4), OPA (2) FC16 (8)	EDR IB (12 | 8) OPA (4), FC16 (24)	EDR IB (16), OPA (8)
NEW 2018	COMING 2019	NEW 2018	To be EOS	COMING 2019

DDN SFA | 製品名と位置付け

位置付け	用途	200NV	400NV	7990	18K
総称	全体をまとめて呼ぶ場合	SFA200NV	SFA400NV	SFA7990	SFA18K
ブロックストレージ	ファイルサーバに接続	SFA200NV	SFA400NV	SFA7990	SFA18K
エンベデッドアプライアンス	コントローラ上にファイルサーバを実装	SFA200NVE	SFA400NVE	SFA7990E	SFA18KE
ExaScalerアプライアンス	ExaScaler(Lustre)エンベデッド	ES200NV	ES400NV	ES7990	ES18K
GridScalerアプライアンス	GridScaler(GPFS)エンベデッド	GS200NV	GS400NV	GS7990	GS18K
A3Iアプライアンス	AI向けエンベデッド	AI200	AI400	AI7990	

DDN® STORAGE

SFA18K

MAXIMUM SCALABILITY WITH NVMe + SAS

FOR CAPACITY & BANDWIDTH WITH DRIVE TYPE FLEXABILITY

SFA18K

Block 90GB/s
Embedded 78GB/s

- **Start in as little as 4U and Scale up & Scale out;**
  - Up to 1872 drives per system:
    - 72 Slots in “head”, 48 NVMe/SAS + 24 SAS
    - Add up to 20x SS9012 Enclosures (90 SAS slots)
- **Flexible Access for File: IB or Ethernet (16 ports), OPA (8 ports)**

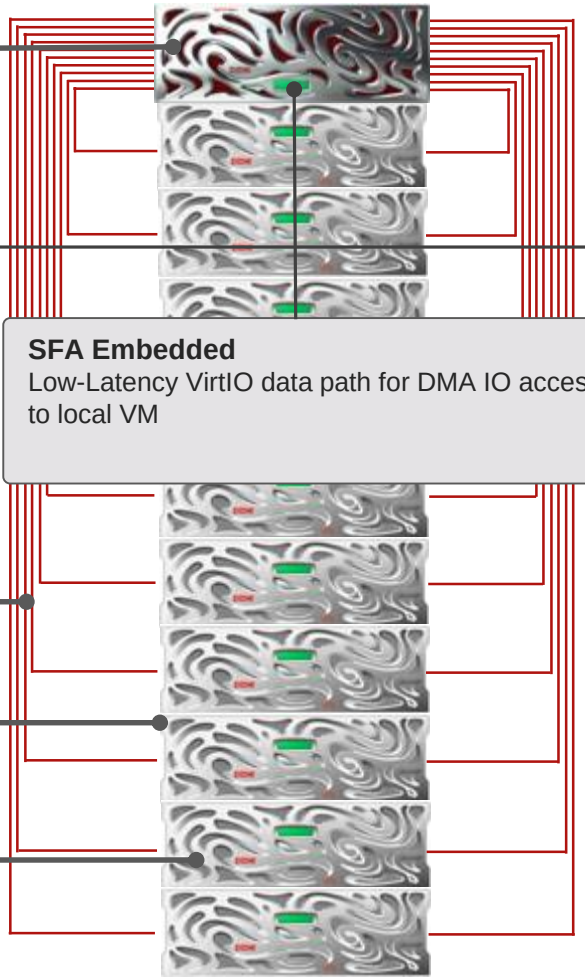
Integrates high-performance, low-latency, RDMA capable networks. EDR InfiniBand and 100Gbps (or 40Gbps) Ethernet and OPA

- **Access for Block: IB (16 ports)**

Block systems attach via EDR IB to external servers

SFA18K Overview

Broad Connectivity Options
EDR IB
40/100 GbE
OmniPath



Real-Time, Multi-CPU RAID Engine
Full Implementation of Real-Time Storage OS, Interrupt-Free; Massively Parallel I/O Delivery System

Flash and NVMe Enabled in 'head'
Up To 48 NVMe devices
Up to 72 SAS SSD

SFA Embedded
Low-Latency VirtIO data path for DMA IO access to local VM

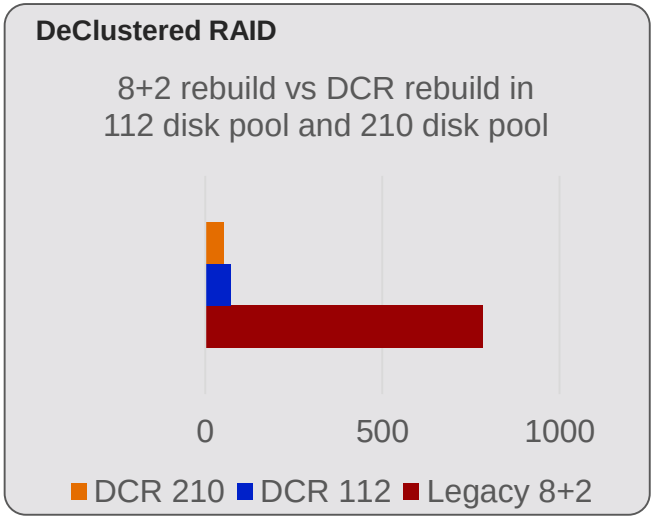
ReACT Cache Management
Real-Time I/O Data-Aware Cache Optimizer, Small IOs Go to Cache; Large & Streaming I/Os Bypass Cache

Storage Fusion Fabric
Highly-Over-Provisioned Backend RAIDed Fabric Withstands Failures of Drives, Enclosures, Cables, etc

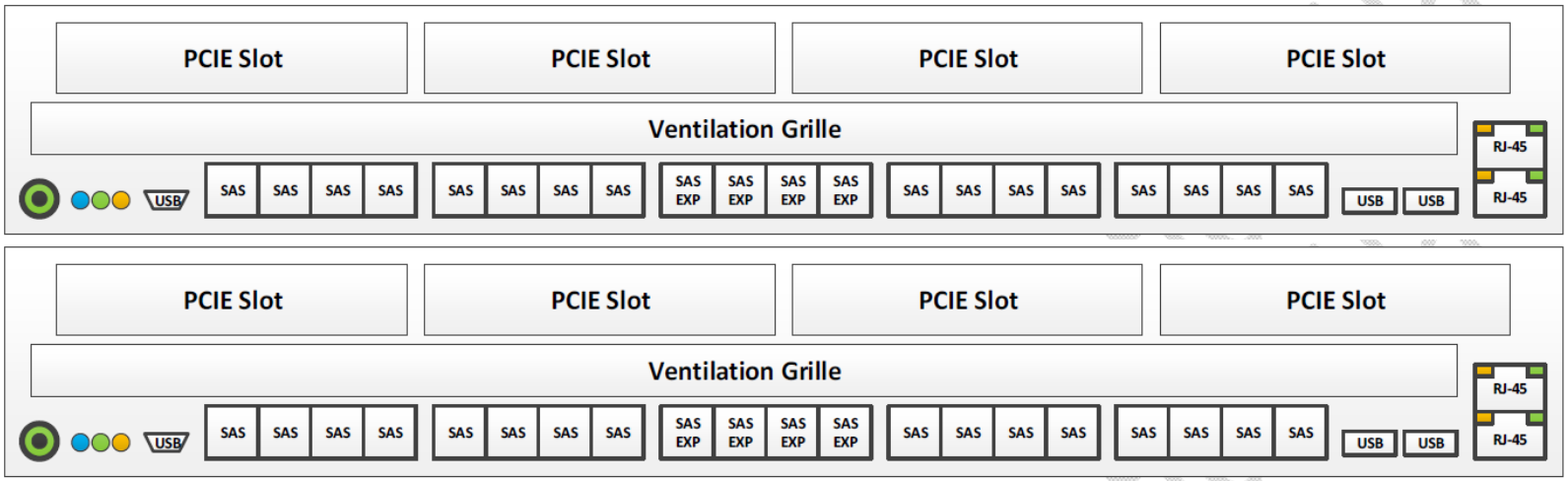
High Density Array
Up to 972 drives in 44U rack. SS9012 enclosure is 90 HDDs / SSDs in 4U

Data Integrity
Real-Time Data Integrity Verification for every I/O

DeClustered RAID



SFA18K Controller



SFA18K CONFIGURATIONS

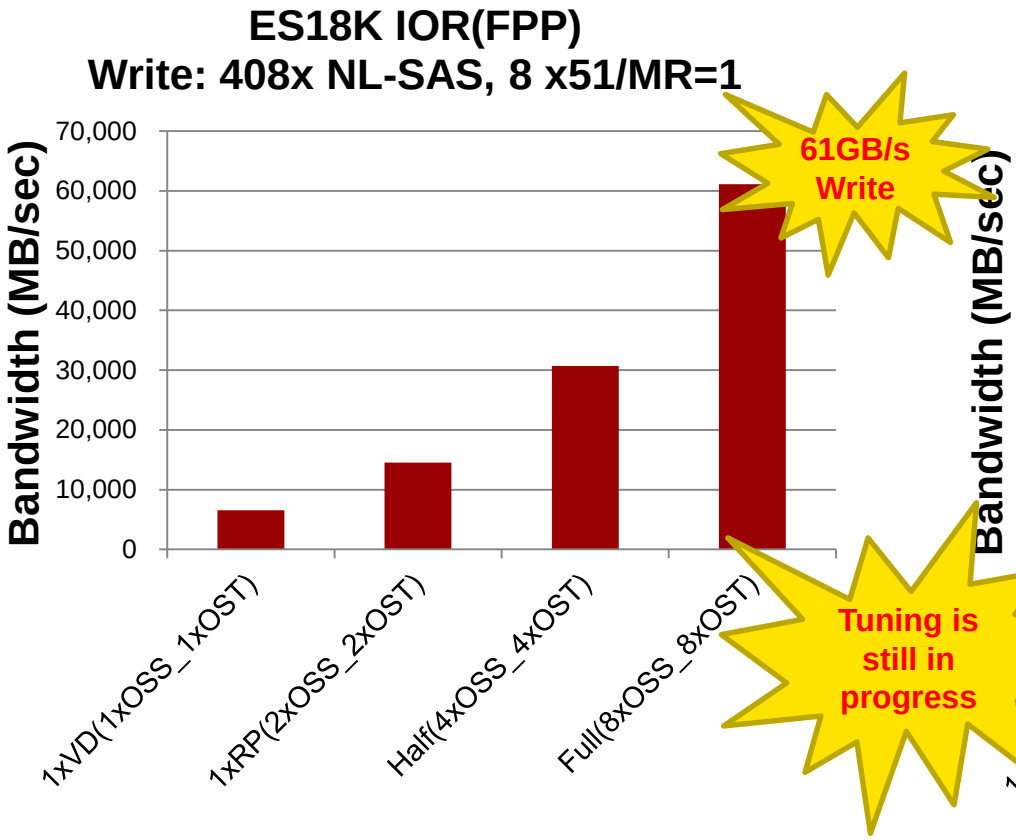
Config	SFA18K Only	+ 1 SS9012	+2 SS9012	+4 SS9012	+5 SS9012	+6 SS9012	+8 SS9012	+10 SS9012	+16 SS9012	+20 SS9012
SFA (BLOCK)										
ES										
GS										
2.5” NVMe or SAS SSD Slots	72 (NVMe:48)	72 (NVMe:48)	72 (NVMe:48)	72 (NVMe:48)	72 (NVMe:48)	72 (NVMe:48)	72 (NVMe:48)	72 (NVMe:48)	72 (NVMe:48)	72 (NVMe:48)
3.5” HDD or SAS SSD Slots	0	90	180	360	450	540	720	900	1440	1800
Enclosure Connection	NA	Direct	Direct	Direct	Direct	Direct	Direct	Direct	Direct Daisy Chain	Daisy Chain
最大物理容量	NA	**1.26PB**	**2.52PB**	**5.04PB**	**6.30PB**	**7.56PB**	**10.08PB**	**12.6PB**	**20.16PB**	**25.20PB**

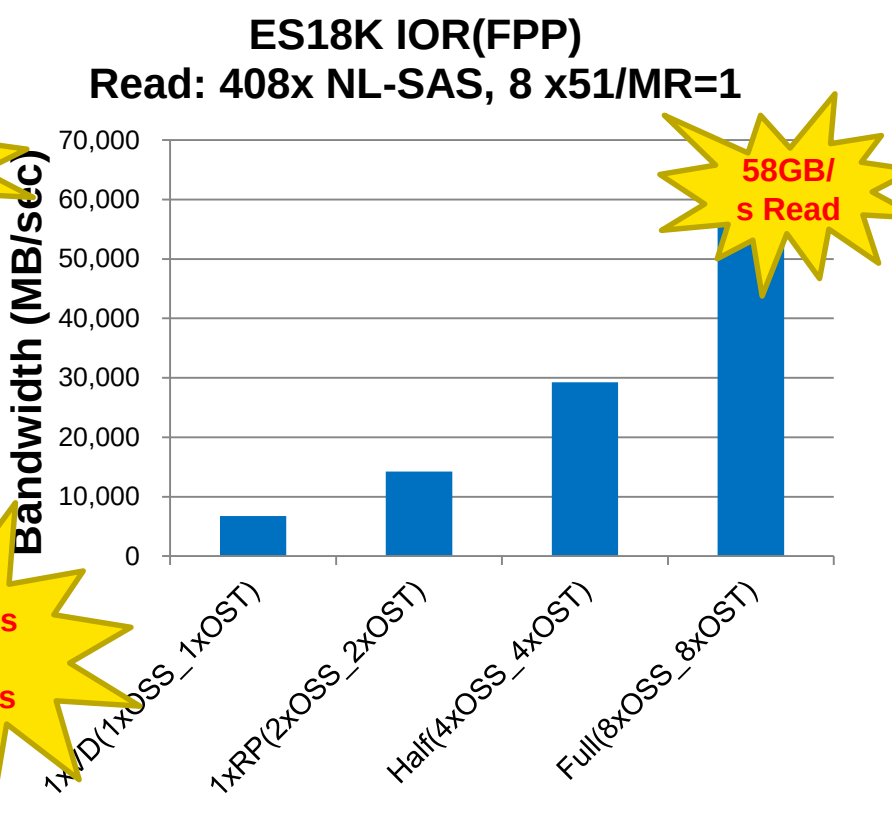
Expansion Options

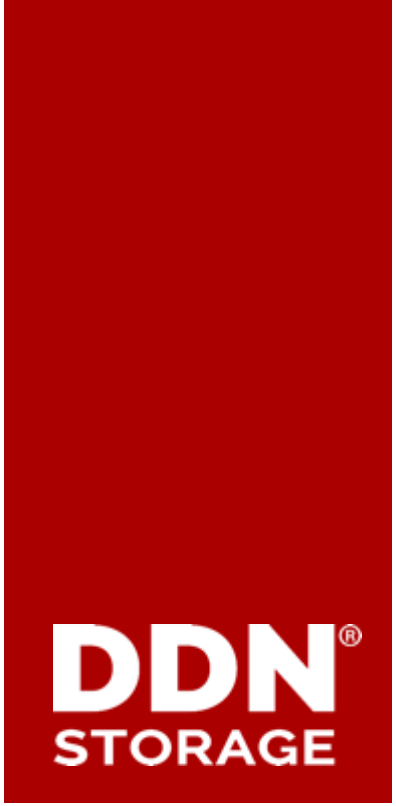
※ 最大物理容量は14TB NL-SAS HDDをSS9012に実装した場合

EXAScaler ES18K Early Performance - IOR 1MB FPP

- EXAScaler IOR Client Benchmarking with an ES18K EDR with 408 NL-SAS Drives as 8 DeClustered RAID Pools of 51 drives.







SFA7990

SFA7990 WORLD'S MOST EFFICIENT HYBRID STORAGE

OPTIMIZED FOR SSD, HDD CAPACITY, MULTICORE & MULTI GPU

SFA7990

Block 20GB/s
Embedded 20GB/s

DDN's ES7990 and GS7990, EXAScaler® and GRIDScaler® appliances, offer scalable, high performance parallel file systems in an integrated package. Designed, deployed and supported by the experts in data intensive workloads, these appliances are the gold standard for eliminating bottlenecks and maximizing application performance.

- **Start at 4U and Scale Up or Scale Out**

Mix Disk and Flash SSD for Hybrid Performance

- **Flexible Access for File: (IB, Ethernet, OPA)**

Integrates high-performance, low-latency, RDMA capable networks. Both EDR InfiniBand and 100Gbps (or 40Gbps) Ethernet plus OPA

- **Flexible Access for Block: (IB, FC)**

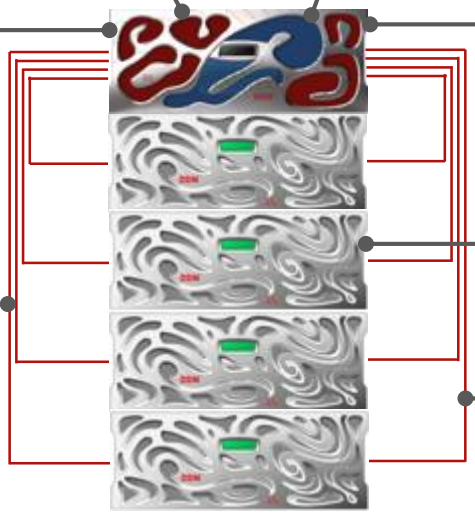
Block systems attach via FC or IB to external servers

SFA7990 OVERVIEW

SFAOS
Mirrored, battery-backed write-back cache
ReACT Cache Management
Fully Declustered Data Placement for rapid rebuilds

EMBEDDED FILESYSTEM APPLIANCES
EXAScaler®
GRIDScaler®

Flexible Connectivity
Block Target: 4x EDR IB OR 8x FC16
Embedded File System: 4 x EDR IB /40/100 GbE OR 2 x OPA



Full SAS Performance for all configurations
Balanced SAS Architecture provides saturation level performance for all expansion options

Redundantly Connected Enclosures
All configurations feature directly attached, redundant links to SFA7990

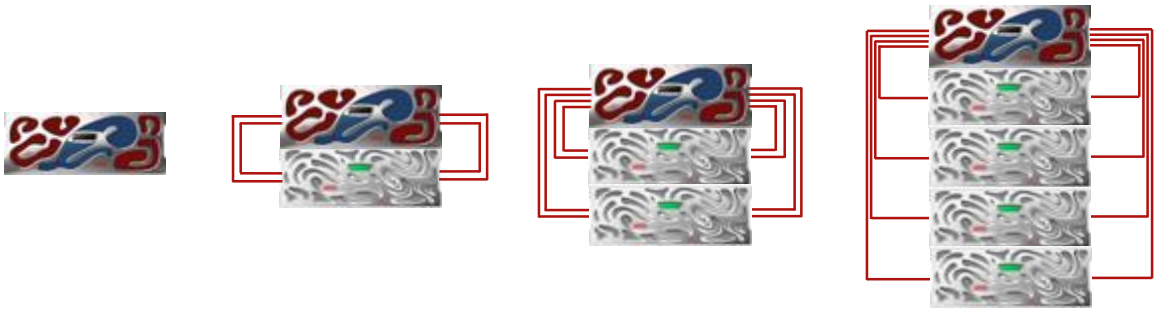
High Density HDD and SAS SSD
SS9012 Enclosures and SFA7990 house up to 90 3.5” Devices.

High Performance SAS
Backend Bandwidth supported by 12Gbps SAS Links from all enclosures to SFA7990

SFA7990 CONFIGURATIONS

	Configuration	SFA7990 Only	+ 1 SS9012	+2 SS9012	+4 SS9012
Supported Solutions	SFA (BLOCK) ES7990 GS7990				
	3.5” HDD or SAS SSD Slots	90	180	270	450
	最大物理容量	1.26PB	2.52PB	3.78PB	6.30PB

Expansion Options

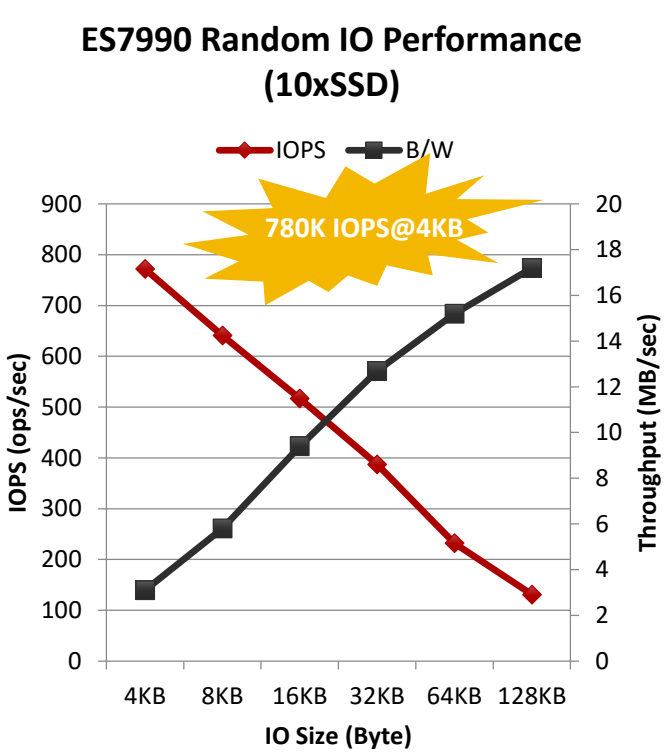


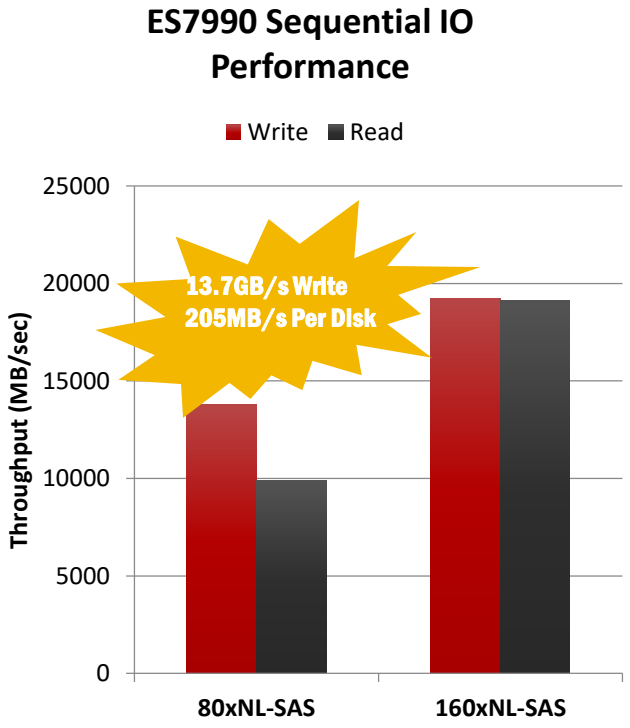
※ 最大物理容量は14TB NL-SAS HDDをSS9012に実装した場合

ES7990 WORLD'S MOST EFFICIENT HYBRID STORAGE

OPTIMIZED FOR SSD, HDD CAPACITY, MULTICORE & MULTI GPU

ES7990 20GB/s

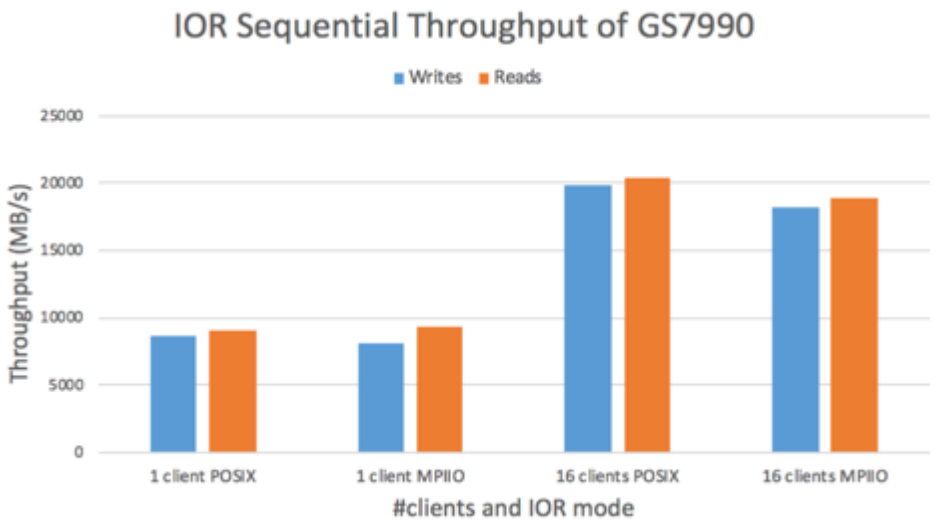




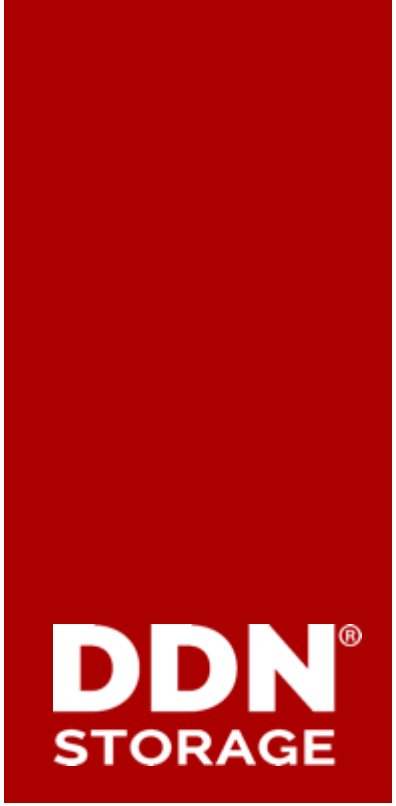
GS7990 WORLD'S MOST EFFICIENT HYBRID STORAGE

OPTIMIZED FOR SSD, HDD CAPACITY, MULTICORE & MULTI GPU

GS7990 20GB/s



* Cached



SFA200/400NV

WORLD'S MOST EFFICIENT HPC FLASH

OPTIMIZED FOR NVMe IOPS, MULTICORE & MULTI GPU

SFA200NV

Block 20GB/s
Embedded 20GB/s

- **Start at 2U and Scale-out**

All NVMe Flash, high density, extreme performance

- **Flexible Access for File: (IB, Ethernet, OPA)**

Integrates high-performance, low-latency, RDMA capable networks. EDR InfiniBand and 100Gbps (or 40Gbps) Ethernet and OPA

- **Flexible Access for Block: (IB, FC)**

Block systems attach via FC or IB to external servers

WORLD'S MOST EFFICIENT HPC FLASH

OPTIMIZED FOR NVMe IOPS, MULTICORE & MULTI GPU

SFA400NV

Block 40GB/s
Embedded 40GB/s

- **Start at 2U and Scale-out**

All NVMe Flash, high density, extreme performance

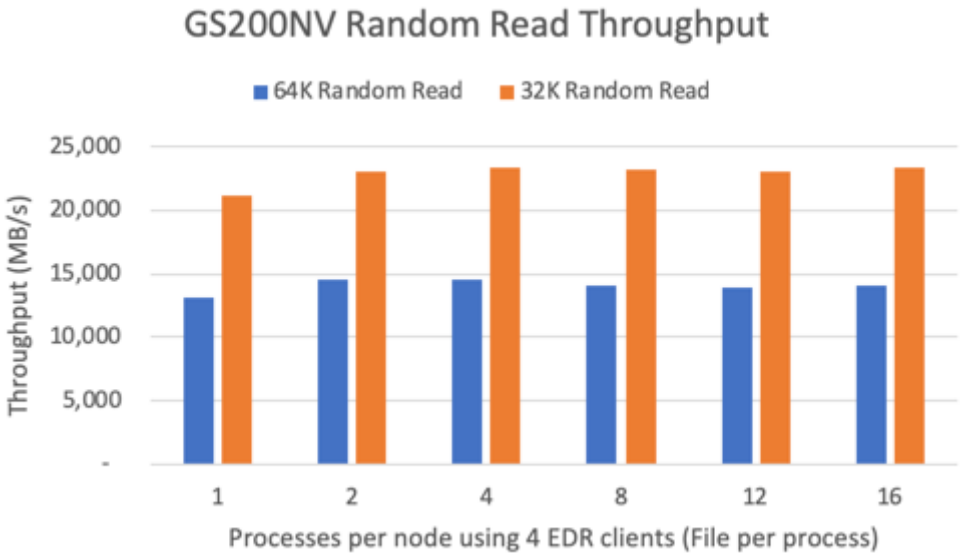
- **Flexible Access for File: (IB, Ethernet, OPA)**

Integrates high-performance, low-latency, RDMA capable networks. EDR InfiniBand and 100Gbps (or 40Gbps) Ethernet and OPA

GS200NV : SCALABLE FILE PLATFORMS

OVER 23GB/s THROUGHPUT EVEN FOR RANDOM READ WORKLOADS

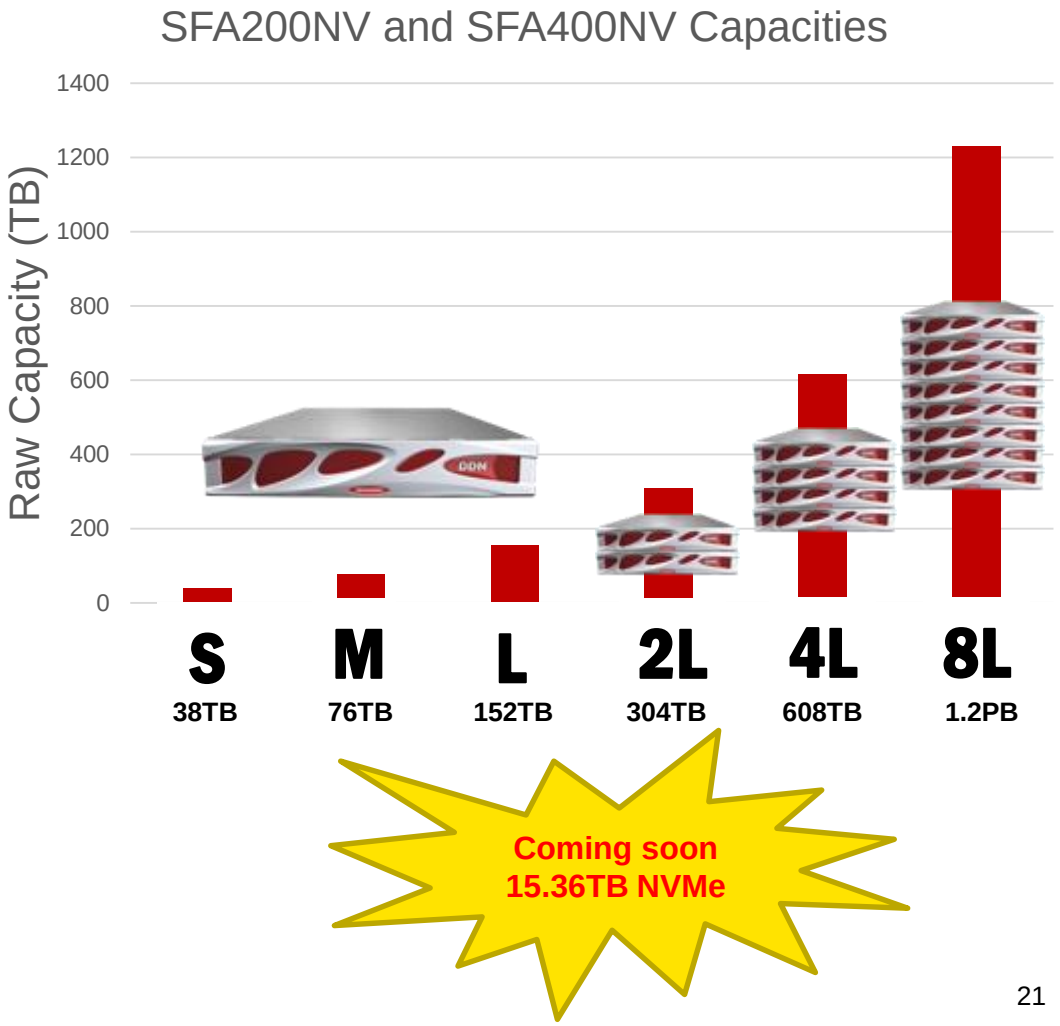
- Saturation level random read performance with small I/O transfers
- Ideal for Analytics platforms requiring Enterprise Features

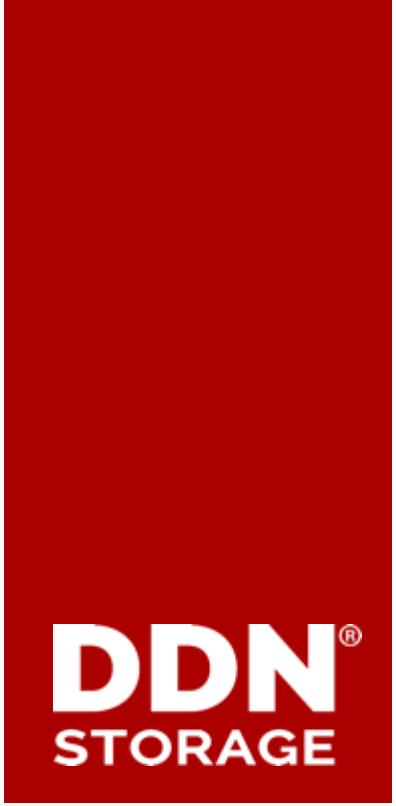


SFA200NV AND SFA400NV | SCALABLE FILE PLATFORMS

Start with a complete All NVMe Flash, Parallel File System from DDN in just 2RU

- Choose your starting option from as small as 38TB or up to 152TB in just 2 RU
- Scale-out with NVMe to Petabytes of Flash and TB/s of performance





A3I

Accelerated, **A**ny-scale **AI** Solutions From DDN

DDN PARALLEL FILE STORAGE APPLIANCES FOR AI AND DL

AI200

All NVME parallel file storage appliance
Optimized for the most intensive workloads
30TB, 60TB, 120TB in 2RU

AI7990

Hybrid parallel file storage appliance
Optimized for capacity, intermix with flash
1PB in 4RU

23GB/s, 395K IOPS • 100Gb Ethernet or EDR InfiniBand • Start with 1, Scale out and mix

DDN

DDN® A³I®

POWERED BY NVIDIA

DDN

DDN® A³I®

with Hewlett Packard Enterprise

DDN A³I WITH DGX-1 SCALABLE REFERENCE ARCHITECTURES

A³I with single DGX-1

A³I with quad DGX-1

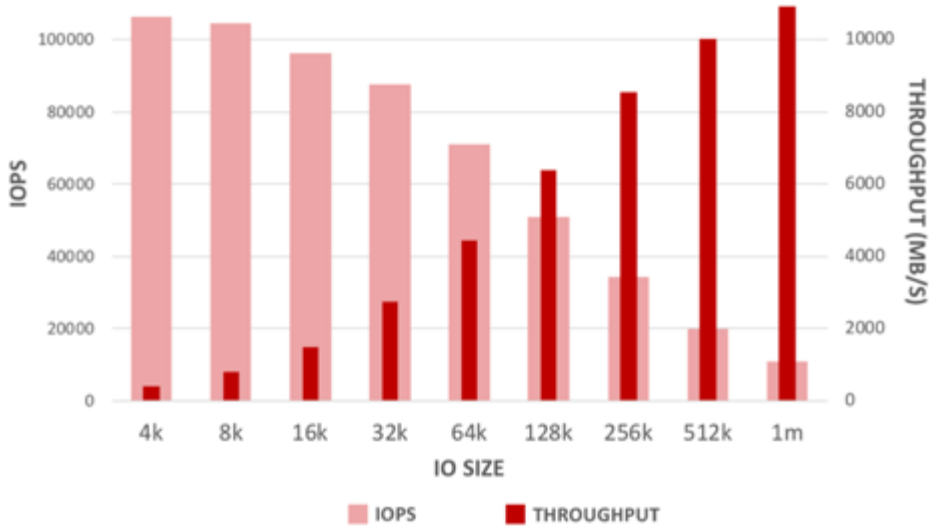
A³I with DGX-1 POD

DDN A³I Solutions: Performance to a single container

Over 100K IOPs and 10GB/s to 1 Container

EXAScaler client demonstrates over 10GB per second to a ***single*** container and over 100K IOPs.

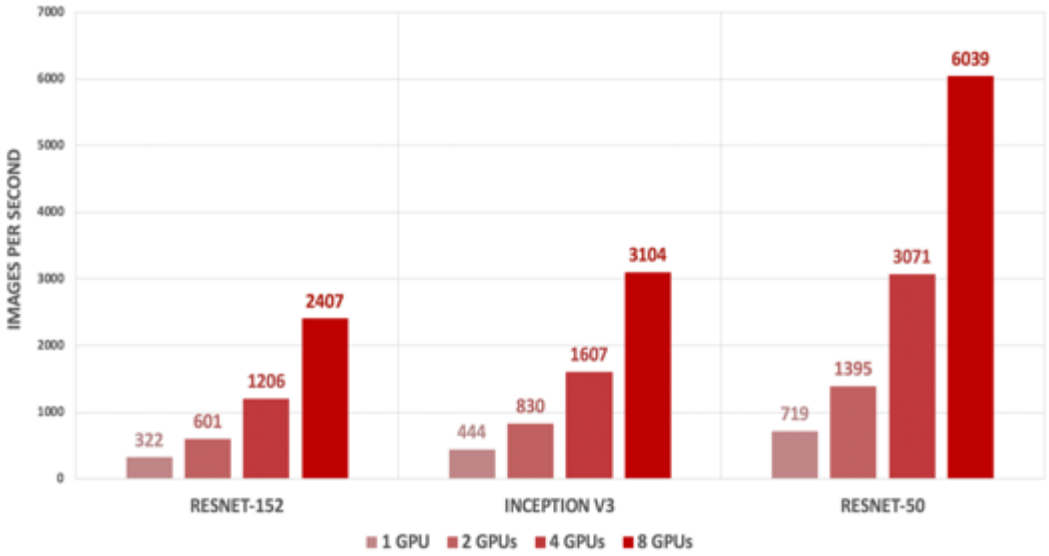
Typical Deep Learning Codes will perform IOs around 128K IOs – and see around 6GB/s with this pattern.



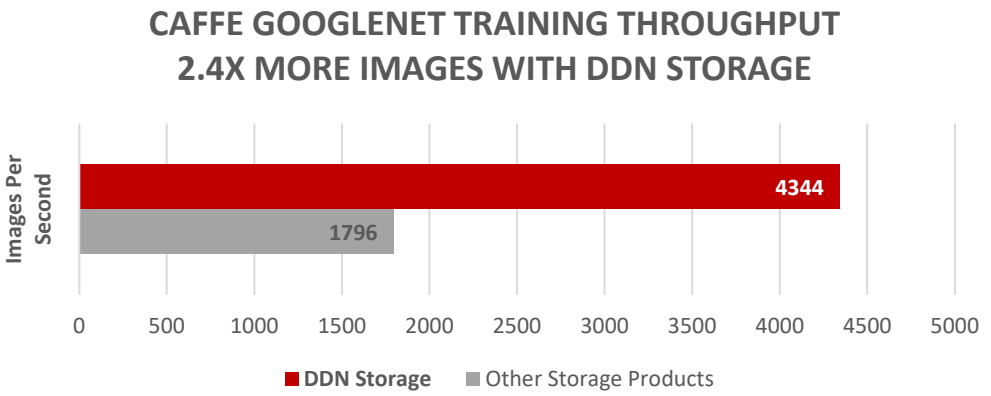
DDN A³I Solutions: Fully-optimized with widely-used DL frameworks

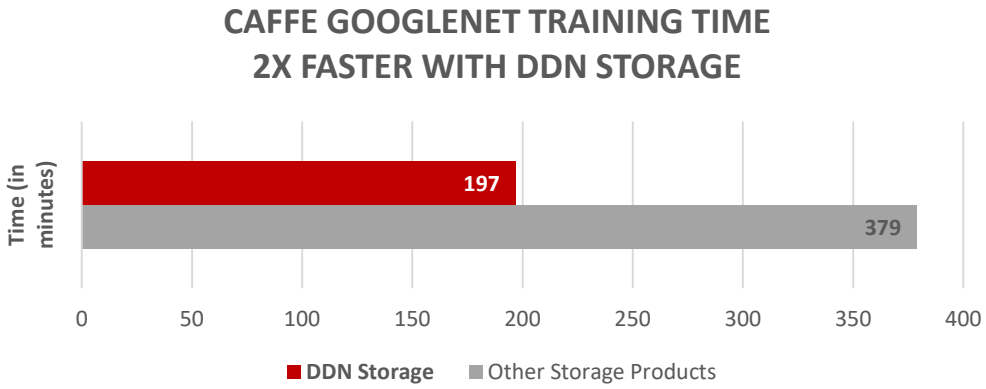
60% Faster - Consistent, Linear AI Performance

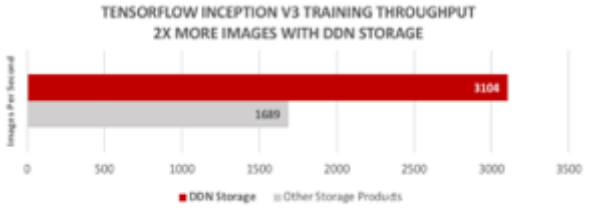
AI benchmarking of AI Frameworks with popular test sets demonstrates images per second rates 60% higher than competing systems

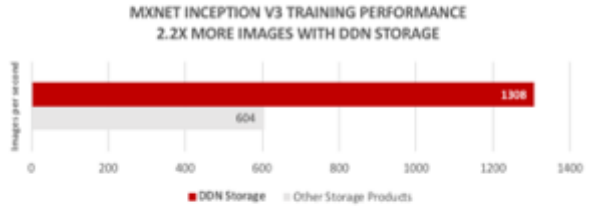


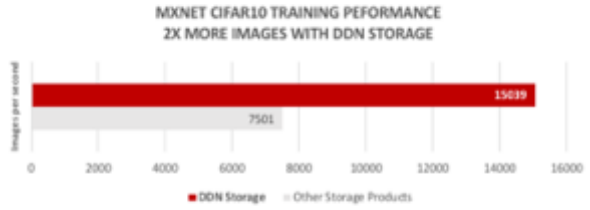
AI Benchmarking: DDN enables and accelerates

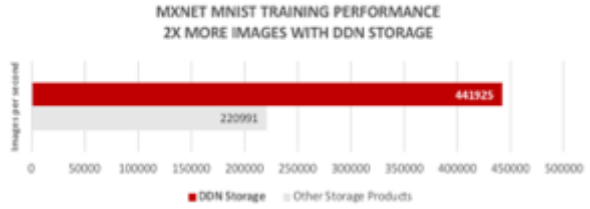












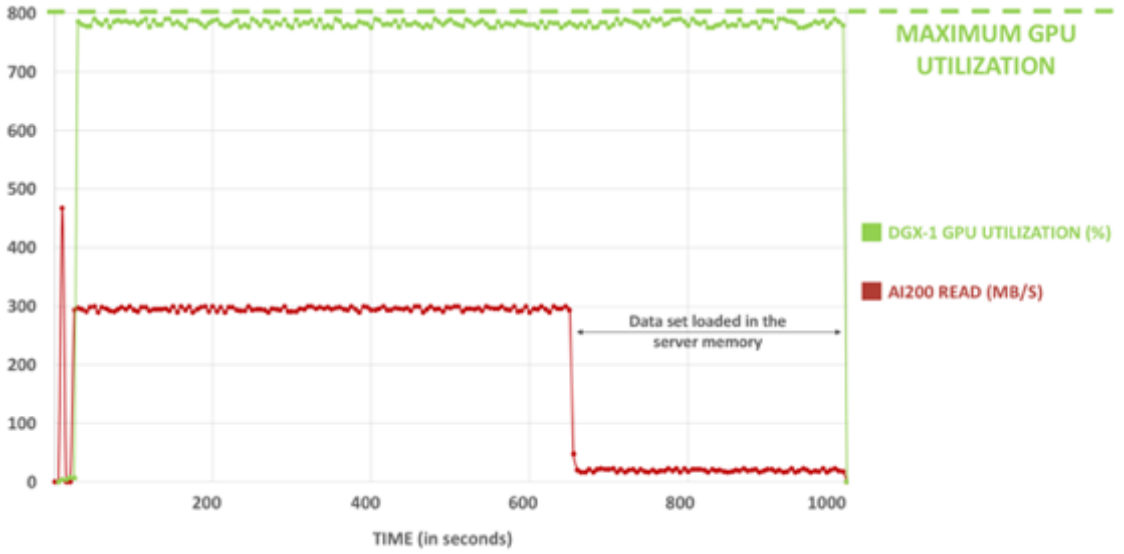
AI Benchmarking: IO during TensorFlow training

Full GPU Saturation, Maximum Productivity

DDN AI200 delivers a steady high-throughput, low-latency stream of data to the DL training application

Achieving full GPU saturation ensures every compute cycle is put to productive use

DDN true end-to-end parallelism enables acceleration at scale for distributed computing with multiple multi-GPU nodes



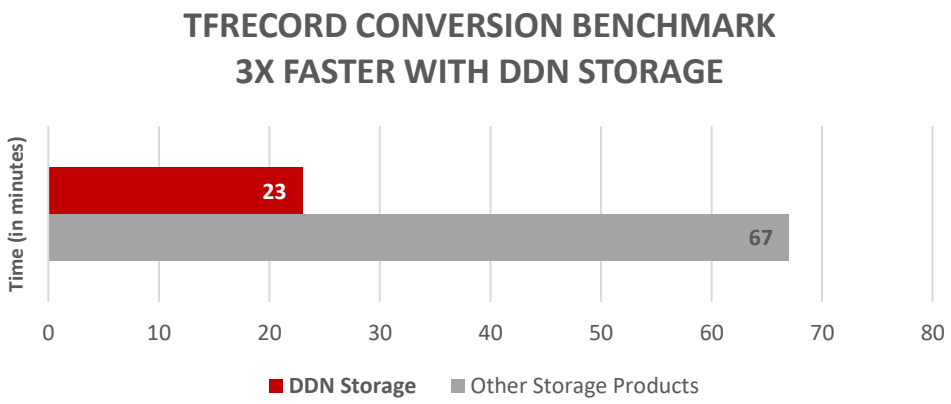
AI Benchmarking: Optimized TensorFlow Data Sets

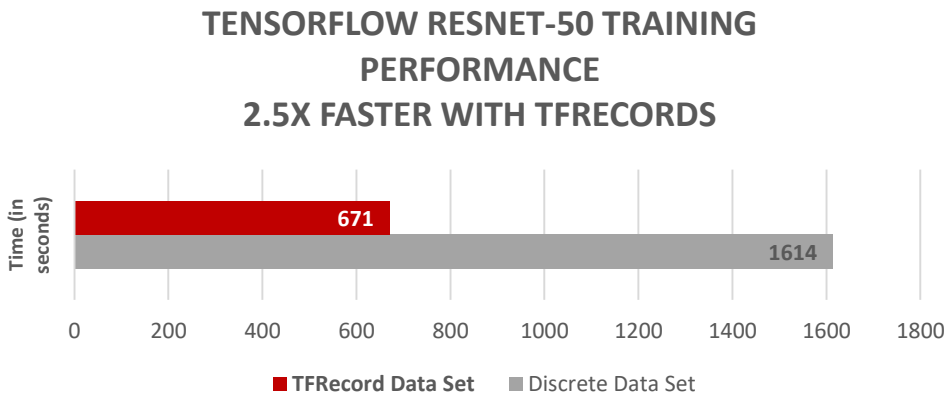
Optimized Data Sets, Accelerated Training

Convert discrete data and metadata asset collections into series of streamlined binary files for TensorFlow

Maximize DL training and validation efficiency while minimizing data management and preparation overhead

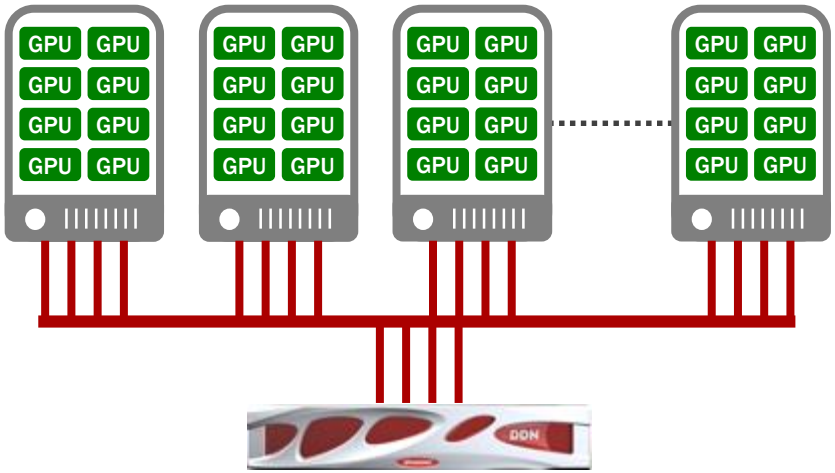
Achieve more accurate results by eliminating manual data shuffle bias



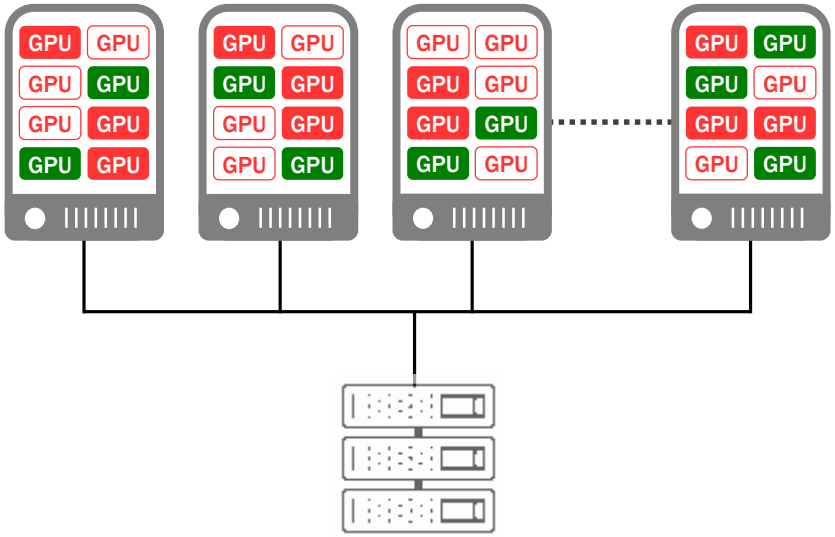


DDN PARALLEL ARCHITECTURE AND PROTOCOL vs NFS

FULL GPU SATURATION
MAXIMUM AI PRODUCTIVITY

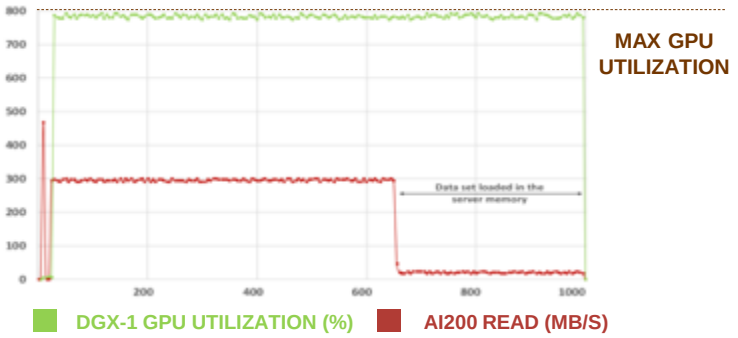


GPU STARVATION
CRIPPLED AI PERFORMANCE



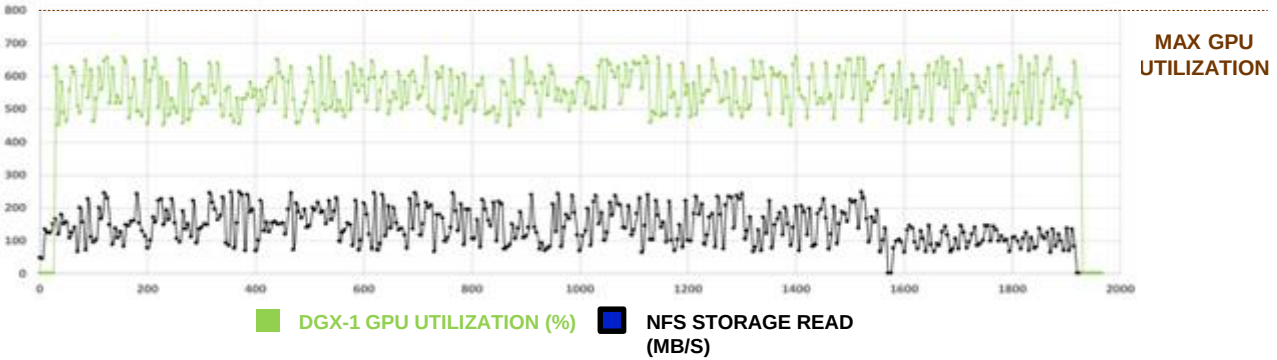
DDN PARALLEL ARCHITECTURE AND PROTOCOL vs NFS

DDN FULL GPU SATURATION



2X FASTER
Consistent DL performance
Achieves full GPU utilization
Efficient loading of data set

NFS GPU STARVATION



DDN A³I SOLUTIONS PERFORMANCE ENHANCEMENTS

AI200 and AI400 Optimizations for AI

Performance improvements for Random Read up delivers up to to around 750K IOPs per RU for AI400 and over 22GB/s per RU

Thank You!

Keep in touch with us

Team-JPSales@ddn.com

Tokyu Bancho Bldg. 8F
6-2 Yonbancho Chiyoda-ku,
Tokyo 102-0081

@ddn_limitless

+81-3-3261-9101
+81-3-3261-9140

company/datadirect-networks

