3S-SERIE

PNY®

AI OPTIMISED STORAGE FOR DEEP LEARNING ACCELERATION AND INFERENCE

Maximizing your DGX systems throughout your AI journey

NOW WITH FULL A100 COMPATIBILITY AND PERFORMANCE

50% FASTER TRAINING

Real life deep learning projects show **a massive 50% improvement** in training times when compared to other solutions. Excellent performance with the standard storage synthetic benchmarks, with bandwidth, latency and IOPS, leaving others behind.

50% LOWER COST

Cost and affordability are a key design focus. By removing the need for expensive storage controllers, costs are dramatically reduced, and more of your investment is spent on GPU and NVMe resource providing greater productivity and ROI.

100% SCALABILITY

With up to 360TB within 2U and a massive 150TB within the 1U, even solutions starting at 30TBs have can scale in stages that suit your project.

FLEXIBLE FAULT TOLERANCE

PNY appliances feature variable RAID protection which provides various levels of resilience. With the recommended configuration being RAID5, with RAID 1, 10, 5 and 6 all being supported.

EXTENDING NVIDIA'S DGX RESOURCE

NVIDIA's DGX range has helped shape the AI landscape and changed future possibilities. However, the DGX range has limited internal space for NVMe flash storage, an essential element for performance and overall capability.

PNY AI Optimised Storage Server creates **a central pool of ultra-low latency NVMe** which can be shared amongst one or multiple DGX servers. Providing each DGX with the ideal level of resource without the need for upfront over investment.

Simply connected via NVIDIA compatible InfiniBand / Ethernet, the unique RDMA protocol ensures the NVMe resource is seen and performs as if it were internal to the DGX.

BLISTERING PERFORMANCE FOR AI WORKFLOWS, AND MORE BUDGET FOR GPU'S

Today's AI servers consume and analyse data at much higher rates than many traditional storage solutions can deliver, resulting in low GPU utilisation and dramatically extending training times decreasing productivity.

PNY, NVIDIA's global partner, has been developing **solutions from the ground up for AI workloads and optimised for the NVIDIA DGX range of AI appliances.** Delivering ultra-low latency and tremendous bandwidth at a price which allows more investment to be made on GPU resource and less on expensive, slower storage.

Ensuring your project's funds are better spent and your team are more productive, by taking full advantage of the DGX compatibility.

START SMALL. SCALE ONLY WHEN NEEDED

With many new A.I. projects and inference solutions requiring only limited amounts of storage, the PNY 3S-Storage range starts from 30TB, while still delivering full performance.

With a 1U delivering up to 150TB and a 2U capable of housing a massive 360TB, starting small and scaling as needed is simple. And should a project require larger capacities, additional expansion units are available.

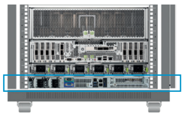
FEATURES AND SPECIFICATIONS

TYPICAL NVIDIA DGX A100 DIRECT ATTACHED TO 1U

NVIDIA DGX A100

PNY 3S-STORAGE

Rear view



Dual HDR/200Gbe
Cable direct to A100

OR

CLUSTERED DGX SOLUTION

Multiple
NVIDIA DGX A100

PNY 3S-STORAGE

↓

Fully **compatible with NVIDIA networking solutions (like Mellanox MQ87XX series)** providing a simple yet blisteringly fast sharable solution.

There is no need for multiple storage nodes or controllers, everything needed is contained and automated within the **single appliance**.

TECHNICAL SUPPORT

With full support and a range of on-site options, you can choose the package most effective for your organisation or project.

CLEAR AND SIMPLE USER INTERFACE

SIMPLE CONNECTIVITY

Connections to either a single deep learning server or multiple servers couldn't be simpler. The interfaces on PNY appliances allows for direct connection to a single server or utilising NVIDIA Mellanox low latency switching technology to connect to multiple servers and create a cluster.

CLUSTER FILESYSTEM FOR MULTIPLE DGX COMPATIBILITY

As solutions evolve and multiple DGX's are used, the storage requires a sharable filesystem,PNY's software platform providing a simple yet blisteringly fast sharable solution design for with NVIDIA's RDMA implementation. Meaning no need to worry as your project evolves and grows.

FLEXIBLE DESIGN, FLEXIBLE SOLUTIONS

	PNY 3S-1050	PNY 3S-2450
Drives	Up to 10 NVMe	Up to 24 NVMe
Storage Capacity (raw)	Up to 150TB	Up to 360TB
Storage Capacity (protected R5)	Up to 135TB	Up to 345TB
Expansion units	Up to 5	Up to 5
Total expansion capacity	Up to 750TB	Up to 2PB
RAID Levels	0, 1, 10, 5, 6 (50, 60)	0, 1, 10, 5, 6 (50, 60)
Bandwidth	23GB/s	23GB/s
IOPS	3M IOPS	3M IOPS
Latency	35ms	35ms
Connectivity	2 x QSP56 HDR InfiniBand /200Gb Ethernet	2 x QSP28 EDR InfiniBand /100Gb Ethernet
Protocols (block and file)	NVMe-oF (InfiniBand or Ethernet) NFS (InfiniBand or Ethernet – RDMA or TCP)	NVMe-oF (InfiniBand or Ethernet) NFS (InfiniBand or Ethernet – RDMA or TCP)
Block to NFS	Auto conversion	Auto conversion
Software Licence	3 years	3 years
Form Factor	1U - 660mm deep	2U - 710mm deep
Warranty	3 years	3 years
Support	3 years 24/7 Software Support	3 years 24/7 Software Support
Support options	Advanced Unified NVIDIA POD monitoring – reseller PS	Advanced Unified NVIDIA POD monitoring - reseller PS

CONTACT YOUR PNY REPRESENTATIVE AT

PNYPRO@PNY.EU

FOR FURTHER CONFIGURATION OPTIONS

GRAPHICS | HPC | AI

PNY TECHNOLOGIES EUROPE
PNYPro@pny.eu
T: +33 (0)5 56 13 75 75

More information: www.pny.eu
Follow us: @PNYProSolutions