

**PNY**®

# AI & DATA CENTER SOLUTIONS





# BENEFITS OF PNY DATA CENTER SOLUTIONS



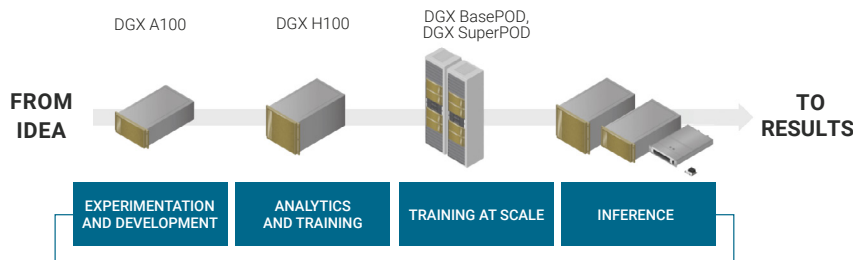
Artificial intelligence has become the go-to approach for solving difficult business challenges. Whether improving customer service, optimizing supply chains, extracting business intelligence, or designing cutting-edge products and services across nearly every industry, AI gives organizations the mechanism to realize innovation. And as a pioneer in AI infrastructure, PNY Data Center Solutions provide the most powerful and complete AI platform for bringing these essential ideas to fruition.

## ONE SYSTEM FOR ALL AI INFRASTRUCTURE

AI INFRASTRUCTURE RE-IMAGINED, OPTIMIZED, AND READY FOR ENTERPRISE AI-AT-SCALE

Flexible AI Infrastructure that adapts to the pace of enterprise:

- One universal building block for the AI data center
- Any workload on any node : any time
- Limitless capacity planning with predictability great performance with scale



## AI DRIVES REAL BUSINESS SUCCESS

Customer services, like Pinterest™, use deep learning for image recognition, curating rich custom experiences for its users. Financial businesses like PayPal™ are using NVIDIA GPU-accelerated deep learning for fraud detection.

Consulting firm Accenture™'s R&D arm and other businesses are using deep learning to detect Internet security threats. Healthcare benefits the most of the application of AI and deep learning.

This sweeping change in the landscape of data-driven business is taking place across all industries. It's changing how farmers herd their animals and leading us to rethink transportation. The first self-driving public transit system is already up and running, giving us a glimpse into futuristic cities that will exist within our lifetimes.

**ARGONNE NATIONAL LABORATORY**  
World's First DGX A100 Supercomputer  
Fighting COVID-19

24-node Cluster of DGX A100 Systems, 192 A100 GPUs,  
Mellanox High-Speed Low-Latency Network Fabric,  
120 PetaFLOPS of AI Computing Power for Scientific Research.



## PNY KEY SELLING POINT

PNY provides the most comprehensive computing, storage and networking portfolio for Datacenters. Its vast expertise and experience enable to support the challenges of the Datacenter coming from a broad range of industries such as Healthcare, Retail, Higher Education and more.

With its constant focus on innovation, quality, reliability, and support of excellence, the company helps its customers to empower and manage their infrastructure in the most technical and financial efficient ways.



**LOGISTICS STRENGTH AND ADAPTABILITY**



**LOCAL TECHNICAL SUPPORT**

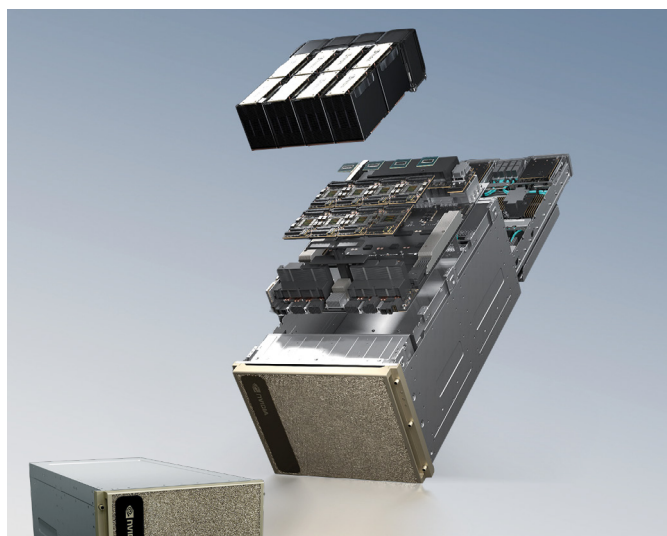


**EXPERIENCE & EXPERTISE**

# NVIDIA® DGX™ SYSTEMS

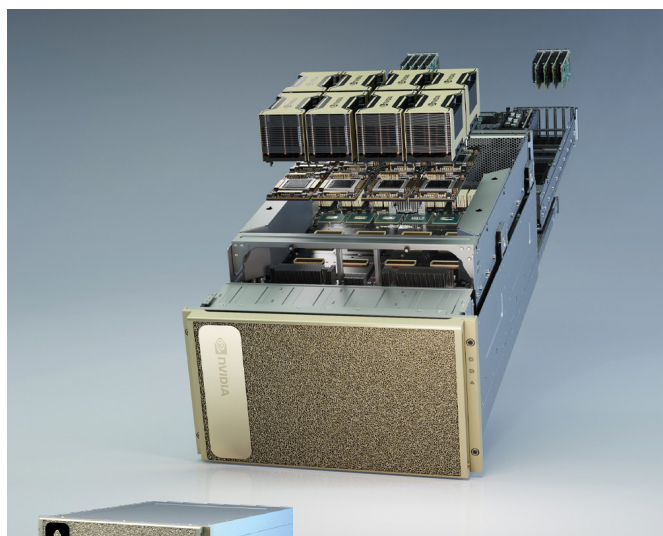
SETTING THE BAR FOR ENTREPRISE AI INFRASTRUCTURE

NVIDIA DGX™ systems are purpose-built to meet the demands of enterprise AI and data science, delivering the fastest start in AI development, effortless productivity, and revolutionary performance for insights in hours instead of months.



## NVIDIA DGX™ H100

The cornerstone of your AI Center of excellence.  
Break through the barriers to AI at scale.  
Leadership-Class Infrastructure on your terms.



## NVIDIA DGX™ A100

AI Training, Inference, and Analytics  
The third generation of the world's most advanced AI system, unifying all AI workloads.

## System Specifications

	NVIDIA DGX H100	NVIDIA DGX A100
GPUS	8x NVIDIA H100 Tensor Core GPUs	8x NVIDIA A100 80 GB GPUs
GPU Memory	640 GB total	640 GB total
Performance	32 petaFLOPS FP8	5 petaFLOPS AI 10 petaOPS INT8
NVIDIA® NVSwitch™	4	6
System Power Usage	~10.2kW max	6,5 kW max
CPU	Dual x86	Dual AMD Rome 7742, 128 cores total, 2.25 GHz (base) 3.4 GHz (max boost)
System Memory	2 TB	2 TB
Networking	4x OSFP ports serving 8x single-port NVIDIA ConnectX-7 VPI 400 Gb/s InfiniBand or 200 Gb/s Ethernet 2x dual-port NVIDIA ConnectX-7 VPI 1x 400 Gb/s InfiniBand 1x 200 Gb/s Ethernet	8x Single-Port Mellanox ConnectX-6 VPI 200 Gb/s HDR InfiniBand 2x Dual-Port Mellanox Connect X- 6 VPI 10/25/50/100/200 Gb/s Ethernet
Management network	10 Gb/s onboard NIC with RJ45 50 Gb/s Ethernet optional NIC Host baseboard management controller (BMC) with RJ45	1 Gb/s onboard NIC with RJ45 Host baseboard management controller (BMC) with RJ45
Storage	OS: 2x 1.92TB M.2 NVME drives Internal Storage: 30 TB (8x 3.84 TB) U.2 NVMe drives	OS: 2x 1.92TB M.2 NVME drives Internal Storage: 30 TB (8x 3.84 TB) U.2 NVMe drives

# NVIDIA DGX BASEPOD™ & SUPERPOD™

BUILDING ENTERPRISE AI  
REQUIRES A SOLUTION  
THAT'S ENTERPRISE-READY

A solution that eliminates the complexity of designing AI infrastructure, integrated into your existing environment without upheaval, and that enables a simplified, fast deployment experience. NVIDIA DGX BasePODs, SuperPODs and PNY AI Cluster solution delivers the world's best platform for enterprise AI innovation, from AI infrastructure building block to Data Center as a product. Start Small, Scale Predictably in Response to Business Demand.



## NVIDIA DGX BasePOD™

A reference architecture that incorporates best practices for compute, networking, storage, power, cooling, and more, in an integrated AI infrastructure design built on NVIDIA DGX, from 2 to 10 nodes.

## NVIDIA DGX SuperPOD™

AI data center infrastructure platform that enables IT to deliver performance without compromise.

DGX SuperPOD offers leadership-class accelerated infrastructure and agile, scalable performance for the most challenging AI and high-performance computing (HPC) workloads.

Become an AI Scaler, from 31 to 127 DGX H100, and more.



## Example for a 127-node DGX SuperPOD configuration

Technology	Component	Description
<a href="#">NVIDIA DGX H100 system with eight 80 GB H100 GPUs</a>	Compute nodes	1016 DGX H100 SXM4 GPUs <ul style="list-style-type: none"> <li>• 89,6 TB of HBM3 memory</li> <li>• 4064 petaFLOPS AI performance</li> <li>• 254 TB System RAM</li> <li>• 3.8 PB local NVMe</li> <li>• 900 GBps NVLink bandwidth per GPU</li> <li>• 7.2 TBps total NVSwitch bandwidth per node</li> </ul>
<a href="#">NVIDIA Quantum QM9700 NDR 400 Gbps InfiniBand</a>	Compute fabric	Rail-optimized, full fat-tree network with eight NDR400 connections per system
	Storage fabric	Fat-tree with eight NDR400 connections per system. The fabric is optimized to match peak performance of the configured storage array
<a href="#">NVIDIA SN4600 switch</a>	In-band management network	64 port 100 Gbps Ethernet switch providing high port density with high performance
<a href="#">NVIDIA SN2201 switch</a>	Out-of-band management network	48 port 1 Gbps Ethernet switch leveraging copper ports to minimize complexity
<a href="#">NVIDIA Unified Fabric Manager Enterprise Edition</a>	Compute/storage fabric management	NVIDIA UFM combines enhanced, real-time network telemetry with AI powered cyber intelligence and analytics to manage scale-out InfiniBand data centers
<a href="#">NVIDIA Base Command Manager</a>		Cluster management for DGX SuperPOD
<a href="#">NVIDIA AI Enterprise</a>	DGX SuperPOD software stack	Best-in-class development tools and frameworks for the AI practitioner and reliable management and orchestration for IT professionals
<a href="#">Magnum IO</a>		The NVIDIA MAGNUM IO™ enables increased performance for AI and HPC
<a href="#">NVIDIA NGC</a>		The NGC catalog provides a collection of GPU-optimized containers for AI and HPC
<a href="#">Slurm</a>	User environment	Slurm is a classic workload manager used to manage complex workloads in a multi-node, batch-style, compute environment



# NVIDIA NETWORKING SOLUTIONS

UNLOCKING SYSTEM PERFORMANCE

Networking technology is at the heart of a new “agile computing” model. NVIDIA® Networking Solutions unlock the full potential of your Data Center.



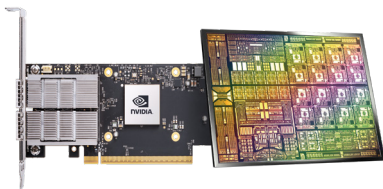
## ACCELERATING COMPUTE AND CLOUD DATA

Modern data centers run distributed applications across a composable infrastructure consisting of disaggregated GPUs, CPUs, memory and storage that require flexible, high-speed connections. NVIDIA networking accelerates and offloads data transfers to ensure compute resources never “go hungry” due to lack of data or bandwidth. Smart adapters and switches reduce latency, increase efficiency, enhance security, and simplify data center automation so applications run faster.

## SMART HIGH PERFORMANCE SOLUTIONS

There is a variety of smart high performance solutions, multi-core processors, network adapters, switches, cables, software and silicon. End-to-End High-Speed Ethernet and InfiniBand Interconnect Solutions.

### Complete Infiniband and Ethernet portfolio



#### SmartNICs and DPU

Ethernet Network Adapters, InfiniBand Host Channel Adapters (HCA) and Data Processing Unit (DPU) for accelerated networking and security for the most advanced cloud and AI workloads.

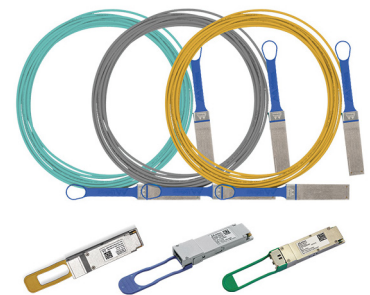
Up to 400Gb/s connectivity per port, 370 million messages/sec, sub 0.6 usec latency, PCIe Gen3.0, Gen4.0 and Gen5.0 support.



#### Switches

Best-in-class Ethernet switch portfolio, in Edge, Leaf, and Spine form factors, ranging in speeds from 10 to 400Gbps. adapters, switches, cables, software and silicon.

NVIDIA InfiniBand switch systems deliver the highest performance and port density for HPC AI, Web 2.0, big data, cloud infrastructure and enterprise data center.



#### Interconnect

LinkX InfiniBand DACs, AOCs and transceivers meet or exceed all the IBTA industry standards for QDR, FRD, EDR, HDR and NDR products.

LinkX Ethernet DACs, AOCs and Optical Transceivers meet or exceed all of the IEEE 802.3xx industry standards for 1G, 10G, 25G, 50G, 100G, 200G and 400G products.



FOR DISCOVERING ALL OUR RECOMMENDATIONS:  
Please visit our [PNY Networking Solutions](https://pny.eu/networkingsolutions/) on:  
[HTTPS://PNY.EU/NETWORKINGSOLUTIONS/](https://pny.eu/networkingsolutions/)

# PNY ADVANTAGE

## “PNY PROVIDES UNSURPASSED SERVICE AND COMMITMENT TO ITS DATA CENTER CUSTOMERS”

- 20 Years expertise selling NVIDIA GPU Solutions
- Strong alliances with technological suppliers
- Dedicated head count for Sales, Marketing and Support
- Local Pre and post sales support
- Direct tech support hotlines
- Pre-sales tools, support and configuration assistance
- Dedicated Field Application Engineers, added-value on site installation
- Published product support and training materials
- Advanced replacement options for mission-critical deployments
- Long product life cycles and availability
- Loyalty partner channel programs
- Dedicated support programs
- Strong logistic and operation abilities
- Equipment loan for strategic opportunities
- PNY LAB: technological centre to support development of AI, HPC and VDI Solutions

## MARKETS WE SERVE



HPC



Data Center



AI/Machine Learning



Clouds



Storage



Security



Telecom



Media & Entertainment

## CONTACT

### PNY Technologies Europe

Zac du Phare  
 9 rue Joseph Cugnot - BP 40181  
 33708 Mérignac Cedex, France

Tel: +33 (0)5 56 13 75 75

### PNY Technologies GmbH

Schumanstraße 18a  
 52146 Würselen  
 Germany

Tel: +49 (0)2405/40848-0

### PNY Technologies Middle East Fze

Jafza View 19 308, Jebel Ali Free Zone,  
 PO Box 263897,  
 Dubai

Tel: +97148814966

e-mail: [SALES@PNY.EU](mailto:SALES@PNY.EU)

[WWW.PNY.EU](http://WWW.PNY.EU)