PNY®

AI & DATA CENTER SOLUTIONS

NVIDIA®

BENEFITS OF PNY DATA CENTER SOLUTIONS

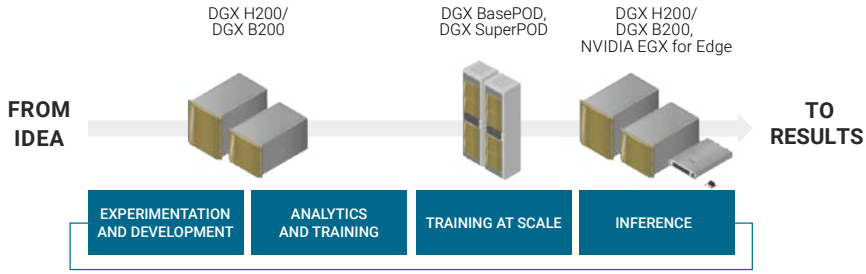
Artificial intelligence has become the go-to approach for solving difficult business challenges. Whether improving customer service, optimizing supply chains, extracting business intelligence, or designing cutting-edge products and services across nearly every industry, AI gives organizations the mechanism to realize innovation. More recently, Generative AI has significantly accelerated the integration of AI across various industries by utilizing AI and Machine Learning algorithms that use existing content to learn and generate new content, such as text, images, music, audio, and videos. It is applied in art, medicine, video games, finance, music, and many other. As a pioneer in AI infrastructure, PNY Data Center Solutions provide the most powerful and complete AI platform for bringing these essential ideas to life. By incorporating generative AI, PNY enhances its capabilities, enabling businesses to harness the full potential of AI-driven.

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



Pioneering Europe AI innovation

Designed to shape Europe's AI future, the Scaleway Supercalculator is equipped with NVIDIA H100 Tensor Core GPUs, a NVIDIA Quantum-2 InfiniBand network platform, and high-performance DDN storage. This machine scales to hundreds or thousands of nodes, effectively addressing the most significant challenges of the next generation of AI applications.

The iliad Group is launching Scaleway's NVIDIA DGX SuperPOD for AI and cloud services, including for its own telecommunication operator, Free, to develop generative AI on many use-cases, such as advanced chatbots for customer service. Additionally, the Kyutai laboratory launched by Iliad, CMA-CGM and Schmidt Futures, will benefit from the Nabu supercomputer. Kyutai aims to spread its progress across the entire AI ecosystem, including the scientific community, developers, companies, and society at large. This extends to startups like Mistral AI, which aims to develop a European LLM model.

PNY Technologies has played a pivotal role in this project, leading the way in AI innovation. As a key player, PNY Technologies supplied, installed, and leveraged its expertise in the NVIDIA SuperPOD infrastructure, building on enduring partnerships, especially with AI NVIDIA leader.

SCALEWAY SUPERCALCULATOR set up with premium NVIDIA AI technologies

Cluster of 127 NVIDIA DGX H100 Tensor Core, 1016 GPUs NVIDIA H100, up to 4021,3 PFLOPS, 1,8 PB of a3i DDN storage, InfiniBand 400 Gb/s.

Scaleway

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

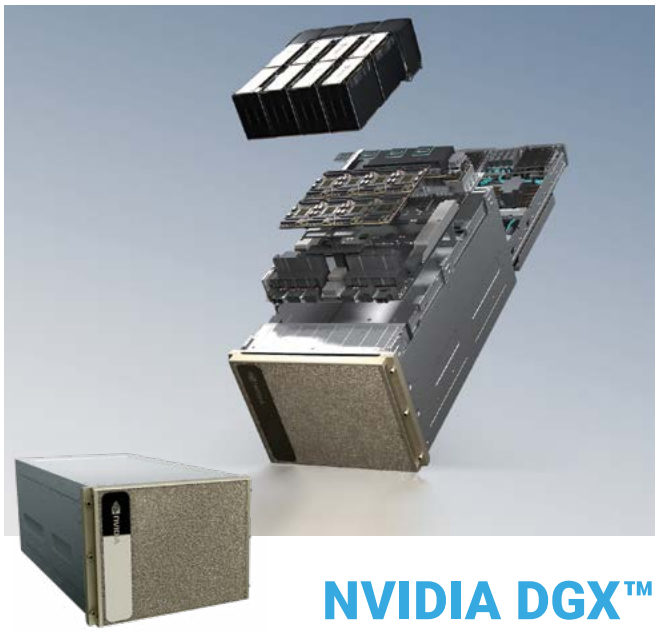
LOCAL TECHNICAL SUPPORT

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™ H200

The gold standard for AI infrastructure

Expand the frontiers of business innovation and optimization with NVIDIA DGX™ H200. DGX H200 is the AI powerhouse that's the foundation of NVIDIA DGX SuperPOD™ and DGX BasePOD™.

NVIDIA DGX™ B200

The foundation for your AI center of excellence

Unified AI platform for develop-to-deploy pipelines for businesses of any size at any stage in their AI journey.

System Specifications

	NVIDIA DGX H200	NVIDIA DGX B200
GPUS	8x NVIDIA H200 Tensor Core GPUs	8x NVIDIA Blackwell GPUs
GPU Memory	1128GB total	1440GB total
Performance	32 petaFLOPS FP8	72 petaFLOPS training and 144 petaFLOPS inference
NVIDIA® NVSwitch™	4	-
System Power Usage	~10,2kW max	~14,3kW max
CPU	Dual Intel® Xeon® Platinum 8480C Processors 112 Cores total, 2.00 GHz (Base), 3.80 GHz (Max Boost)	2 Intel® Xeon® Platinum 8570 Processors 112 Cores total, 2.1 GHz (Base), 4 GHz (Max Boost)
System Memory	2TB	Up to 4TB
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	4x OSFP ports serving 8x single-port NVIDIA ConnectX-7 VPI Up to 400Gb/s InfiniBand/Ethernet 2x dual-port QSFP112 NVIDIA BlueField-3 DPU Up to 400Gb/s InfiniBand/Ethernet
Management network	10 Gb/s onboard NIC with RJ45 50 Gb/s Ethernet optional NIC Host baseboard management controller (BMC) with RJ45	10 Gb/s onboard NIC with RJ45 100 Gb/s dual-port ethernet NIC Host baseboard management controller (BMC) with RJ45
Storage	OS: 2x 1,92TB M.2 NVME drives Internal Storage: 30TB (8x 3.84 TB) U.2 NVMe drives	OS: 2x 1,92TB M.2 NVME drives Internal Storage: 30TB (8x 3.84 TB) U.2 NVMe drives
Software	NVIDIA AI Enterprise: Optimized AI Software NVIDIA Base Command™: Orchestration, Scheduling and Cluster Management Operating system: DGX OS / Ubuntu / Red Hat Enterprise Linux / Rocky	NVIDIA AI Enterprise: Optimized AI Software NVIDIA Base Command™: Orchestration, Scheduling and Cluster Management Operating system: DGX OS / Ubuntu

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 and SuperPODs solution delivers the world's best platform for entreprise 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 16 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 163 DGX H200, and more.

Complete Lifecycle of Expertise

Ever-Improving Advanced Infrastructure Software

Powered by NVIDIA Base Command

NVIDIA DGX SuperPOD™ with DGX GB200 systems

NVIDIA DGX SuperPOD™ with DGX GB200 systems is purpose-built for training and inferencing trillion-parameter generative AI models. Each liquid-cooled rack features 36 NVIDIA GB200 Grace Blackwell Superchips–36 NVIDIA Grace CPUs and 72 Blackwell GPUs–connected as one with NVIDIA NVLink™. Multiple racks connect with NVIDIA Quantum InfiniBand to scale up to tens of thousands of GB200 Superchips.

Maximize Developer Productivity

Massive Supercomputing for Generative AI

Built on NVIDIA Grace Blackwell

FOR DISCOVERING ALL OUR RECOMMENDATIONS:

Please visit our PNY Networking Solutions on:

HTTPS://PNY.EU/NETWORKINGSOLUTIONS/

NVIDIA GB200 Grace Blackwell Superchip

A NEW CLASS OF AI SUPERCHIP

The NVIDIA GB200 Grace Blackwell Superchip connects two high-performance NVIDIA Blackwell Tensor Core GPUs and an NVIDIA Grace CPU using the NVIDIA NVLink®-C2C interconnect that delivers 900 GB/s of bidirectional bandwidth to the two GPUs.

Built with 208 billion transistors, more than 2,5x the amount of transistors in NVIDIA Hopper GPUs, and using TSMC's 4NP process tailored for NVIDIA, Blackwell is the largest GPU ever built. NVIDIA Blackwell achieves the highest compute ever on a single chip, 20 petaFLOPS.

5th Generation NVLink
Scales to 576 GPUs

Secure AI
Full Performance Encryption & TEE

AI Superchip
208B Transistors

2nd Gen Transformer Engine
FP4/FP6 Tensor Core

Decompression Engine
800 GB/sec

RAS Engine
100 % In-System Self-Test

40 petaFLOPS of AI performance

864GB of fast memory

16TB/s of HBM

3,6TB/s of NVLink bandwidth

PNY, YOUR KEY SOLUTION PARTNER IN THE AI ECOSYSTEM

PNY Technologies is leading the way in AI innovation, setting new standards in the industry. As a key partner, PNY offers end-to-end solutions in the AI ecosystem, including compute, storage, networking, and complete software stack. With our deep expertise, we can define what are the best applications and storage strategies that will help you on your journey to AI.

BENEFITS OF AI STORAGE

Data Protection and Security

AI storage solutions come with robust security features. This includes encryption, access controls, and auditing capabilities.

Flexibility and Adaptability

AI storage solutions are designed to be flexible and adaptable to different types of AI workloads.

Advanced Analytics and Data Insights

Optimized AI storage can help support those analytics, and more accurate, insightful, and responsive to real-time data.

Automation of Routine Tasks

AI storage can automate many routine storage management tasks, such as provisioning, data migration, and backup.

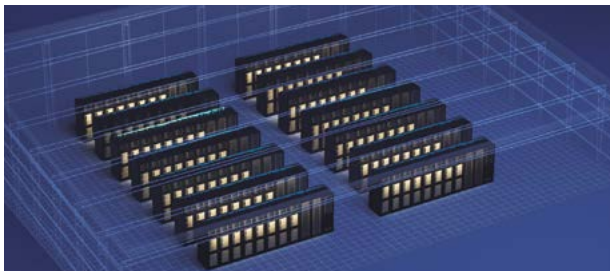
FOR MORE INFORMATION:
Visit: WWW.PNY.EU

ACMK0833

UNLOCK THE FULL POTENTIAL OF YOUR DATACENTER

Networking for AI Data Centers

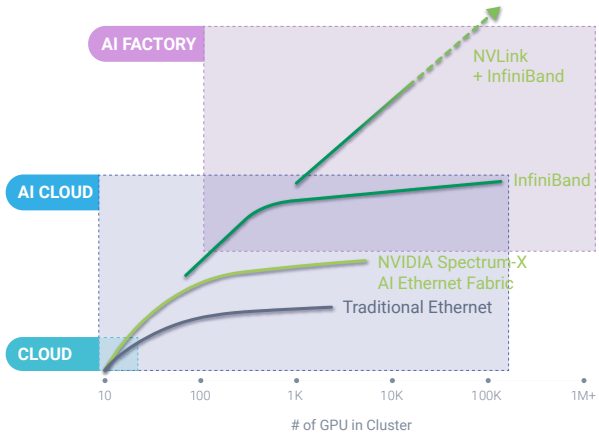
TWO TYPES OF AI DATACENTERS



AI Factories

AI Cloud

THE NETWORK DEFINES THE DATA CENTER



Cloud
- Multi-tenant
- Variety of small-scale workloads
- Traditional Ethernet network can suffice

Generative AI Cloud
- Multi-tenant
- Variety of workloads including larger scale Generative AI
- Traditional Ethernet network for North-South traffic
- NVIDIA Spectrum-X Ethernet for AI fabric (East West)

AI Factories
- Single or few users
- Extremely large AI models
- NVIDIA NVLink and InfiniBand gold standard for AI fabric

Complete Infiniband and Ethernet portfolio

SmartNICs and DPU

Up to 400Gb/s the Ethernet Network Adapters, InfiniBand Host Channel Adapters (HCA) and Data Processing Unit (DPU) provide accelerated networking and security for the most advanced cloud and AI workloads.

Switches

The NVIDIA Spectrum Ethernet switch family spans 1GbE to 800GbE, ideal for AI fabrics with NVIDIA GPUs and end-to-end cloud data center networks. NVIDIA Quantum InfiniBand switches offer 400Gb/s throughput, smart acceleration, and a robust architecture for high-performance computing, AI, and hyperscale cloud infrastructures, with reduced cost and complexity.

Interconnect

Explore NVIDIA's comprehensive line of 100G/200G/400G/800G InfiniBand and Ethernet interconnects, offering low latency, low power, and high reliability for AI and accelerated computing. NVIDIA LinkX® Cables connect within Quantum and Spectrum™ architectures for switch-to-switch, top-of-rack to ConnectX® smart network adapters, and BlueField® DPUs in servers and storage.

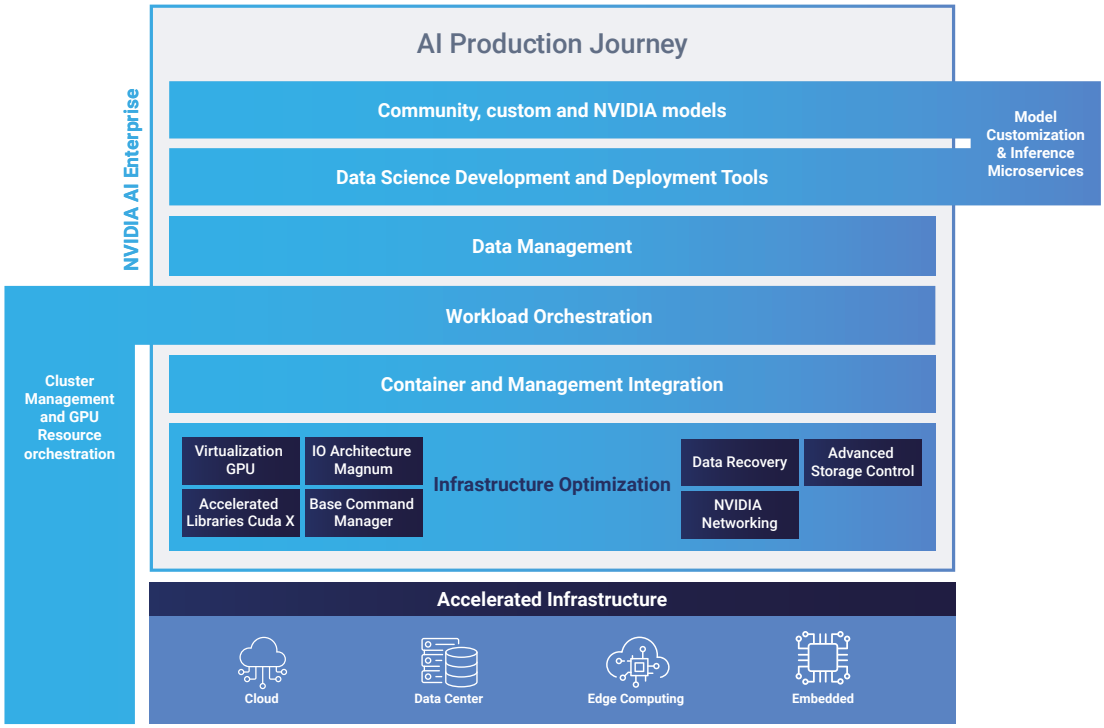
Spectrum™-X

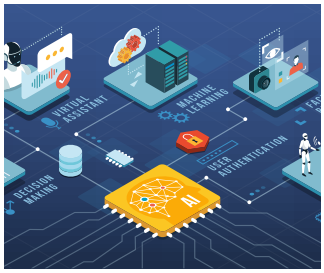
First ethernet networking platform for AI, leveraging the close integration of NVIDIA Spectrum™-4 Ethernet switch with the NVIDIA Bluefield-®3 data processing unit DPU, Spectrum-X achieves peak performance in AI, machine learning, natural language processing, various industry applications.

PNY SOFTWARE SOLUTIONS

OPTIMIZE AI PRODUCTION

At the forefront of AI production, orchestrate massive data processing, optimize hardware resource utilization, and secure every exchange. Unlock AI's potential with cutting-edge software solutions designed for efficiency, scalability, and security.





NVIDIA AI Enterprise

Deploy AI from pilot to production with NVIDIA AI Enterprise platform. Streamline workflows, optimize model performance with microservices, and ensure secure, seamless transitions from prototype to production.

Container and Management Integration

Start with Ubuntu supported by Canonical and get trusted open source and run your entire Artificial Intelligence/Machine Learning lifecycle on a single integrated stack.

Workload Orchestration

Transform your AI infrastructure to accelerate development, optimize resources, and lead the race in AI innovation with our workload orchestration solutions.

Consulting & Development

Embrace and accomplish AI projects with consulting and customized development services, delivering innovative and effective solutions.

BU Software

AI Solutions Aggregator

PNY delivers optimized applications for complex and evolving AI technologies through an aggregation of partners within the same technology landscape, leveraging NVIDIA GPU solutions to ensure high performance and reliability.

AI System & Service Advisor

Empower every stage of the customer lifecycle to deliver significant value, accelerate decision-making, and facilitate fast and easy deployment with our integrated solutions and services, spanning from audit to installation.

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 40 240 240

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: +971 4 8814966

e-mail: **SALES@PNY.EU**

WWW.PNY.EU