Smarter technology for all

Lenovo | NVIDIA

Accelerate AI Training and Inference with Lenovo's AI-Ready Data Center Infrastructure

Artificial Intelligence in the Modern Enterprise

From boosting customer engagement to accelerating health diagnoses to enabling more accurate predictive maintenance, artificial intelligence is transforming every industry. Enterprises know that they need to embrace this change or risk losing out to competitors. But they face challenges with using existing data center infrastructure to power modern AI applications.

IT Challenges for Data Center AI

- The ever-increasing size of AI models, along with the amount of data needed to train them, creates intense resource demands that existing IT compute and networking capabilities cannot meet.

- Modern AI applications are complex, involving many different components that must be orchestrated to work together to obtain useful results. This makes it challenging to manage, monitor, and scale these applications.

- The proliferation of ad-hoc cloud usage or isolated compute silos, especially those based on specialized hardware, results in a lack of standardization that's difficult to manage at scale.
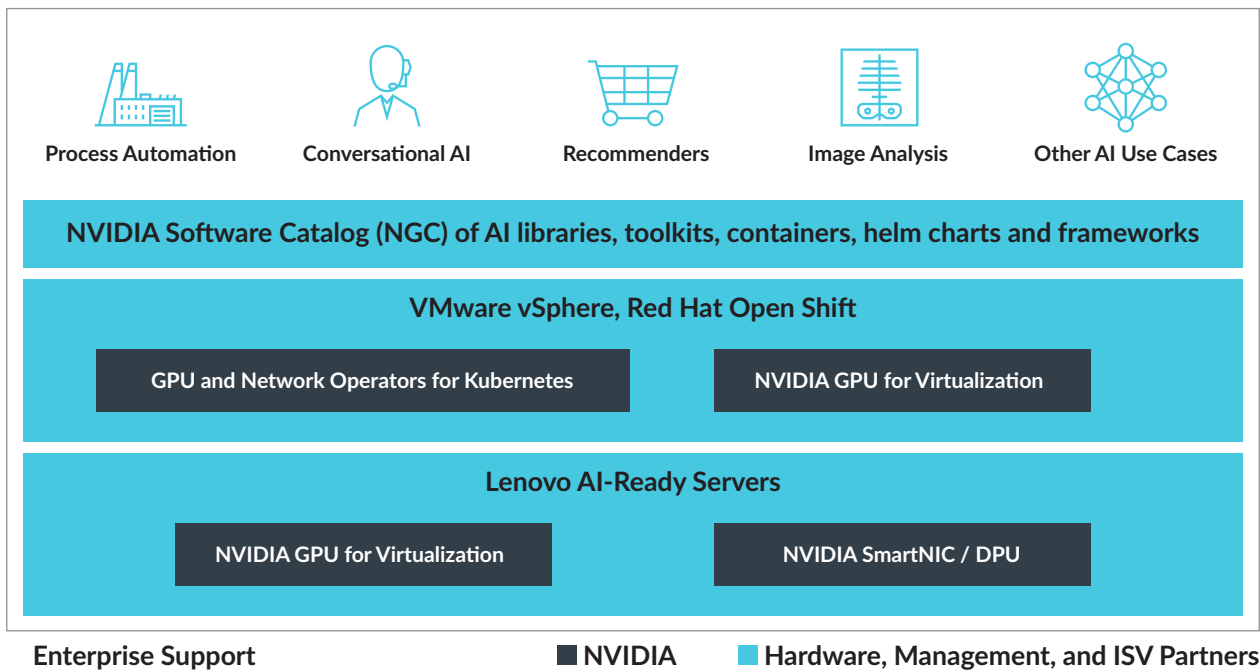
Power your company's AI transformation with an AI-ready vs IT-led infrastructure strategy from Lenovo and NVIDIA that enables faster ROI, increased productivity, and streamlined manageability.

KEY APPLICATIONS / PLATFORMS

- TensorFlow
- PyTorch
- NVIDIA® TensorRT™
- NVIDIA Triton™ Inference Server
- Industry-specific AI frameworks from NVIDIA

PROOF POINTS

- NVIDIA A100 Tensor Core GPU broke eight AI performance records in the MLPerf Training benchmark and was up to 237X faster than CPU in the MLPerf Inference benchmark.

Process Automation | Conversational AI | Recommenders | Image Analysis | Other AI Use Cases

NVIDIA Software Catalog (NGC) of AI libraries, toolkits, containers, helm charts and frameworks

VMware vSphere, Red Hat Open Shift

GPU and Network Operators for Kubernetes

NVIDIA GPU for Virtualization

Lenovo AI-Ready Servers

NVIDIA GPU for Virtualization

NVIDIA SmartNIC / DPU

Enterprise Support   ■ **NVIDIA**   ■ **Hardware, Management, and ISV Partners**

Lenovo AI-Ready Infrastructure and the NVIDIA EGX Platform for Data Center AI

Lenovo, VMware and NVIDIA have partnered to unlock the power of AI for every business by delivering an end-to-end enterprise platform optimized for AI workloads. This fully integrated validated and supported Lenovo AI solution delivers best-in-class AI software, the NVIDIA AI Enterprise Suite, optimized and exclusively certified for the industry's leading virtualization platform, VMware vSphere®. Running on Lenovo AI-Ready infrastructure, this platform accelerates the speed at which developers can build AI Infrastructure. High-performance data analytics enables organizations to scale modern workloads on the same VMware vSphere infrastructure they've already invested in delivering enterprise-class manageability, security and availability.

- Lenovo is the leading tier-1 OEM in the latest MLPerf benchmarks with NVIDIA A100 Tensor Core GPUs. Lenovo was the first OEM to support the NVIDIA A100.

LENOVO NVIDIA-CERTIFIED SYSTEMS

Lenovo AI-Ready servers coupled with NVIDIA GPUs and networking are validated for performance, manageability, security, and scalability and are backed by enterprise-grade support from NVIDIA and Lenovo. With an NVIDIA-Certified System, enterprises can confidently choose performance-optimized hardware solutions to power their accelerated computing workloads—both in smaller configurations and at scale.

Learn more about accelerated servers at nvidia.com/certified-systems

Key Technologies

GPU

NVIDIA Tensor Core technology has brought dramatic speedups to AI training and inference operations, bringing down training times from weeks to hours and providing massive acceleration to inference.

vGPU

NVIDIA virtual GPU (vGPU) software products combine the management and security benefits of server and desktop virtualization with the performance benefits of GPU acceleration.

Networking

NVIDIA ConnectX® SmartNICs and NVIDIA BlueField® data processing units (DPUs) provide a host of software-defined hardware engines for accelerating networking and security. These enable the best of both worlds: best-in-class AI training and inference performance, with all the necessary levels of enterprise data privacy, integrity, and reliability.

Kubernetes

The NVIDIA GPU Operator and NVIDIA Network Operator standardize and automate the deployment of all the necessary components for provisioning Kubernetes clusters. Using Helm charts, containers, and continuous integration and continuous delivery (CI/CD), organizations can deploy updated AI software effortlessly in minutes.

Multi-instance GPU

Multi-Instance GPU (MIG), available on select GPU models, allows one GPU to be partitioned into multiple independent GPU instances. With MIG, infrastructure managers can standardize their GPU-accelerated infrastructure while ensuring optimal use of their GPU resources for all stages of AI deployment, from development to training to production inference.

NGC SOFTWARE CATALOG

The NVIDIA NGC™ software catalog is the hub for performance- optimized deep learning and machine learning applications. NGC simplifies building, sharing, and deploying software, so enterprises can gather insights faster and deliver business value sooner.

BENEFITS OF THE EGX PLATFORM

- **Reduced costs:** Record-setting performance and low total cost of ownership (TCO) enable enterprises to achieve their AI results quickly and efficiently.

- **Increased productivity:** Universal acceleration and a rich end-to-end solution stack enable organizations to maximize utilization and insight with a single infrastructure.

- **Enterprise management:** IT and DevOps integrations allow AI applications to be operated with existing infrastructure management frameworks.

	DEVELOPMENT	TRAINING	PRODUCTION
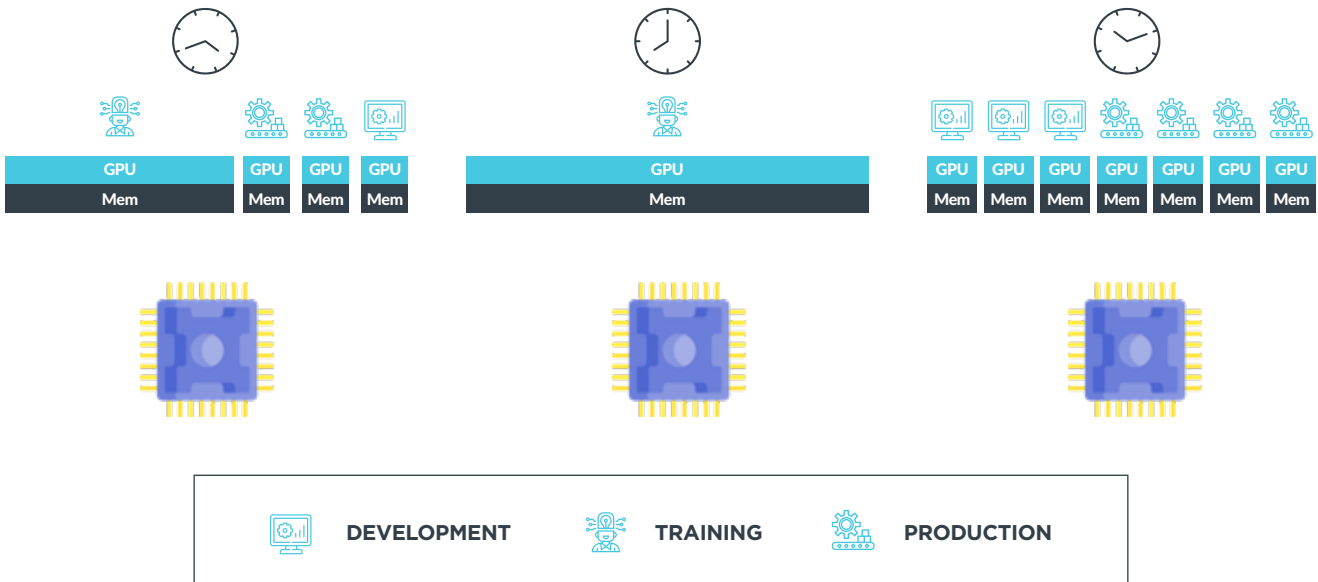
Why Lenovo

Focused on a bold vision to deliver smarter technology for all, Lenovo is developing world-changing technologies that create a more inclusive, trustworthy, and sustainable digital society. By designing, engineering and building the world's most complete portfolio of smart devices and infrastructure, we are also leading an Intelligent Transformation – to create better experiences and opportunities for millions of customers around the world. To find out more visit www.lenovo.com.

Lenovo | NVIDIA