LENOVO HYBRID AI ADVANTAGE™ WITH NVIDIA



QUICK START GUIDE

AI AMBITION IS GROWING. INFRASTRUCTURE HAS TO KEEP UP.

IN THIS GUIDE

Compare **Lenovo Hybrid AI Factory** with NVIDIA infrastructure options and identify the right fit based on workload demands, deployment priorities, and your existing IT environment.

AI is now a core business priority — and a practical engine for innovation, productivity, and growth.

As AI moves into production, workloads can vary dramatically in model size, data sensitivity, latency requirements, utilization, and business criticality. Those differences drive new infrastructure decisions. The right environment depends on what the workload needs to accomplish.

#1

Enhancing, innovating, or reinventing the business with AI is the #1 global business priority.[1]

96%

of organizations plan to increase AI investment over the next 12 months.[1]

NVIDIA

Smarter technology for all

Lenovo

AN INTEGRATED INFRASTRUCTURE FOUNDATION FOR AI

Lenovo Hybrid AI Advantage™ with NVIDIA combines Lenovo AI-ready devices, AI-optimized infrastructure, and AI factory services with NVIDIA accelerated computing, NVIDIA Spectrum-X™ networking, and NVIDIA AI Enterprise software to help organizations turn AI ideas into production outcomes. Lenovo Hybrid AI Factory with NVIDIA provides the infrastructure foundation, designed to support AI from development through enterprise scale. Lenovo Hybrid AI Factory with NVIDIA brings together:



SOFTWARE AND MODELS

NVIDIA AI Enterprise software, including NVIDIA NIM™ microservices, helps teams develop and deploy a broad choice of models — including open community and NVIDIA open models — across enterprise AI environments.



MANAGEMENT AND ORCHESTRATION

Allocate resources, manage workloads, and improve utilization.



COMPUTE

Accelerated computing matched to workload requirements.



NETWORKING

Bandwidth and latency for demanding AI workloads.



STORAGE AND DATA

Fast access to enterprise data and context.



POWER AND LIQUID COOLING

Innovative Lenovo Neptune® liquid cooling for high-density systems.

LENOVO AI SERVICES SUPPORT THE FULL LIFECYCLE

Plan → Integrate → Deploy → Operate → Optimize

RIGHT-SIZE YOUR AI INFRASTRUCTURE

CHOOSE VALIDATED INFRASTRUCTURE SOLUTIONS BASED ON WORKLOAD DEMANDS, DEPLOYMENT PRIORITIES, AND YOUR EXISTING IT ENVIRONMENT.

Lenovo Hybrid AI Factory with NVIDIA provides NVIDIA certified systems built with **Lenovo ThinkSystem** accelerated servers and **Lenovo ThinkAgile** hyperconverged infrastructure (HCI) solutions, helping organizations match infrastructure to AI requirements, accelerate time to value, and retain flexibility as model and workload needs evolve — without designing every environment from scratch.



AI AT EDGE AND BUSINESS LOCATIONS

A more compact enterprise configuration provides dedicated GPU resources for focused inference and AI development requirements.

TYPICAL FIT

- Focused production inference
- Departmental AI and smaller shared workloads
- AI development and testing
- Establishing dedicated AI capacity



AI AT DISTRIBUTED DATA CENTERS

As AI serves more users, agents, and applications, organizations need to scale infrastructure designed for greater GPU capacity and shared inference.

TYPICAL FIT

- Larger production inference environments
- Shared enterprise AI
- Higher concurrent demand
- More complex agents and applications



AI AT LARGE CENTRALIZED/HPC DATA CENTERS

Larger configurations support organizations with the most intensive AI requirements, including advanced model customization, training, and enterprise-scale workloads.

TYPICAL FIT

- Large-scale inference
- Advanced model customization
- Training and fine-tuning
- High-performance AI environments

DEDICATED GPU CAPACITY

Single node/smaller footprint

HIGHER GPU DENSITY

Scale-out networking
Multi-node expansion

HIGH-DENSITY ACCELERATED COMPUTE

Higher-performance networking
Demanding multi-node workloads

ELEVATE YOUR AI CAPABILITIES WITH COMPLEMENTARY WORKSTATION ENVIRONMENTS

Lenovo Workstations powered by NVIDIA RTX™ PRO and GB10 Grace Blackwell Superchip support local AI development, experimentation, prototyping, testing, and validation. They complement Lenovo Hybrid AI Factory configurations by providing dedicated resources for local secure development while supporting AI-driven professional workflows alongside shared enterprise infrastructure.

WHAT SHOULD YOU CONSIDER WHEN CHOOSING AI INFRASTRUCTURE?

USE THESE QUESTIONS TO IDENTIFY THE ENVIRONMENT THAT BEST FITS THE WORKLOAD.

As AI moves into production, the right infrastructure depends on more than model size. Use these questions to evaluate the demands of the workload and the environment around it.

- ☐ WHAT DOES THE WORKLOAD NEED TO DO?

  Consider model size and architecture, including open, proprietary, or custom models, along with inference or customization requirements, expected users, and workload complexity.

- ☐ WHERE DOES THE DATA NEED TO RESIDE?

  Account for data sensitivity, privacy, security, governance, and control requirements.

- ☐ HOW IMPORTANT ARE LATENCY AND PERFORMANCE?

  Determine latency requirements and how close compute should be to users and data.

- ☐ HOW MUCH CAPACITY WILL YOU NEED?

  Consider utilization, concurrency, GPU requirements, and how demand may change over time.

- ☐ WHAT ARE THE OPERATIONAL REQUIREMENTS?

  Factor in networking, storage, power, cooling, management, and the capabilities of your existing IT environment.

- ☐ WHAT DEPLOYMENT MODEL MAKES ECONOMIC SENSE?

  Evaluate infrastructure investment, utilization, token efficiency, operating costs, and the ability to expand efficiently.

Once the workload profile is clear, you select the development or production environment that aligns with your requirements.



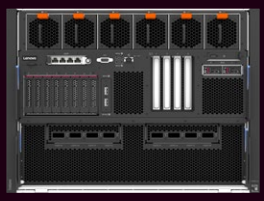
MOVE FASTER WITH LENOVO AI EXPRESS

Lenovo AI Express with NVIDIA accelerates your path to the Lenovo Hybrid AI Factory with NVIDIA in weeks, combining validated quick-start configurations, NVIDIA AI Enterprise software, and Lenovo services, with an order-to-ship goal of just 15 business days for eligible configurations.*

Choose the configuration aligned to your workload profile, then add the software and services your deployment requires — without designing the environment from scratch.

- Validated Lenovo AI building blocks flexible to your workloads
- Easily add NVIDIA AI Enterprise software
- Select from curated services from Lenovo and partners

Build your AI future without the long wait.

SMALL	MEDIUM	LARGE
FOCUSED AI INFERENCING	HIGHER-THROUGHPUT INFERENCING & AGENTIC AI	FULL-SCALE GENERATIVE AI MODELS
		
NVIDIA AI ENTERPRISE SOFTWARE		
Lenovo ThinkSystem SR650a V4 powered by 2x NVIDIA RTX 6000 PRO Blackwell Server Edition GPUs	**Lenovo ThinkSystem SR675 V3** powered by 8x NVIDIA RTX 6000 PRO Blackwell Server Edition GPUs	**Lenovo ThinkSystem SR680 V4** powered by NVIDIA HGX B300
NVIDIA ConnectX-7, NVIDIA Bluefield-3, and NVIDIA Spectrum-X Ethernet networking	NVIDIA ConnectX-8, NVIDIA BlueField-3, and NVIDIA Spectrum-X Ethernet networking	NVIDIA ConnectX-8, NVIDIA BlueField-3, and NVIDIA Spectrum-X Ethernet infrastructure
15 business days order to ship goal*	20 business days order to ship goal*	25 business days order to ship goal*

* For eligible Lenovo AI Express configurations and selected parts in non-restrictive markets. Clock starts after completion of order validation, payment/credit clearance, End User Certification (EUC), and applicable due diligence/export approvals.

A PARTNER YOU CAN TRUST TO BUILD, SCALE, AND OPERATE AI

Moving AI into production raises the stakes. Whether you are validating an initial use case, expanding AI across the enterprise, or supporting demanding production workloads, infrastructure decisions involve more than performance.

Organizations also need reliability, security, operational control, and a clear path for managing increasingly complex workloads. Lenovo helps customers assess where development and validation should occur and which workloads require production infrastructure for deployment and scale.

Lenovo combines validated AI infrastructure and lifecycle services with NVIDIA accelerated computing, networking, and enterprise AI software to help organizations deploy useful AI faster, preserve model choice, and operate with greater confidence.



BUILD ON VALIDATED INFRASTRUCTURE

Lenovo Validated Designs provide tested architectures to help reduce interoperability risk, while **Lenovo EveryScale** pre-integrates and tests hardware before delivery to simplify deployment.



PROTECT DATA AND AI WORKLOADS

Secure by design, Lenovo infrastructure and Lenovo XClarity One management solutions deliver Zero Trust security with hardware-level protection, automated resilience, and centralized visibility across your IT infrastructure. NVIDIA Confidential Computing helps protect sensitive data and models during inference on supported GPUs, while Lenovo Hybrid AI Services help address privacy, compliance, governance, and data-control requirements across AI environments.



SUPPORT ACROSS THE AI LIFECYCLE

Lenovo Hybrid AI Services assists with workload assessment, development and production environment planning, architecture and sizing, deployment, optimization, security, and ongoing management.



TRUSTED INFRASTRUCTURE

#1 in x86 server reliability for 12 consecutive years.[2]



RECOGNIZED AI INFRASTRUCTURE

Lenovo was named a Leader in the Integrated AI Infrastructure Systems quadrant of the 2026 ISG Provider Lens® AI-ready Infrastructure Solutions report.[3]



TRUSTED SUPPLY CHAIN

#5 in Gartner's 2026 Global Supply Chain Top 25.[4]

READY TO IDENTIFY THE RIGHT AI FACTORY FOR YOUR GOALS?

Book an AI Discover workshop today.

Sources

1 Lenovo, "CIO Playbook 2026: The Race for Enterprise AI," January 2026
2 ITIC, "ITIC 2025 Global Server Hardware, Server OS Reliability Report," February 2026
3 Information Services Group, "ISG Provider Lens: AI-ready Infrastructure Solutions," July 2026
4 Gartner, "2026 Global Supply Chain Top 25," June 2026

© Lenovo 2026. All rights reserved. v1.00 October 2026.

NVIDIA

Smarter technology for all

Lenovo