Building the Secure AI Factory: Confidential Computing Meets Lenovo Hybrid AI Infrastructure

Article

At NVIDIA GTC, the AI industry continues to demonstrate how artificial intelligence is moving rapidly from experimentation to production. Enterprises are no longer simply testing models, they are building AI factories capable of producing intelligence on a scale. However, a fundamental challenge remains: Trust.

Most enterprise data that organizations want to use with AI does not reside in the public cloud. It exists across on-premises systems, data centers, and legacy repositories, often containing highly sensitive information such as healthcare records, financial data, proprietary research, and intellectual property. As a result, organizations must balance innovation with strict requirements around data privacy, governance, and security.

This is where Confidential Computing and secure AI infrastructure become essential. The Trust Challenge in Enterprise AI

As enterprises deploy frontier models and agentic AI workloads, a complex trust dilemma emerges between three key stakeholders:

- Model providers who need to protect proprietary model weights and algorithms
- Infrastructure providers responsible for operating the compute platforms
- Data owners who must ensure sensitive enterprise data remains protected

Traditional computing environments leave data in use unencrypted, meaning sensitive information and model intellectual property can potentially be exposed during execution.

Confidential Computing addresses this challenge by enabling workloads to run inside hardware trusted execution environments (TEEs), where data and models remain cryptographically protected even while being processed. This approach allows enterprises to deploy AI with strong assurances that sensitive information and data remain secure.

From AI Experiments to Enterprise AI Factories

As organizations move toward production AI systems, infrastructure requirements expand significantly. AI factories must support:

- Massive volumes of unstructured enterprise data
- High-performance GPU compute
- Scalable AI pipelines for training and inference
- Secure deployment of frontier models
- Consistent governance and data sovereignty

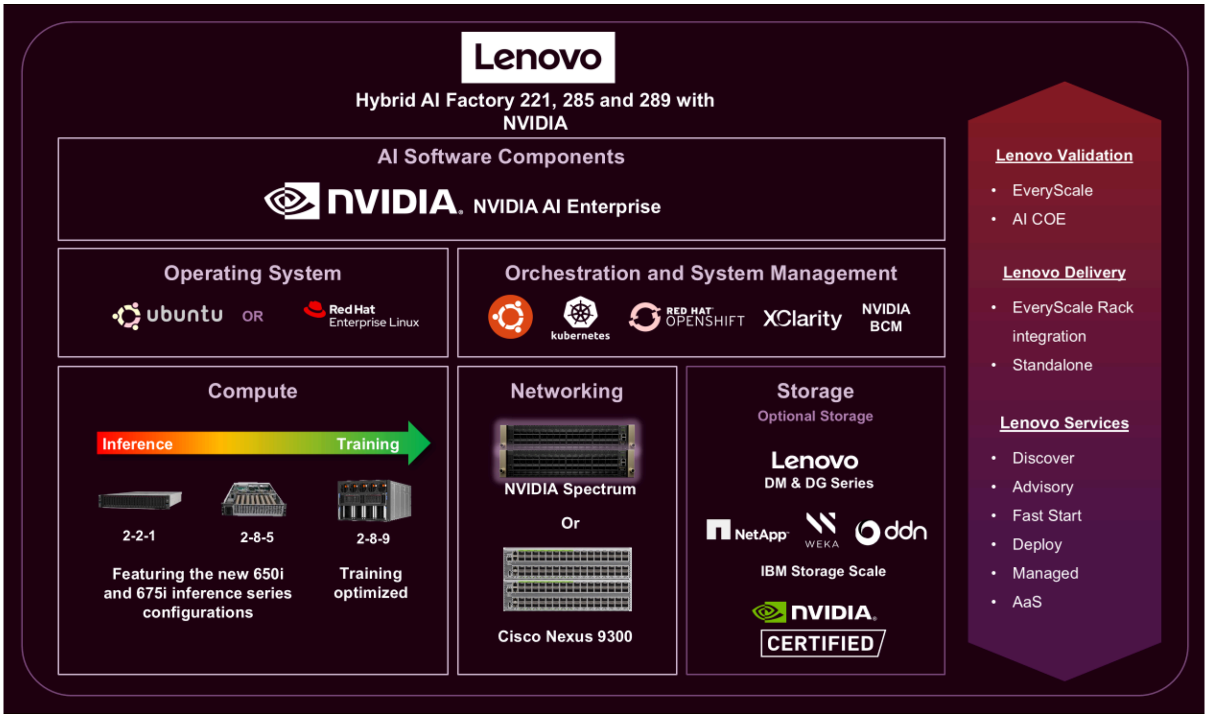
To support this transformation, enterprises need validated architectures that combine performance, scalability, and security. Lenovo’s AI infrastructure portfolio is designed specifically to enable these enterprise AI factories.

Lenovo Hybrid AI Platforms for Enterprise Deployment

One example is the Lenovo Hybrid AI 285 platform, designed to accelerate enterprise AI deployments across hybrid environments. The platform leverages Lenovo ThinkSystem GPU-rich servers, supporting NVIDIA GPUs and the NVIDIA AI Enterprise software to power demanding AI workloads such as large-language-model inference, fine-tuning, and enterprise retrieval-augmented generation (RAG) pipelines.

With a 2-CPU, 8-GPU architecture and high-speed networking, the platform enables organizations to scale from small deployments to large AI clusters while maintaining high performance and operational efficiency.

These validated configurations reduce deployment complexity and help organizations accelerate time-to-value for production AI initiatives.



Lenovo

Hybrid AI Factory 221, 285 and 289 with NVIDIA

AI Software Components

NVIDIA AI Enterprise

Operating System

ubuntu OR Red Hat Enterprise Linux

Orchestration and System Management

kubernetes

RED HAT OPENSHIFT

XClarity

NVIDIA BCM

Compute

Inference

Training

2-2-1

2-8-5

2-8-9

Featuring the new 650i and 675i inference series configurations

Training optimized

Networking

NVIDIA Spectrum

Or

Cisco Nexus 9300

Storage

Optional Storage

Lenovo DM & DG Series

NetApp

WEKA

ddn

IBM Storage Scale

NVIDIA CERTIFIED

Lenovo Validation

- EveryScale
- AI COE

Lenovo Delivery

- EveryScale Rack integration
- Standalone

Lenovo Services

- Discover
- Advisory
- Fast Start
- Deploy
- Managed
- AaS

Figure 1. Lenovo Hybrid AI Factory with NVIDIA

Unlocking Enterprise Data with AI Data Platforms

Enterprise AI systems depend not only on compute performance but also on the ability to access and operationalize vast amounts of enterprise data.

Lenovo's Validated Design for High-Performance AI Data Platforms provides a scalable architecture that enables organizations to securely manage large volumes of unstructured data, including documents, images, audio, and video, while supporting modern AI workflows such as embedding generation and retrieval-augmented generation.

The design integrates high-performance object storage with AI data services to deliver:

- High-throughput data ingestion
- Low-latency retrieval for inference pipelines
- Scalable storage architectures optimized for GPU workloads
- Enterprise-grade security and governance

These capabilities ensure that AI systems can access the data they need while maintaining full control over data sovereignty and compliance requirements.

Securing the Next Generation of AI Infrastructure

As AI continues to transform industries, organizations must build platforms that combine performance, scalability, and trust.

By integrating confidential computing technologies with validated AI infrastructure, enterprises can deploy AI workloads that:

- Protect proprietary model intellectual property
- Safeguard sensitive enterprise data
- Enable secure multi-tenant AI environments
- Support production-scale AI factories

Together, innovations in confidential computing and Lenovo's hybrid AI infrastructure are helping enterprises unlock the full potential of AI, while maintaining the security and governance required for real-world deployments.

Learn more about Lenovo AI infrastructure and hybrid AI factory solutions at NVIDIA GTC.

Visit www.lenovo.com/hybridai to learn more.

Authors

Carlos Huescas is the Worldwide Product Manager for NVIDIA software at Lenovo. He specializes in High Performance Computing and AI solutions. He has more than 15 years of experience as an IT architect and in product management positions across several high-tech companies.

Pierce Beary is the AI Solutions Product Manager focused on datacenter solutions, he previously worked on the WW HPC team as a product manager for Lenovo EveryScale (LeSI). Prior to joining Lenovo in 2021, Pierce held positions at Exxon Mobil, where he worked as a Mechanical Engineer and Engineering Project Manager. Pierce has a Bachelor's degree in Mechanical Engineering from North Carolina State University.

Farah Toosi is a Software Product Manager for NVIDIA enterprise software in Lenovo's Infrastructure Solutions Group. She specializes in AI/ML software infrastructure, GPU orchestration, and enterprise AI platform integrations. She has 8+ years of experience across software products and program management on several high tech companies

Related product families

Product families related to this document are the following:

- Hybrid AI Factory

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service. Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION ”AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary. Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk. Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document, LP2409, was created or updated on March 27, 2026.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP2409
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP2409.

Trademarks