Lenovo Brings the Full NVIDIA Vera Rubin Platform to AI Cloud Providers

Article

At OCP Global Summit 2026, Lenovo is showcasing the latest expansion of Lenovo AI Cloud Gigafactory with the NVIDIA Vera Rubin Platform. Built on the complete NVIDIA Vera Rubin platform, the integrated, rack-scale solution is designed to help cloud service providers (CSPs) and neoclouds rapidly deploy production-ready infrastructure for advanced AI training and inference, turning growing AI demand into revenue-generating capacity.

Lenovo AI Cloud Gigafactory with NVIDIA

For AI cloud providers, the challenge is no longer simply securing high-density GPU infrastructure. Success depends on orchestrating increasingly complex infrastructure requirements to minimize execution risk, accelerate deployment and reach revenue-generating capacity faster.

The **Lenovo AI Cloud Gigafactory with NVIDIA** addresses that challenge through a comprehensive services-led model. The offering bridges Lenovo engineering expertise, global manufacturing, in-house integration and validation, Lenovo Neptune® Liquid Cooling experience and lifecycle services with NVIDIA-accelerated infrastructure, so providers can move from strategy to revenue-ready AI capacity with greater confidence.

"*The AI race is no longer just about deploying GPUs. It's about how quickly providers can transform power, space and infrastructure investments into operational AI capacity and revenue,*" said **Conor Malone, Vice President & GM, Cloud Service Providers, Lenovo**. "*Lenovo AI Cloud Gigafactory with NVIDIA brings together advanced AI infrastructure with Lenovo's global manufacturing and deployment expertise, including capabilities at our North Carolina Smart Campus, to help customers build AI factories faster, scale with confidence, and unlock business value sooner.*"

"The next generation of AI cloud will require full-stack, rack-scale infrastructure designed for increasingly complex reasoning, inference and agentic AI workloads," said **Chris Marriott, Vice President of Enterprise Platforms and Solutions** at **NVIDIA**. *"By bringing the NVIDIA Vera Rubin Platform to Lenovo AI Cloud Gigafactory, Lenovo is helping cloud providers reduce deployment complexity, scale with confidence and turn infrastructure investments into revenue-ready AI capacity faster."*



Figure 1. Lenovo NVIDIA Vera Rubin NVL72

Five Rack-Scale Systems, Delivered as One

As AI demand shifts from training to high-volume inference, reasoning and agentic AI, the unit of AI compute is moving from the individual rack to the full pod. Lenovo's NVIDIA Vera Rubin Platform brings together all five rack-scale systems of the NVIDIA Vera

Rubin platform, aligned with NVIDIA DSX AI Factory Platform reference architecture guidance:

- **NVIDIA Vera Rubin NVL72:** Combines 72 Rubin GPUs and 36 NVIDIA Vera CPUs with NVIDIA NVLink 6 scale-up networking, NVIDIA ConnectX-9 SuperNICs and NVIDIA BlueField-4 DPUs. NVIDIA reports this system can achieve up to 30x higher throughput per megawatt, which directly lowers inference costs by up to 45x per million tokens - all while utilizing one fourth the GPUs to deliver AI intelligence when compared against previous generations.
- **NVIDIA Groq 3 LPX:** A dedicated inference accelerator rack for the low-latency, large-context demands of agentic systems, with 256 LPUs, 128GB of on-chip SRAM, 640 TB/s of scale-up bandwidth, NVIDIA ConnectX-9 based fabric expansion logic for scale-out networking, and NVIDIA BlueField-4 DPUs for scale-in networking.
- **NVIDIA Vera CPU Rack:** A dense, liquid-cooled NVIDIA MGX-based system with 256 Vera CPUs and 128 BlueField-4 DPUs. NVIDIA MGX is a modular reference architecture for accelerated computing, helping system builders design future-compatible systems faster, reducing costs, accelerating time to market, and improving ROI. A single rack can run more than 22,500 concurrent reinforcement learning or agent sandbox environments.
- **NVIDIA CMX Context Memory Storage Platform:** AI-native, rack scale context memory storage built on NVIDIA STX and powered by BlueField-4 STX storage processors. CMX extends GPU memory with a shared, pod-level tier optimized for KV cache.
- **NVIDIA Spectrum-6 SPX Ethernet rack:** High-performance networking racks for connecting compute within the Vera Rubin POD.

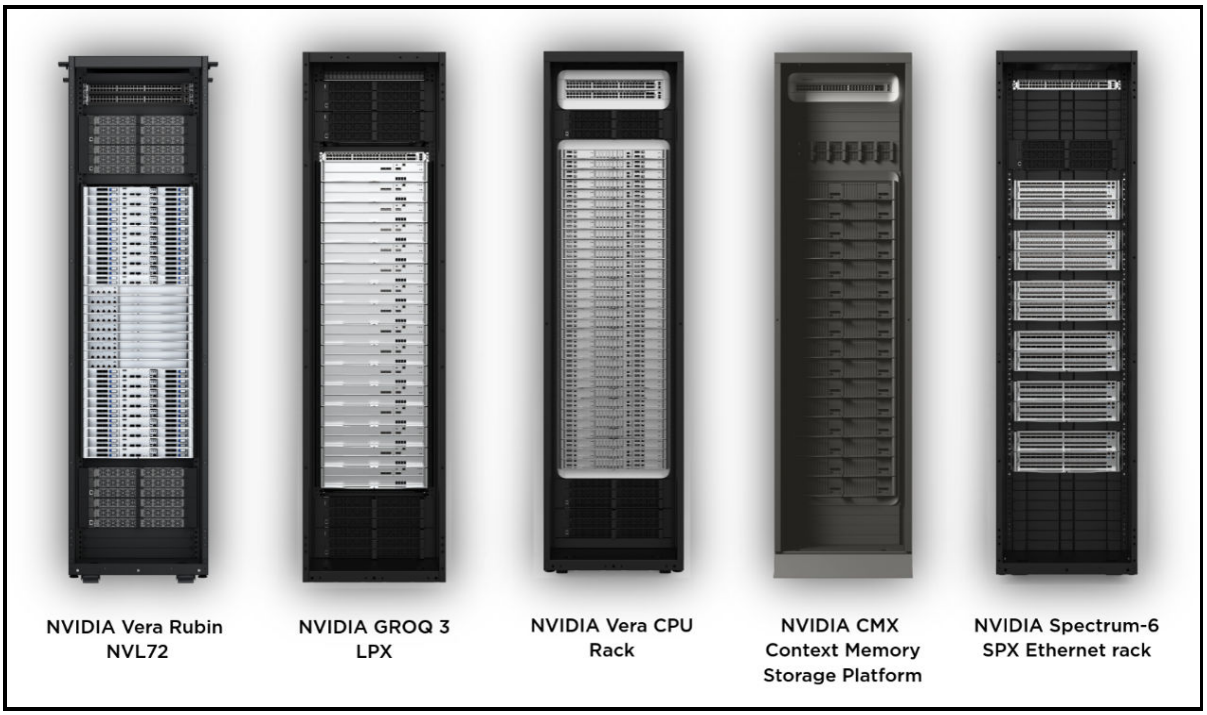


Figure 2. Lenovo's NVIDIA Vera Rubin Platform

Manufacturing at Global Scale, with U.S. Production in North Carolina

Gigafactory-scale AI depends on the ability to build, integrate and validate liquid-cooled racks at volume. Lenovo's global/local manufacturing strategy spans more than 30 factories in 11 markets, pairing global scale with regional production close to customers. Lenovo's supply chain was also recently ranked fifth globally in the Gartner® Supply Chain Top 25 for 2026, its highest-ever placement.

In North America, that scale is anchored by the Lenovo North Carolina Smart Campus (NCSC) in Whitsett,

North Carolina. Following a recent expansion, the broader campus spans 890,000 square feet, with NCSC encompassing 520,000 square feet of that footprint and infrastructure ready for 20 megawatts of power. It is future-proofed for the next two generations of high-performance computing, and Lenovo has added 35% more manufacturing capacity, including Direct Liquid Cooling. Located less than an hour from Lenovo's development labs, NCSC brings production closer to customers for faster delivery, greater flexibility and improved responsiveness.



Figure 3. Lenovo North Carolina Smart Campus (NCSC) in Whitsett, N.C. in the USA

Availability

Lenovo is now shipping NVIDIA Vera Rubin NVL72 globally, enabling CSPs, neoclouds, and AI infrastructure providers to accelerate AI factory deployments today, with early access to the broader NVIDIA Vera Rubin Platform as the ecosystem becomes available.

To further simplify deployment and growth, Lenovo Global Financial Services (LGFS) offers flexible financing and consumption models that help customers preserve capital, accelerate capacity expansion, and align infrastructure investments with business growth.

Lenovo AI Cloud Gigafactory at 2026 OCP Global Summit

Lenovo is showcasing the **NVIDIA Vera Rubin Platform** at Booth D35 during the 2026 OCP Global Summit, held October 12–15 at the San Jose Convention Center. Lenovo experts are on site to discuss AI Cloud Gigafactory design, liquid cooling, manufacturing and deployment for AI cloud builders.

Lenovo is also hosting two key speaking sessions:

- **From Blueprint to Revenue: Engineering and Executing AI Cloud Gigafactories with NVIDIA**

  Monday. October 12, 1:15pm PT at SJCC Concourse Level 211. Featuring John Donovan, ED & GM of AI & Scale Cloud, Lenovo, and Chris Marriott, VP of Enterprise Platforms and Solutions, NVIDIA, will explore how the principles behind gigawatt-scale AI cloud buildouts translate into faster time to revenue for providers and enterprises alike.

- **CLARA: Automated AI Cluster Architecture Design Tool**

  Tuesday, October 13, 1:45pm PT at SJCC Concourse Level 220B. Lenovo's Rainy Yan, AI Cluster Automation Engineer, and Reece Neff, Cluster Solution Engineer, will introduce CLARA, Lenovo's software-defined framework for turning logical AI cluster requirements into physically viable deployments. The session will cover algorithmic topology generation and simulated-annealing floorplan optimization. These techniques minimize cable lengths while meeting strict power and space constraints, reducing design iterations and helping prevent the cascading errors common in manual design.

For more information

For more information about the Lenovo NVIDIA Vera Rubin AI Factory Platform, see the Lenovo CSR4172 V3 NVIDIA Vera Rubin NVL72 datasheet or contact your Lenovo representative.

Author

Nevin Thomas is a marketing professional on the CSP Marketing & Strategy team in the Lenovo Infrastructure Solutions Group, based in Morrisville, NC. He leads product marketing, partner co-marketing, sales enablement and content for Lenovo's North American cloud service provider business, which serves hyperscalers, AI cloud providers and scale cloud providers.

Related product families

Product families related to this document are the following:

- Artificial Intelligence

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.

© Copyright Lenovo 2026. All rights reserved.

This document, LP2549, was created or updated on October 12, 2026.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP2549
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP2549.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at https://www.lenovo.com/us/en/legal/copytrade/.

The following terms are trademarks of Lenovo in the United States, other countries, or both:
Lenovo®
Neptune™

The following terms are trademarks of other companies:

NVIDIA®, NVIDIA Spectrum®, and NVLink® are trademarks of NVIDIA Corporation.

Other company, product, or service names may be trademarks or service marks of others.