



NEWS RELEASE

Cisco Expands Secure AI Factory with NVIDIA for the Rack-Scale Era

2026-08-25

Cisco's full-stack architecture adds massive AI computing power, offering an NVIDIA Cloud Partner compliant reference architecture built for surging neocloud and sovereign cloud demand.

News Summary

- *Cisco, partnering with Supermicro, adds rack-scale AI computing solutions to its portfolio, delivering end-to-end infrastructure to run massive AI workloads, reduce deployment risk, and help customers run workloads with greater efficiency and control over data.*
- *With the addition of high-density server systems, Cisco extends its industry-leading Secure AI Factory architecture to any AI use case, from trillion parameter training models to edge inferencing.*
- *The full-stack architecture will address the quickly growing and evolving demands of neocloud and sovereign cloud customers with an NVIDIA Cloud Partner (NCP) compliant architecture featuring Cisco AI networking systems, backed by industry-leading support and services.*

SAN JOSE, Calif., Aug 25, 2026 /PRNewswire/ -- AI models keep evolving, placing new performance demands on networks. Power and cooling dictate where infrastructure can be built. Customers are demanding AI capabilities that keep their data secure and private. The next era of AI infrastructure can't just deliver raw compute; it must harness that compute through validated, full-stack architectures customers can deploy quickly, run efficiently, and secure with confidence. Today, Cisco (Nasdaq: CSCO) is expanding the [Secure AI Factory with NVIDIA](#) through a partnership with Supermicro, bringing its leadership in high-density, liquid and air-cooled compute to Cisco's industry-leading full-stack, secure AI infrastructure architecture for enterprises, neoclouds and sovereign clouds.

"We are at the beginning of one of the largest datacenter buildouts in history," said Jeetu Patel, President and Chief Product Officer, Cisco. "Every organization is racing to scale AI — but speed only counts if it comes with control of data, managed token costs, and real ROI. It starts with the right infrastructure: compute and networking, delivered as an integrated solution that's easy to deploy and secure from day one."

"AI factories are revenue-generating infrastructure, where compute produces intelligence, and intelligence drives revenue," said Justin Boitano, vice president, Enterprise AI, NVIDIA. "By expanding the Cisco Secure AI Factory with NVIDIA, Cisco is providing enterprises and neoclouds full-stack infrastructure that can get into production faster and generate more value from every watt."

Simple, Secure, Scalable Infrastructure – with More Muscle

Cisco offers a full-stack approach spanning core data centers, local and edge environments, for the enterprise, neocloud and sovereign cloud markets. Through this new partnership, Cisco will offer Supermicro liquid- and air-cooled systems, giving customers access to rack-scale and dense GPU systems – validated and sold as part of the broader Cisco AI infrastructure portfolio. This expansion enables customers to easily manage complex, high-density AI clusters alongside non-AI workloads. Customers will also now be able to deploy rack-to-fabric liquid cooling, featuring Cisco liquid-cooled AI networking systems alongside Supermicro's liquid-cooled servers. This unlocks trillion-parameter training and high-throughput inference use cases with platforms including [NVIDIA Vera Rubin NVL72](#) and [NVIDIA HGX Rubin NVL8](#).

The expanded Cisco Secure AI Factory with NVIDIA offers NVIDIA Cloud Partner (NCP) compliant solutions for neocloud and sovereign cloud customers. Cisco [Silicon One](#)-based switches for the front-end and Cisco's [NVIDIA Spectrum-X based switches](#) for the back-end are unified by Cisco [Nexus One](#), delivering a fully unified networking architecture for the full stack. Cisco is the only NVIDIA technology partner to utilize its own networking switches and network operating system in an NCP compliant solution. Cisco Secure AI Factory with NVIDIA is also expanding its Enterprise Reference Architectures to include the latest generation of NVIDIA AI infrastructure for enterprise deployments.

The Cisco Secure AI Factory with NVIDIA helps customers plan AI investments with confidence. Customer benefits include:

- **Less Risk:** Cisco and Supermicro each bring industry-leading global supply chain expertise to help mitigate delivery timeline challenges, particularly in the face of GPU and memory access challenges. With Cisco Secure AI Factory solutions based on NVIDIA reference architectures, customers can have confidence that their infrastructure is pressure-tested for modern workloads, with security and resiliency built in from the silicon to the agents. Industry-leading global support with a broad partner ecosystem gives customers the confidence to invest.
- **Faster Time to Value:** Once delivered, Cisco and its partners will help customers certify infrastructure quickly. New Cisco Validated Infrastructure Services (CVIS) — aligned with the NVIDIA Infrastructure Services (NVIS) offer — certifies infrastructure is built exactly as it was designed and aligned with reference architectures. Cisco is investing in a dedicated large-scale AI Lab to develop tools and test software to advance CVIS.
- **Simplified Operations:** With NVIDIA AI Enterprise software and AgenticOps through [Cisco Cloud Control](#), customers will use tools they already know, making AI deployment feel like an expansion rather than an overhaul. Customers will have the ability to correlate job health with compute, NIC, optics and network performance metrics, delivering true end-to-end observability.

Availability

Cisco will begin offering Supermicro compute solutions as part of the Secure AI Factory with NVIDIA beginning in October 2026.

Supporting Quotes

"With Cisco Secure AI Factory with NVIDIA, we no longer have to choose between performance, reliability or

ease of management. NCP validation gives us the confidence that our infrastructure is optimized from day one, while the platform's rack-scale capability provides a seamless path to scale our AI operations as our business grows." - James Manning, Co-founder and Chief Executive Officer, Sharon AI.

"Our deep partnership with Cisco is central to our growth, and we rely on their strategic foresight in bringing the right specialized technologies and partner ecosystem into their AI infrastructure portfolio. By combining our strengths with Cisco's secure AI architecture, we can deliver an even more comprehensive full-stack solution, which is exactly what our customers need now." - Neil Anderson, VP and CTO Cloud, Infrastructure, and AI Solutions, WWT

"Our customers span almost every vertical industry, and our priority is always to offer them solutions and services that right-fit their organization. This partnership enables us to offer a robust, full-stack infrastructure spanning high-performance AI compute, networking, power, liquid cooling, and rack-scale solutions that are necessary to provide high-impact, front-end value around AI initiatives. This enhances the focus we have on helping customers securely operate, observe and manage agentic workloads. We look forward to working with Cisco and the newly formed partnership as this solution will help customers meet their demands and desired outcome end-to-end." - Ken Farber, president, ePlus strategy, software and alliances

"This alliance strengthens our full-stack differentiation in the market, while offering customers the security of Cisco's AI infrastructure backed by a proven deployment and services model." - Vikesh Gumpalli, Executive Managing Director, Infrastructure Solutions, NTT DATA North America

Supporting Resources:

- For more information on the Cisco Secure AI Factory with NVIDIA, [click here](#)
- Read the [Announcement FAQ](#) for more detail on today's announcement

About Cisco

Cisco (NASDAQ: CSCO) is the worldwide technology leader that is revolutionizing the way organizations connect and protect in the AI era. For more than 40 years, Cisco has securely connected the world. With its industry leading AI-powered solutions and services, Cisco enables its customers, partners and communities to unlock innovation, enhance productivity and strengthen digital resilience. With purpose at its core, Cisco remains committed to creating a more connected and inclusive future for all. Discover more on [The Newsroom](#) and follow us on X at @Cisco.

View original content to download multimedia: <https://www.prnewswire.com/news-releases/cisco-expands-secure-ai-factory-with-nvidia-for-the-rack-scale-era-302858932.html>

SOURCE Cisco Systems, Inc.