

Cisco and NVIDIA to Help Enterprises Quickly and Easily Deploy and Manage Secure AI Infrastructure

News Summary

- Companies to offer enterprises simplified cloud-based and on-premises AI infrastructure, networking and software, including infrastructure management, secure AI infrastructure, observable end-to-end AI solutions and access to NVIDIA AI Enterprise software that supports the building and deployment of advanced AI and generative AI workloads.
- Cisco and NVIDIA's purpose-built Ethernet networking-based solutions will be sold through Cisco's vast global channel, offering professional services and support through key partners who are committed to helping businesses deploy their GPU clusters via Ethernet infrastructure.
- The collaboration has attracted key customers like ClusterPower, a cloud services provider in Europe, to help drive data center operations with innovative AI/ML solutions that are foundational for its client infrastructure and services.

CISCO LIVE—Cisco and NVIDIA today announced plans to deliver AI infrastructure solutions for the data center that are easy to deploy and manage, enabling the massive computing power that enterprises need to succeed in the AI era.

“AI is fundamentally changing how we work and live, and history has shown that a shift of this magnitude is going to require enterprises to rethink and re-architect their infrastructures,” said Chuck Robbins, Chair and CEO, Cisco. “Strengthening our great partnership with NVIDIA is going to arm enterprises with the technology and the expertise they need to build, deploy, manage, and secure AI solutions at scale.”

“Companies everywhere are racing to transform their businesses with generative AI,” said Jensen Huang, founder and CEO of NVIDIA. “Working closely with Cisco, we’re making it easier than ever for enterprises to obtain the infrastructure they need to benefit from AI, the most powerful technology force of our lifetime.”

A Powerful Partnership

Cisco, with its industry-leading expertise in Ethernet networking and extensive partner ecosystem, together with NVIDIA, the inventor of the GPU that fueled the AI boom, share a vision and commitment to help customers navigate the transitions for AI with highly secure Ethernet-based infrastructure.

Cisco and NVIDIA have offered a broad range of integrated product solutions over the past several years across Webex collaboration devices and data center compute environments to enable hybrid workforces with flexible workspaces, AI-powered meetings and virtual desktop infrastructure. The companies are now deepening their partnership in the data center to enable enterprise customers with scalable and automated AI cluster management, automated troubleshooting, best-in-class customer experiences, and more. Highlights include:

Cisco and NVIDIA Integrated Data Center Solutions Available Today

- **NVIDIA's newest Tensor Core GPUs are available in Cisco's M7 generation of UCS rack and blade servers**, including Cisco UCS X-Series and UCS X-Series Direct, to enable optimal performance across a broad array of AI and data-intensive workloads in the data center and at the edge.
- **[NVIDIA AI Enterprise](#)**, which includes software frameworks, pretrained models and development tools for more secure, stable and supported production AI is now available on Cisco's global price list.
- **Jointly validated reference architectures** through Cisco Validated Designs (CVDs) make it simple to deploy and manage AI clusters at any scale in a wide array of use cases spanning virtualized and containerized environments, with both converged and hyperconverged options. CVDs for FlexPod and FlashStack for Generative AI Inferencing with NVIDIA AI Enterprise will be available this month, with more to follow.
- **Supporting Cisco Networking Cloud:** Cisco simplified AI infrastructure management and operations through both on-premises and cloud-based management with Cisco Nexus Dashboard and Cisco Intersight.
- **Digital Experience Monitoring:** With AI workloads and data in the public cloud, on premises and across multiple data centers, ThousandEyes provides Digital Experience Monitoring to provide AI-driven insights and automated remediation of problems that occur anywhere across the cloud to on-premises networks.
- **The Cisco Observability Platform** uses AI capabilities to contextualize and correlate real-time telemetry across domains, so organizations can better attain the visibility, insights and actions to improve digital experiences.
- **Partners Reducing Risks:** As organizations plan to successfully adopt AI and automation, they will look to Cisco's global ecosystem of partners to advise, support and guide them.

“Building AI compute-based environments can be a daunting task for organizations, and underscores the need for simple, adaptable data center infrastructure that delivers optimized performance,” said Vladimir Ester, chief technology officer and cofounder of ClusterPower. “Cisco and NVIDIA are bringing together the innovations to help enterprises support demand for

more computing power in the data center and navigate the transitions for AI with secure and observable infrastructure.”

“World Wide Technology (WWT) is excited to see NVIDIA and Cisco combining resources to bring proven Ethernet-based AI solutions to market,” said Neil Anderson, vice president of Cloud & Infrastructure Solutions at WWT. “Ethernet has the broadest scale in our client base, and this partnership and resulting offers will accelerate our ability to deliver AI solutions to our clients. Both Cisco and NVIDIA are already a significant part of WWT’s [AI Proving Ground](#) lab where we help clients select and operationalize AI architectures so they can more quickly turn their data into insights and action.”

“As enterprises look to transform their businesses with AI, they must understand the unique demands that AI workloads place on data center infrastructure,” said Vijay Bhagavath, vice president of Cloud & Data Center Networks at IDC Research. “The Cisco and NVIDIA partnership brings together two trusted brands with complementary technologies to enable customers to realize the full potential of AI with a wide range of performance-optimized Ethernet-based infrastructure.”

Availability: 2Q Calendar Year; Solutions sold through Cisco Channel partners

About NVIDIA

Since its founding in 1993, NVIDIA (NASDAQ: NVDA) has been a pioneer in accelerated computing. The company’s invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, ignited the era of modern AI and is fueling industrial digitalization across markets. NVIDIA is now a full-stack computing infrastructure company with data-center-scale offerings that are reshaping industry. More information at <https://nvidianews.nvidia.com/>.

About Cisco

Cisco (NASDAQ: CSCO) is the worldwide technology leader that securely connects everything to make anything possible. Our purpose is to power an inclusive future for all by helping our customers reimagine their applications, power hybrid work, secure their enterprise, transform their infrastructure, and meet their sustainability goals. Discover more on [The Newsroom](#) and follow us on X at [@Cisco](#).

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, performance, features, and availability of NVIDIA’s products and technologies, including NVIDIA AI Enterprise and Tensor Core GPUs; the benefits and impact of the partnership between Cisco and NVIDIA and the availability of the services and offerings; AI fundamentally changing how we work and live; and companies everywhere racing to transform their businesses with generative AI are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners’ products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company’s website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

Many of the products and features described herein remain in various stages and will be offered on a when-and-if-available basis. The statements above are not intended to be, and should not be interpreted as a commitment, promise, or legal obligation, and the development, release, and timing of any features or functionalities described for our products is subject to change and remains at the sole discretion of NVIDIA. NVIDIA will have no liability for failure to deliver or delay in the delivery of any of the products, features or functions set forth herein.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. A listing of Cisco’s trademarks can be found at www.cisco.com/go/trademarks. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company.

NVIDIA and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and/or other countries. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability, and specifications are subject to change without notice.

Alex Shapiro
Enterprise Networking
1-415-608-5044
ashapiro@nvidia.com