



NVIDIA Announces DGX Cloud Lepton to Connect Developers to NVIDIA's Global Compute Ecosystem

- *CoreWeave, Crusoe, Firmus, Foxconn, GMI Cloud, Lambda, Nebius, Nscale, SoftBank Corp. and Yotta Data Services to Bring Tens of Thousands of GPUs to DGX Cloud Lepton Marketplace*
- *NVIDIA Exemplar Clouds Raise the Performance Bar for NVIDIA Cloud Partners*

COMPUTEX—NVIDIA today announced [NVIDIA DGX Cloud Lepton™](#) — an AI platform with a compute marketplace that connects the world's developers building agentic and physical AI applications with tens of thousands of GPUs, available from a global network of cloud providers.

To meet the demand for AI, NVIDIA Cloud Partners ([NCPs](#)) including CoreWeave, Crusoe, Firmus, Foxconn, GMI Cloud, Lambda, Nebius, Nscale, Softbank Corp. and Yotta Data Services will offer NVIDIA Blackwell and other NVIDIA architecture GPUs on the DGX Cloud Lepton marketplace.

Developers can tap into GPU compute capacity in specific regions for both on-demand and long-term computing, supporting strategic and [sovereign AI](#) operational requirements. Leading cloud service providers and GPU marketplaces are expected to also participate in the DGX Cloud Lepton marketplace.

"NVIDIA DGX Cloud Lepton connects our network of global GPU cloud providers with AI developers," said Jensen Huang, founder and CEO of NVIDIA. "Together with our NCPs, we're building a planetary-scale AI factory."

DGX Cloud Lepton helps address the critical challenge of securing reliable, high-performance GPU resources by unifying access to cloud AI services and GPU capacity across the NVIDIA compute ecosystem. The platform integrates with the NVIDIA software stack, including [NVIDIA NIM™](#) and [NeMo™](#) microservices, [NVIDIA Blueprints](#) and [NVIDIA Cloud Functions](#), to accelerate and simplify the development and deployment of AI applications.

For cloud providers, DGX Cloud Lepton provides management software that delivers real-time GPU health diagnostics and automates root-cause analysis, eliminating manual operations and reducing downtime.

Key benefits of the platform include:

- **Improved productivity and flexibility:** Offers a unified experience across development, training and inference, helping boost productivity. Developers can purchase GPU capacity directly from participating cloud providers through the marketplace or bring their own compute clusters, giving them greater flexibility and control.
- **Frictionless deployment:** Enables deployment of AI applications across multi-cloud and hybrid environments with minimal operational burden, using integrated services for inference, testing and training workloads.
- **Agility and sovereignty:** Gives developers quick access to GPU resources in specific regions, enabling compliance with data sovereignty regulations and meeting low-latency requirements for sensitive workloads.
- **Predictable performance:** Provides participating cloud providers enterprise-grade performance, reliability and security, ensuring a consistent user experience.

A New Bar for AI Cloud Performance

[NVIDIA today also announced](#) NVIDIA Exemplar Clouds to help NCPs enhance security, usability, performance and resiliency, using NVIDIA's expertise, reference hardware and software and operational tools.

NVIDIA Exemplar Clouds tap into [NVIDIA DGX™ Cloud Benchmarking](#), a comprehensive suite of tools and recipes for optimizing workload performance on AI platforms and quantifying the relationship between cost and performance.

Yotta Data Services is the first NCP in the Asia-Pacific region to join the NVIDIA Exemplar Cloud initiative.

Availability

Developers can [sign up](#) for early access to NVIDIA DGX Cloud Lepton.

Watch the [COMPUTEX](#) keynote from Huang and learn more at NVIDIA GTC Taipei.

About NVIDIA

[NVIDIA](#) (NASDAQ: NVDA) is the world leader in accelerated computing.

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, performance and availability of NVIDIA's products, services; NVIDIA's collaborations with third parties and the benefits and impact thereof; third parties using or adopting our products and technologies, the benefits and impact thereof; together with cloud partners,

NVIDIA building a virtual global AI factory and additional regional cloud providers being added to the marketplace in the coming months are forward-looking statements within the meaning of Section 27A of the Securities Act of 1933, as amended, and Section 21E of the Securities Exchange Act of 1934, as amended, which are subject to the “safe harbor” created by those sections and 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.

© 2025 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, DGX, DGX Cloud Lepton, NeMo and NVIDIA NIM are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and 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.

Natalie Hereth
NVIDIA Corporation
nhereth@nvidia.com