



NVIDIA DGX Cloud Lepton Connects Europe's Developers to Global NVIDIA Compute Ecosystem

- *Mistral AI, Nebius, Nscale, Firebird, Fluidstack, Hydra Host, Scaleway and Together AI — Along With AWS and Microsoft Azure — Bring Compute Resources to DGX Cloud Lepton Marketplace to Meet AI Demand*
- *Hugging Face Integrates DGX Cloud Lepton Into Training Cluster as a Service, Expanding AI Researcher Access to Scalable Compute for Model Training*
- *NVIDIA and Leading European Venture Capitalists Offer Marketplace Credits to Portfolio Companies to Accelerate Startup Ecosystem*

NVIDIA GTC Paris at VivaTech—NVIDIA today announced the expansion of [NVIDIA DGX Cloud Lepton™](#) — an AI platform featuring a global compute marketplace that connects developers building agentic and physical AI applications — with GPUs now available from a growing network of cloud providers.

Mistral AI, Nebius, Nscale, Firebird, Fluidstack, Hydra Host, Scaleway and Together AI are now contributing [NVIDIA Blackwell](#) and other NVIDIA architecture GPUs to the marketplace, expanding regional access to high-performance compute. AWS and Microsoft Azure will be the first large-scale cloud providers to participate in DGX Cloud Lepton. These companies join CoreWeave, Crusoe, Firmus, Foxconn, GMI Cloud, Lambda and Yotta Data Services in the marketplace.

To make accelerated computing more accessible to the global AI community, [Hugging Face is introducing](#) Training Cluster as a Service. This new offering integrates with DGX Cloud Lepton to seamlessly connect AI researchers and developers building foundation models with the NVIDIA compute ecosystem.

NVIDIA is also working with leading European venture capital firms Accel, Elaia, Partech and [Sofinnova Partners](#) to offer DGX Cloud Lepton marketplace credits to portfolio companies, enabling startups to access accelerated computing resources and scale regional development.

“DGX Cloud Lepton is connecting Europe’s developers to a global AI infrastructure,” said Jensen Huang, founder and CEO of NVIDIA. “With partners across the region, we’re building a network of AI factories that developers, researchers and enterprises can harness to scale local breakthroughs into global innovation.”

DGX Cloud Lepton simplifies the process of accessing reliable, high-performance GPU resources within specific regions by unifying cloud AI services and GPU capacity from across the NVIDIA compute ecosystem onto a single platform. This enables developers to keep their data local, supporting data governance and sovereign AI requirements.

In addition, by integrating with the NVIDIA software suite — including [NVIDIA NIM™](#) and [NeMo™](#) microservices and [NVIDIA Cloud Functions](#) — DGX Cloud Lepton streamlines and accelerates every stage of AI application development and deployment, at any scale. The marketplace works with a new NIM microservice container, which includes support for a broad range of large language models, including the most popular open LLM architectures and more than a million models hosted publicly and privately on Hugging Face.

For cloud providers, DGX Cloud Lepton includes management software that continuously monitors GPU health in real time and automates root-cause analysis, minimizing manual intervention and reducing downtime. This streamlines operations for providers and ensures more reliable access to high-performance computing for customers.

NVIDIA DGX Cloud Lepton Speeds Training and Deployment

Early-access DGX Cloud Lepton customers using the platform to accelerate their strategic AI initiatives include:

- Basecamp Research, which is speeding the discovery and design of new biological solutions for pharmaceuticals, food and industrial and environmental biotechnology by harnessing its 9.8 billion-protein database to pretrain and deploy large biological foundation models.
- EY, which is standardizing multi-cloud access across the global organization to accelerate the development of AI agents for domain- and sector-specific solutions.
- Outerbounds, which enables customers to build differentiated, production-grade AI products powered by the proven reliability of open-source Metaflow.
- Prima Mente, which is advancing neurodegenerative disease research at scale by pretraining large brain foundation models to uncover new disease mechanisms and tools to stratify patient outcomes in clinical settings.
- Reflection, which is building superintelligent autonomous coding systems that handle the most complex enterprise engineering tasks.

Hugging Face Developers Get Access to Scalable AI Training Across Clouds

Integrating DGX Cloud Lepton with Hugging Face’s [Training Cluster as a Service](#) offering gives AI builders streamlined

access to the GPU marketplace, making it easy to reserve, access and use NVIDIA compute resources in specific regions, close to their training data. Connected to a global network of cloud providers, Hugging Face customers can quickly secure the necessary GPU capacity for training runs using DGX Cloud Lepton. [Mirror Physics](#), [Project Numina](#) and [the Telethon Institute of Genetics and Medicine](#) will be among the first Hugging Face customers to access Training Cluster as a Service, with compute resources provided through DGX Cloud Lepton. They will use the platform to advance state-of-the-art AI models in chemistry, materials science, mathematics and disease research.

“Access to large-scale, high-performance compute is essential for building the next generation of AI models across every domain and language,” said Clément Delangue, cofounder and CEO of Hugging Face. “The integration of DGX Cloud Lepton with Training Cluster as a Service will remove barriers for researchers and companies, unlocking the ability to train the most advanced models and push the boundaries of what’s possible in AI.”

DGX Cloud Lepton Boosts AI Startup Ecosystem

NVIDIA is working with Accel, Elaia, Partech and Sofinnova Partners to offer up to \$100,000 in GPU capacity credits and support from NVIDIA experts to eligible portfolio companies through DGX Cloud Lepton.

BioCorteX, Bioptimus and Latent Labs will be among the first to access DGX Cloud Lepton, where they can discover and purchase compute capacity and use NVIDIA software, services and AI expertise to build, customize and deploy applications across a global network of cloud providers.

Availability

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

Watch the [NVIDIA GTC Paris keynote](#) from Huang at VivaTech and explore [GTC Paris sessions](#).

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: DGX Cloud Lepton connecting Europe’s developers to a global AI infrastructure; with partners across the region, NVIDIA building a network of AI factories that developers, researchers and enterprises can harness to scale local breakthroughs into global innovation; the benefits, impact, performance, and availability of NVIDIA’s products, services, and technologies; expectations with respect to NVIDIA’s third party arrangements, including with its collaborators and partners; expectations with respect to technology developments; and other statements that are not historical facts 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 based on management’s beliefs and assumptions and on information currently available to management and 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 and political conditions; NVIDIA’s reliance on third parties to manufacture, assemble, package and test NVIDIA’s products; the impact of technological development and competition; development of new products and technologies or enhancements to NVIDIA’s existing product and technologies; market acceptance of NVIDIA’s products or NVIDIA’s partners’ products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of NVIDIA’s products or technologies when integrated into systems; and changes in applicable laws and regulations, 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 Cloud Lepton, NVIDIA 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