

NVIDIA Launches GPU Cloud Platform to Simplify AI Development

NVIDIA GPU Cloud Platform to Combine Deep Learning Software with World's Fastest GPUs

NVIDIA today announced the [NVIDIA GPU Cloud](#) (NGC), a cloud-based platform that will give developers convenient access -- via their PC, NVIDIA DGX system or the cloud -- to a comprehensive software suite for harnessing the transformative powers of AI.

Speaking at the eighth annual [GPU Technology Conference](#), NVIDIA CEO and founder Jensen Huang said that NGC will make it easier for developers to access the latest, optimized deep learning frameworks and the newest GPU computing resources.

"We're designing a cloud platform that will unleash AI developers, so they can build a smarter world," said Jim McHugh, vice president and general manager of DGX Systems at NVIDIA. "You can do your best work no matter where you are, using our latest technology in the cloud. It's accelerated computing when and where you need it."

Harnessing deep learning presents two challenges for developers and data scientists. One is the need to gather into a single stack the requisite software components -- including deep learning frameworks, libraries, operating system and drivers. Another is getting access to the latest GPU computing resources to train a neural network.

NVIDIA solved the first challenge earlier this year by combining the key software elements within the [NVIDIA DGX-1™ AI supercomputer](#) into a containerized package. As part of the NGC, this package, called the NGC Software Stack, will be more widely available and kept updated and optimized for maximum performance.

To address the hardware challenge, NGC will give developers the flexibility to run the NGC Software Stack on a PC (equipped with a TITAN X or GeForce® GTX 1080 Ti), on a DGX system or from the cloud.

NGC will accelerate and simplify deep learning development by making it easier for developers to conduct deep learning training, experimentation and deployment. Developers will be able to easily design more sophisticated neural networks, process more data, iterate quickly and get to market faster.

NGC will offer the following benefits:

- **Purpose Built:** Designed for deep learning on the world's fastest GPUs.
- **Optimized and Integrated:** The NGC Software Stack will provide a wide range of software, including: Caffe, Caffe2, CNTK, MXNet, TensorFlow, Theano and Torch frameworks, as well as the NVIDIA DIGITS™ GPU training system, the NVIDIA Deep Learning SDK (for example, cuDNN and NCCL), nvidia-docker, GPU drivers and NVIDIA® CUDA® for rapidly designing deep neural networks.
- **Convenient:** With just one NVIDIA account, NGC users will have a simple application that guides people through deep learning workflow projects across all system types whether PC, DGX system or NGC.
- **Versatile:** It's built to run anywhere. Users can start with a single GPU on a PC and add more compute resources on demand with a DGX system or through the cloud. They can import data, set up the job configuration, select a framework and hit run. The output could then be loaded into TensorRT™ for inferencing.

With NGC, developers can build models of any size or type, using a versatile platform that makes it easier to move models from prototyping to deployment. They can increase or decrease computing resources, and only pay for what they need.

NGC is expected to enter public beta by the third quarter. Pricing will be announced at a later date. Learn more at www.nvidia.com/cloud.

Keep Current on NVIDIA

Subscribe to the [NVIDIA blog](#), follow us on [Facebook](#), [Google+](#), [Twitter](#), [LinkedIn](#) and [Instagram](#), and view NVIDIA videos on [YouTube](#) and images on [Flickr](#).

About NVIDIA

[NVIDIA](#)'s (NASDAQ: NVDA) invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined modern computer graphics and revolutionized parallel computing. More recently, GPU deep learning ignited modern AI -- the next era of computing -- with the GPU acting as the brain of computers, robots and self-driving cars that can perceive and understand the world. More information at <http://nvidianews.nvidia.com/>.

Certain statements in this press release including, but not limited to, statements as to the impact, benefits, availability and pricing of NVIDIA GPU Cloud 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 reports NVIDIA files with the Securities and Exchange Commission, or SEC, including its Form 10-K for the fiscal period ended January 29, 2017. 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.

© 2017 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, GeForce, CUDA, NVIDIA DIGITS, NVIDIA DGX-1 and TensorRT 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.

Media Contacts

Ken Brown

+1 408 486 2626

kebrown@nvidia.com