



NVIDIA DGX Cloud

Your own AI supercomputer—in the cloud.



Large language models (LLMs) and generative AI require an AI supercomputer, but many enterprises struggle with the complexity, effort, and time required to deploy and operate infrastructure. These businesses need immediate access to high-performance infrastructure at scale, with enterprise-grade tools and predictable costs that aren't easily obtainable with traditional infrastructure solution providers.

NVIDIA DGX™ Cloud is the world's first AI supercomputer in the cloud, a multi-node AI-training-as-a-service solution designed for the unique demands of enterprise AI. DGX Cloud provides a full-stack AI developer suite, leadership-class infrastructure, and direct access to AI experts, allowing businesses to get started immediately on AI innovation, with no-surprise predictable pricing.

The Best of NVIDIA AI

With NVIDIA DGX Cloud, developers have access to the best of NVIDIA AI innovation. Unlike DIY approaches to AI training, DGX Cloud provides an all-in-one solution with software for streamlining developer workflows. Powered by NVIDIA Base Command™ Platform and featuring a single-pane-of-glass user interface, DGX Cloud delivers a consistent user experience across cloud and on premises. DGX Cloud also includes the NVIDIA AI Enterprise suite, which comes with AI solution workflows, optimized frameworks, and pretrained models to unlock developer productivity.

Your Own AI Supercomputer in the Cloud

DGX Cloud provides enterprises immediate access to an AI supercomputer from a web browser. With DGX Cloud, teams can access multi-node AI training at scale to speed complex model development. If businesses choose to expand to on-premises infrastructure, they can consolidate all of their resources to deliver a hybrid cloud leveraging Base Command Platform, which provides developers a consistent, single-pane view.

Expertise at Your Fingertips

With DGX Cloud, you gain access to NVIDIA AI expertise to help maximize performance and achieve desired results faster. DGX Cloud also offers 24/7 business-critical support, a dedicated technical account manager, and a customer service success manager. Unlike DIY approaches, DGX Cloud provides single-point-of-contact support. Enterprises experience this concierge-like support for the duration of their contract.

Key Features

Software

- > NVIDIA Base Command Platform
- > NVIDIA AI Enterprise software suite
- > Hybrid-cloud support

Hardware

- > Multi-node capable
- > 8 NVIDIA A100 Tensor Core GPUs per node (640GB total)
- > Access to the latest GPU technology
- > 10TB storage per instance
- > 10TB egress per month per tenant
- > The ability to scale up storage and egress bandwidth as needed

Services

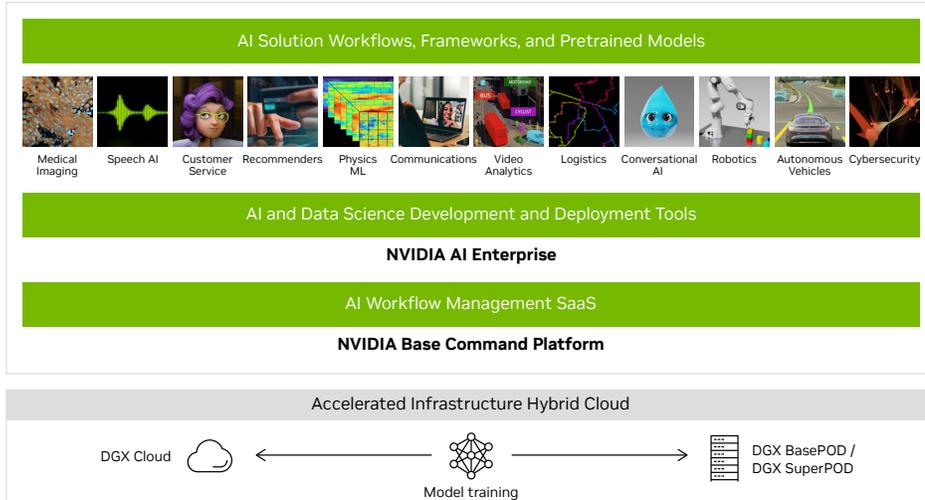
- > Access to NVIDIA AI experts
- > 24/7 business-critical support
- > Technical account manager
- > Customer service success manager
- > Single-point-of-contact support

Predictable Costs, No Surprises

DGX Cloud delivers financial transparency to enterprises, with a single and predictable monthly price that includes software, compute, storage, data egress, AI experts, and enterprise support. Now, businesses can access an AI supercomputer without the concern of spiraling costs.

NVIDIA DGX Cloud - AI Software Stack

Built on NVIDIA AI Enterprise and NVIDIA Base Command Platform



The DGX Cloud offering

Your AI Factory in the Cloud

Enterprises that want to tackle modern AI applications such as LLMs and generative AI need the right infrastructure solution with the simplicity of cloud. With one price, DGX Cloud provides enterprises with immediate access to an AI supercomputer in the cloud, accessible from a web browser, leading-edge software to accelerate AI development, and access to NVIDIA AI expertise and support.

Ready to Get Started?

To learn more about NVIDIA DGX Cloud,
nvidia.com/dgx-cloud

© 2023 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, Base Command, and DGX 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. 2660677. MAR23

