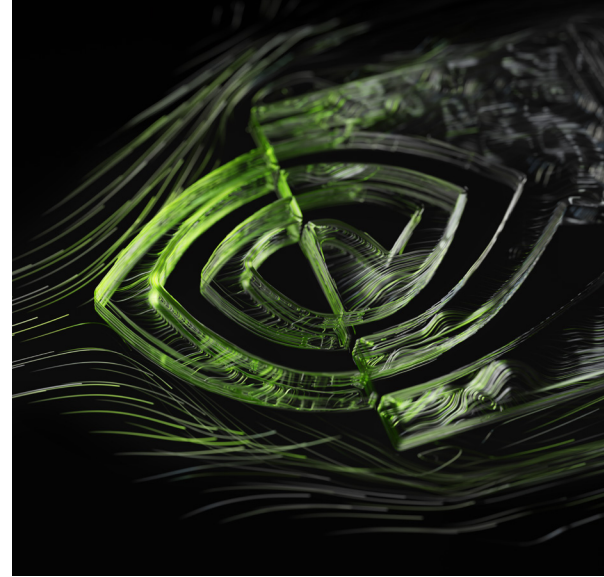




NVIDIA DGX SuperCloud

The fastest path to your AI hybrid cloud.



Enterprises everywhere are transforming supply chains, customer experiences, and clinical outcomes with the power of AI-fueled insights. But to develop the modern AI models that can transform your business, your developers need the infrastructure, tools, and expertise that can achieve faster time to value for AI investment. AI workloads place unparalleled strains on IT, and traditional cloud services are not optimized for the unique demands of prototyping, iterating, and operationalizing AI-infused applications at scale.

NVIDIA DGX SuperCloud™ is the world's first AI platform as a service (PaaS) optimized for the unique demands of enterprise AI development. Powered by an architecture based on the world's most successful AI compute platform, NVIDIA DGX™, business leaders, developers, and IT teams can all benefit from this first-of-its-kind cloud service that places AI supercomputing resources within reach of every AI initiative.

The Cloud-First Way to Get DGX Performance for Enterprise AI

With DGX SuperCloud, you can access the power of NVIDIA DGX in a cloud service. More than record-breaking infrastructure that can tackle your most complex AI models, DGX SuperCloud brings together leadership-class performance with a comprehensive, fully supported software stack optimized for streamlining AI developer workflows—so you can realize more of your valuable AI models in production.

Enterprise-Ready AI Developer Software

NVIDIA AI Enterprise, the broadly adopted software of the NVIDIA AI platform, includes essential frameworks, pre-trained models and enterprise support to streamline each step of the AI workflow, from data processing and AI model training to simulation and large-scale deployment. NVIDIA AI Enterprise enables organizations to develop predictive models to automate business processes and gain rapid business insights with applications such as conversational AI, recommender systems, fraud detection, computer vision and more.

Benefits:

- > The best AI platform with the simplicity of cloud
- > Supercomputing performance without a data center
- > Optimized AI frameworks, data science tools, and pretrained models that speed innovation
- > Turnkey AI platform as a service, managed by NVIDIA
- > Consistent user experience for development across on-prem, hybrid, and cloud environments
- > Enjoy a fast start on AI development and scale in the cloud or on prem as needed

DGX SuperCloud includes:

- > A monthly subscription to eight-GPU cloud-hosted nodes, scalable from a single system to thousands
- > Powered by NVIDIA Base Command Platform
- > Includes the NVIDIA AI Enterprise software suite
- > Provides direct access to NVIDIA AI experts
- > A fully managed service backed by NVIDIA Enterprise Support, a dedicated technical account manager, and a customer success manager

Powered by NVIDIA Base Command Platform

NVIDIA Base Command™ Platform is a comprehensive and intuitive AI workflow management software as a service (SaaS) for your developers. It runs across all infrastructure types (on premises and in cloud), ensuring mobility of workloads and the best economics across on-premises and cloud resources. With an intuitive graphical user and command-line interface and integrated monitoring and reporting dashboards, Base Command Platform simplifies the end-to-end lifecycle management of AI development, including workload management and resource sharing.

Unify AI Development Across Your Hybrid Cloud

Your team is cloud-first, but you still have siloed IT resources that include infrastructure on premises or in colocation data centers. DGX SuperCloud is the first cloud service to deliver a unified control plane that lets your developers manage their workflow across all infrastructure types, including premises-based NVIDIA DGX BasePOD™ and DGX SuperPOD deployments and your DGX SuperCloud. This enables mobility of workloads and delivers the benefits of an enterprise AI hybrid cloud with a single view to manage it all.

Base Command Platform at a glance:

- > Single-GPU, multi-GPU, and multi-node AI training
- > Custom scheduler with quota controls
- > Built-in dataset management and data governance
- > Role-based access control (RBAC), single sign-on (SSO), team collaboration, and sharing
- > Comprehensive job telemetry, including Tensor Core utilization and GPU affinity
- > Detailed showback reporting
- > GPU-optimized software from the NVIDIA NGC™ catalog
- > APIs for MLOps integration
- > Support for interactive jobs to access Jupyter Notebooks and TensorBoard

NVIDIA AI Enterprise at a glance:

- > A cloud-native suite of AI and data analytics software optimized for the development and deployment of AI
- > Reduces development complexity and speeds time to business insights
- > Enterprise reliability, security and scalability
- > Assurance of NVIDIA Enterprise Support including guidance on configuration and performance
- > Priority notifications of the latest security fixes and maintenance releases
- > Long term support for designated software branches

Ready to Get Started?

To learn more about NVIDIA DGX SuperCloud, visit www.nvidia.com/en-us/data-center/dgx-supercloud/ or talk with your NVIDIA representative.

© 2022 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, Base Command, DGX, DGX BasePOD, DGX SuperCloud, DGX SuperPOD, and NGC 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. All other trademarks are property of their respective owners. 2503951. OCT22

