



NVIDIA DGX Foundry

The Development Hub for the World's Most Demanding AI Enterprises

Many businesses find themselves sinking in AI model debt with data science innovation languishing in prototypes and rarely making it to production. Businesses need a solution that can accelerate their AI initiatives with simplified experimentation and workflows and streamlined management of users, jobs, and datasets—without needing to build it themselves.

Need the highest performance infrastructure for complex AI development? Try NVIDIA DGX™ Foundry with NVIDIA Base Command™. It's fully managed by NVIDIA and available for a short, free trial in NVIDIA LaunchPad or with a monthly subscription for your most demanding AI projects.

NVIDIA DGX Foundry, which includes NVIDIA Base Command software and uses NetApp storage, is a world-class infrastructure solution for businesses and their data scientists who need a premium AI development experience without the struggle of building it themselves. Offered as a hosted solution for end-to-end AI development and including access to fully managed NVIDIA infrastructure based on the NVIDIA DGX SuperPOD™ architecture, DGX Foundry is the development hub for the world's most demanding AI enterprises.

Accelerate AI Initiatives

For enterprises to succeed in AI, data scientists need to be able to define, configure, and execute their workloads, collect and inspect results, and iterate to further innovate. At the same time, management demands comprehensive reporting on AI infrastructure utilization to set priorities and plan for the future. NVIDIA DGX Foundry provides an immediately usable, hosted solution for AI development without the overhead and pitfalls of deploying and running a do-it-yourself platform.

Coordinated Access to Premium AI Infrastructure

IT organizations often struggle to provide simple yet powerful tools to all of the AI researchers and scientists in their organization who need them. NVIDIA Base Command, which is cloud-hosted software that comes with the DGX Foundry subscription, offers the easiest way to give them access to the power of NVIDIA DGX systems and NetApp storage in parallel but without interference.



NVIDIA DGX FOUNDRY SUBSCRIPTION DETAILS

Software	<p>NVIDIA Base Command-hosted AI workflow software</p> <p>Includes:</p> <ul style="list-style-type: none"> > Single-GPU, multi-GPU, and multi-node AI training > Custom scheduler with quota controls > Built-in dataset management > Comprehensive job telemetry > 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
Compute	<p>NVIDIA DGX systems</p> <p>Access to a minimum of three nodes</p> <p>Each node is an NVIDIA DGX™ A100 with:</p> <ul style="list-style-type: none"> > 8x NVIDIA A100 Tensor Core GPUs > 640GB of GPU memory > 5 petaFLOPS of AI performance > NVIDIA InfiniBand networking
Storage	<p>A dedicated high-availability (HA) pair of NetApp AFF A800</p>
Networking	<p>Compute nodes: 8x 200Gb/s HDR InfiniBand</p> <p>Storage: 100Gb/s Ethernet</p> <p>Internet access: 10Gb/s</p>
Support	<p>Fully managed service backed by NVIDIA Enterprise Support and a dedicated technical account manager</p>
Data center	<p>Hosted at Equinix</p>

Comprehensive AI Workflow Management

Base Command offers both a comprehensive, cloud-based UI and a complete command line interface to efficiently execute AI workloads with right-sized accelerated resources—from a single GPU to a multi-node cluster—with data sets hosted on integrated, dedicated NetApp storage. A vast library of prebuilt containers with NVIDIA-optimized deep learning and data science frameworks and pretrained models is available via the NVIDIA NGC catalog, which contains curated sets of GPU-optimized software for AI, high-performance computing (HPC), and visualization. This lets users deliver production-ready models and applications sooner and eliminates the overhead of systems management.

Integrated Monitoring and Reporting Dashboards

AI projects are multi-faceted and iterative in nature, requiring constant fine-tuning. Base Command has built-in telemetry directly accessible by the end user. The information can be used to validate their deep learning techniques, workload settings, and resource allocations to enable constant improvement. Base Command also includes reporting and showback capabilities for business leaders who want to measure AI projects against business goals.

AI the Way NVIDIA Does It

NVIDIA's own engineers and researchers have ever-evolving demands as they build state-of-the-art AI, and they rely on the same AI infrastructure every day. NVIDIA is continuously innovating to meet their needs and is committed to providing these enhancements to the Base Command software and NVIDIA DGX Foundry. Your end users get the new functionality that's vetted by hundreds of NVIDIA researchers first, with the efficiency and reliability that our researchers are accustomed to.

[Learn more](#)

To learn more about NVIDIA DGX Foundry, visit nvidia.com/dgx-foundry