Turnkey Enterprise MLOps & Infrastructure Solution
Democratizing the AI enterprise data center with NVIDIA DGX SuperPOD™ in Domino

Domino makes data science at scale a reality

For companies with large teams of code-first data scientists, Domino is an Enterprise MLOps platform that accelerates the process of developing and productionizing data science work. In close collaboration with NVIDIA, Domino eliminates roadblocks to cutting-edge research and experimentation in artificial intelligence and machine learning with GPU workload orchestration on NVIDIA DGX SuperPOD™ systems.

Integrated Solution

Under the hood, Domino automates the DevOps activities required to optimize utilization of the powerful NVIDIA DGX SuperPOD infrastructure, eliminating the low-value configuration and debugging tasks performed by valuable researchers. Domino provides flexible, governable access to GPU resources, and it blends these workloads seamlessly with traditional infrastructure across a single system of record. Now, data scientists have self-service access to GPU resources from their Domino workbench, with governance and access control that satisfy the strict requirements of Enterprise IT organizations. All of the tools for experiment management, collaboration, reproducibility, governance, and operationalizing models are included in Domino.

Orchestrate DGX SuperPOD resources from the workbench to production

The Domino platform, together with DGX SuperPOD systems, supports open, collaborative, reproducible research free of DevOps constraints, and is powered by fast, efficient end-to-end compute. Streamlined use of accelerated DGX SuperPOD compute, accessible in the Domino platform, allows data scientists to focus on model work, while IT supports the sprawl of data science initiatives from a single pane of glass.

Provide Self-Serve Access to GPUs

NVIDIA DGX SuperPOD systems are easily accessible via Domino so data scientists can focus on critical work, and IT teams can eliminate infrastructure configuration and debugging. Resources can be configured within the Domino Compute Grid, rather than depending on IT for one-off tools and deployments, reducing time spent on DevOps.

Scale Across Multi-node Resources with Flexible Compute Frameworks

Domino enables the automatic creation and management of multi-node clusters, releasing them when training is done—eliminating dedicated resources and low utilization. Domino supports ephemeral clusters using Spark, Ray, and Dask.

For example, on 140 nodes, thousands of data scientists can be running GPU notebooks or large workloads concurrently, democratizing the data center for model-driven enterprises.

Drive Utilization of SuperPOD Resources

Domino administrators can easily support a variety of different users and use cases. Provide up thousands of concurrent data scientist sessions with GPUs with NVIDIA Multi-Instance GPU (MIG) technology, run massive multi-node jobs, and host thousands of inference models.

Govern Usage of GPUs by Role

Domino gives IT visibility into who is accessing GPU resources and how they are being used. Permissions can be set to ensure employees without proper entitlements are not burning through valuable resources, while power users have full access to maximize the use of hardware. People and organizations can be assigned access by role (i.e., according to development, validation, or production functions) with different corresponding hardware available.

---

**Figure 1**: Choose the right hardware and software for any task. Use Domino Workspaces to run proprietary and open source tools side by side, governing scalable GPU Hardware Tier access with Domino’s Compute Grid. Here, a JupyterLab IDE is configured to launch with access to a single NVIDIA A100 Tensor Core GPU from a DGX A100 with PyTorch 1.9 and TensorFlow 2.5.
Domino supports the broadest ecosystem of tools and infrastructure

**Built on a robust services layer**

**With all the tools & tech preferred by data scientists**

**Enabling the highest level of sharing on powerful infrastructure**

### Accelerate value from your AI infrastructure

Domino accelerates data science initiatives—at velocity and scale—by complementing NVIDIA DGX SuperPOD with the best-in-class Enterprise MLOps platform. The NVIDIA DGX SuperPOD solution is purpose-built to meet the demands of enterprise AI and data science, delivering the fastest start in AI development, effortless productivity, and revolutionary performance - ushering in the era of cloud-native supercomputing.

Consolidated and centralized support with version control for data science workspaces ensures stable and consistent access to cutting-edge deep learning compute and frameworks such as Keras, TensorFlow, Torch, and TensorRT. That way, when data science teams deploy their on-demand notebooks, they select tailored DGX resources and software for their tasks with administrator controlled permissions. And, the compute environment is fully traceable. The Domino open platform streamlines workflows and scales, taking full advantage of the power of NVIDIA DGX systems and NVIDIA NGC™ optimized containers out-of-the-box.

“Domino makes it easy for our data scientists to rapidly access NVIDIA GPUs so we can support complex use cases like training deep neural networks.”

Mike Johnson, Lead Data Scientist, Lockheed Martin

### A system-of-record for models

Unifying all data science work streams into one common platform makes collaboration effortless in Domino. Version control of all models, tools, and environments is automated. Collaborators see their team's work, then quickly access and fork prior works to progress research, building upon the efforts of their peers. Domino is the system of record for all data science efforts that helps align IT and data science teams and standardize on best practices—one system lets organizations manage and organize all data science work.

**Learn More**

[dominodatalab.com/partners/nvidia](http://dominodatalab.com/partners/nvidia)

Domino

**Website:** dominodatalab.com

**Blog:** blog.dominodatalab.com

**Try:** dominodatalab.com/trial

NVIDIA

**Website:** nvidia.com/DGX

**Blog:**

blogs.nvidia.com/blog/category/enterprise

---

**Figure 2:** A Domino Workspace is used to validate in-notebook access to a single NVIDIA A100 Tensor Core GPU from a DGX A100 for development.