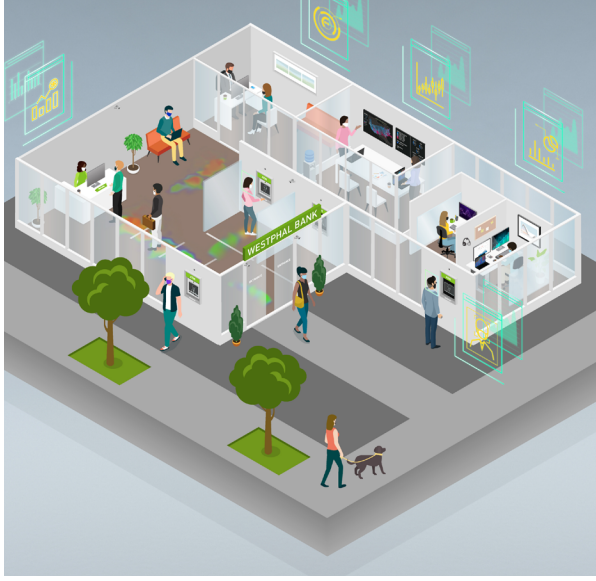
Solution Overview

NVIDIA DGX BasePOD for the Financial Services Industry

Streamline AI development and deployment.



Reducing Costs, Mitigating Risk, and Enhancing Customer Experiences

With today's volatile markets, the ability to generate deeper market insights, calculate risk faster, accelerate fraud detection, and offer conversational AI services can contribute directly to the bottom line. *Business Insider* estimates that the potential for AI-driven cost savings for banks alone will reach $447 billion by 2023, and front- and middle-office AI improvements could represent more than 90 percent of these savings.[1] AI is also proving to be essential in trading. According to Deutsche Bank, 90 percent of equity-futures trades and 80 percent of cash-equity trades are executed by algorithms without any human input.[2]

NVIDIA has made it easier, faster, and more cost-effective for financial institutions to deploy mission-critical AI use cases. By combining the proven performance, scale, and manageability of the NVIDIA DGX BasePOD™ architecture with industry-tailored software and tools from the NVIDIA AI Enterprise software suite, enterprises have a trusted, full-stack platform for building and deploying their AI applications.

To speed the delivery of AI-powered use cases in financial services, NVIDIA has delivered the DGX BasePOD infrastructure solution for the financial services industry (FSI) (figure 1), optimized to streamline AI development and deployment for critical applications such as conversational AI, algorithmic trading, and fraud detection. This solution includes proven, open-source containers and frameworks that have been certified to run securely, both on premises and in the cloud, on the most demanding FSI workloads.

The Value of NVIDIA AI Enterprise

DGX BasePOD includes the **NVIDIA AI Enterprise** software suite, which contains the key building blocks required to develop and deploy domain-specific, end-to-end AI workflows—from data prep and training to inference and deployment. AI practitioners can choose to train on complex neural network models, as well as tree-based machine learning models. The suite's proven, open-source containers, applications, and frameworks include NVIDIA TAO™ Toolkit for document automation and NVIDIA Triton™ Inference Server for streamlining and standardizing AI inference, enabling teams to deploy, run, and scale AI models from any framework on your DGX BasePOD. A broader portfolio of NVIDIA frameworks ease adoption and

Benefits

- Eliminates design complexity
- Accelerates deployment
- Delivers predictable performance at scale
- Includes the NVIDIA software stack proven to optimize financial services applications and development
- Comes with full-stack expertise from NVIDIA Enterprise Support

Accelerate FSI Workloads With NVIDIA DGX BasePOD and NVIDIA AI Enterprise Suite

- Document automation with NVIDIA TAO Toolkit
- Cybersecurity with NVIDIA Morpheus
- Federated learning with NVIDIA FLARE™
- Automated speech recognition and text-to-speech with NVIDIA® Riva
- Monte Carlo simulations for risk analysis with the NVIDIA HPC SDK
- Virtual assistants and ESG analysis with NVIDIA NeMo
- AI inferencing with NVIDIA Triton Inference Server

1. *Business Insider.* Winning Strategies for AI in Banking.
2. *The Economist.* The Stockmarket Is Now Run by Computers, Algorithms, and Passive Managers.

accelerate key FSI workloads, including cybersecurity; Monte Carlo simulation for risk analysis; natural language processing; virtual assistant; environmental, social, and governance (ESG); and more. This combination gives organizations access to a fully integrated solution of AI-accelerated software and hardware that lets them quickly deploy, streamline, and accelerate their AI workloads. And because enterprise-class support is included, organizations get the transparency of open-source backed by the assurance that the global NVIDIA Enterprise Support team will help AI projects stay on track.

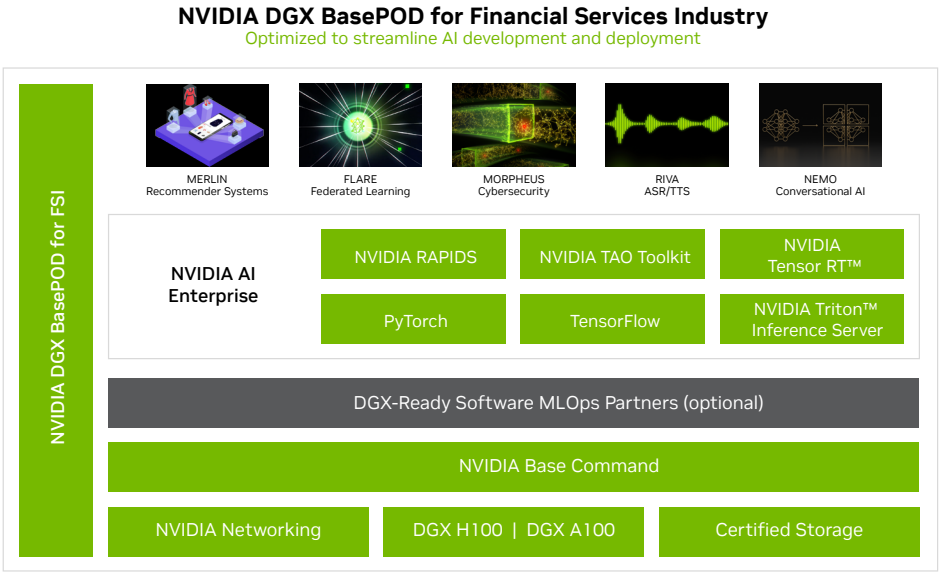


Figure 1. DGX BasePOD for the financial services industry

Powered by NVIDIA Base Command

Included with DGX BasePOD is **NVIDIA Base Command™**, a proven platform that includes AI infrastructure management tools and acceleration libraries to ensure a high-performance, highly efficient environment. Base Command provides cluster and workload management, support for multiple operating environments to optimize AI workflows across hybrid infrastructure, and the **NVIDIA Magnum IO™** portfolio of infrastructure acceleration technologies. This fully integrated solution delivers the highest performance and utilization in the industry. By providing the AI software, compute power, tools, and support needed, it gives organizations of any size access to enterprise-class, accelerated infrastructure, so they can focus on creating business value from AI.

Effortless AI Training and Cluster Management Across Hybrid Infrastructure

Many companies have trouble scaling AI across on-prem and cloud instances due to incompatibilities with their software control planes. **NVIDIA DGX Cloud** is a high-performance, multi-node AI-training-as-a-service solution for unifying AI development across the enterprise, from cloud to on prem. Using **NVIDIA Base Command Platform,** which powers DGX Cloud, data scientists and MLOps leaders get a single-pane-of-glass view into dataset management and accelerated computing utilization across any infrastructure configuration, from on-prem to hybrid-cloud environments. Having one platform across instances, regardless of location, increases operational efficiency, lowers costs, maximizes compute utilization, and accelerates the creation of valuable AI-enabled applications.

"Figure 2. NVIDIA DGX Cloud's AI software stack.

A Strong Ecosystem of Proven Partners

NVIDIA DGX BasePOD for FSI solutions are certified by NVIDIA and include a qualified and proven ecosystem of storage partners. They use the Magnum IO portfolio for intelligent data center input/output (IO) and include technologies like NVIDIA GPUDirect® Storage, which provides the highest-performance IO directly to the GPUs powering the AI infrastructure, accelerating jobs like image processing. DGX BasePOD for FSI, fully integrated and tested with the partner ecosystem, also simplifies the deployment of on-prem accelerated AI infrastructure for enterprise IT organizations.

Supported by NVIDIA

With NVIDIA DGX BasePOD, both AI practitioners and IT administrative teams have access to NVIDIA experts globally. This provides coordinated support across the full solution, including partner products, control over upgrade and maintenance schedules with long-term support (LTS) options, and access to instructor-led customer training and knowledge base resources.

Pro Tip

Enable Your Hybrid-Cloud Journey With DGX BasePOD

With NVIDIA DGX Cloud, teams can access multi-node AI training in a convenient service to speed model development. Begin your AI journey with DGX Cloud, and then easily scale with DGX BasePOD on premises for a unified hybrid AI cloud. Combining NVIDIA DGX™-based infrastructure, whether on premises or in the cloud, enterprises can use the **NVIDIA Base Command Platform** to manage and orchestrate AI workloads in the hybrid cloud.

Ready to Get Started?

To learn more about NVIDIA DGX BasePOD, visit **nvidia.com/dgx-basepod**

To learn more about NVIDIA AI Enterprise, visit **www.nvidia.com/ai-enterprise-suite**