SOLUTION BRIEF

SUPERCHARGE AI IN FINANCIAL SERVICES WITH NVIDIA DGX BASEPOD AND NETAPP STORAGE

Reduce costs, mitigate risk, and enhance customer experiences

The opportunity

In today's fast-paced and ever-changing markets, staying ahead of the game is crucial for financial institutions. But how can you gain that competitive edge? The answer lies in the transformative power of artificial intelligence (AI). AI has the potential to revolutionize the way that you do business, from generating deeper market insights, to detecting fraud faster, to offering personalized services through conversational AI.

Imagine being able to analyze vast amounts of data and to identify unusual patterns with lightning speed. AI can supercharge your fraud detection efforts, enabling you to stay one step ahead of the bad actors. It can also enhance your credit decisions and risk management by analyzing complex datasets and market trends, providing you with valuable insights that help shape your strategies.

By processing an avalanche of real-time data and by executing trades at lightning speed, AI empowers you to seize opportunities before they slip away. The days of relying solely on human intuition are fading into the past. According to Deutsche Bank, a staggering 90% of equity futures trades and 80% of cash equity trades are now executed by algorithms without any human intervention.[1]

The financial rewards of embracing AI are equally compelling. Business Insider estimates that AI-driven cost savings for banks alone could reach an $447 billion in 2023.[2] The potential for savings is immense, with front- and middle-office AI improvements accounting for more than 90% of these savings.

By investing in AI infrastructure, you future-proof your institution, keeping you competitive in an industry that is evolving at breakneck speed.

The challenges

Financial institutions like yours face critical challenges stemming from siloed data and incomplete information, which can significantly impede the potential of AI-driven outcomes. But you can overcome these challenges to avoid severe consequences. You must prioritize data management practices like cleansing, standardization, and validation to reduce the potential for compromised decision-making, missed opportunities, and ineffective fraud detection and risk assessment. And you need accurate AI models to help decrease risk exposure and to provide sufficient customer insights.

A unified hybrid data architecture is a must to facilitate scalability and innovation, enabling your institution to adapt to AI's dynamic nature. You also must follow adequate governance practices to avoid exposing your organization to security breaches, regulatory noncompliance, and reputational harm. And to prevent wasted investments, escalated costs, and suboptimal AI performance, you need efficient resource utilization.

Furthermore, your organization must optimize collaboration and productivity among your teams to promote progress, innovation, and the delivery of personalized services. Your financial institution risks falling behind competitors who successfully tackle these challenges, leading to decreased market share and eroded customer trust.

KEY BENEFITS

- Simplified design
- Faster deployment
- Predictable performance at scale
- NVIDIA software stack, proven to optimize financial services applications and development
- Full-stack expertise from NVIDIA Enterprise Support
- Lower TCO for high-performing flash storage
- Faster model to production with near-instantaneous data clones with the NetApp® ONTAP® AI proven architecture
- Model traceability, model determinism, and simplified versioning with NetApp ONTAP AI Snapshot™ copies
- Elevated data security FIPS-140-2 encryption, data isolation with multitenancy, and automatic ransomware detection with ONTAP AI

[1] The Economist, The stockmarket is now run by computers, algorithms and passive managers, October 5, 2019.
[2] Business Insider, Winning Strategies for AI in Banking.

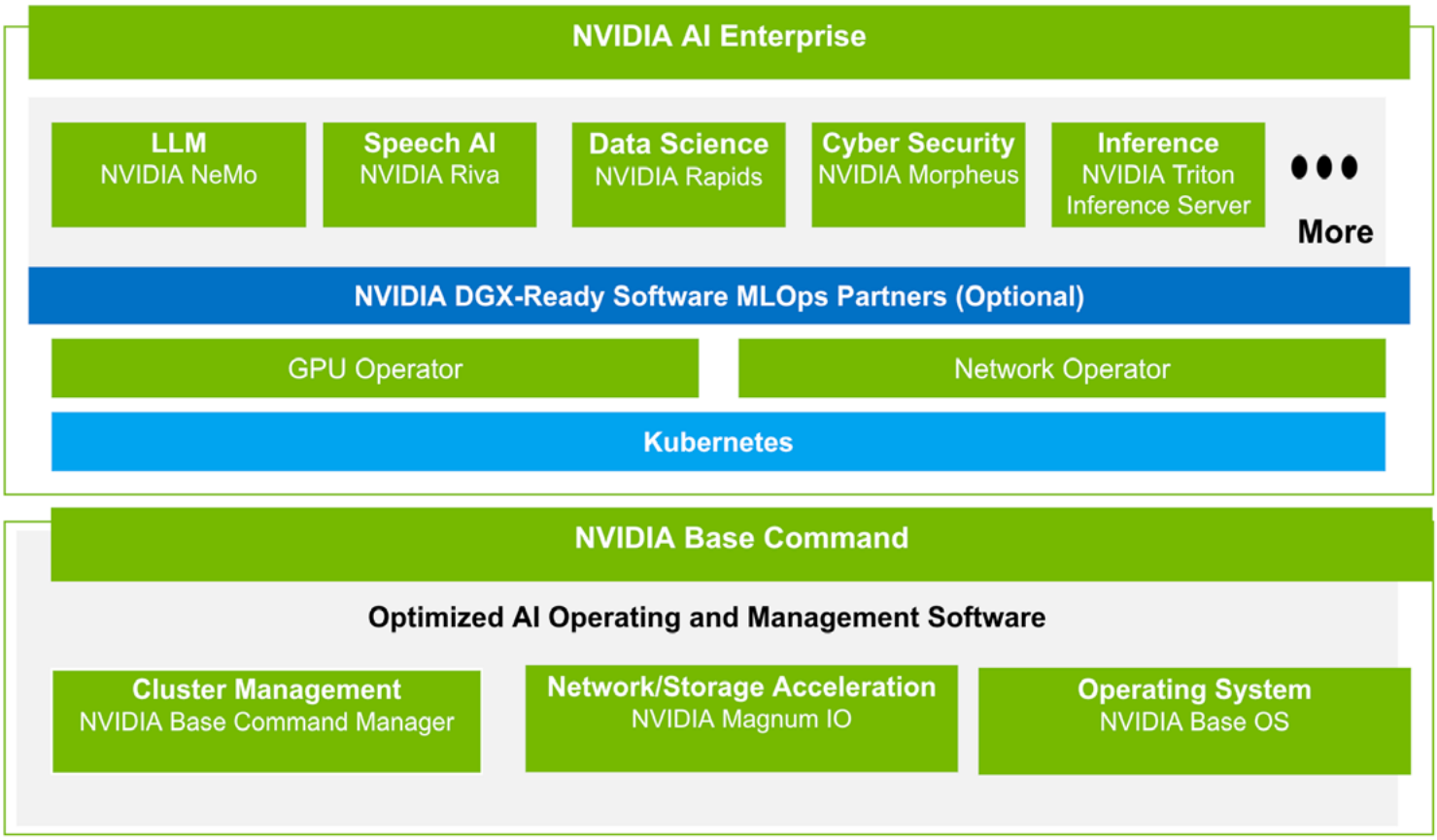

Figure 1: DGX BasePOD for the financial services industry.

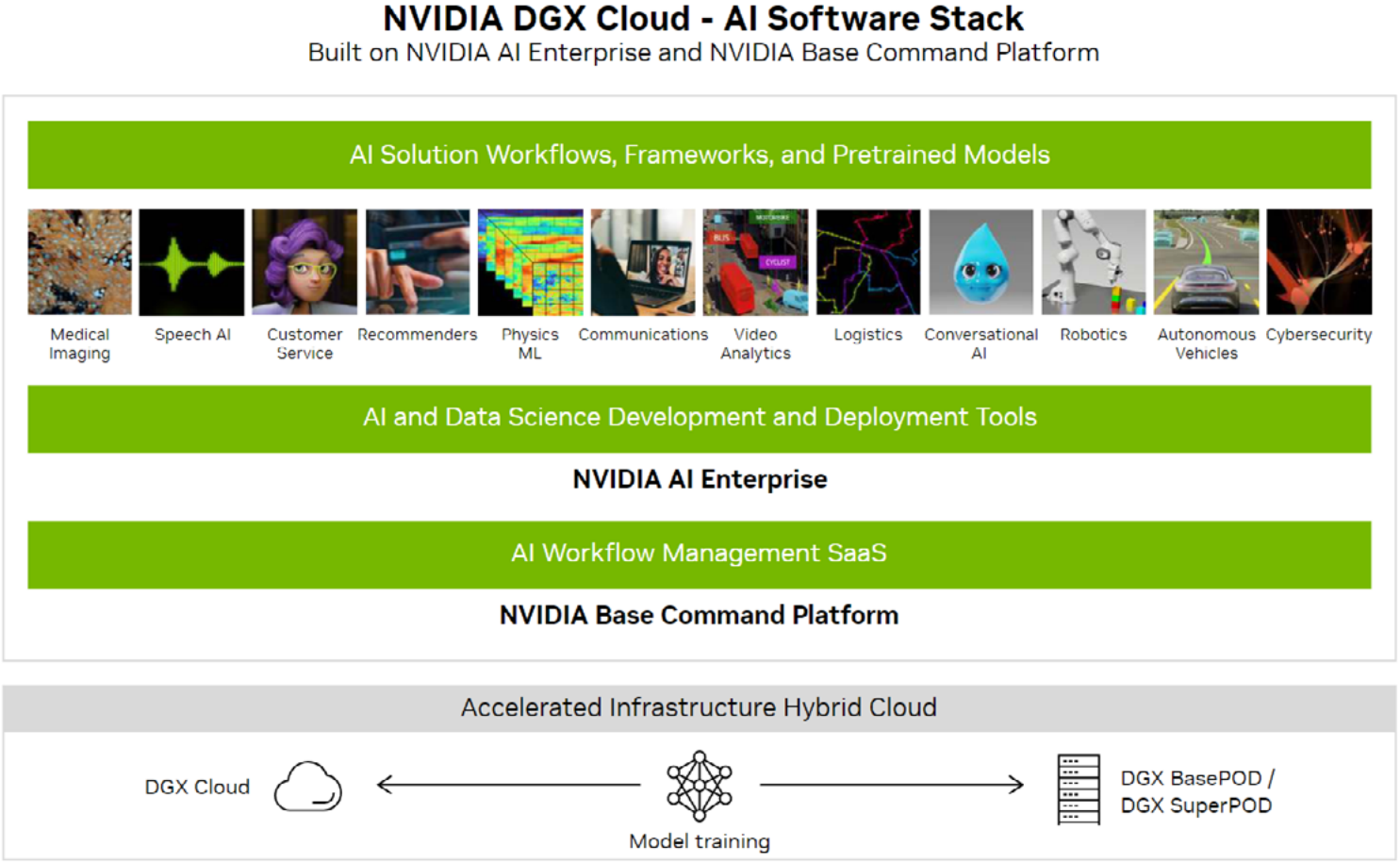

Figure 2: NVIDIA DGX Cloud AI software stack.

The consequences of not addressing these challenges include missed business opportunities, compromised security and compliance, inefficient resource allocation, reduced competitiveness, and diminished customer satisfaction. To thrive in the AI-driven financial landscape, your institution must confront these challenges head-on, unlocking AI's full potential and enabling informed decisions, operational efficiency, and personalized services. Embracing the transformative power of AI is pivotal for shaping the future of the financial services industry and for helping you stay ahead in a rapidly evolving market.

Reduce costs, mitigate risk, and enhance customer experiences

Together, NetApp and NVIDIA have made it easier, faster, and more cost-effective for financial institutions like yours 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, your enterprise has a trusted, full-stack platform for building and deploying your AI applications. And by combining NVIDIA technology with NetApp all-flash hybrid cloud–enabled storage arrays, you can choose the clouds that best fit your data science workloads.

High-performance NetApp AFF and C-Series storage systems enable point-in-time instantaneous NetApp Snapshot copies and clones, which enable ultrafast iteration and traceability of models. Built-in data efficiencies and multiprotocol support lower your TCO without sacrificing performance on NetApp storage. You get unparalleled protection and risk reduction on NetApp storage with robust multitenancy. NVIDIA handles the high-performance compute while NetApp technology makes it simple to securely optimize your data pipelines.

Enable your hybrid cloud journey with DGX BasePOD

With NVIDIA DGX Cloud, your teams can access multinode 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. By combining NVIDIA DGX™-based infrastructure, whether on premises or in the hybrid cloud, your enterprise can use the NVIDIA Base Command Platform to manage and to orchestrate AI workloads in the hybrid cloud.

To speed the delivery of AI-powered use cases in financial services, NVIDIA offers the **DGX BasePOD infrastructure solution for the financial services industry** (Figure 1). This solution is optimized to streamline AI development and deployment for critical applications such as conversational AI, algorithmic trading, and fraud detection. It 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 workloads.

The value of NVIDIA AI Enterprise

DGX BasePOD includes the NVIDIA AI Enterprise software suite,* which contains the key building blocks that you need to develop and to deploy domain-specific, end-to-end AI workflows—from data preparation and training to inference and deployment. AI practitioners can choose to train on complex neural network models and on tree-based machine learning (ML) models. The suite's proven, open-source containers, applications, and frameworks include NVIDIA TAO™ Toolkit for document automation and NVIDIA Triton™ Inference Server to streamline and to standardize AI inference. So, your teams can deploy, run, and scale AI models from any framework on your DGX BasePOD. A broader portfolio of NVIDIA frameworks eases adoption and accelerates key workloads in financial services, including cybersecurity; Monte Carlo simulation for risk analysis; natural language processing; virtual assistants; environmental, social, and governance (ESG) standards; and more. This combination gives your organization access to a fully integrated solution of AI-accelerated software and hardware that lets you quickly deploy, streamline, and accelerate your AI workloads. And because enterprise-class support is included, you get the transparency of open source backed by the assurance that the global NVIDIA Enterprise Support team will help your AI projects stay on track.

workloads with NVIDIA DGX BasePOD, NVIDIA AI Enterprise Suite,* and NetApp DataOps Toolkit

- Documented automation with NVIDIA TAO Toolkit
- Cybersecurity with NVIDIA Morpheus
- Federated learning with NVIDIA FLARE™
- Automated speech recognition (ASR) and text-to-speech (TTS) with NVIDIA® Riva*
- Monte Carlo simulations for risk analysis with the NVIDIA HPC Software Development Kit (SDK)
- Virtual assistants and ESG analysis with NVIDIA NeMo
- Increased I/O throughput on the ONTAP AI stack with NVIDIA Magnum IO
- Orchestration of AI inferencing of NVIDIA Triton Inference Server with ONTAP AI
- Ability to allocate MIG instances to IDEs and to instantiate Triton Inference Servers with NetApp DataOps Toolkit
- Near-instantaneous self-service data science workspace creation, workspace cloning, and volume caches

* For details, see the NVIDIA AI Enterprise packaging, pricing, and licensing guide.

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 help you maintain a high-performance, highly efficient environment. Base Command provides cluster and workload management, support for multiple operating environments to optimize AI workflows across your 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. Because it provides the AI software, compute power, tools, and support necessary, no matter what size your organization is, you get access to enterprise-class, accelerated infrastructure so that you can focus on creating business value from AI.

The foundation for your AI hybrid cloud
Because of incompatibilities with their software control planes, many companies have trouble scaling AI across on-premises and hybrid cloud instances. NVIDIA DGX Cloud is a high-performance, multinode AI-training-as-a-service solution to unify AI development across the enterprise, from hybrid cloud to on premises. By using NVIDIA Base Command Platform, which powers DGX Cloud, your data scientists and MLOps leaders get a single-pane-of-glass view into dataset management and accelerated compute utilization across any infrastructure configuration, from on-premises to hybrid cloud environments. And by having one platform across instances, regardless of location, you increase operational efficiency, lower costs, maximize compute utilization, and speed the creation of valuable AI-enabled applications.

A robust partner relationship between NetApp and NVIDIA
NVIDIA DGX BasePOD for financial services includes a qualified and proven ecosystem to solve one of the toughest challenges for your data scientists: data access and movement between environments in a highly regulated industry. For instance, due to regulatory requirements, certain data is restricted from leaving the network and so must be trained cost-effectively within the network.

NetApp DataOps Toolkit enables the rapid creation of self-service data science workspaces, efficient workspace cloning and volume caches across clouds, seamless orchestration of the NVIDIA Triton Inference Server, and automated installation of NVIDIA AI Enterprise. This streamlined process effectively enhances the productivity of your data scientists by reducing the time that they spend on extract, transform, load (ETL) tasks. Furthermore, NetApp ONTAP Snapshot copies can help you meet regulatory requirements such as dataset-to-model traceability, model determinism, and simplified versioning. And to create immutable datasets and to enhance telemetry for reporting and chargebacks, your data scientists can use volume Snapshot copies and instantaneous clones and set up caches.

NVIDIA DGX BasePOD for financial services, certified by NVIDIA and NetApp, offers a reliable ecosystem to tackle the challenge of securely moving and accessing data in highly regulated industries to help your organization produce highly accurate AI models. With NetApp's data-centric workflows, you can effortlessly transfer data and improve data accessibility across various zones, regions, and first-party cloud storage services, giving your data scientists a consistent experience both on premises and in the cloud.

Supported by NVIDIA
With NVIDIA DGX BasePOD, both your AI practitioners and your IT administrative teams have access to NVIDIA experts globally. You get 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.

Contact Us

About NetApp
NetApp is the intelligent data infrastructure company combining unified data storage, integrated data services, and CloudOps solutions to turn a world of disruption into opportunity for every customer. NetApp creates silo-free infrastructure, then harnesses observability and AI, to enable the best data management. As the only enterprise-grade storage service natively embedded in the world's biggest clouds, our data storage delivers seamless flexibility and our data services create a data advantage through superior cyber-resilience, governance, and applications agility. Our CloudOps solutions provide continuous optimization of performance and efficiency through observability and AI. No matter the data type, workload or environment, transform your data infrastructure to realize your business possibilities with NetApp. www.netapp.com

NetApp