

AI-READY ENTERPRISE PLATFORM

Unleashing AI for Every Enterprise

Artificial intelligence (AI) is transforming every industry, whether it's by improving customer relationships in financial services, streamlining manufacturer supply chains, or helping doctors deliver better outcomes for patients. IDC projects that by 2024, 60 percent of G2000 will expand the use of AI and machine learning (ML) across all business-critical horizontal functions like marketing, legal, HR, procurement, and supply chain logistics.

While organizations understand the power of AI to transform their business and keep them competitive, the technology is still fairly unfamiliar to many IT and line of business stakeholders. Unlike traditional enterprise applications, AI apps are relatively new for enterprise IT departments. They're anchored in rapidly-evolving, open-source, bleeding-edge code and lack proven approaches that meet the rigors of scaled enterprise production settings. Many organizations struggle to identify the right foundation that will set them up for success and maximize the business value of AI.

AI-Ready Platform from NVIDIA

The NVIDIA AI Enterprise software suite enables organizations to harness the power of AI, even if they don't have AI expertise today. Optimized to streamline AI development and deployment, NVIDIA AI Enterprise includes proven, open-source containers and frameworks, certified to run on common data center platforms from VMware and Red Hat, mainstream NVIDIA-Certified Systems™ configured with GPUs or CPU-only, and on the public cloud. Since support is included, organizations benefit from the transparency of open source and the assurance that they can get help from the global NVIDIA Enterprise Support team to keep their AI projects on track. With the NVIDIA AI Enterprise software suite, AI is accessible to organizations of any size, and provides the compute power, tools, and support that organizations need to focus on creating AI business value.

KEY CHALLENGES FOR ENTERPRISE AI WORKLOADS

- > **Risk** – Pulling together an end-to-end AI solution from disparate products—and integrating them with existing infrastructure—is difficult.
- > **Performance** – Fast time-to-deployment and high performance are critical for AI, ML, and data analytics workloads.
- > **Scaling** – Going from proof-of-concept to enterprise deployment requires effective scaling—through efficient use of resources—to ensure manageability, availability, and infrastructure cost management.

BENEFITS FOR ENTERPRISES

- > **Optimized to simplify AI** development and deployment with included AI frameworks and containers, enabling insights to be gathered faster and business value to be delivered sooner.
- > **Certified to deploy anywhere**, including on popular data center platforms from VMware and Red Hat, mainstream NVIDIA-Certified Systems, and in the public cloud.
- > **Supported by NVIDIA AI experts** so AI projects stay on track, with included priority notifications, long-term support, and customized support upgrade options.



Figure 1. With the AI-Ready Platform, end users can access the software they need to build successful AI projects, and IT admins can support the projects using familiar tools.

Key Enabling Technologies

The NVIDIA AI-Ready Enterprise platform is comprised of the following building blocks:

NVIDIA AI Enterprise Software Suite

NVIDIA AI Enterprise is an end-to-end, cloud-native suite of AI and data analytics software, optimized so every organization can succeed with AI. It's also certified to deploy anywhere—from the enterprise data center to the public cloud—and includes global enterprise support and training.

NVIDIA AI frameworks and containers enable performance-optimized data science, training, and inference. They also simplify the building, sharing, and deployment of AI applications, so enterprises can gather insights faster and deliver business value sooner. For example, NVIDIA RAPIDS™ makes it possible for organizations to streamline the data science process with up to 70X faster performance, while improving cost-effectiveness by up to 20X. Additionally, NVIDIA TAO Toolkit boosts AI development by up to 10X, the NVIDIA Triton™ Inference Server allows organizations to improve TCO by 97X, and NVIDIA A100 enables them to achieve 100X lower latency when compared to CPU.

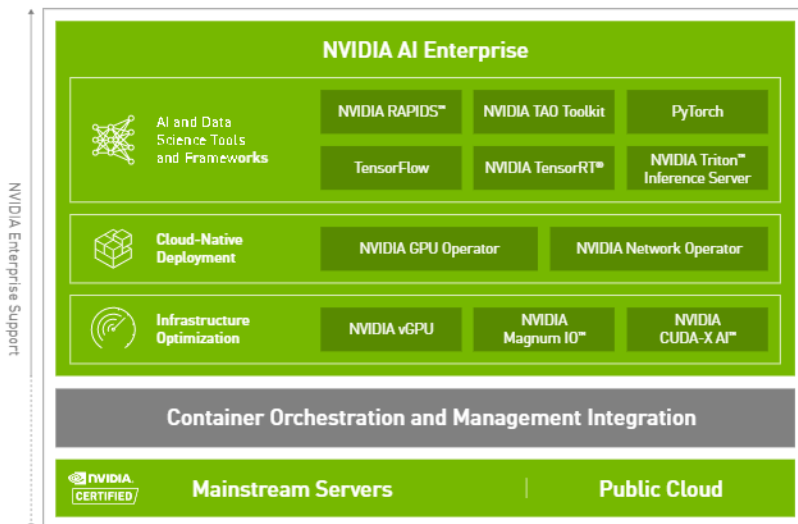


Figure 2. The NVIDIA AI Enterprise software suite includes the applications, frameworks, and tools used by AI researchers, data scientists, and developers, as well as tools for cloud-native deployments and infrastructure optimizations.

Container Orchestration and Management Software

NVIDIA AI Enterprise can be deployed anywhere—including in a virtualized enterprise data center with VMware vSphere with Tanzu, on mainstream NVIDIA-Certified Servers, with Red Hat's OpenShift container platform, in the public cloud, and even on CPU-only servers. Tested and certified by NVIDIA with broadly-adopted container, management, and server platforms, enterprises have the flexibility to deploy NVIDIA AI Enterprise with confidence, in just about any data center environment—on-premises, in a colocation facility, or on a private cloud. Organizations with a hybrid cloud strategy also have the flexibility to run NVIDIA AI Enterprise on GPU-accelerated public cloud instances on AWS, Azure, and Google with full NVIDIA support.

NVIDIA-Certified Systems

The NVIDIA AI Enterprise Suite is certified to run on NVIDIA-Certified Systems, which include the following:

- > NVIDIA Ampere architecture-based GPUs, the Tensor Core technology included in the Ampere architecture, deliver dramatic speedups to AI operations, reduced training times from weeks to hours, and provide massive inference acceleration.
- > NVIDIA® ConnectX® SmartNICs and the NVIDIA BlueField® data processing unit (DPU) provide a host of software-defined hardware engines for accelerating networking and security. These enable the best of both worlds: best-in-class AI training and inference performance with all the necessary levels of enterprise data privacy, integrity, and reliability.
- > NVIDIA Converged Accelerators combine the powerful performance of the NVIDIA Ampere architecture with the enhanced security and latency-reduction capabilities of the NVIDIA BlueField-2 DPU. Enterprises can use converged accelerators to create faster, more efficient, and secure AI systems in data centers and at the edge.

NVIDIA Enterprise Support

With NVIDIA AI Enterprise, organizations get the transparency of open source with the assurance of full enterprise support and platform certification. They can extend their team to include NVIDIA experts, get support ticket prioritization and coordinated support across the full solution and partner products until resolution, control upgrades and maintenance schedules with long term support (LTS) options, and access the latest customer training and knowledge base resources.

NVIDIA AI Enterprise support includes:

- > Full enterprise-grade support for every deployment option, including bare metal, virtualized, containerized, GPU and CPU, and public cloud.
- > Access to NVIDIA AI experts during local business hours for guidance on configuration and performance, including access to engineering.
- > Priority notifications related to the latest security fixes and maintenance releases.
- > Long-term support for up to 3-years for designated software branches.
- > Customized support upgrade options, including a designated technical account manager (TAM) and Business Critical support for 24x7 live agent access.

NVIDIA-CERTIFIED SYSTEMS

- > Confidently deploy scalable hardware and software solutions that securely and optimally run accelerated workloads.
- > Learn more about accelerated servers at [nvidia.com/certified-systems](https://www.nvidia.com/certified-systems)



NVIDIA AI Enterprise Trial Programs

NVIDIA offers the following trial programs that enable enterprise customers to evaluate products for free, based on their existing infrastructure:

> NVIDIA LaunchPad

The NVIDIA LaunchPad program provides worldwide enterprises and organizations with immediate, short-term access to the NVIDIA AI Enterprise software suite running on private accelerated computing infrastructure. It also includes a set of curated hands-on labs for AI practitioners and IT staff. IT administrators learn best practices for deploying NVIDIA AI Enterprise, and AI practitioners learn how to optimize training and inferencing workloads using AI tools and frameworks.

> Evaluation Software

NVIDIA AI Enterprise evaluation software is available for customers with existing NVIDIA-Certified Systems who are ready to start a proof-of-concept (POC) project for deployment at scale.

Learn more

To learn more about NVIDIA AI for Enterprise and NVIDIA-Certified Systems, visit:

[nvidia.com/en-us/data-center/products/ai-enterprise/](https://www.nvidia.com/en-us/data-center/products/ai-enterprise/)

[nvidia.com/certified-systems](https://www.nvidia.com/certified-systems)