Solution Overview

NVIDIA AI Enterprise

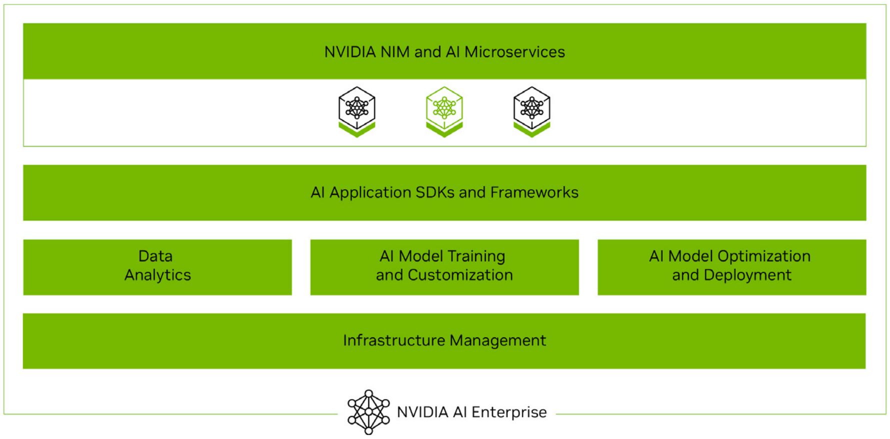
The software platform for production AI.

Deployment Challenges for Enterprise AI

Developing and maintaining an AI software stack can be intricate and demanding, presenting obstacles to managing AI workflows from initial prototype stages to full-scale production. With the presence of thousands of open-source packages and dependencies, security risk mitigation becomes crucial. Expertise is required to achieve optimal price/performance ratios. Additionally, total cost of ownership (TCO) can be adversely affected by infrastructure silos and the inability to run on the most cost-effective platform.

NVIDIA AI Enterprise

NVIDIA AI Enterprise is a cloud-native software platform that streamlines the development and deployment of production-grade AI solutions, including generative AI, computer vision, speech AI, and more.


NVIDIA NIM and AI Microservices
AI Application SDKs and Frameworks
Data Analytics
AI Model Training and Customization
AI Model Optimization and Deployment
Infrastructure Management
NVIDIA AI Enterprise

NVIDIA AI Enterprise Software Platform

Key Components:

- > **NVIDIA NIM™** is a set of easy-to-use microservices designed for secure, reliable deployment of high-performance AI model inferencing across the cloud, data center, and workstations.
- > **NVIDIA Blueprints** are reference AI workflows, pretrained for specific use cases, that can be customized by enterprise developers and the partner ecosystem.
- > **NVIDIA NeMo™** is an end-to-end framework for building, customizing, and deploying enterprise-grade generative AI models.
- > **Base Command Manager Essentials** streamlines infrastructure provisioning, workload management, and resource monitoring, across data center and cloud.

Easy-to-use microservices provide optimized model performance with enterprise-grade security, support, and stability. Additionally, the software platform includes:

- SDKs and frameworks that support AI applications across many domains.
- Tools and libraries to accelerate data analytics and AI model training and customization.
- Tools to manage AI clusters at scale—edge and data center, bare-metal and virtualized.

Enterprise-Grade Features

NVIDIA AI Enterprise addresses the complexities of building and maintaining a high-performance, secure, cloud-native AI software platform.

Security and software lifecycle management:

- Monthly patches for bug fixes and CVEs (common vulnerabilities and exposures).
- Production-grade software maintained for up to three years with guaranteed API stability.

Enterprise support:

- Global enterprise-grade support for production deployments.
- Service-level agreement (SLA) response times and timely resolution provided by NVIDIA experts.

End-to-end manageability:

- Management software for large-scale AI infrastructure.
- Cloud-native integrations that work with all major Kubernetes platforms.

Portability from pilot to production:

- Reduced risk associated with differences between environments.
- Ability to choose the most cost-effective platform for each use case.

Creating and Deploying Enterprise AI Solutions

NVIDIA AI Enterprise simplifies the process of deploying a variety of AI solutions. **NVIDIA Blueprints and AI workflows** are available as starting points for use cases to be created, customized, and deployed using NIM microservices and other GPU-accelerated libraries and frameworks.

Digital Humans for Customer Service

Multimodal PDF Data Extraction for Enterprise RAG

Route Optimization

Security Vulnerability Analysis

AI Chatbots

Deploy Anywhere

Run NVIDIA AI-enabled solutions across the cloud, in the data center, and on workstations for a true develop-once, deploy-anywhere experience.

NVIDIA-Certified Systems

NVIDIA AI Enterprise is supported on over 400 NVIDIA-Certified Systems™, available from a wide range of equipment manufacturers. Explore **NVIDIA-Certified Systems**.

Cloud

NVIDIA AI Enterprise enables organizations to efficiently and cost-effectively build and deploy public cloud or hybrid cloud AI solutions with **AWS**, **Azure**, **Google Cloud**, **Oracle Cloud**, and other NVIDIA cloud partners.

Container Orchestration

NVIDIA AI Enterprise includes support for container orchestration with VMware Tanzu, Red Hat OpenShift, HPE Ezmeral, Google Kubernetes Engine (GKE), Amazon Elastic Kubernetes Service (EKS), and upstream Kubernetes.

Ready to Get Started?

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

To sign up for a free 90-day evaluation license, visit:
nvidia.com/ai-enterprise-eval

To experience NVIDIA NIM microservices through the API catalog with a UI-based playground and access to free NVIDIA-managed API endpoints, visit https://build.nvidia.com/explore/discover/

To get hands-on experience with NVIDIA AI Enterprise, apply for a free lab through NVIDIA LaunchPad at: nvidia.com/try-ai Or contact Sales at: nvidia.com/ai-enterprise-sales