

New NVIDIA AI Enterprise Release Lights Up Data Centers

Software suite for enterprise AI adds production support for VMware vSphere with Tanzu, validates Domino Data Lab Enterprise MLOps software, and expands roster of NVIDIA-Certified Systems and channel partners to serve industries worldwide.

Author: John Fanelli

With a new year underway, NVIDIA is helping enterprises worldwide add modern workloads to their mainstream servers using the latest release of the NVIDIA AI Enterprise software suite.

NVIDIA AI Enterprise 1.1 is now generally available. Optimized, certified and supported by NVIDIA, the latest version of the software suite brings new updates including production support for containerized AI with the NVIDIA software on VMware vSphere with Tanzu, which was previously only available on a trial basis. Now, enterprises can run accelerated AI workloads on vSphere, running in both Kubernetes containers and virtual machines with NVIDIA AI Enterprise to support advanced AI development on mainstream IT infrastructure.

Among the top customer-requested features in NVIDIA AI Enterprise 1.1 is production support for running on VMware vSphere with Tanzu, which enables developers to run AI workloads on both containers and virtual machines within their vSphere environments. This new milestone in the AI-ready platform curated by NVIDIA and VMware provides an integrated, complete stack of containerized software and hardware optimized for AI, all fully managed by IT.

NVIDIA will soon add VMware vSphere with Tanzu support to the NVIDIA LaunchPad program for NVIDIA AI Enterprise, available at nine Equinix locations around the world. Qualified enterprises can test and prototype AI workloads at no charge through curated labs designed for the AI practitioner and IT admin. The labs showcase how to develop and manage common AI workloads like chatbots and recommendation systems, using NVIDIA AI Enterprise and VMware vSphere, and soon with Tanzu.

“Organizations are accelerating AI and ML development projects and VMware vSphere with Tanzu running NVIDIA AI Enterprise easily empowers AI development requirements with modern infrastructure services,” said Matt Morgan, vice president of Product Marketing, Cloud Infrastructure Business Group at VMware. “This announcement marks another key milestone for VMware and NVIDIA in our sustained efforts to help teams leverage AI across the enterprise.”

While enterprises are eager to use containerized development for AI, the complexity of these workloads requires orchestration across many layers of infrastructure. NVIDIA AI Enterprise 1.1 provides an ideal solution for these challenges as an AI-ready enterprise platform.

“AI is a very popular modern workload that is increasingly favoring deployment in containers. However, deploying AI capabilities at scale within the enterprise can be extremely complex, requiring enablement at multiple layers of the stack, from AI software frameworks, operating systems, containers, VMs, and down to the hardware,” said Gary Chen, research director, Software Defined Compute at IDC. “Turnkey, full-stack AI solutions can greatly simplify deployment and make AI more accessible within the enterprise.”

The 1.1 release of NVIDIA AI Enterprise also provides validation for the Domino Data Lab Enterprise MLOps Platform with VMware vSphere with Tanzu. This new integration enables more companies to cost-effectively scale data science by accelerating research, model development, and model deployment on mainstream accelerated servers.

“This new phase of our collaboration with NVIDIA further enables enterprises to solve the world’s most challenging problems by putting models at the heart of their businesses,” said Thomas Robinson, vice president of Strategic Partnerships at Domino Data Lab. “Together, we are providing every company the end-to-end platform to rapidly and cost-effectively deploy models enterprise-wide.”

Amidst the new release of NVIDIA AI Enterprise, the industry ecosystem is expanding with the first NVIDIA-Certified Systems from Cisco and Hitachi Vantara, as well as a growing roster of newly qualified system integrators offering solutions for the software suite.

The first Cisco system to be NVIDIA-Certified for NVIDIA AI Enterprise is the Cisco UCS C240 M6 rack server with NVIDIA A100 Tensor Core GPUs. The two-socket, 2RU form factor can power a wide range of storage and I/O-intensive applications, such as big data analytics, databases, collaboration, virtualization, consolidation and high-performance computing.

“At Cisco we are helping simplify customers’ hybrid cloud and cloud-native transformation. NVIDIA-Certified Cisco UCS servers, powered by Cisco Intersight, deliver the best-in-class AI workload experiences in the market,” said Siva Sivakumar, vice president of product management at Cisco. “The certification of the Cisco UCS C240 M6 rack server for NVIDIA AI Enterprise allows customers to add AI using the same infrastructure and management software deployed throughout their data center.”

The first NVIDIA-Certified System from Hitachi Vantara compatible with NVIDIA AI Enterprise is the Hitachi Advanced Server DS220 G2 with NVIDIA A100 Tensor Core GPUs. The general-purpose, dual-processor server is optimized for performance and capacity, and delivers a balance of compute and storage with the flexibility to power a wide range of solutions and applications.

“For many enterprises, cost is an important consideration when deploying new technologies like AI-powered quality control, recommender systems, chatbots and more,” said Dan McConnell, senior vice president, Product Management at Hitachi Vantara. “Accelerated with NVIDIA A100 GPUs and now certified for NVIDIA AI Enterprise, Hitachi Unified Compute Platform (UCP) solutions using the Hitachi Advanced Server DS220 G2 gives customers an ideal path for affordably integrating powerful AI-ready infrastructure to their data centers.”

A broad range of additional server manufacturers offer NVIDIA-Certified Systems for NVIDIA AI Enterprise. These include Atos, Dell Technologies, GIGABYTE, New H3C Information Technologies, Hewlett Packard Enterprise, Inspur Electronic Information, Lenovo and Supermicro, all of whose systems feature NVIDIA A100, NVIDIA A30 or other NVIDIA GPUs. Customers can also choose to deploy NVIDIA AI Enterprise on their own servers or on as-a-service bare metal infrastructure from Equinix Metal across nine regions globally.

AMAX, Colfax International, Exxact Corporation and Lambda are the newest system integrators qualified for NVIDIA AI Enterprise, joining a global ecosystem of channel partners that includes Axians, Carahsoft Technology Corp., Computacenter, Insight Enterprises, NTT, Presidio, Sirius, SoftServe, SVA System Vertrieb Alexander GmbH, TD SYNnex, Trace3 and World Wide Technology.

Enterprises interested in experiencing development with NVIDIA AI Enterprise can apply for instant access to no cost using curated labs via the NVIDIA LaunchPad program, which also features labs using NVIDIA Fleet Command for edge AI, as well as NVIDIA Base Command for demanding AI development workloads.

Original URL: <https://blogs.nvidia.com/blog/2022/01/19/ai-enterprise-release/>