SMARTER TECHNOLOGY FOR ALL

vmware® **NVIDIA**. Lenovo

AI-Ready Enterprise Platform

Unlocking AI for every enterprise with an end-to-end platform from Lenovo, NVIDIA and VMware

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. While most organizations know they need to invest in AI to secure their future, they struggle with finding the strategy and platform that can enable success.

Unlike traditional enterprise applications, AI apps are a relatively recent development for many IT departments. They're anchored in rapidly evolving, open-source, bleeding-edge code and lack proven approaches that meet the rigors of scaled production settings in enterprises. In fact, just 53 percent of projects make it from pilot to production, with the complexity of integrating AI solutions with existing infrastructure among the top three barriers to AI implementation.

Key challenges for enterprise AI workloads

- **Risk** – It's difficult to pull together an end-to-end AI solution from disparate products and integrate them with existing infrastructure.

- **Performance** – High performance is critical for AI, machine learning, and data analytics workloads. Fast time to deployment is also critical.

- **Scaling** – Going from proof of concept to enterprise deployments requires effective scaling—through efficient use of resources to ensure manageability and availability—and infrastructure cost management.

AI-Ready Platform

Lenovo VMware and NVIDIA have partnered to unlock the power of AI for every business by delivering an end-to-end enterprise platform optimized for AI workloads. This fully integrated validated and supported Lenovo AI solution delivers best-in-class AI software, the NVIDIA AI Enterprise Suite, optimized and exclusively certified for the industry's leading virtualization platform, VMware vSphere®. Running on Lenovo AI-Ready infrastructure. This platform accelerates the speed at which developers can build AI Infrastructure and high-performance data analytics, enables organizations to scale modern workloads on the same VMware vSphere infrastructure they've already invested in, and delivers enterprise-class manageability, security and availability.

Key Enabling Technologies Validated Architecture

The solution consists of the following building blocks.
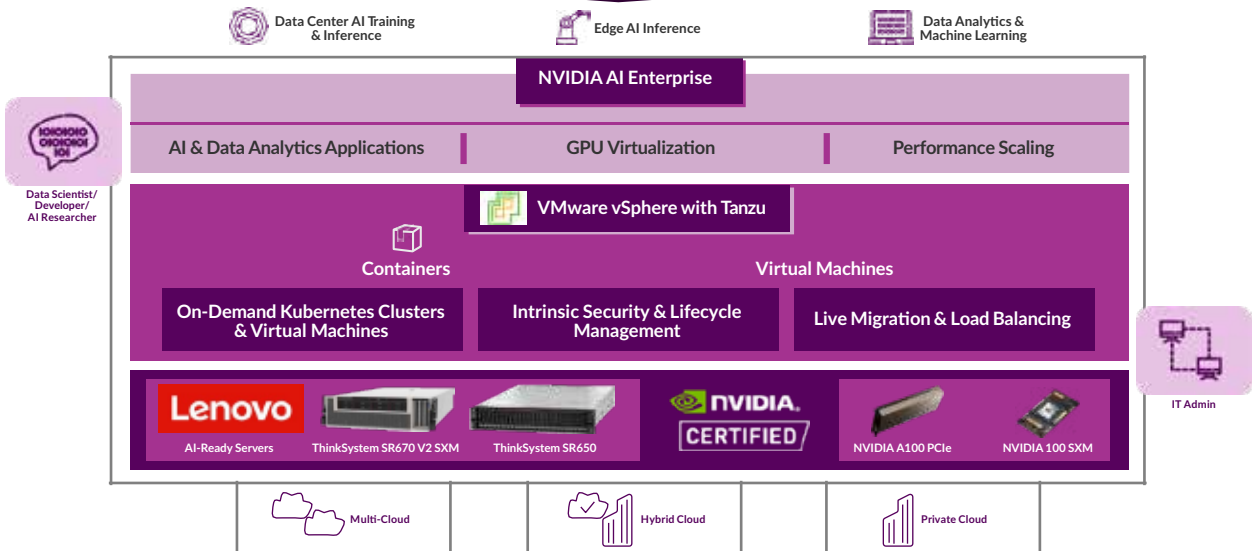
Lenovo ThinkSystem AI-Ready Infrastructure

The Lenovo AI-Ready Infrastructure supporting NVIDIA AI Enterprise with VMWare vSphere Enterprise Plus are optimum high performing, scalable servers designed for AI workloads with enterprise grade reliability. The Lenovo AI-Ready Infrastructure includes the ThinkSystem SR650 V2 and the ThinkSystem SR670 V2 providing a two-socket design with the latest Intel Xeon 3rd generation scalable processors in a a 2U and 3U footprint, respectively. The ThinkSystem SR650 V2 supports up to three NVIDIA A100 GPUs or up to eight NVIDIA T4 GPUs while the ThinkSystem SR670 V2 supports up to 8 NVIDIA A100 GPUs with NVLink Bridge or NVIDIA A100 HGX 4-GPU for the highest performance workloads. The ThinkSystem SR665 delivers the latest AMD EPYC™ processors in a 2U2S design supporting up three NVIDIA A100 GPUs or eight NVIDIA T4 GPUs. The Lenovo AI-Ready Infrastructure provides a solid foundation for:

- Transforming physical resources into services, using validated designs for hybrid cloud

- Performing analytics on streaming data, using validated designs for Big Data

- Increasing productivity of virtualized transactional systems, using validated designs for OLTP databases.

Fully integrated, validated and supported Lenovo-AI solution

Data Center AI Training & Inference

Edge AI Inference

Data Analytics & Machine Learning

NVIDIA AI Enterprise

Data Scientist/ Developer/ AI Researcher

| AI & Data Analytics Applications | GPU Virtualization | Performance Scaling |

VMware vSphere with Tanzu

Containers

Virtual Machines

| On-Demand Kubernetes Clusters & Virtual Machines | Intrinsic Security & Lifecycle Management | Live Migration & Load Balancing |

LENOVO — AI-Ready Servers — ThinkSystem SR670 V2 SXM — ThinkSystem SR650 — NVIDIA CERTIFIED — NVIDIA A100 PCIe — NVIDIA 100 SXM

IT Admin

Multi-Cloud   Hybrid Cloud   Private Cloud

VMware Tanzu support for NVIDIA AI Enterprise coming soon.

NVIDIA AI Enterprise Suite

NVIDIA AI Enterprise is an end-to-end, cloud native suite of AI and data science applications and frameworks optimized and exclusively certified by NVIDIA to run on VMware vSphere with NVIDIA-Certified Systems. It includes key enabling technologies and software from NVIDIA for rapid deployment, management and scaling of AI workloads in the modern hybrid cloud. NVIDIA AI Enterprise is licensed and supported by NVIDIA.

VMware vSphere with Tanzu

VMware vSphere® with VMware Tanzu™ is the industry's leading virtualization platform that runs over 70 million workloads across hundreds of thousands of organizations worldwide. VMware vSphere transforms bare-metal servers (including CPU- and GPU-based resources) into centrally managed AI and machine learning infrastructure pools that can quickly provision virtual machines and containers on demand.

VMware vSphere optimizes performance, increases availability, tightens security, streamlines maintenance, and reduces costs to create an agile, efficient, resilient and intrinsically secure infrastructure platform that supports existing workloads and next-gen applications such as AI.

VMware vSphere has been enhanced to optimally run the NVIDIA AI Enterprise software suite leveraging NVIDIA's latest NVIDIA Ampere architecture-based GPUs.

Benefits for Enterprises

- **Compute-intensive workloads** – ideal platform for hyperconverged infrastructure (HCI) or software-defined storage (SDS).

- **Comprehensive security** – Lenovo ThinkShield is an end-to-end approach to security that begins with development and continues through supply chain and the full life cycle of the device.

- **Easily manage** – Lenovo XClarity Administrator is a virtualized application that centrally manages ThinkSystem servers, storage, and networking. Via reusable patterns and policies, it ramps up and scales infrastructure provisioning and maintenance. It serves as a central integration point to extend your data center management processes to physical IT.

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 such as the NVIDIA A100 Tensor Core GPU. The Tensor Core technology included in the Ampere architecture has brought dramatic speedups to AI operations, bringing down training times from weeks to hours and providing massive acceleration to inference.

- NVIDIA® Mellanox® 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.