Lenovo ThinkSystem SR680a V3

Maximum Acceleration for Complex AI

Designed for Performance

The Lenovo ThinkSystem SR680a V3 GPU system delivers massive computational performance for Artificial Intelligence (AI), High-Performance Computing (HPC), and graphical and simulation workloads across various industries.

AI has arrived, and new solutions and possibilities for the most complex problems capture imaginations globally. Here are a few ways AI has captured the attention of the IT industry and the world at large:

- Generative AI platforms like Chat GPT, Bard, Claude & Microsoft Bing AI
- Chatbots for customer service interfaces
- Natural language processing (NLP) for call centers
- Computer vision for retail customer experience
- Fraud & threat detection in the financial industry
- Medical imaging and diagnostic applications

The possibilities seem endless as more solutions leverage the capabilities of AI and the demand for accelerated computing via GPUs increases. The ThinkSystem SR680a V3 GPU system delivers an optimized, enterprise-grade solution for deploying accelerated AI and HPC workloads in production.

Choice for Maximum Acceleration Flexibility

The ThinkSystem SR680a V3 offers a choice of acceleration platforms featuring NVIDIA H100/H200, and support for future GPUs. With high-speed interconnects between the GPUs, the system delivers unparalleled computational power for demanding AI and HPC workloads.

- Two 5th Gen Intel® Xeon® Scalable processors
- 8 GPUs with high-speed interconnect
    - NVIDIA H100/H200 SXM with NVLink interconnect
- Support for high-speed networking directly connected to the GPU complex
- 32x DDR5 slots, max frequency, up to 4TB
- Up to 16 high-speed 2.5" NVMe drives

The ThinkSystem SR680a V3 is built on two 5th Gen Intel® Xeon® Scalable processors, is designed to support both NVIDIA and AMD GPU platforms, and is designed to fit into an industry-standard 19" rack.

intel. **XEON** PLATINUM

Massive Computational Performance

NVIDIA H100/H200 Tensor Core GPU platforms deliver unprecedented acceleration — at every scale — to power the world's highest-performing elastic data centers for AI and HPC applications. The H100/H200 can efficiently scale up or be partitioned into seven isolated GPU instances. A second-generation Multi-Instance GPU (MIG) provides a unified platform that enables elastic data centers to dynamically adjust to shifting workload demands.

Lenovo Makes "AI for All" Possible

Lenovo delivers AI to data deployed inside your data center and at the edge. We have the industry's most comprehensive AI portfolio with over 80 products designed for AI workloads. Lenovo partners with over 50 AI Innovator partners globally to deliver turnkey Hybrid Cloud and Edge AI solutions.

Unique in the industry, Lenovo can deliver AI "from pocket to cloud" with servers, workstations, and AI-enabled devices. Whether you are a public cloud provider or building your AI model on-prem, Lenovo provides systems and solutions that scale to meet your needs. We bring AI to your data where and when you need it the most, in a truly hybrid approach: public, private, or personal.

Lenovo Professional Services

Offering a breadth of services, solutions and platforms, Lenovo's AI Professional Services help businesses of all sizes navigate the AI landscape, find the right solutions, and put AI to work for their organizations quickly, cost-effectively and at scale. Lenovo experts bring AI from concept to reality — from designing AI roadmaps to deploying platforms.

Specifications

Form Factor	8U rack
Processor	2x 5th Gen Intel® Xeon® Scalable processors, up to 350W
Memory	Up to 4TB using 32x DDR5 DIMMs with maximum frequency at 5600MHz
GPU	Supports 8x high-performance GPUs: • 8x NVIDIA HGX™ H100/H200 GPUs with NVLink interconnects at 900 GB/s
I/O Expansion	Up to 10x PCIe Gen5 x16 FHHL adapters 8x in front connected to PCIe switch for GPU connectivity 2x in rear connected to CPU for CPU connectivity
Storage	Up to 16x 2.5" hot-swap NVMe SSDs Up to 2x M.2 for boot (RAID via VROC)
Power	Up to 8x hot-swap power supplies, allowing full N+N redundancy
Cooling	Air cooled with N+1 hot-swap fan solution
Management	XClarity Controller2 (XCC2), which provides advanced service-processor control, monitoring, and alerting functions. The XCC2 consolidates the service processor functionality, super I/O, video controller, and remote presence capabilities into a single chip on the server system board.
OS Support	RHEL, Ubuntu, ESXi

About Lenovo

Lenovo (HKSE: 992) (ADR: LNVGY) is a US$62 billion revenue global technology powerhouse, ranked #171 in the Fortune Global 500, employing 77,000 people around the world, and serving millions of customers every day in 180 markets. Focused on a bold vision to deliver smarter technology for all, Lenovo is expanding into new growth areas of infrastructure, mobile, solutions and services. This transformation is building a more inclusive, trustworthy, and sustainable digital society for everyone, everywhere.

For More Information

To learn more about the Lenovo ThinkSystem SR680a V3, contact your Lenovo representative or Business Partner or visit lenovo.com/thinksystem. For detailed specifications consult the SR680a V3 Product Guide.

NEED STORAGE?	Learn more about Lenovo Storage lenovo.com/systems/storage
NEED SERVICES?	Learn more about Lenovo Services lenovo.com/systems/services
GET IT AS A SERVICE?	Learn more about Lenovo TruScale lenovo.com/truscale