ThinkSystem NVIDIA H200 141GB GPUs
Product Guide

The NVIDIA H200 Tensor Core GPU supercharges generative AI and high-performance computing (HPC) workloads with game-changing performance and memory capabilities. H200 is the newest addition to NVIDIA's leading AI and high-performance data center GPU portfolio, bringing massive compute to data centers.

The NVIDIA H200 141GB 700W GPU is offered in either the SXM5 form factor or as a PCIe double-wide GPU adapter. Four or eight SXM5 GPU modules are implemented with a fully-connected NVLink topology in supported ThinkSystem servers. SXM5 GPUs are either air-cooled or water-cooled, depending on the server. PCIe double-wide GPUs are air-cooled, and can be implemented using 2-way or 4-way NVLink bridges.

Leveraging the power of H200 multi-precision Tensor Cores, an eight-way HGX H200 provides over 32 petaFLOPS of FP8 deep learning compute and over 1.1TB of aggregate HBM memory for the highest performance in generative AI and HPC applications.

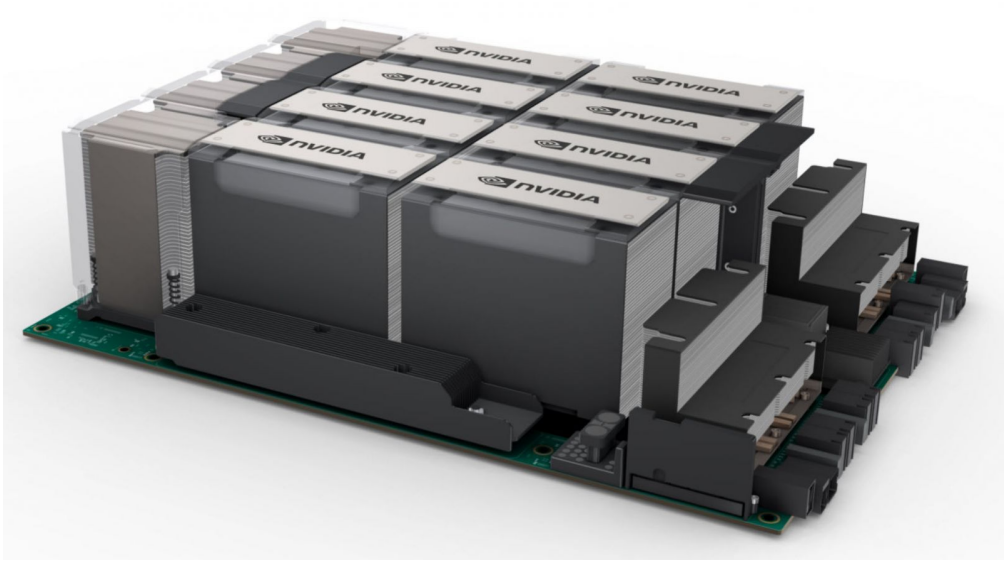

Figure 1. ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Board in the ThinkSystem SR680a V3 server

Did you know?

To maximize compute performance, H200 is the world's first GPU with HBM3e memory with 4.8TB/s of memory bandwidth, a 1.4X increase over H100. H200 also expands GPU memory capacity nearly doubled to 141GB. The combination of faster and larger HBM memory accelerates performance of computationally intensive generative AI and HPC applications, while meeting the evolving demands of growing model sizes.

Part number information

The following table shows the part numbers for the 8-GPU and 4-GPU boards. The feature codes contain all H200 GPUs in the SXM form factor plus the NVLink high-speed interconnections between the GPUs.

The table also indicates which GPUs include a 5-year subscription to NVIDIA AI Enterprise Software (NVAIE).

Table 1. Ordering information

Part number	Feature code	Description	Includes NVAIE*	NVIDIA part number	Controlled GPU status
HGX form factor GPUs					
CTO only	C2ER	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Liquid Cooled Board	No	935-24287-2741-000	Controlled
CTO only	C1HM	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Board	No	935-24287-2740-000	Controlled
CTO only	C3V2	ThinkSystem NVIDIA HGX H200 141GB 700W 4-GPU Board	No	935-23087-2741-000	Controlled
Double-wide PCIe adapter form factor					
4X67A97315	C3V3	ThinkSystem NVIDIA H200 NVL 141GB PCIe GPU Gen5 Passive GPU	Included*	900-21010-2740-030	Controlled
4X67A97320	C3V1	ThinkSystem NVIDIA 2-way bridge for H200 NVL	-	900-23945-2700-000	-
4X67A97322	C3V0	ThinkSystem NVIDIA 4-way bridge for H200 NVL	-	900-23946-2700-000	-

* ThinkSystem NVIDIA H200 NVL 141GB PCIe GPU Gen5 Passive GPU, C3V3 includes a 5-year subscription to NVIDIA AI Enterprise Software (NVAIE). See the NVIDIA AI Enterprise Software section.

The NVIDIA H200 GPU is Controlled which means the GPU is not offered in certain markets, as determined by the US Government.

Features

The NVIDIA H200 Tensor Core GPU supercharges generative AI and HPC with game-changing performance and memory capabilities. As the first GPU with HBM3e, H200's faster, larger memory fuels the acceleration of generative AI and LLMs while advancing scientific computing for HPC workloads.

NVIDIA HGX™ H200, the world's leading AI computing platform, features the H200 GPU for the fastest performance. An eight-way HGX H200 provides over 32 petaflops of FP8 deep learning compute and 1.1TB of aggregate high-bandwidth memory for the highest performance in generative AI and HPC applications.

Key AI and HPC workload features:

- Unlock Insights With High-Performance LLM Inference
  In the ever-evolving landscape of AI, businesses rely on large language models to address a diverse range of inference needs. An AI inference accelerator must deliver the highest throughput at the lowest TCO when deployed at scale for a massive user base. H200 doubles inference performance compared to H100 when handling LLMs such as Llama2 70B.
- Optimize Generative AI Fine-Tuning Performance
  Large language models can be customized to specific business case needs with fine-tuning, low-rank adaptation (LoRA), or retrieval-augmented generation (RAG) methods. These methods bridge the gap between general pretrained results and task-specific solutions, making them essential tools for industry and research applications.

NVIDIA H200's Transformer Engine and fourth-generation Tensor Cores speed up fine-tuning by 5.5X over A100 GPUs. This performance increase allows enterprises and AI practitioners to quickly optimize and deploy generative AI to benefit their business. Compared to fully training foundation models from scratch, fine-tuning offers better energy efficiency and the fastest access to customized solutions needed to grow business.

- Industry-Leading Generative AI Training
  The era of generative AI has arrived, and it requires billion-parameter models to take on the paradigm shift in business operations and customer experiences.

  NVIDIA H200 GPUs feature the Transformer Engine with FP8 precision, which provides up to 5X faster training over A100 GPUs for large language models such as GPT-3 175B. The combination of fourth-generation NVLink, which offers 900GB/s of GPU-to-GPU interconnect, PCIe Gen5, and NVIDIA Magnum IO™ software, delivers efficient scalability from small enterprise to massive unified computing clusters of GPUs. These infrastructure advances, working in tandem with the NVIDIA AI Enterprise software suite, make the NVIDIA H200 the most powerful end-to-end generative AI and HPC data center platform.

- Supercharged High-Performance Computing
  Memory bandwidth is crucial for high-performance computing applications, as it enables faster data transfer and reduces complex processing bottlenecks. For memory-intensive HPC applications like simulations, scientific research, and artificial intelligence, H200's higher memory bandwidth ensures that data can be accessed and manipulated efficiently, leading to up to a 110X faster time to results.

  The NVIDIA data center platform consistently delivers performance gains beyond Moore's Law. And H200's breakthrough AI capabilities further amplify the power of HPC+AI to accelerate time to discovery for scientists and researchers working on solving the world's most important challenges.

- Reduced Energy and TCO
  In a world where energy conservation and sustainability are top of mind, the concerns of business leaders and enterprises have evolved. Enter accelerated computing, a leader in energy efficiency and TCO, particularly for workloads that thrive on acceleration, such as HPC and generative AI.

  With the introduction of H200, energy efficiency and TCO reach new levels. This cutting-edge technology offers unparalleled performance, all within the same power profile as H100. AI factories and at-scale supercomputing systems that are not only faster but also more eco-friendly deliver an economic edge that propels the AI and scientific community forward. For at-scale deployments, H200 systems provide 5X more energy savings and 4X better cost of ownership savings over the NVIDIA Ampere architecture generation.

Key features of the Hopper architecture:

- NVIDIA H200 Tensor Core GPU

  H200 is the world's most advanced chip ever built. It features major advances to accelerate AI, HPC, memory bandwidth, interconnect, and communication at data center scale.

- Transformer Engine

  The Transformer Engine uses software and Hopper Tensor Core technology designed to accelerate training for models built from the world's most important AI model building block, the transformer. Hopper Tensor Cores can apply mixed FP8 and FP16 precisions to dramatically accelerate AI calculations for transformers.

- NVLink Switch System

  The NVLink Switch System enables the scaling of multi-GPU input/output (IO) across multiple servers. The system delivers up to 9X higher bandwidth than InfiniBand HDR on the NVIDIA Ampere architecture.

- NVIDIA Confidential Computing

  NVIDIA Confidential Computing is a built-in security feature of H100 and H200 GPUs where users can protect the confidentiality and integrity of their data and applications in use while accessing the GPUs.

- Second-Generation Multi-Instance GPU (MIG)

  The Hopper architecture's second-generation MIG supports multi-tenant, multi-user configurations in virtualized environments, securely partitioning the GPU into isolated, right-size instances to maximize quality of service (QoS) for 7X more secured tenants.

- DPX Instructions

  Hopper's DPX instructions accelerate dynamic programming algorithms by 40X compared to CPUs and 7X compared to NVIDIA Ampere architecture GPUs. This leads to dramatically faster times in disease diagnosis, real-time routing optimizations, and graph analytics.

Technical specifications

The following table lists the NVIDIA H200 GPU specifications.

Table 2. GPU specifications

Specification	**NVIDIA H200 SXM**	**NVIDIA H200 NVL PCIe**
Form Factor	SXM5	DW PCIe
FP64	34 teraFLOPS	30 teraFLOPS
FP64 Tensor Core	67 teraFLOPS	60 teraFLOPS
FP32	67 teraFLOPS	60 teraFLOPS
TF32 Tensor Core	495 / 989 teraFLOPS*	418 / 835 teraFLOPS*
BFLOAT16 Tensor	990 / 1,979 teraFLOPS*	836 / 1,671 teraFLOPS*
FP16 Tensor Core	990 / 1,979 teraFLOPS*	836 / 1,671 teraFLOPS*
FP8 Tensor Core	1,979 / 3,958 teraFLOPS*	1,570 / 3,341 teraFLOPS*
INT8 Tensor Core	1,979 / 3,958 TOPS*	1,570 / 3,341 TOPS*
GPU Memory	141 GB HBM3e	141 GB HBM3e
GPU Memory Bandwidth	4.8 TB/s	4.8 TB/s
Total Graphics Power (TGP) or Continuous Electrical Design Point (EDPc)	700W	600W
Multi-Instance GPUs	Up to 7 MIGS @ 18 GB	Up to 7 MIGS @ 16.5 GB
Interconnect	NVLink: 900 GB/s PCIe Gen5: 128 GB/s	NVLink: 900 GB/s PCIe Gen5: 128 GB/s

* Without / with structural sparsity enabled

Server support

The following tables list the ThinkSystem servers that are compatible.

Table 3. Server support (Part 1 of 5)

Part Number	Description	AMD V3				2S Intel V3/V4						Multi Node V3			1S V3			
		SR635 V3 (7D9H / 7D9G)	SR655 V3 (7D9F / 7D9E)	SR645 V3 (7D9D / 7D9C)	SR665 V3 (7D9B / 7D9A)	ST650 V3 (7D7B / 7D7A)	SR630 V3 (7D72 / 7D73)	SR650 V3 (7D75 / 7D76)	SR630 V4 (7DG8 / 7DG9)	SR650 V4 (7DGC / 7DGD)	SR650a V4 (7DGC / 7DGD)	SD535 V3 (7DD8 / 7DD1)	SD530 V3 (7DDA / 7DD3)	SD550 V3 (7DD9 / 7DD2)	ST45 V3 (7DH4 / 7DH5)	ST50 V3 (7DF4 / 7DF3)	ST250 V3 (7DCF / 7DCE)	SR250 V3 (7DCM / 7DCL)
C2ER	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Liquid Cooled Board	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
C1HM	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Board	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
C3V2	ThinkSystem NVIDIA HGX H200 141GB 700W 4-GPU Board	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
4X67A97315	ThinkSystem NVIDIA H200 NVL 141GB PCIe GPU Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	2	N	N	N	N	N	N	N

Table 4. Server support (Part 2 of 5)

Part Number	Description	4S 8S Intel V3/V4					GPU Rich						Edge					
		SR850 V3 (7D97 / 7D96)	SR860 V3 (7D94 / 7D93)	SR950 V3 (7DC5 / 7DC4)	SR850 V4 (7DJT / 7DJS)	SR860 V4 (7DJQ / 7DJN)	SR670 V2 (7Z22 / 7Z23)	SR675 V3 (7D9Q / 7D9R)	SR680a V3 (7DHE)	SR680a V3 B200 (7DM9)	SR685a V3 (7DHC)	SR780a V3 (7DJ5)	SE100 (7DGR)	SE350 (7Z46 / 7D1X)	SE350 V2 (7DA9)	SE360 V2 (7DAM)	SE450 (7D8T)	SE455 V3 (7DBY)
C2ER	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Liquid Cooled Board	N	N	N	N	N	N	N	N	N	N	1	N	N	N	N	N	N
C1HM	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Board	N	N	N	N	N	N	N	1[1]	N	1[1]	N	N	N	N	N	N	N
C3V2	ThinkSystem NVIDIA HGX H200 141GB 700W 4-GPU Board	N	N	N	N	N	N	1	N	N	N	N	N	N	N	N	N	N
4X67A97315	ThinkSystem NVIDIA H200 NVL 141GB PCIe GPU Gen5 Passive GPU	N	N	N	N	N	N	8	N	N	N	N	N	N	N	N	N	N

1. Contains 8 separate GPUs connected via high-speed interconnects

Table 5. Server support (Part 3 of 5)

Part Number	Description	Super Computing							1S Intel V2			2S Intel V2			AMD V1			
		SC750 V4 (7DDJ)	SC777 V4 (7DKA)	SD665 V3 (7D9P)	SD665-N V3 (7DAZ)	SD650 V3 (7D7M)	SD650-I V3 (7D7L)	SD650-N V3 (7D7N)	ST50 V2 (7D8K / 7D8J)	ST250 V2 (7D8G / 7D8F)	SR250 V2 (7D7R / 7D7Q)	ST650 V2 (7Z75 / 7Z74)	SR630 V2 (7Z70 / 7Z71)	SR650 V2 (7Z72 / 7Z73)	SR635 (7Y98 / 7Y99)	SR655 (7Y00 / 7Z01)	SR645 (7D2Y / 7D2X)	SR665 (7D2W / 7D2V)
C2ER	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Liquid Cooled Board	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
C1HM	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Board	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
C3V2	ThinkSystem NVIDIA HGX H200 141GB 700W 4-GPU Board	N	N	N	1	N	N	1	N	N	N	N	N	N	N	N	N	N
4X67A97315	ThinkSystem NVIDIA H200 NVL 141GB PCIe GPU Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Table 6. Server support (Part 4 of 5)

Part Number	Description	Dense V2				4S V2		8S	4S V1			1S Intel V1			
		SD630 V2 (7D1K)	SD650 V2 (7D1M)	SD650-N V2 (7D1N)	SN550 V2 (7Z69)	SR850 V2 (7D31 / 7D32)	SR860 V2 (7Z59 / 7Z60)	SR950 (7X11 / 7X12)	SR850 (7X18 / 7X19)	SR850P (7D2F / 2D2G)	SR860 (7X69 / 7X70)	ST50 (7Y48 / 7Y50)	ST250 (7Y45 / 7Y46)	SR150 (7Y54)	SR250 (7Y52 / 7Y51)
C2ER	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Liquid Cooled Board	N	N	N	N	N	N	N	N	N	N	N	N	N	N
C1HM	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Board	N	N	N	N	N	N	N	N	N	N	N	N	N	N
C3V2	ThinkSystem NVIDIA HGX H200 141GB 700W 4-GPU Board	N	N	N	N	N	N	N	N	N	N	N	N	N	N
4X67A97315	ThinkSystem NVIDIA H200 NVL 141GB PCIe GPU Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Table 7. Server support (Part 5 of 5)

Part Number	Description	2S Intel V1								Dense V1			
		ST550 (7X09 / 7X10)	SR530 (7X07 / 7X08)	SR550 (7X03 / 7X04)	SR570 (7Y02 / 7Y03)	SR590 (7X98 / 7X99)	SR630 (7X01 / 7X02)	SR650 (7X05 / 7X06)	SR670 (7Y36 / 7Y37)	SD530 (7X21)	SD650 (7X58)	SN550 (7X16)	SN850 (7X15)
C2ER	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Liquid Cooled Board	N	N	N	N	N	N	N	N	N	N	N	N
C1HM	ThinkSystem NVIDIA HGX H200 141GB 700W 8-GPU Board	N	N	N	N	N	N	N	N	N	N	N	N
C3V2	ThinkSystem NVIDIA HGX H200 141GB 700W 4-GPU Board	N	N	N	N	N	N	N	N	N	N	N	N
4X67A97315	ThinkSystem NVIDIA H200 NVL 141GB PCIe GPU Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N

Operating system support

For SXM GPUs, operating system support is based on that of the supported servers. See the SR680a V3 server product guide for details: https://lenovopress.lenovo.com/lp1909-thinksystem-sr680a-v3-server

The following tables list the OS support for the PCIe adapter.

Table 8. Operating system support for ThinkSystem NVIDIA H200 NVL 141GB PCIe Gen5 Passive GPU, 4X67A97315

Operating systems	SR675 V3
Microsoft Windows Server 2022	Y
Microsoft Windows Server 2025	Y
Red Hat Enterprise Linux 8.10	Y
Red Hat Enterprise Linux 9.4	Y
Red Hat Enterprise Linux 9.5	Y
SUSE Linux Enterprise Server 15 SP6	Y
Ubuntu 24.04 LTS	Y

NVIDIA GPU software

This section lists the NVIDIA software that is available from Lenovo.

- NVIDIA AI Enterprise Software
- NVIDIA HPC Compiler Software

As listed in Table 1 in the Part number information section, the PCIe adapter version of the H200 includes a 5-year subscription to NVIDIA AI Enterprise Software (NVAIE).

NVIDIA AI Enterprise Software

Lenovo offers the NVIDIA AI Enterprise (NVAIE) cloud-native enterprise software. NVIDIA AI Enterprise is an end-to-end, cloud-native suite of AI and data analytics software, optimized, certified, and supported by NVIDIA to run on VMware vSphere and bare-metal with NVIDIA-Certified Systems™. It includes key enabling technologies from NVIDIA for rapid deployment, management, and scaling of AI workloads in the modern hybrid cloud.

NVIDIA AI Enterprise is licensed on a per-GPU basis. NVIDIA AI Enterprise products can be purchased as either a perpetual license with support services, or as an annual or multi-year subscription.

- The perpetual license provides the right to use the NVIDIA AI Enterprise software indefinitely, with no expiration. NVIDIA AI Enterprise with perpetual licenses must be purchased in conjunction with one-year, three-year, or five-year support services. A one-year support service is also available for renewals.
- The subscription offerings are an affordable option to allow IT departments to better manage the flexibility of license volumes. NVIDIA AI Enterprise software products with subscription includes support services for the duration of the software's subscription license

The features of NVIDIA AI Enterprise Software are listed in the following table.

Table 9. Features of NVIDIA AI Enterprise Software (NVAIE)

Features	**Supported in NVIDIA AI Enterprise**
Per GPU Licensing	Yes
Compute Virtualization	Supported
Windows Guest OS Support	No support
Linux Guest OS Support	Supported
Maximum Displays	1
Maximum Resolution	4096 x 2160 (4K)
OpenGL and Vulkan	In-situ Graphics only
CUDA and OpenCL Support	Supported
ECC and Page Retirement	Supported
MIG GPU Support	Supported
Multi-vGPU	Supported
NVIDIA GPUDirect	Supported
Peer-to-Peer over NVLink	Supported
GPU Pass Through Support	Supported
Baremetal Support	Supported
AI and Data Science applications and Frameworks	Supported
Cloud Native ready	Supported

Note: Maximum 10 concurrent VMs per product license

The following table lists the ordering part numbers and feature codes.

Table 10. NVIDIA AI Enterprise Software (NVAIE)

Part number	**Feature code 7S02CTO1WW**	**NVIDIA part number**	**Description**
AI Enterprise Perpetual License			

Part number	Feature code 7S02CTO1WW	NVIDIA part number	Description
7S02001BWW	S6YY	731-AI7004+P3CMI60	NVIDIA AI Enterprise Perpetual License and Support per GPU Socket, 5 Years
7S02001EWW	S6Z1	731-AI7004+P3EDI60	NVIDIA AI Enterprise Perpetual License and Support per GPU Socket, EDU, 5 Years
AI Enterprise Subscription License			
7S02001FWW	S6Z2	731-AI7003+P3CMI12	NVIDIA AI Enterprise Subscription License and Support per GPU Socket, 1 Year
7S02001GWW	S6Z3	731-AI7003+P3CMI36	NVIDIA AI Enterprise Subscription License and Support per GPU Socket, 3 Years
7S02001HWW	S6Z4	731-AI7003+P3CMI60	NVIDIA AI Enterprise Subscription License and Support per GPU Socket, 5 Years
7S02001JWW	S6Z5	731-AI7003+P3EDI12	NVIDIA AI Enterprise Subscription License and Support per GPU Socket, EDU, 1 Year
7S02001KWW	S6Z6	731-AI7003+P3EDI36	NVIDIA AI Enterprise Subscription License and Support per GPU Socket, EDU, 3 Years
7S02001LWW	S6Z7	731-AI7003+P3EDI60	NVIDIA AI Enterprise Subscription License and Support per GPU Socket, EDU, 5 Years

Find more information in the NVIDIA AI Enterprise Sizing Guide.

NVIDIA HPC Compiler Software

Table 11. NVIDIA HPC Compiler

Part number	Feature code 7S09CTO6WW	Description
HPC Compiler Support Services		
7S090014WW	S924	NVIDIA HPC Compiler Support Services, 1 Year
7S090015WW	S925	NVIDIA HPC Compiler Support Services, 3 Years
7S09002GWW	S9UQ	NVIDIA HPC Compiler Support Services, 5 Years
7S090016WW	S926	NVIDIA HPC Compiler Support Services, EDU, 1 Year
7S090017WW	S927	NVIDIA HPC Compiler Support Services, EDU, 3 Years
7S09002HWW	S9UR	NVIDIA HPC Compiler Support Services, EDU, 5 Years
7S090018WW	S928	NVIDIA HPC Compiler Support Services - Additional Contact, 1 Year
7S09002JWW	S9US	NVIDIA HPC Compiler Support Services - Additional Contact, 3 Years
7S09002KWW	S9UT	NVIDIA HPC Compiler Support Services - Additional Contact, 5 Years
7S090019WW	S929	NVIDIA HPC Compiler Support Services - Additional Contact, EDU, 1 Year
7S09002LWW	S9UU	NVIDIA HPC Compiler Support Services - Additional Contact, EDU, 3 Years
7S09002MWW	S9UV	NVIDIA HPC Compiler Support Services - Additional Contact, EDU, 5 Years
HPC Compiler Premier Support Services		
7S09001AWW	S92A	NVIDIA HPC Compiler Premier Support Services, 1 Year
7S09002NWW	S9UW	NVIDIA HPC Compiler Premier Support Services, 3 Years
7S09002PWW	S9UX	NVIDIA HPC Compiler Premier Support Services, 5 Years
7S09001BWW	S92B	NVIDIA HPC Compiler Premier Support Services, EDU, 1 Year
7S09002QWW	S9UY	NVIDIA HPC Compiler Premier Support Services, EDU, 3 Years

Part number	Feature code 7S09CTO6WW	Description
7S09002RWW	S9UZ	NVIDIA HPC Compiler Premier Support Services, EDU, 5 Years
7S09001CWW	S92C	NVIDIA HPC Compiler Premier Support Services - Additional Contact, 1 Year
7S09002SWW	S9V0	NVIDIA HPC Compiler Premier Support Services - Additional Contact, 3 Years
7S09002TWW	S9V1	NVIDIA HPC Compiler Premier Support Services - Additional Contact, 5 Years
7S09001DWW	S92D	NVIDIA HPC Compiler Premier Support Services - Additional Contact, EDU, 1 Year
7S09002UWW	S9V2	NVIDIA HPC Compiler Premier Support Services - Additional Contact, EDU, 3 Years
7S09002VWW	S9V3	NVIDIA HPC Compiler Premier Support Services - Additional Contact, EDU, 5 Years

Regulatory approvals

The NVIDIA H200 GPU has the following regulatory approvals:

- RCM
- BSMI
- CE
- FCC
- ICES
- KCC
- cUL, UL
- VCCI

Warranty

The NVIDIA H200 GPU assumes the server's base warranty and any warranty upgrades.

Seller training courses

The following sales training courses are offered for employees and partners (login required). Courses are listed in date order.

1. **Think AI Weekly: ISG & SSG Better Together: Uniting AI Solutions & Services for Smarter Outcomes**
   2025-08-01 | 55 minutes | Employees Only

   View this session to hear from our speakers Allen Holmes, AI Technologist, ISG and Balaji Subramaniam, AI Regional Leader-Americas, SSG.

   Topics include:

   • An overview of ISG & SSG AI CoE Offerings with Customer Case Studies
   • The Enterprise AI Deal Engagement Flow with ISG and SSG
   • How sellers can leverage this partnership to differentiate with Enterprise clients.
   • NEW COURSE: From Inception to Execution: Evolution of an AI Deal

   Tags: Artificial Intelligence (AI), Sales, Services, Technology Solutions, TruScale Infrastructure as a Service

   Published: 2025-08-01
   Length: 55 minutes

   Start the training:
   Employee link: Grow@Lenovo

   Course code: DTAIW145
2. **Think AI Weekly: Third-Party Due Diligence Requirements for GPU Opportunities**
   2025-07-24 | 46 minutes | Employees Only

   View this session to hear from Tanya Roychowdhury, Legal Counsel Director and Andrea Fazio, Third-party Due Diligence Project Manager as the explain:

   - What are the requirements?
   - Why are they important?
   - What this means to sales

   Tags: Artificial Intelligence (AI), DataCenter Products, NVIDIA, Sales, Technical Sales

   Published: 2025-07-24
   Length: 46 minutes

   Start the training:
   Employee link: Grow@Lenovo

   Course code: DTAIW143

3. **ThinkSystem Supercomputing Servers Powered by NVIDIA**
   2025-06-27 | 30 minutes | Employees and Partners

   This course offers you information about the Lenovo SC777 V4 Neptune server, the first Lenovo server to use an Arm processor from NVIDIA. By the end of this course, you'll be able to list three features of the ThinkSystem SC777 V4 Neptune server, list three features of the ThinkSystem N1380 Neptune enclosure, describe two customer benefits of the ThinkSystem SC777 V4 Neptune server, and list four workload environments to which the SC777 V4 server is well suited.

   Tags: DataCenter Products, NVIDIA, ThinkSystem

   Published: 2025-06-27
   Length: 30 minutes

   Start the training:
   Employee link: Grow@Lenovo
   Partner link: Lenovo 360 Learning Center

   Course code: SXXW2545

4. **VTT AI: NVIDIA and Lenovo: Data Center Platform Overview**
   2025-06-10 | 77 minutes | Employees Only

   Please join this session to hear Steve Stein, Senior Product Marketing Manager, NVIDIA and Naman Malhotra, Senior Product Manager, Lenovo as they present these topics:

   •NVIDIA Accelerated Computing Portfolio
   •Use Cases and Positioning
   •Lenovo Platforms and Solutions

   Tags: Artificial Intelligence (AI), Nvidia, Server

   Published: 2025-06-10
   Length: 77 minutes

   Start the training:
   Employee link: Grow@Lenovo

   Course code: DVAI216

5. **VTT AI: Introducing the Lenovo Hybrid AI 285 Platform April 2025**
   2025-04-30 | 60 minutes | Employees Only

   The Lenovo Hybrid AI 285 Platform enables enterprises of all sizes to quickly deploy AI infrastructures supporting use cases as either new greenfield environments or as an extension to current infrastructures. The 285 Platform enables the use of the NVIDIA AI Enterprise software stack. The AI Hybrid 285 platform is the perfect foundation supporting Lenovo Validated Designs.
   • Technical overview of the Hybrid AI 285 platform
   • AI Hybrid platforms as infrastructure frameworks for LVDs addressing data center-based AI solutions.
   • Accelerate AI adoption and reduce deployment risks

   Tags: Artificial Intelligence (AI), Nvidia, Technical Sales, Lenovo Hybrid AI 285

   Published: 2025-04-30
   Length: 60 minutes

   Start the training:
   Employee link: Grow@Lenovo

   Course code: DVAI215

6. **Lenovo Cloud Architecture VTT: Supercharge Your Enterprise AI with NVIDIA AI Enterprise on Lenovo Hybrid AI Platform**
   2025-04-17 | 75 minutes | Employees and Partners

   Join us for an in-depth webinar with Justin King, Principal Product Marketing Manager for Enterprise AI exploring the power of NVIDIA AI Enterprise, delivering Generative and Agentic AI outcomes deployed with Lenovo Hybrid AI platform environments.
   In today's data-driven landscape, AI is evolving at high speed, with new techniques delivering more accurate responses. Enterprises are seeking not just an understanding but also how they can achieve AI-driven business outcomes.
   With this, the demand for secure, scalable, and high-performing AI operations-and the skills to deliver them-is top of mind for many. Learn how NVIDIA AI Enterprise, a comprehensive software suite optimized for NVIDIA GPUs, provides the tools and frameworks, including NVIDIA NIM, NeMo, and Blueprints, to accelerate AI development and deployment while reducing risk-all within the control and security of your Lenovo customer's hybrid AI environment.

   Tags: Artificial Intelligence (AI), Cloud, Data Management, Nvidia, Technical Sales

   Published: 2025-04-17
   Length: 75 minutes

   Start the training:
   Employee link: Grow@Lenovo
   Partner link: Lenovo 360 Learning Center

   Course code: DVCLD221
7. **AI VTT: GTC Update and The Lenovo LLM Sizing Guide**
   2025-03-12 | 86 minutes | Employees Only

   Please view this session that is two parts. Part one is Robert Daigle, Director, Global AI Solutions and Hande Sahin-Bahceci, AI Solutions Marketing Leader explaining the upcoming announcements for NVIDIA GTC. Part Two is Sachin Wani, AI Data Scientist explaining the Lenovo LLM Sizing Guide with these topics:

   • Minimum GPU requirements for fine-tuning/training and inference
   • Gathering requirements for the customer's use case
   • LLMs from a technical perspective

   Tags: Artificial Intelligence (AI), Technical Sales

   Published: 2025-03-12
   Length: 86 minutes

   Start the training:
   Employee link: Grow@Lenovo

   Course code: DVAI214

8. **Partner Technical Webinar - NVIDIA Portfolio**
   2024-11-06 | 60 minutes | Employees and Partners

   In this 60-minute replay, Jason Knudsen of NVIDIA presented the NVIDIA Computing Platform. Jason talked about the full portfolio from GPUs to Networking to AI Enterprise and NIMs.

   Tags: Artificial Intelligence (AI), Nvidia

   Published: 2024-11-06
   Length: 60 minutes

   Start the training:
   Employee link: Grow@Lenovo
   Partner link: Lenovo 360 Learning Center

   Course code: 110124

9. **Q2 Solutions Launch TruScale GPU Next Generation Management in the AI Era Quick Hit**
   2024-09-10 | 6 minutes | Employees and Partners

   This Quick Hit focuses on Lenovo announcing additional ways to help you build, scale, and evolve your customer's private AI faster for improved ROI with TruScale GPU as a Service, AI-driven systems management, and infrastructure transformation services.

   Tags: Artificial Intelligence (AI), Services, TruScale

   Published: 2024-09-10
   Length: 6 minutes

   Start the training:
   Employee link: Grow@Lenovo
   Partner link: Lenovo 360 Learning Center

   Course code: SXXW2543a

10. **VTT AI: The NetApp AIPod with Lenovo for NVIDIA OVX**
2024-08-13 | 38 minutes | Employees and Partners

AI, for some organizations, is out of reach, due to cost, integration complexity, and time to deployment. Previously, organizations relied on frequently retraining their LLMs with the latest data, a costly and time-consuming process. The NetApp AIPod with Lenovo for NVIDIA OVX combines NVIDIA-Certified OVX Lenovo ThinkSystem SR675 V3 servers with validated NetApp storage to create a converged infrastructure specifically designed for AI workloads. Using this solution, customers will be able to conduct AI RAG and inferencing operations for use cases like chatbots, knowledge management, and object recognition.

Topics covered in this VTT session include:
•Where Lenovo fits in the solution
•NetApp AIPod with Lenovo for NVIDIA OVX Solution Overview
•Challenges/pain points that this solution solves for enterprises deploying AI
•Solution value/benefits of the combined NetApp, Lenovo, and NVIDIA OVX-Certified Solution

Tags: Artificial Intelligence (AI), Nvidia, Sales, Technical Sales, ThinkSystem

Published: 2024-08-13
Length: 38 minutes

Start the training:
Employee link: Grow@Lenovo
Partner link: Lenovo 360 Learning Center

Course code: DVAI206

11. **Partner Technical Webinar - NVIDIA Smart Spaces**
2024-07-24 | 60 minutes | Employees and Partners

In this 60-minute replay, Alex Pazos, NVIDIA BDM for Smart Spaces, reviewed the NVIDIA AI for Smart Spaces framework and use cases. Alex reviewed the Metropolus Framework and the Smart Spaces ecosystem. Then he reviewed several use cases including sports stadiums, warehouses, airports, and roadways.

Tags: Artificial Intelligence (AI)

Published: 2024-07-24
Length: 60 minutes

Start the training:
Employee link: Grow@Lenovo
Partner link: Lenovo 360 Learning Center

Course code: 071924

12. **Guidance for Selling NVIDIA Products at Lenovo for ISG**
    2024-07-01 | 25 minutes | Employees and Partners

    This course gives key talking points about the Lenovo and NVIDIA partnership in the Data Center. Details are included on where to find the products that are included in the partnership and what to do if NVIDIA products are needed that are not included in the partnership. Contact information is included if help is needed in choosing which product is best for your customer. At the end of this session sellers should be able to explain the Lenovo and NVIDIA partnership, describe the products Lenovo can sell through the partnership with NVIDIA, help a customer purchase other NVIDIA product, and get assistance with choosing NVIDIA products to fit customer needs.

    Tags: Artificial Intelligence (AI), Nvidia

    Published: 2024-07-01
    Length: 25 minutes

    Start the training:
    Employee link: Grow@Lenovo
    Partner link: Lenovo 360 Learning Center

    Course code: DNVIS102

13. **Think AI Weekly: Lenovo AI PCs & AI Workstations**
    2024-05-23 | 60 minutes | Employees Only

    Join Mike Leach, Sr. Manager, Workstations Solutions and Pooja Sathe, Director Commercial AI PCs as they discuss why Lenovo AI Developer Workstations and AI PCs are the most powerful, where they fit into the device to cloud ecosystem, and this week's Microsoft announcement, Copilot+PC

    Tags: Artificial Intelligence (AI), ThinkStation

    Published: 2024-05-23
    Length: 60 minutes

    Start the training:
    Employee link: Grow@Lenovo

    Course code: DTAIW105

Related links

For more information, refer to these documents:

- ThinkSystem and ThinkAgile GPU Summary:
  https://lenovopress.lenovo.com/lp0768-thinksystem-thinkagile-gpu-summary
- ServerProven compatibility:
  https://serverproven.lenovo.com/
- NVIDIA H200 product page:
  https://www.nvidia.com/en-us/data-center/h200/
- NVIDIA Hopper Architecture page
  https://www.nvidia.com/en-us/data-center/technologies/hopper-architecture/
- ThinkSystem SR680a V3 product guide
  https://lenovopress.lenovo.com/lp1909-thinksystem-sr680a-v3-server

Related product families

Product families related to this document are the following:

- GPU adapters

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service. Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION ”AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary. Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk. Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document, LP1944, was created or updated on July 28, 2025.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP1944
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP1944.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at https://www.lenovo.com/us/en/legal/copytrade/.

The following terms are trademarks of Lenovo in the United States, other countries, or both:
Lenovo®
ServerProven®
ThinkAgile®
ThinkSystem®

The following terms are trademarks of other companies:

AMD is a trademark of Advanced Micro Devices, Inc.

Intel® is a trademark of Intel Corporation or its subsidiaries.

Linux® is the trademark of Linus Torvalds in the U.S. and other countries.

Microsoft®, Windows Server®, and Windows® are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.