ThinkSystem NVIDIA H800 PCIe Gen5 GPUs
Product Guide (withdrawn product)

The ThinkSystem NVIDIA H800 PCIe Gen5 GPU delivers high performance, scalability, and security for every workload. It uses breakthrough innovations in the NVIDIA Hopper™ architecture to deliver industry-leading conversational AI, speeding up large language models by 30X over the previous generation.

The H800 is offered in China, Hong Kong and Macau, however the H100 is not offered in these markets. The differences between the H100 and H800 are the NVLink interface speed and the FP64 performance.

The NVIDIA H800 GPU features fourth-generation Tensor Cores and the Transformer Engine with FP8 precision, further extending NVIDIA's market-leading AI leadership with up to 9X faster training and an incredible 30X inference speedup on large language models. For high-performance computing (HPC) applications, The GPUs triple the floating-point operations per second (FLOPS) of FP64 and add dynamic programming (DPX) instructions to deliver up to 7X higher performance.

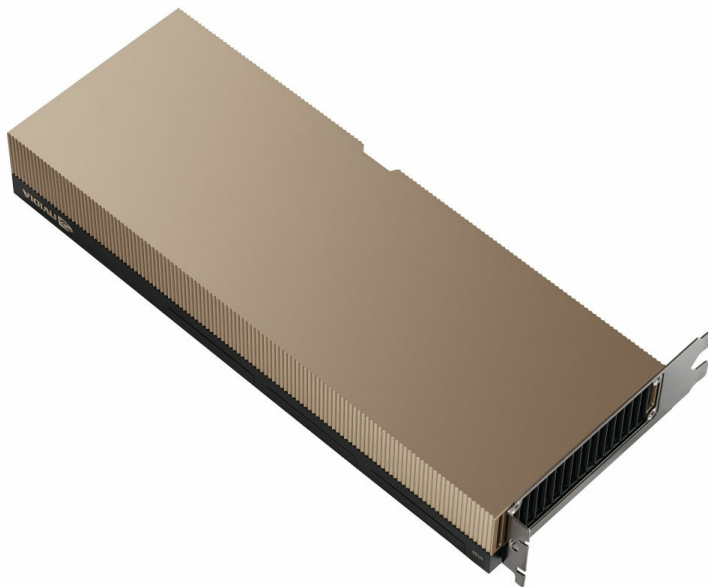The following figure shows the NVIDIA H800 GPU in double-wide PCIe adapter form factor.

Figure 1. ThinkSystem NVIDIA H800 PCIe Gen5 GPU

Did you know?

The NVIDIA H800 NVL Tensor Core GPU is optimized for Large Language Model (LLM) Inferences, with its high compute density, high memory bandwidth, high energy efficiency, and unique NVLink architecture.

Part number information

The following table shows the part numbers for the NVIDIA H800 GPU.

Withdrawn: The GPUs listed below are now withdrawn from marketing.

Table 1. Ordering information

Part number	Feature code	Description
Double-wide PCIe adapters		
4X67A89326	BXEJ	ThinkSystem NVIDIA H800 NVL 94GB PCIe Gen5 Passive GPU
4X67A86451	BUJW	ThinkSystem NVIDIA H800 80GB PCIe Gen5 Passive GPU
NVLink bridge (order 3 per pair of GPUs)		
4X67A71309	BG3F	ThinkSystem NVIDIA Ampere NVLink 2-Slot Bridge

The PCIe option part numbers includes the following:

- One GPU with full-height (3U) adapter bracket attached
- Documentation

Features

The ThinkSystem NVIDIA H800 PCIe Gen5 GPU delivers high performance, scalability, and security for every workload. The GPU uses breakthrough innovations in the NVIDIA Hopper™ architecture to deliver industry-leading conversational AI, speeding up large language models (LLMs) by 30X over the previous generation.

The NVIDIA H800 GPU for mainstream servers comes with a five-year software subscription, including enterprise support, to the NVIDIA AI Enterprise software suite, simplifying AI adoption with the highest performance. This ensures organizations have access to the AI frameworks and tools they need to build accelerated AI workflows such as AI chatbots, recommendation engines, vision AI, and more.

The NVIDIA H800 GPU features fourth-generation Tensor Cores and the Transformer Engine with FP8 precision, further extending NVIDIA's market-leading AI leadership with up to 9X faster training and an incredible 30X inference speedup on large language models. For high-performance computing (HPC) applications, the GPU triples the floating-point operations per second (FLOPS) of FP64 and adds dynamic programming (DPX) instructions to deliver up to 7X higher performance. With second-generation Multi-Instance GPU (MIG), built-in NVIDIA confidential computing, and NVIDIA NVLink Switch System, the NVIDIA H800 GPU securely accelerates all workloads for every data center from enterprise to exascale.

Key features of the NVIDIA H800 GPU:

- NVIDIA H800 Tensor Core GPU

  Built with 80 billion transistors using a cutting-edge TSMC 4N process custom tailored for NVIDIA's accelerated compute needs, H800 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 Hopper that makes NVIDIA H800 the world's first accelerator with confidential computing capabilities. Users can protect the confidentiality and integrity of their data and applications in use while accessing the unsurpassed acceleration of H800 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 GPU processing specifications and performance of the NVIDIA H800 GPU.

Table 2. Specifications of the NVIDIA H800 GPU

Feature	H800 80GB PCIe adapter	H800 NVL 94GB PCIe adapter
GPU Architecture	NVIDIA Hopper	NVIDIA Hopper
NVIDIA Tensor Cores	528 fourth-generation Tensor Cores per GPU	TBD
NVIDIA CUDA Cores (shading units)	18,432 FP32 CUDA Cores per GPU	TBD
Peak FP64 performance	0.8 TFLOPS	TBD
Peak FP64 Tensor Core performance	0.8 TFLOPS	TBD
Peak FP32 performance	51 TFLOPS	TBD
Peak Tensor Float 32 (TF32) performance	756 TFLOPS*	TBD
Peak FP16 performance	1,513 TFLOPS*	TBD
Peak Bfloat16 (BF16) performance	1,513 TFLOPS*	TBD
Peak FP8 performance	3,026 TFLOPS*	TBD
INT8 Integer Performance	3,026 TOPS*	TBD
GPU Memory	80 GB HBM2e	94 GB HBM3
Memory Bandwidth	2 TB/sec	3.9 TB/sec
ECC	Yes	Yes
Interconnect Bandwidth	NVLink: 400 GB/sec PCIe Gen5: 128 GB/sec	NVLink: 400 GB/sec PCIe Gen5: 128 GB/sec
System Interface	PCIe Gen 5.0, x16 lanes	PCIe Gen 5.0, x16 lanes
Form Factor	PCIe full height/length, double width	PCIe full height/length, double width
Multi-Instance GPU (MIG)	Up to 7 GPU instances, 10GB each	Up to 7 GPU instances, 12GB each
Max Power Consumption	350 W	400 W
Thermal Solution	Passive	Passive
Compute APIs	CUDA, DirectCompute, OpenCL, OpenACC	CUDA, DirectCompute, OpenCL, OpenACC

* With structural sparsity enabled

Server support

The following tables list the ThinkSystem servers that are compatible.

NVLink server support: The NVLink Ampere bridge is supported with additional NVIDIA A-series and H-series GPUs. As a result, there are additional servers listed as supporting the bridge that don't support the H800 GPU.

Table 3. Server support (Part 1 of 4)

Part Number	Description	SR635 V3 (7D9H / 7D9G)	SR655 V3 (7D9F / 7D9E)	SR645 V3 (7D9D / 7D9C)	SR665 V3 (7D9B / 7D9A)	ST650 V2 (7D7B / 7D7A)	SR630 V3 (7D72 / 7D73)	SR650 V3 (7D75 / 7D76)	SR850 V3 (7D97 / 7D96)	SR860 V3 (7D94 / 7D93)	SR950 V3 (7DC5 / 7DC4)	SD535 V3 (7DD8 / 7DD1)	SD530 V3 (7DDA / 7DD3)	SD550 V3 (7DD9 / 7DD2)	SR670 V2 (7Z22 / 7Z23)	SR675 V3 (7D9Q / 7D9R)	SR680a V3 (7DHE)	SR685a V3 (7DHC)	ST250 V3 (7DCF / 7DCE)	SR250 V3 (7DCM / 7DCL)
		2S AMD V3				**2S Intel V3**			**4S 8S Intel V3**			**Multi Node**			**GPU Rich**				**1S V3**	
4X67A89326	ThinkSystem NVIDIA H800 NVL 94GB PCIe Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N	N	N	8	N	N	N	N
4X67A86451	ThinkSystem NVIDIA H800 80GB PCIe Gen5 Passive GPU	N	N	N	N	N	N	3	2	4	N	N	N	N	N	8	N	N	N	N
4X67A71309	ThinkSystem NVIDIA Ampere NVLink 2-Slot Bridge	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	Y	N	N	N	N

Table 4. Server support (Part 2 of 4)

Part Number	Description	SE350 (7Z46 / 7D1X)	SE350 V2 (7DA9)	SE360 V2 (7DAM)	SE450 (7D8T)	SE455 V3 (7DBY)	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)
		Edge					**Super Computing**					**1S Intel V2**			**2S Intel V2**		
4X67A89326	ThinkSystem NVIDIA H800 NVL 94GB PCIe Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
4X67A86451	ThinkSystem NVIDIA H800 80GB PCIe Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
4X67A71309	ThinkSystem NVIDIA Ampere NVLink 2-Slot Bridge	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Table 5. Server support (Part 3 of 4)

		AMD V1					Dense V2				4S V2		8S	4S V1			1S Intel V1			
Part Number	**Description**	SR635 (7Y98 / 7Y99)	SR655 (7Y00 / 7Z01)	SR655 Client OS	SR645 (7D2Y / 7D2X)	SR665 (7D2W / 7D2V)	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)
4X67A89326	ThinkSystem NVIDIA H800 NVL 94GB PCIe Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
4X67A86451	ThinkSystem NVIDIA H800 80GB PCIe Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
4X67A71309	ThinkSystem NVIDIA Ampere NVLink 2-Slot Bridge	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Table 6. Server support (Part 4 of 4)

		2S Intel V1								Dense V1			
Part Number	**Description**	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)
4X67A89326	ThinkSystem NVIDIA H800 NVL 94GB PCIe Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N
4X67A86451	ThinkSystem NVIDIA H800 80GB PCIe Gen5 Passive GPU	N	N	N	N	N	N	N	N	N	N	N	N
4X67A71309	ThinkSystem NVIDIA Ampere NVLink 2-Slot Bridge	N	N	N	N	N	N	N	N	N	N	N	N

Operating system support

The following table lists the supported operating systems.

Tip: These tables are automatically generated based on data from Lenovo ServerProven.

Table 7. Operating system support for ThinkSystem NVIDIA H800 80GB PCIe Gen5 Passive GPU, 4X67A86451

Operating systems	SR650 V3 (4th Gen Xeon)	SR650 V3 (5th Gen Xeon)	SR675 V3	SR850 V3	SR860 V3
Microsoft Windows 10	N	Y	N	N	N
Microsoft Windows 11	N	Y	N	N	N
Microsoft Windows Server 2019	Y	Y	Y	Y [1]	Y [1]
Microsoft Windows Server 2022	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.6	Y	N	Y	Y	Y
Red Hat Enterprise Linux 8.7	Y	N	Y	Y	Y
Red Hat Enterprise Linux 8.8	Y	Y	N	Y	Y
Red Hat Enterprise Linux 9.0	Y	N	Y	Y	Y
Red Hat Enterprise Linux 9.1	Y	N	Y	Y	Y
Red Hat Enterprise Linux 9.2	Y	Y	N	Y	Y
SUSE Linux Enterprise Server 15 SP4	Y	N	Y	Y	Y
SUSE Linux Enterprise Server 15 SP5	Y	Y	N	Y	Y
Ubuntu 20.04.5 LTS	N	N	Y	Y	Y
Ubuntu 22.04 LTS	Y	N	Y	Y	Y
Ubuntu 22.04.3 LTS	N	Y	N	N	N
VMware vSphere Hypervisor (ESXi) 7.0 U3	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 8.0	Y	N	N	N	N
VMware vSphere Hypervisor (ESXi) 8.0 U1	Y	N	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 8.0 U2	Y	Y	Y	Y	Y

[1] For limitation, please refer Support Tip TT1591

NVIDIA GPU software

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

- NVIDIA vGPU Software (vApps, vPC, RTX vWS, and vCS)
- NVIDIA Omniverse Software (OVE)
- NVIDIA AI Enterprise Software
- NVIDIA HPC Compiler Software

NVIDIA vGPU Software (vApps, vPC, RTX vWS, and vCS)

Lenovo offers the following virtualization software for NVIDIA GPUs:

- **Virtual Applications (vApps)**

  For organizations deploying Citrix XenApp, VMware Horizon RDSH or other RDSH solutions. Designed to deliver PC Windows applications at full performance. NVIDIA Virtual Applications allows users to access any Windows application at full performance on any device, anywhere. This edition is suited for users who would like to virtualize applications using XenApp or other RDSH solutions. Windows Server hosted RDSH desktops are also supported by vApps.

- **Virtual PC (vPC)**

  This product is ideal for users who want a virtual desktop but need great user experience leveraging PC Windows® applications, browsers and high-definition video. NVIDIA Virtual PC delivers a native experience to users in a virtual environment, allowing them to run all their PC applications at full performance.

- **NVIDIA RTX Virtual Workstation (RTX vWS)**

  NVIDIA RTX vWS is the only virtual workstation that supports NVIDIA RTX technology, bringing advanced features like ray tracing, AI-denoising, and Deep Learning Super Sampling (DLSS) to a virtual environment. Supporting the latest generation of NVIDIA GPUs unlocks the best performance possible, so designers and engineers can create their best work faster. IT can virtualize any application from the data center with an experience that is indistinguishable from a physical workstation — enabling workstation performance from any device.

- **Virtual Compute Server (vCS)**

  NVIDIA Virtual Compute Server (vCS) enables data centers running on Red Hat Enterprise Linux, Red Hat Virtualization, and other supported KVM-based hypervisors to accelerate server virtualization with the latest NVIDIA data center GPUs, so that the most compute-intensive workloads, such as artificial intelligence, deep learning, and data science, can be run in a virtual machine (VM) powered by NVIDIA vGPU technology.

The following license types are offered:

- **Perpetual license**

  A non-expiring, permanent software license that can be used on a perpetual basis without the need to renew. Each Lenovo part number includes a fixed number of years of Support, Upgrade and Maintenance (SUMS).

- **Annual subscription**

  A software license that is active for a fixed period as defined by the terms of the subscription license, typically yearly. The subscription includes Support, Upgrade and Maintenance (SUMS) for the duration of the license term.

- **Concurrent User (CCU)**

  A method of counting licenses based on active user VMs. If the VM is active and the NVIDIA vGPU software is running, then this counts as one CCU. A vGPU CCU is independent of the connection to the VM.

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

Table 8. NVIDIA vGPU Software

Part number	Feature code 7S02CTO1WW	Description
NVIDIA vApps		

Part number	Feature code 7S02CTO1WW	Description
7S020003WW	B1MP	NVIDIA vApps Perpetual License and SUMS 5Yr, 1 CCU
7S020004WW	B1MQ	NVIDIA vApps Subscription License 1 Year, 1 CCU
7S020005WW	B1MR	NVIDIA vApps Subscription License 3 Years, 1 CCU
7S02003DWW	S832	NVIDIA vApps Subscription License 4 Years, 1 CCU
7S02003EWW	S833	NVIDIA vApps Subscription License 5 Years, 1 CCU
NVIDIA vPC		
7S020009WW	B1MV	NVIDIA vPC Perpetual License and SUMS 5Yr, 1 CCU
7S02000AWW	B1MW	NVIDIA vPC Subscription License 1 Year, 1 CCU
7S02000BWW	B1MX	NVIDIA vPC Subscription License 3 Years, 1 CCU
7S02003FWW	S834	NVIDIA vPC Subscription License 4 Years, 1 CCU
7S02003GWW	S835	NVIDIA vPC Subscription License 5 Years, 1 CCU
NVIDIA RTX vWS		
7S02000FWW	B1N1	NVIDIA RTX vWS Perpetual License and SUMS 5Yr, 1 CCU
7S02000GWW	B1N2	NVIDIA RTX vWS Subscription License 1 Year, 1 CCU
7S02000HWW	B1N3	NVIDIA RTX vWS Subscription License 3 Years, 1 CCU
7S02000XWW	S6YJ	NVIDIA RTX vWS Subscription License 4 Years, 1 CCU
7S02000YWW	S6YK	NVIDIA RTX vWS Subscription License 5 Years, 1 CCU
7S02000LWW	B1N6	NVIDIA RTX vWS EDU Perpetual License and SUMS 5Yr, 1 CCU
7S02000MWW	B1N7	NVIDIA RTX vWS EDU Subscription License 1 Year, 1 CCU
7S02000NWW	B1N8	NVIDIA RTX vWS EDU Subscription License 3 Years, 1 CCU
7S02003BWW	S830	NVIDIA RTX vWS EDU Subscription License 4 Years, 1 CCU
7S02003CWW	S831	NVIDIA RTX vWS EDU Subscription License 5 Years, 1 CCU
NVIDIA vCS		
7S02000ZWW	S6YL	NVIDIA Virtual Compute Server Subscription, 1 GPU (Max 10 CC VMs), 1 Year
7S020010WW	S6YM	NVIDIA Virtual Compute Server Subscription, 1 GPU (Max 10 CC VMs), 3 Years
7S020011WW	S6YN	NVIDIA Virtual Compute Server Subscription, 1 GPU (Max 10 CC VMs), 5 Years
7S020012WW	S6YP	NVIDIA Virtual Compute Server Subscription, 1 GPU (Max 10 CC VMs), EDU, 1 Year
7S020013WW	S6YQ	NVIDIA Virtual Compute Server Subscription, 1 GPU (Max 10 CC VMs), EDU, 3 Years
7S020014WW	S6YR	NVIDIA Virtual Compute Server Subscription, 1 GPU (Max 10 CC VMs), EDU, 5 Years

NVIDIA Omniverse Software (OVE)

NVIDIA Omniverse™ Enterprise is an end-to-end collaboration and simulation platform that fundamentally transforms complex design workflows, creating a more harmonious environment for creative teams.

NVIDIA and Lenovo offer a robust, scalable solution for deploying Omniverse Enterprise, accommodating a wide range of professional needs. This document details the critical components, deployment options, and support available, ensuring an efficient and effective Omniverse experience.

Deployment options cater to varying team sizes and workloads. Using Lenovo NVIDIA-Certified Systems™ and Lenovo OVX nodes which are meticulously designed to manage scale and complexity, ensures optimal performance for Omniverse tasks.

Deployment options include:

- Workstations: NVIDIA-Certified Workstations with A5000 or A6000 Ada GPUs for desktop environments.
- Data Center Solutions: Deployment with Lenovo OVX nodes or NVIDIA-Certified Servers equipped with L40, L40S or A40 GPUs for centralized, high-capacity needs.

NVIDIA Omniverse Enterprise includes the following components and features:

- Platform Components: Kit, Connect, Nucleus, Simulation, RTX Renderer.
- Foundation Applications: USD Composer, USD Presenter.
- Omniverse Extensions: Connect Sample & SDK.
- Integrated Development Environment (IDE)
- Nucleus Configuration: Workstation, Enterprise Nucleus Server (supports up to 8 editors per scene); Self-Service Public Cloud Hosting using Containers.
- Omniverse Farm: Supports batch workloads up to 8 GPUs.
- Enterprise Services: Authentication (SSO/SSL), Navigator Microservice, Large File Transfer, User Accounts SAML/Account Directory.
- User Interface: Workstation & IT Managed Launcher.
- Support: NVIDIA Enterprise Support.
- Deployment Scenarios: Desktop to Data Center: Workstation deployment for building and designing, with options for physical or virtual desktops. For batch tasks, rendering, and SDG workloads that require headless compute, Lenovo OVX nodes are recommended.

The following part numbers are for a subscription license which is active for a fixed period as noted in the description. The license is for a named user which means the license is for named authorized users who may not re-assign or share the license with any other person.

Table 9. NVIDIA Omniverse Software (OVE)

Part number	Feature 7S02CTO1WW	Description
7S02003ZWW	SCX0	NVIDIA Omniverse Enterprise Subscription per GPU, 1 Year
7S020042WW	SCX3	NVIDIA Omniverse Enterprise Subscription per GPU, 3 Years
7S020041WW	SCX2	NVIDIA Omniverse Enterprise Subscription per GPU, INC, 1 Year
7S020040WW	SCX1	NVIDIA Omniverse Enterprise Subscription per GPU, EDU, 1 Year
7S020043WW	SCX4	NVIDIA Omniverse Enterprise Subscription per GPU, EDU, 3 Years

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 10. 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 11. NVIDIA AI Enterprise Software (NVAIE)

Part number	Feature code 7S02CTO1WW	Description
AI Enterprise Perpetual License		
7S020019WW	S6YW	NVIDIA AI Enterprise Perpetual License and Support per GPU, 1 Year
7S02001AWW	S6YX	NVIDIA AI Enterprise Perpetual License and Support per GPU, 3 Years
7S02001BWW	S6YY	NVIDIA AI Enterprise Perpetual License and Support per GPU, 5 Years
7S02001CWW	S6YZ	NVIDIA AI Enterprise Perpetual License and Support per GPU, EDU, 1 Year
7S02001DWW	S6Z0	NVIDIA AI Enterprise Perpetual License and Support per GPU, EDU, 3 Years
7S02001EWW	S6Z1	NVIDIA AI Enterprise Perpetual License and Support per GPU, EDU, 5 Years
AI Enterprise Subscription License		
7S02001FWW	S6Z2	NVIDIA AI Enterprise Subscription License and Support per GPU, 1 Year

Part number	Feature code 7S02CTO1WW	Description
7S02001GWW	S6Z3	NVIDIA AI Enterprise Subscription License and Support per GPU, 3 Years
7S02001HWW	S6Z4	NVIDIA AI Enterprise Subscription License and Support per GPU, 5 Years
7S02001JWW	S6Z5	NVIDIA AI Enterprise Subscription License and Support per GPU, EDU, 1 Year
7S02001KWW	S6Z6	NVIDIA AI Enterprise Subscription License and Support per GPU, EDU, 3 Years
7S02001LWW	S6Z7	NVIDIA AI Enterprise Subscription License and Support per GPU, EDU, 5 Years

Find more information in the NVIDIA AI Enterprise Sizing Guide.

NVIDIA HPC Compiler Software

Table 12. 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
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

Part number	Feature code 7S09CTO6WW	Description
7S09002VWW	S9V3	NVIDIA HPC Compiler Premier Support Services - Additional Contact, EDU, 5 Years

Auxiliary power cables

The power cables needed for the H800 SXM GPUs are included with the supported servers.

The H800 PCIe GPU option part number does not ship with auxiliary power cables. Cables are server-specific due to length requirements. For CTO orders, auxiliary power cables are derived by the configurator. For field upgrades, cables will need to be ordered separately as listed in the table below.

Table 13. Auxiliary power cables for H800

Auxiliary power cable needed with the SR650 V3, SR655 V3, SR665 V3, SR665, SR650 V2

400mm 16-pin (2x6+4) cable
Option:
SR650 V3: 4X67A82883, ThinkSystem SR650 V3 GPU Full Length Thermal Option Kit*
SR655 V3: 4X67A86438, ThinkSystem SR655 V3 GPU Enablement Kit*
SR665 V3: 4X67A85856, ThinkSystem SR665 V3 GPU Full Length Thermal Option Kit*
SR665, SR650 V2: 4X97A85028, ThinkSystem 400mm 2x6+4 GPU Power Cable
Feature: BRWK
SBB: SBB7A66338
Base: SC17B33047
FRU: 03KM846

* The option part numbers are for thermal kits and include other components needed to install the GPU. See the SR650 V3 product guide or SR655 V3 product guide or SR665 V3 product guide for details.

Auxiliary power cable needed with the SR675 V3

235mm 16-pin (2x6+4) cable
Option: 4X97A84510, ThinkSystem SR675 V3 Supplemental Power Cable for H100 GPU Option
Feature: BSD2
SBB: SBB7A65299
Base: SC17B39301
FRU: 03LE554

Auxiliary power cable needed with the SR850 V3, SR860 V3

200mm 16-pin (2x6+4) cable
Option: 4X97A88016, ThinkSystem SR850 V3/SR860 V3 H100 GPU Power Cable Option Kit
Feature: BW28
SBB: SBB7A72759
Base: SC17B40604
FRU: 03LF915

Auxiliary power cable needed with the SR670 V2

215mm 16-pin (2x6+4) cable
Option: 4X97A85027, ThinkSystem SR670 V2 H100/L40 GPU Option Power Cable
Feature: BRWL
SBB: SBB7A66339
Base: SC17B33046
FRU: 03KM845

Regulatory approvals

The NVIDIA H800 GPU has the following regulatory approvals:

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

Operating environment

The NVIDIA H800 GPU has the following operating characteristics:

- Ambient temperature
    - Operational: 0°C to 50°C (-5°C to 55°C for short term*)
    - Storage: -40°C to 75°C

- Relative humidity:
    - Operational: 5-85% (5-93% short term*)
    - Storage: 5-95%

* A period not more than 96 hours consecutive, not to exceed 15 days per year.

Warranty

One year limited warranty. When installed in a Lenovo server, the GPU assumes the server's base warranty and any warranty upgrades.

Related publications

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 Hopper Architecture page
  https://www.nvidia.com/en-us/data-center/technologies/hopper-architecture/

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, LP1814, was created or updated on October 31, 2023.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP1814
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP1814.

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:

Intel® and Xeon® are trademarks 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.