ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU

Product Guide

The ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU delivers universal acceleration and energy efficiency for video, AI, virtual workstations, and graphics applications in the enterprise, in the cloud, and at the edge. And with NVIDIA's AI platform and full-stack approach, the NVIDIA L4 GPU is optimized for video and inference at scale for a broad range of AI applications to deliver the best in personalized experiences.

The following figure shows the ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU.

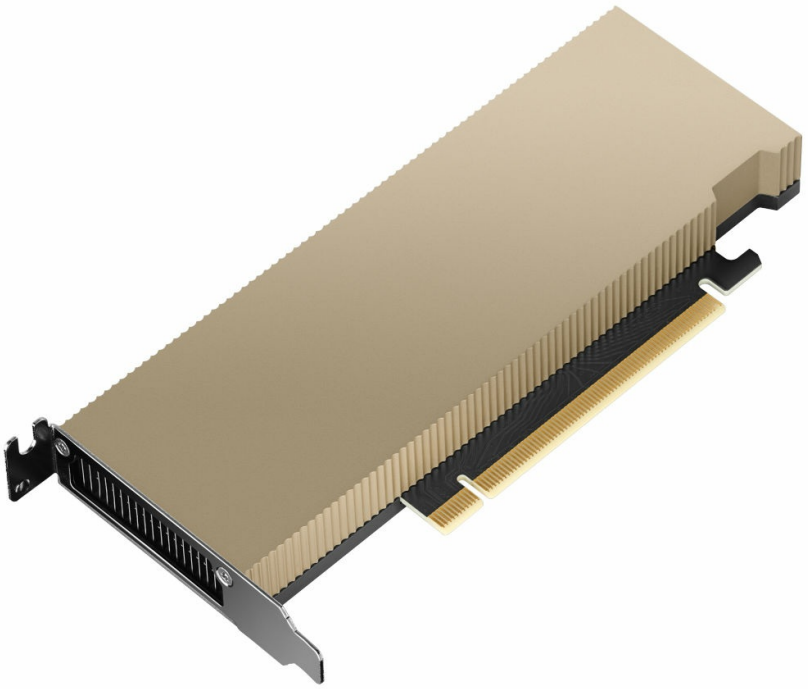

Figure 1. ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU

Did you know?

The NVIDIA L4 Tensor Core GPU powered by the NVIDIA Ada Lovelace architecture is a universal, energy-efficient accelerator designed to meet AI needs across video, visual computing, graphics, virtualization, and numerous applications, including cloud gaming, simulation, and data science. It's a true universal GPU in a low-profile form factor that delivers a cost-effective, energy-efficient solution for high throughput and low latency in every server, from the edge to the data center to the cloud.

Part number information

The following table shows the part numbers for the GPU.

Table 1. Ordering information

Part number	Feature code	Description
4X67A84824	BS2C	ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU

The option part number includes the following:

- One NVIDIA L4 GPU with full-height (3U) adapter bracket attached
- Low-profile (2U) adapter bracket
- Documentation

Features

The NVIDIA Ada Lovelace L4 Tensor Core GPU delivers universal acceleration and energy efficiency for video, AI, virtualized desktop, and graphics applications in the enterprise, in the cloud, and at the edge. With NVIDIA's AI platform and full-stack approach, L4 is optimized for video and inference at scale for a broad range of AI applications, including recommendations, voice-based AI avatar assistants, generative AI, visual search, and contact center automation to deliver the best personalized experiences.

As the most efficient NVIDIA accelerator for mainstream, servers equipped with L4 enable up to 120X higher AI Video performance over CPU solutions, while providing 2.7X more generative AI performance, and over 4X more graphics performance versus the previous generation NVIDIA T4. NVIDIA L4's versatility and energy-efficient, single-slot, low-profile form factor make it ideal for global deployments, including edge locations.

Real-time AI Video Pipeline Performance

Transform video applications with the power of NVIDIA L4. Whether streaming live to millions of viewers, enabling users to build creative stories, delivering immersive AR/VR experiences, servers equipped with L4's hardware-accelerated video encoders and decoders allow hosting up to 1040 AV1 video-streams at 720p30 for mobile users concurrently (Measured performance: AV1 low-latency encode with p1 preset).

With the fourth-generation Tensor Core technology with added FP8 precision support and 1.5X larger GPU memory, NVIDIA L4 GPUs paired with the CV-CUDA library take video content understanding to a new level. L4 delivers 120X higher AI Video performance than CPU based solutions for the entire end-to-end pipeline, enabling enterprises to gain real-time insights to provide personalized content, improve search relevance, detect objectionable content, and implement smart space solutions.

Generative AI Performance

Generative AI application capabilities in image and text generation make customer lives more convenient and experiences more immersive across all industries. NVIDIA L4 supercharges computationally intensive generative AI inference by delivering up to 2.7X higher performance compared to the NVIDIA T4. And, with 50% more memory capacity, L4 enables larger image generation up to 1024x768, not possible on the T4 GPU.

High Energy Efficiency and Low TCO

As AI and video become more pervasive, the demand for efficient, cost effective computing is increasing more than ever. NVIDIA L4 GPUs deliver up to 120X better AI video performance, resulting in up to 99% better energy efficiency and lower total cost of ownership compared to traditional CPU-based infrastructure. This enables enterprises to reduce rack space and significantly lower the overall carbon footprint while making their data centers capable of scaling to many more users.

The energy saved by switching from CPUs to NVIDIA L4s in a 2MW data center can power over 2,000 homes for one year or the carbon offset from 172,000 trees grown over 10 years. (results from the EPA calculator using 935MW savings)

Optimized Graphics Performance

With third-generation RT cores and AI-powered DLSS 3, NVIDIA L4 delivers over 4X higher performance for AI-based avatars, NVIDIA Omniverse™ virtual worlds, cloud gaming, and virtual workstations enabling creators to build real-time cinematic- quality graphics and incredibly detailed scenes for an immersive visual experience not possible with CPUs.

Sustainable Workload Acceleration

The NVIDIA L4 is an integral part of the NVIDIA data center platform. Built for video, AI, virtual workstation (vWS), graphics, simulation, data science, and data analytics, the platform accelerates over 3,000 applications and is available everywhere at scale, from data center to edge to cloud, delivering both dramatic performance gains and energy-efficiency opportunities.

Optimized for mainstream deployments, L4 delivers a low-profile form factor operating in a 72W low-power envelope, making it an efficient, cost-effective solution for any server or cloud instance from NVIDIA's partner ecosystem.

Enterprise Ready: AI Software Streamlines Development and Deployment

Enterprise adoption of AI is now mainstream, and organizations require end-to-end, AI-ready infrastructure that will future-proof them for this new era. NVIDIA AI Enterprise is an end-to-end, cloud-native suite of AI and data analytics software optimized to help every organization excel at AI and certified to deploy anywhere, from the enterprise data center to the cloud. It comes with included global enterprise support to ensure AI projects stay on track.

Optimized to streamline AI development and deployment, NVIDIA AI Enterprise includes proven, open-source containers and frameworks that are certified to run on common data center platforms and mainstream NVIDIA-Certified Systems™ with NVIDIA L4 GPUs. Since support is included, organizations get the transparency of open source and the assurance of global NVIDIA Enterprise Support with AI expertise for both their AI practitioners and IT administrators.

NVIDIA AI Enterprise software is a license addition for NVIDIA L4 Tensor Core GPUs, making AI accessible to nearly every organization with the highest performance in training, inference, and data science. NVIDIA AI Enterprise together with NVIDIA L4 simplifies the building of an AI-ready platform, accelerates AI development and deployment, and delivers performance, security, and scalability to gather insights faster and achieve business value sooner.

Learn about all the AI workloads you can run on L4 with free, hands-on NVIDIA AI Enterprise labs through NVIDIA LaunchPad.

Technical specifications

The NVIDIA L4 GPU has the following specifications:

- Form factor
  - PCIe Low Profile adapter (69mm x 169mm)
  - NVIDIA Form Factor 5.5
- Host interface:
  - PCIe 4.0 x16
  - MSI-X interrupt messaging protocol (MSI not supported)
  - PCIe Lane Polarity Inversion and Lane Reversal
- Single Root I/O Virtualization (SR-IOV) support
  - 256 virtual functions (VFs)
  - ARI Forwarding

- Hardware Root of Trust
  - Secure boot
  - Secure firmware upgrade
  - Firmware rollback protection
  - Support for in-band firmware update disable (established after each GPU reset)
  - Secure application processor recovery

The following table lists the GPU processing specifications and performance of the NVIDIA L4.

Table 2. Specifications of the ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU

Feature	**Specification**
GPU Architecture	NVIDIA Ada Lovelace
Peak FP32 performance (non-Tensor)	30.3 TFLOPS
Peak FP16 Tensor performance	121 TFLOPS, 242 TFLOPS*
Peak Tensor Float 32 (TF32) performance	60 TFLOPS, 120 TFLOPS*
Peak Bfloat16 (BF16) performance with FP32 Accumulate	121 TFLOPS, 242 TFLOPS*
Peak FP8 Tensor performance	242.5 TFLOPS, 485 TFLOPS*
Peak Integer Performance	INT8: 242.5 TOPS, 485 TOPS*
GPU Memory	24 GB GDDR6
Memory Bandwidth	300 GB/s
ECC	Yes
NVIDIA NVLink	No support
System Interface	PCIe Gen 4, x16 lanes
Form Factor	PCIe low profile (169mm x 69mm)
Multi-Instance GPU (MIG)	No support
Max Power Consumption	72 W
Thermal Solution	Passive
vGPU Software Support	NVIDIA vPC/vApps, NVIDIA RTX Virtual Workstation (vWS)
Display connectors	None
Graphics APIs	DirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.6, Vulkan 1.3
Compute APIs	CUDA 12.0, Direct Compute, OpenCL 3.0

* With structural sparsity enabled

Server support

The following tables list the ThinkSystem servers that are compatible.

Table 3. Server support (Part 1 of 4)

Part Number	Description	2S AMD V3				2S Intel V3			4S 8S Intel V3			Multi Node			GPU Rich				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)	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)	ST50 V3 (7DF4 / 7DF3)	ST250 V3 (7DCF / 7DCE)	SR250 V3 (7DCM / 7DCL)
4X67A84824	ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU	4	5	2	5	8	3	8	4	8	N	1	1	2	8	N	N	N	N	N	N

Table 4. Server support (Part 2 of 4)

Part Number	Description	Edge					Super Computing					1S Intel V2			2S Intel V2		
		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)
4X67A84824	ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU	1	N	2	4	6	N	N	N	N	N	N	N	N	8	3	8

Table 5. Server support (Part 3 of 4)

Part Number	Description	AMD V1					Dense V2				4S V2		8S	4S V1			1S Intel V1			
		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)
4X67A84824	ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU	N	N	N	N	8	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Table 6. Server support (Part 4 of 4)

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)
4X67A84824	ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU	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 L4 24GB PCIe Gen4 Passive GPU, 4X67A84824 (Part 1 of 2)

Operating systems	SE360 V2	SE450	SE455 V3	SE350	SD530 V3	SD535 V3	SD550 V3	SR630 V3 (4th Gen Xeon)	SR630 V3 (5th Gen Xeon)	SR635 V3	SR645 V3	SR650 V3 (4th Gen Xeon)	SR650 V3 (5th Gen Xeon)	SR655 V3	SR665 V3	SR850 V3	SR860 V3	ST650 V3 (4th Gen Xeon)	ST650 V3 (5th Gen Xeon)
Microsoft Windows 10	N	N	N	N	N	N	N	N	N	Y	Y	N	Y	Y	Y	N	N	N	N
Microsoft Windows 11	N	N	N	N	N	N	N	N	N	Y	Y	N	Y	Y	Y	N	N	N	N
Microsoft Windows Server 2016	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Microsoft Windows Server 2019	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y [1]	Y [1]	Y	Y
Microsoft Windows Server 2022	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 7.9	N	Y	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Red Hat Enterprise Linux 8.3	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Red Hat Enterprise Linux 8.4	N	Y	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Red Hat Enterprise Linux 8.5	N	Y	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Red Hat Enterprise Linux 8.6	Y	Y	Y	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
Red Hat Enterprise Linux 8.7	N	Y	N	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
Red Hat Enterprise Linux 8.8	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.9	N	N	N	N	Y	N	Y	N	N	Y	Y	N	N	Y	Y	N	N	N	N
Red Hat Enterprise Linux 9.0	Y	Y	Y	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
Red Hat Enterprise Linux 9.1	N	Y	N	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
Red Hat Enterprise Linux 9.2	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 9.3	N	N	N	N	Y	N	Y	N	N	Y	Y	N	N	Y	Y	N	N	N	N
SUSE Linux Enterprise Server 15 SP3	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
SUSE Linux Enterprise Server 15 SP4	Y	Y	Y	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
SUSE Linux Enterprise Server 15 SP5	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Ubuntu 18.04.5 LTS	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Ubuntu 18.04.6 LTS	N	Y	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Ubuntu 20.04 LTS	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Ubuntu 20.04.5 LTS	Y	Y	Y	Y	N	Y	N	N	N	Y	Y	N	N	Y	Y	Y	Y	N	N
Ubuntu 22.04 LTS	Y	Y	N	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
VMware vSphere Hypervisor (ESXi) 6.7 U3	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
VMware vSphere Hypervisor (ESXi) 7.0 U3	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 8.0	N	Y	N	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	N	N	N	N
VMware vSphere Hypervisor (ESXi) 8.0 U1	Y	Y	Y	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
VMware vSphere Hypervisor (ESXi) 8.0 U2	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

[1] For limitation, please refer Support Tip TT1591

Table 8. Operating system support for ThinkSystem NVIDIA L4 24GB PCIe Gen4 Passive GPU, 4X67A84824 (Part 2 of 2)

Operating systems	SR630 V2	SR650 V2	SR670 V2	ST650 V2	SR665
Microsoft Windows 10	N	N	N	N	N
Microsoft Windows 11	N	N	N	N	N
Microsoft Windows Server 2016	Y	Y	Y	Y	Y
Microsoft Windows Server 2019	Y	Y	Y	Y	Y
Microsoft Windows Server 2022	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 7.9	Y	Y	Y	Y	Y [1]
Red Hat Enterprise Linux 8.3	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.4	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.5	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.6	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.7	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.8	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.9	N	N	N	N	N
Red Hat Enterprise Linux 9.0	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 9.1	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 9.2	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 9.3	N	N	N	N	N
SUSE Linux Enterprise Server 15 SP3	Y	Y	Y	Y	Y
SUSE Linux Enterprise Server 15 SP4	Y	Y	Y	Y	Y
SUSE Linux Enterprise Server 15 SP5	Y	Y	Y	Y	Y
Ubuntu 18.04.5 LTS	Y	Y	Y	Y	N
Ubuntu 18.04.6 LTS	N	N	N	N	N
Ubuntu 20.04 LTS	Y	Y	N	N	N
Ubuntu 20.04.5 LTS	N	N	N	N	N
Ubuntu 22.04 LTS	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 6.7 U3	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 7.0 U3	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 8.0	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 8.0 U1	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 8.0 U2	Y	Y	Y	Y	Y

[1] The OS is not supported with EPYC 7003 processors.

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)

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.

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 9. NVIDIA vGPU Software

Part number	Feature code 7S02CTO1WW	Description
NVIDIA vApps		
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

Part number	Feature code 7S02CTO1WW	Description
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 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 10. 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 11. 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)

Features	Supported in NVIDIA AI Enterprise
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 12. NVIDIA AI Enterprise Software (NVAIE)

Part number	Feature code 7S02CTO1WW	Description
AI Enterprise Perpetual License		
7S02001BWW	S6YY	NVIDIA AI Enterprise Perpetual License and Support per GPU, 5 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
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 13. 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

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

Auxiliary power cables

The NVIDIA L4 does not require an auxiliary power cable.

Regulatory approvals

The NVIDIA L4 GPU has the following regulatory approvals:

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

Operating environment

The NVIDIA L4 GPU has the following operating characteristics:

- Ambient temperature
  - Operational: 0°C to 50°C (-5°C to 55°C for short term*)
  - Non-operational: -40°C to 75°C
- Humidity: 5-85% relative humidity

* 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.

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: 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

   Published: 2024-05-23
   Length: 60 minutes
   Employee link: Grow@Lenovo
   Course code: DTAIW105
2. **VTT Cloud Architecture: NVidia Using Cloud for GPUs and AI**
   2024-05-22 | 60 minutes | Employees Only

   Join JD Dupont, NVIDIA Head of Americas Sales, Lenovo partnership and Veer Mehta, NVIDIA Solution Architect on an interactive discussion about cloud to edge, designing cloud Solutions with Nvidia GPUs and minimizing private\hybrid cloud OPEX with GPUs. Discover how you can use what is done at big public cloud providers for your customers. We will also walk through use cases and see a demo you can use to help your customers.

   Published: 2024-05-22
   Length: 60 minutes
   Employee link: Grow@Lenovo
   Course code: DVCLD212

3. **Partner Technical Webinar - Nvidia Update**
   2024-05-13 | 60 minutes | Employees and Partners

   In this 60-minute replay, Veer Mehta, Nvidia Solutions Architect gave an Nvidia AI update for Lenovo. Veer reviewed the highlights from the Nvidia GTC. He also reviewed the Nvidia hardware and software offerings that Lenovo sells.

   Published: 2024-05-13
   Length: 60 minutes
   Employee link: Grow@Lenovo
   Partner link: Lenovo Partner Learning
   Course code: 051024

4. **Generative AI Overview Foundational**
   2024-02-16 | 17 minutes | Employees Only

   It seems the whole world is excited about Generative AI, and while some of it is just hype, it has become clear that Generative AI has the potential to revolutionize many aspects of our personal and professional lives. In this brief NVIDIA course, we'll explore one aspect of the Generative AI excitement, the value you get from Generative AI technology. We will discuss what Generative AI is, how it works, and how enterprises are planning to use this technology.

   By the end of this course, you will be able to discuss the Generative AI market trends and the challenges in this space with your customers. And you will be able to explain what Generative AI is and how the technology works to help enterprises unlock new opportunities for business.

   Published: 2024-02-16
   Length: 17 minutes
   Employee link: Grow@Lenovo
   Course code: DAINVD106

5. **Industry Use Cases in Modern Computing Foundational**
   2024-02-16 | 9 minutes | Employees Only

   As GPU powered computing continues to improve exponentially, applications that were once science fiction are becoming best practice. This is an introductory NVIDIA course that explores some exciting industry focused use cases that are providing companies with faster time to insight, productivity at scale and a great ROI.

   By the end of this course, you will be able to explain how companies in a few key industry verticals are benefiting from a variety of accelerated compute use cases.

   Published: 2024-02-16
   Length: 9 minutes
   Employee link: Grow@Lenovo
   Course code: DAINVD105

6. **Introduction to Artificial Intelligence Foundational**
   2024-02-16 | 10 minutes | Employees Only

   This NVIDIA course aims to answer questions such as, what is AI and why are enterprises so interested in it? and how does AI happen, why are GPUs so important for it, and what does a good AI solution look like?

   By the end of this training, you should be able to describe AI and relate it to some common enterprise use cases. You'll know the difference between training and inference and be able to visualize a typical AI workflow. More importantly, you'll understand the difficulties of traditional CPU-based AI and appreciate why businesses would benefit greatly by adopting GPU-accelerated workflows. Finally, you'll also understand what features contribute to an awesome AI solution and why customers respect and enjoy NVIDIA's solutions.

   Published: 2024-02-16
   Length: 10 minutes
   Employee link: Grow@Lenovo
   Course code: DAINVD104

7. **GPU Fundamentals Foundational**
   2024-02-16 | 10 minutes | Employees Only

   This NVIDIA course introduces you to two devices that a computer typically uses to process information, the CPU and the GPU. We'll discuss their differences and look at how the GPU overcomes the limitations of the CPU. Once you understand the power and advantages of GPU processing, we will talk about the value GPUs bring to modern-day enterprise computing.

   By the end of this course, you should know the difference between serial and parallel processing. You will be able to explain what a GPU is in very simple terms and explain the value that GPUs bring to enterprises. Additionally, you'll become familiar with the typical GPU-accelerated enterprise workloads and list one or two use cases under them. By the time you exit this course, you should be able to target various GPU-accelerated computing opportunities with the right NVIDIA GPU.

   Published: 2024-02-16
   Length: 10 minutes
   Employee link: Grow@Lenovo
   Course code: DAINVD103

8. **Partner Technical Webinar – NVidia**
   2023-12-11 | 60 minutes | Employees and Partners

   In this 60-minute replay, Brad Davidson of Nvidia will help us recognize AI Trends, and Discuss Industry Verticals Marketing.

   Published: 2023-12-11
   Length: 60 minutes
   Employee link: Grow@Lenovo
   Partner link: Lenovo Partner Learning
   Course code: 120823

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:
  http://www.lenovo.com/us/en/serverproven
- NVIDIA L4 product page:
  https://www.nvidia.com/en-us/data-center/l4/

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, LP1717, was created or updated on August 29, 2023.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP1717
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP1717.

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®, DirectX®, 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.