ThinkSystem NVIDIA A30 24GB PCIe Gen4 Passive GPU

Product Guide

The NVIDIA A30 offers versatile compute acceleration for mainstream enterprise servers. With NVIDIA Ampere architecture Tensor Cores and Multi-Instance GPU (MIG), it delivers speedups securely across diverse workloads, including AI inference at scale and HPC applications. The A30 combines fast memory bandwidth and low-power consumption in a PCIe form factor to enable an elastic data center and delivers maximum value for enterprises.

The third-generation Tensor Core technology supports a broad range of math precisions providing a unified workload accelerator for data analytics, AI training, AI inference, and HPC. Accelerating both scale-up and scale-out workloads on one platform enables elastic data centers that can dynamically adjust to shifting application workload demands. This simultaneously boosts throughput and drives down the cost of data centers.

Figure 1. ThinkSystem NVIDIA A30 24GB PCIe Gen4 Passive GPU

Did you know?

The NVIDIA A30 Tensor Core GPU delivers a versatile platform for mainstream enterprise workloads, like AI inference, training, and HPC. With TF32 and FP64 Tensor Core support, as well as an end-to-end software and hardware solution stack, A30 ensures that mainstream AI training and HPC applications can be rapidly addressed.

Part number information

The following table shows the part numbers for the NVIDIA A30 GPU.

Table 1. Ordering information

Part number	Feature code	Description
4X67A76581	BQZR	ThinkSystem NVIDIA A30 24GB PCIe Gen4 Passive GPU w/o CEC
CTO only	BJHG	ThinkSystem NVIDIA A30 24GB PCIe Gen4 Passive GPU
4X67A71309	BG3F	ThinkSystem NVIDIA Ampere NVLink 2-Slot Bridge

The PCIe option part numbers includes the following:

- One NVIDIA A30 GPU with full-height (3U) adapter bracket attached
- Documentation

GPUs without a CEC chip: The NVIDIA A30 GPU is offered without a CEC chip (look for "w/o CEC" in the name). The CEC is a secondary Hardware Root of Trust (RoT) module that provides an additional layer of security, which can be used by customers who have high regulatory requirements or high security standards. NVIDIA uses a multi-layered security model and hence the protection offered by the primary Root of Trust embedded in the GPU is expected to be sufficient for most customers. The CEC defeatured products still offer Secure Boot, Secure Firmware Update, Firmware Rollback Protection, and In-Band Firmware Update Disable. Specifically, without the CEC chip, the GPU does not support Key Revocation or Firmware Attestation. CEC and non-CEC GPUs of the same type of GPU can be mixed in field upgrades.

Features

The ThinkSystem NVIDIA A30 24GB PCIe Gen4 Passive GPU offers the following features:

- Third-Generation NVIDIA Tensor Core: Performance and Versatility

  Compared to the NVIDIA T4 Tensor Core GPU, the third-generation Tensor Cores on NVIDIA A30 deliver over 20X more AI training throughput using TF32 without any code changes and over 5X more inference performance. In addition, A30 adds BFLOAT16 to support a full range of AI precisions.

- TF32: Higher Performance for AI Training, Zero Code Changes

  A30 supports a new precision, TF32, which works just like FP32 while providing 11X higher floating point operations per second (FLOPS) over the prior-generation V100 for AI without requiring any code changes. NVIDIA's automatic mixed precision (AMP) feature enables a further 2X boost to performance with just one additional line of code using FP16 precision. A30 Tensor Cores also include support for BFLOAT16, INT8, and INT4 precision, making A30 an incredibly versatile accelerator for both AI training and inference.

- Double-Precision Tensor Cores: The Biggest Milestone Since FP64 for HPC

  A30 brings the power of Tensor Cores to HPC, providing the biggest milestone since the introduction of double-precision GPU computing for HPC. The third generation of Tensor Cores in A30 enables matrix operations in full, IEEE-compliant, FP64 precision. Through enhancements in NVIDIA CUDA-X™ math libraries, a range of HPC applications that need double-precision math can now see boosts of up to 30% in performance and efficiency compared to prior generations of GPUs.

- Multi-Instance GPU: Multiple Accelerators in One GPU

  With Multi-Instance GPU (MIG), each A30 can be partitioned into as many as four GPU instances, fully isolated at the hardware level with their own high-bandwidth memory, cache, and compute cores. MIG works on NVIDIA AI Enterprise with VMware vSphere and NVIDIA Virtual Compute Server (vCS) with hypervisors such as Red Hat RHEL/RHV.

- PCIe Gen 4: Double the Bandwidth and NVLINK Between GPU pairs

  A30 supports PCIe Gen 4, which doubles the bandwidth of PCIe Gen 3 from 15.75 GB/sec to 31.5 GB/sec, improving data transfer speeds from CPU memory for data-intensive tasks and datasets. For additional communication bandwidth, A30 supports NVIDIA NVLink between pairs of GPUs, which provides data transfer rates up to 200 GB/sec.

- 24 GB of GPU Memory

  A30 features 24 GB of HBM2 memory with 933 GB/s of memory bandwidth, delivering 1.5X more memory and 3X more bandwidth than T4 to accelerate AI, data science, engineering simulation, and other GPU memory-intensive workloads.

- Structural Sparsity: 2X Higher Performance for AI

  Modern AI networks are big, having millions and in some cases billions of parameters. Not all of these parameters are needed for accurate predictions, and some can be converted to zeros to make the models “sparse” without compromising accuracy. Tensor Cores in A30 can provide up to 2X higher performance for sparse models. While the sparsity feature more readily benefits AI inference, it can also improve the performance of model training.

Technical specifications

The following table lists the NVIDIA A30 GPU specifications.

Table 2. A30 specifications

Feature	Specification
GPU Architecture	NVIDIA Ampere
NVIDIA Tensor Cores	224 third-generation Tensor Cores per GPU
NVIDIA CUDA Cores (shading units)	3804 FP32 CUDA Cores per GPU
Double-Precision Performance	FP64: 5.2 TFLOPS FP64 Tensor Core: 10.3 TFLOPS
Single-Precision Performance	FP32: 10.3 TFLOPS Tensor Float 32 (TF32): 82 TFLOPS, 165 TFLOPS*
Half-Precision Performance	FP16: 165 TFLOPS, 330 TFLOPS*
Bfloat16	165 TFLOPS, 330 TFLOPS*
Integer Performance	INT8: 330 TOPS, 661 TOPS* INT4: 661 TOPS, 1321 TOPS*
GPU Memory	24 GB HBM2
Memory Bandwidth	933 GB/sec
ECC	Yes
Interconnect Bandwidth	64 GB/sec
System Interface	PCIe Gen 4, x16 lanes
Form Factor	PCIe full height/length, double width
Multi-Instance GPU (MIG)	Up to 4 GPU instances
Max Power Consumption	165 W
Thermal Solution	Passive
Compute APIs	CUDA, DirectCompute, OpenCL, OpenACC

* 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	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)
4X67A76581	ThinkSystem NVIDIA A30 24GB PCIe Gen4 Passive GPU w/o CEC	N	3	N	3	N	N	3	N	N	N	N	N	N	8	8	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	Y	Y	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)
4X67A76581	ThinkSystem NVIDIA A30 24GB PCIe Gen4 Passive GPU w/o CEC	N	N	N	2[1]	N	N	N	N	N	N	N	N	N	N	N	3
4X67A71309	ThinkSystem NVIDIA Ampere NVLink 2-Slot Bridge	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

1. Double-wide GPUs are only supported in the SE450 with the 360mm chassis; not supported in the 300mm chassis

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)
4X67A76581	ThinkSystem NVIDIA A30 24GB PCIe Gen4 Passive GPU w/o CEC	N	2	N	N	3	N	N	N	N	N	4	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)

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)
4X67A76581	ThinkSystem NVIDIA A30 24GB PCIe Gen4 Passive GPU w/o CEC	N	N	N	N	N	N	2	4	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 A30 24GB PCIe Gen4 Passive GPU, 4X67A76581

Operating systems	SE450	SR650 V3 (4th Gen Xeon)	SR650 V3 (5th Gen Xeon)	SR655 V3	SR665 V3	SR675 V3	SR650 V2	SR670 V2	SR860 V2	SR655	SR665	SR650 (Xeon Gen 2)	SR650 (Xeon Gen 1)
Microsoft Windows 10	N	N	Y	Y	Y	N	N	N	N	Y [1]	N	N	N
Microsoft Windows 11	N	N	Y	Y	Y	N	N	N	N	Y	N	N	N
Microsoft Windows Server 2016	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
Microsoft Windows Server 2019	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Microsoft Windows Server 2022	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 7.7	N	N	N	N	N	N	N	N	N	Y [2]	Y [2]	Y	Y
Red Hat Enterprise Linux 7.8	N	N	N	N	N	N	N	N	N	Y [2]	Y [2]	Y	Y
Red Hat Enterprise Linux 7.9	Y	N	N	N	N	N	Y	Y	Y	Y [2]	Y [2]	Y	Y
Red Hat Enterprise Linux 8.0	N	N	N	N	N	N	N	N	N	Y [2]	N	Y	Y
Red Hat Enterprise Linux 8.1	N	N	N	N	N	N	N	N	N	Y [2]	Y [2]	Y	Y
Red Hat Enterprise Linux 8.2	N	N	N	N	N	N	Y	Y	Y	Y [2]	Y [2]	Y	Y
Red Hat Enterprise Linux 8.3	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.4	Y	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.5	Y	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.6	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.7	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.8	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 8.9	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N
Red Hat Enterprise Linux 9.0	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 9.1	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 9.2	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Red Hat Enterprise Linux 9.3	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N
SUSE Linux Enterprise Server 15	N	N	N	N	N	N	N	N	N	N	N	Y	Y
SUSE Linux Enterprise Server 15 SP1	N	N	N	N	N	N	N	N	N	Y [2]	Y [2]	Y	Y
SUSE Linux Enterprise Server 15 SP2	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
SUSE Linux Enterprise Server 15 SP3	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
SUSE Linux Enterprise Server 15 SP4	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
SUSE Linux Enterprise Server 15 SP5	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Ubuntu 18.04.6 LTS	Y	N	N	N	N	N	N	N	N	N	N	N	N
Ubuntu 20.04 LTS	N	N	N	N	N	N	Y	N	N	N	N	N	N
Ubuntu 20.04.5 LTS	Y	N	N	Y	Y	Y	N	N	N	N	N	N	N
Ubuntu 22.04 LTS	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Operating systems	SE450	SR650 V3 (4th Gen Xeon)	SR650 V3 (5th Gen Xeon)	SR655 V3	SR665 V3	SR675 V3	SR650 V2	SR670 V2	SR860 V2	SR655	SR665	SR650 (Xeon Gen 2)	SR650 (Xeon Gen 1)
VMware vSphere Hypervisor (ESXi) 7.0	N	N	N	N	N	N	N	N	N	Y [2]	Y [2]	Y	Y
VMware vSphere Hypervisor (ESXi) 7.0 U1	N	N	N	N	N	N	N	N	Y	Y [2]	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 7.0 U2	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 7.0 U3	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 8.0	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 8.0 U1	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
VMware vSphere Hypervisor (ESXi) 8.0 U2	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

[1] ISG will not sell/preload this OS, but compatibility and cert only.

[2] 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)
- 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 8. NVIDIA vGPU Software

Part number	Feature code 7S02CTO1WW	Description
NVIDIA vApps		
7S020003WW	B1MP	NVIDIA vApps SUMS ONLY 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 SUMS 5Yr ONLY, 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 SUMS ONLY 5Yr, 1 CCU
7S02000GWW	B1N2	NVIDIA RTX vWS Subsc Lic 1Yr 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 SUMS ONLY 5Y, 1CCU
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 RTX 6000 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		
7S02001BWW	S6YY	NVIDIA AI Enterprise Perpetual License and Support per GPU, 5 Years

Part number	Feature code 7S02CTO1WW	Description
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 12. NVIDIA HPC Compiler

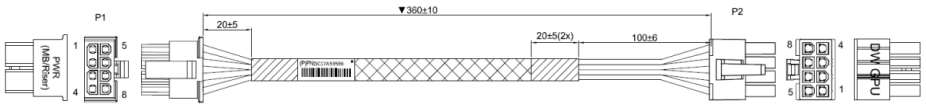
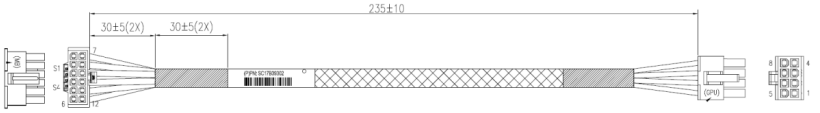
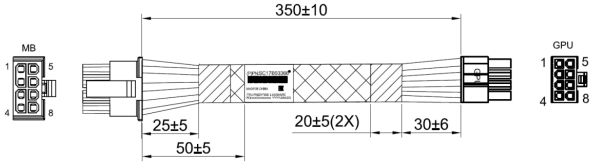
Part number	Feature code 7S09CTO6WW	Description
HPC Compiler Support Services		
7S090014WW	S924	NVIDIA HPC Compiler Support Services, 1 Year
7S090015WW	S925	NVIDIA HPC Compiler Support Services, 3 Years
7S09002GWW	S9UQ	NVIDIA HPC Compiler Support Services, 5 Years
7S090016WW	S926	NVIDIA HPC Compiler 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

Part number	Feature code 7S09CTO6WW	Description
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

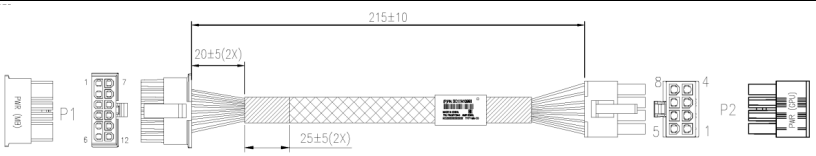
Auxiliary power cables

The A30 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 A30 (click images to show larger versions)

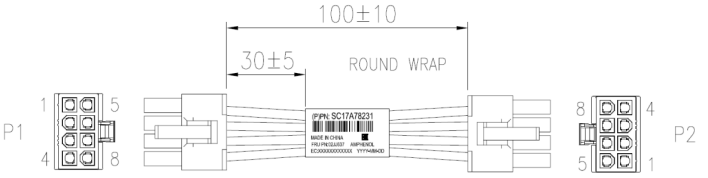
Auxiliary power cables needed with the SR665, SR650 V2, SR650 V3, SR655 V3, SR665 V3
360mm 8pin (2x4) cable **Option***: SR665: 4M17A80478 or 4M17A11759 SR650 V2: 4H47A38666 or 4H47A80491 SR650 V3: 4X67A82883 SR655 V3: 4X67A86438 SR665 V3: 4X67A85856 **Feature**: BAD8 **SBB**: SBB7A49792 or SBB7A21691 **Base**: SC17A59596 **FRU**: 02YE420  * The option part numbers are for thermal kits and include other components needed to install the GPU. See the server product guide for details.
Auxiliary power cable needed with the SR675 V3
235mm 8pin (2x4) cable **Option**: 4X97A84509, ThinkSystem SR675 V3 GPU Power Cable Kit **Feature**: BRGU **SBB**: SBB7A65297 **Base**: SC17B39302 **FRU**: 03LE555 
Auxiliary power cable needed with the SE450
350mm 8pin (2x4) cable **Option**: 4X97A81810, ThinkEdge SE450 GPU Power Cable **Feature**: BMHY **SBB**: SBB7A50523 **Base**: SC17B03366 **FRU**: 02YF808 
Auxiliary power cable needed with the SR670 V2

215mm 8pin (2x4) cable
Option: 4X97A69627, ThinkSystem SR670 V2 GPU Option Power Cable
Feature: BFNL
SBB: SBB7A40281
Base: SC17A10990
FRU: 02YE945



Auxiliary power cables needed with the SR860 V2

190mm 8pin (2x4) cable
Option: 4X97A76342, GPU Riser to GPU Power Cable, 190mm
Feature: BAX5
SBB: SBB7A17004
Base: SC17A78231
FRU: 02JJ637

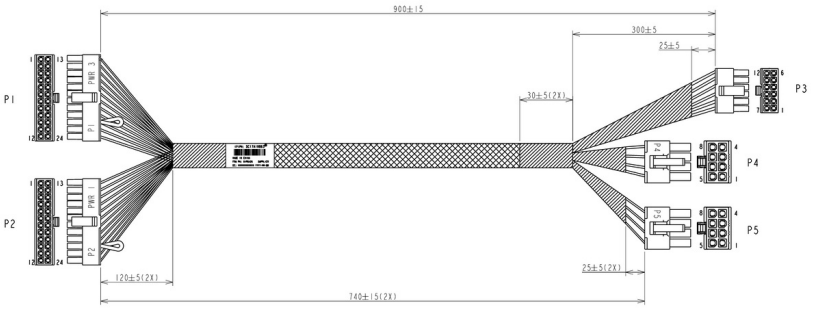


Auxiliary power cables supplied with the SR670 (configure-to-order only)

900mm SR670 Cage 1 power cable
Feature: B3Y3
SBB: SBB7A10375
Base: SC17A10876
FRU: 01PG448

Notes:

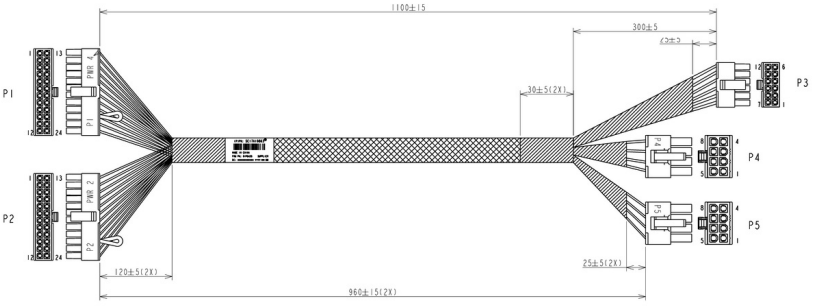
- Feature & SBB also include PCIe data cable
- Cable also supplies power to the riser



1100mm SR670 Cage 2 power cable
Feature: B3Y2
SBB: SBB7A10374
Base: SC17A10863
FRU: 01PG426

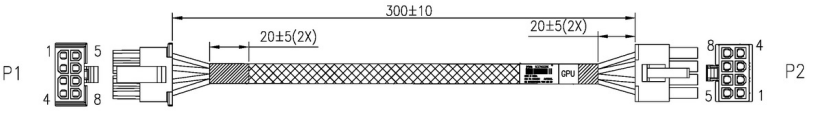
Notes:

- Feature & SBB also include PCIe data cable
- Cable also supplies power to the riser



Auxiliary power cables needed with the SR650

300mm 8pin (2x4) cable
Option: 4XH7A08794, ThinkSystem SR650 GPU Cable Kit
Feature: AUSR
SBB: SBB7A00299
Base: SC17A02296
FRU: 01KN066



Regulatory approvals

The NVIDIA A30 GPU has the following regulatory approvals:

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

Operating environment

The NVIDIA A30 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 A30 product page:
  https://www.nvidia.com/en-us/data-center/products/a30-gpu/
- NVIDIA Ampere Architecture page
  https://www.nvidia.com/en-us/data-center/ampere-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, LP1774, was created or updated on July 9, 2023.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP1774
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP1774.

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® 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.