



NVIDIA L40S - Technical Specifications

Architecture	NVIDIA Ada Lovelace Architecture
Foundry	TSMC
Process Size	4nm NVIDIA Custom Process
Transistors	76.3 billion
Die Size	608.44 mm
CUDA Parallel Processing Cores	18,176
NVIDIA Tensor Cores (4th Gen)	568
NVIDIA RT Cores (3rd Gen)	142
Peak FP32 TFLOPS (non-Tensor)	91.6
Peak FP16 Tensor TFLOPS with FP16 Accumulate	366.5 733*
Peak FP8 Tensor TFLOPs with FP16 Accumulate	733 1466*
Peak TF32 Tensor TFLOPS	183 366*
Peak BF16 Tensor TFLOPS with FP32 Accumulate	366.5 733*
Peak INT8 Tensor TOPS	733 1466*
Peak INT4 Tensor TOPS	733 1466*
RT Core Performance TFLOPS	212
GPU Memory	48 GB GDDR6 with ECC
Memory Interface	384-bit
Memory Bandwidth	864 GB/s
Interconnect	x16 PCIe Gen4 (no NVLink)
Max Power Consumption	350W
Graphics Bus	PCI Express 4.0 x16
Display Connectors	DP 1.4a (4) Supports NVIDIA Mosaic and Quadro® Sync ¹
Display Max Resolution / Quantity	Up to four 5K Monitors @ 60Hz per card or dual 8K displays @ 60Hz ² Each display port can support 4K @ 120hz with 30-bit color
Form Factor	4.4" H x 10.5" L - Dual Slot

Product Weight	1051 grams
Thermal Solution	Passive
vGPU Software Support ³	NVIDIA vPC/vApps, NVIDIA RTX Virtual Workstation (vWS)
vGPU Profiles Supported	See Virtual GPU Licensing Guide
Virtual Display Max Resolution / Quantity	Up to four 5K displays or two 8K displays for 8Q+ profiles
Graphics APIs	DirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.6 ⁴ , Vulkan 1.3 ⁴
Compute APIs	CUDA 12.0, Direct Compute, OpenCL 3.0
NVIDIA® 3D Vision® and 3D Vision Pro	Support via 3-pin mini DIN
Frame lock	Compatible (with Quadro Sync II)
Power Connector	1x PCIe CEM5 16-pin
NVENC NVDEC	3x ENC 3x DEC (includes AV1 encode and decode)
NEBS Ready	Level 3
Secure Boot with Root of Trust	Supported

Peak performance rates are based on GPU Boost Clock.

*Effective TOPS / TFLOPS using the new Sparsity Feature

1 Quadro Sync II card sold separately. Mosaic supported on Windows 10 and Linux.

2 Dual 8K monitors requires monitors that support DisplayPort 1.4 DSC for 8K resolution via a single cable.

3 L40 is configured for virtualization by default with physical display connectors disabled. The display outputs can be enabled via management software tools. Virtualization software support is anticipated in early 2023.

4 API is based on a published Khronos specification and is expected to pass the Khronos conformance testing process when available. Current conformance status can be found at www.khronos.org/conformance.