

NVIDIA® A800 40GB ACTIVE

SPECIFICATIONS

Product	NVIDIA A800 40GB Active
Architecture	NVIDIA Ampere
Foundry	TSMC
Process Size	7 nm NVIDIA Custom Process
Die Size	826 mm
CUDA® Cores	6912
Streaming Multiprocessors	108
Tensor Cores Gen 3	432
FP64 Performance	9.7 TFLOPS
FP32 Performance	19.5 TFLOPS
TF32 Tensor Core	311.8 TFLOPS*
BFLOAT16 Tensor Core	312 TFLOPS 624 TFLOPS*
FP16 Tensor Core	312 TFLOPS 624 TFLOPS*
INT8 Tensor Core	1247.4 TOPS*
INT4 Tensor Core	1248 TOPS 2496 TFLOPS*
NVLink	2-Way Low Profile, 2-Boards
NVLink Bandwidth	400 GB/s
GPU Memory	40GB HBM2
Memory Interface	5120-bit
Memory Bandwidth	1555.2 GB/s
MIG Support	Up to 7 MIG Instances
System Interface	PCIe 4.0 x16
Display Support	None Provided, use companion NVIDIA T1000 or RTX A4000 board for video output
Thermal Solution	Active
Form Factor	4.4" H x 10.5" L, Dual-Slot
Power Connector	CEM5 16-pin
Maximum Power Consumption	240W

*Structural sparcity enabled.

1 3-year software subscription and enterprise support for NVIDIAS AI Enterprise license. Activation required.