

NVIDIA A16 16GB vs NVIDIA A40 48GB

Side by side datacenter GPU specifications from the Flopper.io database.

OVERVIEW

	NVIDIA A16 16GB	NVIDIA A40 48GB
FP32 TFLOPS	4.5	37.4
Memory	16 GB	48 GB
Memory Bandwidth	200 GB/s	696 GB/s
TDP	—	300 W

PERFORMANCE BY PRECISION

Precision	NVIDIA A16 16GB peak TFLOPS (dense)	NVIDIA A40 48GB peak TFLOPS (dense)
FP64	—	—
FP32	4.5	37.4
TF32	9.0	74.8
BF16	—	149.7
FP16	17.9	149.7
FP8	—	—
FP6	—	—
FP4	—	—
INT8	35.9	299.3
INT4	—	598.7

SPECIFICATIONS

	NVIDIA A16 16GB	NVIDIA A40 48GB
Architecture	Ampere	Ampere
Process Node	—	—
Launch Year	2021	2021
Form Factor	PCIe	PCIe
Memory Type	GDDR6	GDDR6
Interface Width	—	384-bit

	NVIDIA A16 16GB	NVIDIA A40 48GB
Interconnect	—	—
TDP	—	300 W
FP32 Efficiency	—	0.12 TFLOPS/W