



NVIDIA A16 16GB vs NVIDIA GeForce RTX 3090 24GB

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

OVERVIEW

	NVIDIA A16 16GB	NVIDIA GeForce RTX 3090 24GB
FP32 TFLOPS	4.5	35.6
Memory	16 GB	24 GB
Memory Bandwidth	200 GB/s	936 GB/s
TDP	—	350 W

PERFORMANCE BY PRECISION

Precision	NVIDIA A16 16GB peak TFLOPS (dense)	NVIDIA GeForce RTX 3090 24GB peak TFLOPS (dense)
FP64	—	0.6
FP32	4.5	35.6
TF32	9.0	35.6
BF16	—	71.0
FP16	17.9	71.0
FP8	—	—
FP6	—	—
FP4	—	—
INT8	35.9	284.0
INT4	—	568.0

SPECIFICATIONS

	NVIDIA A16 16GB	NVIDIA GeForce RTX 3090 24GB
Architecture	Ampere	Ampere
Process Node	—	8nm
Launch Year	2021	2020
Form Factor	PCIe	PCIe
Memory Type	GDDR6	—
Interface Width	—	—

	NVIDIA A16 16GB	NVIDIA GeForce RTX 3090 24GB
Interconnect	—	—
TDP	—	350 W
FP32 Efficiency	—	0.10 TFLOPS/W