NVIDIA A100 PCIe 40GB vs NVIDIA GeForce RTX 3080 12GB

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

OVERVIEW

	NVIDIA A100 PCIe 40GB	NVIDIA GeForce RTX 3080 12GB
FP32 TFLOPS	**19.5**	**30.6**
Memory	**40 GB**	**12 GB**
Memory Bandwidth	**1.6 TB/s**	**—**
TDP	**250 W**	**350 W**

PERFORMANCE BY PRECISION

Precision	NVIDIA A100 PCIe 40GB peak TFLOPS (dense)	NVIDIA GeForce RTX 3080 12GB peak TFLOPS (dense)
FP64	9.7	0.5
FP32	19.5	30.6
TF32	156.0	30.6
BF16	312.0	61.3
FP16	312.0	61.3
FP8	—	—
FP6	—	—
FP4	—	—
INT8	624.0	245.2
INT4	1248.0	490.3

SPECIFICATIONS

	NVIDIA A100 PCIe 40GB	NVIDIA GeForce RTX 3080 12GB
Architecture	Ampere	Ampere
Process Node	7nm	8nm
Launch Year	2020	2022
Form Factor	PCIe	PCIe
Memory Type	—	GDDR6X
Interface Width	—	384-bit

	NVIDIA A100 PCIe 40GB	NVIDIA GeForce RTX 3080 12GB
Interconnect	–	–
TDP	250 W	350 W
FP32 Efficiency	0.08 TFLOPS/W	0.09 TFLOPS/W