NVIDIA GeForce RTX 3080 12GB

Ampere | PCIe | 2022 | 8nm

OVERVIEW

30.6	12 GB	—	350 W
FP32 TFLOPS	VRAM GDDR6X	Memory Bandwidth	TDP

PERFORMANCE METRICS

Precision	Peak (dense)	2:4 Sparse	Bits	Type
FP64	0.5	N/A	64	Standard
FP32	30.6	N/A	32	Standard
TF32	30.6	61.3	32	Standard
BF16	61.3	122.6	16	Standard
FP16	61.3	122.6	16	Standard
FP8	–	not published	—	Standard
FP6	–	not published	—	Standard
FP4	–	not published	—	Standard
INT8	245.2	490.3	8	Standard
INT4	490.3	980.6	4	Standard

MEMORY SPECIFICATIONS

Capacity	12 GB
Type	GDDR6X
Bandwidth	—
Interface Width	384-bit

POWER & EFFICIENCY

TDP	350 W
Max Power	350 W
FP32 Efficiency	0.09 TFLOPS/W
FP16 Efficiency	0.18 TFLOPS/W

HARDWARE DETAILS

Vendor: NVIDIA
Architecture: Ampere
Process Node: 8nm
Form Factor: PCIe
Launch Year: 2022
Variant: 12GB

COMPUTE ENGINE

CUDA Cores: 8,960
Tensor Cores: 280 (3rd Gen)
SMs: 70
Base Clock: 1.26 GHz
Boost Clock: 1.71 GHz

INTERCONNECT

PCIe: 4.0 x16

CHIP DESIGN

Transistors: 28.3 billion

Die Size: 628 mm^2

ENTERPRISE FEATURES

ECC: No

Sparsity: Yes