



NVIDIA GeForce RTX 4070 12GB

Ada Lovelace PCIe 2023 4nm

OVERVIEW

29.1 FP32 TFLOPS	12 GB VRAM GDDR6X	504 GB/s Memory Bandwidth	200 W TDP
---------------------	----------------------	------------------------------	--------------

PERFORMANCE METRICS

Precision	Peak (dense)	2:4 Sparse	Bits	Type
FP64	0.5	N/A	64	Standard
FP32	29.1	N/A	32	Standard
TF32	29.1	58.2	32	Standard
BF16	58.3	116.6	16	Standard
FP16	58.3	116.6	16	Standard
FP8	116.6	233.2	8	Standard
FP6	—	not published	—	Standard
FP4	—	not published	—	Standard
INT8	233.2	466.4	8	Standard

MEMORY SPECIFICATIONS

Capacity	12 GB	TDP	200 W
Type	GDDR6X	Max Power	200 W
Bandwidth	504 GB/s	FP32 Efficiency	0.15 TFLOPS/W
Interface Width	192-bit	FP16 Efficiency	0.29 TFLOPS/W

POWER & EFFICIENCY

HARDWARE DETAILS

Vendor: NVIDIA	Architecture: Ada Lovelace	Process Node: 4nm
Form Factor: PCIe	Launch Year: 2023	Variant: 12GB

COMPUTE ENGINE

CUDA Cores: 5,888	Tensor Cores: 184 (4th Gen)	SMs: 46
Base Clock: 1.92 GHz	Boost Clock: 2.48 GHz	

INTERCONNECT

PCIe: 4.0

CHIP DESIGN

Transistors: 35.8 billion Die Size: 295 mm²

ENTERPRISE FEATURES

Sparsity: Yes