

NVIDIA GB200 Grace Blackwell Superchip 372GB vs NVIDIA GeForce RTX 5070 12GB

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

OVERVIEW

	NVIDIA GB200 Grace Blackwell Superchip 372GB	NVIDIA GeForce RTX 5070 12GB
FP32 TFLOPS	160.0	30.9
Memory	372 GB	12 GB
Memory Bandwidth	16.0 TB/s	672 GB/s
TDP	—	250 W

PERFORMANCE BY PRECISION

Precision	NVIDIA GB200 Grace Blackwell Superchip 372GB peak TFLOPS (dense)	NVIDIA GeForce RTX 5070 12GB peak TFLOPS (dense)
FP64	80.0	0.5
FP32	160.0	30.9
TF32	2500.0	30.9
BF16	5000.0	61.7
FP16	5000.0	123.5
FP8	10000.0	246.9
FP6	10000.0	—
FP4	20000.0	493.9
INT8	10000.0	246.9

SPECIFICATIONS

	NVIDIA GB200 Grace Blackwell Superchip 372GB	NVIDIA GeForce RTX 5070 12GB
Architecture	Blackwell	Blackwell
Process Node	4nm	4nm
Launch Year	2024	2025
Form Factor	Superchip	PCIe
Memory Type	HBM3e	GDDR7

	NVIDIA GB200 Grace Blackwell Superchip 372GB	NVIDIA GeForce RTX 5070 12GB
Interface Width	—	192-bit
Interconnect	NVLink-C2C	—
TDP	—	250 W
FP32 Efficiency	—	0.12 TFLOPS/W