

NVIDIA GeForce RTX 5090 32GB vs NVIDIA GB200 NVL72 GPU 186GB

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

OVERVIEW

	NVIDIA GeForce RTX 5090 32GB	NVIDIA GB200 NVL72 GPU 186GB
FP32 TFLOPS	104.8	80.0
VRAM	32 GB	186 GB
Memory Bandwidth	1.8 TB/s	8.0 TB/s
TDP	575 W	—

PERFORMANCE BY PRECISION

Precision	NVIDIA GeForce RTX 5090 32GB peak TFLOPS (dense)	NVIDIA GB200 NVL72 GPU 186GB peak TFLOPS (dense)
FP64	1.6	40.0
FP32	104.8	80.0
TF32	209.5	1250.0
BF16	419.1	2500.0
FP16	419.1	2500.0
FP8	838.2	5000.0
FP6	—	5000.0
FP4	1676.0	10000.0
INT8	838.2	5000.0

SPECIFICATIONS

	NVIDIA GeForce RTX 5090 32GB	NVIDIA GB200 NVL72 GPU 186GB
Architecture	Blackwell	Blackwell
Process Node	4nm	4nm
Launch Year	2025	2024
Form Factor	PCIe	SXM
Memory Type	GDDR7	HBM3e

	NVIDIA GeForce RTX 5090 32GB	NVIDIA GB200 NVL72 GPU 186GB
Interface Width	512-bit	—
Interconnect	—	NVLink
TDP	575 W	—
FP32 Efficiency	0.18 TFLOPS/W	—