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

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

OVERVIEW

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

PERFORMANCE BY PRECISION

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

SPECIFICATIONS

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

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