

Parameter	A100 (Ampere)	H100 (Hopper)	RTX 4090D 48G (Custom)	L40	L40S	RTX 6000 Ada
Architecture	Ampere	Hopper	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace
Process Node	7nm TSMC	4nm TSMC	4nm TSMC	4nm TSMC	4nm TSMC	4nm TSMC
FP32 Base TFLOPS	19.5	51.22	82.6	90.5	91.6	91.1
FP32 Tensor TFLOPS	156	756-989	660.6/1321.2	181.05	212	1457
FP64 Base TFLOPS	9.7	25.61	5.17*	-	-	-
Memory Configuration	80GB HBM2e	80GB HBM3	48GB GDDR6X	48GB GDDR6	48GB GDDR6	48GB GDDR6
Memory Bandwidth	2,039 GB/s	3,430 GB/s	937 GB/s	864 GB/s	864 GB/s	960 GB/s
Memory ECC	Yes	Yes	Yes	Yes	Yes	Yes
CUDA Cores	6,912	14,592	14,592	18,176	18,176	18,176
CUDA Core Gen	3rd Gen	4th Gen	3rd Gen (Ada)	3rd Gen (Ada)	3rd Gen (Ada)	3rd Gen (Ada)
Tensor Core Count	432	456	456	568	568	568
Tensor Cores Gen	3rd Gen	4th Gen	4th Gen	4th Gen	4th Gen	4th Gen
RT Core Count	N/A	N/A	114	142	142	142
RT Cores Gen	N/A	N/A	3rd Gen	3rd Gen	3rd Gen	3rd Gen
TDP	400W	700W	450W	300W	350W	300W
PCIe Interface	Gen4	Gen5	Gen4	Gen4	Gen4	Gen4
NVLink Support	Yes	Yes	No	No	No	No
MIG Support	Yes	Yes	No	No	No	No
vGPU Support	Yes	Yes	No	Yes	Yes	Yes
Display Outputs	N/A	N/A	4x DP 1.4a	4x DP 1.4a	4x DP 1.4a	4x DP 1.4a
Form Factor	SXM/PCIe	PCIe	Dual-slot PCIe	Dual-slot PCIe	Dual-slot PCIe	Dual-slot PCIe
Key Features	Earliest 80GB HBM	FP64 3x A100	vRAM Upgrade	NEBS Level 3	Structural Sparsity	vGPU
Typical Use Case	HPC/AI	HPC/AI	AI Inference/Fine-Tune & HPC	Data Center	Generative AI	Professional Viz