NVIDIA® A100 LIQUID COOLED PCIe

SPECIFICATIONS

Product	NVIDIA A100 Liquid Cooled PCIe
Architecture	Ampere
Process Size	7nm | TSMC
Transistors	54 Billion
Die Size	826 mm2
CUDA Cores	6912
Streaming Multiprocessors	108
Tensor Cores | Gen 3	432
Multi-Instance GPU (MIG) Support	Yes, up to seven instances per GPU
FP64	9.7 TFLOPS
Peak FP64 Tensor Core	156 TFLOPS | 312 TFLOPS Sparsity
Peak FP32	19.5 TFLOPS
TF32 Tensor Core	156 TFLOPS | 312 TFLOPS Sparsity
Peak FP16 Tensor Core	312 TFLOPS | 624 TFLOPS Sparsity
Peak INT8 Tensor Core	624 TOPS | 1248 TOPS Sparsity
INT4 Tensor Core	1248 TOPS | 2496 TOPS Sparsity
NVLink	2-way | Standard or Wide Slot Spacing
NVLink Interconnect	600 GB/s Bidirectional
GPU Memory	80 GB HBM2e
Memory Interface	5120-bit
Memory Bandwidth	1555 GB/s
System Interface	PCIe 4.0 x16
Thermal Solution	Liquid Cooled
vGPU Support	NVIDIA AI Enterprise
Power Connector	PCIe 16-pin
Total Board Power	300 W