

# NVIDIA® A2

## SPECIFICATIONS

<b>GPU Architecture</b>	NVIDIA Ampere
<b>CUDA Cores</b>	1280
<b>Tensor Cores</b>	40   Gen 3
<b>RT Cores</b>	108 Gen 2
<b>Peak FP32</b>	4.5 TFLOPS
<b>Peak TF32 Tensor Core</b>	9 TFLOPS   18 TFLOPS Sparsity
<b>Peak FP16 Tensor Core</b>	18 TFLOPS   36 TFLOPS Sparsity
<b>INT8</b>	36 TOPS   72 TOPS Sparsity
<b>INT4</b>	72 TOPS   144 TOPS Sparsity
<b>GPU Memory</b>	16 GB GDDR6 ECC
<b>Memory Bandwidth</b>	200 GB/s
<b>Thermal Solution</b>	Passive
<b>Maximum Power Consumption</b>	40-60 Watt   Configurable
<b>System Interface</b>	PCIe Gen 4.0 x8