PNY NVIDIA A2 Module 16GB GDDR6 ECC, PCI-E 4.0 x16, LP

Kod producenta: TCSA2MATX-PB



CUDA Cores 1280
Tensor Cores 40
RT Cores 10

FP32 Performance 4.5 TFLOPS

GPU Memory 16 GB

GPU Memory Type GDDR6 ECC

Memory Bandwidth 200 GB/s

Memory Interface 128-bit

System Interface PCI-E 4.0 x8

NVENC | NVDEC 1x Video Encoder | 2x Video

Decoder

Max Power Consumption 60 W
Thermal Management Passive

NVIDIA A2

Unprecedented Acceleration for World's Highest-Performing Elastic Data Centers

The NVIDIA A2 Tensor Core GPU provides entry-level inference with low power, a small footprint, and high performance for intelligent video analytics (IVA) or NVIDIA AI at the edge. Featuring a low-profile PCIe Gen4 card and a low 40–60 watt (W) configurable thermal design power (TDP) capability, the A2 brings versatile inference acceleration to any server.

A2's versatility, compact size, and low power exceed the demands for edge deployments at scale, instantly upgrading existing entry-level CPU servers to handle inference. Servers accelerated with A2 GPUs deliver up to 20X higher inference performance versus CPUs and 1.3x more efficient IVA deployments than previous GPU generations — all at an entry-level price point.

NVIDIA-Certified systems with the NVIDIA A2, A30, and A100 Tensor Core GPUs and NVIDIA AI—including the NVIDIA Triton Inference Server, open source inference service software—deliver breakthrough inference performance across edge, data center, and cloud. They ensure that AI-enabled applications deploy with fewer servers and less power, resulting in easier deployments and faster insights with dramatically lower costs.

Strona firmowa produktu:

https://www.superstorage.pl/pny-nvidia-a2-module-16gb-gddr6-ecc-pci-40-x16-lp-p-5818.html