

NVIDIA® T4

SPECIFICATIONS

GPU Architecture	NVIDIA Turing
Use Case	Universal Deep Learning Accelerator
NVIDIA GPU	TU104-895
GPU Clocks	585 MHz Base 1590 MHz Maximum Boosts
Turing Tensor Cores	320
NVIDIA CUDA Cores	2560
Peak FP32	8.1 TFLOPS
Mixed Precision FP16/FP32	65 TFLOPS
INT8	130 TOPS
INT4	260 TOPS
GPU Memory	16 GB GDDR6
ECC	Yes Enabled by default
Memory Interface	256-bit
Maximum Memory Clock	5001 MHz
Memory Bandwidth	300 GB/s
CODECs Supported	H.264 H.265
720p Encoding Streams	22 Simultaneously in HQ mode
1080p Encoding Streams	10 Simultaneously
Ultra HD 2160p Streams	2-3 Simultaneously
Thermal Solution	Passive
Operating Temperature	0 to 50 degrees Centigrade
Operating Humidity	5% to 90% relative humidity
Maximum Power Consumption	70 W
System Interface	PCIe Gen 3.0 x16 32 GB/s PCIe Gen 3.0 x8 also supported
Form Factor	Low-Profile PCIe Single Slot
Physical Dimensions	6.61" L x 2.71" H
Compute APIs	CUDA NVIDIA TensorRT Open CL ONYX
Graphics APIs	DirectX 2 OpenGL 4.6 Vulkan 1.2