



## NVIDIA® TESLA® M4 GPU ACCELERATOR

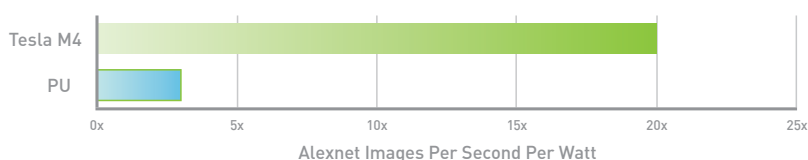
### The World's First Accelerator for the Hyperscale Data Center

Exploding volumes of user-generated data are redefining what's required for hyperscale data centers. Today's cloud applications harness valuable data to deliver smarter, real-time experiences using modern video and image processing and deep learning techniques. These applications can benefit greatly from GPU acceleration in the data center.

The NVIDIA Tesla M4 is the world's first accelerator designed for hyperscale servers, enabling customers to keep up with ever-growing amount of data. It's engineered to accelerate application throughput in a small, low-power design, slashing data center costs by half and deliver up to 7x more power-efficient processing than CPUs for deep learning inference at 20 images/sec/watt and video workloads.

### Hyperscale Application Advantage:

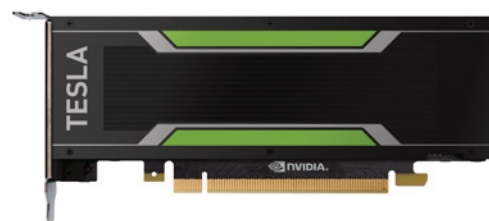
#### 6.5x More Power Efficient



#### 3x Faster Deep Learning Inference



PU E5-2699v3



### FEATURES

NVIDIA GPU Boost™, which delivers up to 2.2 Teraflops of single-precision performance

Small, low-power design for hyperscale servers

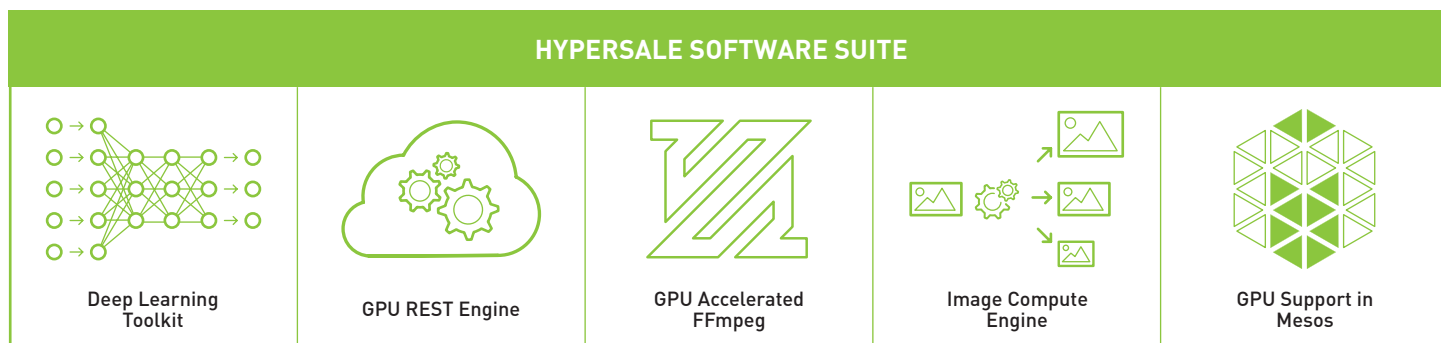
Server qualification to deliver maximum uptime in the data center

### SPEIFICATIONS

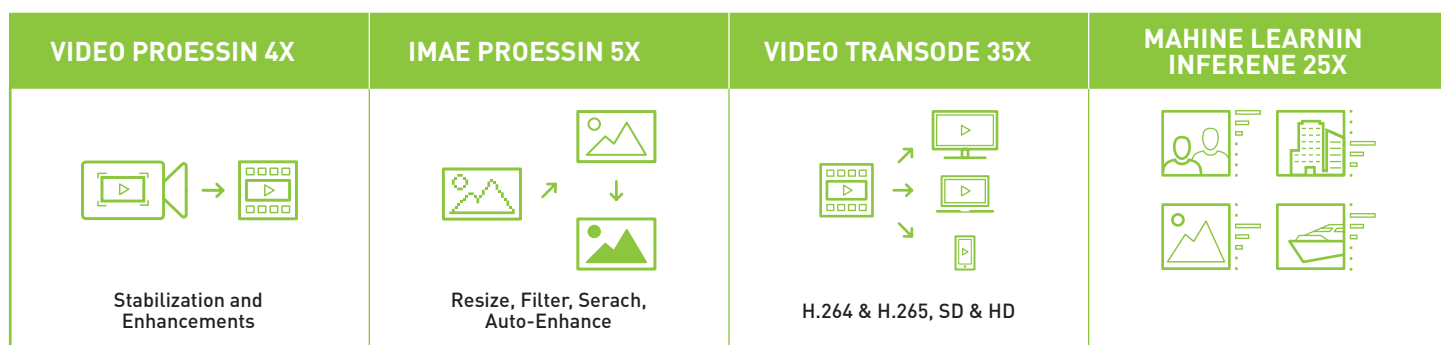
GPU Architecture	NVIDIA Maxwell™
NVIDIA CUDA® Cores	1024
Single-Precision Performance	2.2 Teraflops with NVIDIA GPU Boost
Double-Precision Performance	.07 Teraflops with NVIDIA GPU Boost
GPU Memory	4 GB GDDR5
Memory Bandwidth	88 GB/s
System Interface	PCIe Gen3
Max Power Consumption	50W-75W
Thermal Solution	Passive
Form Factor	Low Profile
Compute APIs	NVIDIA CUDA, DirectCompute, OpenCL, OpenACC

## HYPERSCALE WORKLOADS ACCELERATED WITH NVIDIA TESLA M4

The Tesla M4 accelerator and NVIDIA Hyperscale suite provide a powerful foundation of best-in-class modern hyperscale data centers. The NVIDIA Tesla Hyperscale Suite includes GPU Rest Engine for real-time accelerated services, GPU-accelerated FFMPEG for optimizing video processing, Image Compute Engine for efficient and dynamic image resizing, and GPU Inference Engine for modern deep learning workloads.



## Slash Data Center Costs By Half And Deliver Up To 5x More Performance For Hyperscale Workloads With Tesla M4.



**THINKMATE**  
HIGH PERFORMANCE COMPUTING

