

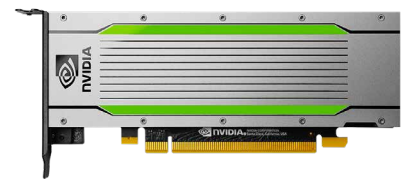


## NVIDIA T4 FOR VIRTUALIZATION



### Flexibly Power Any Virtual Workload

The NVIDIA T4 data center GPU now supports virtualization workloads. Based on the latest NVIDIA Turing™ architecture, this solution can be deployed with Tesla T4 – the most universal GPU to date capable of running any workload. By providing the computation required to deliver real-time ray tracing, the same GPU computing platform used by designers and engineers can now also be used by artists to create photorealistic imagery that features light bouncing off surfaces just as it would in real life. This computing platform also enables AI-enhanced graphics, video, and image processing to speed creative workflows, as well as Tensor Cores to accelerate deep learning inferencing workloads. And as the modern digital workplace becomes more graphics intensive, this platform can be used to provide a native-PC user experience for virtualized office productivity applications. With support for Live Migration of GPU accelerated VMs and other high performance technologies, this universal GPU accelerator provides the ultimate flexibility for cost-effectively scaling VDI.



#### SPECIFICATIONS

GPU Architecture	<b>NVIDIA Turing</b>
NVIDIA Turing Tensor Cores	<b>320</b>
NVIDIA CUDA® Cores	<b>2,560</b>
RT Cores	<b>40</b>
Memory Size	<b>16 GB GDDR6</b>
Memory BW	<b>Up to 320 GB/sec</b>
vGPU Profiles	<b>1 GB, 2 GB, 4 GB, 8 GB, 16 GB</b>
Form Factor	<b>PCIe 3.0 single slot (half height &amp; length)</b>
Power	<b>70 W</b>
Thermal	<b>Passive</b>

## Faster Virtual Workstation Performance

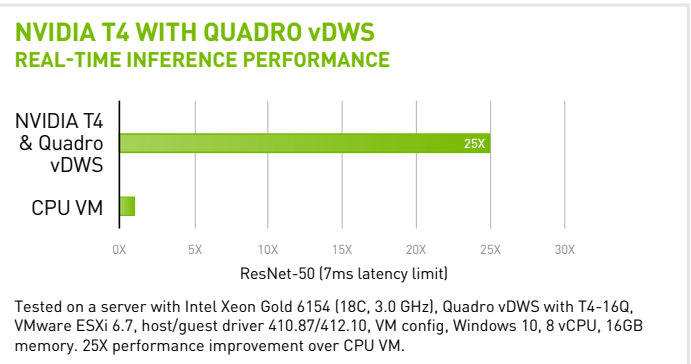
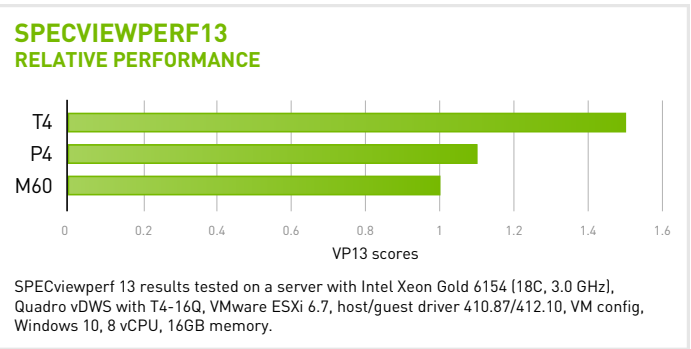
With the T4 and NVIDIA Quadro® Virtual Data Center Workstation (Quadro vDWS) software, virtual workstation users can achieve up to 2X performance (compared to the M60) and access a 16GB framebuffer (double the Tesla P4), which means they can work on larger models and achieve their best designs, faster.

T4 powered Quadro virtual workstations running deep learning inferencing workloads can perform up to 25X faster than a VM driven by a CPU-only server. NVIDIA GPU Cloud (NGC) containers simplify the installation process for IT and reduce the risks of implementing deep learning workloads.

The T4 is an RTX-capable GPU, supporting the enhancements of the RTX platform. When combined with Quadro vDWS, virtual workstations can achieve real-time ray tracing performance of up to 5 Giga Rays per second. Bring creations to market faster with accelerated batch rendering, and speed creative workflows with AI-enhanced denoising. With RTX, artists working in Quadro virtual workstations can create photorealistic designs with accurate shadows, reflections, and refractions and can do so on any device, from anywhere.

## Power Efficient Virtual Desktops

Knowledge workers using modern productivity applications and viewing video on multiple and 4K



monitors require increased graphics support to enjoy a native-PC user experience in a virtual environment. With its 16GB framebuffer and compact form factor, two T4 GPUs can deliver the same user density as a single Tesla M10 with a 32GB framebuffer. Because the T4 consumes only 70W of power, it can deliver this density with lower power consumption. The T4 also supports VP9 video decode and H.265 (HVEC) encode/decode and provides more than 1TB of system memory. For data centers tasked with running a wide range of graphics and compute intensive workloads, the flexibility of T4 makes it an ideal solution.

To learn more about NVIDIA virtual GPU technology, visit [www.nvidia.com/virtualgpu](http://www.nvidia.com/virtualgpu)