NVIDIA TESLA GPUs FOR VIRTUALIZATION

NVIDIA virtual GPU (vGPU) software enables delivery of graphics-rich virtual desktops and workstations accelerated by NVIDIA® Tesla® GPUs, the most powerful data center GPUs on the market today. With NVIDIA virtual GPU software, Tesla GPU resources can be divided so the GPUs are shared across multiple virtual machines, or multiple Tesla GPUs can be allocated to a single virtual machine to power the most demanding workflows. NVIDIA virtual GPU software runs on Tesla GPUs based on NVIDIA Turing™, Volta™, Pascal™, and Maxwell™ architectures.

Choose the right virtual GPU software edition for your use case:

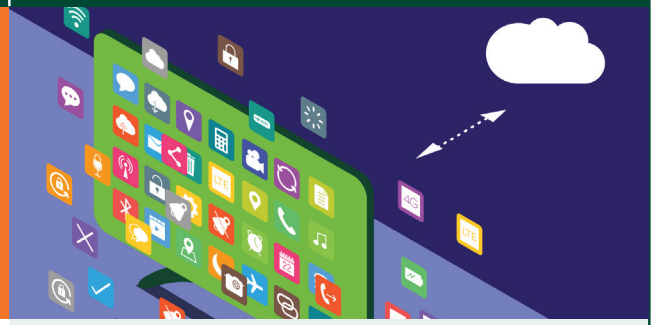
NVIDIA QUADRO® vDWS

NVIDIA Quadro Virtual Data Center Workstation (Quadro vDWS) is targeted for designers, architects, engineers, and artists. When paired with a powerful Tesla GPU, users can virtualize any application from the data center with an amazing user experience—including ANSYS Discovery Live, ESRI ArcGIS Pro, Siemens NX, Dassault Systèmes SOLIDWORKS, Autodesk Revit, and more—allowing you to deliver workstation-class performance on any device.

NVIDIA GRID® vPC

NVIDIA GRID Virtual PC (GRID vPC) targets mobile professionals and knowledge workers running virtual desktops optimized for Windows 10 and office applications. Software developers can also enjoy a modern software development environment, using 2D electronic design automation (EDA) tools and Linux applications. Healthcare providers and financial traders also benefit from increased productivity with multiple high-resolution monitor support.

NVIDIA GRID® vAPPS



NVIDIA GRID Virtual Apps (GRID vApps) is used to launch applications on any device without having to present a full, virtualized desktop to a user. Remote desktop session host (RDSH) solutions can be paired with a more powerful GPU to run more graphics-intensive applications or paired with a less powerful GPU to run general-purpose applications and have more users share a virtual machine.

NVIDIA Tesla GPUs for Virtualization

	TESLA V100	TESLA P100	TESLA P40	TESLA T4	TESLA P4	TESLA P6	M10	M60
								
GPU	1 NVIDIA Volta	1 NVIDIA Pascal	1 NVIDIA Pascal	1 NVIDIA Turing	1 NVIDIA Pascal	1 NVIDIA Pascal	4 NVIDIA Maxwell	2 NVIDIA Maxwell
CUDA Cores	5,120	3,584	3,840	2,560	2,560	2,048	2,560 (640 per GPU)	4,096 (2,048 per GPU)
Tensor Cores	640	--	--	320	--	--	--	--
RT Cores	--	--	--	40	--	--	--	--
Guaranteed QoS (GPU Scheduler)	✓	✓	✓	✓	✓	✓	--	--
Live Migration			✓	✓	✓	✓	✓	✓
Multi-vGPU	✓	✓	✓	✓	✓	✓	✓	✓
Memory Size	32/16 GB HBM2	16 GB HBM2	24 GB GDDR5	16 GB GDDR6	8 GB GDDR5	16 GB GDDR5	32 GB GDDR5 (8 GB per GPU)	16 GB GDDR5 (8 GB per GPU)
vGPU Profiles	1 GB, 2 GB, 4 GB, 8 GB, 16 GB, 32 GB	1 GB, 2 GB, 4 GB, 8 GB, 16 GB	1 GB, 2 GB, 3 GB, 4 GB, 6 GB, 8 GB, 12 GB, 24 GB	1 GB, 2 GB, 4 GB, 8 GB, 16 GB	1 GB, 2 GB, 4 GB, 8 GB	1 GB, 2 GB, 4 GB, 8 GB, 16 GB	0.5 GB, 1 GB, 2 GB, 4 GB, 8 GB	0.5 GB, 1 GB, 2 GB, 4 GB, 8 GB
Form Factor	PCIe 3.0 dual slot and SXM2 (rack servers)	PCIe 3.0 dual slot (rack servers)	PCIe 3.0 dual slot (rack servers)	PCIe 3.0 single slot (rack servers)	PCIe 3.0 single slot (rack servers)	MXM (blade servers)	PCIe 3.0 dual slot (rack servers)	PCIe 3.0 dual slot (rack servers)
Power	250 W /300 W (SXM2)	250 W	250 W	70 W	75 W	90 W	225 W	240 W / 300 W (225 W option)
Thermal	passive	passive	passive	passive	passive	bare board	passive	active/passive
Use Case	Ultra-high-end rendering, simulation, 3D design with Quadro vDWS; ideal upgrade path for P100	High-end rendering, simulation, 3D design with Quadro vDWS	Mid-range to high-end rendering, 3D design and engineering workflows with Quadro vDWS	Entry-level to high-end 3D design and engineering workflows with Quadro vDWS	Entry-level to mid-range 3D design and engineering workflows with Quadro vDWS; two P4 GPUs are an ideal upgrade path for one M60	For customers requiring GPUs in a blade server form factor; ideal upgrade path for M6	Knowledge workers using modern productivity apps and Windows 10 requiring best density and total cost of ownership (TCO), multi-monitor support with NVIDIA GRID vPC/vApps	Mid-range to high-end 3D design and engineering workflows with Quadro vDWS, supported on a wide range of certified servers

WHAT MAKES NVIDIA VIRTUAL GPUs POWERFUL

EXCEPTIONAL USER EXPERIENCE
Ultimate user experience, with the ability to support both compute and graphics workloads for every vGPU

BEST USER DENSITY
Industry's highest user-density solution with support for up to 24 virtual desktops per physical GPU. Lower TCO with up to eight vGPU profiles for the most flexibility to provision resources to match your users' needs

CONTINUOUS INNOVATION
Regular cadence of new software releases to ensure you stay on top of the latest features and enhancements

PREDICTABLE PERFORMANCE
Consistent performance with guaranteed quality of service, whether on premises or in the cloud

OPTIMAL MANAGEMENT AND MONITORING
End-to-end management and monitoring for real-time insight into GPU performance. Broad partner integrations so you can use the tools you know and love

BROADEST ECOSYSTEM SUPPORT
Support for all major hypervisors. Most extensive portfolio of professional apps certifications with Quadro drivers

NVIDIA