





## NVIDIA GPUs FOR VIRTUALIZATION

NVIDIA virtual GPU software enables powerful GPU performance from the enterprise data center and public/private clouds for any workload, from virtual desktops (VDI) to data science and AI. Installed on a server with an NVIDIA GPU, the NVIDIA vGPU software creates virtual GPUs that can be shared between multiple virtual machines, running on any device, anywhere. This enables the IT department standardized on software-defined, hyper converged infrastructure, to leverage the management and security benefits of virtualization as well as the performance of NVIDIA GPUs required for modern graphics and compute workloads.

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

NVIDIA Virtual Compute Server	NVIDIA RTX Virtual Workstation	NVIDIA Virtual PC	NVIDIA Virtual Apps
			
<p>NVIDIA Virtual Compute Server (vCS) software virtualizes NVIDIA GPUs to accelerate compute-intensive workloads, including over 600 GPU-accelerated applications for AI, deep learning, data science, and high-performance computing (HPC). vCS gives data center admins the ability to manage GPU clusters with standard server virtualization management applications, maximizing GPU utilization and ensuring security.</p>	<p>NVIDIA RTX™ Virtual Workstation (vWS) is targeted for designers, architects, engineers, and artists. When paired with a powerful NVIDIA 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.</p>	<p>NVIDIA Virtual PC (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.</p>	<p>NVIDIA Virtual Apps (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.</p>

# NVIDIA GPUs Recommended for Virtualization

	A100	V100	A40	RTX 8000	RTX 6000	T4	M10
GPU Architecture	1 NVIDIA Ampere	1 NVIDIA Volta	1 NVIDIA Ampere	1 NVIDIA Turing	1 NVIDIA Turing	1 NVIDIA Turing	4 NVIDIA Maxwell
RTX Technology	—	—	✓	✓	✓	✓	—
Guaranteed QoS (GPU Scheduler)	✓	✓	✓	✓	✓	✓	—
Live Migration	✓	✓	✓	✓	✓	✓	✓
Multi-vGPU	✓	✓	✓	✓	✓	✓	✓
Memory Size	40/80GB HBM2	32/16GB HBM2	48GB GDDR6	48GB GDDR6	24GB GDDR6	16GB GDDR6	32GB GDDR5 (8GB per GPU)
vGPU Profiles	4GB, 5GB, 8GB, 10GB, 16GB, 20GB, 40GB, 80GB	1GB <sup>1</sup> , 2GB <sup>1</sup> , 4GB, 8GB, 16GB, 32GB	1GB <sup>1</sup> , 2GB <sup>1</sup> , 3GB <sup>1</sup> , 4GB, 6GB, 8GB, 12GB, 16GB, 24GB, 48	1GB <sup>1</sup> , 2GB <sup>1</sup> , 3GB <sup>1</sup> , 4GB, 6GB, 8GB, 12GB, 16GB, 24GB, 48GB	1GB <sup>1</sup> , 2GB <sup>1</sup> , 3GB <sup>1</sup> , 4GB, 6GB, 8GB, 12GB, 24GB	1GB <sup>1</sup> , 2GB <sup>1</sup> , 4GB, 8GB, 16GB	0.5GB, 1GB, 2GB, 4GB, 8GB
Form Factor	SXM4, and PCIe 4.0 dual slot	PCIe 3.0 dual slot and SXM2	PCIe 4.0 dual slot	PCIe 3.0 dual slot	PCIe 3.0 dual slot	PCIe 3.0 single slot	PCIe 3.0 dual slot
Power	400/250W	250W/300W (SXM2)	300W	250W	250W	70W	225W
Thermal	passive	passive	passive	passive	passive	passive	passive
vGPU Software Support	vCS	vWS, vPC, vApps, vCS	vWS, vPC, vApps, vCS	vWS, vPC, vApps, vCS	vWS, vPC, vApps, vCS	vWS, vPC, vApps, vCS	vWS, vPC, vApps
Use Case	AI, data analytics, and HPC	Ultra-high-end rendering, simulation, 3D design with vWS; compute-intensive AI, deep learning, HPC workloads with vCS	Mid-range to high-end 3D design and creative workflows with vWS; virtualized AI with NVIDIA vCS; upgrade path for RTX 8000, RTX 6000	High-end rendering, 3D design, and creative workflows with vWS; AI and data science workloads with vCS	Mid-range to high-end rendering, 3D design, and creative workflows with vWS	Entry-level to midrange 3D design and engineering workflows with vWS; high-density, low-power GPU acceleration for knowledge workers with vPC software	Knowledge workers using modern productivity apps and Windows 10 requiring best density and total cost of ownership (TCO), multi-monitor support with NVIDIA vPC/vApps

The following NVIDIA GPUs are also supported for virtualization: NVIDIA V100S, RTX A6000, RTX 8000 (active), RTX 6000 (active), P40, and P6.

<sup>1</sup> Not supported by vCS. Minimum profile size supported by vCS is 4GB.

## WHAT MAKES NVIDIA VIRTUAL GPU<sub>s</sub> POWERFUL



### EXCEPTIONAL USER EXPERIENCE

Ultimate user experience, with the ability to support both compute and graphics workloads.



### CONTINUOUS INNOVATION

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



### 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.



### BEST USER DENSITY

Industry's highest user-density solution with support for up to 32 virtual desktops per physical GPU. Lower TCO with more than ten vGPU profiles for the most flexibility to provision resources to match your users' needs.



### PERFORMANCE

Consistent near bare-metal performance, whether on premises or in the cloud.



### BROADEST ECOSYSTEM SUPPORT

Support for all major hypervisors. Most extensive portfolio of professional apps certifications with NVIDIA RTX Enterprise Drivers.

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

© 2021 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, CUDA, NVIDIA RTX and NVIDIA Maxwell, Ampere, Turing, Volta, and Pascal are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. All other trademarks and copyrights are the property of their respective owners. JAN21

