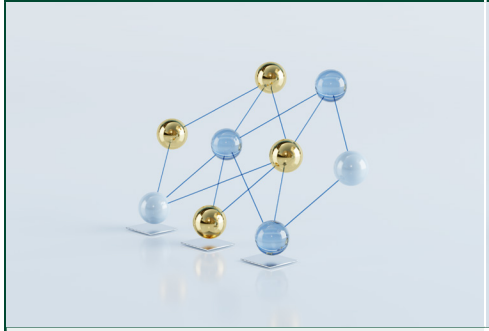
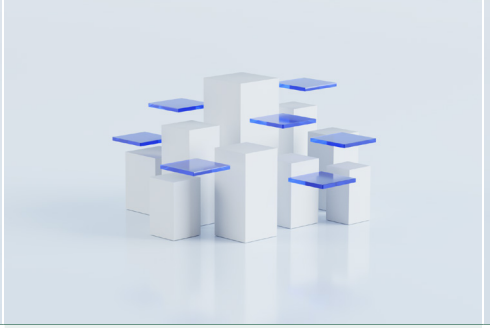
NVIDIA GPUs FOR VIRTUALIZATION

NVIDIA virtual GPU (vGPU) software enables powerful GPU performance from the enterprise data center, as well as public and 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 and hyperconverged infrastructure to leverage the management and security benefits of virtualization –and get the performance of NVIDIA GPUs 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
			
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 security.	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.	NVIDIA Virtual PC (vPC) targets mobile professionals and knowledge workers who run virtual desktops that are optimized for Windows 10 and Microsoft 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 Virtual Apps (vApps) are 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 GPUs Recommended for Virtualization (Ampere Architecture)

	A100	A30	A40	A16
				
GPU Architecture	1 NVIDIA Ampere	1 NVIDIA Ampere	1 NVIDIA Ampere	4 NVIDIA Ampere
RTX Technology	—	—	✓	✓
Guaranteed QoS (GPU Scheduler)	✓	✓	✓	✓
Live Migration	✓	✓	✓	✓
Multi-vGPU	✓	✓	✓	✓
Memory Size	40/80GB HBM2	24GB HBM2	48GB GDDR6	64 GB GDDR6 (16 GB per GPU)
vGPU Profiles	4GB, 5GB, 8GB, 10GB, 16GB, 20GB, 40GB, 80GB	4 GB, 6GB, 8GB, 12GB, 24GB	1GB[1], 2GB[1], 3GB[1], 4GB, 6GB, 8GB, 12GB, 16GB, 24GB, 48GB	1GB[1], 2GB[1], 4GB, 8GB, 16GB
Form Factor	SXM4, and PCIe 4.0 dual slot	PCIe 4.0 dual slot	PCIe 4.0 dual slot	PCIe 4.0 dual slot
Power	400/250W	165W	300W	250W
Thermal	passive	passive	passive	passive
vGPU Software Support	vCS	vCS	vWS, vPC, vApps, vCS	vWS, vPC, vApps, vCS
Use Case	AI, data analytics, and HPC	AI inferencing, data analytics, and HPC	Midrange to high-end 3D design and creative workflows with vWS; upgrade path for RTX 8000, RTX 6000, or T4	Knowledge worker virtual desktops using modern productivity apps and multimedia with NVIDIA vPC/vApps. Entry NVIDIA RTX Virtual Workstations, upgrade path for M10 or T4.

The following NVIDIA GPUs are also supported for virtualization: NVIDIA V100/V100S, RTX A6000, RTX A5000, RTX 8000, RTX 6000, T4, P40, P6, and M10.

[1] Not supported by vCS. Minimum profile size supported by vCS is 4GB.

WHAT MAKES NVIDIA VIRTUAL GPUS POWERFUL

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

BEST USER DENSITY
Industry's highest user-density solution with support for up to 64 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.

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

PERFORMANCE
Consistent near bare-metal performance, 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 NVIDIA RTX Enterprise Drivers.

To learn more about NVIDIA virtual GPU technology, visit **www.nvidia.com/virtualgpu**