

NVIDIA Professional Graphics Solutions

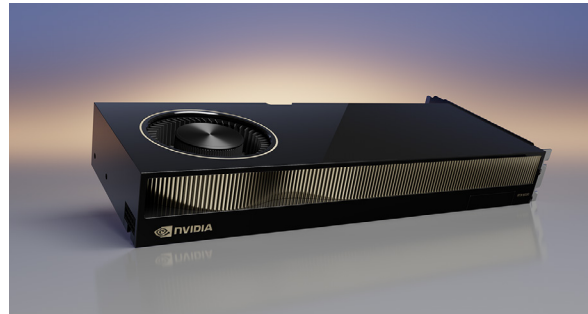
You need to do great things—create and collaborate from anywhere, on any device, without the distractions of slow performance, poor stability, or application incompatibility. With NVIDIA RTX™, you can unleash your vision and enjoy ultimate creative freedom.

NVIDIA RTX professional visualization products power a wide range of laptop, desktop, and data center solutions. Leverage the latest advancements in real-time ray tracing, AI, virtual reality (VR), and interactive, photorealistic rendering to develop revolutionary products, tell vivid visual stories, and design groundbreaking architecture like never before. Support for advanced features, frameworks, and SDKs across all of our products gives you the power to tackle the most challenging visual computing tasks, no matter the scale.



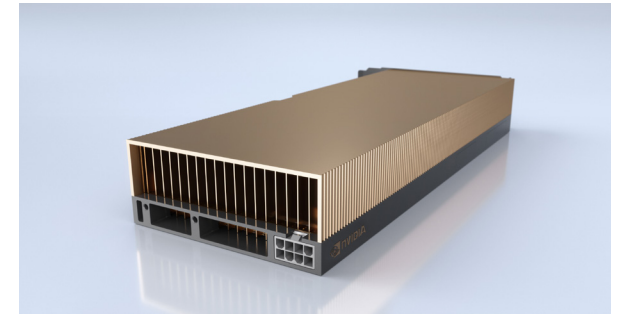
NVIDIA Professional Laptop GPUs

Professionals today increasingly need to work on complex workflows like VR, 8K video editing, and photorealistic rendering on the go. NVIDIA RTX laptop GPUs deliver world-class performance in a portable form factor combining the latest advancements in real-time ray tracing, advanced shading, and AI-based capabilities, so professionals can tackle demanding workflows from anywhere.



NVIDIA Desktop Workstation GPUs

NVIDIA RTX and Quadro-powered desktop workstations are designed and built specifically to drive the most challenging workloads of artists, designers, and engineers. Large GPU memory boosts application performance to accelerate your workflows and tackle the heaviest of workloads. NVIDIA RTX workstation solutions deliver significant business impact for demanding industries like manufacturing, media and entertainment, and energy.



NVIDIA Data Center GPUs

Demand for visualization, rendering, data science, and simulation continues to grow as businesses tackle larger, more complex workloads. Scale up your visual compute infrastructure and tackle graphics-intensive workloads, complex designs, photorealistic renders, and augmented and virtual environments at the edge with NVIDIA GPUs. Optimized for reliability in enterprise data centers, NVIDIA GPUs feature both active and passive thermal solutions to fit into a variety of servers.

NVIDIA Professional Graphics Solutions

New	GPU Specifications							Performance			Display Technology				Virtual Reality		Options				
	NVIDIA CUDA® Processing Cores ¹	NVIDIA RT Cores	Tensor Cores	GPU Memory	Peak Memory Bandwidth	NVIDIA® NVLink®	Floating-Point Performance, Single Precision (TFLOPS, Peak) ²	Tensor Performance (TFLOPS, Peak) ³	Error-Correcting Code (ECC) Memory	Maximum Active Displays	DisplayPort 1.4 ⁴	HDMI via Adaptors, HDMI	NVIDIA SLI ⁵	High-Dynamic Range (HDR) ⁶	NVIDIA Mosaic Technology	VR Ready ⁷	Variable Rate Shading	GPUDirect® for Video	Graphics Synchronization with Sync II	3D Stereo ⁸	Encode/Decode ⁹
Laptop GPUs																					
•	NVIDIA RTX 5000 Ada Generation	9,728	76 (3rd Gen)	304 (4th Gen)	16GB	576GB/s		42.6	681.8	• ⁹	4*	Yes*	Yes*		•	•	•	•	•	•	•
•	NVIDIA RTX 4000 Ada Generation	7,424	58 (3rd Gen)	232 (4th Gen)	12GB	432GB/s		33.6	538.0	• ⁹	4*	Yes*	Yes*		•	•	•	•	•	•	•
•	NVIDIA RTX 3500 Ada Generation	5,120	40 (3rd Gen)	160 (4th Gen)	12GB	432GB/s		23.0	368.6	• ⁹	4*	Yes*	Yes*		•	•	•	•	•	•	•
•	NVIDIA RTX 3000 Ada Generation	4,608	36 (3rd Gen)	144 (4th Gen)	8GB	256GB/s		19.9	318.6	• ⁹	4*	Yes*	Yes*		•	•	•	•	•	•	•
•	NVIDIA RTX 2000 Ada Generation	3,072	24 (3rd Gen)	96 (4th Gen)	8GB	256GB/s		14.5	231.6		4*	Yes*	Yes*		•	•	•	•	•	•	•
	NVIDIA RTX A5500	7,424	58 (2nd Gen)	232 (3rd Gen)	16GB	512GB/s		24.7	197.8	• ⁹	4*	Yes*	Yes*		•	•	• ¹⁰	•	•	•	•
	NVIDIA RTX A4500	5,888	46 (2nd Gen)	184 (3rd Gen)	16GB	512GB/s		18.5	148.4	• ⁹	4*	Yes*	Yes*		•	•	•	•	•	•	•
	NVIDIA RTX A3000 12GB	4,096	32 (2nd Gen)	128 (3rd Gen)	12GB	336GB/s		14.1	113.0	• ⁹	4*	Yes*	Yes*		•	•	•	•	•	•	•
	NVIDIA RTX A2000 8GB	2,560	20 (2nd Gen)	80 (3rd Gen)	8GB	224GB/s		9.3	74.3		4*	Yes*	Yes*		•	•	• ¹⁰	•	•	•	•
•	NVIDIA RTX A1000 6GB	2,560	20 (2nd Gen)	80 (3rd Gen)	6GB	168GB/s		9.3	74.6		4*	Yes*	Yes*		•	•	• ¹⁰	•	•	•	•
	NVIDIA RTX A1000	2,048	16 (2nd Gen)	64 (3rd Gen)	4GB	224GB/s		7.5	59.7		4*	Yes*	Yes*		•	•	• ¹⁰	•	•	•	•
	NVIDIA RTX A500	2,048	16 (2nd Gen)	64 (3rd Gen)	4GB	112GB/s		7.0	56.0		4*	Yes*	Yes*		•	•	•	•	•	•	•
	NVIDIA T600	896			4GB	192GB/s		3.0			4*	Yes*	Yes*		•	•				•	•
	NVIDIA T550	1,024			4GB	112GB/s		3.7												•	•
Desktop GPUs																					
•	NVIDIA RTX 6000 Ada Generation	18,176	142 (3rd Gen)	568 (4th Gen)	48GB	960 GB/s		91.1	1,457.0	• ⁹	4	4	4		•	•	•	•	•	•	•
•	NVIDIA RTX 5000 Ada Generation	12,800	100 (3rd Gen)	400 (4th Gen)	32GB	576GB/s		65.3	1,044.4	• ⁹	4	4	4		•	•	•	•	•	•	•
•	NVIDIA RTX 4500 Ada Generation	7,680	60 (3rd Gen)	240 (4th Gen)	24GB	432GB/s		39.9	637.8	• ⁹	4	4	4		•	•	•	•	•	•	•
•	NVIDIA RTX 4000 Ada Generation	6,144	48 (3rd Gen)	192 (4th Gen)	20GB	360GB/s		26.7	427.6	• ⁹	4	4	4		•	•	•	•	•	•	•
•	NVIDIA RTX 4000 SFF Ada Generation	6,144	48 (3rd Gen)	192 (4th Gen)	20GB	280GB/s		19.2	306.8	• ⁹	4	4	4		•	•	•	•	•	•	•
	NVIDIA RTX A6000	10,752	84 (2nd Gen)	336 (3rd Gen)	48GB	768GB/s	•	38.7	309.7	• ⁹	4	4	4	•	•	•	•	•	•	•	•
	NVIDIA RTX A5500	10,240	80 (2nd Gen)	320 (3rd Gen)	24GB	768GB/s	•	34.1	272.8	• ⁹	4	4	4	•	•	•	•	•	•	•	•
	NVIDIA RTX A5000	8,192	64 (2nd Gen)	256 (3rd Gen)	24GB	768GB/s	•	27.8	222.2	• ⁹	4	4	4	•	•	•	•	•	•	•	•
	NVIDIA RTX A4500	7,168	56 (2nd Gen)	224 (3rd Gen)	20GB	640GB/s	•	23.7	189.2	• ⁹	4	4	4	•	•	•	•	•	•	•	•
	NVIDIA RTX A4000	6,144	48 (2nd Gen)	192 (3rd Gen)	16GB	448GB/s		19.2	153.4	• ⁹	4	4	4		•	•	•	•	•	•	•
	NVIDIA RTX A2000 12GB	3,328	26 (2nd Gen)	104 (3rd Gen)	12GB	288GB/s		7.9	63.9	• ⁹	4	4	4		•	•	•	•	•	•	•
	NVIDIA T1000 NVIDIA T1000 8GB	896			4GB or 8GB	160GB/s		2.5			4	4	4		•	•				•	•
	NVIDIA T400 4GB	384			4GB	80GB/s		1.0			4 ¹¹	3	3		•	•				•	•
Data Center GPUs																					
•	NVIDIA L40S	18,176	142 (3rd Gen)	568 (4th Gen)	48GB	864GB/s		91.6	1,466.0	• ⁹	4	4	4	•	•	•	•	•	•	•	•
•	NVIDIA L40	18,176	142 (3rd Gen)	568 (4th Gen)	48GB	864GB/s		90.5	724.1	• ⁹	4	4	4	•	•	•	•	•	•	•	•
	NVIDIA A40	10,752	84	336	48GB	696GB/s	•	37.4	299.4	• ⁹	4	3	3	•	•	•	•	•	•	•	•
	NVIDIA A10	9,216	72	288	24GB	600GB/s		31.2	249.9	• ⁹					•	•				•	•
	NVIDIA A16	4x 1,280	4x 10	4x 40	4x 16GB	4x 232GB/s		4x 4.5	4x 17.8	• ⁹					•	•				•	•

* Check with OEM manufacturer for specific display topology.

- CUDA parallel processing cores cannot be compared between GPU generations due to several important architectural differences that exist between streaming multiprocessor designs.
- Peak rates are based on GPU Boost clock.
- Effective TFLOPS using the sparsity feature. NVIDIA Ada Lovelace architecture using FP8 matrix multiply with FP16 or FP32 accumulate; NVIDIA Ampere architecture using FP16 matrix multiply with FP16 or FP32 accumulate.
- Feature support varies by system-level implementation. Check with your workstation OEM vendor for system specific configurations. Feature support varies by system-level implementation. Check with your workstation OEM vendor for system specific configurations.
- SLI functionality is provided via NVLink.
- Supported adaptors are required for HDMI.
- Supports multi-view rendering (MVR) feature.
- For more details on GPU-specific video encode/decode format support, refer to: <https://developer.nvidia.com/video-encode-and-decode-gpu-support-matrix-new>
- Ensures data integrity and reliability by eliminating soft errors on direct random-access memory (DRAM) only.
- Support for configuration at 60W TGP and above. NVIDIA RTX A1000 and RTX A2000 8GB Laptop GPUs, VR-Ready at 60W TGP and above.
- NVIDIA T400 4GB desktop GPUs can drive four displays via multi-stream transport (MST).

For more information on NVIDIA professional graphics solutions, visit: www.nvidia.com/en-us/design-visualization/rtx/

© 2023 NVIDIA Corporation. All rights reserved. NVIDIA and the NVIDIA logo, CUDA, GPUDirect, NVLink, , RTX, SLI, Turing, Ampere, and Ada Generation are trademarks or registered trademarks of the respective owners with which they are associated. Features, pricing, availability, and specifications are subject to change without notice. 2869400. AUG23

