

AI HARDWARE ATLAS

CIPHER LACUNA

NVIDIA + AMD / 53 products / Updated 2026-09-26

A = announced / launched. R = availability or vendor target. Blank dates are not established.

GeForce RTX 5070 NVIDIA 12 GB GDDR7 / Per GPU / 0.672 TB/s A: 2025-01-06 R: 2025-03-05 (Released)	GeForce RTX 3080 10GB NVIDIA 10 GB GDDR6X / Per GPU / 0.76 TB/s A: 2020-09-01 R: 2020-09-17 (Released)	DGX Station (GB300) NVIDIA 784 GB coherent total / Shared CPU + GPU / 7.1 TB/s A: 2025-03-18 R: -
GeForce RTX 5070 Ti NVIDIA 16 GB GDDR7 / Per GPU / 0.896 TB/s A: 2025-01-06 R: 2025-02-20 (Released)	GeForce RTX 3070 Ti NVIDIA 8 GB GDDR6X / Per GPU A: 2021-05-31 R: 2021-06-10 (Released)	RTX PRO 6000 Server Edition NVIDIA 96 GB GDDR7 ECC / Per GPU / 1.597 TB/s A: 2025-03-18 R: -
GeForce RTX 5080 NVIDIA 16 GB GDDR7 / Per GPU / 0.96 TB/s A: 2025-01-06 R: 2025-01-30 (Released)	GeForce RTX 3070 NVIDIA 8 GB GDDR6 / Per GPU / 0.448 TB/s A: 2020-09-01 R: 2020-10-29 (Released)	RTX PRO Server (8-GPU example) NVIDIA 8 x 96 GB GPU memory / System total / 13 TB/s A: 2025-05-18 R: -
GeForce RTX 5090 NVIDIA 32 GB GDDR7 / Per GPU / 1.792 TB/s A: 2025-01-06 R: 2025-01-30 (Released)	GeForce RTX 3060 Ti NVIDIA 8 GB GDDR6 / 8 GB GDDR6X / Per GPU A: 2020-12-01 R: 2020-12-02 (Released)	Instinct MI350P PCIe AMD 144 GB HBM3E / Per GPU / 4 TB/s A: 2026-07-23 R: -
GeForce RTX 4090 NVIDIA 24 GB GDDR6X / Per GPU / 1.008 TB/s A: 2022-09-20 R: 2022-10-12 (Released)	GeForce RTX 3060 12GB NVIDIA 12 GB GDDR6 / Per GPU A: 2021-01-12 R: 2021-02-25 (Released)	H200 NVIDIA 141 GB HBM3e / Per GPU / 4.8 TB/s A: - R: 2024-Q2 (Vendor availability window)
GeForce RTX 4080 SUPER NVIDIA 16 GB GDDR6X / Per GPU / 0.736 TB/s A: 2024-01-08 R: 2024-01-31 (Released)	GeForce RTX 3060 8GB NVIDIA 8 GB GDDR6 / Per GPU A: - R: -	B200 Blackwell NVIDIA 180 GB HBM3e / Per GPU / 8 TB/s A: 2024-03-18 R: -
GeForce RTX 4080 NVIDIA 16 GB GDDR6X / Per GPU / 0.7168 TB/s A: 2022-09-20 R: 2022-11-16 (Released)	GeForce RTX 3050 8GB NVIDIA 8 GB GDDR6 / Per GPU A: 2022-01-04 R: 2022-01-27 (Released)	DGX B200 (8 GPUs) NVIDIA 8 x 180 GB HBM3e / System total / 64 TB/s A: 2024-03-18 R: -
GeForce RTX 4070 Ti SUPER NVIDIA 16 GB GDDR6X / Per GPU / 0.672 TB/s A: 2024-01-08 R: 2024-01-24 (Released)	GeForce RTX 3050 6GB NVIDIA 6 GB GDDR6 / Per GPU A: 2024-02-02 R: 2024-02-02 (Released)	B300 Blackwell Ultra NVIDIA 288 GB HBM3e / Per GPU / 8 TB/s A: 2025-03-18 R: -
GeForce RTX 4070 Ti NVIDIA 12 GB GDDR6X / Per GPU / 0.504 TB/s A: 2023-01-03 R: 2023-01-05 (Released)	Radeon RX 9070 AMD 16 GB GDDR6 / Per GPU / 0.64 TB/s A: 2025-02-28 R: 2025-03-06 (Released)	Instinct MI300X AMD 192 GB HBM3 / Per GPU / 5.3 TB/s A: 2023-12-06 R: -
GeForce RTX 4070 SUPER NVIDIA 12 GB GDDR6X / Per GPU / 0.504 TB/s A: 2024-01-08 R: 2024-01-17 (Released)	Radeon RX 9070 XT AMD 16 GB GDDR6 / Per GPU / 0.64 TB/s A: 2025-02-28 R: 2025-03-06 (Released)	Instinct MI325X AMD 256 GB HBM3E / Per GPU / 6 TB/s A: 2024-10-10 R: -
GeForce RTX 4070 NVIDIA 12 GB GDDR6 / 12 GB GDDR6X / Per GPU / 0.504 TB/s A: 2023-04-12 R: 2023-04-13 (Released)	DGX Spark (GB10) NVIDIA 128 GB unified / Shared CPU + GPU / 0.273 TB/s A: 2025-03-18 R: 2025-10-15 (Orders opened)	Instinct MI350X AMD 288 GB HBM3E / Per GPU / 8 TB/s A: 2025-06-12 R: -
GeForce RTX 4060 Ti 16GB NVIDIA 16 GB GDDR6 / Per GPU / 0.288 TB/s A: 2023-05-18 R: 2023-07 (Released)	Ryzen AI Max+ PRO 395 AMD Up to 128 GB unified* / Shared CPU + GPU / 0.256 TB/s A: 2025-01-06 R: 2025-Q1 (Vendor availability window)	Instinct MI355X AMD 288 GB HBM3E / Per GPU / 8 TB/s A: 2025-06-12 R: -
GeForce RTX 4060 Ti 8GB NVIDIA 8 GB GDDR6 / Per GPU / 0.288 TB/s A: 2023-05-18 R: 2023-05-24 (Released)	RTX PRO 4500 Blackwell NVIDIA 32 GB GDDR7 ECC / Per GPU / 0.896 TB/s A: 2025-03-18 R: 2025-Summer (Vendor target (summer 2025))	GB300 NVL72 NVIDIA 72 GPUs / ~20 TB HBM / Rack total / 576 TB/s A: - R: 2025-H2 (Partner rollout)
GeForce RTX 4060 NVIDIA 8 GB GDDR6 / Per GPU / 0.272 TB/s A: 2023-05-18 R: 2023-06-29 (Released)	RTX PRO 5000 Blackwell NVIDIA 48 GB GDDR7 ECC / Per GPU / 1.344 TB/s A: 2025-03-18 R: 2025-Summer (Vendor target (summer 2025))	Vera Rubin NVL72 NVIDIA 20,736 GB HBM4 total (72 x 288 GB) / Rack total / 1400 TB/s A: - R: 2026-H2 (Partner rollout expected)
GeForce RTX 3090 Ti NVIDIA 24 GB GDDR6X / Per GPU / 1.008 TB/s A: 2022-01-04 R: 2022-03-29 (Released)	RTX PRO 5000 Blackwell 72GB NVIDIA 72 GB GDDR7 ECC / Per GPU / 1.344 TB/s A: 2025-12-18 R: 2025-12-18 (Generally available)	Vera Rubin NVL144 CPX NVIDIA 100 TB fast memory / rack / Rack total / 1700 TB/s A: 2025-09-09 R: - (Not on NVIDIA's current roadmap)
GeForce RTX 3090 NVIDIA 24 GB GDDR6X / Per GPU / 0.936 TB/s A: 2020-09-01 R: 2020-09-24 (Released)	RTX PRO 6000 Blackwell NVIDIA 96 GB GDDR7 ECC / Per GPU / 1.792 TB/s A: 2025-03-18 R: 2025-04 (Vendor distribution window)	Instinct MI455X AMD 432 GB HBM4 / Per GPU / 23.3 TB/s A: 2026-07-23 R: -
GeForce RTX 3080 Ti NVIDIA 12 GB GDDR6X / Per GPU A: 2021-05-31 R: 2021-06-03 (Released)	Radeon AI PRO R9700 AMD 32 GB GDDR6 / Per GPU / 0.64 TB/s A: 2025-05-20 R: 2025-Q3 (Standalone cards)	Helios (MI455X) AMD 31,104 GB HBM4 total (72 x 432 GB) / Rack total / 1670 TB/s A: - R: 2026-H2 (Volume expected)
GeForce RTX 3080 12GB NVIDIA 12 GB GDDR6X / Per GPU A: 2022-01 R: 2022-01 (Released)	Radeon PRO W7900 AMD 48 GB GDDR6 ECC / Per GPU / 0.864 TB/s A: 2023-04-13 R: -	

COMPARE LIKE FOR LIKE

Dedicated GPU VRAM, shared system memory and rack totals have different meanings. Capacity is not speed.
Each row links to an official source. The website includes power, architecture, detailed notes and all sources.