



NVIDIA DGX H200

Also known as: DGX H200

DGX H200 is NVIDIA's factory-integrated Hopper-generation system: eight H200 SXM GPUs (141 GB HBM3e each) on NVLink 4, dual Xeon hosts, and ConnectX-7 fabric in an 8U chassis with NVIDIA-direct support. It is the prior-generation counterpart to DGX B200, sharing the H100 chassis and board layout but with the higher-capacity, higher-bandwidth H200 GPU swapped in. Buy DGX H200 for a validated, single-vendor-supported node, particularly where an existing H100/H200 DGX or HGX fleet needs a drop-in match; buy an HGX H200 OEM server (Supermicro, Dell, HPE, Lenovo) when you want to control the CPU, storage or networking BOM or already run your own cluster ops.

8x H200 SXM GPUs · Hopper, NVLink 4	1,128 GB HBM3e Total GPU memory	~10.2 kW System max power · 200-240V input
8U rackmount Form factor	8x ConnectX-7 Networking · up to 400 Gb/s InfiniBand per port	287.6 lb (130.45 kg) max Weight

COMPUTE		01/05
GPUs	8x NVIDIA H200 SXM5 · Hopper architecture, 141 GB HBM3e each	
GPU-to-GPU fabric	NVLink 4 · 900 GB/s per GPU bidirectional	
CPUs	2x Intel Xeon Platinum 8480C · 56 cores each	
MEMORY & STORAGE		02/05
Total GPU memory	1,128 GB HBM3e	
System memory	2 TB · 32 DIMMs	
OS storage	2x 1.92 TB NVMe M.2 · RAID 1	
Data cache storage	8x 3.84 TB NVMe U.2 SED · RAID 0, ~30.7 TB raw	
NETWORKING		03/05
Compute fabric NICs	8x ConnectX-7 single-port · 4 OSFP ports, up to 400 Gb/s InfiniBand	
In-band management/storage	dual-port Ethernet cards	

POWER & COOLING		04/05
System max power	~10.2 kW · reported across multiple sources	
Power supplies	6x 3.3 kW · 4+2 redundant, 200-240V each	
Cooling	Air-cooled · front-to-back	
Airflow	1,105 CFM · at 80% fan PWM	
Heat output	38,557 BTU/hr	
Operating temperature	5 C to 30 C (41 F to 86 F)	
PHYSICAL		05/05
Form factor	8U rackmount	
Height	14 in	
Width	19 in	
Depth	35.3 in	
Weight	287.6 lb (130.45 kg) max	
PROCUREMENT		
Typical lead time		
8 to 16 weeks from order to delivery, reported, varies by region and demand.		
Pricing		
\$400,000 to \$550,000 per unit, reported and street-sourced (one reseller lists \$550,000 flat; other market surveys cite the \$400,000 to \$500,000 range for complete 8-GPU systems). No public NVIDIA list price exists.		
Price is quote-only through NVIDIA's partner network; expect a formal quote cycle rather than a checkout price. H200 street pricing has been softening through 2026 as Blackwell (B200/B300) supply ramps and takes the top-end workload share, per reseller and market commentary. The DGX premium over an equivalent HGX H200 OEM build buys NVIDIA-direct support and validated firmware, not extra GPU throughput; teams with existing OEM support contracts often land the same GPUs cheaper via HGX. H200 is electrically and mechanically a drop-in for existing H100 DGX/HGX racks and PDUs, which shortens integration time versus a first Blackwell deployment.		

FIELD NOTES

- At ~10.2 kW max, one DGX H200 still exceeds a standard 5 to 8 kW colo rack's power allocation on its own; plan high-density power delivery even though it draws noticeably less than a DGX B200.
- The H200's 8U chassis is shorter than the B200's 10U but the weight is close (287.6 lb vs 313.9 lb), so rack floor loading and lift equipment planning don't get meaningfully easier by choosing H200 over B200.
- Because H200 shares the H100 chassis and 4+2 PSU layout, an existing H100 DGX SuperPOD rack can usually absorb H200 nodes without a facilities redesign, which is the main reason to pick H200 over waiting for or refitting to Blackwell.
- 141 GB per GPU (vs 80 GB on H100) mainly buys headroom for larger KV caches and batch sizes in inference; if the workload is training-bound and memory capacity isn't the bottleneck, the H200 premium over H100 is often not worth it versus deploying B200 instead if budget allows.
- If your fleet is already HGX-based and OEM-supported, an HGX H200 server from your existing vendor is usually the lower-friction and lower-cost path to the same GPU; reserve DGX H200 for teams that need NVIDIA's own support SLA or SuperPOD reference-architecture compliance.

QUESTIONS WE GET ON THIS PART

How much does a DGX H200 cost?

NVIDIA does not publish a list price. Reported street pricing runs roughly \$400,000 to \$550,000 for a complete 8-GPU system, with one reseller listing it flat at \$550,000. Expect a formal quote process and regional/volume variation.

DGX H200 vs DGX B200: which should we buy?

H200 (Hopper) draws less power (~10.2 kW vs ~14.3 kW), fits the same 8U/H100-era rack infrastructure, and costs less. B200 (Blackwell) delivers materially higher training and inference throughput via NVLink 5 and newer tensor cores, at a higher power and price point. Choose H200 for a lower-risk fit into existing H100 infrastructure or budget constraints, B200 for new deployments chasing peak performance per GPU.

Can a DGX H200 use existing DGX H100 rack power and cooling?

Largely yes. The H200 keeps the H100's 8U chassis, 6x 3.3 kW PSU layout (4+2 redundant), and similar airflow profile (1,105 CFM), so an H100-provisioned rack typically does not need a facilities redesign to host H200 nodes.

Is DGX H200 still worth buying with Blackwell available?

Yes for specific cases: it is still shipping as of mid-2026, priced below DGX B200, and is a drop-in for existing H100/H200 racks. It makes less sense for a greenfield deployment with no legacy infrastructure and no power constraint, where B200's higher throughput per dollar of power usually wins.

REFERENCES

- [1] <https://docs.nvidia.com/dgx/dgxm100-user-guide/introduction-to-dgxm100.html>
- [2] <https://itctshop.com/product/nvidia-dgx-h200/>
- [3] <https://medium.com/@UVATION/nvidia-dgx-h200-power-consumption-key-facts-requirements-a1793249f95f>
- [4] <https://www.sunbirdcdm.com/blog/nvidia-h200-power-requirements-can-your-racks-support-them>
- [5] <https://intuitionlabs.ai/articles/nvidia-ai-gpu-pricing-guide>
- [6] <https://www.nvidia.com/en-us/data-center/h200/>
- [7] <https://www.trgdatacenters.com/resource/nvidia-dgx-buyers-guide-everything-you-need-to-know/>
- [8] <https://docs.nvidia.com/dgx/dgxm100-user-guide/power-capping.html>

Latest version of this datasheet: <https://gpusmith.com/hardware/systems/nvidia-dgx-h200>. Specifications compiled from the references above; verify against vendor documentation before purchase.