

# H200 Graphics Card AI Server



## Overview

DGX H200 is a fully integrated hardware and software solution on which to build your AI Center of Excellence. It includes NVIDIA Base Command™ and the NVIDIA AI Enterprise software suite, plus expert advice from NVIDIA DGXperts. The NVIDIA H200 GPU supercharges generative AI and high-performance computing (HPC) workloads with game-changing performance and memory capabilities. Where the H100 tops out at 80 GB of HBM3, the H200 jumps to 141 GB of HBM3e with 4. It is designed specifically for workloads. The NVIDIA H200 GPU, fully leveraged through Nor-Tech's AI and HPC server integration expertise, enables researchers, engineers, and business decision makers to take advantage of GPU-optimized simulation software that can be deployed on NVIDIA GPU-accelerated clusters interconnected with NVIDIA. The NVIDIA H200 is the latest evolution in AI and HPC computing, offering unmatched performance for deep learning, large-scale AI model training, and high-performance computing (HPC). With HBM3e memory, enhanced Tensor Cores, and NVLink support, the H200 accelerates AI-driven workloads across. For instance, a fully-equipped DGX system with eight H200 GPUs is quoted at roughly \$400K-\$500K (). In contrast, entry-level accelerators (e. This large price spread reflects differences in architecture, memory. The Aivres KR6288-X2/E2 server is purpose-built for demanding AI and HPC workloads, delivering the performance, reliability, and scalability required for large-scale training, inference, and data-intensive scientific computing.

## Article Content

Aivres: NVIDIA HGX H200 GPU Server

The NVIDIA H200 is built for large-scale AI training, enabling faster convergence and higher throughput. Teams can train bigger models, shorten cycles, and achieve state-of-the-art accuracy across

Rent NVIDIA H200 GPU Servers - AI & LLM Inference

Rent NVIDIA H200 GPU Servers with 141GB HBM3e memory for LLMs, AI inference, and HPC. Available in Germany, France, Canada & Netherlands.

NVIDIA Data Center GPU Specs: A Complete

Introduction and Background Graphical processors have transcended their gaming origins to become the workhorses of the AI and HPC era. NVIDIA pioneered this

China reportedly approves imports of Nvidia's H200

Quizzes Hardware Graphics Cards China reportedly approves imports of Nvidia's H200 GPU, but the US government may cap exports to

NVIDIA DGX H200 Power Consumption: Key Facts

Planning for an NVIDIA DGX H200? This deep dive into its power consumption, thermal output, and infrastructure requirements will help you

NVIDIA Data Center GPUs Explained: From A100 to

Understand NVIDIA data center GPUs for AI inference. Compare T4, L4, A100, H100, H200, and B200 on use cases, memory, and pricing to choose

NVIDIA GeForce RTX 3090 vs NVIDIA H200 NVL

AI GPU We compared a Desktop platform GPU: 24GB VRAM GeForce RTX 3090 and a GPU: 141GB VRAM H200 NVL to see which GPU has better performance in key specifications, benchmark tests,

Top 12 NVIDIA GPUs for AI Training & Inference in 2026

Compare the top 12 NVIDIA GPUs for AI in 2026, including H100, H200, B200, GB200, and RTX cards for training, inference, and LLM workloads

900-21010-0040-000

NVIDIA H200 NVL PCIE GPU 141GB HBM3E 4.8TB/S MEMORY BANDWIDTH PCI-E 5.0 X16 GRAPHICS PROCESSING UNIT VIDEO CARD -

AI & HPC Servers with NVIDIA H200 GPUs

The Nor-Tech optimized NVIDIA H200 GPU is for HPC & AI simulation that run the broadest set of scientific simulations (3,000 accelerated apps) and machine learning networks with proven year-over

GPU Servers For AI, Deep / Machine Learning & HPC

Dive into Supermicro's GPU-accelerated servers, specifically engineered for AI, Machine Learning, and High-Performance Computing.

↳ GPU Server in 2026: L40S, A30, RTX Pro or H100/H200?

Learn when A30, L40S or RTX Pro is enough for inference, rendering, VDI and AI workloads, and when H100/H200 is worth the extra cost. We explain how to evaluate VRAM, power,

GPU Requirements 2026: Llama 4 = 1× H100 (\$2.50/hr), DeepSeek

GPU Requirements 2026: Llama 4 = 1× H100 (\$2.50/hr), DeepSeek V3.2 = 8× H200 (\$36/hr), Qwen 3.5 27B = 1× H100 Back to Blog Written by Mitrasish, Co-founder May 14, 2026 GPU

China's Largest Retailer, JD, Is Selling Banned NVIDIA RTX

NVIDIA's "Banned" consumer graphics cards such as the RTX 5090 & RTX PRO 6000 Blackwell are being sold at China's largest retailer.

Original NEW NVIDIA H200 141GB NVL HBM3e PCIE

141GB HBM3e. □□Nvidia Tesla A100 40GB PCIE SXM4 GPU Deep Learning AI 900-21001-0000-000. Nvidia Tesla V100 32GB SXM2 PCIE GPU

NVIDIA DGX H200

DGX H200 sets a new bar for compute density, packing 32 petaFLOPS of AI performance into a 8U form factor, replacing legacy compute infrastructure with

How China's AI chips stack up against Nvidia's H200

China now has a number of AI chipmakers, as the country looks to wean itself off reliance on foreign technology. Here is how Chinese rivals

NVIDIA AI GPU Prices: H100 (\$27K-\$40K) & H200

High sticker prices: Advanced AI GPUs can cost tens of thousands of dollars each; e.g. an 8×H100 server board was estimated at \$216K (). The H200 (an

NVIDIA AI GPU Pricing: A Guide to H100 & H200 Costs

In short, NVIDIA hardware underpins most advanced AI training today. However, this leadership comes at a steep cost. Cutting-edge GPUs (e.g. Hopper-based H100, Blackwell-based H200) have very

Nvidia H200 GPU: Specs, VRAM, Price, and AI

This guide covers the full H200 specs, VRAM capacity and bandwidth, SXM vs NVL variant differences, current pricing, how it compares to

NVIDIA H200 GPU Servers | AI & High-Performance Computing

Crystal Cloud provides fully integrated NVIDIA H200 GPU servers, available for leasing and hosting. Our Supermicro and Lenovo servers, configured with AMD or Intel CPUs, deliver scalable, high-efficiency

Trump: Nvidia Can Sell H200 AI Chips to China, But

The US government has approved the sale of Nvidia's H200 AI graphics chip systems to China, with certain stipulations. "I have informed

7 Best GPU for LLM in 2026 (Including Local LLM Setups)

Best GPU for LLM workloads explained with a practical framework covering VRAM, context length, and how Fluence reduces total cost.

## Contact Us

For more information, pricing, or custom solutions, please contact us:

Website: <https://www.saastisfy.fr>

Email: [sales@saastisfy.fr](mailto:sales@saastisfy.fr)

Phone: +33 6 52 81 47 39

Address: 75 Rue de Rivoli, 75001 Paris, France

This document is for informational purposes only. Specifications subject to change without notice.

