



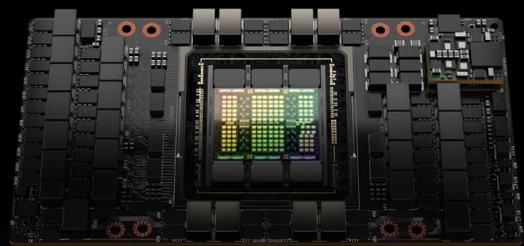
NVIDIA H800 Tensor Core GPU

NVIDIA Hopper 架构的 数据中心 GPU

NVIDIA® H800 Tensor Core GPU 可加速每个数据中心，为 AI 和数据分析应用提供动力支持。NVIDIA H800 GPU 采用第四代 Tensor Core 和支持 FP8 精度的 Transformer 引擎，可加速训练大型语言模型。NVIDIA Hopper™ 架构采用第二代多实例 GPU (MIG) 技术，支持在虚拟环境中实现多租户、多用户配置，安全地将 GPU 划分为独立的、大小合适的实例，为受到保障的租户最大化地提升服务质量 (QoS)。NVIDIA 机密计算是 Hopper 全新的内置安全特性，使 NVIDIA H800 成为全球领先的带有机密计算能力的加速器。当用户在使用 H800 GPU 时，可以保护其数据和应用程序的机密信息和完整性。同时，Hopper 的 DPX 指令在用于疾病诊断、实时路线优化和图形分析时，可加速动态规划算法。

为企业级 AI 准备就绪？

NVIDIA H800 Tensor Core GPU PCIe 规格为主流服务器提供 5 年软件订阅，包括企业级技术支持和 NVIDIA AI Enterprise 软件套件，以帮助简化 AI 应用落地。这可确保组织能够访问所需的 AI 应用框架和工具，以构建 H800 加速的 AI workflows，例如 AI 聊天机器人、推荐引擎、视觉 AI 等。了解更多 NVIDIA H800 的 NVIDIA AI Enterprise 软件订阅和相关支持权益，请访问[此处](#)。



规格

	H800 SXM	H800 PCIe
FP64	1 TFLOP	0.8 TFLOP
FP64 Tensor Core	1 TFLOP	0.8 TFLOP
FP32	67 TFLOPS	51 TFLOPS
TF32 Tensor Core	989 TFLOPS*	756 TFLOPS*
BFLOAT16 Tensor Core	1979 TFLOPS*	1513 TFLOPS*
FP16 Tensor Core	1979 TFLOPS*	1513 TFLOPS*
FP8 Tensor Core	3958 TFLOPS*	3026 TFLOPS*
INT8 Tensor Core	3958 TOPS*	3026 TOPS*
GPU 显存	80GB	80GB
GPU 显存带宽	3.35TB/s	2TB/s
解码器	7 NVDEC 7 JPEG	7 NVDEC 7 JPEG
最大热设计功耗 (TDP)	最高 700 瓦 (可配置)	300 – 350 瓦 (可配置)
多实例 GPU	支持最大 7 个 MIG 实例，每个实例 10 GB 显存	
外形规格	SXM	PCIe 双插槽风冷式
互连技术	NVLink™: 400GB/s PCIe 5.0: 128GB/s	NVLink: 400GB/s PCIe 5.0: 128GB/s
服务器选项	搭载 8 个 GPU 的 NVIDIA HGX™ H800 合作伙伴和 NVIDIA 认证系统和搭载 8 个 GPU 的 NVIDIA DGX™ H800 NVIDIA 认证系统 (NVIDIA-Certified Systems™)	搭载 1 至 8 个 GPU 的合作伙伴和 NVIDIA 认证系统
NVIDIA AI Enterprise 软件套件	附加	已包含

* 采用稀疏技术。非稀疏规格降低 1/2。

