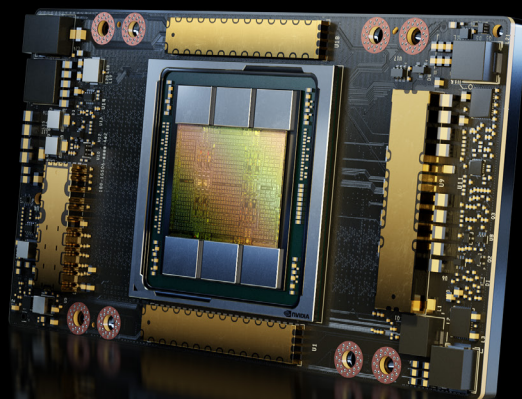




NVIDIA A800 TENSOR CORE GPU

NVIDIA Ampere 架构的数据中心 GPU



NVIDIA A800 Tensor Core GPU 可加速弹性数据中心，为 AI 和数据分析应用提供动力支持。A800 可高效扩展，也可借助多实例 GPU (MIG) 技术划分为 7 个独立的 GPU 实例，从而提供统一的平台，助力弹性数据中心根据不断变化的工作负载需求动态地进行调整。

NVIDIA A800 Tensor Core 技术支持广泛的数学精度，可针对每个工作负载提供单个加速器。新一代 A800 80GB 显存带宽高达 2TB/s，缩短了针对 AI 模型和大型数据集的解决方案实施时间。

A800 是完整 NVIDIA 数据中心解决方案的一部分，该解决方案由硬件、网络、软件、库以及 NVIDIA NGC™ 目录中经优化的 AI 模型和应用等叠加而成。作为端到端 AI 平台，A800 可助力研究人员获得真实的结果，并将解决方案部署到生产环境中。

NVIDIA A800 TENSOR CORE GPU 规格 (SXM4 和 PCIe 外形规格)

| | A800 80GB PCIe | A800 80GB SXM |
|------------------------|--|---|
| FP64 | 9.7 TFLOPS | |
| FP64 Tensor Core | 19.5 TFLOPS | |
| FP32 | 19.5 TFLOPS | |
| Tensor Float 32 (TF32) | 156 TFLOPS 312 TFLOPS* | |
| BFLOAT16 Tensor Core | 312 TFLOPS 624 TFLOPS* | |
| FP16 Tensor Core | 312 TFLOPS 624 TFLOPS* | |
| INT8 Tensor Core | 624 TOPS 1248 TOPS* | |
| GPU 显存 | 80GB HBM2e | 80GB HBM2e |
| GPU 显存带宽 | 1935GB/s | 2039GB/s |
| 最大热设计功耗 (TDP) | 300W | 400W*** |
| 多实例 GPU | 最多 7 个 MIG 每个 10GB | 最多 7 个 MIG 每个 10GB |
| 外形规格 | PCIe (双插槽风冷式 或单插槽液冷式) | SXM |
| 互连技术 | 搭载 2 个 GPU 的 NVIDIA® NVLink® 桥接器: 400GB/s** PCIe 4.0: 64GB/s | NVLink: 400GB/s PCIe 4.0: 64GB/s |
| 服务器选项 | 搭载 1 至 8 个 GPU 的 合作伙伴认证 系统和 NVIDIA 认证系统 (NVIDIA- CertifiedSystems™) | 搭载 4 个或 8 个 GPU 的 NVIDIA HGX™ A800 合作伙伴认证 系统和 NVIDIA 认证 系统 |

* 采用稀疏技术

** SXM4 GPU 通过 HGX A800 服务器主板连接；PCIe GPU 通过 NVLink 桥接器可连接多达两个 GPU

*** 400W TDP (适用于标准配置)。HGX A800-80GB 自定义散热解决方案 (CTS) SKU 可支持高达 500W 的 TDP