

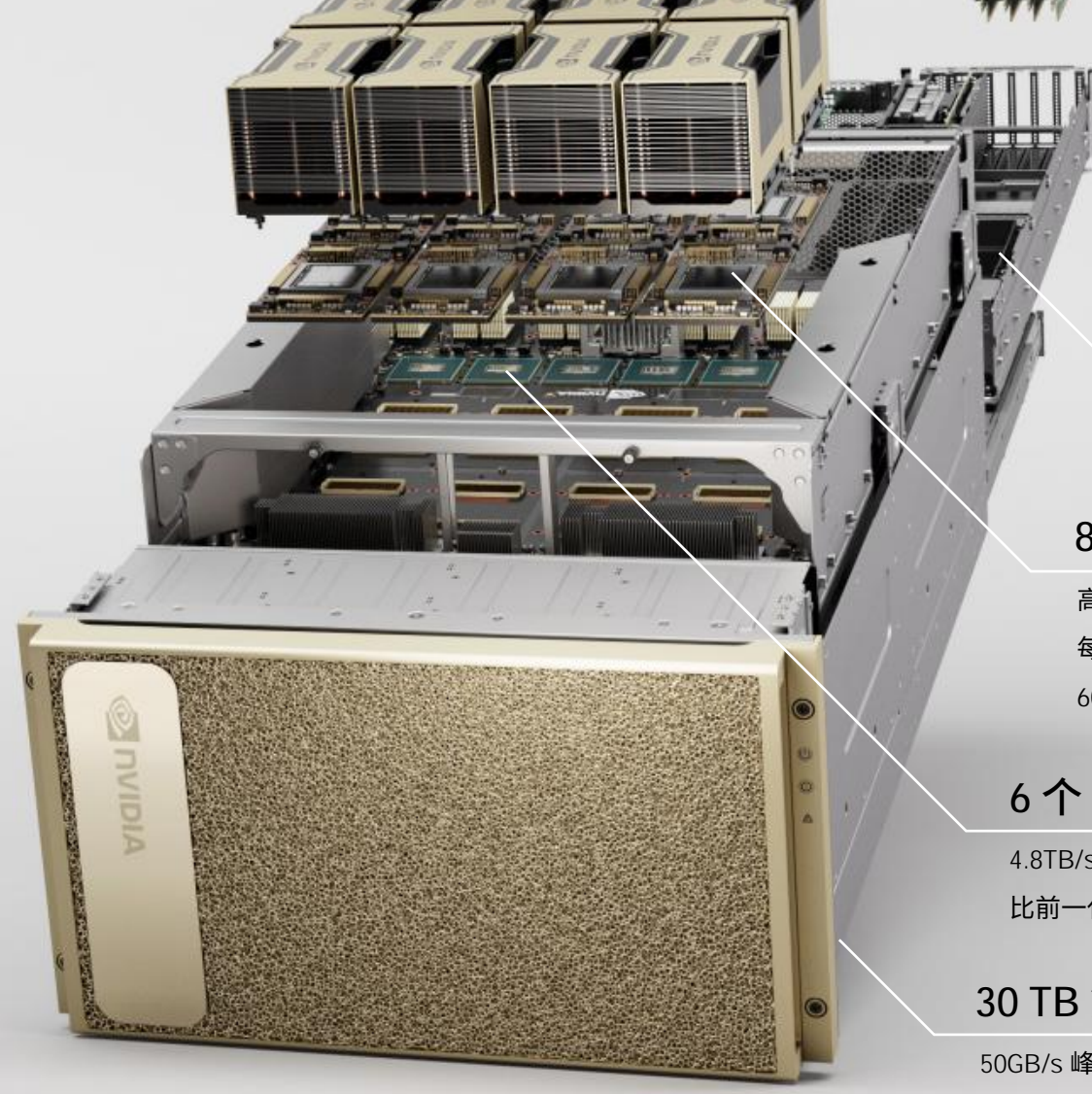
# NVIDIA DGX A100

AI 基础架构通用系统



## AI 数据中心的基石

经历重塑和优化的 AI 基础架构，为企业级 AI 准备



10 个 NVIDIA® Mellanox® ConnectX-6 200Gb/s 网络接口  
500GB/s 峰值双向带宽

双路 64 核 AMD Rome CPU  
2TB RAM  
以 3.2 倍核心数量支持超密集 AI 作业

8 个 NVIDIA A100 Tensor Core GPU  
高达 640GB 的总 GPU 显存  
每个 GPU 支持 12 个 NVIDIA NVLink™  
600GB/s GPU 间双向带宽

6 个 NVIDIA NVSwitch™  
4.8TB/s 双向带宽  
比前一代 NVSwitch 快两倍

30 TB 第 4 代 NVME SSD  
50GB/s 峰值带宽  
比第三代 NVME SSD 快两倍

1 万亿个  
晶体管

1 千米  
PCB 线路长度

1 百万个  
钻孔

3 万个  
组件

50 万个  
接线引脚

5 PETAFLOPS  
AI 性能

## 为创新者带来颠覆性的性能

NVIDIA DGX SUPERPOD: 大规模释放各个企业的 AI 潜能



NVIDIA DGX™ A100 系统

NVIDIA Mellanox Quantum™ HDR 200 Gb/s InfiniBand 智能交换机

千兆字节级全闪存存储

模块化构建，拥有 20 个可扩展节点单元

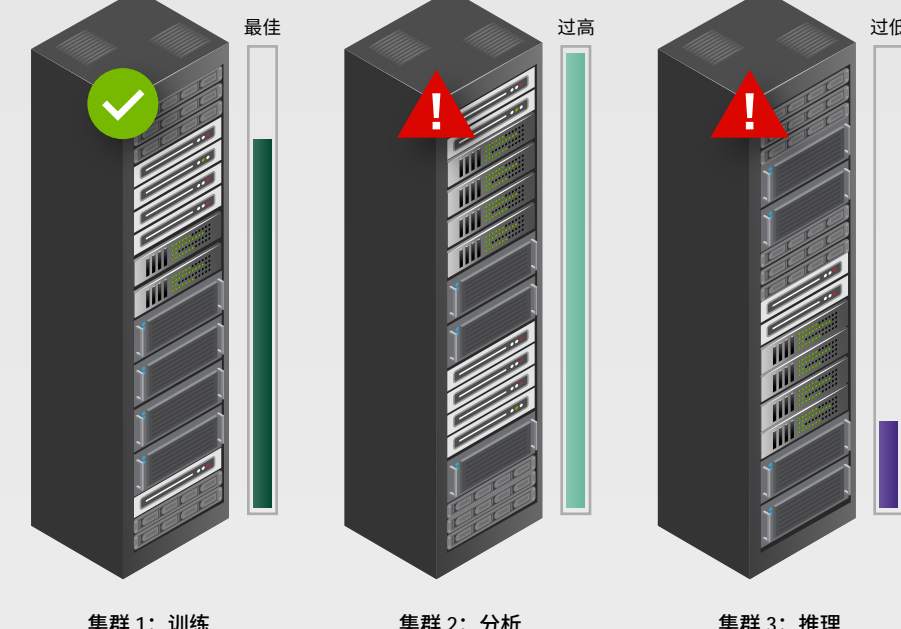
经 NVIDIA NGC™ 优化的软件

无需数月时间，仅数周内即可完成部署

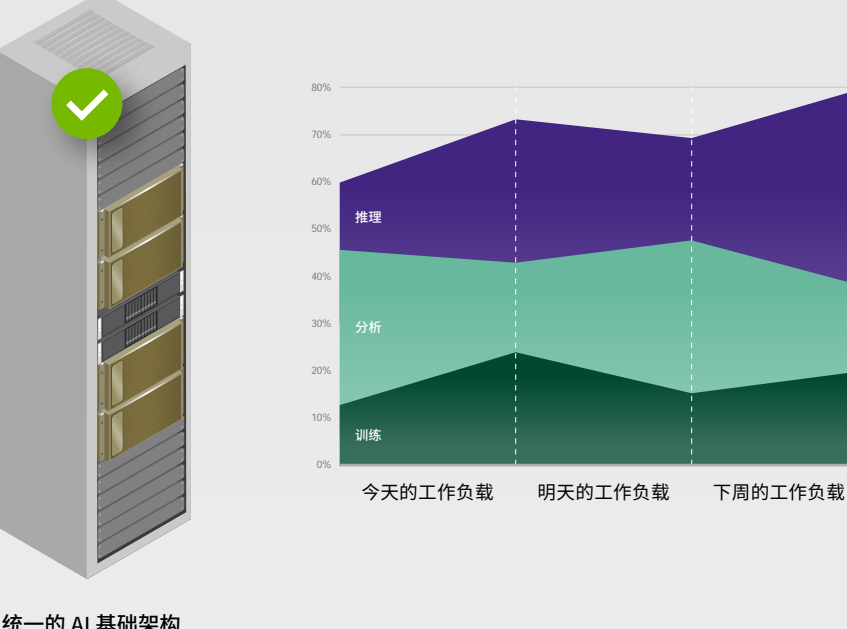
## 加速各类型 AI 工作负载

在同一个系统中运行数据分析、训练和推理工作负载

传统基础架构存在限制  
基础设施孤岛会使 AI 工作负载缺少资源，或浪费性能。



DGX A100 基础架构系统敏捷  
在任意系统上运行任意工作负载，大幅提升利用率。



## 适用于数据科学团队的强劲工具

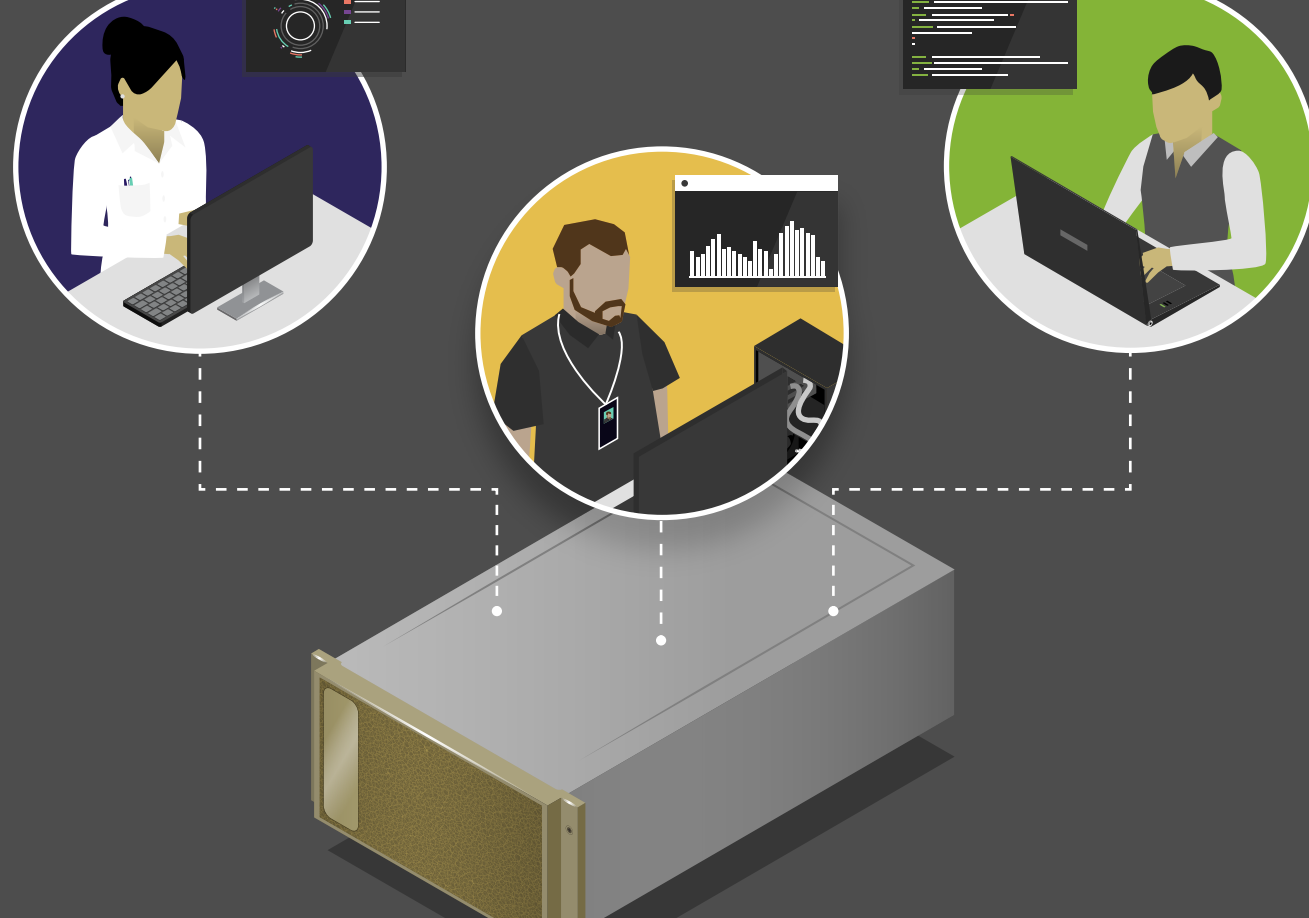
赋予每一位开发者探索的动力

在单个 DGX A100 上集成的多实例 GPU (MIG) 可为:

25 位开发者提供的性能，相当于每位开发者拥有

2 个专用 GPU 或

6 个专用 28 核双路 CPU 服务器



## 直接获取 NVIDIA DGXPert 全球团队的支持

借助每个 DGX 系统获得专业的 AI 知识与技能



NVIDIA 与您一起走过每一步

设计 / 计划 / 构建 / 测试 / 部署 / 运营 / 监控

10 多年  
AI 创新经验

100 多款  
经 GPU 优化的 NGC 软件和工具

10000 多名  
遍布全球的资深 AI 从业者

## 加速 AI 发展，加快投资回报率 (ROI) 增长

部署灵活的 AI 基础架构，掌控资本支出 (CAPEX)



当前的 AI 数据中心

1100 万美元 | 25 个机架 | 630kW

DGX A100 数据中心

100 万美元 | 1 个机架 | 32.5kW  
5 个 DGX A100 系统，用于实现 AI 训练和推理

立即释放您所在企业的 AI 潜能

[www.nvidia.cn/data-center/dgx-a100/](http://www.nvidia.cn/data-center/dgx-a100/)