**存储服务器DGX A100技术白皮书**

****

产品特征：

扩展企业 AI 的挑战

每家企业都需要利用人工智能 (AI) 实现转型，以在这个充满挑战的时代求得生存，继而实现蓬勃发展。但长期以来，传统方法所采用的计算架构较为缓慢，而且总是分开处理分析、训练和推理工作负载，所以企业需要一种适用于 AI 基础架构的平台对此加以改进。传统方法不仅复杂、成本高、扩展速度受限，而且对现代 AI 束手无策。 因此，企业、开发者、数据科学家和研究人员都需要一个新平台，以便统一处理所有 AI 工作负载、简化基础架构以及提高投资回报率 (ROI)。

适用于各种 AI 工作负载的通用系统

作为一种通用系统，NVIDIA DGX™ A100 可处理各种 AI 工作负载，包括分析、训练和推理。DGX A100 设立了全新计算密度标准，不仅在 6U 外形尺寸下封装了 5 petaFLOPS 的 AI 性能，而且用单个统一系统取代了传统的计算基础架构。此外，凭借 NVIDIA A100 Tensor Core GPU 中的多实例 GPU (MIG) 功能，DGX A100 首次实现了强大算力的精细分配，使得管理员可针对特定工作负载分配大小合适的资源。DGX A100 提供了多达 640GB 的 GPU 显存总量，可将大型训练作业的性能提升高达三倍，并将 MIG 实例的大小翻倍，因而能够处理极为复杂的大型作业，同时轻松应对简单、小型的任务。DGX A100 利用 NGC 中经过优化的软件运行 DGX 软件堆栈，同时将密集计算能力和全面的工作负载灵活性相结合，因此成为单节点部署以及通过 NVIDIA DeepOps 部署大规模 Slurm 和 Kubernetes 集群的理想之选。。

安全性更高的企业 AI 系统

NVIDIA DGX A100 采用多层级架构保护所有主要的软硬件组件，确保 AI 企业处于稳定的安全状态。DGX A100 内置安全机制，覆盖基板管理控制器 (BMC)、 CPU 载板、GPU 载板、自加密驱动和安全启动，可帮助 IT 人员专注于 AI 操作，而不必花时间评估和应对安全威胁。

通过 NVIDIA Mellanox 实现卓越的数据中心可扩展性

NVIDIA DGX A100 配备所有 DGX 系统中速度领先的 I/O 架构，是 NVIDIA DGX SuperPOD™ 等大型 AI 集群的基础构件，而后者为可扩展的 AI 基础架构描绘了企业蓝图。DGX A100 具有八个用于实现集群的单端口 NVIDIA Mellanox® ConnectX®-6 VPI HDR InfiniBand 适配器，以及多达两个用于存储和网络连接的双端口 ConnectX-6 VPI 以太网适配器，二者的速度均可达到 200Gb/s。借助大规模 GPU 加速计算与精尖网络硬件和软件优化的强强联合，DGX A100 可扩展至数百乃至数千个节点，从而攻克对话式 AI 和大规模图像分类等更艰巨的挑战。

产品参数：

|  |  |  |
| --- | --- | --- |
| 产品型号 |  | NVIDIA A100 |
| GPU个数 |  | 8 个 NVIDIA A100 80GB GPU |
| GPU显存 |  | 共 640GB |
| 性能 |  | 5 petaFLOPS AI 10 petaOPS INT8 |
| NVIDIA NVSwitch |  | 6 |
| 系统功率 |  | 最大 6.5 千瓦 |
| CPU |  | 两个 AMD Rome 7742、共 128 个核心、  2.25 GHz（基准频率）、3.4 GHz（最大加速频率） |
| 系统内存 |  | 2TB |
| 网络 |  | 8 个单端口 Mellanox  ConnectX-6 VPI  200Gb/s HDR  InfiniBand 两个双端口  Mellanox  ConnectX-6 VPI 10/25/50/100/  200Gb/s 以太网 |
| 存储 |  | 操作系统 ：两个 1.92TB M.2 NVMe  驱动器内部存储 ： 30TB（8个3.84TB）U.2NVMe 驱动器 |
| 软件 |  | Ubuntu Linux 操作系统同时支持 ： Red Hat Enterprise Linux CentOS |
| 系统重量 |  | 最大 123.16 千克 |
| 包装后的系统重量 |  | 最大 163.16 千克 |
| 系统尺寸 |  | 高度 ：264.0 毫米  宽度 ：最大 482.3 毫米  长度 ：最大 897.1 毫米 |
| 运行温度范围 |  | 5-30 ºC |

如需更多信息

请访问中科云达官方网站：www.cloudta.com.cn 或拨打免费服务电话：400-801-5670

销售：sales@cloudta.com.cn

中科云达公司保留对产品规格或其它产品信息（包括但不限于产品重量、外观、尺寸或其它物理因素）不经通知予以更改

的权利；文中所提到的信息，如因产品升级或其它原因而导致的变更，恕不另行通知。本文中所涉及的产品图片均以产品实物为准。



中科云达（北京）科技有限公司

地址：北京市昌平区回龙观镇高新三街1号1幢1层102

[更多产品信息请访问：](http://www.chinasupercloud.com/)www.cloudta.com.cn

客户服务热线：400-801-5670

关注中科云达官方微信可快速了解更多资讯