



# NVIDIA DGX SUPERPOD 企业解决方案

## 实现大规模 AI 创新的捷径

NVIDIA DGX SuperPOD™ 企业解决方案是业界首个支持任何组织大规模实施 AI 的基础架构解决方案。这一全套解决方案基于 NVIDIA DGX SuperPOD 参考架构搭建，专门面向需要快速实现 AI 创新的企业而设计，为企业提供了可快速部署的行业领先基础架构，并由完整的端到端服务和生命周期服务提供支持，而这一切均由 NVIDIA 作为强大后盾。

NVIDIA DGX SuperPOD 企业解决方案可在短短几周而非长达几个月内提供全方位服务体验和经行业验证的成果。这一解决方案不仅仅是硬件的集合，更是一个全堆栈平台，其中包含业内领先的计算、存储、网络、基础架构管理和数据科学工作流程工具。这些组件经过优化，可结合使用并带来大规模出色性能，同时还可提供高端递送服务，确保顺利部署和运行。

### 攻克大规模、多节点 AI 基础架构挑战

NVIDIA DGX SuperPOD 企业解决方案的设计宗旨是要攻克大规模 AI 所面临的重要挑战，实现无可比拟的多系统训练。由于配置规模不断扩大，且越来越多的节点实现了并行计算，因此扩展 GPU 间的通信变得相当复杂，进而使得传统大型计算集群十分受限。随着系统的扩张，这将导致性能收益越来越低。NVIDIA DGX SuperPOD 企业解决方案针对多节点 AI 基础架构的独特需求优化了系统中的每个组件，从而解决了这一扩展问题。基于同一 DGX SuperPOD 架构搭建的 Selene 是 NVIDIA 自有的 NVIDIA DGX SuperPOD 部署解决方案，也是全球速度和能效方面均无比出色的超级计算机，正如 TOP500 和 Green500 榜单所示。<sup>1</sup> 其在多个 MLPerf 基准测试结果中也独占鳌头。<sup>2</sup>

### 智能适应并集成到您的业务中

数据科学团队需要合适的工具、平台和基础架构来简化 AI 工作流程，加速获得见解。IT 团队需要寻找合适的合作伙伴来帮助扩展现有基础架构，以及应对高性能计算、网络结构、存储架构和 AI 软件的复杂性，这些都是扩展 AI 所不可或缺的。因此他们需要顺应业务需求和实施时间限制的灵活部署方法。NVIDIA 的专业服务团队可帮助针对任何环境优化 NVIDIA DGX SuperPOD 企业解决方案，包括根据您的独特需求量身打造灵活的部署方案。

### NVIDIA 支持的贯穿整个生命周期的专业知识

除架构设计之外，企业还需采用更便捷的方式，确保加速计算基础架构对业务发挥更大效用。企业需要围绕其 IT 环境获得全套、快速且经优化的实施体验，确保数据科学家在首日即可投入工作。借助 NVIDIA DGX SuperPOD 企业解决方案，企业可以从数据中心规划全套服务和基础架构交付专业知识中获益，从而加速规模调整、安装、训练和持续优化等各个部署阶段——所有这一切均由 NVIDIA 以及我们的 DGX SuperPOD 企业版解决方案合作伙伴提供支持。

### DGX SUPERPOD 企业解决方案

#### 硬件和软件

- > 100-700 PFLOPS AI 系统
- > 20-140 NVIDIA DGX™ A100 系统
- > 1-10PB 高性能存储
- > 200 Gbps NVIDIA® Mellanox® 网络结构
- > NVIDIA CUDA-X™ 和 DGX 软件堆栈
- > ML0ps 工具

#### 生命周期服务\*

##### 规划和部署\*\*

- > 容量规划
- > 数据中心设计
- > 性能预估
- > 站点评估 / 准备
- > 安装
- > 安装后测试
- > 配置和管理

#### 训练和优化

- > 应用性能测试
- > 站点文档包
- > 用户和 DevOps 训练
- > 基于工作负载的 NVIDIA 深度学习学院培训
- > 定制系统运行手册
- > 移交会议

\* NVIDIA 与合作伙伴服务组合

\*\* 企业本地或 DGX-Ready 数据中心内部署



## NVIDIA DGX SuperPOD 企业解决方案

### 单一解决方案中的高性能基础架构 - 面向 AI 优化

NVIDIA DGX SuperPOD 企业解决方案将经过设计优化的 AI 计算、网络结构、存储和软件集于一身。其依托 NVIDIA DGX™ A100 开展计算，这是面向所有 AI 工作负载的通用系统，可提供远超以往的计算密度、性能和灵活性。每个 NVIDIA DGX A100 系统都具有高达 640GB 的 GPU 显存总量，且配备世界顶尖的加速器 NVIDIA A100 Tensor Core GPU，支持企业将训练、推理和分析整合到一个统一且易于部署的 AI 基础架构中。

作为 DGX SuperPOD 企业解决方案专用的高性能网络结构，NVIDIA Mellanox 采用创新型 NVIDIA Mellanox InfiniBand™ 网络内计算技术，其中包括 NVIDIA Mellanox Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)™ 技术和拥塞控制技术。这一强大结合实现了出色的性能和可扩展性，且有助降低运营成本和基础架构的复杂度。

AI 超级计算机还需要超快速的存储器以在尖峰容量下运行。在结构精妙的系统中，存储解决方案需要并行处理各种数据类型（例如文本、表格数据、音频和视频），且要始终保持高性能以应对 AI 数据的巨大深度和多样性。NVIDIA DGX SuperPOD 企业解决方案经认证的存储器由我们精心挑选，并已针对 AI 工作负载的独特需求进行测试，此外还针对您的环境进行了专门优化，以确保成功。

为扩展 AI，企业需要在 IT 和 DevOps 方法中集成经优化的软件和数据科学工作流。MLOps 软件简化了 AI 应用交付流程，使数据科学团队和 IT 部门能够更有效地管理用户、模型、数据集、试验等更多资源，同时加速应用的持续交付。DGX SuperPOD 企业解决方案包含来自 NVIDIA NGC™ 目录的经全面优化的 AI 软件，并提供来自 NVIDIA DGX-Ready Software 合作伙伴的 MLOps 软件，可帮助组织管理、扩展和加速 AI 及数据科学。此软件堆栈提供了简化的机器学习流程，支持数据科学从业人员和 IT 及 DevOps 团队合作取得出色性能，同时加速生产应用的部署。

### 我们的经验助力您获得成功

DGX SuperPOD 企业解决方案整合了数千名 NVIDIA 研究人员和工程师在设计和使用 AI 超级计算机方面积累的丰富经验，他们将通过此平台将创新技术推向市场。这些全球 AI 专家团队每天都会使用 DGX SuperPOD，随时都可助您实现 AI 抱负。

如需了解有关 NVIDIA DGX SuperPOD 企业解决方案的详细信息，请访问 <https://www.nvidia.cn/data-center/dgx-superpod/>

1 访问 [top500.org](https://top500.org) 获取详情 | 2 访问 [mlperf.org](https://mlperf.org) 获取详情

© 2020 NVIDIA Corporation. 保留所有权利。NVIDIA、NVIDIA 徽标、CUDA-X、CUDA-X AI、DGX A100、DGX SuperPOD、Mellanox、Mellanox InfiniBand、Mellanox SHARP 和 NGC 均为 NVIDIA Corporation 在美国和其他国家或地区的商标或注册商标。其他公司和产品名称可能是其各关联公司的商标。其他所有商标均为其各自所有者的资产。2020 年 11 月

