

IT 是 AI 的根基

AI 正在改变各个行业，但很多公司在没有 IT 主导策略的情况下就开始了转型。其结果是步入创新的孤岛，无法有效扩展。正确的 AI 基础架构策略能够吸引人才、整合资源并推动创新。

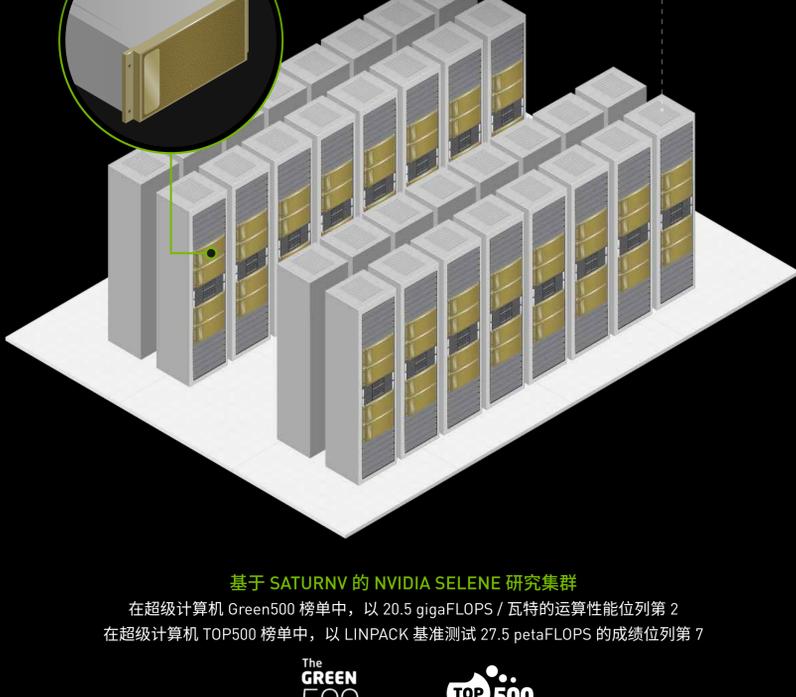
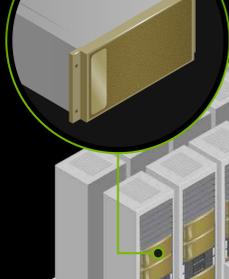
大规模部署 AI 对数据中心提出了前所未有的要求。如要利用机器学习和深度学习助力公司业务转型，您需要针对 AI 独特需求而优化的基础架构。



我们已经建立了世界上最先进的 AI 基础架构，并将其命名为“SATURNV”。此架构集成了我们在世界大型数据中心设计、部署可扩展 AI 方面积累的全部技术。

NVIDIA DGX™ A100 AI 数据中心基础架构的通用系统

总性能增长至 4.6 ExaFLOPS



基于 SATURNV 的 NVIDIA SELENE 研究集群

在超级计算机 Green500 榜单中，以 20.5 gigaFLOPS / 瓦特的运算性能位列第 2
在超级计算机 TOP500 榜单中，以 LINPACK 基准测试 27.5 petaFLOPS 的成绩位列第 7



工作中的可扩展 AI

在 NVIDIA，基于 NVIDIA DGX 的 IT 基础架构为 AI 最重要的工作提供动力。NVIDIA DGX 基础架构和革命性的性能推动了 AI 跨行业快速发展，如助力自动驾驶汽车发展、最大限度地提高产品质量、提升客户满意度，以及推动 AI 研究取得新突破等。每一天，NVIDIA 开发者、操作人员及客户都能通过该基础架构提供的全面覆盖其业务的革命性性能及轻松提升的生产力而获益。

NVIDIA RTX™ 显卡 适用于创意应用的实时光线追踪和 AI 功能



性能再创新高

在领先的 AI 性能基准套件 MLPerf v0.7 训练中，NVIDIA DGX SuperPOD 和 DGX 系统在商用系统的所有 8 个大规模基准中均创下了世界纪录。这一拔得头筹的基础架构解决方案基于 DGX SuperPOD 参考架构构建，并在不到一个月的时间内完成了部署。

推荐 (DLRM) --->

3.33 分钟

通过了解用户与服务项目（如产品或广告）之间的交互方式和内容，在社交媒体或电子商务网站等面向用户的服务中提供个性化结果。

自然语言处理 (BERT) --->

0.81 分钟

根据一段文本中不同单词之间的关系理解文本。能够支持回答问题、解释句子和许多与语言相关的其他用例。

强化学习 (MINIGO) --->

17.07 分钟

评估不同的可能行为，在 19x19 棋盘的围棋这一策略游戏时，赢得最多奖励。

翻译 (NON-RECURRENT) - TRANSFORMER --->

0.62 分钟

使用前馈神经网络将文本翻译为另一种语言。

翻译 (RECURRENT) - GNMT --->

0.71 分钟

使用循环神经网络 (RNN) 将文本翻译为另一种语言。

目标检测 --->

10.46 分钟

(重量级) Mask R-CNN

0.82 分钟

(轻量级) SSD

在图像或视频中查找真实对象的实例（如人脸、自行车和建筑物等），并对每个对象指定边界框。

图像分类 (RESNET-50 V1.5) --->

0.76 分钟

将标签从固定的类别集中分配给输入图像，适用于自动驾驶汽车等计算机视觉问题。

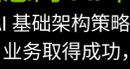
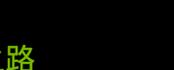
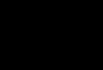
测试平台：搭载 NVIDIA DGX A100 多用途集群的 NVIDIA DGX SuperPOD，DGX A100 配置：双路 Epyc 7742、1TB 系统内存、Mellanox ConnectX-6 VPI HDR InfiniBand/以太网，以及 8 个 NVIDIA A100 Tensor Core GPU

最大规模：所有 MLPerf v0.7 结果均基于 NVIDIA DGX A100（八个 A100）测试得出。MLPerf ID 最大规模：ResNet50 v1.5: 0.7-37, Mask R-CNN: 0.7-28, SSD: 0.7-33, GNMT: 0.7-34, 转换器: 0.7-34, MiniGo: 0.7-34, BERT: 0.7-38, DLRM: 0.7-17.

MLPerf 名称和徽标均为商标。详情请参见 www.mlperf.org

构建您的世界级 AI 基础架构

您可以充分利用 NVIDIA 模块化参考架构 NVIDIA DGX SuperPOD™，该架构以 SATURNV 中的见解为基础，并由 DGX A100 系统和我们生态系统中值得信赖的 IT 解决方案供应商提供支持。



引领您的 AI 转型之路

我们可以帮助您制定并执行 AI 基础架构策略。CIO 有机会通过开发和执行 GPU 加速的 AI 基础架构战略使 AI 业务取得成功，这一战略将使 IT 部门成为现在及将来实现成功转型的可信赖合作伙伴。

www.nvidia.cn/data-center/dgx-superpod