

加速数据工程

数据科学基础，通过 GPU 加速数据 workflow

数据工程是数据科学的基础，为分析和建模奠定了基础。为了让组织机构能够从结构化和非结构化数据中提取信息和见解，快速访问准确且完整的数据集至关重要。处理不同来源的海量数据需要复杂的基础设施和专业技能。当扩展到数百万乃至数万亿个数据点时，看似微不足道的低效有可能导致巨大的时间和金钱成本。

在此课程中，我们将探讨 GPU 如何改进数据 workflow，以及如何运用先进的数据工程技术和工具，显著加速性能。通过更快速的数据通道创建仪表板和机器学习 (ML) 模型，用户可以随时获得最新信息。

课程时长	8 小时 (课后 6 个月内可以继续访问和使用课件和实验资源)。
课程模式	讲师授课，及每位学员使用云端完全配置的 GPU 加速工作站实验练习。
课程价格	微信添加 DLI 小助手 (微信号 NVIDIALearn)，沟通培训需求。
学员评测方式	基于技能的编码测试，评估学员通过交互式仪表板，对百万数据点进行高效筛选的能力。
培训证书	成功完成本课程和测试后，学员将获得 NVIDIA DLI 培训证书，证明在相关领域的技能，为职业发展提供证明。
预备知识	<ul style="list-style-type: none"> • 中级 Python 知识 (list comprehension、objects) • 熟悉 pandas • 统计学基础 (mean、median、mode)
课程语言	英文
技术	pandas, cuDF, Dask, NVTabular, Plotly



学习此课程的硬件要求	您需要一台笔记本电脑或台式机，且能够运行最新版 Chrome 或 Firefox 浏览器。我们为您提供在云端完全配置的 GPU 加速工作站的专用访问权限。
-------------------	---

学习目标

参加本次培训，您将学习到：

- 数据在计算机内的移动方式。如何在 CPU、DRAM、磁盘内存与 GPU 之间构建适当的平衡。
- 硬件如何读取和操控不同的文件格式。
- 通过多个 GPU，如何使用 NVTabular 扩展 ETL 工作流。
- 如何构建交互式 Plotly 仪表盘，让使用者秒筛数百万的数据点。

为何选择 NVIDIA 深度学习培训中心 (DLI) 的实战培训

- 随时随地访问云端完全配置的 GPU 加速工作站来动手实践。
- 获得实战经验指导，使用通用、行业标准的软件、工具和框架。
- 学习如何在广泛的行业中构建深度学习和加速计算应用程序，如自动驾驶汽车、数字内容创作、游戏开发、医疗医学及金融。
- 学习与行业领导者（例如洛杉矶儿童医院、梅奥医院和普华永道）合作设计的课程，获取现实应用的专业知识。
- 获得 NVIDIA 官方全球开发者培训证书，证明在相关领域的能力，助力职业发展。

课程大纲

议题	说明
课程介绍 (15 分钟)	> 讲师介绍 > 登录课件平台



<p>硬件上的数据 (60 分钟)</p>	<p>探索不同硬件处理数据方法的优缺点，以及支持这些方法的框架</p> <ul style="list-style-type: none"> > Pandas > CuDF > Dask
<p>休息 (15 分钟)</p>	
<p>采用 NVTabular 的 ETL (120 分钟)</p>	<p>通过大数据推荐系统，学习如何使用 NVTabular 将 ETL 工作流从 1 个 GPU 扩展至多个</p> <ul style="list-style-type: none"> > 将原始 Json 转换为分析就绪的 Parquet 文件 > 学习如何快速向数据集添加特征，例如 Categorify 和 Lambda 运算符
<p>休息 (60 分钟)</p>	
<p>数据可视化 (120 分钟)</p>	<p>扮演气象专家角色，学习如何在地图上绘制降水数据</p> <ul style="list-style-type: none"> > 学习如何使用描述性统计数据和直方图等图形来评估数据质量 > 学习如何提升内存使用效率，让用户快速通过图形界面筛选数据
<p>休息 (15 分钟)</p>	
<p>结课项目：数据侦探 (15 分钟)</p>	<p>用户抱怨仪表盘速度太慢。运用本课中学到的技术，查找和消除后端代码的效率问题</p>
<p>总结 (15 分钟)</p>	<ul style="list-style-type: none"> > 回顾所学要点并回答问题 > 完成评估并获得证书 > 填写培训调查表 > 了解如何设置您自己的 AI 应用程序开发环境

相关课程

- 加速数据科学基础
- 构建智能推荐系统
- 加速计算基础 —— CUDA Python
- 更多课程, 请访问 www.nvidia.cn/dli

培训咨询和购买

- 微信添加 DLI 小助手, 微信号 [NVIDIALearn](#) 。

