

加速数据科学基础

为了提升客户保留率、预测客户购买行为，或降低金融服务风险，企业或组织通常需要在不浪费关键资源的前提下，从大量的数据中筛选、处理并采集能提供深刻见解的信息。传统的由 CPU 驱动的数据科学分析工作流可能会很缓慢冗长，而借助 GPU 您的团队可以快速理解数据以驱动业务决策。

在本课程中，数据科学家和开发人员将学习如何构建和执行端到端的由 GPU 加速的数据科学工作流，从而能快速地探索数据、迭代算法并将工作成果转换成产品。有了 RAPIDS 加速数据科学库，开发人员就可直接应用各种由 GPU 加速的机器学习算法，包括 XGBoost、KNN、DBSCAN 和逻辑回归等，来执行大规模的数据分析。

所有参加培训的人员都可以访问云端的完全配置的 GPU 加速服务器，获得 DLI 认证讲师的指导，并有机会结识其他开发人员。完成课程和评估测试后，参与者可以获得培训证书，证明相关技能来助力职业发展。

课程时长	8 小时（课后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	基于代码
培训证书	成功完成本课程和测试后，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	<p>Python 使用经验，包含 Pandas 和 NumPy。</p> <ul style="list-style-type: none"> 为获得 pandas 经验，我们建议学习 Kaggle 上的 pandas 课程。 为获得使用 Python 的数据科学经验，我们建议学习 Kaggle 上的机器学习课程。
课程语言	英文
技术、工具、库和框架	RAPIDS, cuDF, XGBoost, cuML, cuGraph, Dask, cuPy, pandas, NumPy, Bokeh, data science, data analytics, machine learning, deep learning

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

在本课程中，您将学习如何：

- 使用 cuDF 和 Apache Arrow 数据帧进行 GPU 加速的数据准备和特征提取。
- 使用 XGBoost 和其它多种cuML里的算法实现由 GPU 加速的多领域的机器学习任务。
- 使用 cuGraph 执行由GPU加速的数据图形可视化，在较短的时间内实现大规模的数据分析。
- 使用 cuGraph 例程快速实现大规模的图形分析。

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
介绍 (15 分钟)	> 讲师介绍 > 登录课程

<p>GPU 加速的数据处理 (120 分钟)</p>	<p>获取和准备多个数据集(其中一些将大于内存), 以供后续的多个机器学习实验使用。</p> <ul style="list-style-type: none"> > 使用 cuDF 和 Dask cuDF 直接将数据读取到单个和多个 GPU。 > 使用 cuDF 在 GPU 上为机器学习任务准备有关人口、道路网络和诊所的信息。
<p>午休 (60 分钟)</p>	
<p>GPU 加速的机器学习 (120 分钟)</p>	<p>使用准备好的数据, 练习一些基本的机器学习技术。</p> <ul style="list-style-type: none"> > 基于 cuML 算法, 应用有监督和无监督的 GPU 加速算法。 > 在多个 GPU 上, 用 Dask 训练 XGBoost 模型。 > 使用 cuGraph 在 GPU 上创建和分析图形数据。
<p>休息 (15 分钟)</p>	
<p>项目实践: 数据分析以拯救英国 (120 分钟)</p>	<p>应用新的 GPU 加速的数据处理方法, 对人口规模数据进行分析, 来避免影响整个英国人口的 (模拟的) 流行病。</p> <ul style="list-style-type: none"> > 使用 RAPIDS 集成多个大型数据集, 并对真实世界进行分析。 > 根据模拟的流行病的每日数据, 以已有的分析为基础进行迭代处理。
<p>总结 (15 分钟)</p>	<ul style="list-style-type: none"> > 回顾所学关键内容 > 完成测试, 获取证书 > 填写调查表

相关课程

- 加速数据工程
- 构建智能推荐系统
- 更多课程, 请访问 www.nvidia.cn/dli



购买培训和咨询

- 微信扫码添加 DLI 小助手，微信号 [NVIDIALearn](#)。

