



# 加速数据科学基础



## 课程概述

数据科学是指使用科学方法、流程、算法和系统来分析数据并从中提取洞察。它使组织能够将数据转化为有价值的资源，从而实现更明智的决策、更高效的运营以及更好的客户体验。在本课程中，您将学习如何使用 GPU 加速工具更快地开展数据科学工作，从而获得更具可扩展性、更可靠且更具成本效益的结果。

## 学习目标

### 参加本次培训，您将学到：

- > 使用 cuDF 加速 pandas、Polars 和 Dask，高效分析各种规模的数据集
- > 使用多种机器学习算法 (包括 XGBoost) 解决不同的数据科学问题
- > 在 Triton Inference Server 上部署机器学习模型，以实现最佳性能
- > 学习并应用强大的图算法，使用 NetworkX 和 cuGraph 分析复杂网络
- > 在超大规模数据集上执行多种分析任务，以遏制影响英国的模拟疫情爆发

完成本课程后，您将能够更高效地完成各类数据科学任务，从而支持更多迭代周期，并显著提升整体生产力。

## 课程概要

课程时长	8 小时，课后 6 个月内可以继续访问课件和实验 (实验资源用量有限额)。
课程模式	讲师授课，每位学员使用完全配置的云端实验环境进行实战练习。
预备知识	<ul style="list-style-type: none"> <li>&gt; 需要具备 Python 使用经验，最好熟悉 pandas 和 NumPy。</li> <li>&gt; 用于满足先修条件的推荐资源：<a href="#">Kaggle: pandas 教程</a>、<a href="#">Kaggle: 机器学习入门</a>。</li> </ul>
使用的工具、库和框架	RAPIDS、cuDF、XGBoost、cuML、cuGraph、Dask、cuPy、pandas、Polars、NumPy 和 Bokeh
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。
课程语言	中文
学习此课程的硬件要求	您需要一台能够上网的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实验环境的专用访问权限。
课程价格	微信添加 DLI 小助手 (微信号 NVIDIALearn)，沟通培训需求。

## 课程大纲

<b>课程介绍</b> (30 分钟)	<ul style="list-style-type: none"><li>&gt; 讲师介绍</li><li>&gt; 登录 DLI 学习平台</li><li>&gt; 课程总览</li></ul>
<b>GPU 加速的数据处理</b> (120 分钟)	<p><b>摄取并准备多个数据集 (其中部分数据规模大于内存), 用于本课程后续的多项机器学习练习:</b></p> <ul style="list-style-type: none"><li>&gt; 使用 pandas、Polars、cuDF 和 Dask 将数据直接读取到单 GPU 或多 GPU 上。</li><li>&gt; 使用 cuDF 在 GPU 上为机器学习任务准备人口、道路网络和诊所相关数据。</li></ul>
<b>休息 (60 分钟)</b>	
<b>GPU 加速的机器学习</b> (120 分钟)	<p><b>将多种关键机器学习技术应用于第一部分中准备好的数据:</b></p> <ul style="list-style-type: none"><li>&gt; 使用 cuML 提供的有监督与无监督 GPU 加速算法。</li></ul>
<b>图分析</b> (120 分钟)	<p><b>执行图分析任务:</b></p> <ul style="list-style-type: none"><li>&gt; 使用 cuGraph 在 GPU 上创建并分析图数据。</li></ul>
<b>休息 (15 分钟)</b>	
<b>项目: 用数据分析守护英国</b> (60 分钟)	<p><b>结合人口级别的数据, 应用新学到的 GPU 加速数据处理与分析技能, 帮助遏制影响整个英国人口的模拟疫情:</b></p> <ul style="list-style-type: none"><li>&gt; 使用 RAPIDS 集成多个超大规模数据集, 并开展真实世界的数据分析。</li><li>&gt; 随着模拟疫情在每个模拟日产生新数据, 不断调整和迭代分析方案。</li></ul>
<b>评估测试和总结</b> (30 分钟)	<ul style="list-style-type: none"><li>&gt; 回顾所学要点</li><li>&gt; 运用所学内容, 通过课程测试</li><li>&gt; 获取 NVIDIA 培训证书</li></ul>
<b>下一步</b>	<p><b>学习更多 DLI 相关课程:</b></p> <ul style="list-style-type: none"><li>&gt; <a href="#">深度学习新手入门</a></li><li>&gt; <a href="#">深度学习基础 —— 理论与实践入门</a></li><li>&gt; <a href="#">为大语言模型添加新知识</a></li><li>&gt; <a href="#">构建基于大语言模型 (LLM) 的应用</a></li><li>&gt; <a href="#">构建基于扩散模型的生成式 AI 应用</a></li><li>&gt; <a href="#">模型并行 —— 构建和部署大型神经网络</a></li></ul>

## 为何选择 NVIDIA 深度学习培训中心 (DLI) 的实战培训?

- > 学习 NVIDIA 与技术专家和行业领导者合作开发的课程, 获取全球同步、技术领先和现实可用的专业开发技能和经验。
- > 学习使用行业通用、标准的软件、工具和框架进行端到端的应用开发, 能够在广泛的行业中构建基于生成式 AI、代理式 AI、大语言模型、加速计算、深度学习、数据科学、机器人开发等应用。
- > 系统化地学习理论, 并使用云端完全配置的实验环境同步边练, 高效提升实战开发能力。
- > 获得 NVIDIA 培训证书, 并进一步考取 NVIDIA 认证, 助力职业发展, 强化团队关键技能。

## 准备好开始学习了吗?

查询更多 DLI 课程, 请访问 [nvidia.cn/training](https://nvidia.cn/training)

如有疑问, 请通过微信联系 DLI 小助手 (微信号 NVIDIALearn)