NVIDIA Deep Learning Institute (DLI)
Accelerated Data Science Teaching Kit

Quick Start Guide

1. Introduction

This guide introduces you to the main teaching content found in the **Accelerated Data Science Teaching Kit** and provides basic instructions to access other features of the kit.

NVIDIA co-developed this Teaching Kit with Professor Polo Chau from Georgia Institute of Technology (Georgia Tech) and Professor Xishuang Dong from Prairie View A&M University for use in variety of academic teaching disciplines that benefit from accelerated Data Science. This comprehensive package contains everything needed to teach a full-term curriculum course in Data Science and analytics with GPUs.

The Accelerated Data Science Teaching Kit covers introductory, intermediate, and advanced Data Science and analytics topics. The content includes modules exploring concepts and programming for data analysis technologies and machine learning applications. Topics such as data collection and pre-processing, scalable and distributed computing, GPU-accelerated machine learning, and data visualization are covered in current or future versions of the kit. Some emerging technologies and techniques for implementing solutions for projects and lab exercises focus on NVIDIA GPU-accelerated algorithms and Data Science using the RAPIDS open-source computing framework. The content also includes topics in cultural awareness; from addressing issues such as bias and fairness in Data Science, to highlighting important hidden figures and unique challenges faced by underrepresented groups of people.

All materials are provided in electronic form for ease of use as-is or modified to meet the needs of your specific course.

1.1 Does the Kit Include GPU Compute Resources?

The Accelerated Data Science Teaching Kit includes access to ***free online DLI courses and certificate opportunities for students*** - a value of up to $90 per person per course (see "2.5 NVIDIA DLI Online Courses and Certificates" below).

The kit also includes Google Colab credits for you as an educator. Instructions on setting up Colab and redeeming the credits, or setting up your own AWS or other cloud environment, are linked in the next Section 1.2.

Students can also access CUDA-enabled GPU resources in other ways such as through local GPU cards or access to other remote clusters with GPUs provided by your school.

1.2 System Requirements and Environment Setup

Thanks to the development of NVIDIA GPUs and the RAPIDS framework, accelerated data science is more efficient than ever in terms of both time and resource cost. NVIDIA CUDA-enabled GPUs and RAPIDS are the most one needs to develop and run all the Teaching Kit labs.

All the information and requirements about what you need to get started with RAPIDS and GPUs can be found on the Getting Started with RAPIDS webpage.

Google Colab Environment Setup

Many of the Teaching Kit labs are designed around the Jupyter notebook style of interactive programming, utilizing the Google Colab platform to host the notebooks and provide the cloud-based GPU compute.

Google Colab and NVIDIA have teamed up to provide educators with the compute they need to teach the labs and exercises. As a qualified Teaching Kit educator, you are eligible to receive free Google Colab compute credits! Instructions for redemption can be found in *Free Google Colab Credits.pdf* in the Teaching Kit download package. Students can use the Colab free tier to access runtimes that include NVIDIA T4 GPUs and pre-installed RAPIDS zero code change libraries.

- RAPIDS on Google Colab

AWS and Other Cloud Environment Setup

- Amazon Web Services (AWS)
- Microsoft Azure
- Google Cloud

NOTE: *You should have received an email invitation to the Teaching Kit's private BitBucket repository that contains the most recent version of the lab solutions and associated data sets. Please contact NVDLI@nvidia.com if you have not yet received this email invitation or if it has expired.*

2. Accelerated Data Science Teaching Kit Content

Not all Teaching Kit content types apply to every module. For example, *Module 5: Data Integration* does not have a lab because it is a relatively small module.

2.1 Syllabus

The syllabus is available at https://developer.nvidia.com/data-science-teaching-kit-syllabus (also pointed to by *syllabus.html)*. It outlines the content for every module and links to the suggested online DLI courses and course sections that fit into the applicable modules. It also outlines modules coming in future releases of the Teaching Kit, and will also host the lecture

videos when they become available in future releases (see "2.4 Lecture Recordings" below).

2.2 Lecture Slides

The PowerPoint .ppt lecture slides are designed for instructor-led teaching. In future releases of the Teaching Kit, these files will also include embedded audio narration with examples of how you might present the slides (accessible in slideshow mode) to your students. This is also useful because students can watch the lectures on their own time, thereby adding a "flipped" classroom aspect to your course.

2.3 Lecture Recordings

You and your students are able to view available lecture videos linked from the online syllabus. These recordings contain the same content as the lecture slides with the embedded audio thereby adding a "flipped" classroom aspect to your course. More videos will become available in future releases of the kit.

2.4 Labs, Solutions and Data Sets

The labs/solutions are designed as one- to two-week hands-on programming assignments for students.

The *instructions* folder found in each module's lab folder in the Teaching Kit .zip file contains the lab instructional documents for students in either.pdf format, or .ipynb (Jupyter notebook) for more exploratory labs. Each of these documents usually begins with a description of the lab, objectives, prerequisites and detailed instructions. Some instruction documents also contain example point allocation for each section of the lab to provide an example grading rubric and an idea of how much relative time students can be expected to spend on each section.

The *solutions* folder found in each module's lab folder in the lab solutions repository contains at least one example implementation of approaches to solving the lab provided as .html, .py (Python) and .ipynb (Jupyter notebook) files. Some associated data sets are also found in the solutions folder.

The most recent version of the labs, solutions and data sets are in the Accelerated Data Science Teaching Kit Bitbucket repository. ***Additionally, the repository contains instructions for running the sample solution codes.***

2.5 NVIDIA DLI Online Courses and Certificates

The Teaching Kit includes access to **free** online DLI courses and certificate opportunities for your students – **a value of up to $90 per person per course**. DLI training reinforces AI, Data Science and Accelerated Computing concepts presented in the Teaching Kit and trains students on how to apply those concepts to end-to-end projects. Through built-in assessments, students can earn certificates that prove subject matter competency which can be leveraged for academic and professional career growth. Each course presents a self-paced learning environment with access to a GPU-accelerated workstation in the cloud. All students need is a web browser and Internet connection to get started. *Although these courses are designed to be taken online self-paced, you are free to administer them to your students in a live setting.*

The primary recommended course (with certificate opportunity) for students learning through this Teaching Kit is Accelerating End-to-End Data Science Workflows.

The online syllabus suggests students can also take sections of this course during your university semester course, or the entire full-day course upon near-completion of your semester course. Other online DLI courses are suggested in specific modules as well.

To enable these or any other online DLI courses for your students please send your developer.nvidia.com account email address to NVDLI@nvidia.com with subject line "DS Teaching Kit DLI Online Course Access". You will then receive information about how to give free access to your students.

Detailed descriptions of all available DLI courses can be found at www.nvidia.com/dli.

3. About the NVIDIA Deep Learning Institute (DLI)

The NVIDIA Deep Learning Institute (DLI) offers hands-on training for developers, data scientists, and researchers looking to solve challenging problems with deep learning and accelerated computing. Through built-in assessments, students can earn certifications that prove subject matter competency and can be leveraged for professional career growth.

Become a DLI Certified Instructor

Join the University Ambassador Program to teach DLI courses at your university to students, faculty, and researchers at no cost. Educators can apply at developer.nvidia.com/dli/cip.

Attend Instructor-led Training

In addition to online, self-paced courses, DLI offers all fundamentals and industry-specific courses as in-person workshops led by DLI-certified instructors. View upcoming workshops near you at www.nvidia.com/dli.