Fundamentals of Accelerated Data Science with RAPIDS

Whether you work at a software company that needs to improve customer retention, a financial services company that needs to mitigate risk, or a retail company interested in predicting customer purchasing behavior, your organization is tasked with preparing, managing, and gleaning insights from large volumes of data without wasting critical resources. Traditional CPU-driven data science workflows can be cumbersome, but with the power of GPUs, your teams can make sense of data quickly to drive business decisions.

In this Deep Learning Institute (DLI) workshop, developers will learn how to build and execute end-to-end GPU-accelerated data science workflows that enable them to quickly explore, iterate, and get their work into production. Using the RAPIDS accelerated data science libraries, developers will apply a wide variety of GPU-accelerated machine learning algorithms, including XGBoost, cuGRAPH's single-source shortest path, and cuML's KNN, DBSCAN, and logistic regression to perform data analysis at scale.

All workshop attendees get access to fully configured, GPU-accelerated servers in the cloud, guidance from a DLI-certified instructor, and the opportunity to network with other developers, data scientists, and researchers. Attendees can earn a certificate to prove subject matter competency and support professional growth.

Prerequisites:	Experience with Python, ideally including pandas and NumPy.
	To gain experience with pandas, we suggest this **pandas course** on Kaggle.
	To gain experience with data science using Python, we suggest this **machine learning course** on Kaggle.
	To get experience accelerating data science workflows, we suggest the **Accelerating Data Science Workflows with RAPIDS** course with DLI.
Technologies:	RAPIDS, cuDF, XGBoost, cuML, cuGraph, Dask, cuPy, pandas, NumPy, Bokeh, data science, data analytics, machine learning, deep learning
Price:	$10,000 for each group of 20 participants (excludes tax, if applicable)
Duration:	Approximately 8 hours

Learning Objectives

In this workshop, developers will learn how to:
> Implement GPU-accelerated data preparation and feature extraction using cuDF and Apache Arrow data frames
> Apply a broad spectrum of GPU-accelerated machine learning tasks using XGBoost and a variety of cuML algorithms
> Execute GPU-accelerated graph analysis with cuGraph, achieving massive-scale analytics in small amounts of time
> Rapidly achieve massive-scale graph analytics using cuGraph routines

Why DLI Hands-On Training?

> Build deep learning, accelerated computing, and accelerated data science applications for industries such as autonomous vehicles, healthcare, manufacturing, media and entertainment, robotics, smart cities, and more.
> Gain real-world expertise through content designed in collaboration with industry leaders, such as the Children's Hospital of Los Angeles, Mayo Clinic, PwC, and Uber.
> Access content anywhere, anytime with a fully configured, GPU-accelerated workstation in the cloud.
> Earn an NVIDIA DLI certificate to demonstrate subject matter competency and support career growth.
> Work with the most widely used, industry-standard software, tools, and frameworks.

Workshop Outline

Introduction (15 mins)	
GPU-Accelerated Data Manipulation (120 mins)	Ingest and prepare several datasets (some larger-than-memory) for use in multiple machine learning exercises later in the workshop. > Read data directly to single and multiple GPUs with cuDF and Dask cuDF. > Prepare population, road network, and clinic information for machine learning tasks on the GPU with cuDF.
Lunch (60 mins)	
GPU-Accelerated Machine Learning (120 mins)	Apply several essential machine learning techniques to the data that was prepared in the first section. > Use supervised and unsupervised GPU-accelerated algorithms with cuML. > Train XGBoost models with Dask on multiple GPUs. > Create and analyze graph data on the GPU with cuGraph.
Break (15 mins)	
Project: Data Analysis to Save the UK (120 mins)	Apply new GPU-accelerated data manipulation and analysis skills with population-scale data to help stave off a simulated epidemic affecting the entire UK population. > Use RAPIDS to integrate multiple massive datasets and perform real-world analysis. > Pivot and iterate on your analysis as the simulated epidemic provides new data for each simulated day.
Assessment and Q&A (15 mins)	

Next Steps

Connect with your NVIDIA contact to schedule an onsite workshop for your team, or submit your request at www.nvidia.com/requestdli and the DLI team will be in touch.

Related Training

If your organization is interested in applying accelerated data science to healthcare, we recommend the online, self-paced course **Data Science Workflows for Deep Learning in Medical Applications**. Your team will learn how to organize and augment a medical images dataset and validate these techniques by using a convolutional neural network (CNN). Get started here.

Additional Resources

DLI offers other hands-on training and educational resources in data science, deep learning, and accelerated computing, including:

> Self-paced, online courses on accelerated data science, deep learning, accelerated computing, and more at www.nvidia.com/dli

> Instructor-led workshops on deep learning for computer vision, multi-GPUs, healthcare, industrial inspection, robotics, intelligent video analytics, and more at www.nvidia.com/dli

> Blogs, webinars, and other resources on data science at www.nvidia.com/datascience