

# FUNDAMENTALS OF ACCELERATED COMPUTING WITH CUDA C\_C++

Duration: 1 day

Delivery Type: Classroom (Hands-on labs)

## OVERVIEW

This workshop teaches the fundamental tools and techniques for accelerating C/C++ applications to run on massively parallel GPUs with CUDA. Work your way through dozens of hands-on coding exercises using a live, GPU-enabled development environment in the cloud. Learn how to write code to be executed by a GPU accelerator, configure code parallelization using the CUDA thread hierarchy, manage and optimize memory migration between the CPU and GPU accelerator, leverage command line and visual profilers to guide your work, and utilize concurrent streams for instruction-level parallelism. Finish by implementing the workflow that you have learned on a new task — accelerating a fully functional, but CPU-only, particle simulator for observable massive performance gains. At the end of the workshop, you will have access to additional resources to create new GPU-accelerated applications on your own.

Upon successful completion of the workshop, participants will receive NVIDIA DLI Certification to recognize subject matter competency and support professional career growth.

## WHY DEEP LEARNING INSTITUTE HANDS-ON TRAINING?

- Learn how to build deep learning and accelerated computing applications across a wide range of industry segments such as autonomous vehicles, digital content creation, finance, game development, and healthcare
- Obtain guided hands-on experience using the most widely-used, industry-standard software, tools, and frameworks
- Gain real-world expertise through content designed in collaboration with industry leaders including the Children's Hospital Los Angeles, Mayo Clinic, and PwC
- Earn NVIDIA DLI Certification to demonstrate your subject matter competency and support professional career growth
- Access content anywhere, anytime with a fully-configured, GPU-accelerated workstation in the cloud

## PREREQUISITES

Basic C/C++ competency, including familiarity with variable types, loops, conditional statements, functions, and array manipulations. No previous knowledge of CUDA programming is assumed.

## COURSE OBJECTIVE

At the conclusion of the workshop, you will have an understanding of the fundamental tools and techniques for GPU-accelerating C/C++ applications with CUDA, and will be able to:

- Expose and express data and instruction-level parallelism in C/C++ applications using CUDA.
- Utilize CUDA managed memory and optimize memory migration using asynchronous prefetching.
- Write GPU-accelerated CUDA C/C++ applications, or refactor existing CPU-only applications, using a profile-driven approach.

## COURSE OUTLINE

	Components	Description
Introduction (15 mins)	<ul style="list-style-type: none"><li>• Getting started</li></ul>	Instructor introduction and environment setup
Accelerating Applications with CUDA C/C++ (120 mins)	<ul style="list-style-type: none"><li>• Write, compile, and run GPU code</li><li>• Control parallel thread hierarchy</li><li>• Allocate and free memory for the GPU</li></ul>	First steps with the CUDA programming paradigm: Learn the essential syntax and concepts to be able to write GPU-enabled C/C++ applications with CUDA.
Break (60 mins)		
Managing Accelerated Application Memory with CUDA C/C++ (120 mins)	<ul style="list-style-type: none"><li>• Profile CUDA code with command line profiler</li><li>• Go deep on Unified Memory</li><li>• Optimize Unified Memory management</li></ul>	Synergistic learning of the command line profiler and CUDA managed memory, with a focus on observation-driven application improvements and a deep understanding of managed memory behavior.
Break (15 mins)		
Asynchronous Streaming and Visual Profiling for Accelerated Applications with CUDA C/C++ (90 mins)	<ul style="list-style-type: none"><li>• Profile CUDA code with the NVIDIA Visual Profiler</li><li>• Use concurrent CUDA streams</li></ul>	Use the NVIDIA Visual Profiler to identify opportunities for improved memory management and instruction-level parallelism. Learn how to implement instruction-level parallelism with CUDA streams.
Break (15 mins)		
Assessment (30 mins)	<ul style="list-style-type: none"><li>• Accelerate a CPU-only nbody particle simulator</li></ul>	Leverage all your learning to accelerate a CPU-only nbody particle simulator, observing massive performance gains and earning certification.
Next Steps (15 mins)	<ul style="list-style-type: none"><li>• Workshop survey</li><li>• Setting up your own GPU-enabled environment</li><li>• Advanced project</li></ul>	Learn how to setup a CUDA and GPU-enabled environment to begin work on your own projects. Optionally, accelerate a Mandelbrot set application for practice.