Fundamentals of Accelerated Computing with CUDA C/C++

This workshop teaches the fundamental tools and techniques for accelerating C/C++ applications to run on massively parallel GPUs with CUDA®. You'll learn how to write code, configure code parallelization with CUDA, optimize memory migration between the CPU and GPU accelerator, and implement the workflow that you've 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'll have access to additional resources to create new GPU-accelerated applications on your own.

Duration:	8 hours
Price:	Contact us for pricing. During the workshop, each participant will have dedicated access to a fully configured, GPU-accelerated workstation in the cloud.
Assessment type:	Code-based
Certificate:	Upon successful completion of the assessment, participants will receive an NVIDIA DLI certificate to recognize their subject matter competency and support professional career growth.
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.
Languages:	English, Japanese, Chinese
Tools, libraries, and frameworks:	nvprof, nvpp

Learning Objectives

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

- Write code to be executed by a GPU accelerator
- 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
- Leverage command line and visual profilers to guide your work
- Utilize concurrent streams for instruction-level parallelism
- Write GPU-accelerated CUDA C/C++ applications, or refactor existing CPU-only applications, using a profile-driven approach

Why Deep Learning Institute Hands-On Training?

- Learn to build deep learning and accelerated computing applications for industries such as autonomous vehicles, finance, game development, healthcare, robotics, and more.
- Obtain hands-on experience with the most widely used, industry-standard software, tools, and frameworks.
- Gain real-world expertise through content designed in collaboration with industry leaders such as the Children's Hospital of Los Angeles, Mayo Clinic, and PwC.
- Earn an NVIDIA DLI certificate to demonstrate your subject matter competency and support career growth.
- Access content anywhere, anytime with a fully configured, GPU-accelerated workstation in the cloud.

Workshop Outline

TOPIC	DESCRIPTION
Introduction (15 mins)	> Meet the instructor. > Create an account at **courses.nvidia.com/join**
Accelerating Applications with CUDA C/C++ (120 mins)	**Learn the essential syntax and concepts to be able to write GPU-enabled C/C++ applications with CUDA:** > Write, compile, and run GPU code. > Control parallel thread hierarchy. > Allocate and free memory for the GPU.
Break (60 mins)	
Managing Accelerated Application Memory with CUDA C/C++ (120 mins)	**Learn the command line profiler and CUDA managed memory, focusing on observation-driven application improvements and a deep understanding of managed memory behavior:** > Profile CUDA code with the command line profiler. > Go deep on unified memory. > Optimize unified memory management.
Break (15 mins)	
Asynchronous Streaming and Visual Profiling for Accelerated Applications with CUDA C/C++ (120 mins)	**Identify opportunities for improved memory management and instruction-level parallelism:** > Profile CUDA code with the NVIDIA Visual Profiler. > Use concurrent CUDA streams.
Final Review (15 mins)	> Review key learnings and wrap up questions. > Complete the assessment to earn a certificate. > Take the workshop survey.

This content is also available as a self-paced, online course. Visit **www.nvidia.com/dli** for more information.