

## 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.

<b>Duration:</b>	8 hours
<b>Price:</b>	\$10,000 for groups of up to 20 (price increase for larger groups). During the workshop, each participant will have dedicated access to a fully configured, GPU-accelerated workstation in the cloud.
<b>Assessment type:</b>	Code-based
<b>Certificate:</b>	Upon successful completion of the assessment, participants will receive an NVIDIA DLI certificate to recognize their subject matter competency and support professional career growth.
<b>Prerequisites:</b>	Basic C/C++ competency, including familiarity with variable types, loops, conditional statements, functions, and array manipulations. No previous knowledge of CUDA programming is assumed.
<b>Languages:</b>	English, Japanese, Chinese
<b>Tools, libraries, and frameworks:</b>	Nsight Systems

### 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
<b>Introduction</b> (15 mins)	<ul style="list-style-type: none"> <li>&gt; Meet the instructor.</li> <li>&gt; Create an account at <a href="https://courses.nvidia.com/join">courses.nvidia.com/join</a></li> </ul>
<b>Accelerating Applications with CUDA C/C++</b> (120 mins)	Learn the essential syntax and concepts to be able to write GPU-enabled C/C++ applications with CUDA: <ul style="list-style-type: none"> <li>&gt; Write, compile, and run GPU code.</li> <li>&gt; Control parallel thread hierarchy.</li> <li>&gt; Allocate and free memory for the GPU.</li> </ul>
<b>Break</b> (60 mins)	
<b>Managing Accelerated Application Memory with CUDA C/C++</b> (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: <ul style="list-style-type: none"> <li>&gt; Profile CUDA code with the command line profiler.</li> <li>&gt; Go deep on unified memory.</li> <li>&gt; Optimize unified memory management.</li> </ul>
<b>Break</b> (15 mins)	
<b>Asynchronous Streaming and Visual Profiling for Accelerated Applications with CUDA C/C++</b> (120 mins)	Identify opportunities for improved memory management and instruction-level parallelism: <ul style="list-style-type: none"> <li>&gt; Profile CUDA code with Nsight Systems.</li> <li>&gt; Use concurrent CUDA streams.</li> </ul>
<b>Final Review</b> (15 mins)	<ul style="list-style-type: none"> <li>&gt; Review key learnings and wrap up questions.</li> <li>&gt; Complete the assessment to earn a certificate.</li> <li>&gt; Take the workshop survey.</li> </ul>

This content is also available as a self-paced, online course. Visit [www.nvidia.com/dli](https://www.nvidia.com/dli) for more information.