



Tutorial Title: Fundamentals of Accelerated Computing with CUDA C/C++

Tutorial Length: 4 Hours

Instructor Profile:

Dr. Ayaz ul Hassan Khan received his BS CS from NED, MS CS from LUMS and PhD in Computer Science and Engineering with the specialization in Parallel Computing from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia. He has fifteen years of experience in both industry and academics. He has hands-on experience in Network Programming, Embedded Systems Programming, and Parallel Programming. As part of his PhD dissertation work, he has got extensive experience in optimizing parallel programs for Graphics Processing Units (GPUs) using CUDA. He has explored several optimizations for CUDA programs that are helpful for the efficient implementation of parallel numerical algorithms for solving linear systems that is a key issue in the design of large scale simulation systems. In addition to that he has also got hands on experience in setting up GPU servers including the installations of hardware, operating system, remote access setup, run-time profilers, CUDA compiler and debugging tools. His current areas of interest include Parallel and Distributed Computing, High Performance Computing, Deep Learning and Big Data Analytics. He has published 8 journals and 10 conference papers/posters in the field of his research areas in recent years. Check out the following for details: <https://sites.google.com/site/ayazresearch/>

Tutorial Description:

This tutorial 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.

Upon successful completion of the assessment, participants will receive an NVIDIA DLI certificate to recognize their subject matter competency and support professional career growth.

Tutorial Format	Hands-on Coding using NVIDIA DLI Cloud Platform
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.
Tools, libraries, and frameworks:	nvprof, nvpp
Target Audience	Researchers, Students, Faculty Members, Software Developers
Content Level	Beginner



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:

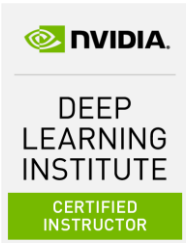
- > Write code to be executed by a GPU accelerator
- > Expose and express data and instruction-level parallelism in C/C++ applications using CUDA
- > Utilize CUDA-managed memory and optimize memory migration using asynchronous prefetching
- > Leverage command line and visual profilers to guide your work
- > 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?

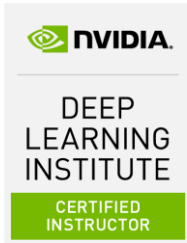
- > Learn to build deep learning and accelerated computing applications for industries such as autonomous vehicles, finance, game development, healthcare, robotics, and more.
- > Obtain hands-on experience with the most widely used, industry-standard software, tools, and frameworks.
- > Gain real-world 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 (10 mins)	<ul style="list-style-type: none">> Meet the instructor.> Create an account at courses.nvidia.com/join
Accelerating Applications with CUDA C/C++ (100 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">> Write, compile, and run GPU code.> Control parallel thread hierarchy.> Allocate and free memory for the GPU.



Break (15 mins)		
Managing Accelerated Application Memory with CUDA C/C++ (100 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"> > Profile CUDA code with the command line profiler. > Go deep on unified memory. > Optimize unified memory management. 	
Final Review (15 mins)	<ul style="list-style-type: none"> > 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.		



Dr. Ayaz ul Hassan Khan

(HEC Approved PhD Supervisor)
A-314 Block 2 Gulistan-e-Jauhar,
Karachi, Pakistan
Phone: +923318470261
ayazhk@gmail.com

Education

King Fahd University of Petroleum and Minerals PHD, Computer Science and Engineering, 3.875 CGPA	Sep 2010 - Dec 2014
Lahore University of Management Sciences MS, Computer Science (major: Networking and Distributed Systems), 3.61 CGPA	Sep 2003 - Dec 2004
NED University of Engineering and Technology BCIT, Computer Science and IT, 2 nd Position (90%)	Jan 1999 - Feb 2003

Additional Qualifications/Short Courses

- Oracle Certified Professional (OCP) DBA Track.
- Microsoft Certified Technology Specialist
- Strategic Time Management
- Strategic Visions
- Strategic Negotiation Skills
- Vision Retreat
- SAP: Introduction to ABAP Programming Course
- SAP Basics Course
- 4MAT Instructional Design Training
- Fundamentals of Accelerated Computing with CUDA C/C++ (NVIDIA Deep Learning Institute)
- DLI Platform Course for Instructors (NVIDIA Deep Learning Institute)

Teaching Experience

Assistant Professor, Habib University - Pakistan	Jan 2020 – Present
Assistant Professor, Karachi Institute of Economics and Technology - Pakistan	Jan 2018 – Jan 2020
Assistant Professor, Qassim University – Saudi Arabia	Aug 2015 - Jan 2018
Assistant Professor, FAST – NU - Pakistan	Aug 2009 – Jul 2010
Visiting Faculty, Mohammad Ali Jinnah University - Pakistan	Jun 2007 – May 2009
Assistant Professor, Mohammad Ali Jinnah University – Pakistan	Aug 2006 – Jun 2007

Taught Courses:

- Operating Systems
- Parallel Computing (MS/PhD Level)
- Parallel and Distributed Computing
- GPGPU Programming (MS/PhD Level)
- Concepts of Algorithms and Computer Programming



- Discrete Structures
- Distributed Systems and Parallel Processing
- Distributed Systems (MS Level)
- Computer Organization and Assembly Language
- Object Oriented Programming
- Advance Computer Programming
- Advance Operating Systems
- Programming Methodology

Conducted Trainings/Short Courses

- Strategic Time Management
- 4MAT Instructional Design
- CUDA Programming Workshop

Industry Experience

Software Engineer – Database Architecture/Application Development
TRG (The Resource Group) – Pakistan

Jul 2007 – Jul 2009

Software Engineer – Application Development
Wavetec Private Limited – Pakistan

Jan 2005 - Jul 2006