Fundamentals of Accelerated Computing with CUDA Python

This workshop teaches you the fundamental tools and techniques for running GPU-accelerated Python applications using CUDA® and the NUMBA compiler GPUs. You'll work though dozens of hands-on coding exercises, and at the end of the training, implement a new workflow to accelerate a fully functional linear algebra program originally designed for CPUs, observing impressive performance gains. After the workshop ends, you'll have additional resources to help you create new GPU-accelerated applications on your own.

Duration:	8 hours
Price:	$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.
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 Python competency, including familiarity with variable types, loops, conditional statements, functions, and array manipulations. NumPy competency, including the use of ndarrays and ufuncs. No previous knowledge of CUDA programming is required.
Languages:	English
Tools, libraries, and frameworks:	Numba, NumPy

Learning Objectives

At the conclusion of the workshop, you'll have an understanding of the fundamental tools and techniques for GPU-accelerated Python applications with CUDA and Numba:

> GPU-accelerate NumPy ufuncs with a few lines of code.

> Configure code parallelization using the CUDA thread hierarchy.

> Write custom CUDA device kernels for maximum performance and flexibility.

> Use memory coalescing and on-device shared memory to increase CUDA kernel bandwidth.

> Generate random numbers on the GPU.

> Learn intermediate GPU memory management techniques.

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**
Introduction to CUDA Python with Numba (120 mins)	> Begin working with the Numba compiler and CUDA programming in Python. > Use Numba decorators to GPU-accelerate numerical Python functions. > Optimize host-to-device and device-to-host memory transfers.
Break (60 mins)	
Custom CUDA Kernels in Python with Numba (120 mins)	> Learn CUDA's parallel thread hierarchy and how to extend parallel program possibilities. > Launch massively parallel custom CUDA kernels on the GPU. > Utilize CUDA atomic operations to avoid race conditions during parallel execution.
Break (15 mins)	
RNG, Multidimensional Grids, and Shared Memory for CUDA Python with Numba (120 mins)	> Use xoroshiro128+ RNG to support GPU-accelerated Monte Carlo methods. > Learn multidimensional grid creation and how to work in parallel on 2D matrices. > Leverage on-device shared memory to promote memory coalescing while reshaping 2D matrices.
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.