Summer School

Tyson Whitehead

May 9, 2018

Courses

	Option 1	Option 2	Option 3
Monday	Octave	Metagenome	Python
Tuesday	MPI	CUDA	R
Wednesday	MPI	CUDA	Cloud/Singularity
Thursday	OpenMP	C++	ParaView
Friday	Debugging/profiling	Fortran	sklearn/Tensorflow

Octave (Monday Option 1)

	Option 1	Option 2	Option 3
Monday	Octave	Metagenome	Python

Instructor

James Desjardins

Prerequisites

Attendees should install Octave 4.0 or later. Research computing experience is helpful. Familiarity with the Software Carpentry Octave/Matlab course material is sufficient to follow everything.

Description

This course explores generalizable strategies of using Octave on the Graham cluster to execute procedures established in Matlab. This training session also provides a hands on exploration of profiling tools for monitoring the performance of operations in Octave and compares different forms of simple operations.

Metagenome (Monday Option 2)

	Option 1	Option 2	Option 3
Monday	Octave	Metagenome	Python

Instructor: Armin Sobhani

Prerequisites

Basic Unix competence and basic knowledge of bioinformatics applications are required.

Description

In this one-day hands-on session, a typical metagenomics pipeline will be explored to introduce common tools used in bioinformatics analysis and how they can be utilized in an HPC environment.

Python (Monday Option 3)

	Option 1	Option 2	Option 3
Monday	Octave	Metagenome	Python

Instructor

Pawel Pomorski

Prerequisites

Some programming experience with Python will be useful. Experience with another language is also sufficient.

Description

This course will teach the student about various strategies to make a Python program run faster. It covers tools like NumPy, which provides highly efficient numerical libraries. For cases where a library is not available, the course will cover Cython and other tools which compile Python code. Some of the methods which make it possible to run parallel Python programs will also be briefly previewed.

MPI (Tuesday and Wednesday Option 1)

	Option 1	Option 2	Option 3
Tuesday	MPI	CUDA	R
Wednesday	MPI	CUDA	Cloud/Singularity

Instructors

Jemmy Hu and Ge Baolai

Prerequisites

Basic C/C++ and/or Fortran knowledge; experience editing and compiling code in a Linux environment.

Description

MPI is a standardized and portable message-passing interface for parallel computing clusters. In this two-day session, through lectures interspersed with hands-on labs, the students will learn the basics of MPI programming. Examples and exercises will be based on parallelization of common computing problems.

CUDA (Tuesday and Wednesday Option 2)

	Option 1	Option 2	Option 3
Tuesday	MPI	CUDA	R
Wednesday	MPI	CUDA	Cloud/Singularity

Instructor

Pawel Pomorski and Sergey Mashchenko

Prerequisites

C/C++ scientific programming, and editing and compiling code in a Linux environment. Experience with CUDA/OpenMP a plus.

Description

Introductory course to graphics processing unit (GPU) programming. The structuring of data and computations that makes full use of the GPU will be discussed in detail with many hands on examples. Some new features on Graham/Cedar GPUs will be covered. Sufficient for students to get started developing their own GPU applications.

R (Tuesday Option 3)

	Option 1	Option 2	Option 3
Tuesday	MPI	CUDA	R

Instructor

Marcelo Ponce

Prerequisites

Attendees should install R or Rstudio. Basic programming experience in any language.

Description

Introduction to R covering: the interface, primitive data types, lists, vectors, matrices, and data frames (crucial in data analysis). More advanced topics also covered include: basics statistics and function creation; *pply family functions; and basics of scripting. Time depending: data management strategies (i.e., saving results, workspaces and installing packages) and basic plotting.

Cloud/Singularity (Wednesday Option 3)

	Option 1	Option 2	Option 3
Wednesday	MPI	CUDA	Cloud/Singularity

Instructor

Edward Armstrong

Prerequisites

None.

Description

Traditionally, high performance clusters have restricted access to the internet at large due to security concerns. With the advent of cloud computing, users can now enable web access to their HPC work flows. This workshop will cover the basics of deploying a cloud instance to setting up your environment, and more advanced topics such as running a web server and setting up interactive connections between the cluster and the cloud.

OpenMP (Thursday Option 1)

	Option 1	Option 2	Option 3
Thursday	OpenMP	C++	ParaView

Instructor

Jemmy Hu

Prerequisites

Basic C/C++ and/or Fortran knowledge; experience editing and compiling code in a Linux environment.

Description

OpenMP is a compiler extension for parallel programming on shared memory systems. This course teaches the the key concepts through demos and hand-on exercises. The syllabus is: introduction, shared-memory model; cluster environment, compilers and tools; directives, latest specifications; synchronization, library routines, environment variables; and case studies.

C++ (Thursday Option 2)

	Option 1	Option 2	Option 3
Thursday	OpenMP	C++	ParaView

Instructor

Paul Preney

Prerequisites

C++98 procedural/object-oriented code experience or other intermediate/advanced object-oriented/concurrent code experience.

Description:

This course is targeted at researchers and students who are using C++. It will focus on understanding and leveraging a number of modern C++ language: move semantics, constant expressions (incl. compile-time computations), variadic template expressions (incl. C++17 fold expressions), using concurrency and parallel constructs, and calling Fortran code from C++ and vice versa.

ParaView (Thursday Option 3)

	Option 1	Option 2	Option 3
Thursday	OpenMP	C++	ParaView

Instructor

Tyson Whitehead and Weiguang Guan

Prerequisites:

Attendees should install ParaView. Some programming experience is helpful but not necessary.

Description:

ParaView is an open-source, multi-platform data analysis and visualization application. The first half of the course starts with basic usage and continuing up to temporal analysis and animation. The second half covers getting data into ParaView. This includes loading CSV, raw data, and converting to formats recognized by ParaView (e.g., VTK legacy formats and VTK XML-based formats).

Debugging/profiling (Friday Option 1)

	Option 1	Option 2	Option 3
Friday	Debugging/profiling	Fortran	sklearn/Tensorflow

Instructor

Sergey Mashchenko

Prerequisites

Basic knowledge of one or more parallel programming platforms (MPI, OpenMP, and/or CUDA).

Description

This course will introduce the tools for debugging and profiling parallel programs available in SHARCNET. Through a series of hands-on exercises we will learn to find and fix common bugs in MPI, OpenMP, and CUDA programs, and improve efficiency of a parallel code

Fortran (Friday Option 2)

	Option 1	Option 2	Option 3
Friday	Debugging/profiling	Fortran	sklearn/Tensorflow

Instructor

Baolai Ge

Prerequisites

Knowing a programming language (C/C++) and MATLAB helpful. Understanding numerical methods useful for following examples.

Description

Fortran has gained many modern features while still remaining a top candidates for scientific computing. This course covers: basics and the 2018 standard; compilers/development environments; data I/O; array slicing and loop constructs; random numbers; multiple source files; calling C/C++ and interoperability; coarrays and co-routines for easy parallel process on distributed memory systems; and more.

sklearn/Tensorflow (Friday Option 3)

	Option 1	Option 2	Option 3
Friday	Debugging/profiling	Fortran	sklearn/Tensorflow

Machine learning with sklearn and Tensorflow

Instructor

Weiguang Guan and Jose Nandez

Prerequisites

Helpful to know Python basics and some calculus and linear algebra.

Description

This course will look at machine learning through real-world problems such as predictive analytics, recognition of handwritten digits and classification of flowers on Graham. The first half will be 'regular' machine learning with sklearn and the second half deep learning with Tensorflow. Different network architectures and training parameters (for the best performance) will also be explored.