高性能与大数据计算培训
High-Performance and Data-Intensive Computing Training

刘奘 (Yan Liu) Ph.D.

National Center for Supercomputing Applications (NCSA)

Department of Geography and Geographic Information Science

University of Illinois at Urbana-Champaign

Urbana, Illinois, U.S.A.

Alumnus of Department of Computer Science (95' and 98')

Wuhan University

Organizing Committee

- School of Computer Science
  - Associate Professor Zhaohui Cai
  - Associate Professor Lian He
  - Professor Shubo Liu
  - Professor Chanle Wu
  - Professor Libing Wu
  - Dr. Gang Ye
- High-Performance Computing Center, Information Technology Services
  - Dr. Zhengguo Hong
  - Ming Xin
  - Dr. Meng Zhang
  - Yi Zhang
- School of Remote Sensing and Information Engineering
  - Professor Kun Qin
- School of Cyber Science and Engineering
  - Professor Jianming Fu

Goals

- Understand Supercomputing and Data-Intensive Computing
- Understand Parallelisms and Computational Thinking
- Understand Parallel Programming Models
  - Embarrassingly Parallel Computing
  - OpenMP
  - MPI
- Understand Programming Models for Data-Intensive Computing
  - Hadoop
  - Spark
- Understand Container Technologies
  - Use containers on HPC clusters
  - Use Jupyter for online interactive computing

Training Schedule

- A crash course on HPC and Big Data computing
- Dates: July 09-14 (6 days)
- Everyday we have an intensive schedule for lectures and hands-on exercises
- Lecture materials
  - Most of the training materials are extracted from public presentations and open source code packages
    - **Do not reinvent the wheel!**

Prerequisites

- Knowledge and skills
  - Basic understanding of concepts in operating systems and parallel and distributed computing
  - Familiar with Linux/Unix systems. Familiar with shell (e.g., *bash*, *csh*) scripting
  - Basic understanding of parallel computing
  - Familiar with C, C++, or Python programming
- Accounts on HPC cluster
  - Virtual cluster: 202.114.96.44

Day 1

09:00 - 09:50	Introduction to Supercomputing and Data-Intensive Computing
10:00 - 10:50	Hands-on: Accessing WHU Cluster Computing Environment
11:00 - 12:00	Resource Management and Job Submission
14:00 - 14:50	Hands-on: Cluster Computing 101
15:00 - 15:50	Embarrassingly Parallel Computing
16:00 - 17:00	Hands-on: Tools for Embarrassingly Parallel Computing

Day 2

09:00 - 09:50	Introduction to Scientific Workflows on HPC
10:00 - 10:50	Hands-on: Native Workflow Support in Job Schedulers
11:00 - 12:00	Exascale Computing: Opportunities and Challenges
14:00 - 14:50	Manycore and Multicore Computing
15:00 - 15:50	OpenMP Programming
16:00 - 17:00	Hands-on: OpenMP Programming

Day 3

09:00 - 09:50	Introduction to Message Passing Interface (MPI)
10:00 - 10:50	Hands-on: MPI basics
11:00 - 12:00	MPI Point-to-point Communication
14:00 - 14:50	Hands-on: MPI Point-to-point Communication Examples
15:00 - 15:50	MPI Collective Communication
16:00 - 17:00	Hands-on: MPI Collective Communication Examples

Day 4

09:00 - 09:50	MPI Non-blocking Communication
10:00 - 10:50	Hands-on: Blocking vs. Non-blocking Communication
11:00 - 12:00	MPI IO
14:00 - 14:50	Hands-on: MPI IO exercises
15:00 - 15:50	MPI One-sided Communication
16:00 - 17:00	Other Advanced MPI Features

Day 5

09:00 - 09:50	Introduction to Big Data Computing
10:00 - 10:50	Hands-on: MapReduce Programming
11:00 - 12:00	Introduction to Spark
14:00 - 14:50	Hands-on: Spark Basics
15:00 - 15:50	Spark Programming
16:00 - 17:00	Hands-on: Spark for Big Data Analysis

Day 6

09:00 - 09:50	Container Technologies
10:00 - 10:50	Hands-on: Docker Container Exercises
11:00 - 12:00	Containers on HPC
14:00 - 14:50	Hands-on: Singularity
15:00 - 15:50	Jupyter: Bridging HPC and Researchers for Reproducible and Interactive Analysis
16:00 - 17:00	Hands-on: Jupyter Environment for Interactive Computing

Accounts on WHU HPC Cluster

- Virtual cluster for this training
  - 202.114.96.44
  - Get your account from WHU HPC team
- Personal computer requirements
  - An *ssh* client (with X capability preferred)
  - An *scp* client for file transfer

Thanks!

Q&A