SDSC

Cyberinfrastructure-Enabled Machine Learning
Summer Institute

June 18, June 22-25 2021

CIML SI21 Main Sessions
Logistics and Introductions
Mary Thomas

SDSC SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego

Welcome to the FIRST CIML Summer Institute!

- Focus is on scalable machine learning.

- Great response:
  - Over 120 applicants; 90+ institutions; ~50% grads, 35% fac/staff; 28% MSI
  - 50+ invited; 40+ attending

- Event Website: https://na.eventscloud.com/website/22773/home/

- GitHub: https://github.com/ciml-org/ciml-summer-institute-2021

- Please be on time so we can stay on schedule.

What is CIML?

- NSF CyberTraining Grant: *Developing a Best Practices Training Program in Cyberinfrastructure-Enabled Machine Learning Research (CIML)*
- Objectives: Scalable Machine Learning
  - To create generalized machine learning training and project materials that run on large-scale NSF funded cyberinfrastructure resources such as XSEDE
  - Targeted towards researchers and educators who are using machine learning (ML) and big data analytics methods for their domain specific applications or instructional material
  - To develop a community of machine learning and data analytics CI Users (CIU) and Contributors (CIC) who actively contribute to the training material repository and incorporate the materials into their projects and courses.
  - Synthesize the training material into a domain independent CIML workflow system that can be used for creating applications that run on the NSF HPC ecosystem.

Logistics

- Friday, June 18th was "Prep Day"

- We focussed on making sure you can connect to Expanse, run jobs, launch notebooks.

- We will use Slack for chatting/communicating

- We will use Zoom for
  - All presentations and group discussions
  - Breakout rooms for hands-on sessions
  - To avoid Zoom fatigue, we'll have lots of breaks

- When speakers have 5 mins left, we will make a clicker sound (demo)

- WebSite: https://na.eventscloud.com/website/22773/home/

- GitHub: https://github.com/ciml-org/ciml-summer-institute-2021

Day 1 Agenda:
Cyberinfrastructure & HPC (Tues, 06/22/21)

TIME (PST)	TOPIC	PRESENTER
8:00 AM - 9:00 AM	2.1 Welcome, Orientation, & Introductions	Mary Thomas
9:00 AM – 9:40 AM	2.2 Introduction HPC/Cyberinfrastructure	Robert Sinkovits
9:40 AM – 10:20 AM	2.3 CPU Computing - Hardware, architecture, and software infrastructure	Mary Thomas
10:20 AM – 11:00 AM	2.4 GPU Computing - Hardware architecture and software infrastructure	Andreas Goetz
11:00 AM – 11:45 AM	Break/Lunch	
11:45 PM – 12:25 PM	2.5 Data Management and File Systems	Manu Shantharam
12:25 PM – 1:05 PM	2.6 Introduction to Singularity: Containers for Scientific and High-Performance Computing	Marty Kandes
1:05 PM – 1:35 PM	2.7 Reproducibility in Science and Machine Learning	Peter Rose

CIML Instructors

Andreas Goetz, Ph.D.
Director of Computational Chemistry Laboratory

Marty Kandes, Ph.D.
Computational and Data Science Research Specialist

Mai Nguyen, Ph.D.
Lead for Data Analytics

Paul Rodriguez, Ph.D.
Research Analyst

Peter Rose, Ph.D.
Director of Structural Bioinformatics Laboratory

Manu Shanthanam, Ph.D.
Senior Computational Scientist

Robert Sinkovits, Ph.D.
Director of Scientific Computing Applications

Mary Thomas, Ph.D.
Computational Data Scientists, HPC Trainer

Let's get to know each other

1. Name

2. Institution/Company & Department

3. How do you like to spend your time when not at work?

4. What have you binged watched or read?

Basic Information

- Expanse User Guide:
  - https://www.sdsc.edu/support/user_guides/expanse.html

- You need to have an Expanse account in order to access the system. There are a few ways to do this:
  - Submit a proposal through the XSEDE Allocation Request System
  - PI on an active allocation can add you to their allocation (if you are collaborators working on the same project).
  - Request a trial account, instructions @ https://portal.xsede.org/allocations/startup.

- Online repo and information:
  - https://github.com/sdsc-hpc-training-org/expanse-101
  - https://hpc-training.sdsc.edu/expanse-101/

Resources

- Expanse User Guide
  - https://www.sdsc.edu/support/user_guides/expanse.html
- GitHub Repo for this webinar: clone code examples for this tutorial – clone example code:
  - https://github.com/sdsc-hpc-training-org/expanse-101
- SDSC Training Resources
  - https://www.sdsc.edu/education_and_training/training
  - https://github.com/sdsc-hpc-training/webinars
- XSEDE Training Resources
  - https://www.xsede.org/for-users/training
  - https://cvw.cac.cornell.edu/expanse/

We hope you all have a great workshop!