

Welcome to TACC Summer Machine Learning Institute: Overview

Weijia Xu

Research Scientist, Group Manager
Scalable Computational Intelligence
Texas Advanced Computing Center

The University of Texas at Austin

August, 2021

TACC AT A GLANCE



- **Personnel**
 - 150 Staff (~70 PhD)
- **Facilities**
 - 12 MW Data center capacity
 - Two office buildings, Three Datacenters, two visualization facilities, and a chilling plant.
- **Systems and Services**
 - A Billion compute hours per year
 - 5 Billion files, 50 Petabytes of Data, Hundreds of Public Datasets
- **Capacity & Services**
 - HPC, HTC, Visualization, Large scale data storage, Cloud computing
 - Consulting, Curation and analysis, Code optimization, Portals and Gateways, Web service APIs, Training and Outreach



Machine Learning & HPC

- High Performance Computing empowers machine learning
- Mass storage
 - More data available
 - More past “experience”
- Faster computations
 - More and faster memory
 - Make complex solution practical through scaling
 - Improve performance
 - Enable online learning

Scalable Computational Intelligence Group (SCI) @ TACC

- Machine learning and data analysis support
 - R (Rstudio), Python (Jupyter, DASK), Scala (Zeppelin)...
 - Scikit-learn, Tensorflow, Keras, Pytorch ...
 - Hadoop, Spark, Hive, Hbase, Drill, Storm...
 - NLP, Elastic Search, Solr,
- Research & Development :
 - Big data analysis
 - High performance machine learning
 - Data science and domain collaboration

Main Aims of Institute

- Introduction of basic tools and methods for machine learning and data analysis.
- Resources and support at TACC for machine learning.
- How to access and use resources at TACC to solve practical problems.

Instructors



Amit Gupta (Mon)



Kelly Pierce (Tue)



Weijia Xu (Wed)



David Walling (Mon, Thur.)



Zhao Zhang (Friday)



Charlee Dey

Schedules at a Glance

- Institute page: shorturl.at/amzCG
 - Check for updates and additional information throughout the week.
- Day 1 – introduction and getting started with TACC
- Day 2 – Supervised Learning methods and examples
- Day 3 – Unsupervised Learning methods and examples
- Day 4 – Deep Learning methods and examples
- Day 5 – Saleable deep learning and consulting.

Institute Format

- Two 90 minutes zoom session each day
 - Morning Session 9:30 ~ 11:00
 - Afternoon Session 14:00 ~15:30
 - <https://zoom.us/j/98431479273?pwd=ZzVUSWsvcjVYenRpMDhOWIU5bDA3dz09>
- Break and Practice
 - Resource reservation runs from 10:00 to 16:00 each day
 - Frontera, 8th fastest machine in the world.

Frontera Supercomputer

- Primary compute system:
 - 39PF PetaFlops Peak Performance --
8,000+ nodes of Intel Cascade Lake
 - Mon, Tue, Wed,
- Storage: DataDirect Networks
 - 50+ PB disk, 3PB of Flash, 1.5TB/sec peak I/O rate.
- Front end for data movers, workflow, API
- Two subsystems with GPU:
 - Longhorn 112 IBM Power9 nodes with 4 V100 GPU card each.
 - 360 node with four RTX5000 GPU each
- Thur., Fri.



Today

- Morning session focuses on getting started with TACC resources
 - Computing resources at TACC
 - Basic system environments on the cluster
 - How to access TACC resources remotely
- Two parallel sessions in the afternoon
 - Introduction of Python programming language and using Jupyter Notebook on TACC resources
 - <https://zoom.us/j/98431479273?pwd=ZzVUSWsvcjVYenRpMDhOWIU5bDA3dz09>
 - Introduction of R programming language and using RStudio on TACC resources **Starting at 1:30 today**
 - <https://zoom.us/j/92700907989?pwd=aVlkOVFnTUwyOHUyUGpiRWVrRUIFUT09>