Tutorial Agenda

Ian Foster, University of Chicago and Argonne National Lab
Dennis Gannon, School of Informatics and Computing, Indiana University, Bloomington Indiana
Vani Mandava, Microsoft Research, Redmond Washington

Part 1. The Cloud and Interactive Scientific Discovery

Starting at 8:30

- 0.5 hour. Introduction and Startup.
  - Create cloud accounts.
  - Download additional tutorial materials
  - Download and install storage explorer.
- 1.0 hour. Orienting in the Cloud Universe
  - Public vs private vs research clouds.
  - Cloud portals and SDKs (AWS, Azure, Google, JetStream)
  - Storage systems: blob stores including S3, Azure blob storage, OpenStack Swift, SQL and NoSQL storage including Google Big Table, AWS DynamoDB, AWS RDS, Azure DocumentDB
  - Hands-on lab 1. Create a storage account and upload a file with the storage explorer.
  - Hands-on lab 1b if lab 1 for the python jupyter folks.
- 0.5 hour. Break
- 1 hour. Virtual Machine and Containers
  - Compute Infrastructure: Virtual Machines and how to launch them and attach storage
  - Hands-on lab 2. Install the Microsoft Linux Data Science VM
  - Introduction to Containers
- 12:00 Lunch

Part 2. Scaling Science in the Cloud

This section focuses on higher level services in the cloud.

1:00 pm tutorial resumes.

- 1.5 hours. Parallelism in the cloud (discussion and demo)
  - Map Reduce
  - Spark and Hadoop
  - Hands-on lab 3. A simple Spark exercise: K-means clustering
  - Kubernetes and Mesos and container services.
  - Microservice concepts and demo
  - Serverless functions and lab 4
- ½ hour break at 2:30.
- 1 hour. Data Analytics
  - Survey discussion
  - Hands-on lab part 2 of lab 3
- 1 Hour. Machine learning and event stream analysis.
  - Survey discussion
  - AzureML and Hands-on lab 5.