**High Throughput, Cluster and Grid Computing Outline**

1. Introduction: High Throughput, Cluster and GRID computing
   1. Previous computer experience
   2. Overview of distributed computing and its uses
   3. Linux operating system
   4. Working at the command line
   5. Bash - Either review or new
2. Building a cluster from the ground up
   1. Installing Linux on one node
   2. Installing Linux on many nodes
3. Configuration and Management of one to many nodes
   1. Network Installation
   2. Puppet
   3. Cobbler
4. Topics include:
   1. Hierarchy of a cluster – Head node and compute nodes
   2. Networking, package management, updating
   3. Security
   4. Cluster monitoring - Nagios, Ganglia
5. Virtual Machines and Containers
   1. Problem: Code compiled on one platform and you want to run elsewhere
   2. KVM - Virtual machines
   3. Singularity - Containers
6. File Systems
   1. Local file systems
   2. Distributed file systems -
   3. NFS - File based distributed file system
   4. Hadoop - Block based distributed file system
   5. CVMFS - Readonly global file system
7. Batch Systems
   1. Installing a batch system on one node
   2. Batch system => More efficient system usage
   3. Condor Batch
   4. SLURM Batch
8. Introduction to Jupyter Notebooks
9. Using Python and Jupyter Notebooks
10. Using R and Jupyter Notebooks
11. Accessing data in Jupyter Notebooks
12. Running your first application
13. Running locally
14. Running remotely
15. Managing UIDs
16. Introduction to batch systems
17. CPU bound
18. I/O bound
19. Introduction to XSEDE
20. Packaging applications for remote execution
21. How to get the results
22. What happens when it fails
23. Kaggle - <https://www.kaggle.com/>
24. Stanford Large Network Dataset Collection - <http://snap.stanford.edu/data/>
25. Running the final project on XSEDE

Kaggle - https://www.kaggle.com/

High demand computing

Scaling to many node

Ray tracing on a cluster

Animation Graphics processing

Weather data processing

Traffic flow simulation and analysis

Creating a portable workflow

Pegasus

Payloads

Running on someone else’s cluster