How to Use a Supercomputer

Shelley Knuth
shelley.knuth@colorado.edu

Peter Ruprecht
peter.ruprecht@colorado.edu

www.rc.colorado.edu

Outline

- Today we will discuss:
  - Who Research Computing is
  - Resources provided
  - Accessing resources - Stampede
  - Data Storage and access

What does Research Computing do?

- We manage
  - Shared large scale compute resources
  - Large scale storage
  - High-speed network without firewalls – ScienceDMZ
  - Software and tools
- We provide
  - Consulting support for building scientific workflows on the RC platform
  - Training
  - Data management support in collaboration with the Libraries

How Does CU-RC Compare to Other HPC Centers?

- Generally, we're a bit smaller than most supercomputing centers!
- Staff of ~10-15 (including students)
- Operate the Janus supercomputer, a large data repository, and other large scale computing resources
- Have 700 users across 30 institutes and departments on campus

SUPERCOMPUTER RESOURCES

What Is a Supercomputer?

- Many supercomputers are one large computer made up of many smaller computers and processors – a "cluster"
- With a supercomputer, all these different computers talk to each other through a communications network
  - On Janus and Stampede – InfiniBand
- Each different computer is called a **node**
- Each node has processors/cores
  - Carry out the instructions of the computer

Why Use a Supercomputer?

- Supercomputers give you the opportunity to solve problems that are too complex for the desktop
  - Might take hours, days, weeks, months, years
  - If you use a supercomputer, might only take minutes, hours, days, or weeks
- Useful for problems that require large amounts of memory

Computers and Cars - Analogy

≈

Computers and Cars - Analogy

≈

World's Fastest Supercomputers

www.top500.org

Rank	Site	Name	TeraFlops
1	National Super Computer Center (Guangzhou, China)	Tianhe-2	54902.4
2	Oak Ridge National Laboratory (United States)	Titan	27112.5
3	DOE/NNSA/LLNL (United States)	Sequoia	20132.7
4	RIKEN Advanced Institute for Computational Science (Japan)	K	11280.4
5	DOE/Argonne National Lab (United States)	Mira	10066.3
6	Swiss National Supercomputing Centre (Switzerland)	Piz Daint	7788.9
7	Texas Advanced Computing Center (United States)	Stampede	8520.1
8	Forschungszentrum Juelich (Germany)	JUQUEEN	5872.0
9	DOE/NNSA/LLNL (United States)	Vulcan	5033.2
10	Government (Undisclosed) (United States)	Undisclosed	3143.5

What Does It Mean to Be Fast?

- Titan can do 27 trillion calculations per second
  - A regular PC can perform 17 billion per second
  - Not enough for some people!
- Size of a basketball court
- Computers this size require lots of electricity and are expensive to run and maintain!

Stampede

- 6400 nodes
- 250 TB of aggregate memory
- The nodes have 2 Intel Xeon E5 processors and an Intel Xeon Phi Coprocessor
  - 32 GB memory
- 16 Large memory nodes with 32 cores/node and 1 TB of memory

Credit: TACC

Stampede

- 128 compute nodes have single GPU
  - Designed for visualization
- Connected to 14 PB file storage
- Tape archive system with 60 PB capacity
- Designed to provide large memory, large data transfer, and GPU capabilities
- Nodes connected by Infiniband

Stampede

- 4 Login nodes
- Three locations for storage
  - $HOME, $WORK, $SCRATCH
  - Lustre
- Modules are used to access software available on the systems

Supercomputing Resources

- With appropriate connections, researchers at USGS can get access to some of the large supercomputers around the world
- The Extreme Science and Engineering Discovery Environment (XSEDE) can provide assistance with setting up accounts (Janice and Jeff are “Campus Champions”)
- Today we will use Stampede

Supercomputer Applications

- Climate modeling
- Predict seismic activity
- Forecast hurricanes
- Mapping the universe

EXAMPLE

Login and Modules example

- **Log in:**
  - Login to XSEDE User Portal (portal.xsede.org)
  - Go to accounts and access specific machines from User
  - Can also access from command line

    ```
    ssh username@stampede.tacc.xsede.org
    ```
  - Talk to Campus Champion about getting accounts

- **List and load modules**
  - `module list`
  - `module avail`
  - `module load netcdf/4.2.1.1`
  - `module load ncview/2.1.1`

DATA STORAGE AND TRANSFER

Storage Spaces - Stampede

- **Home Directories**
  - Store source code
  - Not for direct computation
  - 5 GB quota on stampede
  - Backed up
- **$WORK Space**
  - 400 GB quota
  - Start job runs from here
  - Large file storage
  - Not backed up

Storage Spaces – Stampede – Continued

- **Lustre Parallel Scratch Filesystem**
  - 8.5 PB
  - Output from running jobs should go here
  - Store large files briefly
  - Files purged after 10 days

Keeping Lustre Happy

- Janus Lustre is tuned for large parallel I/O operations.
- Creating, reading, writing, or removing many small files simultaneously can cause performance problems.
- Don’t put more than 10,000 files in a single directory.
- Avoid “ls –l” in a large directory.
- Avoid shell wildcard expansions (*) in large directories.

Data Repositories

- New initiatives from funding agencies to make data publicly available
- Recent push in development of data repositories across the globe
- Wide range – for geosciences:
  - National Water Information System
  - Earth Resources Observation and Science
  - South Florida Information Access

http://www.usgs.gov/datamanagement/preserve/repositories.php

Research Data Storage: PetaLibrary

- NSF Major Research Instrumentation grant
- Long term storage option
- Keep data on spinning disk or tape
- Provide expertise and services around this storage
  - Data management
  - Consulting
- No HIPAA, FERPA data
- Infrastructure guaranteed for 5 years

Data Sharing and Transfers

- ***Globus tools***: Globus Online and gridftp
  - https://www.globus.org
  - Web-mediated drag-and-drop transfers
  - Extremely efficient for large transfers to distant locations
  - Easier external access for collaborators
- ***SSH***: scp, sftp, rsync
  - Adequate for smaller transfers

Questions?

- More examples to come after presentations!

Peter.Ruprecht@colorado.edu
Shelley.Knuth@colorado.edu