



HPCC New User Training

Getting Started on HPCC Resources

(Part 1/2)

Misha Ahmadian

High Performance Computing Center

Summer 2022



- HPCC Training Courses
 - Please check the website for upcoming User Training workshops
 - <http://www.depts.ttu.edu/hpcc/about/training.php>
- ShortCourse Survey
 - Looking forward to have your feedback on this Training Workshop
 - *You will receive a survey in your inbox from TTU ShortCourse*
- The PowerPoint slides are available online
 - <http://www.depts.ttu.edu/hpcc/about/training.php>

About this Training Workshop?



TEXAS TECH UNIVERSITY
Information Technology Division

Hands-on training that covers all you need to know about HPCC RedRaider Cluster:

- Who are the audiences:
 - ✓ HPCC Users (TTU Students/Faculties/Staff/External Researchers)
 - ✓ None-HPCC users (TTU Students/Faculties/Staff/ External Researchers)
- Is this training sufficient?
 - Could be enough for intermediate to advance users
 - We encourage beginners to review the slides along with the online user guide documents: <https://www.depts.ttu.edu/hpcc/userguides/index.php>

About this Training Workshop?



TEXAS TECH UNIVERSITY
Information Technology Division

Course Schedule:

- The HPCC New User Training is offered twice per semester.
- Each session will be held for 4 hours per day (10 am – 12 pm) and (1 pm – 3 pm) with a 1-hour lunch break.

Preferred Requirements:

- Attendees are preferred to have their HPCC account ready:
 - If you did not request or receive your new HPCC account, please stay with us!
- Windows/Linux/Mac laptop or desktop

About this Training Workshop?



TEXAS TECH UNIVERSITY
Information Technology Division

How to make the best out of this Training?

- Take your notes during the sessions, but not too many!
- Follow the instructor and try the commands and examples under your HPCC accounts.
- A few exercises will be provided during each session, which will help you to practice and learn.



Part 1:

- ❖ Introduction to High Performance Computing
- ❖ HPCC Resources
- ❖ Logging and using the Clusters
- ❖ Transferring Data



Introduction to High Performance Computing



Simplest Programming Model: Serial Computing



TEXAS TECH UNIVERSITY
Information Technology Division

- Software is often written and optimized to run **serially**.
 - Execution occurs on a single computer using a single CPU core.
 - A problem is broken down into a linear series of instructions.
 - Instructions are executed one after another.
 - Only one instruction may execute at any given time.
 - While simple, this model may not make full use of modern multi-core processors.

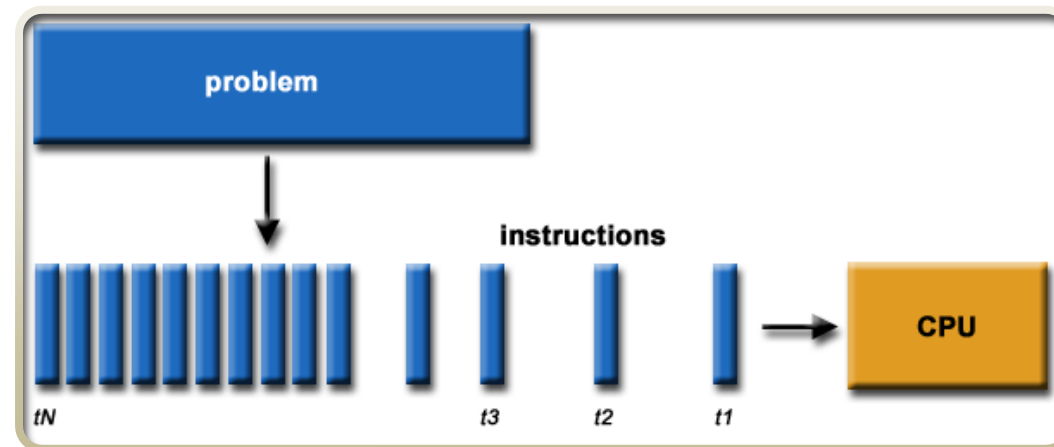
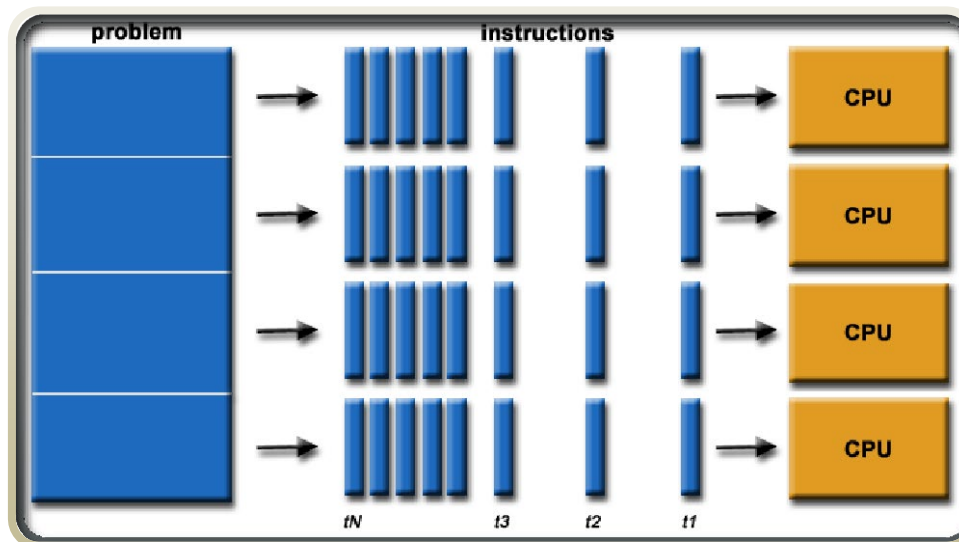


Image provided by Lawrence Livermore National Labs (computing.llnl.gov)



More Advanced Usage: Parallel Computing

- **Parallel Computing** is the simultaneous use of multiple compute resources to solve a computational problem.
 - Execution occurs across multiple CPU cores.
 - A problem is broken into discrete parts that can be solved concurrently.
 - Each part is further broken down into a series of instructions, executed one after another.
 - Instructions from each part execute simultaneously on different CPU cores.



Classes of Parallel Computers



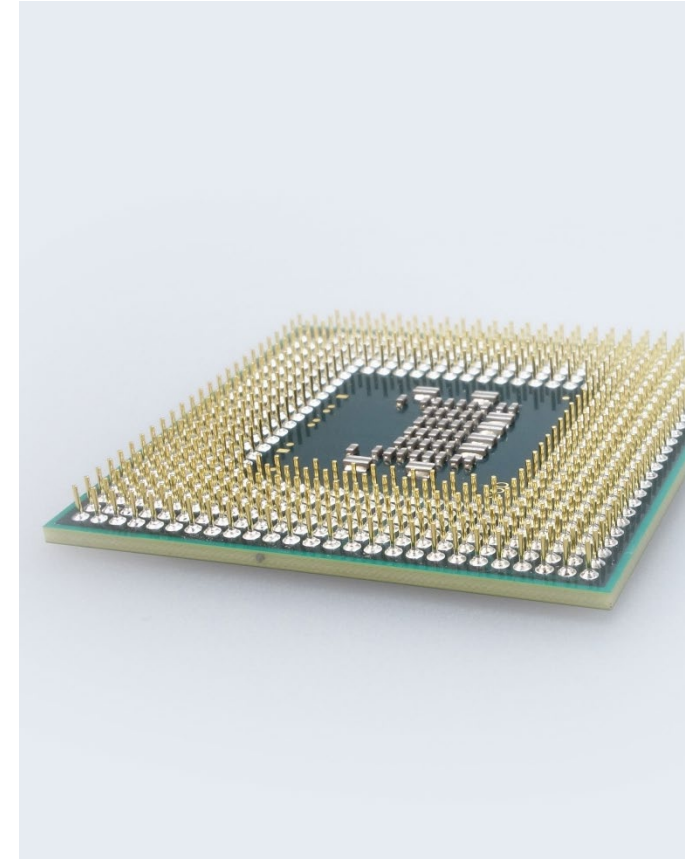
TEXAS TECH UNIVERSITY
Information Technology Division

- **Multi-core Computing:**

- Multi-core processors contain multiple ‘processing units’ (called cores) on a single chip.
- Allows for parallel execution across cores – each able to reach the same system resources (RAM, Keyboard, Monitor, etc...).

- **Symmetric Multiprocessor (SMP):**

- A symmetric multiprocessor is a computer system with multiple identical processors.
- Each processor likely has multiple cores
- Allows for parallel execution across cores – each able to reach the same system resources (RAM, Keyboard, Monitor, etc...).



Classes of Parallel Computers



TEXAS TECH UNIVERSITY
Information Technology Division

- **Clusters:**
 - Groups of loosely coupled computers working together closely.
 - Processes can be spread across multiple nodes, but processes are unable to reach the same system resources (RAM, Keyboard, Monitor, etc...).
- **Massively Parallel Processors (MPP):**
 - A group of tightly coupled computers working together closely across a specialized high-speed interconnect.
 - Processes can be spread across multiple nodes, but processes are unable to reach the same system resources (RAM, Keyboard, Monitor, etc...).
 - Common computing paradigm for campus-based HPC centers.



Classes of Parallel Computers



TEXAS TECH UNIVERSITY
Information Technology Division

- **National-Scale Supercomputers:**

- Highly scaled forms of parallel computing run for the National Science Foundation, Department of energy, National Institutes of health, etc.
- Organized much like larger versions of the TTU HPCC clusters.
- Awards for time require proposals and are evaluated on a competitive basis.

- **Grid and Cloud Computing:**

- Highly distributed forms of parallel computing.
- Clusters or single resources are spread across multiple sites using the Internet for connectivity.
- Commercial clouds are often more expensive to use than dedicated fully-utilized on-premises clusters for HPC.





Why does programming style matter?

- Different problems are suited to each of the major programming models.
 - **Serial programming:**
 - Executes serially using a single core/thread
 - “*Single-core machines*”
 - Good for problems that don't require inter-thread or inter-process communication.
 - **Multi-core / Multi-threaded Programming:**
 - Executes in parallel using multiple cores/threads
 - All threads are running on the same machine and access the same RAM
 - “*Multicore & Symmetric Multiprocessing*”
 - Needed for problems that require different threads or processes to share information.
 - **Massively Parallel / Distributed Programming:**
 - Executes in parallel using multiple machines
 - “*Clusters, Massive Parallel Processors, & Grid/Cloud*”
 - Needed for problems that require harnessing multiple large-scale resources.

Match Your Programming Style To The Problem!



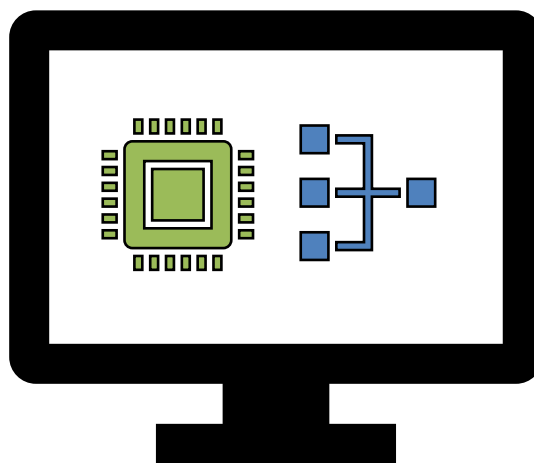
TEXAS TECH UNIVERSITY
Information Technology Division

- **If your program is not written to use a certain model, it will not necessarily “*just work*” in that model.**
 - Running serial code on a 128-core machine will use 1 core and leave 127 cores sitting idle.
 - Attempting to run multi-threaded code written to be used in just one node across 10 nodes instead without adjustments will result in 1 node being overutilized and 9 nodes sitting idle.
- **Not all multi-threading/MPP is equal!**
 - Try to understand how your program works at a small scale before attempting to “*scale up*”.
 - Keep in mind that programming language, developer decisions and even user input data can greatly alter how well an application scales.
 - Many existing codes need to be tuned or configured to run optimally.





HPCC Resources



HPCC Resources: RedRaider Primary Cluster



TEXAS TECH UNIVERSITY
Information Technology Division

- **RedRaider Cluster Commissioned in 2021:**

- **Nocona partition (240 CPU nodes)**

- 2x AMD EPYC ROME 7702 processors / node
- 30,720 Cores (128 cores/node)
- 120 TB total RAM (512 GB/node)
- HDR 200 Gbps InfiniBand fabric
- 804 Teraflops (81.4% efficiency)

- **Matador partition (20 GPU nodes)**

- 40 NVIDIA Tesla V100 GPUs (2 V100 / node)
- 2x Intel Xeon Cascade lake 6248 processors / node
- 800 CPU Cores (40 cores/node)
- 7.5 TB total RAM (384 GB/node)
- HDR 100 Gbps InfiniBand fabric
- 226 Teraflops (80.6% Efficiency)



HPCC Resources: RedRaider Primary Cluster



TEXAS TECH UNIVERSITY
Information Technology Division

- **Toreador partition (11 nodes)**
 - 33 NVIDIA Tesla A100 GPUs (3 A100 / node)
 - 2x AMD EPYC ROAM processors / node
 - 176 CPU Cores (16 cores/node)
 - 2.1 TB total RAM (192 GB/node)
 - HDR 100 Gbps InfiniBand fabric
 - *May expect higher waiting time*

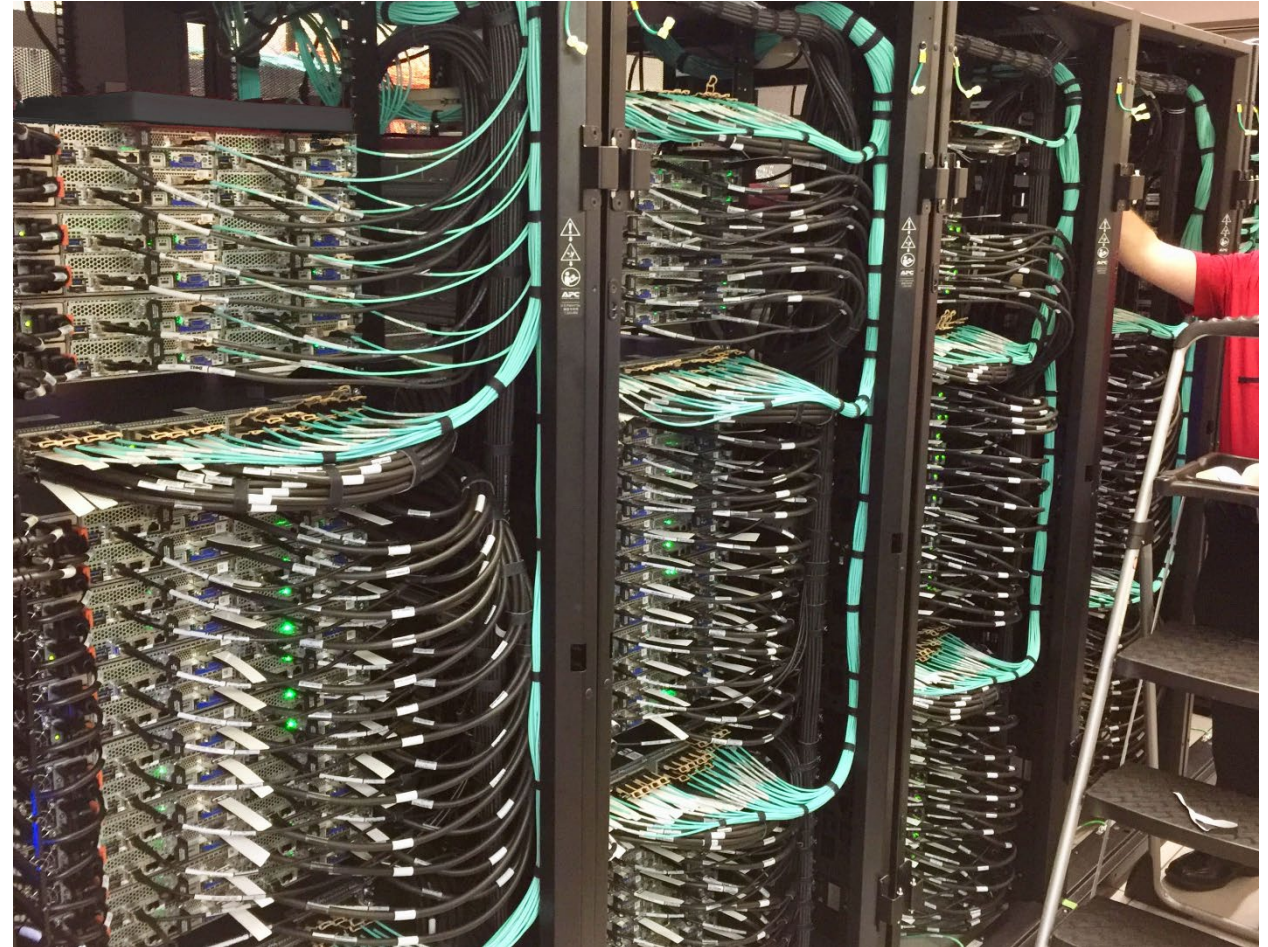


HPCC Resources



TEXAS TECH UNIVERSITY
Information Technology Division

- **Quanah partition (467 nodes)**
 - Commissioned in 2017
 - 2x Intel Xeon E5-2695v4 Broadwell Processors/node
 - *16,812 total cores* (36 cores/node)
 - 87.56 TB total RAM (192 GB/node)
 - Non-blocking Omni-Path (100 Gbps) fabric
 - Benchmarked at **485 Teraflops**
- **XLQuanah partition (16 nodes)**
 - Commissioned in 2022
 - 2x Intel Xeon E5-2695v4 Broadwell Processors/node
 - *512 total cores* (36 cores/node)
 - 4 TB total RAM (256 GB/node)
 - InfiniBand (53 Gbps) fabric
 - *This partition requires a request for special access*



HPCC Resources: Storage



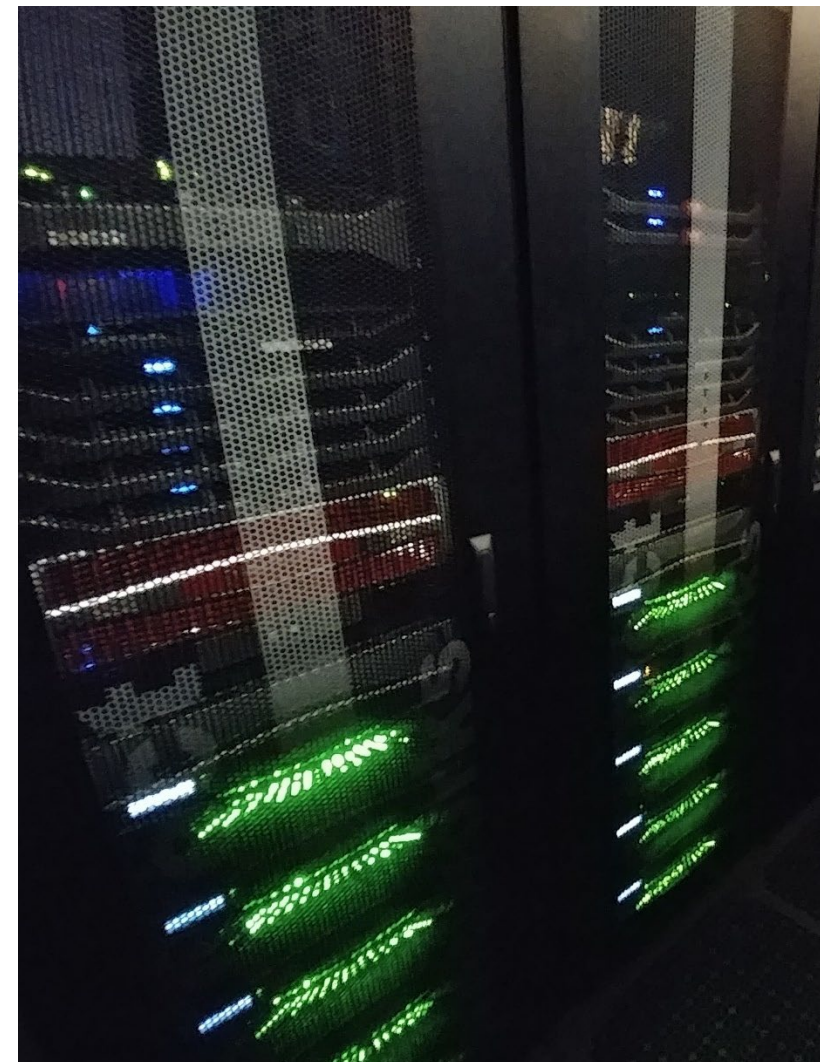
TEXAS TECH UNIVERSITY
Information Technology Division

Cluster-Wide Storage System:

- 6.1 PB of total storage space
- 200 Gbps HDR Fabric
- Based on Lustre parallel file system
- Quota/Backup/Purge policies per storage area (free to all accounts):

Area	Quota	File Limit	Backup	Purge
/home/<eraider>	300 GB	1,000,000	Yes	No
/lustre/work/<eraider>	700 GB	1,000,000	No	No
/lustre/scratch/<eraider>	None	None	No	Monthly

- Researchers/groups may purchase additional dedicated storage space:
 - With Backup: \$80/TB/Year
 - Without Backup: \$40/TB/Year





TEXAS TECH UNIVERSITY

Information Technology Division

RedRaider Cluster Software Environment

	Quanah	Nocona/Matador/Toreador/XLQuanah
Operating System	<ul style="list-style-type: none">CentOS 7.4 **	<ul style="list-style-type: none">CentOS 8.1
Job Resource Manager	<ul style="list-style-type: none">Slurm 20.11.7	<ul style="list-style-type: none">Slurm 20.11.7
Package Build Env	<ul style="list-style-type: none">RPM Build	<ul style="list-style-type: none">Spack v0.15
Software Deployment Env	<ul style="list-style-type: none">Lmod 7.7.14	<ul style="list-style-type: none">Lmod 8.2.10
Available C/C++/Fortran /MPI Compilers	<ul style="list-style-type: none">GCC 4.8.5 (Default)GCC 5.4.0GCC 7.3.0Intel 18.0.3.222impi 2018.3.222OpenMPI 1.10.[6-7]MVAPICH 2.2	<ul style="list-style-type: none">GCC 8.3.1 (Default)GCC 9.2.0GCC 10.1.0 (Recommended)AOCC/AOCL (Coming Soon)Intel compiler/MKL/MPI 2019OpenMPI 3.1.6, 4.0.4MVAPICH (Coming Soon)
GPU Libraries	<ul style="list-style-type: none">N/A	<ul style="list-style-type: none">CUDA 11.0 (default)Cudnn 8.0.1 (default)

The complete list of available software packages on the RedRaider cluster is available on the [HPCC website](#).

RedRaider Cluster Software Environment



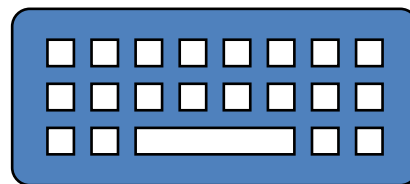
TEXAS TECH UNIVERSITY
Information Technology Division

- List of currently installed software packages on the RedRaider Cluster:

abyss	charm	ete3	git	ImageMagick	libxc	ncbi-rmblastn	ngs	paraview	py-astunparse	py-six	samtools	trinity
ampl	charmpp	exonerate	gnu	impi	libxsmm	ncbi-vdb	nvhpc	pear	py-cython	py-termcolor	scalapack	udunits
ansys	clustalw	fastStructure	gnu7	intel	lumerical	nccl	nwchem	perl	py-gast	python	singularity	vasp
augustus	cmake	fastx-toolkit	gnuplot	intel-mkl	mafft	ncl	octave	phdf5	py-google-pasta	python2	sparsehash	vasp-vtst
bamtools	cuda	fftw	grads	intel-mpi	maftools	nco	openblas	phyluce	py-grpcio	python3	stacks	vcftools
bcftools	cudnn	fftw3	gromacs	intel-tbb	maker	netcdf	openfoam	picard	py-keras-preprocessing	py-wheel	stokes	voro++
bedtools	cufflinks	fmriprep	gromacs-serial	java	matlab	netcdf-c	openjdk	picrust	py-matplotlib	py-wrapt	swig	vsearch
bedtools2	cvmfs	fsl	gsl	kokkos	mkl	netcdf-cxx	openmpi	povray	py-numpy	qiime	tcad	xanim
boost	dos2unix	gatk	gurobi	kokkos-nvcc-wrapper	molden	netcdf-cxx4	openmpi3	presto	py-opt-einsum	R	tcl	xcpEngine
bowtie2	drVM	gcc	hdf5	lammps	mpi4py27	netcdf-fortran	orca	proj	py-protobuf	repeatmasker	tcocoffee	zlib
bwa	eigen	gdal	hpl	lapack	mpi4py34	netcdf-serial	osu-micro-benchmarks	protobuf	pyrad	rmblast	tensorflow	zstd
cadence	elpa	geant4	hpx	libgtextutils	mvapich2	netlib-lapack	PAML	prun	py-scipy	root	tensorflow-mpi	
cdo	emboss	geos	htslib	libint	namd	netlib-scalapack	parallel-netcdf	py-abs-py	py-setuptools	rosetta	totalview	



Logging and Using the Cluster



Getting Started



TEXAS TECH UNIVERSITY
Information Technology Division

- Account Request:
 - Faculty/Staff account
 - Student account
 - Research Partner account
 - <http://www.depts.ttu.edu/hpcc/accounts/index.php>
- User Guides:
 - <http://www.depts.ttu.edu/hpcc/userguides/index.php>
- More details about HPCC equipment:
 - <http://www.depts.ttu.edu/hpcc/operations/equipment.php>



Getting Started



TEXAS TECH UNIVERSITY
Information Technology Division

- Logging to HPCC Resources:
 - <http://tinyurl.com/ttu-hpcc-login>
- On or Off Campus?
 - On Campus: Wired TTU network & TTUnet wireless network
 - Off Campus: Any other network connection, including
 - TTUHSC networks
 - TTUguest wireless network
- Logging in from Off Campus:
 - Log in via the SSH gateway
 - Establish a VPN - <https://goo.gl/4LbuWG>
 - Neither system is owned or maintained by HPCC



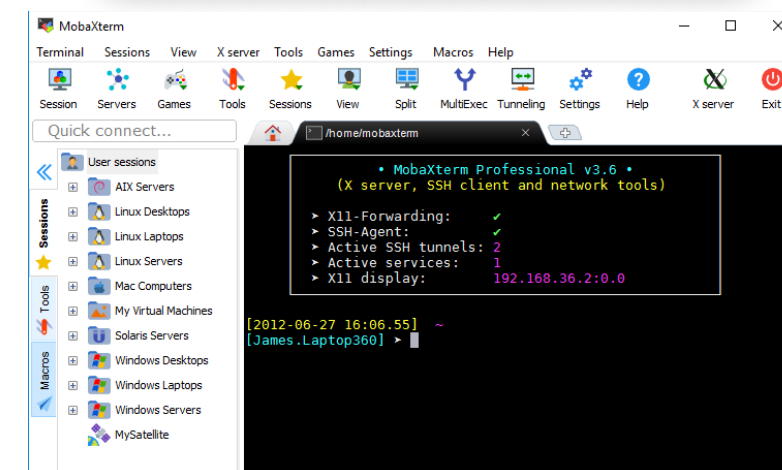
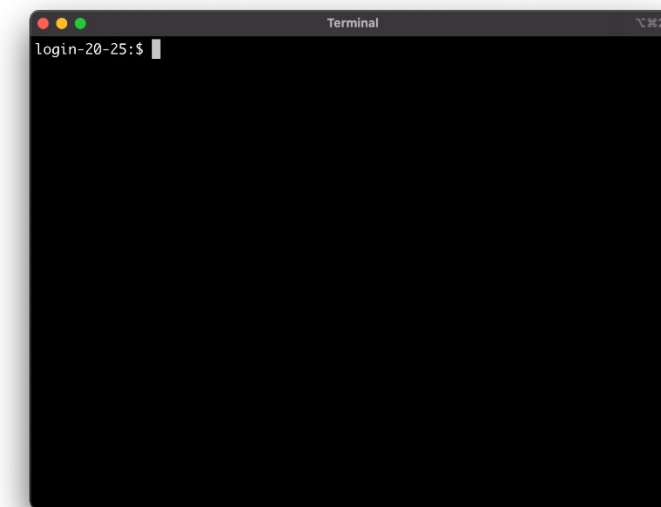


TEXAS TECH UNIVERSITY

Information Technology Division

Logging to RedRaider Cluster

- Mac/Linux Users:
 - SSH (Secure Shell): Freely available on Linux/Unix/macOS and used via the Terminal.
- ```
ssh eraider@login.hpcc.ttu.edu
ssh eraider@quanah.hpcc.ttu.edu
```
- Windows Users:
    - MobaXterm (Recommended): <https://mobaxterm.mobatek.net>
    - Putty: <https://www.putty.org>
  - Once logged in:
    - RedRaider has two login nodes: (**login-20-25**, **login-20-26**)
    - The load-balancer lands your SSH session on one of these nodes.
    - Quanah login node currently still available, login.hpcc.ttu.edu preferred.





## Information Technology Division

[illegible]

## ← Space quota and file limit usage

# Environment Settings



TEXAS TECH UNIVERSITY  
Information Technology Division

- Hostname pattern of the nodes in the RedRaider cluster:
  - Login nodes:
    - `Login-20-[25-26]`
    - `quanah`
  - CPU worker nodes:
    - `cpu-<rack#>-<chassis#>`
    - E.g., `cpu-23-10`
  - GPU worker nodes:
    - `gpu-<rack#>-chassis#>`
    - E.g., `gpu-20-7`

- **Interactive GUI using Linux/Mac.**
  - Mac users must install [XQuartz](#). Linux Users can use the Terminal.
  - Logging to the cluster using "-Y -X" with your normal ssh command:
    - `ssh -Y -X eraider@login.hpcc.ttu.edu`
  - Run a test command like `xclock`.
- **Interactive GUI using Windows.**
  - Install MobaXterm. (Consult MobaXTerm web site for details.)
  - Open a new tab in MobaXterm.
  - Logging to the cluster using "-Y -X" with your normal ssh command:
    - `ssh -Y -X eraider@login.hpcc.ttu.edu`
  - Run a test command like `xclock`.



# Environment Settings



TEXAS TECH UNIVERSITY  
Information Technology Division

- Lmod Modules:
  - The primary way to change your user environment.
  - Load/Unload a particular set of cluster-wide installed software packages into your environment or job submissions.
  - Makes life easier by modifying the proper environment variables for you such as PATH, LD\_LIBRARY\_PATH, etc.
  - User Guide:
    - [https://www.depts.ttu.edu/hpcc/userguides/general\\_guides/software\\_environment.php](https://www.depts.ttu.edu/hpcc/userguides/general_guides/software_environment.php)

```
Available modules in Nocona partition
login-20-26:~$ module avail
=====
Texas Tech University
High Performance Computing Center
RedRaider Cluster
=====

Current Arch: [nocona]

--- /opt/apps/nfs/spack/var/spack/environments/nocona/modules/linux-centos8-x86_64/Core ---
cmake/3.17.3 git/2.27.0 intel/19.1.2 totalview/2021.1.16
gcc/9.2.0 intel-mkl/2019.5.281 matlab/R2020b
gcc/10.1.0 intel-mpi/2019.10.317 openfoam/v65
gcc/10.2.0 (D) intel-tbb/2020.2 perl/5.30.3

----- /opt/apps/nfs/RedRaider-modules -----
matador/0.15.4 nocona/0.15.4 (L) quanah/0.15.4 toreador/0.15.4

Where:
D: Default Module
L: Module is loaded
login-20-26:~$
```




- Lmod Module Commands:

| Command                                                        | Description                                    |
|----------------------------------------------------------------|------------------------------------------------|
| <code>module avail</code>                                      | List all the available modules                 |
| <code>module list</code>                                       | List all the modules currently loaded          |
| <code>module load &lt;module_name&gt;</code>                   | Load a module in your environment              |
| <code>module unload &lt;module_name&gt;</code>                 | Unload a module from your environment          |
| <code>module swap &lt;old_module&gt; &lt;new_module&gt;</code> | Replace the old module with the new one        |
| <code>module spider &lt;keyword&gt;</code>                     | Search for a module in the Lmod hierarchy list |
| <code>module purge</code>                                      | Unload all the modules currently loaded        |
| <code>module help &lt;module_name&gt;</code>                   | Show the description of the module             |
| <code>module whatis &lt;module_name&gt;</code>                 | Show a brief info about the module             |
| <code>module show &lt;module_name&gt;</code>                   | Show a complete info about the module          |

# Search Available Modules on the RedRaider




TEXAS TECH UNIVERSITY  
Information Technology Division

 High Performance Computing Center

Resources ▾ Operations ▾ About HPCC ▾ RedRaider Cluster ▾ HPCC Support



TTU / High Performance Computing Center

 MATADOR NOCONA QUANAH TOREADOR

## HPCC RedRaider Cluster Software Packages

HPCC users may find the most up-to-date list of available software packages and dependencies on the RedRaider cluster on this web page. [For more info on how to use Modules, please refer to the Software Environment Setup Guide.](#)


**Search for a software package:**

1. Use the search box above to look for any specific compiler, math library, or scientific software package and check the availability on the HPCC RedRaider cluster.
  - The list of the available software package(s) will appear for each partition.
  - Commonly, results may show the different versions of the same software package built by various compilers.
  - [For more details regarding each partition on the RedRaider cluster, click here.](#)
2. Click on the search results leads you to the full path of the software package in the HPCC software hierarchy stack
  - You will be able to see the software package's description and its complete LMOD module command.
3. Use the delete button  in the search box to clear the search, or click on the return button  to go back to the search result if you are not there already.

**Traverse the current software package hierarchy available for each partition:**

1. Select the target partition from the top menu.
  - [For more details regarding each partition on the RedRaider cluster, click here.](#)
2. Choose the desired independent software package or compiler from the first level of the stack.
  - The description box will appear for the selected package below the stack.
  - You will have a choice to select the preferred version(s) of the software/compiler from the next level on the stack.
  - The bottom-right box below the stack shows more details about using the software module(s) correctly.
3. If dependent software packages are available, they will show up at the next level on the software stack.
  - If applicable, follow steps 2 and 3 to traverse deeper into the hierarchy stack.

If your desired software package or a specific software version did not show up here, you may follow the instructions on the [Request Software](#) web page and place your new software request. Please note that HPCC policies and restrictions on software installation described in this link apply to all new software requests.

To see this information again, click on the  button in the partition header at the top.

[HPCC RedRaider Software Packages](#)



TEXAS TECH UNIVERSITY

Information Technology Division

# Modules Tips and Recommendations

- **Try to keep all your module load commands as part of your job submission scripts instead of adding them to your `.bashrc` file.**
  - Makes debugging and changing between experiments and cluster partitions easier.
  - Prevents collisions or accidentally running jobs in the wrong environment.
- **Please note that **Quanah** (Intel nodes), **Nocona** (AMD nodes), **Matador** (V100 GPU nodes), and **Toreador** (A100 GPU nodes) have a different set of modules.**
  - More details about this later in this training.
- **Always include the version number of a module in the module load command.**
  - Makes version tracking easier.
  - Prevents unanticipated changes in version during an experiment.
  - Example: Use `module load nwchem/6.6-intel` instead of just `module load nwchem`



# Exercise #1

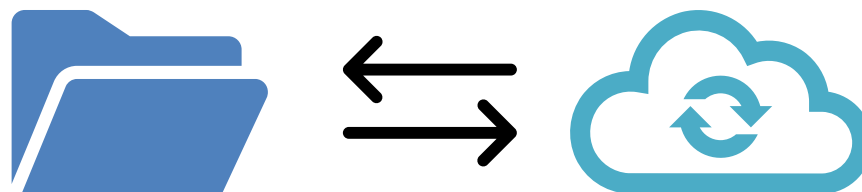


TEXAS TECH UNIVERSITY  
Information Technology Division

1. Log in to the “`login.hpcc.ttu.edu`” using your eraider account.
2. List all the available modules on the cluster:
  - What partition does the Lmod module support?
3. Load the ‘`gcc/10.1.0`’ module
4. Do you see any changes in the list of available modules? Are any new modules on this list?
5. Do you see ‘`openmpi/4.0.4`’ in the updated list? If so, then load it, please.
  - Do you see any changes again?
  - What modules have you loaded so far? Can you get a list of them?
6. Now try to load “`intel/19.1.2`”. Is there any problem? What is the solution?
7. Purge all the currently loaded modules. Could you get rid of the partition module?
8. How about modules on “`quanah.hpcc.ttu.edu`”?



# Transferring Data





TEXAS TECH UNIVERSITY

Information Technology Division

# Transferring Data (Using Globus)

- Whenever possible, refrain from using:
  - `scp`, `sftp`, `rsync`, or any other data transfer tool
  - Okay for very small transfers, but for large ones, see below.
- Transfer files using Globus Connect personal client:
  - User Guide: <http://tinyurl.com/hpcc-data-transfer>
  - Globus Connect service is well connected to the campus network.
  - Globus Connect service eliminates the data transfer load from the cluster login nodes.
  - Globus Connect personal client works on **Linux, Mac and Windows** and is easy to control through a simple web GUI.
  - Numerous other sites (including TACC) support Globus Connect data transfers.



# Transferring Data (Using Globus)



TEXAS TECH UNIVERSITY  
Information Technology Division

Please view our updated [Privacy Policy](#) and [Terms of Service](#).

File Manager The HPCC end point is: **ttuhpcc#TTUTerra**

Collection **ttuhpcc#TTUTerra** **3** Path **/~/training/** **Start**

**1** → Panels

**2** **TTU\_Dell\_optiplex** **2** Path **/~/C:/[redacted]/Globus/**

Transfer & Timer Options **Start** **5** ←

☐ select all

| NAME        | LAST MODIFIED      | SIZE |
|-------------|--------------------|------|
| exercise1   | 2/1/2022, 2:11 PM  |      |
| exercise2   | 2/1/2022, 11:48 AM |      |
| exercise4   | 2/1/2022, 1:57 PM  |      |
| exercise5   | 2/1/2022, 2:33 PM  |      |
| exercise6   | 2/1/2022, 3:14 PM  |      |
| file.txt    | 2/1/2022, 11:09 AM | 39   |
| long.txt    | 7/7/2021, 2:17 PM  | 52   |
| testvim.txt | 2/1/2022, 2:25 PM  | 10   |
| vasp.out    | 2/2/2022, 9:25 AM  | 24   |

Share

| NAME                                                                           | LAST MODIFIED       | SIZE     |
|--------------------------------------------------------------------------------|---------------------|----------|
| <input checked="" type="checkbox"/> Col7.0.2_2021-04-30_linux64.ifc_src.tar.gz | 11/30/2021, 4:10 PM | 57.60 MB |
| <input type="checkbox"/> redraider.shared.R2021b.tar.gz                        | 11/30/2021, 3:44 PM | 23.07 KB |

**4** ←

# Coming in Part 2 of this Training ...



TEXAS TECH UNIVERSITY  
Information Technology Division

## Part 2:

- ❖ Resource Allocation and Job Submission with SLURM
- ❖ Software builds and installation
- ❖ HPCC Policies
- ❖ Getting Help



# Lunch Break

Let's get back at 1:00 pm



TEXAS TECH UNIVERSITY  
Information Technology Division™