Introduction to the MSU HPCC

High Performance Computing Center

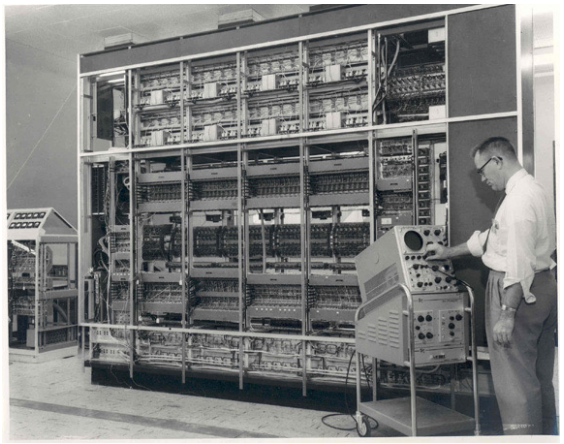
Dirk Colbry
colbrydi@msu.edu

Thanks to Dr. Mahmoud Parvizi for the slides

Agenda

- ***Overview of the HPCC***
- Logging on to the HPCC
- Simple example
  - Navigating Files
  - Module System
  - Submitting a job
- Where to get help

Computing at MSU



ICER

CMSE

MISTIC

Computer Science Department

MERIT

HPCC

Computational Mathematics Science and Engineering Department

1957

1963–1973

1967

1968

1971

1978

2004

2009

2015

CDC 3600

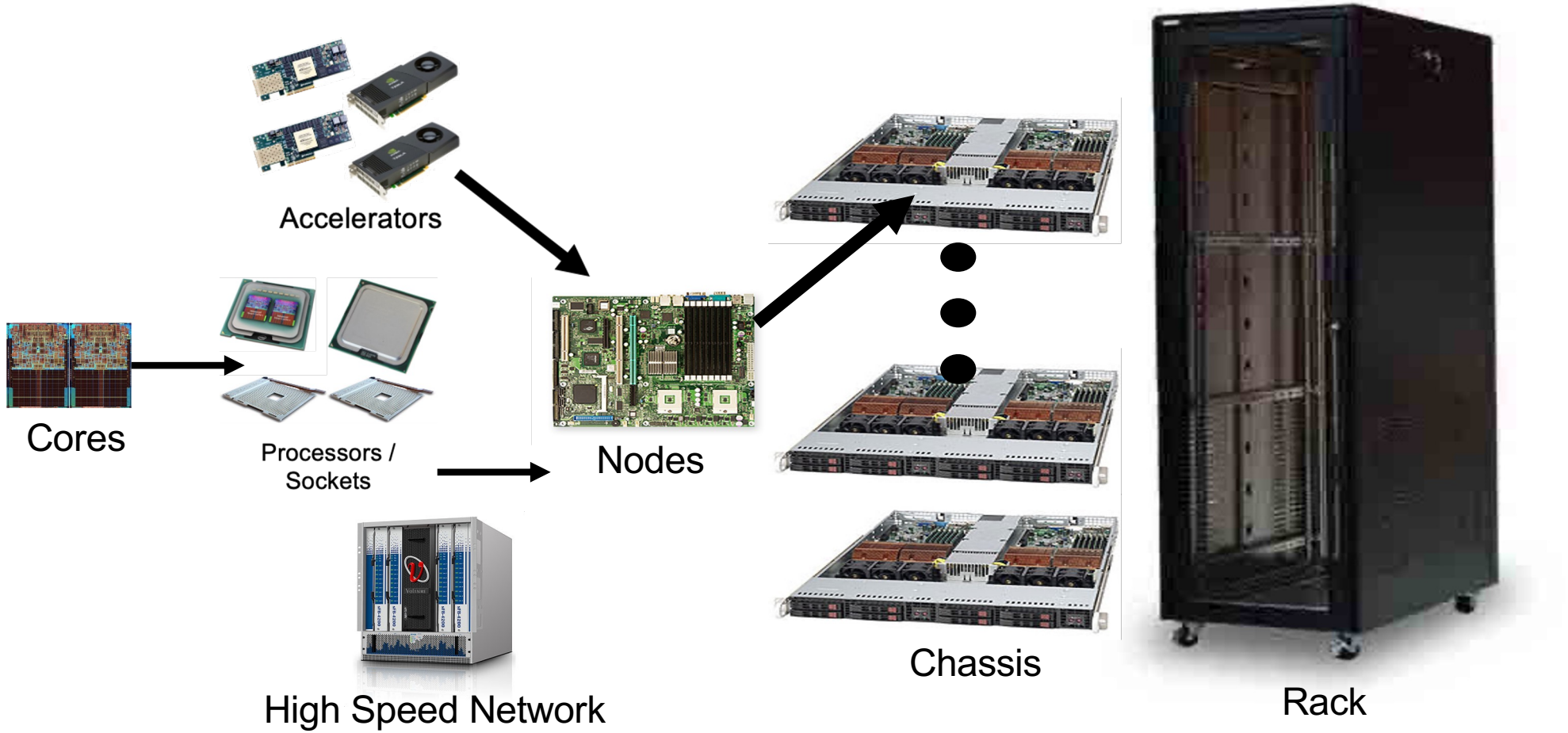
CDC 6500

Cyber 750

ICER

Creating a Cluster

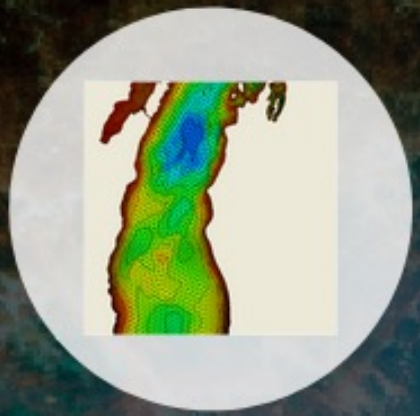
aka Supercomputer

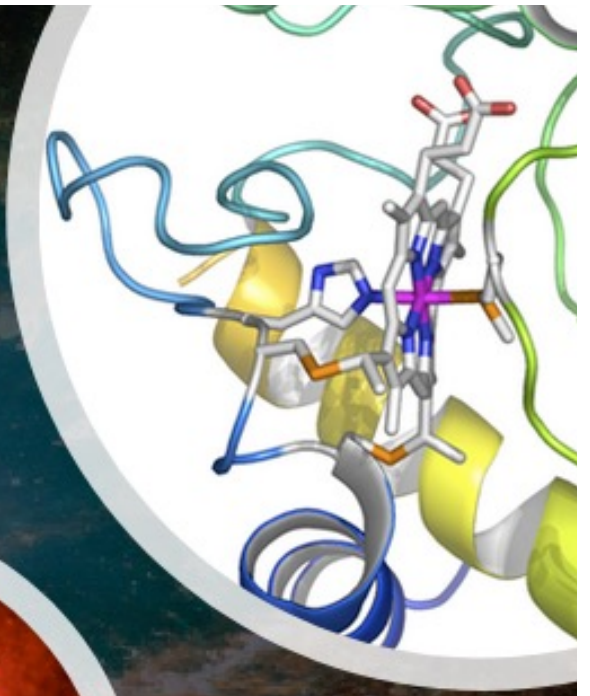


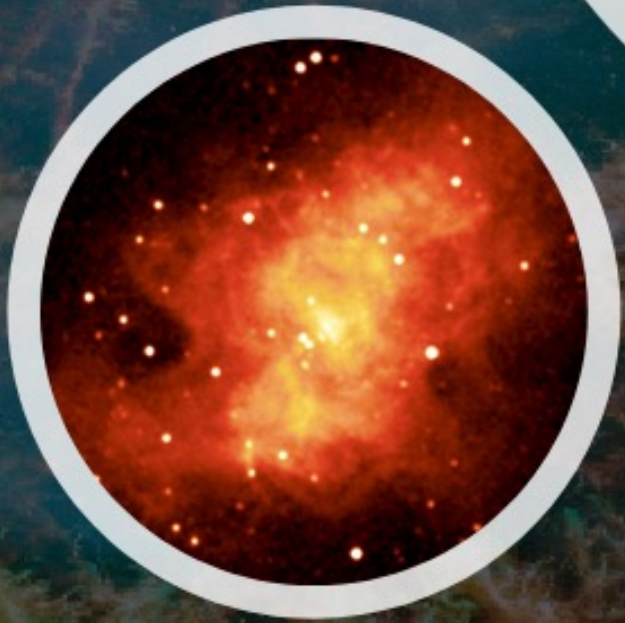
Data Center

What Problems are we Solving?

- Boundary Simulations
- Data Analysis
- Search (aka Optimization)

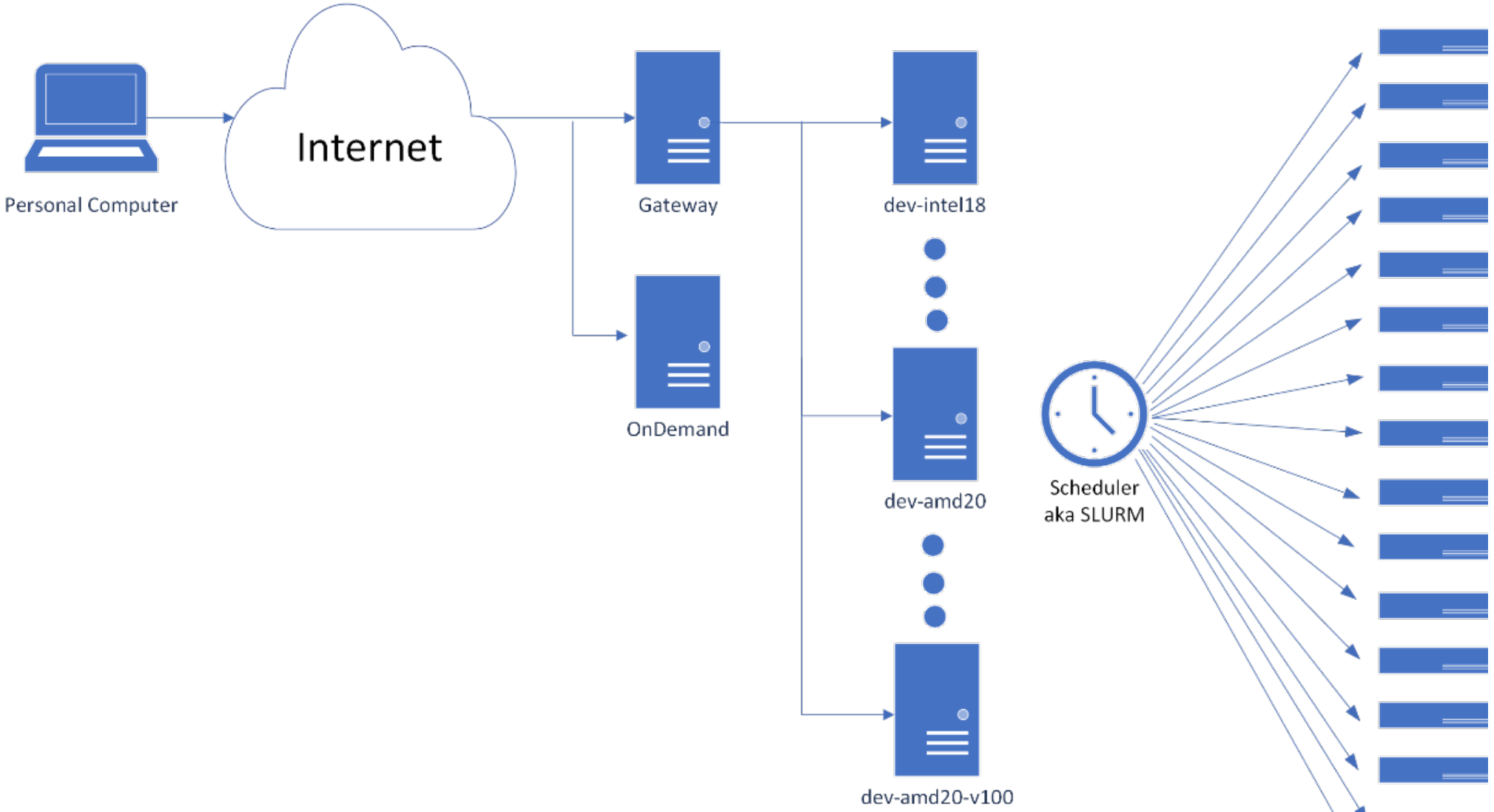






Agenda

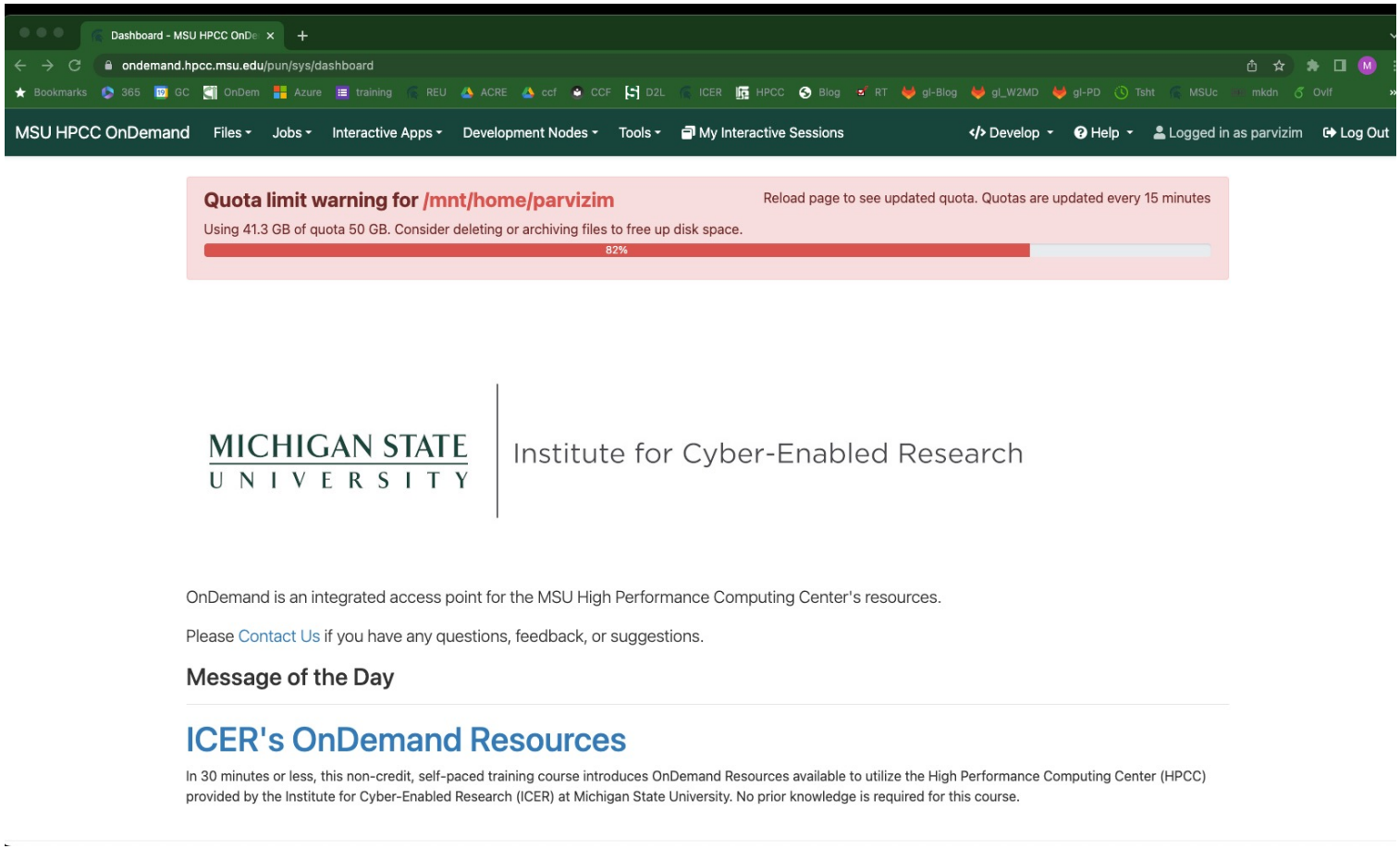
- Overview of the HPCC
- ***Logging on to the HPCC***
- Simple example
  - Navigating Files
  - Module System
  - Submitting a job
- Where to get help




```
ssh USERNAME@gateway.hpcc.msu.edu
```

HPCC OnDemand

HPCC OnDemand ondemand.hpcc.msu.edu



Agenda

- Overview of the HPCC
- Logging on to the HPCC
- ***Simple example***
  - Navigating Files
  - Module System
  - Submitting a job
- Where to get help

Getting a simple example

1. Open a “terminal” on the HPCC
2. Load the powertools module

```
module load powertools
```

3. See all of the examples

```
getexample
```

4. Run the “get example” command

```
getexample helloHPCC
```

5. Change to the “helloworld” directory

```
cd helloHPCC
```

6. Run the example

```
cat README
```

Compiling and Running the Example

1. Compile the code (follow the README)

```
cc -o example -O example_calc_e.c
```

2. Test the example on the dev node

```
./example
```

3. Review the submission script

```
cat example.sb
```

4. Run the the example on the “cluster”

```
sbatch example.sb
```

5. Review the output

```
cat slurm*
```

Agenda

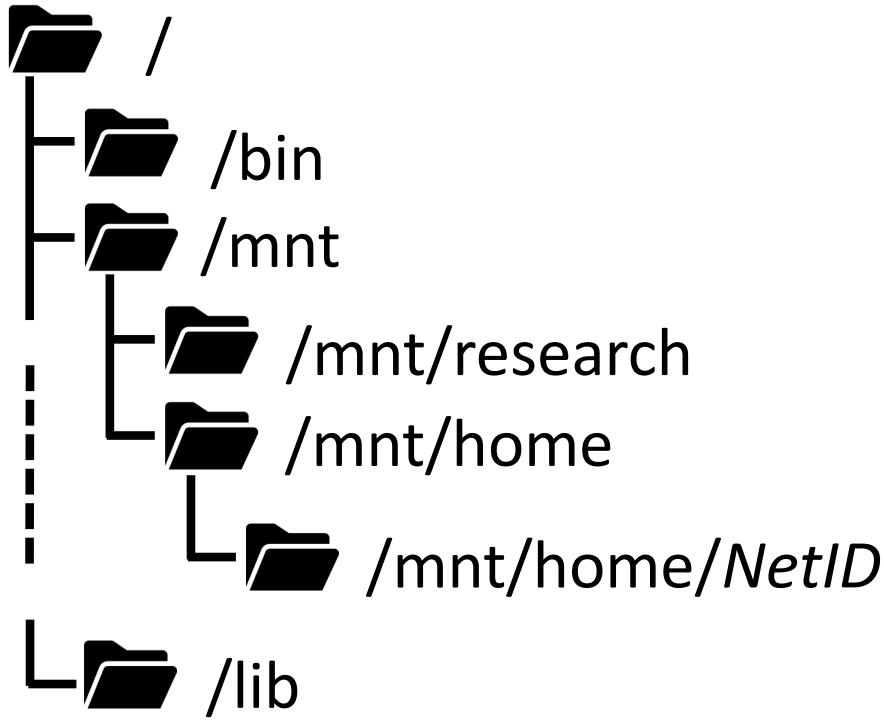
- Overview of the HPCC
- Logging on to the HPCC
- Simple example
  - ***Navigating Files***
  - Module System
  - Submitting a job
- Where to get help

Linux

Linux on the HPCC

A **clustered filesystem** is a hierarchical collection of files accessible to all compute nodes of a cluster

- **File:** A formatted collection of bytes referenced by the OS
- **Directory:** Any file containing another file
- **Filesystem:** Method use by OS to store and retrieve files



Linux

Linux on the HPCC

Directories have a **tree-like** structure

- **Examples:**

/ -- the "root"

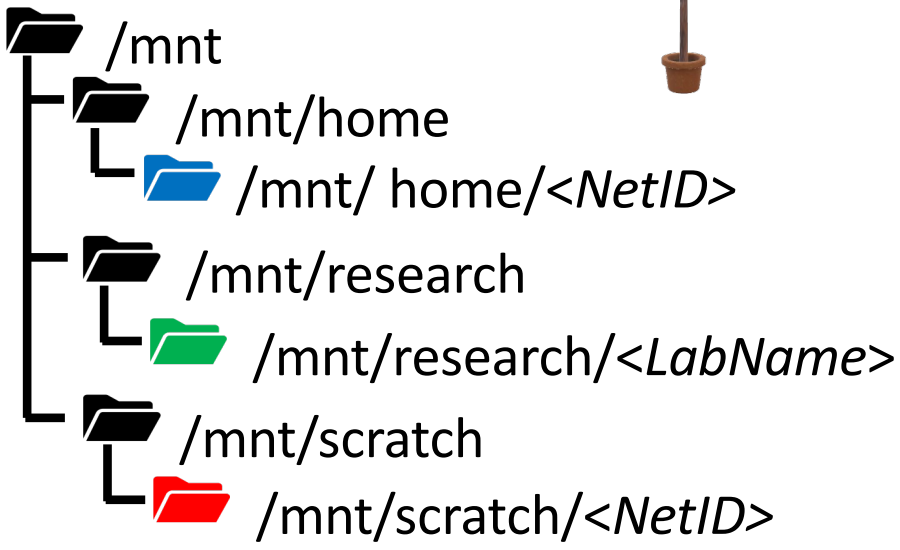
- **/bin** -- shared executables
- **/mnt** -- all groups and users
  - **/mnt/research** -- all research groups
  - **/mnt/home** -- all users
    - **/mnt/home/*NetID*** -- your personal directory
- **/lib** -- shared library of programs

Linux

Linux on the HPCC

User Spaces are directories users can write files to using any nodes in a cluster

- **Home:** Personal files and default login directory (50Gb storage)
- **Research:** Group research files (50Gb - 1Tb storage)
- **Scratch:** Temporary working files (~800 Tb total storage)



Linux

Linux on the HPCC

Exercise: use the **quota** command to display the details of your user spaces

Type in Your Terminal:

[user@computer] $ quota

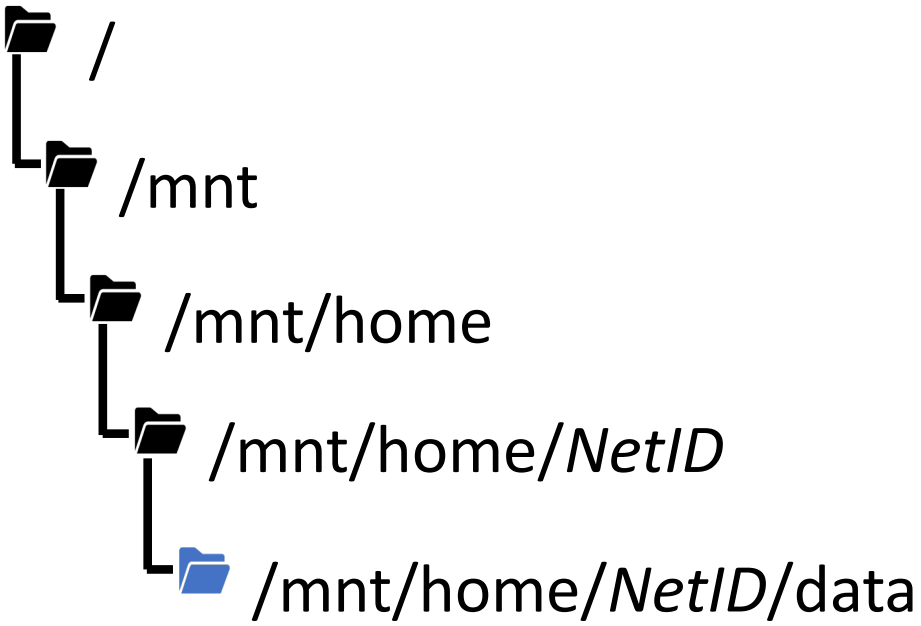
```
parvizim@dev-intel18:~ — ssh parvizim@hpcc.msu.edu — 123×49

Development nodes are a shared system; for information about performance
considerations please see: https://wiki.hpcc.msu.edu/x/N4JnAg
===

[parvizim@dev-intel18 ~]$ quota
Home Directory:                Space      Space      Space      Space      Files      Files      Files
Files
                               Quota      Used       Available  % Used     Quota      Used       Available
% Used
-------------------------------------------------------------------------------------------------------
------------

/mnt/home/parvizim             50G        7G         43G        14%        1048576    87299      961277
8%



Research Groups:               Space      Space      Space      Space      Files      Files      Files
Files
                               Quota      Used       Available  % Used     Quota      Used       Available
% Used
-------------------------------------------------------------------------------------------------------
------------

TOPMED                         4096G      3733G      363G       91%        4194304    1558       4192746
0%
UKBB                           9216G      8242G      974G       89%        9437184    6897       9430287
0%
helpdesk                       12288G     9888G      2400G      80%        52428800   44415439   8013361
85%



Temporary Filesystems:
-------------------------------------------------------------------------------------------------------
------------

/mnt/scratch (/mnt/gs18)       Space Quota  Space Used  Space Free  Space % Used Filess Quota Files Used  Files Free
Files % Used
                               51200G       1G          51199G      0%           1048576      2           1048574
0%

/mnt/ls15 (legacy scratch)     Inodes Used  Quota       Free
                               1            1000000     999999

[parvizim@dev-intel18 ~]$
```

Linux

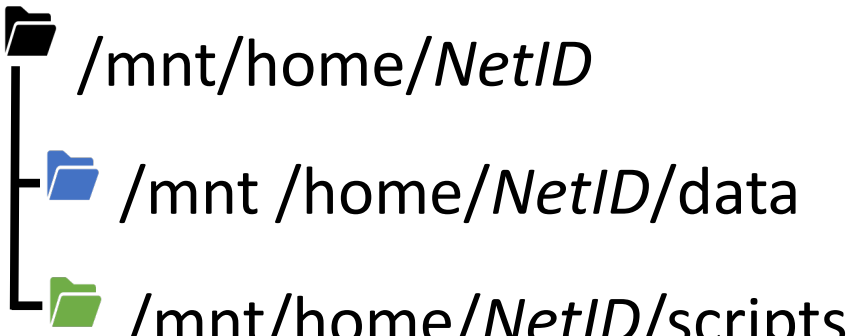
Linux on the HPCC

A **path to a file** is a list of the files containing the file of interest

Absolute **/mnt/home/*NetID*/data**



Relative **. . /scripts**



Path Shortcuts

Shortcut	Description	Example
.	Current folder (single dot)	./command
..	Parent folder (two dots)	cat ../.bashrc
~	Home Directory (tilde)	cd ~/Documents/
-	Previous Directory (dash)	cd -

Note: If you want a file to be “hidden” have its name statr with a dot
ex: .bashrc

If you want to see hidden files use the command “ls -a” (List all)

Linux

Linux on the HPCC

Exercise: Find path with **pwd**, or print working directory, command

```
[parvizim@dev-intel18 ~]$
[parvizim@dev-intel18 ~]$
[parvizim@dev-intel18 ~]$ pwd
/mnt/home/parvizim
[parvizim@dev-intel18 ~]$
```

Type in Your Terminal:

[user@computer] $ pwd

Linux

Linux on the HPCC

Exercise: List files with **ls**, or list information, command

```
[parvizim@dev-intel18 ~]$
[parvizim@dev-intel18 ~]$ ls
Documents
[parvizim@dev-intel18 ~]$
```

Type in Your Terminal:

[user@computer] $ ls

Linux

Linux on the HPCC

Exercise: Enter ‘Documents’ with **cd**, or change directory, command and list the contents

```
[[parvizim@dev-intel18 ~]$
[[parvizim@dev-intel18 ~]$ cd Documents; ls
MATLAB  RT  Workshops
[parvizim@dev-intel18 Documents]$
```

Type in Your Terminal:

[user@computer] $ cd Documents; ls

Transferring files to/from HPCC

	Small Files < 10MB	Med Files < 2 GB	Lots of Big Files	Interface	Upload To HPCC	Download To PC
OnDemand	Yes	No	No	Web	Yes	Yes
wget / curl	Yes	Maybe	No	Command Line	Yes	No
git	Yes	Maybe	No	Command Line	Yes	Maybe*
scp/rsync*	Yes	Yes	Maybe	Command Line	Yes	Yes
MobaXTerm	Yes	Yes	Maybe	Windows App	Yes	Yes
Globus*	Yes	Yes	Yes	Web	Yes	Yes

* Requires Software Install on your PC

```
scp FILENAME USERNAME@rsync.hpcc.msu.edu:~
```

Linux

Linux on the HPCC

File transfer GUI **Globus** provided by MSU

https://www.globus.org/data-transfer

Agenda

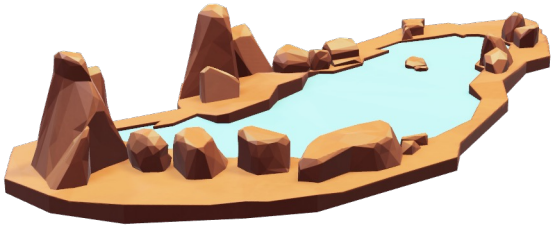
- Overview of the HPCC
- Logging on to the HPCC
- Simple example
  - Navigating Files
  - ***Module System***
  - Submitting a job
- Where to get help

Linux

Linux on the HPCC

The HPCC **module system** helps manage the software environment you need to run your computations

- **Environment:** User specified software applications and their dependencies



- **Dependency:** Any file needed by an executable software application

- **Module:** User loaded software that comprise an environment

Linux

Linux on the HPCC

Examples of HPCC modules include **compilers** and **libraries**

- **Compiler:** Software that translates code e.g., source to machine (GCC, intel, CUDA)

- **Library:** Collection of software resources used by the compiler and other executables; e.g., Math (BLAS, LaPACK)

Linux

Linux on the HPCC

Exercise: List default HPCC modules with the **module list** command

```
[parvizim@dev-intel18 ~]$
[parvizim@dev-intel18 ~]$ module list

Currently Loaded Modules:
  1) GCCcore/6.4.0     7) OpenBLAS/0.2.20                    13) CMake/3.11.1     19) libffi/3.2.1
  2) binutils/2.28     8) FFTW/3.3.7                         14) ncurses/6.0      20) Python/3.6.4
  3) GNU/6.4.0-2.28    9) ScaLAPACK/2.0.2-OpenBLAS-0.2.20    15) libreadline/7.0  21) Java/1.8.0_152
  4) OpenMPI/2.1.2    10) bzip2/1.0.6                        16) Tcl/8.6.8        22) MATLAB/2018a
  5) tbb/2018_U3      11) zlib/1.2.11                        17) SQLite/3.21.0
  6) imkl/2018.1.163  12) Boost/1.67.0                       18) GMP/6.1.2
```

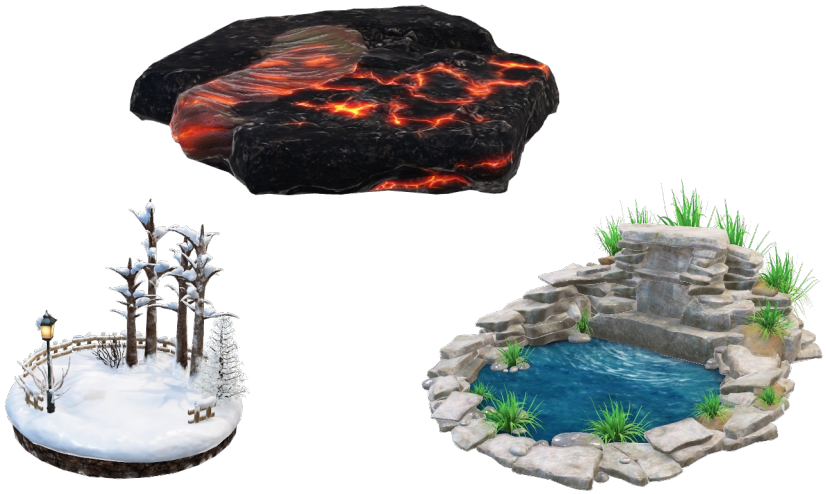
Type in Your Terminal:

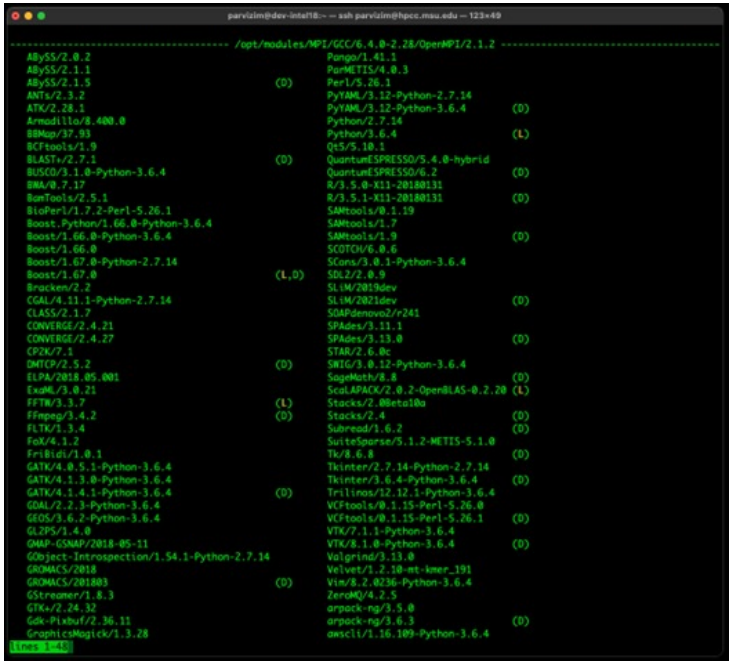
[user@computer] $ module list

Linux

Linux on the HPCC

Exercise: Find all HPCC modules with the **module avail** command





Type in Your Terminal:

[user@computer] $ module avail

Linux

Linux on the HPCC

Exercise: Load HPCC modules with the **module load** command

```
[parvizim@dev-intel18 ~]$ module load R/4.0.2
Lmod has detected the following error:  These module(s) or extension(s) exist but cannot be loaded as requested:
"R/4.0.2"
   Try: "module spider R/4.0.2" to see how to load the module(s).


[parvizim@dev-intel18 ~]$
```

Type in Your Terminal:

[user@computer] $ module load R/4.0.2

Linux

Linux on the HPCC

Exercise: Find specific modules with the **module spider** command

Type in Your Terminal:

```
[parvizim@dev-intel18 ~]$
[parvizim@dev-intel18 ~]$ module spider R

----------------------------------------------------------------------------
  R:
----------------------------------------------------------------------------
    Description:
      R is a free software environment for statistical computing and graphics.

     Versions:
        R/3.3.1
        R/3.4.3-X11-20160819
        R/3.4.3-X11-20171023
        R/3.4.3xF
        R/3.4.3xS
        R/3.4.4-X11-20180131
        R/3.5.0-X11-20180131
        R/3.5.1-X11-20180131
        R/3.5.1-X11-20180604-UR
        R/3.5.1-X11-20180604
        R/3.6.0-X11-20180604
        R/3.6.2-X11-20180604
        R/3.6.2
        R/3.6.3
        R/4.0.0-X11-20180604
        R/4.0.0
        R/4.0.2.bak
        R/4.0.2.test
        R/4.0.2-X11-20180604
        R/4.0.2
        R/4.0.3
        R/4.1.0
        R/4.1.2
        R/4.2.2
     Other possible modules matches:
        ADMIXTURE  AMDuProf  APR  APR-util  Abaqus_parallel  AdapterRemoval  Advisor  Amber  AmrPlusPlus  Armadillo  ...

----------------------------------------------------------------------------
  To find other possible module matches execute:

      $ module -r spider '.*R.*'

----------------------------------------------------------------------------
  For detailed information about a specific "R" package (including how to load the modules) use the module's full name.
```

[user@computer] $ module spider R

Linux

Linux on the HPCC

Exercise: Load HPCC modules with the **module load** command

```
[parvizim@dev-intel18 ~]$ module load GCC/8.3.0  OpenMPI/3.1.4 R/4.0.2

Lmod is automatically replacing "GNU/6.4.0-2.28" with "GCC/8.3.0".


Lmod is automatically replacing "GNU/6.4.0-2.28" with "GCC/8.3.0".

--------------------------------------------------------------------------------
  --------------------

The following dependent module(s) are not currently loaded: OpenBLAS/0.2.20 (required by: ScaLAPACK/2.0.2
-OpenBLAS-0.2.20, Boost/1.67.0, Python/3.6.4, ScaLAPACK/2.0.2-OpenBLAS-0.2.20, Boost/1.67.0, Python/3.6.4
)
--------------------------------------------------------------------------------
  --------------------


Inactive Modules:
  1) CMake/3.11.1     2) imkl/2018.1.163     3) tbb/2018_U3

Due to MODULEPATH changes, the following have been reloaded:
  1) Boost/1.67.0     2) GMP/6.1.2     3) libffi/3.2.1     4) zlib/1.2.11

The following have been reloaded with a version change:
  1) FFTW/3.3.7 => FFTW/3.3.8              8) ScaLAPACK/2.0.2-OpenBLAS-0.2.20 => ScaLAPACK/2.0.2
  2) GCCcore/6.4.0 => GCCcore/8.3.0        9) Tcl/8.6.8 => Tcl/8.6.9
  3) Java/1.8.0_152 => Java/11.0.2        10) binutils/2.28 => binutils/2.32
  4) OpenBLAS/0.2.20 => OpenBLAS/0.3.7    11) bzip2/1.0.6 => bzip2/1.0.8
  5) OpenMPI/2.1.2 => OpenMPI/3.1.4       12) libreadline/7.0 => libreadline/8.0
  6) Python/3.6.4 => Python/3.7.4         13) ncurses/6.0 => ncurses/6.1
  7) SQLite/3.21.0 => SQLite/3.29.0

[parvizim@dev-intel18 ~]$
```

Type in Your Terminal:

[user@computer] $ module load GCC/8.3.0 OpenMPI/3.1.4 R/4.0.2

Linux

Linux on the HPCC

Exercise: Unload HPCC modules with the **module unload** command

```
parvizim — parvizim@dev-intel18:~ — ssh -XY parvizim@hpcc.msu.edu — 102×27
Currently Loaded Modules:
  1) GCCcore/6.4.0      9) ScaLAPACK/2.0.2-OpenBLAS-0.2.20  17) SQLite/3.21.0
  2) binutils/2.28     10) bzip2/1.0.6                      18) GMP/6.1.2
  3) GNU/6.4.0-2.28    11) zlib/1.2.11                      19) libffi/3.2.1
  4) OpenMPI/2.1.2     12) Boost/1.67.0                     20) Python/3.6.4
  5) tbb/2018_U3       13) CMake/3.11.1                     21) Java/1.8.0_152
  6) imkl/2018.1.163   14) ncurses/6.0                      22) MATLAB/2018a
  7) OpenBLAS/0.2.20   15) libreadline/7.0                  23) powertools/1.2
  8) FFTW/3.3.7        16) Tcl/8.6.8

[parvizim@dev-intel18 ~]$ module unload powertools/1.2; module list

Currently Loaded Modules:
  1) GCCcore/6.4.0      9) ScaLAPACK/2.0.2-OpenBLAS-0.2.20  17) SQLite/3.21.0
  2) binutils/2.28     10) bzip2/1.0.6                      18) GMP/6.1.2
  3) GNU/6.4.0-2.28    11) zlib/1.2.11                      19) libffi/3.2.1
  4) OpenMPI/2.1.2     12) Boost/1.67.0                     20) Python/3.6.4
  5) tbb/2018_U3       13) CMake/3.11.1                     21) Java/1.8.0_152
  6) imkl/2018.1.163   14) ncurses/6.0                      22) MATLAB/2018a
  7) OpenBLAS/0.2.20   15) libreadline/7.0
  8) FFTW/3.3.7        16) Tcl/8.6.8

[parvizim@dev-intel18 ~]$
```

Type in Your Terminal:

[user@computer] $ module unload powertools; module list

Linux

Linux on the HPCC

Exercise: Unload all HPCC modules with **module purge** command

```
[parvizim@dev-intel18 ~]$
[parvizim@dev-intel18 ~]$ module list

Currently Loaded Modules:
  1) GCCcore/6.4.0      7) OpenBLAS/0.2.20                 13) CMake/3.11.1      19) libffi/3.2.1
  2) binutils/2.28      8) FFTW/3.3.7                      14) ncurses/6.0       20) Python/3.6.4
  3) GNU/6.4.0-2.28     9) ScaLAPACK/2.0.2-OpenBLAS-0.2.20 15) libreadline/7.0  21) Java/1.8.0_152
  4) OpenMPI/2.1.2     10) bzip2/1.0.6                     16) Tcl/8.6.8         22) MATLAB/2018a
  5) tbb/2018_U3       11) zlib/1.2.11                     17) SQLite/3.21.0
  6) imkl/2018.1.163   12) Boost/1.67.0                    18) GMP/6.1.2



[parvizim@dev-intel18 ~]$ module purge; module list
No modules loaded
[parvizim@dev-intel18 ~]$
```

Type in Your Terminal:

[user@computer] $ module purge; module list

Linux

Linux on the HPCC

Exercise: Reload default HPCC modules with the **logout** command

```
[[parvizim@dev-intel18 ~]$
[[parvizim@dev-intel18 ~]$ module purge; module list
No modules loaded
[[parvizim@dev-intel18 ~]$ logout
Connection to dev-intel18 closed.

Currently Loaded Modules:
  1) gateway/1.0



[[parvizim@gateway-03 ~]$ ssh dev-intel18
Last login: Mon Jan 23 21:39:23 2023 from gateway-03.dmz

===
Please note that processes on development nodes are limited to two hours of
CPU time; for longer-running jobs, please submit to the queue.

Development nodes are a shared system; for information about performance
considerations please see: https://docs.icer.msu.edu/development_nodes/
===

[[parvizim@dev-intel18 ~]$ module list

Currently Loaded Modules:
  1) GCCcore/6.4.0     7) OpenBLAS/0.2.20                 13) CMake/3.11.1     19) libffi/3.2.1
  2) binutils/2.28     8) FFTW/3.3.7                      14) ncurses/6.0      20) Python/3.6.4
  3) GNU/6.4.0-2.28    9) ScaLAPACK/2.0.2-OpenBLAS-0.2.20 15) libreadline/7.0  21) Java/1.8.0_152
  4) OpenMPI/2.1.2    10) bzip2/1.0.6                     16) Tcl/8.6.8        22) MATLAB/2018a
  5) tbb/2018_U3      11) zlib/1.2.11                     17) SQLite/3.21.0    23) powertools/1.2
  6) imkl/2018.1.163  12) Boost/1.67.0                    18) GMP/6.1.2


[parvizim@dev-intel18 ~]$
```

Type in Your Terminal:

[user@computer] $ logout

Linux

Linux on the HPCC

Exercise: See what the module is doing with the **module show** command

Type in Your Terminal:

[user@computer] $ module show matlab

```
[parvizim@dev-intel18 ~]$
[parvizim@dev-intel18 ~]$ module purge; module list
No modules loaded
[parvizim@dev-intel18 ~]$ logout
Connection to dev-intel18 closed.

Currently Loaded Modules:
  1) gateway/1.0



[parvizim@gateway-03 ~]$ ssh dev-intel18
Last login: Mon Jan 23 21:39:23 2023 from gateway-03.dmz

===
Please note that processes on development nodes are limited to two hours of
CPU time; for longer-running jobs, please submit to the queue.

Development nodes are a shared system; for information about performance
considerations please see: https://docs.icer.msu.edu/development_nodes/
===

[parvizim@dev-intel18 ~]$ module list

Currently Loaded Modules:
  1) GCCcore/6.4.0     7) OpenBLAS/0.2.20                13) CMake/3.11.1     19) libffi/3.2.1
  2) binutils/2.28     8) FFTW/3.3.7                     14) ncurses/6.0      20) Python/3.6.4
  3) GNU/6.4.0-2.28    9) ScaLAPACK/2.0.2-OpenBLAS-0.2.20 15) libreadline/7.0 21) Java/1.8.0_152
  4) OpenMPI/2.1.2    10) bzip2/1.0.6                    16) Tcl/8.6.8        22) MATLAB/2018a
  5) tbb/2018_U3      11) zlib/1.2.11                    17) SQLite/3.21.0    23) powertools/1.2
  6) imkl/2018.1.163  12) Boost/1.67.0                   18) GMP/6.1.2


[parvizim@dev-intel18 ~]$
```

Agenda

- Overview of the HPCC
- Logging on to the HPCC
- Simple example
  - Navigating Files
  - Module System
  - ***Submitting a job***
- Where to get help

Two Components of a Submission Script

- Resource Requests
  - Commands to communicate to the Scheduler
  - Resources you will need (time, memory, etc.)
  - Other settings (mail, names, etc.)
- List of Commands
  - What you would typically type on a command line

```
#!/bin/bash
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --time=00:06:00
#SBATCH --mem=1gb

# output information about how this job is running
using bash commands

echo "This job is running on $HOSTNAME on `date`"

# run the "hello" program in this directory

time ./example
```

Linux

Linux on the HPCC

Exercise: Write a bash script ‘my_job.sb’ to schedule a SLURM job that runs your ‘hello.c’ script

```
#!/bin/bash                                    #Tell the shell to interpret bash

##### SLURM Resource Requests #####

#SBATCH --time=0-00:10                         #How long the job will run (days-hours:minutes)
#SBATCH --nodes=1                              #How many compute nodes the job needs
#SBATCH --ntasks=1                             #How many concurrent tasks the job needs
#SBATCH --cpus-per-task=1                      #How many CPUs each task needs
#SBATCH --mem-per-cpu=1G                       #How much memory each CPU needs

##### SLURM Administrative Settings #####

#SBATCH --job-name HelloWorld                  #Name the job for convenience
#SBATCH --output=%x-%j.SLURMout                #Name the output file (JobName-JobNumber.SLURMout)
#SBATCH --mail-type=ALL                        #Tell SLURM to email you when job starts, stops, error
#SBATCH --mail-user=                           #Provide SLURM your email address


##### bash  Commands to Run #####

module purge                            #unload all modules

module load GNU/8.2.0-2.31.1            #load the GNU compiler

cd /mnt/home/                           #Navigate to the directory contaning hello.c

gcc hello.c -o hello                    #Run the command to compile hello.c

./hello                                 #Run the compiled executable hello
```

Linux

Linux on the HPCC

Exercise: Submit ‘my_job.sb’ to SLURM with the **sbatch** command

```
[parvizim@dev-intel18 ~]$ sbatch my_job.sb
Submitted batch job 3479290
[parvizim@dev-intel18 ~]$
```

Type in Your Terminal:

[user@computer] $ sbatch my_job.sb

Other Useful Commands

- sq – Show the contents of your queue
- qs – Show the status of your queue
- scancel ########- Cancel a job with number #
- js –j ####### - Show job stats

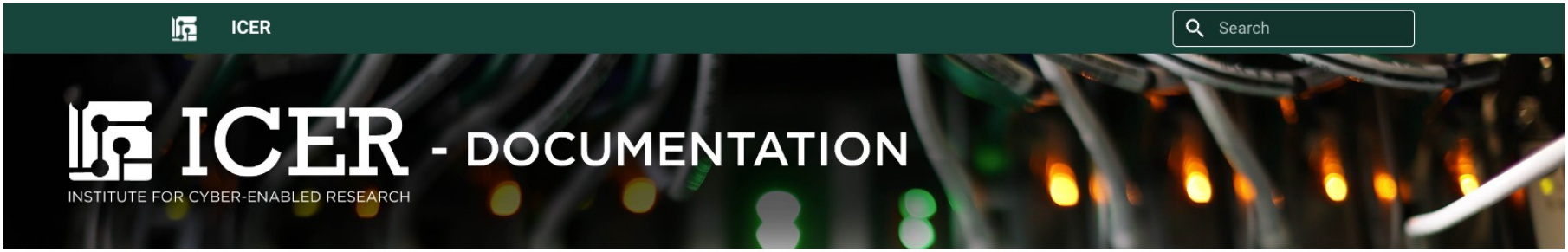
Agenda

- Overview of the HPCC
- Logging on to the HPCC
- Simple example
  - Navigating Files
  - Module System
  - Submitting a job
- ***Where to get help***

Linux

Linux on the HPCC

HPCC Documentation docs.icer.msu.edu



MSU HPCC User Documentation

- Home
- Access the HPCC
- HPCC resources
- Computing environment
- File systems
- File transfer
- Software module system
- Powertools
- Job submission by SLURM
- Software tutorials
- Training material
- FAQ
- Lab Notebooks
- Tags

Home

Getting Access to the HPCC

For potential users with an MSU NetID, accounts must be requested by a MSU tenure-track faculty member. Researchers at partner institutions should use the mechanism specified by their institution's agreement with MSU. For more information, see: Obtain an HPCC Account and on the ICER website.

CPU and GPU Time Limits

Non-buyin users are limited to 500,000 CPU hours (30,000,000 minutes) and 10,000 GPU hours (600,000 minutes) every year (from January 1st to December 31st). More information is available at Job Policies.

Table of contents

- Getting Access to the HPCC
- CPU and GPU Time Limits
- Buy-in Options
- Questions?
- Online Helpdesk Hours
- HPCC Workshops and Training
- Acknowledgements

Office Hours

- Mondays and Thursdays 1-2pm
- Microsoft Teams
  - https://docs.icer.msu.edu/virtual_help_desk/

Another Example

Use the rest of the time (if any) to explore other curated examples. Ask Questions...

1. See all of the examples

   getexample

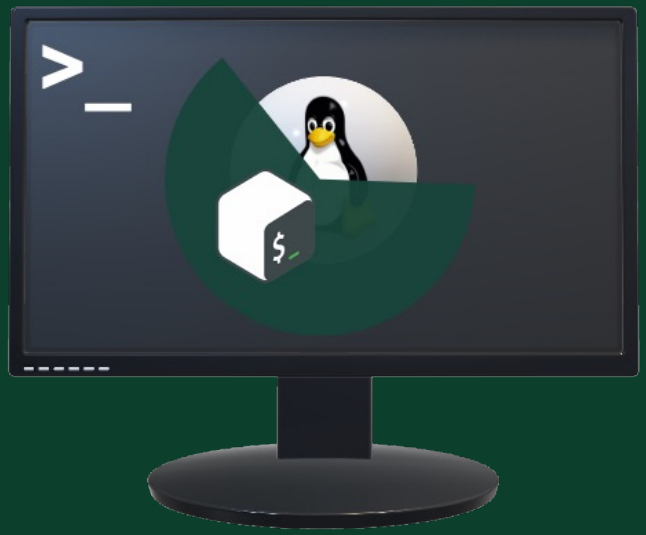
2. Copy an example

   getexample helloMPI

3. Review the example

   cd helloMPI

   cat README



Contact ICER

icer.msu.edu/contact

ICER
INSTITUTE FOR CYBER-ENABLED RESEARCH

Thanks

ICER user support team

Mahmoud Parvizi
Nanye Long
Xiaoge Wang
Andrew Fullard
Claire Kopenhafer
Nicholas Panchy
Craig Gross
Pat Bills

ICER
INSTITUTE FOR CYBER-ENABLED RESEARCH