UF Research Computing

Unix Command-Line
Modules
Scheduler

Matt Gitzendanner <magitz@ufl.edu>
Alex Moskalenko <om@hpc.ufl.edu>

UF Research Comuputing, UF HPC Training Session

Outline

- ✦ Basic Unix Command-Line
- ✦ Environment Modules
- ✦ Scheduler submission scripts

UF Research Computing/HPC Center

Tools

ssh client to connect to submit.hpc.ufl.edu

e.g.: Terminal, PuTTY

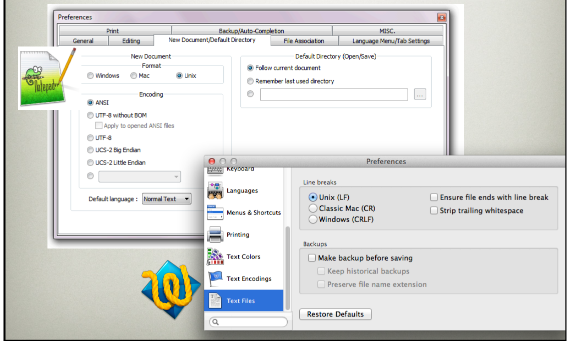
SFTP client to move files to/from your computer

e.g.: Cyberduck, FileZilla

Text editor to prepare files
Especially on Windows, be sure to convert DOS line breaks to Unix, and *don't use Word*
Both have SFTP built in

e.g.: TextWrangler, Notepad++

Unix line breaks



New Document
Format
○ Windows ○ Mac ● Unix
Encoding
● ANSI
○ UTF-8 without BOM
☐ Apply to opened ANSI files
○ UTF-8
○ UCS-2 Big Endian
○ UCS-2 Little Endian
Default language : Normal Text

Default Directory (Open/Save)
● Follow current document
○ Remember last used directory

Line breaks
● Unix (LF)
○ Classic Mac (CR)
○ Windows (CRLF)
☐ Ensure file ends with line break
☐ Strip trailing whitespace

Backups
☐ Make backup before saving
☐ Keep historical backups
☐ Preserve file name extension

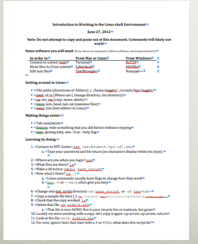
Command Line

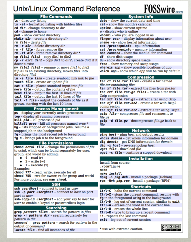
- ✦ Rather than a mouse, Unix/Linux use text entered at the command line
- ✦ Takes some getting used to, but not too hard
- ✦ MUCH better for complex or repetitive tasks—can be easily scripted

---

Learn by doing…

Handouts
Also available on Wiki




Unix/Linux Command Reference
FOSSwire.com

Environment Modules

Notice the velvet.pbs script

```
module load velvet
mkdir assembly
velveth assembly/ 21 -fasta –short test_reads.fa
velvetg assembly/
```

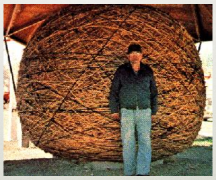
What's this module load thing?

And how is it that we don't need to put the full path to velvet in the command?

The biggest ball of twine in…

App versions

Library versions



Compilers

ENV vars

Dependencies

Paths

Python versions

MPI implementations

Modules

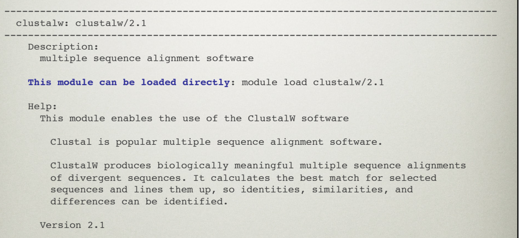
- ✦ Users are free of the gory details
- ✦ Many application versions coexist
- ✦ Easy switching of compilers/MPIs
- ✦ Easy addition of dependencies
- ✦ Automated environment setup (paths, variables, licenses, etc.…)
- ✦ Powerful search and help system
- ✦ Stable pipelines

Focus on Science, not busywork

Basic usage - Discover

- module spider
- module spider clu
- module spider netcdf/3.6.3
- module avail
- module avail clu (what does it find?)
- module key clu

Discover – `module spider clu`



```
-------------------------------------------------------------------------------------
  clustalw: clustalw/2.1
-------------------------------------------------------------------------------------
    Description:
      multiple sequence alignment software

    This module can be loaded directly: module load clustalw/2.1

    Help:
      This module enables the use of the ClustalW software

        Clustal is popular multiple sequence alignment software.

        ClustalW produces biologically meaningful multiple sequence alignments
        of divergent sequences. It calculates the best match for selected
        sequences and lines them up, so identities, similarities, and
        differences can be identified.

      Version 2.1
```

Discover variants

```
[magitz@submit1 2012-06-27]$ module spider raxml/7.3.0.0307

Rebuilding cache file, please wait ... done


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

  raxml: raxml/7.3.0.0307

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

    Description:

      Maximum Likelihood Tree Inference Tool - Standard version


    This module can be loaded directly: module load raxml/7.3.0.0307


    Additional variants of this module can also be loaded after the loading the

     following modules:

      intel/11.1, openmpi/1.4.3
```

Discover – `module avail`

```
[magitz@submit1 2012-06-27]$ module load intel/11.1

[magitz@submit1 2012-06-27]$ module avail


------------- /apps/lmod/modulefiles/compiler/intel/11_1 -------------

  amber/11                  gamess/11-Aug11    mrbayes/3.2.1

  amber/12 (default)        gulp/403           mvapich2/1.5.1

  bayesphylogenies/2.0      hdf5/1.8.7         nfft/3.1.3

  cuda/4.0                  lammps/17Sep11     openmpi/1.3.4

  dock/6.5                  libxc/1.1.0        openmpi/1.4.3 (default)


------------ /apps/lmod/modulefiles/core ---------------------------------
```

Long list of core module files

Basic usage - Use

- `module list`
- `module load clustalw`
- `module load python/2.6.5`
- `module add`
- `module del, rm, unload clustalw` – pick one!
- `module load intel/10.1 openmpi/1.2.7 mrbayes`

Use – load/unload

- `module load clustalw`
- `module list`

```
Currently Loaded Modules:
  1) clustalw/2.1
```

- `module unload clustalw` OR `module purge`
- `module list`

```
Warning: No modules installed
```

Basic usage

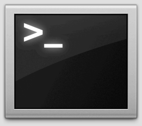
```
[magitz@submit1 2012-06-27]$ module spider mrbayes/3.2.1
Rebuilding cache file, please wait ... Done
----------------------------------------------------------------------
  mrbayes: mrbayes/3.2.1
----------------------------------------------------------------------
    Description:
      Bayesian inference of phylogeny

    This module can only be loaded through the following modules:
      intel/11.1
      intel/11.1, openmpi/1.4.3
...



[magitz@submit1 2012-06-27]$ module load intel/11.1 openmpi mrbayes
[magitz@submit1 2012-06-27]$ module list
Currently Loaded Modules:
  1) intel/11.1  2) openmpi/1.4.3  3) mrbayes/3.2.1
```

Job Scheduling and Usage

PBS

Cluster basics

Head node(s)



Login to head node

Scheduler

Tell the scheduler what you want to do

Compute resources

Your job runs on the cluster

Scheduling a job

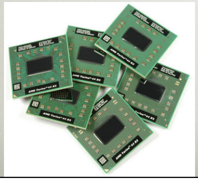
- ✦ Need to tell scheduler what you want to do
  - How many **CPUs** you want and how you want them grouped
  - How much **RAM** your job will use
  - Information about **how long** your job will run
  - The **commands** that will be run

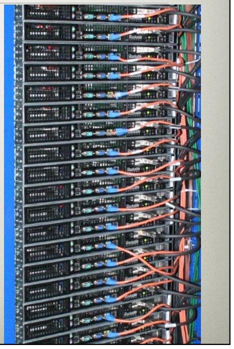
Scheduler

Tell the scheduler what you want to do

Nodes and processors

```
#PBS –l nodes=1:ppn=4
#PBS –l nodes=2:ppn=8
```





RAM

```
#PBS –l pmem=900mb
```

- Lots to consider, but do your best at estimating RAM needed for job
- Over about 2GB of RAM, “costs” toward CPU allocation
- Wasted RAM leads to idle CPUs and low job throughput

Walltime

```
#PBS –l walltime=00:50:00
```

- Fairly straight forward
- As with all resource requests, accuracy helps ensure ***your*** jobs and all other jobs will run sooner

Scheduler

Tell the scheduler what you want to do

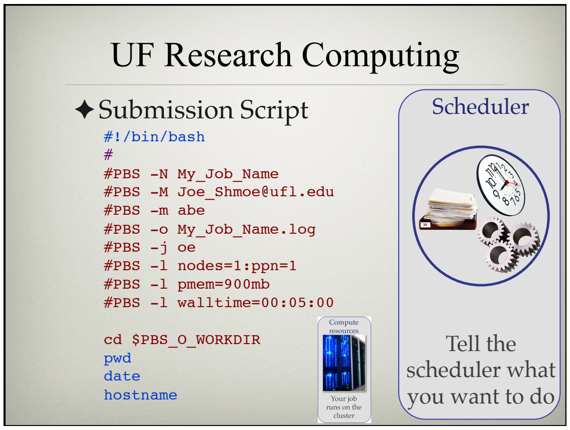
UF Research Computing

✦ Ordinary Shell Script

```
#!/bin/bash

pwd
date
hostname
```

Read the manual for your application of choice.

Commands typed on the command line can be put in a script.

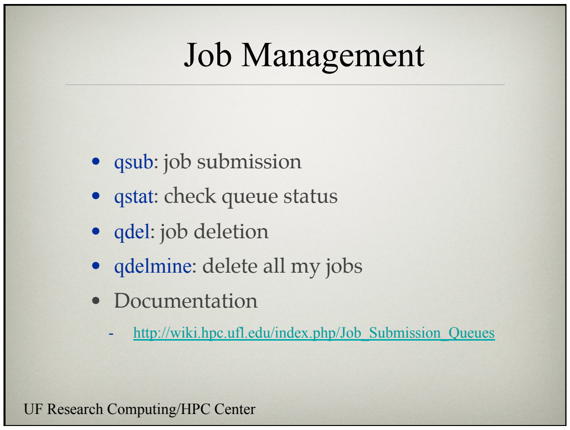


UF Research Computing

✦ Submission Script

```
#!/bin/bash
#
#PBS -N My_Job_Name
#PBS -M Joe_Shmoe@ufl.edu
#PBS -m abe
#PBS -o My_Job_Name.log
#PBS -j oe
#PBS -l nodes=1:ppn=1
#PBS -l pmem=900mb
#PBS -l walltime=00:05:00

cd $PBS_O_WORKDIR
pwd
date
hostname
```

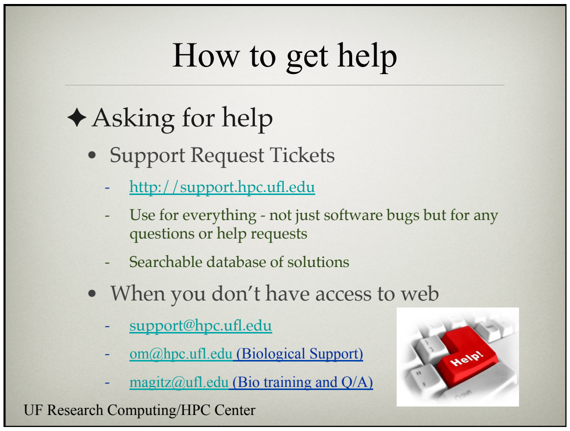
Tell the scheduler what you want to do



Job Management

- qsub: job submission
- qstat: check queue status
- qdel: job deletion
- qdelmine: delete all my jobs
- Documentation
  - http://wiki.hpc.ufl.edu/index.php/Job_Submission_Queues

UF Research Computing/HPC Center

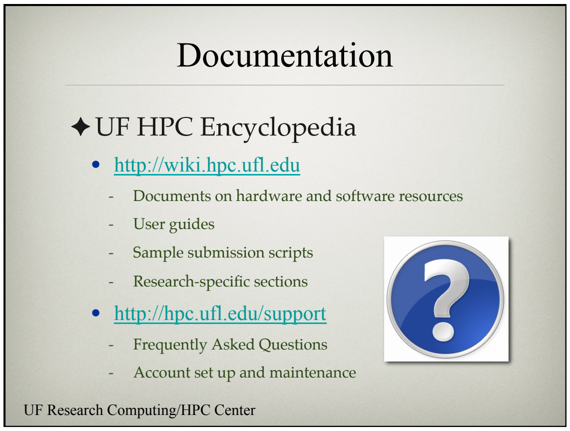


How to get help

✦ Asking for help

- Support Request Tickets
  - http://support.hpc.ufl.edu
  - Use for everything - not just software bugs but for any questions or help requests
  - Searchable database of solutions
- When you don't have access to web
  - support@hpc.ufl.edu
  - om@hpc.ufl.edu (Biological Support)
  - magitz@ufl.edu (Bio training and Q/A)

UF Research Computing/HPC Center

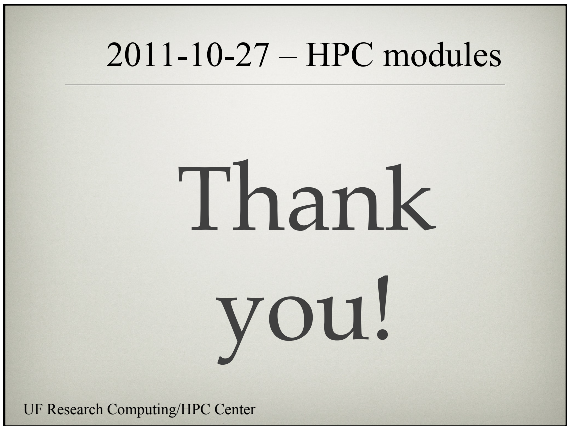


Documentation

✦ UF HPC Encyclopedia

- http://wiki.hpc.ufl.edu
  - Documents on hardware and software resources
  - User guides
  - Sample submission scripts
  - Research-specific sections
- http://hpc.ufl.edu/support
  - Frequently Asked Questions
  - Account set up and maintenance

UF Research Computing/HPC Center



2011-10-27 – HPC modules

Thank you!

UF Research Computing/HPC Center