O2 – Tips and Tricks

RC INFORMATIONAL SESSION

Presenters: Research Computing Consultants (RCCs)

HARVARD
MEDICAL SCHOOL

RESEARCH COMPUTING

How to know if you're requesting too much RAM or CPUs on your jobs?

O2_jobs_reports:

- Customizable queries allow you to select specific time intervals, partitions, jobs states and more.

- Three available output formats

- Creates a comprehensive report for your jobs or everyone in your Lab

Create your own modulefiles

Here's an example modulefile (hmmer/3.2.2.lua)

```
help([[
For detailed instructions, go to:
    http://hmmer.org

PDF documentation current available at:
    http://eddylab.org/software/hmmer/Userguide.pdf
]])

whatis("Version: 3.3.2")
whatis("Keywords: sequence, HMM, hidden, Markov, model, homolog, alignment")
whatis("URL:  http://hmmer.org ")
whatis("Description: biosequence analysis using profile hidden Markov models")


prepend_path( "PATH", "/n/app/hmmer/3.3.2-gcc-9.2.0/bin")

prepend_path( "MANPATH", "/n/app/hmmer/3.3.2-gcc-9.2.0/share/man")


family("hmmer")
```

`help()` can let you document the purpose of the module, as well as any helpful usage tips/whatever you'd like to include for informational purposes. This text shows up when using `module help <modulename>`.

`whatis()` allows you to specify metadata for your module.

`prepend_path()` allows you to modify environment variables that manage paths, such as PATH, MANPATH, etc. `append_path()` does something similar, but appends to the existing environment variable instead of prepends.

`family()` allows you to create mutually exclusive versions for a given software type (e.g., only one version of "hmmer" can be loaded at a time).

Strictly speaking, every element of this file is optional. For function, likely only one or more instances of `prepend_path()` are needed to set up a working module. For more advanced utilities to use in modulefile creation, refer to the Lmod documentation or contact rchelp@hms.harvard.edu for assistance.

Creating your own modulefiles, cont.

- O2 modulefile general hierarchy:

```
Core/<program>/<version>.lua
Linux/<program>/<version>.lua
Compiler/<compiler>/<version>/<program>/<version>.lua
MPI/<compiler>/<version>/<mpi>/<version>/<program>/<version>.lua
```

- We suggest for custom modulefile organization to at least have a central `software` location, with software titles underneath, e.g.,
    - `software`
        - `software1`
            - `version1.lua`
            - `version2.lua`
        - `software2`
            - `version1.lua`
            - `version2.lua`
        - Etc.

- Do whatever makes the most sense for yourself/the lab
- Once your modulefile hierarchy location/path is determined, add it to MODULEPATH via the following:
    - `module use /path/to/software`
    - Where `/path/to/software` is the location of your software hierarchy
    - It may be useful to add it to your `.bashrc` file to prevent having to run the above command every time you log in.
- Note that if you install a version of software that matches a version on O2, yours will take precedent.

HARVARD
MEDICAL SCHOOL | RESEARCH COMPUTING

Are you running out of space? Or seeing errors with your jobs?

- Perhaps you have hit a [storage limit](#)!

- Common error message: "No space left on device"

- Being over a storage limit will prevent you from:

  ◦ Adding/modifying files
  ◦ Writing hidden configuration/"housekeeping" files

- Use quota-v2 to check your storage usage and limits!

- If you're over a storage limit, remove files or move them to another storage location.

Are you running out space? Or seeing errors with your jobs? Cont.

- Use quota-v2 to check your storage usage and limits!

Your usage

Group usage

Exceeded limits

```
[rc@login06 ~ 2023-05-02 14:48:29]$ quota-v2
                        Usage for rc (As of 2023-05-02 14:00:00)
```

path	type	username	usage	storage limit	last update
/home	user	orchestra_rc	3.30 GiB	100.00 GiB	2023-05-02 14:00:00
/n/scratch3/users	user	orchestra_rc		10.00 TiB	2023-05-02 14:00:00
/n/standby/hms/it/rccg	group		10.53 TiB	14.00 TiB	2023-05-02 14:00:00

```
                              Exceeded quotas
```

path	type	username	usage	storage limit	last update	grace remaining
/n/standby/hms/it/rccg_q1	group		502.93 MiB	500.00 MiB	2023-05-02 14:00:00	None
/n/standby/hms/it/rccg_q2	group		502.93 MiB	500.00 MiB	2023-05-02 14:00:00	None
/n/standby/hms/it/rccg_q6	group		1.18 GiB	500.00 MiB	2023-05-02 14:00:00	None

```
You have lost write access to one or more of the above paths. Please clear them up to regain write access.
[rc@login06 ~ 2023-05-02 14:49:13]$
```

Available on the O2 command line (see above) or on the O2Portal.

RESEARCH COMPUTING
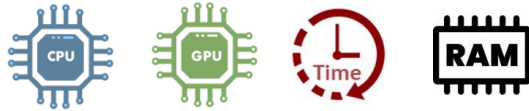
HARVARD
MEDICAL SCHOOL

O2 Portal – An alternative or complement to the command line

- Access to O2 resources

- Access to popular applications

- Transfer data to/from O2

- Move data between O2 filesystems

- View/edit files

Tmux - why is important?

- Tmux is a terminal multiplexer in Linux

  ◦ Enable pseudo-terminal instances
  ◦ Detach/Attach from a pseudo-terminal

- Use cases

  ◦ Data Transfer on the O2 transfer cluster
  ◦ Continue to run a command *after* disconnecting from a session

Loops - save time and streamline a computational workflow

```
for fq in *fq; do

    sbatch -p short –t 10:0  --wrap "fastq_to_fasta -i $fq -o ${fq%.fq}.fa"

Done


for file in *fq; do

  batch="$batch $file"

  [ ${#batch} -gt  100 ]  &&  sbatch -p short –t 1:0:0 --wrap "for fq in $batch; do fastq_to_fasta -i \$fq -o \${fq%.fq}.fa; done"

  Batch=""

done

 [ -z "$batch" ] || sbatch -p short –t 1:0:0 --wrap "for fq in $batch; do fastq_to_fasta -i \$fq -o \${fq%.fq}.fa; done"
```

Don'ts

- Do not run your analysis on the login nodes

- Do not submit short [seconds] jobs to O2

- Do not request too much memory

- Do not be afraid to ask for help! Contact at rchelp@hms.harvard.edu