### Quick Links

<table>
<thead>
<tr>
<th>HPC at NYU</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPC Newsletter</td>
</tr>
<tr>
<td>Getting started on Mercer</td>
</tr>
<tr>
<td>Getting an account</td>
</tr>
<tr>
<td>Logging in</td>
</tr>
<tr>
<td>Windows</td>
</tr>
<tr>
<td>Mac/Linux</td>
</tr>
<tr>
<td>Clusters and Storage</td>
</tr>
<tr>
<td>Mercer (HPC)</td>
</tr>
<tr>
<td>Dumbo (Hadoop)</td>
</tr>
<tr>
<td>Transferring data to/from the clusters</td>
</tr>
<tr>
<td>Writing and submitting jobs with qsub</td>
</tr>
<tr>
<td>Available software</td>
</tr>
</tbody>
</table>
Running jobs on the NYU HPC clusters

Queues
Writing a job script

Submitting a Job
Basic qsub options and directives
Setting resource limits
Running MPI jobs
Running GPU jobs
Running R jobs
Running Matlab, Gaussia n, etc jobs
Working interactively
Job dependencies and delaying starting
Running many similar jobs
Setting job priorities

Where in the queue is my job, and why?

**Monitor jobs with qstat**

```
qs
ta
t in
t m or
t e
d
t et
t ail
```

**See what is running where with pbsstop**

When will my job start?
Why hasn't my job started?
Canceling a job

Basic R jobs

Multiple R versions exist in HPC environment. To check what are available, on mercer:

```bash
$ module avail r
```

```
--------------------------------------------- /share/apps/modules/modulefiles
---------------------------------------------
  r/intel/3.0.3                       ray/openmpi/intel/20160114
  reproZip/intel/1.0.3                rpy2/intel/2.5.6
  r/intel/3.1.2                       rdkit/intel/201409.2        requests/2.7.0
  rsem/intel/1.2.15                   rseqc/intel/2.3.9
  r/intel/3.2.0                       recon/intel/1.08            rmblast/2.2.28
  ribopicker/0.4.3                    rseqc/intel/2.3.9
  r/intel/3.2.2                       recon/intel/1.08            rmblast/2.2.28
  rstudio/0.98.1028                   rendertoolbox3/2.1-18       rose/20151118
  randfold/intel/2.0                  repeat_modeler/1.0.8
  rtax/0.984                          repeat_masker/4.0.5
  raxml/intel/7.3.0                   repeat_masker/4.0.5
  rosetta/intel/54167                 ruby/gnu/2.1.1
  raxml/intel/8.0.23                  repeat_modeler/1.0.8
  rosetta/openmpi/intel/2014.35.57232 repeat_masker/4.0.5
  raxml/intel/8.2.5                   repeat_modeler/1.0.8
  rosetta/intel/54167                重复计数器/intel/1.0.5
```

Suppose we want to use 3.2.2, run these commands:

```bash
$ module purge
$ module list
No Modulefiles Currently Loaded.
$ module load r/intel/3.2.2
$ module list
Currently Loaded Modulefiles:
  1) intel/14.0.2  4) cairo/gnu/1.12.16  7) mpfr/gnu/3.1.2  10)
inpc/intel/8.34    13) openssl/gnu/1.0.1g 16) r/intel/3.2.2
  2) zlib/intel/1.2.8 5) expat/intel/2.1.0    8) mpc/gnu/1.0.2
  3) bzip2/intel/1.0.6 6) gmp/gnu/5.1.3   9) gcc/4.8.2
  14) curl/intel/7.38.0
  15) openmpi/intel/1.6.5
$ R
```

We first clean up the environment by doing 'module purge'. Then we load the R version selected, check what are available in current environment. We can see that R 3.2.2 is indeed loaded along with its dependency modules. Let's try this basic R example. We name it "example.R":

```r
df <- data.frame(x=c(1,2,3,1), y=c(7,19,2,2))
df
indices <- order(df$x)
order(df$x)
df[indices,]
df[rev(order(df$y)),]
```
Below is the screen scene while running it on mercer:

```r
$ R
R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-centos-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> df <- data.frame(x=c(1,2,3,1), y=c(7,19,2,2))
> df
   x y
 1 1  7
 2 2 19
 3 3  2
 4 1  2
> indices <- order(df$x)
> order(df$x)
[1] 1 4 2 3
> df[indices,]
   x y
 1 1  7
 4 1  2
 2 2 19
 3 3  2
> df[rev(order(df$y)),]
   x y
 2 2 19
 1 1  7
 4 1  2
 3 3  2
> quit()

Save workspace image? [y/n/c]: n
```

What is shown above is a simple demo case on login nodes. For real interactive analysis scenario, users are encouraged to run on compute nodes using the `qsub` command (with `-I` and `-X` switched on) to request dedicated resources, e.g.:

```bash
$ qsub -I -X -l nodes=1:ppn=4 -l mem=4GB -l walltime=2:00:00
qsub: waiting for job 7777264 to start
qsub: job 7777264 ready

$ module load r/intel/3.2.2
$ R
```

Besides running our analysis interactively, long running and big data crunching jobs ought to be submitted to the batch system PBS. The
"example.R" can be submitted to PBS to run in batch mode with a job script as:

```
#!/bin/bash

#PBS -V
#PBS -N RTest
#PBS -l nodes=1:ppn=1,mem=2GB,walltime=00:10:00
#PBS -M NetID@nyu.edu
#PBS -m abe

module purge
module load r/intel/3.2.2
cd /share/NetID//r-workdir
R --no-save -q -f example.R > example.out 2>&1
```

The output and log of this exemplar PBS job are available in /share/apps/examples/r/basic/, which is accessible on mercer. More examples to be added.