Introduction to the Ginsburg High Performance Computing Cluster

Columbia Biostatistics Computing Club

Ting-Hsuan Chang

Mar 10, 2025

Table of Contents

1. Getting Started

2. Installing R Packages

3. Running R on Ginsburg
   - Preparing files
   - Transferring files
   - Job submission and output

When to use HPC?

- You want to evaluate a new statistical method by running it 1000 times on simulated data, but each run is time-consuming. Instead of running these simulations sequentially, you can use HPC to execute many of them **in parallel**, which will significantly reduce the total runtime.
- If you have very large datasets, opening/analyzing them will probably crash your computer.

Contact your advisor or Ying(?) to set up an account.

Your account will be associated with a group, e.g., biostats.

Access Ginsburg through terminal:

```
$ ssh tc3255@ginsburg.rcs.columbia.edu
```

Check out the Ginsburg and Slurm documentations.

Table of Contents

After logging in to Ginsburg, create a directory to install R packages. You only need to do this once.

```
$ mkdir /burg/biostats/users/tc3255/rpackages
```

Start R:

```
$ module load R
$ R
```

Add this directory to library paths (where R looks for packages):

```
> .libPaths("/burg/biostats/users/tc3255/rpackages/")
```

Check the library paths:

```
> .libPaths()
[1] "/burg/biostats/users/tc3255/rpackages"
[2] "/burg/opt/R-4.4.2/lib64/R/library"
```

Install a package (e.g., “sm”):

```
> install.packages("sm", lib="/burg/biostats/users/tc3255/rpackages")
```

Test to see if package can be called:

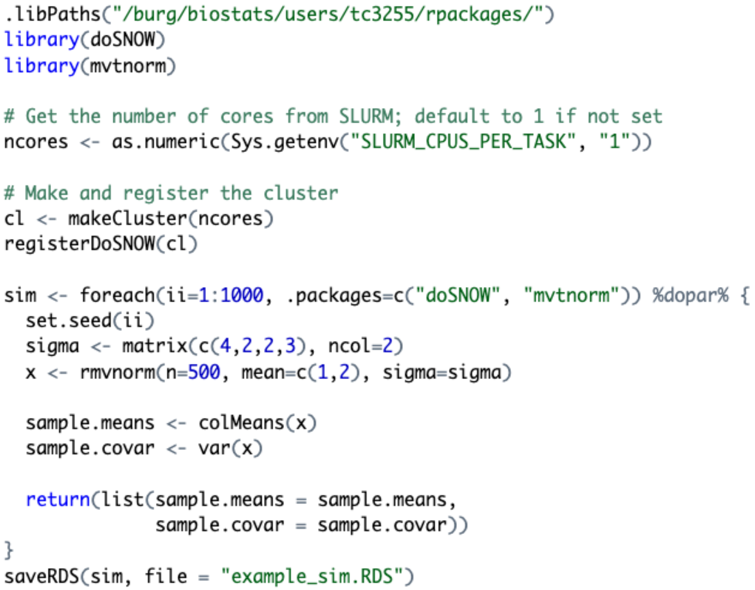
```
> library(sm)
```

Add the following line to the top of your .R file to make sure R knows where to look for your installed packages.

```
.libPaths("/burg/biostats/users/tc3255/rpackages/")
```

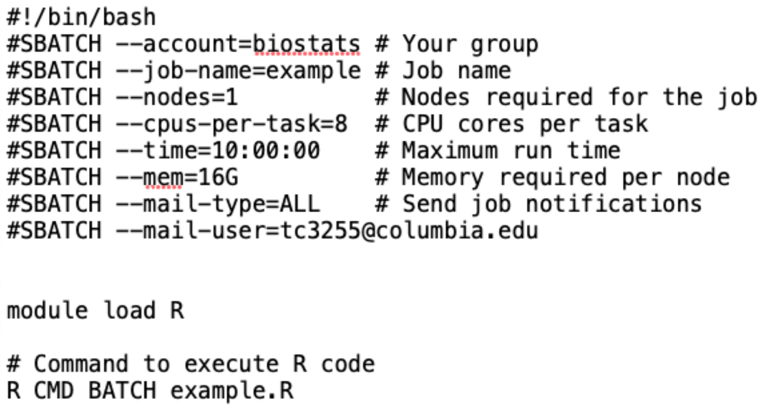
Table of Contents

R code (“example.R”)



```
.libPaths("/burg/biostats/users/tc3255/rpackages/")
library(doSNOW)
library(mvtnorm)

# Get the number of cores from SLURM; default to 1 if not set
ncores <- as.numeric(Sys.getenv("SLURM_CPUS_PER_TASK", "1"))

# Make and register the cluster
cl <- makeCluster(ncores)
registerDoSNOW(cl)

sim <- foreach(ii=1:1000, .packages=c("doSNOW", "mvtnorm")) %dopar% {
  set.seed(ii)
  sigma <- matrix(c(4,2,2,3), ncol=2)
  x <- rmvnorm(n=500, mean=c(1,2), sigma=sigma)

  sample.means <- colMeans(x)
  sample.covar <- var(x)

  return(list(sample.means = sample.means,
              sample.covar = sample.covar))
}
saveRDS(sim, file = "example_sim.RDS")
```

Job script ("example.sh")



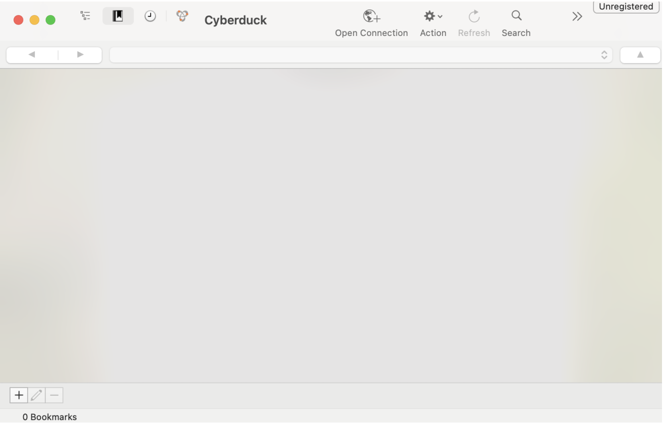
```
#!/bin/bash
#SBATCH --account=biostats # Your group
#SBATCH --job-name=example # Job name
#SBATCH --nodes=1          # Nodes required for the job
#SBATCH --cpus-per-task=8  # CPU cores per task
#SBATCH --time=10:00:00    # Maximum run time
#SBATCH --mem=16G          # Memory required per node
#SBATCH --mail-type=ALL    # Send job notifications
#SBATCH --mail-user=tc3255@columbia.edu


module load R

# Command to execute R code
R CMD BATCH example.R
```

See Ginsburg documentation for basic job directives.

Transferring files between local and HPC

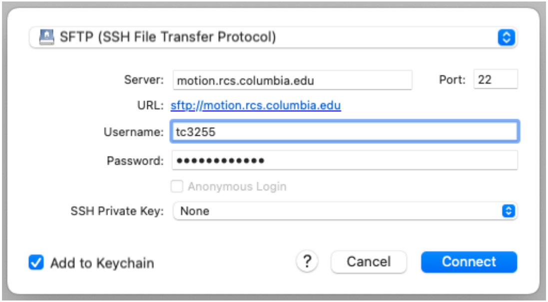
Cyberduck

Open Cyberduck, click “Open Connection” on the top.



Choose SFTP

Server: motion.rcs.columbia.edu



SFTP (SSH File Transfer Protocol)

Server: motion.rcs.columbia.edu

Port: 22

URL: sftp://motion.rcs.columbia.edu

Username: tc3255

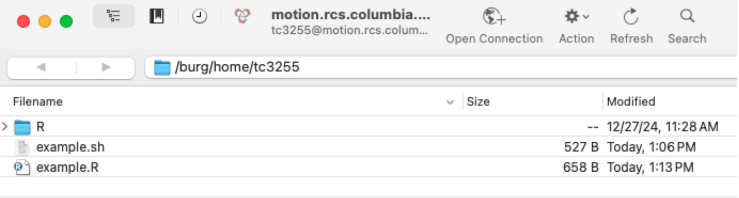
Password: •••••••••••••

☐ Anonymous Login

SSH Private Key: None

☑ Add to Keychain

Upload your R code and job script to Cyberduck.



/burg/home/tc3255

Filename	Size	Modified
R	--	12/27/24, 11:28 AM
example.sh	527 B	Today, 1:06 PM
example.R	658 B	Today, 1:13 PM

Job submission

Once your local files are transferred to HPC, log in to Ginsburg and submit your job script:

```
$ sbatch example.sh
```

Check job status:

```
$ squeue -u tc3255
```

Cancel a job:

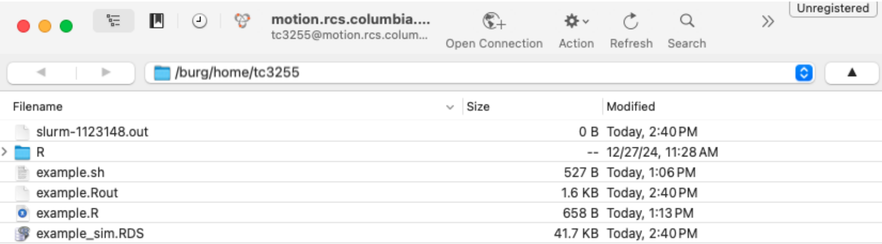
```
$ scancel your-job-id
```

Show information about a completed job:

```
$ sacct -j your-job-id
```

Job output

Refresh your cyberduck directory.



Filename	Size	Modified
slurm-1123148.out	0 B	Today, 2:40 PM
R	--	12/27/24, 11:28 AM
example.sh	527 B	Today, 1:06 PM
example.Rout	1.6 KB	Today, 2:40 PM
example.R	658 B	Today, 1:13 PM
example_sim.RDS	41.7 KB	Today, 2:40 PM

If your job failed, download .Rout to check for errors.

“example_sim.RDS” contains simulation results from the example.