Class 11:
Great Lakes and parallel statistical computing in R

Edward L. Ionides

Outline

Objectives for this class

- To log in to Great Lakes and run a batch R script.
- To adapt `doParallel` and `foreach` for the cluster.

Requirements

We follow Section 1.2 of the Great Lakes user guide. As preliminaries, you need:

- A Slurm account. You should already have a primary account, `stats_dept1`, and a smaller backup account for if you exhaust your resourses, `stats_dept2`.
- A Great Lakes cluster login account. If you have not yet filled in the form at `https://arc-ts.umich.edu/greatlakes/user-guide/` then do so.
- A umich internet address. Use the umich VPN if you are not on campus.

Connecting to Great Lakes with macOS or Linux

1. Open a Terminal window. On a Mac, this can be done using Control-Spacebar and typing Terminal.
2. Type

   ```
   ssh uniqname@greatlakes.arc-ts.umich.edu
   ```

   where `uniqname` is your uniqname.
3. Login with your Kerberos level-1 password, and Duo two-factor authentication.

This creates a remote command shell on Great Lakes.

Connecting to Great Lakes with Windows

This is essentially the same as for macOS.

1. Follow instructions to install PuTTY at https://documentation.its.umich.edu/node/350
2. Launch PuTTY and enter `greatlakes.arc-ts.umich.edu` as the host name, then click open. If you receive a "PuTTY Security Alert" pop-up, this is completely normal, click the "Yes" option. This will tell PuTTY to trust the host the next time you want to connect to it. From there, a terminal window will open; you will be required to enter your UMICH uniqname and then your Kerberos level-1 password in order to log in. Please note that as you type your password, nothing you type will appear on the screen; this is completely normal. Press "Enter/Return" key once you are done typing your password.
3. Complete the request for Duo two-factor authentication.

This creates a remote command shell on Great Lakes.

Moving files on and off Great Lakes: scp

On Mac or Linux, you can use `scp` which has similar syntax to `cp`. To copy `myfile` on your laptop to a subdirectory `mydir` of your home directory on greatlakes:

```
scp myfile uniqname@greatlakes-xfer.arc-ts.umich.edu:mydir
```

To copy an entire directory, use the `-r` flag for recursive copy:

```
scp -r mydir uniqname@greatlakes-xfer.arc-ts.umich.edu:
```

These commands can also be reversed to copy files from Great Lakes to your machine. The following copies `mydir` back to the current working directory:

```
scp -r uniqname@greatlakes-xfer.arc-ts.umich.edu:mydir .
```

You will need to authenticate via Duo to complete the file transfer. On Windows, you can use WinSCP or FileZilla.

Cluster batch workflow

1. You create a batch script and submit it as a job
2. Your job is scheduled, and it enters the queue
3. When its turn arrives, your job will execute the batch script
4. Your script has access to all applications and data
5. When your script completes, anything it sent to standard output and error are saved in files stored in your submission directory
6. You can ask that email be sent to you when your jobs starts, ends, or fails
7. You can check on the status of your job at any time, or delete it if it's not doing what you want
8. A short time after your job completes, it disappears

Useful batch commands

Submit a job

```
sbatch sample.sbat
```

Query job status

```
squeue -j jobid squeue -u uniqname
```

Delete a job

```
scancel jobid
```

Check a job script and estimate its start time

```
sbatch --test-only sample.sbat
```

More Slurm commands to try

`sacct -u user`	show recent job history
`seff jobid`	show cpu utilization for jobid

Acknowledgments

- This lesson is prepared for the STATS 810, Fall 2020.
- It builds on the Greal Lakes User Guide and notes by Charles Antonelli and John Thiels.

- Compiled on October 7, 2020.