CAC Help

These rough notes offers an introduction to CAC. More information can be found on the following websites:

<http://research-it.berkeley.edu/services/high-performance-computing/using-r-savio>

<https://thecoatlessprofessor.com/programming/working-with-r-on-a-cluster/>

<https://hcc-docs.unl.edu/display/HCCDOC/Submitting+R+jobs>

<https://cac.queensu.ca/wiki/index.php/Hardware:Frontenac>

1. Log onto the cluster:

Code: ssh -X hpc4300@login.cac.queensu.ca

\*Use your own hpc number though…

The terminal will then prompt you to enter your password.

1. Create a new directory (folder):

mkdir NewFolderName

1. Here are some nice general commands to know…

ls : List items in your current directory:

pwd :print the current working directory:

rm [filename]: delete the filename from your directory

cd : Change the directory. Example : cd /global/home/hpcg1578/Peter

cat : Print the contents of a file. Example: Cat namefile.txt

1. How to transfer files from your computer onto the CAC directory (For windows):

Download a program called WINSCP (This program is like FileZilla).

This program allows you to easily drag and drop files between the two directories.

1. How do you write your \*.sh files?

* I would first download Notepad ++
* When you save, always make sure to have the extension .sh

1. Submit your \*.sh files onto CAC:

* sbatch yourjob.sh

1. Sometimes when you submit your \*.sh jobs, you may encounter an error message such as:

“sbatch: error: Batch script contains DOS line breaks (\r\n)

“sbatch: error: instead of expected UNIX line breaks (\n)

* One possible solution is this code:

dos2unix yourjob.sh

* Now, try running your job again.

1. How to Run R jobs

* Check out test1.sh as well as Cluster\_Test1.R for a very simple example.
* Make sure though that your R script is in the same directory as your shell script. If your R script is in a different directory, then change your code to: [path to R file]/Rfile
* You could also use Rscript instead of R CMD BATCH in your \*.sh scripts…

1. A way to check on the status of the jobs you have submitted:

* squeue --start | grep hpc4300

1. How to install R packages onto the CAC cluster.

If the R codes you input contain calls such as library() or require() then it won’t work in the cluster. You will need to install the R packages yourself into the CAC directory. Here’s how to do it:

1. First, create a directory inside the CAC domain where you will save all of your R packages. (mkdir Foldername)
2. Type “R” (without quotations) into the CAC cluster. Your screen should now resemble what you would normally see when running R
3. It may be helpful to look at the arguments of the install.packages() function.

In CAC, type: install.packages(“R package you want”, lib=”Path to folder where you want to save your R packages”)

Note: This may take a while…

1. Close the terminal and login again to CAC
2. In all of your R script files, you will need to add the following line at the top:

.libPaths(“Path to the folder that has your R packages in CAC”)

For example, I made a folder called “RPackages” which contains all of my downloaded R packages on CAC. My path is : .libPaths("/global/home/hpc4300/RPackages")

* ILLUSTRATION: Please see test2.sh and Cluster\_Test2.R for an example. I use the “SKAT” package.
* Some R packages are very special (Ex: MDMR), and extra steps will be required to install these packages. This code may help:

1. At prompt type: module load gcc

(This changes the c compiler, which R uses to install packages)

1. Open R and then try installing MDMR again using install.packages(“MDMR”). If it installed successfully, library(MDMR) should work.
2. How to perform array jobs and how to call variables from shell script to your R script.

* Check out test4.sh and test4.R

1. For coding in bash, the syntax is a little different however all concepts in your usual programming are the same. The following website is helpful in figuring out the syntax:

<http://tldp.org/HOWTO/Bash-Prog-Intro-HOWTO.html>

* I’ve also compiled some very rough notes titled “Shell scripting tutorial.docx”.

1. How to run the ms program:

Here is a sample command I’ve used:

./ms 200 1 -t 5.0 -T

Here,

* 200 denotes the number of haplotypes I want to simulate
* 1 denotes the number of simulations I want to run
* -t denotes
* 5.0 denotes some mutation rate
* -T means I want the output to include tree data

To save the output of the ms file into some directory, use the following code:

./ms 200 1 -t 5.0 -T > /global/home/hpc4300/results.txt

\*Change to your own directory. Note that results.txt is the name and file type I give to the output file saved.