Computing resources in the Department of Statistics and Actuarial Science

Brad McNeney

Stat and Act Sci, SFU

Fall 2014

Computing Resources Overview

- Topics for discussion:
  - Network
  - Support
  - Equipment
- Goals:
  - Share information on what's available
  - Discuss what we need

Network

- Wired:
  - Historically a Unix network that existed as a sub-network of the Faculty of Applied Sciences network (FASnet)
  - Later diversified (Linux/Mac/Windows) and migrated to the campus-wide network
- Wireless
  - Coverage is spotty, but improving (?)

Support – History

- Has evolved over the years.
  - Originally had a dedicted network administrator shared with Math
    - Affiliated with FASnet's Network Support Group (NSG), but very independent.
  - In 2005 support more integrated with the NSG.
  - By 2010 we had no dedicated person and received support from the NSG through a centralized help ticket system called `gripe`.
  - Recently the NSG has been moved to SFU IT Services and renamed the Research Computing Group.
  - Our desktop support has been taken over by the Faculty of Science IT Support team.

Support – Current Status

- `http://stat.sfu.ca/it-support.html`
  - Science IT support team: `http://www.sfu.ca/science/faculty-support/IT-Support.html`, or, directly at `http://support1.mbb.sfu.ca/SelfService/Create.html?Queue=20`
  - Unix support through the Research Computing Group: email `research-support@sfu.ca` (no longer `gripe@fas.sfu.ca`).

Equipment – Personal Computers

- Personal laptops
  - Software (JMP, SAS, etc.) can be downloaded from
    http://www.sfu.ca/itservices/technical/software.html
  - Nate has some advice on SAS ...
- Desktop computers in faculty offices and student labs
  - Windows & Mac in faculty offices
  - Windows in P9309 and K9501 (Math/Stat Unified Lab Environment)
    - R/RStudio, SAS, various distributions of TeX, PuTTY
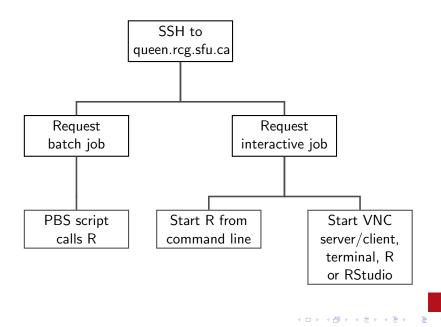 - Macs in IRMACS
    - Software?

Equipment – Servers

- The Department used to maintain servers for remote login.
- All servers have been merged with The Colony, a local cluster.
  - The Colony is intended for relatively short jobs and/or interactive computing
  - Using The Colony is the focus of the rest of the seminar.
  - Reference: `http://stat.sfu.ca/research/computing/getting-started-with-the-cluster.html`
- We also have access to WestGrid, the Western Canada HPC resource, for large-scale computation
  - For information about WestGrid, see
    `https://www.westgrid.ca/faq/prospective_users`
    `https://www.westgrid.ca/support/quickstart/new_users`
  - Faculty can apply for an account – see the FAQ "How do I get an account at a regional HPC consortium (like WestGrid)?"
  - Students need a faculty sponsor.
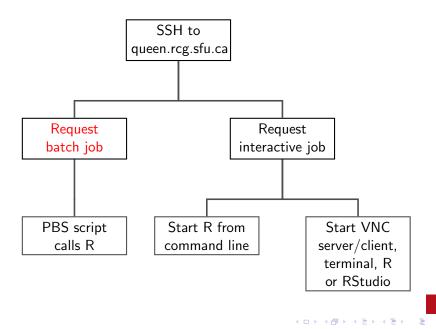
Colony Cluster – E.G., to Run R

SSH to queen.rcg.sfu.ca

- Before logging in to queen for the first time,
    - Run the password change script located at
      `https://my.sfu.ca/ChangePassword`
    - Join the `colony-users` email list: go to
      `https://maillist.sfu.ca/`, log in with your SFU ID, Go to list:
      colony-users, subscribe
- Log in with an application like PuTTY from a Windows computer,
  or ssh from a Mac or Linux terminal.

Colony Cluster to Run R

Request Batch Job

- Batch jobs are specified by a script that requests computing resources and executes the command (e.g., R).
- From the command-line on queen, change to the directory in which your batch script resides and use the qsub command; e.g.:

  ```
  cd Cluster
  qsub ex2.pbs
  ```

  to change to the subdirectory Cluster and submit the batch file ex2.pbs
    - Source available at
      http://people.stat.sfu.ca/~mcneney/Cluster/ex2.pbs.
    - The pbs extension stands for Portable Batch System, the software that schedules jobs.
- When the job is submitted you will get a job number.
- Check on the status of your job with qstat or checkjob <jobnumber>.

PBS Script Resource Requests

- Example from ex2.pbs

  ```
  #PBS -l nodes=16:ppn=1,pmem=1gb,walltime=0:10:00
  ```

  requests 16 nodes, 1 processor per node, each node has at least
  1GB memory, and the job will take at most 10 minutes.

- You must also identify your user group. For us this is
  `colony-users`

  ```
  #PBS -W group_list=colony-users
  ```

- The cluster can email you when your job begins (b), ends (e) or is
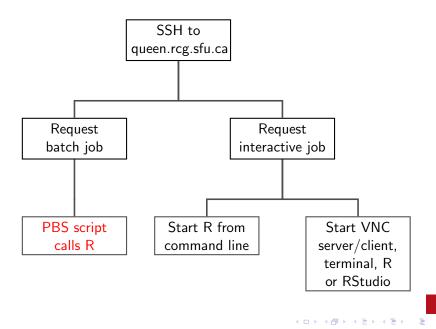 aborted (a):

  ```
  #PBS -m bea
  #PBS -M <yourSFUID>@sfu.ca
  ```

- You can also give your job a name, and specify the files to receive
  output and error messages:

  ```
  #PBS -N ex2run
  #PBS -o ex2run.out
  #PBS -e ex2run.err
  ```

Colony Cluster to Run R

PBS Script to Run R

- After the resource request lines of the PBS script come the commands to run your job.
  - what you'd do if you had just logged in to the cluster node yourself
- Software on the cluster is maintained in a collection of "modules"
  - See the Colony wiki for more on modules
  - Type `module avail` to get a list of all available modules.
  - To run R you need to load the R module:
    ```
    module load LANG/R/3.0.2
    ```
- Now change to the directory where your R scrip resides (assume this is the directory from which the script was submitted)
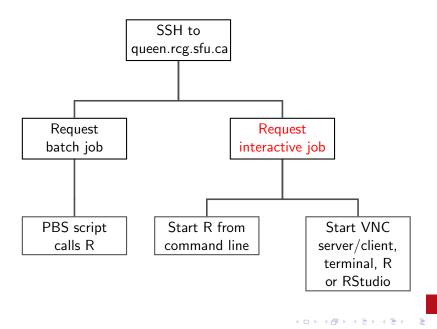 ```
  cd $PBS_O_WORKDIR
  ```
- Ready to run your R script
  ```
  R --vanilla -f ex2.R
  ```

# Parallel Programming in R

- The `parallel` library has functions to handle low-level communcation between the nodes/processors that you requested in your PBS file.
- For example, suppose you requested 4 nodes with 1 processor per node:
  - `cl = makeCluster(4,type="MPI")` will start four "worker" instances of R on the 4 nodes, establish communcation between them and save necessary information for future reference in the object `cl`.
- `parallel` also has functions to distribute the work across the worker processes.
- One example from ex2.R (available at http://people.stat.sfu.ca/∼mcneney/Cluster/ex2.R):
  - Basic idea: Use a parallel version of the `lapply()` function.
  - If X is a list, then `lapply(X,FUN)` applies FUN to each element of X.
  - Similarly, `clusterApply(cl,X,FUN)` applies FUN to each element of X, with each application executed on a different node of cl.

# Colony Cluster to Run R

# Request Interactive Job

- Use qsub `-I` to request an interactive job.
  - specify the resources, etc, right in the qsub command; e.g.:

    ```
    qsub -I -V -l walltime=01:00:00,nodes=1:ppn=1,pmem=1gb\
    -W group_list=colony-users
    ```
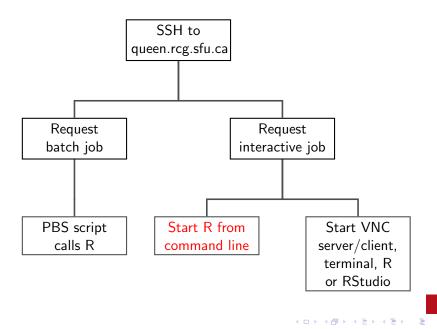
    (the backslash is the Unix line continuation character) to request one processor on one node or

    ```
    qsub -I -V -l walltime=01:00:00,nodes=1:ppn=4,pmem=1gb\
    -W group_list=colony-users
    ```

    to request four processors on one node, etc.

Colony Cluster to Run R

SSH to
queen.rcg.sfu.ca

Request
batch job

Request
interactive job

PBS script
calls R

Start R from
command line

Start VNC
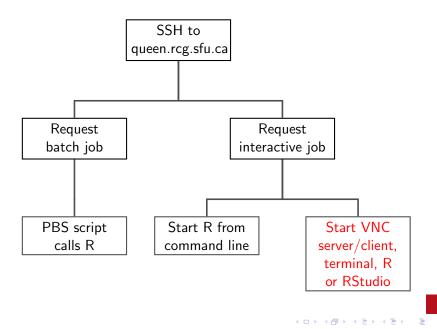server/client,
terminal, R
or RStudio

Start R from Command Line

- When your job starts you will be logged in to a node on the cluster.
- Act as if you logged in yourself; e.g.:
  ```
  cd Cluster
  module load LANG/R/3.0.2
  R
  ```

Colony Cluster to Run R

SSH to
queen.rcg.sfu.ca

Request
batch job

Request
interactive job

PBS script
calls R

Start R from
command line

Start VNC
server/client,
terminal, R
or RStudio

Start VNC Server, etc. (Experimental)

- Virtual Network Computing (VNC) allows sharing of desktops between computers.
- Steps:
  1. Start a VNC server on your cluster node with `vncserver` (see `http://stat.sfu.ca/research/computing/vncserver.html` for more information)
     - Record the desktop name.
  2. Start a VNC client on your computer (see `http://stat.sfu.ca/research/computing/vncclient.html` for more information)
     - This causes a desktop on the server to appear on your computer.
  3. Start a terminal on the remote desktop, and compute as usual
     - Now you can run R and see R graphics, run RStudio, etc.

Partial Recap of Resources

- Support http://stat.sfu.ca/it-support.html
  - Science IT support team: http://www.sfu.ca/science/faculty-support/IT-Support.html, or, directly at http://support1.mbb.sfu.ca/SelfService/Create.html?Queue=20
  - Cluster support through the Research Computing Group: email research-support@sfu.ca
- Department's Computing page: http://www.stat.sfu.ca/research/computing.html
- Department's getting started with the cluster page: http://stat.sfu.ca/research/computing/getting-started-with-the-cluster.html
- Colony documentation: http://www.rcg.sfu.ca/documentation/ and click on "The Colony HPC Cluster".
- WestGrid getting started page: https://www.westgrid.ca/support/quickstart/new_users