BF528 - Submitting Jobs on SCC

Writing and Executing Scripts

Script: a file that contains a sequence of one or more commands

- Shell Script : **sh script_name.sh**
- Rscript : **Rscript script_name.R**
- Python : **python script_name.py**

Environment configuration and executing jobs

- **Modules** – Used to load applications not automatically loaded by the system, including alternative versions of applications.
  - Check the available modules
    [kkarri@scc4 newdir]$ module avail R
  - Load a module in current environment
    [kkarri@scc4 newdir]$ module load R/3.4.0
  - Unload a module
    [kkarri@scc4 newdir]$ module unload R/3.4.0
- To check the version of a tool or software
  - **[kkarri@scc4 newdir]$ which R**

Running Jobs

Job: unit of computation, e.g. execute a single program

Two types of jobs:

- **Interactive job –** running interactive shell: run GUI applications, code debugging, benchmarking of serial and parallel code performance;
- **Batch job –** job command specified in a script and run on a cluster node with no user interaction

Most of your jobs will be **batch jobs** submitted as **qsub scripts**

Anatomy of a qsub script

Non-Interactive “Batch” Scripts (qsub)

Script Interpreter

Scheduler Directives

Task Commands

```
#!/bin/bash -l

#$ -P rcs
#$ -N test
#$ -j y
#$ -m bae


echo "==============================================="
echo "Starting on : $(date)"
echo "Running on node : $(hostname)"
echo "Current directory : $(pwd)"
echo "Current job ID : $JOB_ID"
echo "Current job name : $JOB_NAME"
echo "==============================================="

module load R
sleep 10

echo "==============================================="
echo "Finished on : $(date)"
echo "=========== ==========================================="
```

Simple qsub script example

```
#!/bin/bash -l

# Set SCC project
#$ -P my_project

# Specify hard time limit for the job.
#   The job will be aborted if it runs longer than this time.
#   The default time is 12 hours
#$ -l h_rt=24:00:00

module load python/3.6.0
python -V
```

More information available on:

http://www.bu.edu/tech/support/research/computing-resources/tech-summary/

Batch Jobs – qsub and qstat

Use the Open Grid Scheduler (OGS) command **qsub** to submit the compiled program to the batch system:

```
[kkarri@scc4 stranded]$ qsub -P bf528 stranded_transcriptome.qsub
```

NB: '`-P <project name>`' is a required argument!

Check the status of your job with **qstat**

```
[kkarri@scc4 stranded]$ qstat -u kkarri
job-ID  prior   name       user         state submit/start at     queue                          slots ja-task-ID
-----------------------------------------------------------------------------------------------------------------
```

Understanding qstat output

```
scc % qstat -u userID
job-ID  prior   name       user         state submit/start at     queue                          slots ja
-----------------------------------------------------------------------------------------------------------
1000001 0.10000 network_gr userID       r     05/19/2014 09:17:53 b@scc-bd7.scc.bu.edu              1
1000002 0.00000 network_gr userID       qw    05/19/2014 09:14:53                                  16
```

Understanding qstat output	
Queue Column	Description
job-ID	Lists the job ID. All jobs have a unique job ID that can be used to reference a particular job.
prior	Priority of the job. This number is set by the SCC batch scheduling system and its value is set based on a "fair share" system to make sure all researches have reasonably equal access to the SCC resources.
name	Name of the job. A user can specify the name of the job using the `-N` option to the **`qsub`** command. Otherwise the system will use the batch script (or binary executable) name as a job's name.
userID	User SCC login ID.
state	The state of the job: (r) – running; (qw) – waiting to run; (hqw) – on hold, waiting to run; (Eqw) – job in error state; (s) – suspended; (t) – transfering.
submit/start	Time when the job was submitted. When the job is running, this field is updated with the time the job started.
queue	The queue name and the node ID on which the job is running.
slots	The number of slots the job requested.
task	The task ID if the job was submitted as a job array. For all other types of jobs this field is empty.

General Directives	
Directive	**Description**
-l h_rt=*hh:mm:ss*	Hard run time limit in *hh:mm:ss* format. The default is 12 hours.
-P *project_name*	Project to which this jobs is to be assigned. This directive is **mandatory** for all users associated with any Med.Campus project.
-N *job_name*	Specifies the job name. The default is the script or command name.
-o *outputfile*	File name for the stdout output of the job.
-e *errfile*	File name for the stderr output of the job.
-j y	Merge the error and output stream files into a single file.
-m *b|e|a|s|n*	Controls when the batch system sends email to you. The possible values are – when the job begins (b), ends (e), is aborted (a), is suspended (s), or never (n) – default.
-M *user_email*	Overwrites the default email address used to send the job report.
-V	All current environment variables should be exported to the batch job.
-v *env=value*	Set the runtime environment variable *env* to *value*.
-hold_jid *job_list*	Setup job dependency list. *job_list* is a comma separated list of job ids and/or job names which must complete before this job can run. See Advanced Batch System Usage for more information.

Delete single or multiple jobs

qdel command and **Job id** can request to delete a job

```
[kkarri@scc4 newdir]$ qdel 3992851
kkarri has deleted job 3992851
```

Delete all of your running and queued jobs

```
[kkarri@scc4 newdir]$ qdel -u kkarri
kkarri has deleted job 3992852
kkarri has deleted job 3992853...
```

SCC OnDemand Demonstration: qsub

For Next Time

Review the workshop on cluster computing and shell scripting:

Workshop 6. Shell Scripts and Cluster Computing

Group assignments announced next class