Shell Scripting and Scheduling Jobs using Cluster Computing Resources

Laura Gutierrez Funderburk, Alex Razoumov

Simon Fraser University

July 2019

compute canada | calcul canada

SFU

SIMON FRASER UNIVERSITY

SFU
SIMON FRASER UNIVERSITY

About Laura

- Mathematics Major, SFU
- DevOps Engineer for iReceptor, SFU

SFU
SIMON FRASER UNIVERSITY

compute canada | calcul canada

About Alex

- PhD in computational astrophysics, UBC
- WestGrid Training & Visualization Coordinator, Compute Canada

SFU
SIMON FRASER UNIVERSITY

compute canada | calcul canada

How did I first learn about scheduling jobs using CC

- Volunteer in Dr. Cedric Chauve's Lab as Bioinformatician from May 2017 - July 2018

- DevOps work with iReceptor in Dr. Felix Breden's Lab September 2018 - Today

SFU
SIMON FRASER UNIVERSITY

compute canada | calcul canada

How did it change my work

- It helped me be more efficient
- Reduced error-rate when performing repetitive tasks
- Automate multiple processes at large scale (1000s to 1,000,000s)

compute canada | calcul canada

SFU
SIMON FRASER UNIVERSITY

Tell me about you

- Name
- What you do
- One challenge you face in your work
- What you hope to learn from this workshop

SFU
SIMON FRASER UNIVERSITY

compute canada | calcul canada

Recap

- Unix Shell: Program whose goal is to provide a user interface which allows users to type commands

- Commands: Sequences of lines of text (entered by a user, or read from a file or data streams). Interpreted by Unix-like operating systems for r/w/x

compute canada | calcul canada

SFU
SIMON FRASER UNIVERSITY

Recap

- Bash is a language for job control in computing
- Bash is a language interpreted by Unix-like operating systems
- A shell script is a computer program designed to be run by the Unix shell

compute canada | calcul canada

SFU
SIMON FRASER UNIVERSITY

Sample Script

A script that prints "Hello world!" to the screen

```
#!/bin/sh
echo "Hello world"
```

compute canada | calcul canada

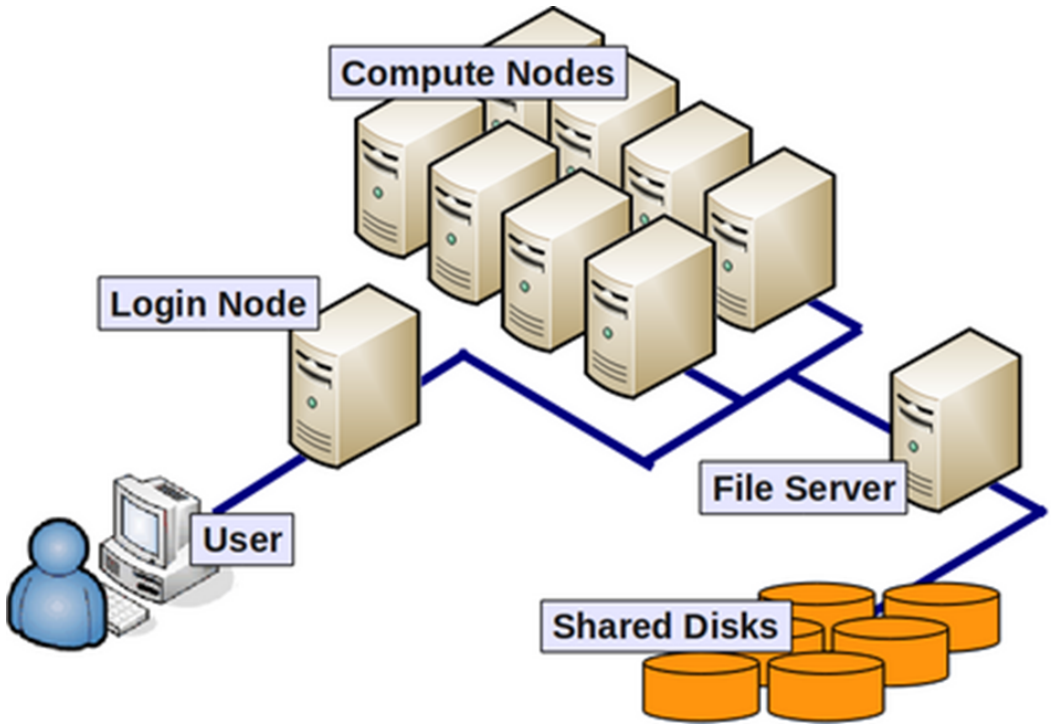
SFU
SIMON FRASER UNIVERSITY

What is a computer cluster?

- Computer cluster: set of loosely or tightly connected computers that work together
- They can be viewed as a single system
- Computer clusters have each node set to perform the same task, controlled and scheduled by software

compute canada | calcul canada

SFU
SIMON FRASER UNIVERSITY



SFU
SIMON FRASER UNIVERSITY

compute canada | calcul canada

How do we connect to a computer cluster

- Secure shelling (SSH): remotely logging in to computers running the Linux operating system
- Need to have a SSH client program installed on your machine
- macOS and Linux: command-line SSH client pre-installed. Windows: various graphical SSH clients you can use (MobaXTerm or Putty)

Key components when scheduling a job

File : simple_job.sh

```
#!/bin/bash
#SBATCH --time=00:01:00
#SBATCH --account=def-someuser
echo 'Hello, world!'
sleep 30
```

Key components when scheduling a job

Submitting our job

```
$ sbatch --time=00:30:00 simple_job.sh
```

Key components when scheduling a job

There are many kinds of jobs which are adequate depending on your script and what you want to do
To learn more, visit
`https://docs.computecanada.ca/wiki/Running_jobs`
If you are a Compute Canada user, contact support@computecanada.ca (or ask Alex at the end of this workshop)

Kinds of jobs we will explore in this workshop

Serial and array job

File : array_job.sh

```
#!/bin/bash
#SBATCH --account=def-someuser
#SBATCH --time=0-0:5
#SBATCH --array=1-10
./myapplication $SLURM_ARRAY_TASK_ID
```

What we are going to do today

- Go over a shell script prepared for today
- Secure shell into a virtual machine that allows us to submit jobs
- Modify a script to perform several tasks
- Submit a job via the sbatch command on a virtual machine
- Go over output results

compute canada | calcul canada

SFU
SIMON FRASER UNIVERSITY

Time to practice

(UBUNTU) Press CTRL + ALT + T (or enter Terminal in search bar)
(MacOS) Search for Terminal in Applications (or Spotlight)
(Windows) If you have MobaXterm or Putty installed, open it, otherwise visit `https://sfu.syzygy.ca/`

compute canada | calcul canada

SFU
SIMON FRASER UNIVERSITY

Final Remarks

- We did a recap of what Bash and shell scripting are
- We learned basic components necessary when writing and scheduling a job on a computer cluster
- We modified a shell script and submitted a job using a virtual machine

compute canada | calcul canada

SFU
SIMON FRASER UNIVERSITY