

C3G Analysis Workshop: RNA-Seq

Concepts from *personal* to *super computer*

January 22-23, 2019

Canadian Centre for
Computational
Genomics

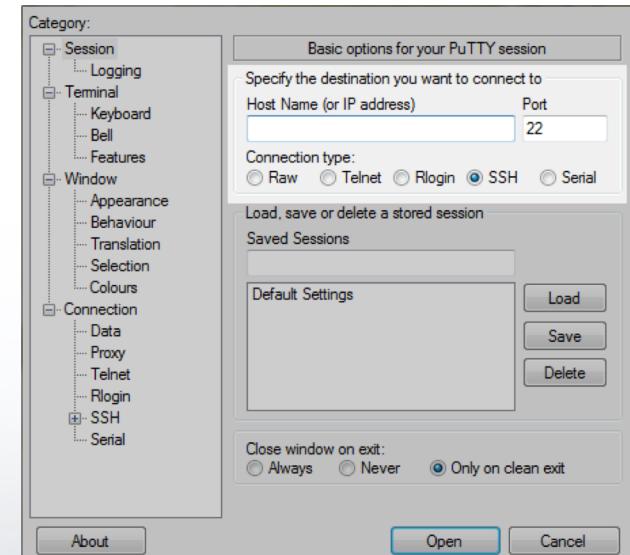


Test your account

Open a bash terminal and type:

ssh my_cc_account@mp2.ccs.usherbrooke.ca

On older windows system, use putty:



Overview



Calcul Québec, Calcul/Compute Canada

A simple computer

- User Interface
- A system that counts
- A system that stores

Supercomputers

- Pipelines
- Genpipes

Pause

Exercises



compute
canada | calcul
canada

https://docs.computecanada.ca/wiki/Compute_Canada_Documentation

Each researcher in
Canada has a 10 TB
space and 50 Core-
years on all sites

Arbutus/UVic



Cedar/SFU



Compute • Calcul
Ontario

Graham/Waterloo



Calcul

Ontario



Calcul Québec

Beluga/ETS-Mcgill

Niagara/Uoft

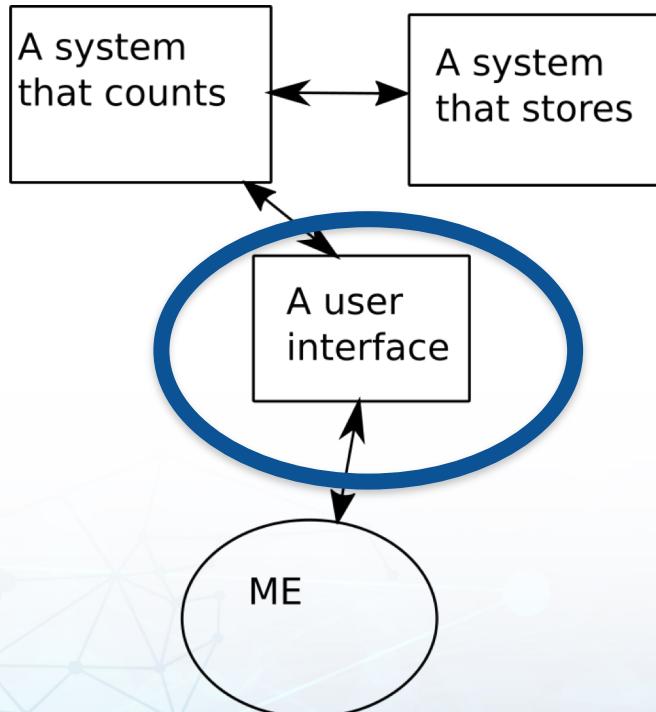


- CCF national data centres support researchers across Canada





What is a computer?



Here is a rather
simplified version of
one



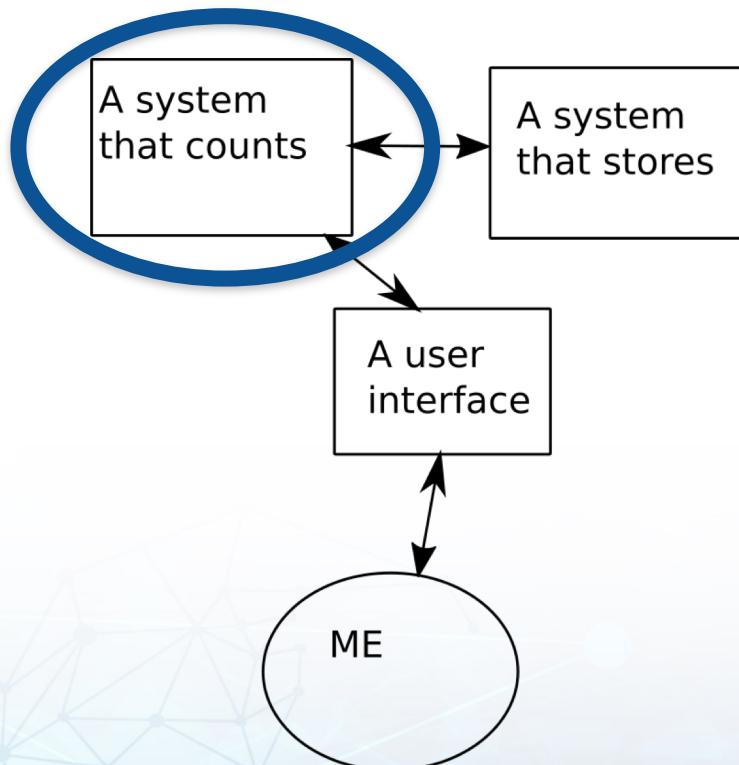
A User Interface (UI)

GUI

```
Farm
Billio
Ease
Tarif
farm@moins:~$ pwd
/Billio/home/poq
farm@moins:~$ ls
cqvfschecksum.soft.muglc  Downloads      Music    Public    snap      Templates   truite   truite.zip
cqvfschecksum.cvmfs-config.computecanada.ca  Desktop      install_test  old_repo  PycharmProjects  Snapshots  test
core                           Documents      java_error_in_PYCHARM.hprof  Pictures   R           sop-vlc.sh  truite.png
farm@moins:~$ ls -l
total 451896
drwxr-xr-x  5 poq poq   4096 Jul 19 11:48 c3g
drwxr-xr-x 10 poq poq   4096 Jun 12 16:23 container_cvmfs
-rw-----  1 poq poq 257409924 Apr  3 14:35 core
drwxr-xr-x  4 poq poq   4096 Jun 26 13:47 cq
drwxr-xr-x  1 poq poq  2574 Apr 10 13:34 cq_et_cie.kdbx
drwxr-xr-x  1 poq poq   4096 Jun 12 12:38 cvmfscchecksum.cvmfs-config.computecanada.ca
drwxr-xr-x  2 poq poq   4096 Jul 19 11:06 Desktop
drwxr-xr-x  6 poq poq   4096 Jul 24 13:25 Documents
drwxr-xr-x  5 poq poq   4096 Jul 24 14:06 Downloads
drwxr-xr-x  6 poq poq   4096 Mar 29 12:38 install_test
-rw-r--r--  1 poq poq 265169749 Jun 12 15:32 java_error_in_PYCHARM.hprof
drwxr-xr-x  2 poq poq   4096 Mar 27 13:54 Music
drwxr-xr-x  6 poq poq   4096 Apr 10 13:44 old_repo
drwxr-xr-x  2 poq poq   4096 Jul 24 13:32 Pictures
drwxr-xr-x  2 poq poq   4096 Jul 27 13:49 Public
drwxr-xr-x  1 poq poq   4096 Jul 12 11:44 pycharmProjects
drwxr-xr-x  3 poq poq   4096 Jul  3 16:49 R
drwxr-xr-x  3 poq poq   4096 May 24 10:35 snap
drwxr-xr-x  2 poq poq   4096 May 24 10:39 Snapshots
-rw-r--r--  1 poq poq 181 Jul 10 10:21 sop-vlc.sh
drwxr-xr-x  2 poq poq   4096 Mar 27 13:54 Templates
drwxr-xr-x  7 poq poq   4096 Jun 27 13:50 test
-rw-r--r--  1 poq poq 16188 Apr 25 11:47 titi
drwxr-xr-x  2 poq poq   4096 Jun  6 10:49 truite
-rw-r--r--  1 poq poq 29875 Jul 12 18:25 truite.png
-rw-r--r--  1 poq poq 1403 Jun  1 13:13 truite.txt
-rw-r--r--  3 poq poq   4096 Apr  3 21:15 truite.zip
drwxrwxr-x  5 poq poq   4096 May 24 09:55 venv
drwxr-xr-x  2 poq poq   4096 Mar 27 13:54 Videos
farm@moins:~$
```

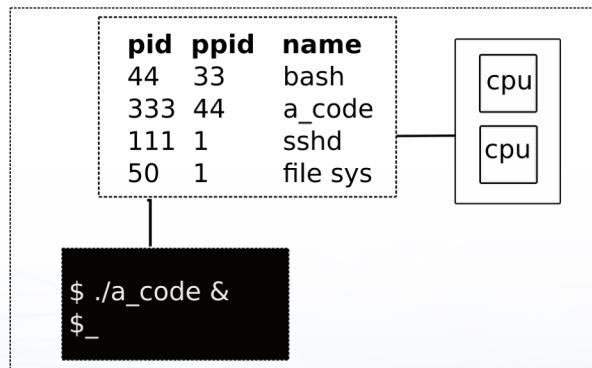
UI

Here it is again



A system that counts

- When you type commands, you will type them in the terminal, or for that matter every time you click on a new icon on your computer, a script/binary/software is *executed*
- Once executed, it *transforms* into a process

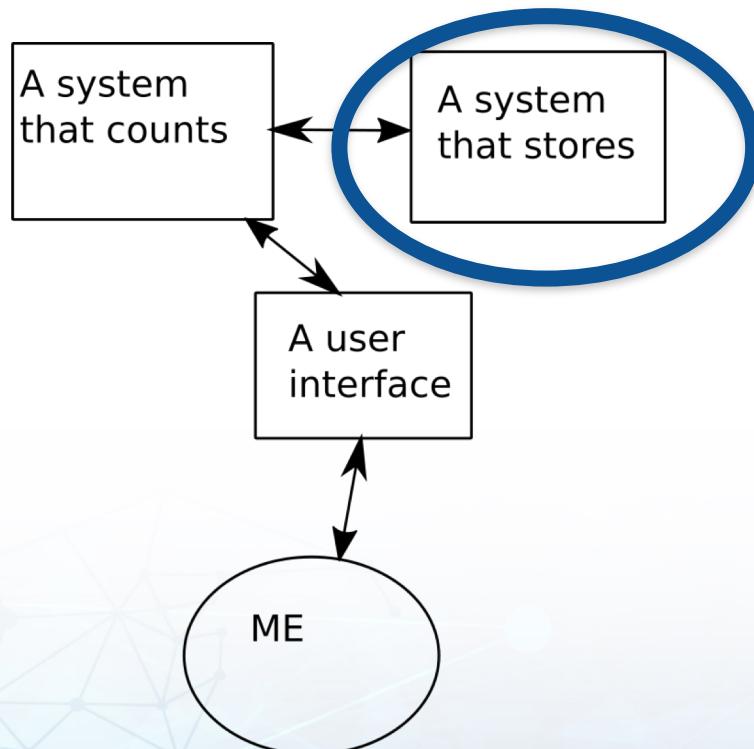


- This process has ways to talk to the external world: streams. There are three "standard" streams called `stdin`, `stdout` and `stderr`.

One more time



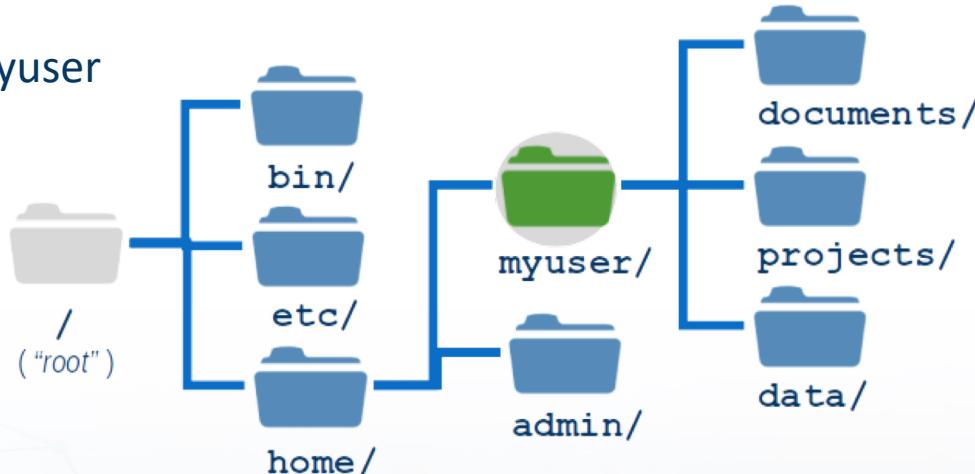
Canadian Centre
for
Computational
Genomics



A system that stores

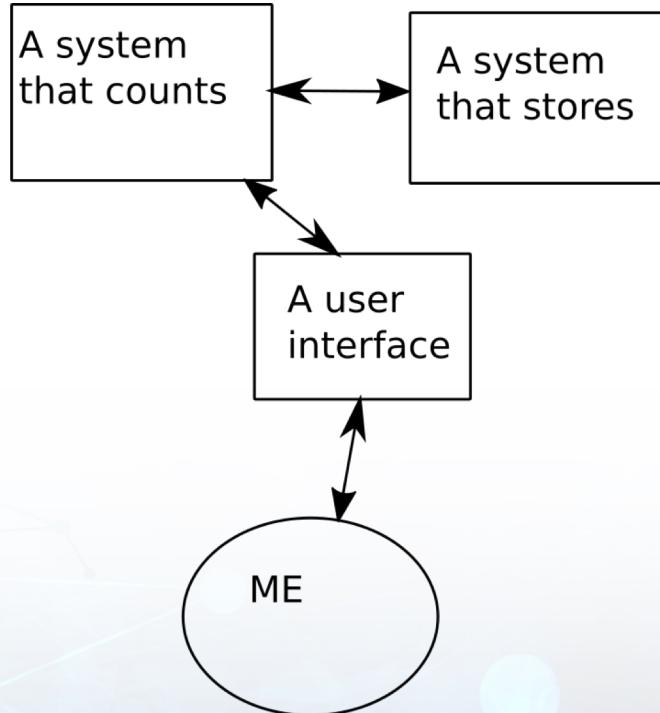
Some definitions

- Root -> /
- \$HOME -> /home/myuser
- Full path
- Working directory
- Relative path



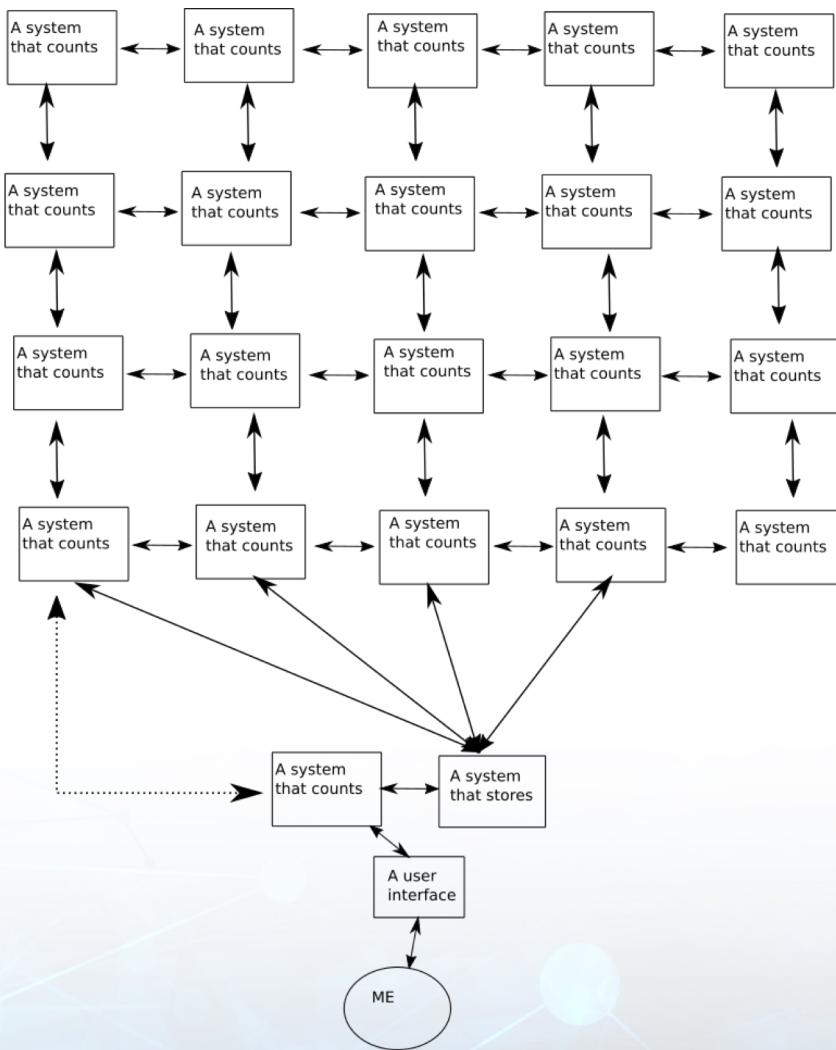
What is a supercomputer?

Simple computer



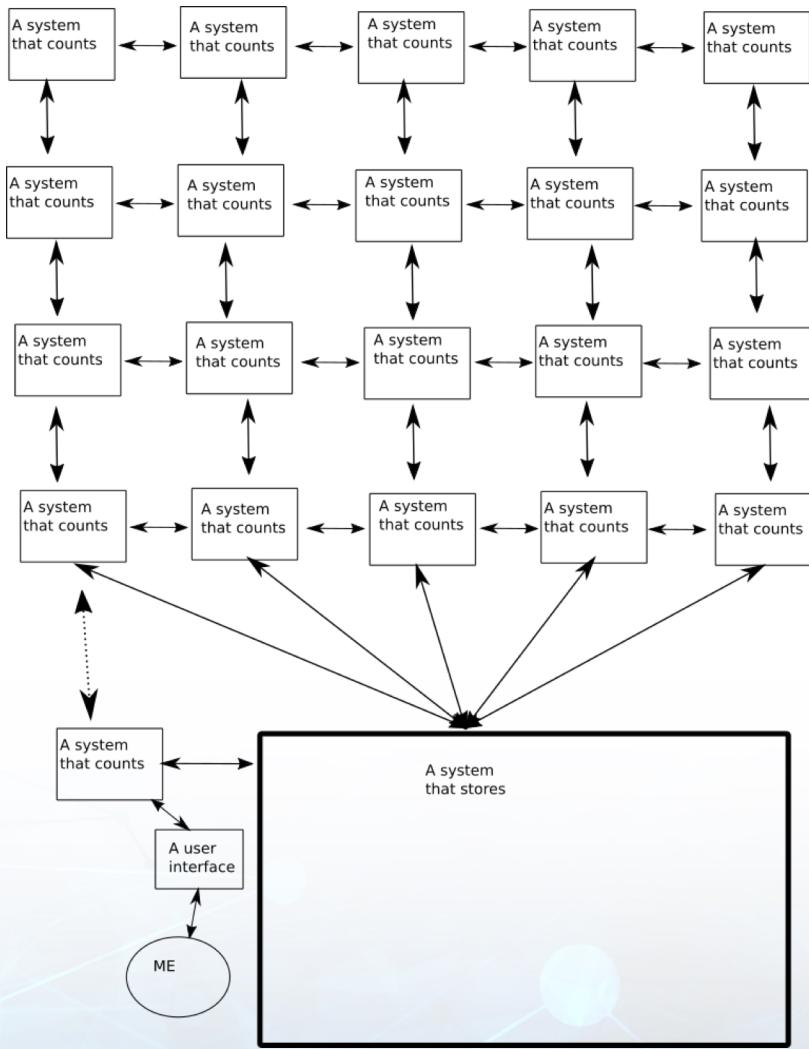


Supercomputer for Data science





Supercomputer for Data science

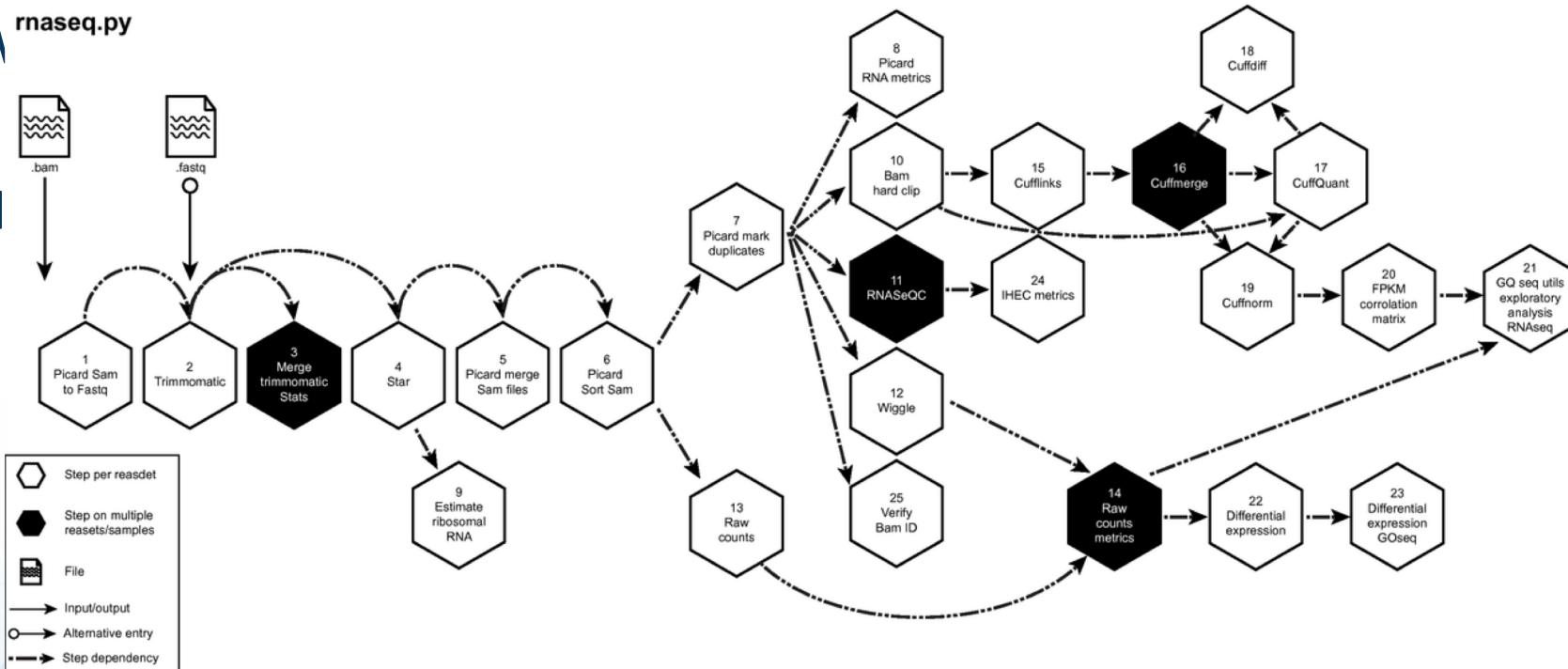




Canadian Centre
for
Computational
Genomics

What can we do with an HPC

rnaseq.py



How to make the HPC work?

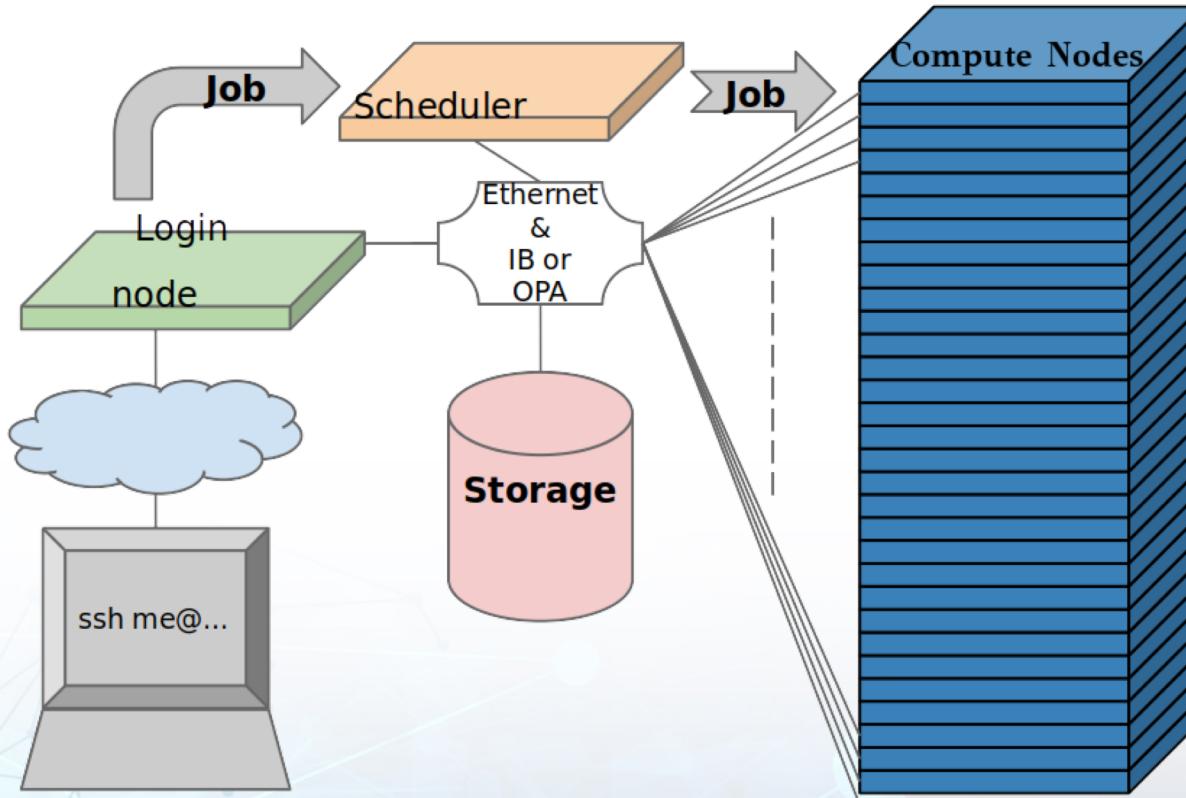


Canadian Centre
for
Computational
Genomics

For many tasks, you **don't** have to think that a supercomputer is super... In fact, **GenPipes** is taking care of that part for you!

- You run **Genpipes** on a single (simple) computer and it creates a script that will control the supercomputer for you
- You just ask the supercomputer to execute it and then look at your data being generated.

More HPC details

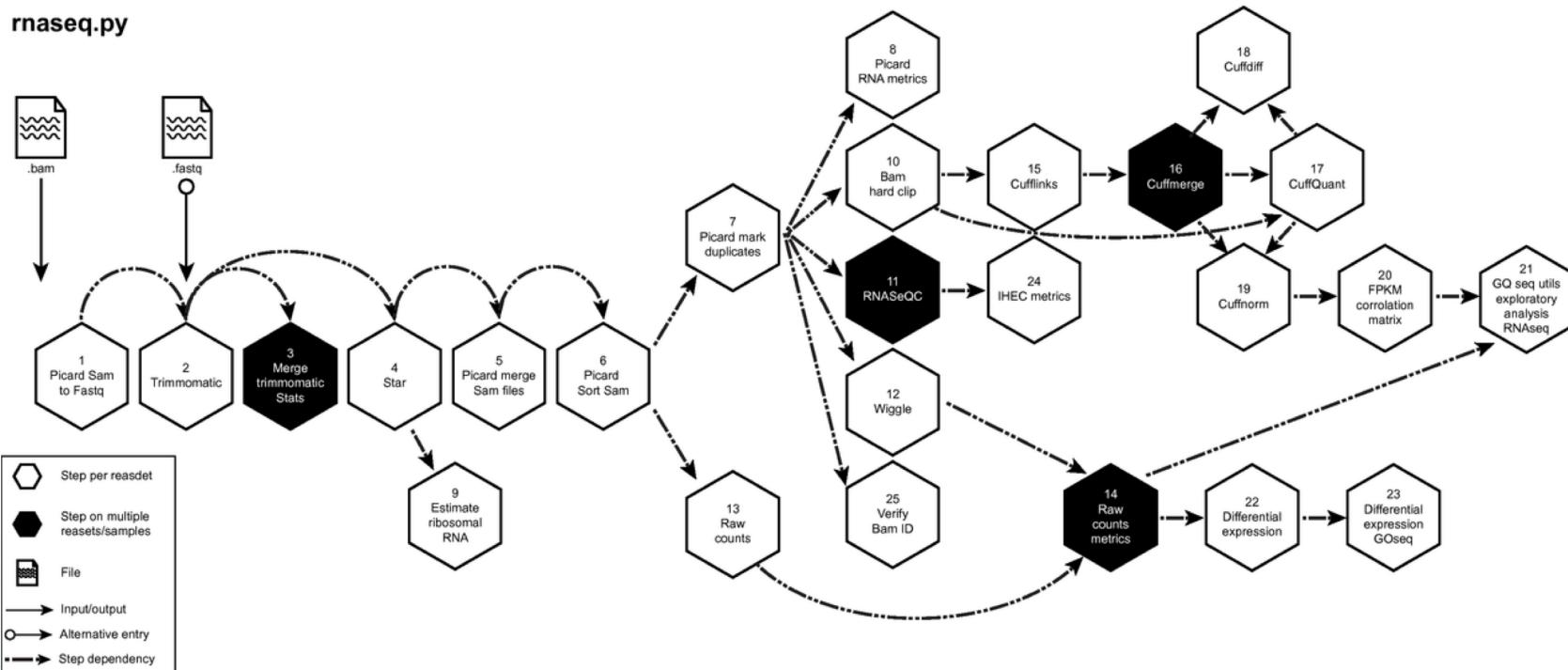


What is this workshop about?



Canadian Centre for
Computational
Genomics

`rnaseq.py`





Canadian Centre
for
Computational
Genomics



Thank you!



compute | calcul
canada | canada

