

McGill

Leveraging Canadian Bioinformatics with Galaxy Vz in a HPC center

David Anderson de Lima Morais^{1,2}, Carol Gauthier^{1,2}, Michel Barrette¹, David Bujold², Maxime Caron², Alain Veilleux¹, Guillaume bourque¹

C a N A D A

Centre de Calcul Scientifique, Université de Sherbrooke, Quebec, Canada
 McGill University and Genome Quebec Innovation Center, Montreal, Canada



OVERVIEW

- Presently, Canada possesses 13 PhDs, 17 Masters and more than 20 Bachelor/certificate bioinformatics
 programs forming hundreds of bioinformaticians¹.
- Genome Canada is a non-profit organization responsible for managing large-scale research and is divided into 6 genome centers across the country².
- Compute Canada is the Canadian National HPC platform that integrates data centers, resources and tools for the Canadian academic research community³.
- Mammouth supercomputer⁴, provides a Galaxy environment for the Canadian scientific community. Our hybrid approach (cloud/HPC) consists of deploying Galaxy on a virtual machine (hosted on an interactive node) in a way that allows for the launching of jobs on Mammouth's computing nodes and using simple API connectors and file system mounts.



GALAXY ENVIRONMENT

- Galaxy is installed in a Vz (G) inside Mammouth's interactive node (F).
- Users with a Compute Canada's account (A) have access to Galaxy (G) and their data (H).
- The interactive server (F) will also provide Vz containing private and public versions of the UCSC browser.
- Pipeline jobs launched by Galaxy (Galaxy user) are intercepted by the Galaxy-Mammouth's API and launched directly on the compute node (E) as the logged user.
- The public and private data as well as all softwares used by Galaxy are stored on the interactive node and NFS mounted in the Vzs.
- The whole interactive node + Vzs environment can be easily replicated to any other HPC center in Canada (I).



A

Compute Canada

Database (CCDB)

Users, projects,

allocations, stats,

resources, etc...

Genomic Services Portal

• Single sign-on.

B

- Credentials fetched from the Compute Canada Database (CCDB).
- Same login on HPC resources, Galaxy, UCSC browsers and other tools.
 Map users to the Genome Quebec Innovation Center Database (Nanuq)
- Create or use mapped projects from Compute Canada Database (CCDB).
- Control access to your private data and tools (sharing with other users).
- Control the location of data across Data Centers.
- Navigate private and public metadata (Samples/Assay grid, etc..)
- On demand genomic Virtual Server (Vz or Vm)
 - Choose data sets to access.
- Manage data location if needed.
- Select tools and services wanted (Galaxy, UCSC browser, etc.)
- Start virtual server and access the service.
- Save and reuse virtual server images on demand.

• Access resources usage stats and other Compute Canada information.



ACKNOWLEDGMENTS

Marc-Étienne Rousseau, System Architect, CBRAIN Project. Jean-François Landry, Universite de Sherbrooke. Universite de Sherbrooke. McGill University and Genome Center. Compute Canada. Genome Quebec.

REFERENCES

- 1- http://bioinformatics.ca/
- 2- http://www.genomecanada.ca/
- 3- http://computecanada.ca/
- 4- http://computecanada.ca/index.php/en/resources/compute-resources-from-ccdb-data
- 5- http://cbrain.mcgill.ca/