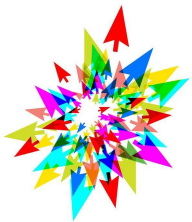
Introduction to High-Performance Computing

July 28, 2021 Darcy Quesnel



compute canada | calcul canada

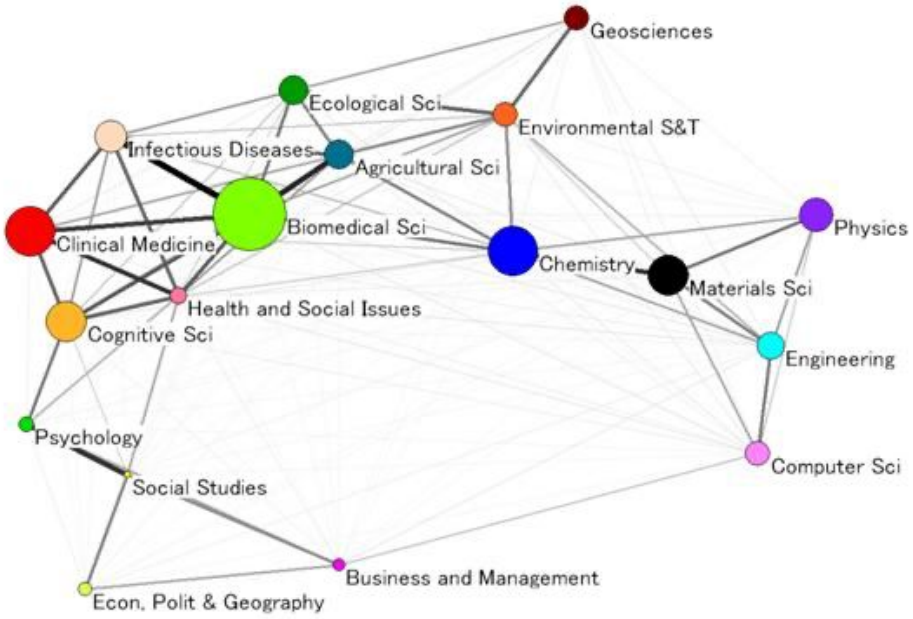
Calcul Québec

Based on slides from Pier-Luc St-Onge of Calcul Quebec

Calcul Quebec Partners and Sponsors

Research Domains

- Genomics
- Molecular Dynamics
- Finite Element Analysis
- Computational Fluid Dynamics
- Astronomy and Astrophysics
- Geospatial Data Analysis
- Visualization and Image Analysis
- Artificial Intelligence / Deep Learning
- Humanities and Social Sciences



Advanced Research Computing

- *High-Performance Computing (HPC)*, sometimes referred to as *Advanced Research Computing (ARC)*
- Large parallel- and cluster-computing resources
- Research data management
- Cloud computing
- Project web portals

Comparison

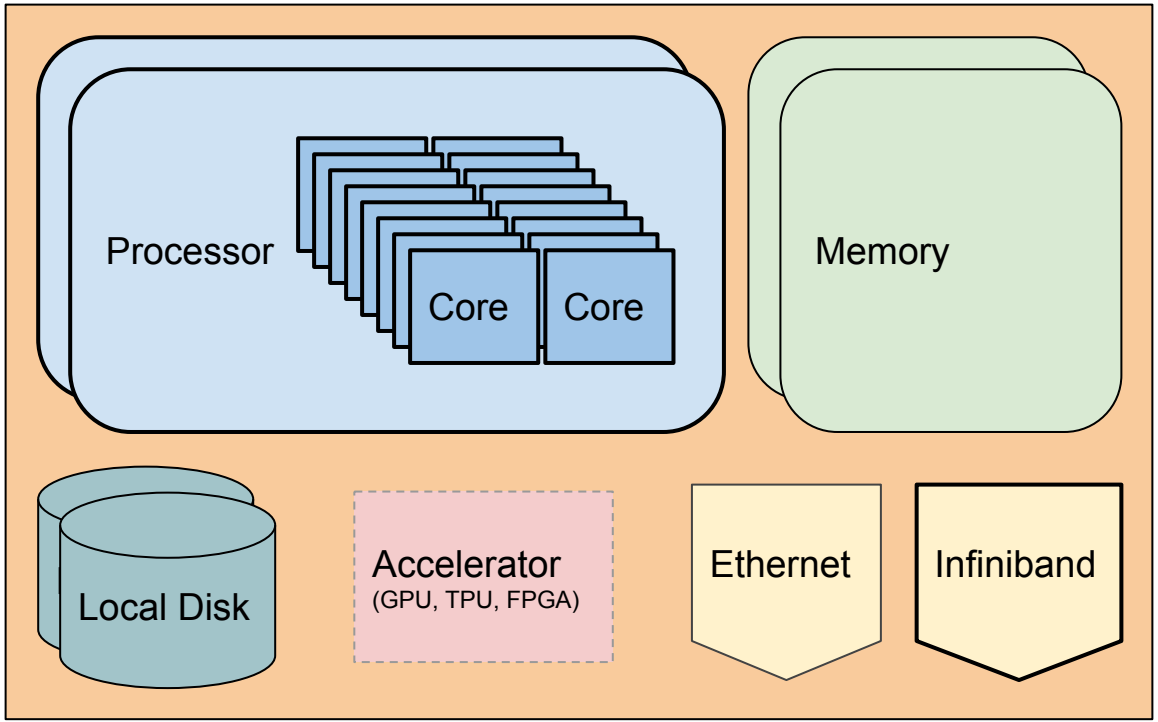
Your Laptop or Desktop

- 2 to 4 cores per processor
- 4 to 32 GB RAM per machine
- 128 GB to 4 TB storage per machine
- Local storage / External USB storage
- Wifi and/or Ethernet for networking
- Gaming GPU
- Direct access
- General usage
  - *Easy to develop and test applications and pipelines on*

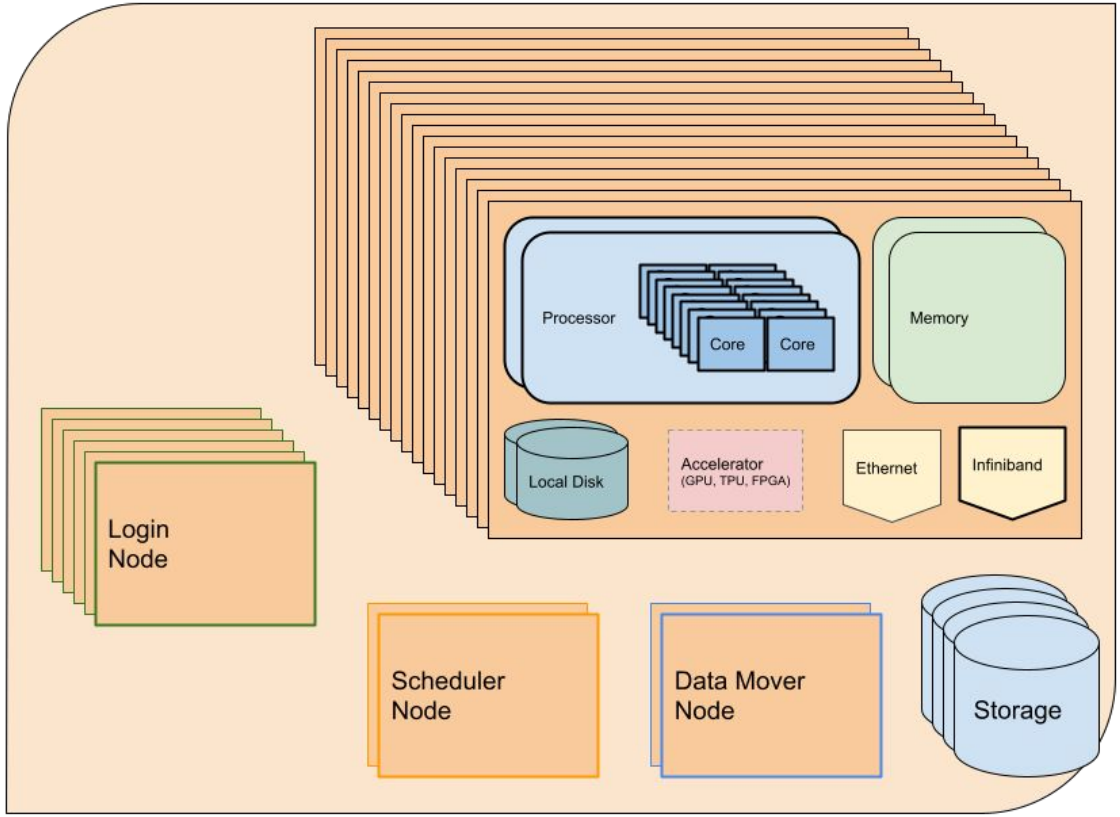
HPC Compute Node

- 16 to 64 cores per processor
- 32 to 3072 GB per node
- 150 TB to 16 PB storage per cluster
- Networked cluster storage
- Ethernet and fast network interfaces
- Accelerators (GPU, ...)
- Access by a submission queue
- Multi-node unattended tasks
  - *Possible to run applications and pipelines with a large amount of data and long runtimes*

Zoom in to an HPC Compute Node

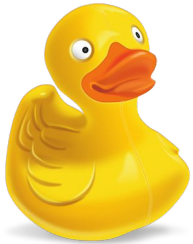


Zoom out on the HPC Cluster



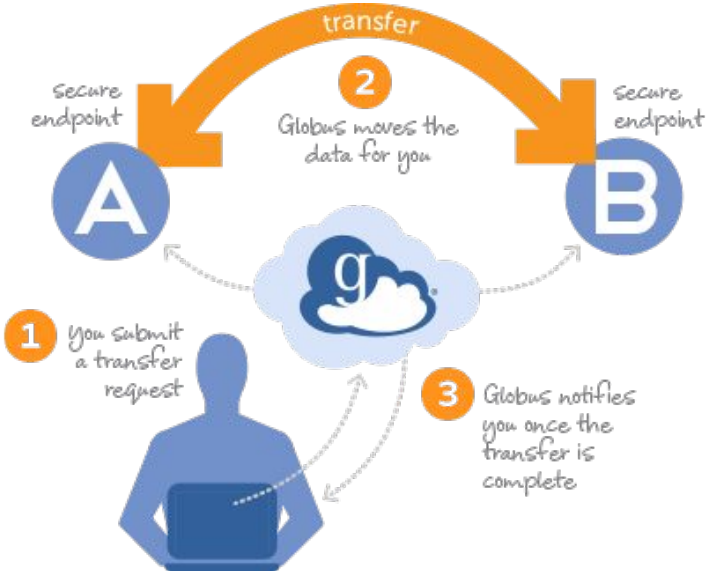
Software Environment on an HPC Cluster

- The operating system is usually Linux-based
  - CentOS, Scientific Linux, Ubuntu, ...
- To connect to a cluster, we use a secure remote shell
  - `ssh`, MobaXterm, ...
- To transfer data securely, we use secure remote copy
  - `scp`, Cyberduck, ...
  - Other tools for large amounts of data, such as Globus
- To develop code, we can use command-line tools
  - Command-line development tools
  - Shell-based text editors
  - Scientific software and libraries
  - Version control software
- Web portals may provide access to domain-specific web tools



Transferring Data

- Use command-line tools to transfer data to or from a cluster
  - `scp`, `rsync`, ...
- Use GUI programs to transfer data if it's more comfortable
  - Cyberduck, WinSCP, ...
- Use Globus for large, long-running, unattended data transfers
  - Web interface
  - Third-party transfers
  - Parallel streams
  - Long-running transfer jobs managed by the service



Scientific Software Modules

- Use a module system to add command-line software packages and libraries to your environment
  - Compilers
    - GCC, LLVM, ...
  - Interpreters
    - Python, R, ...
  - Libraries
    - OpenMPI, Boost, OpenBLAS, ...
  - Other tools
    - FreeSurfer, FSL, ...
  - These are all tuned for the node architectures
  - It's possible to ask for other software packages to be installed and made available

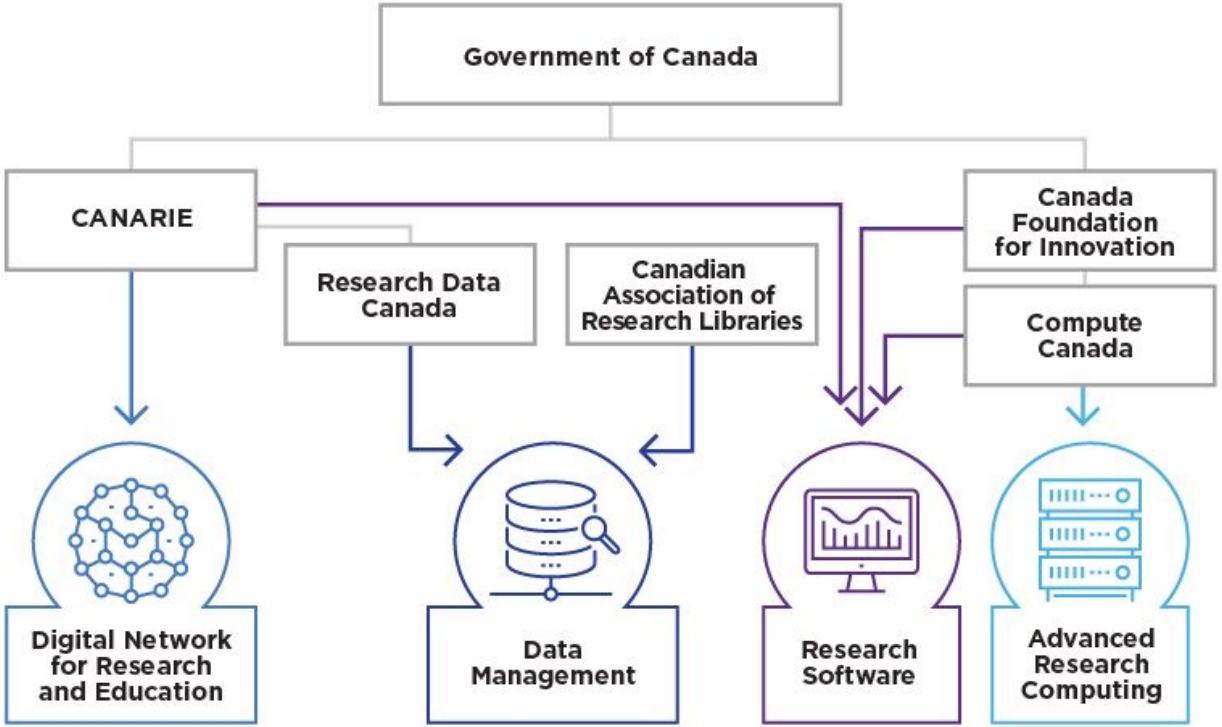
Job Scheduling System

- Sometimes called *batch* systems
- Large, long-running jobs **should not run on a *login node*,** but on the *compute nodes*
- Submit your job script to the scheduler and specify the resources you need
  - Number of processors
  - Amount of memory
  - Presence of specific accelerators
  - Job duration
  - Account name
  - Script to run on compute nodes
  - Where you want your results
- The scheduler uses a *fair-share* policy

Cloud Systems

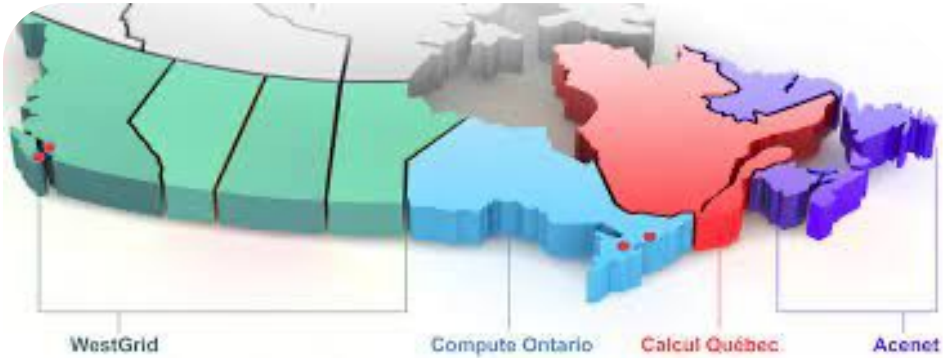
- Infrastructure as a service
  - Allocate virtual machines, networks, and storage resources on the fly
  - VMs are blank slates to be configured and managed by the user
  - VMs have the software and libraries that the user installs
  - For example, a user can configure a web server for a project-oriented portal

Compute Canada



Compute Canada and Calcul Québec

- Compute Canada is a consortium of regional partners
  - ACENET, Calcul Québec, Compute Ontario, and WestGrid
  - About 200 analysts and administrators across many Canadian universities



Calcul Quebec

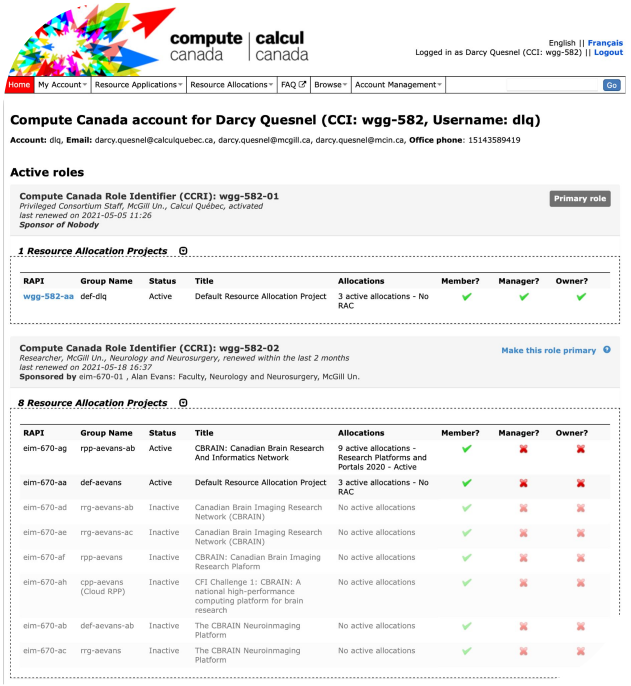
- Calcul Quebec
  - About 50 people
    - HPC/ARC analysts
      - User support, workshops, ...
    - System administrators
    - Scientific and operational directors
    - Spread across *Laval*, *Sherbrooke*, *McGill*, *UdeM* and *ÉTS*
  - HPC Resources
    - Béluga HPC is housed at *ÉTS*

Compute Canada National Resources

Name / Institution	**Total Cores**	**Total GPUs**	**Nodes**
Niagara / UofT	61 920	-	40 cores/node, 188 GB/node
Cedar / SFU	58 416	584	32 cores/node, 125 GB/node
Graham / UWaterloo	36 160	320	32 cores/node, 125 GB/node
Béluga / McGill	34 880	688	40 cores/node, 186 GB/node

Accessing Compute Canada Resources

- All user registration and management is done through CCDB
  - A professor can register for an account
    - Create an account with CCDB
    - Request a role of principal investigator
  - A student, postdoc, or research staff can register for an account
    - Create an account with CCDB
    - Request a sponsored role with the registered principal investigator
    - Once the account is active, connect to any Compute Canada national system



Online Resources

- computecanada.ca
- calculquebec.ca
- docs.computecanada.ca
- calculquebec.eventbrite.com
- support@computecanada.ca