Advanced research computing at Western and beyond

Resources, support and training courses

September 26, 2024, Noon - 1:00pm

SHARCNET™
Shared Hierarchical Academic Research Computing Network

Academic/research computing at Western

Campus IT groups - academic computing

- Science: https://sts.uwo.ca/
- Social science: https://ssts.uwo.ca/
- Engineering: https://www.eng.uwo.ca/itg/
- More https://wts.uwo.ca/

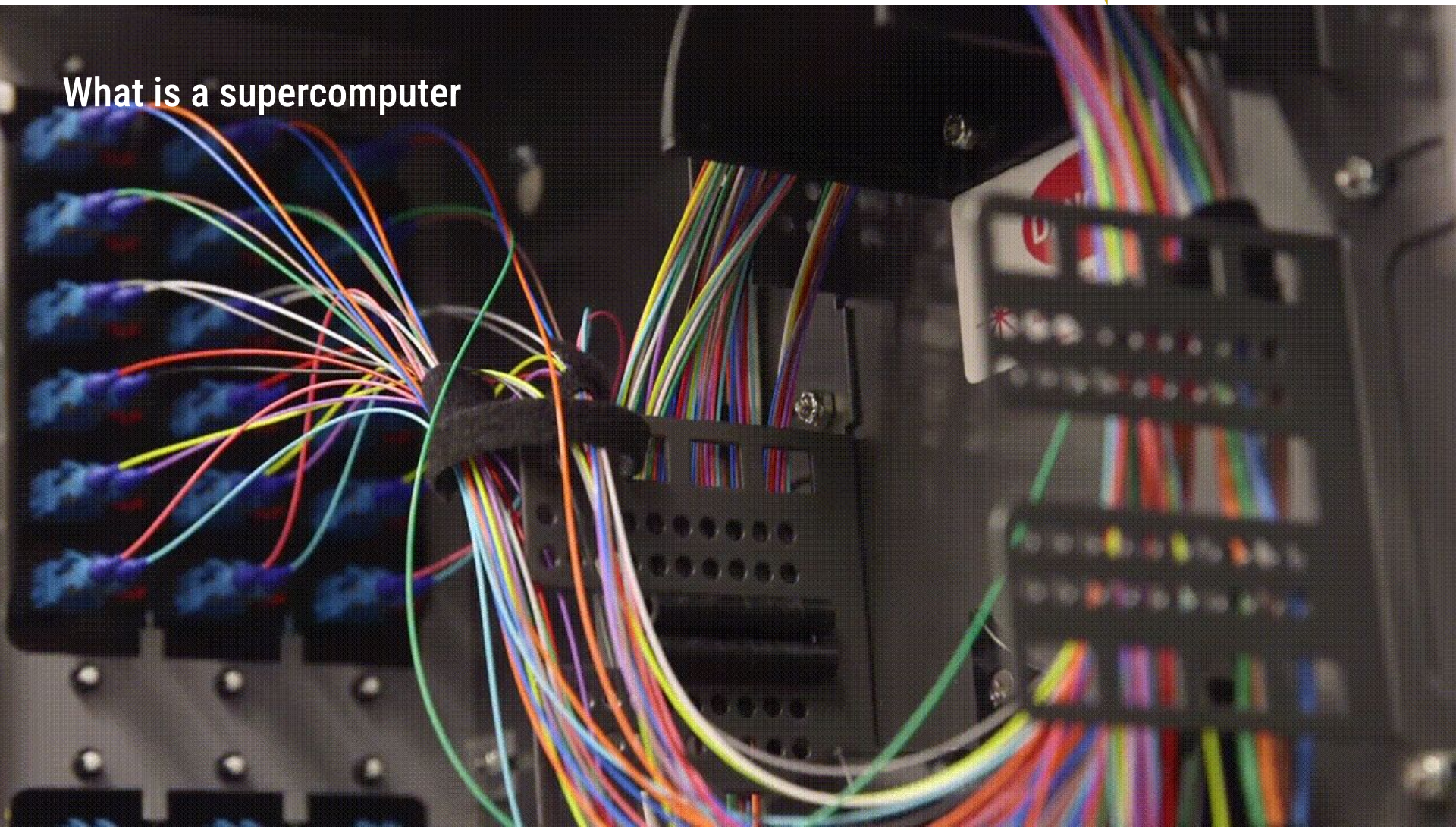
SHARCNET - HPC

- Advanced research computing (ARC)
- Research data management (RDM, e.g. for librarians)
- Training on ARC, data science, machine learning
- More https://www.sharcnet.ca/

Supercomputing at Western, SHARCNET and beyond

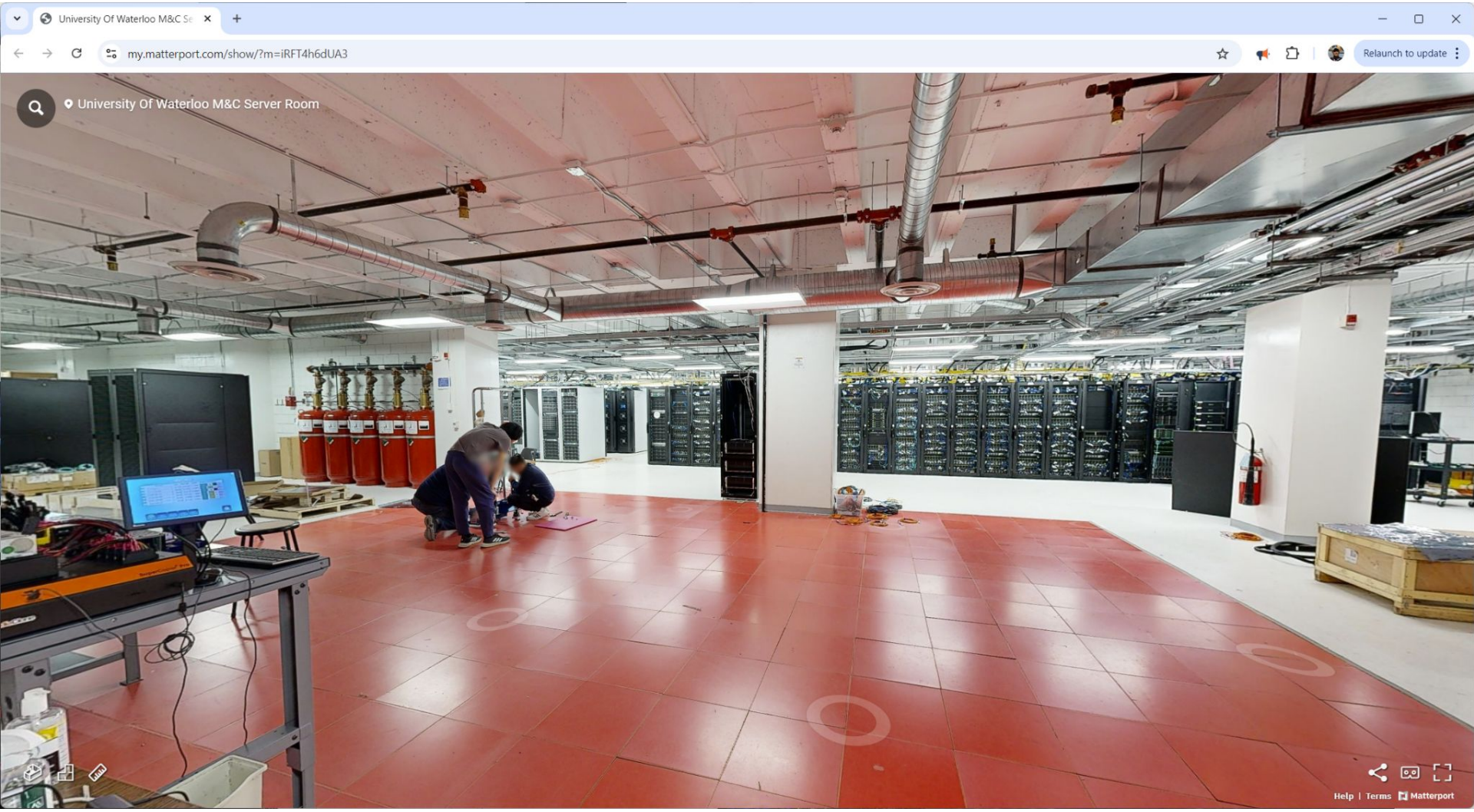
- Supercomputing at Western, SHARCNET and beyond
  - What are supercomputers and what are available (updates)
  - Why do you need supercomputers
  - Who are using them
  - How to access supercomputers (clusters) and clouds
  - Accessing and managing files
  - Running programmes - *submitting jobs*
  - Where to get help
- PIs applying for compute, storage and cloud resources
- What every graduate student should know
- Introduction to advanced research computing courses
- Q & A

What is a supercomputer



A tour from here

Supercomputing: *What and where are they*

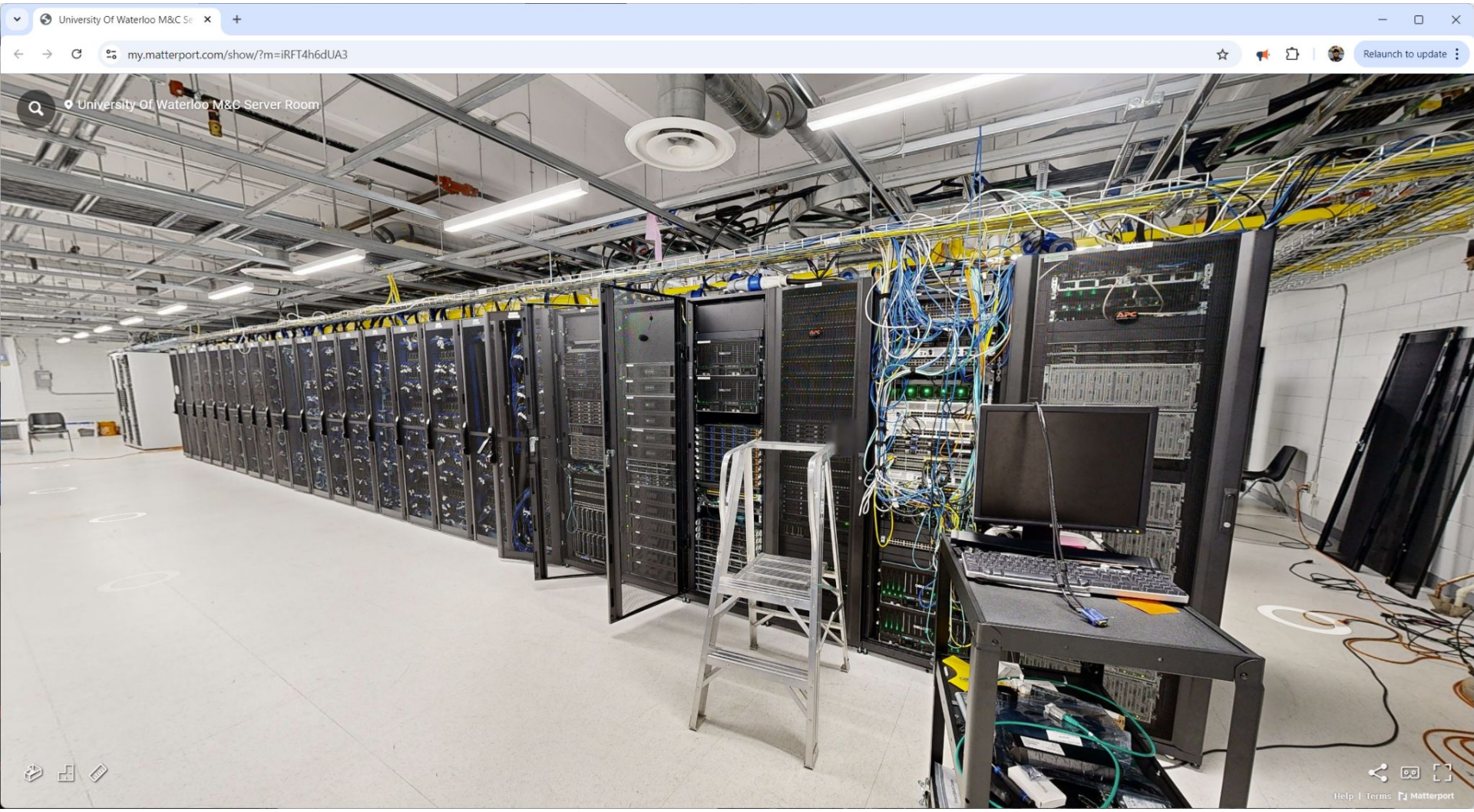


Western UNIVERSITY · CANADA

SHARCNET™

Digital Research Alliance of Canada

Supercomputing: *What and where are they*



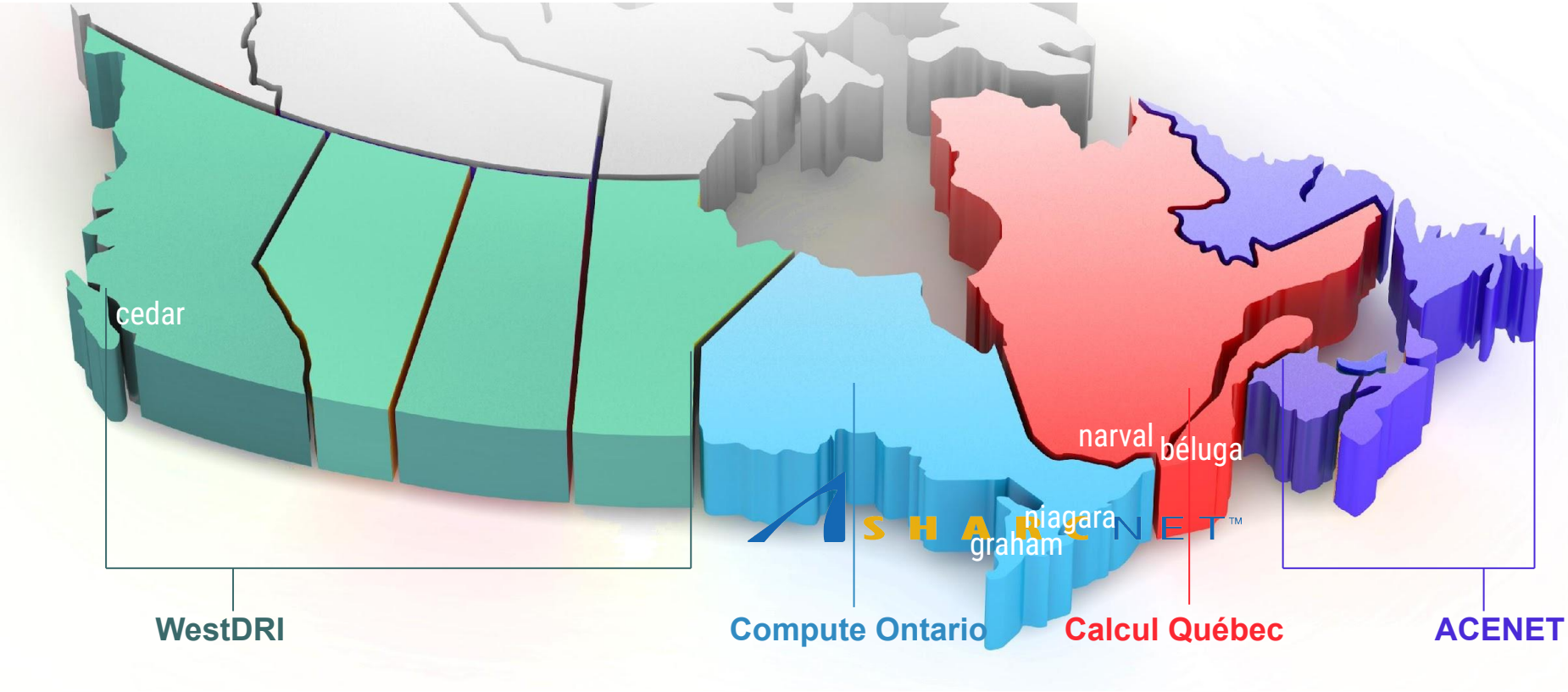
Western UNIVERSITY · CANADA

SHARCNET™

Digital Research Alliance of Canada

Digital Research Alliance of Canada

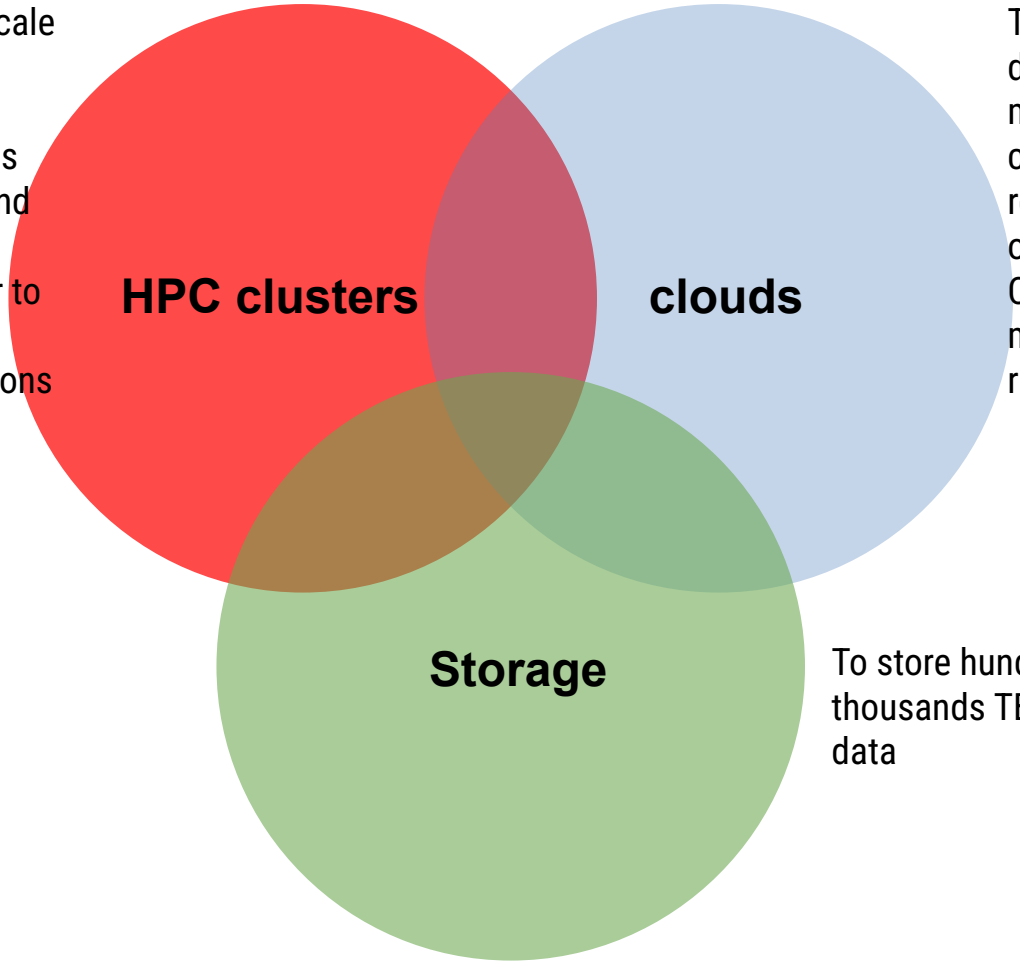
Supercomputing: *What and where are they*



One can access all national supercomputers across the country, for free.

Supercomputing: *What and where are they*

To run large and exascale simulations that need dedicated access to hundreds of thousands CPU and GPU cores and low latency, fast interconnect fabric; or to run hundreds of thousands of simulations simultaneously



To run web services, databases or use virtual machines (VMs) with full control. The CPU and GPU resources are typically over committed, that is, a CPU core or GPU device might be shared by the running VMs.

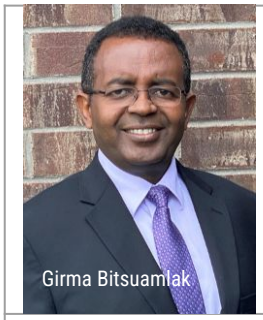
To store hundreds of thousands TB of research data

Supercomputing: *What and where are they*

Why use supercomputing resources

- You do not have many cores or much memory;
- You do not have the GPUs of the type needed;
- You need a huge amount of disk space, e.g. hundreds of TB;
- You need to run large scale of simulations that need hundreds of cores;
- You need to run large amount of simulations concurrently instead of one after another;
- You need to run a web services;
- You need to run a SQL database;
- You need to run programs on a cloud;
- All these services are free

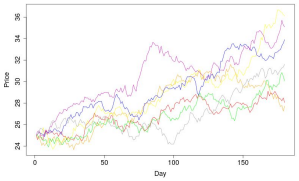
Supercomputing: *Diverse user groups*



Civil Engineering and WindEEE
Experimental and computational wind engineering.

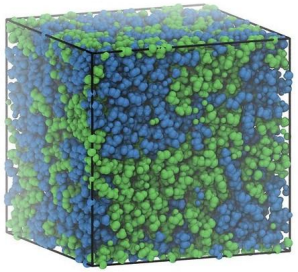
https://windeee.ca/

Economics
Computational finance, with focus on stock pricing.

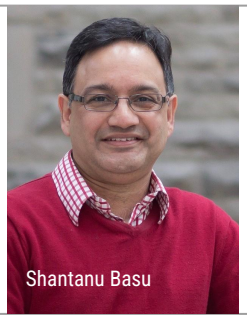


https://publish.uwo.ca/~lstentof/

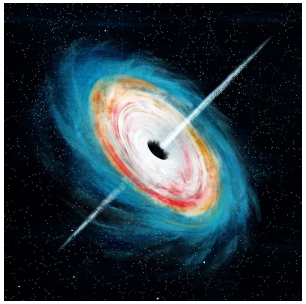
Physics and Chemistry
Computational physics and chemistry, with applications in biological soft matter.



https://softsimu.net/mikko/



Astrophysics
Large scale simulation of the formation of the universe and supermassive black holes in the early universe.



https://physics.uwo.ca/~basu/

Brain and Mind Institute
Medical imaging, computational brain research.

https://www.uwo.ca/bmi/investigators/ali-khan.html

Western UNIVERSITY · CANADA

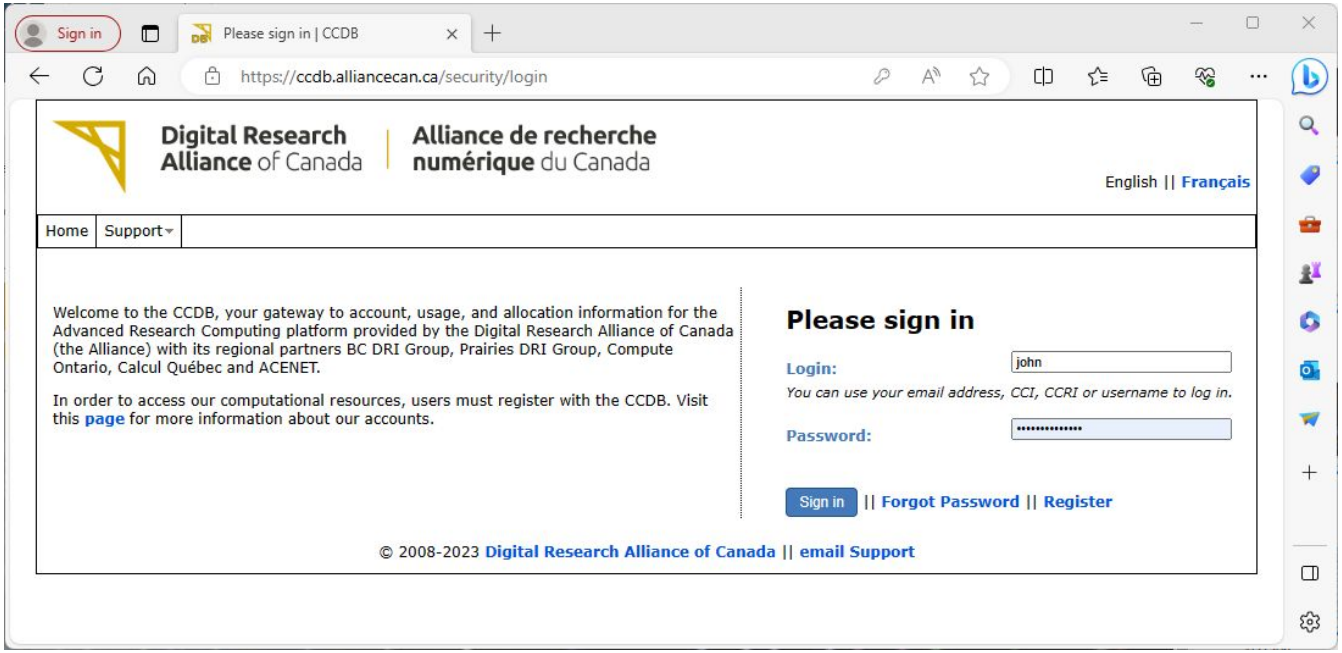
SHARCNET™

Digital Research Alliance of Canada

Supercomputing: *Getting an account*

Sign up for an account for FREE at

https://ccdb.alliancecan.ca/



Supercomputing: *Getting an account*

Some conditions that apply

- Your supervisor should have an account.
- Students, postdoc, visiting scholars and other research staff can sign up for an account with supervisor's role ID (CCRI, e.g. abc-123-02)
- This account allows you to access all the supercomputers and clouds across the country.
- *Multi-factor authentication (MFA) is used to login to any of the systems as you do at Western.*

Supercomputing: *Resources*

Clusters across the country

- cedar.alliancecan.ca
- graham.alliancecan.ca
- niagara.alliancecan.ca
- beluga.alliancecan.ca
- narval.alliancecan.ca

Cloud services

- arbutus.cloud.alliancecan.ca
- cedar.cloud.alliancecan.ca
- graham.cloud.alliancecan.ca
- beluga.cloud.alliancecan.ca

Storage space

- **home** - personal data: 50G, backed up.
- **project** - long term, group storage: 1T per group, up to 40T by request; backed up.
- **scratch** - group, temporary storage: 20T per user, up to 200T by request; old files are removed in 60 days.
- **nearline** (tapes) - archive

Supercomputing: *Resources*

Current resources across the country[1]

Resource	Cedar	Graham	Niagara	Béluga	Narval	Total
CPU cores	100,400	41,548	80,640	39,120	83,216	**344,924**
GPUs	1,352	520	N/A	688	632	**3,192**
Storage[2]	23PB	16PB	3.5PB	25PB	19PB	-
Source: https://docs.alliancecan.ca/wiki/National_systems						

1. The counts may not reflect the actual numbers in service.
2. Listed here are project spaces only. Each cluster has home and scratch, which are relatively smaller. Some systems also have nearline, a tape based storage system for archive.

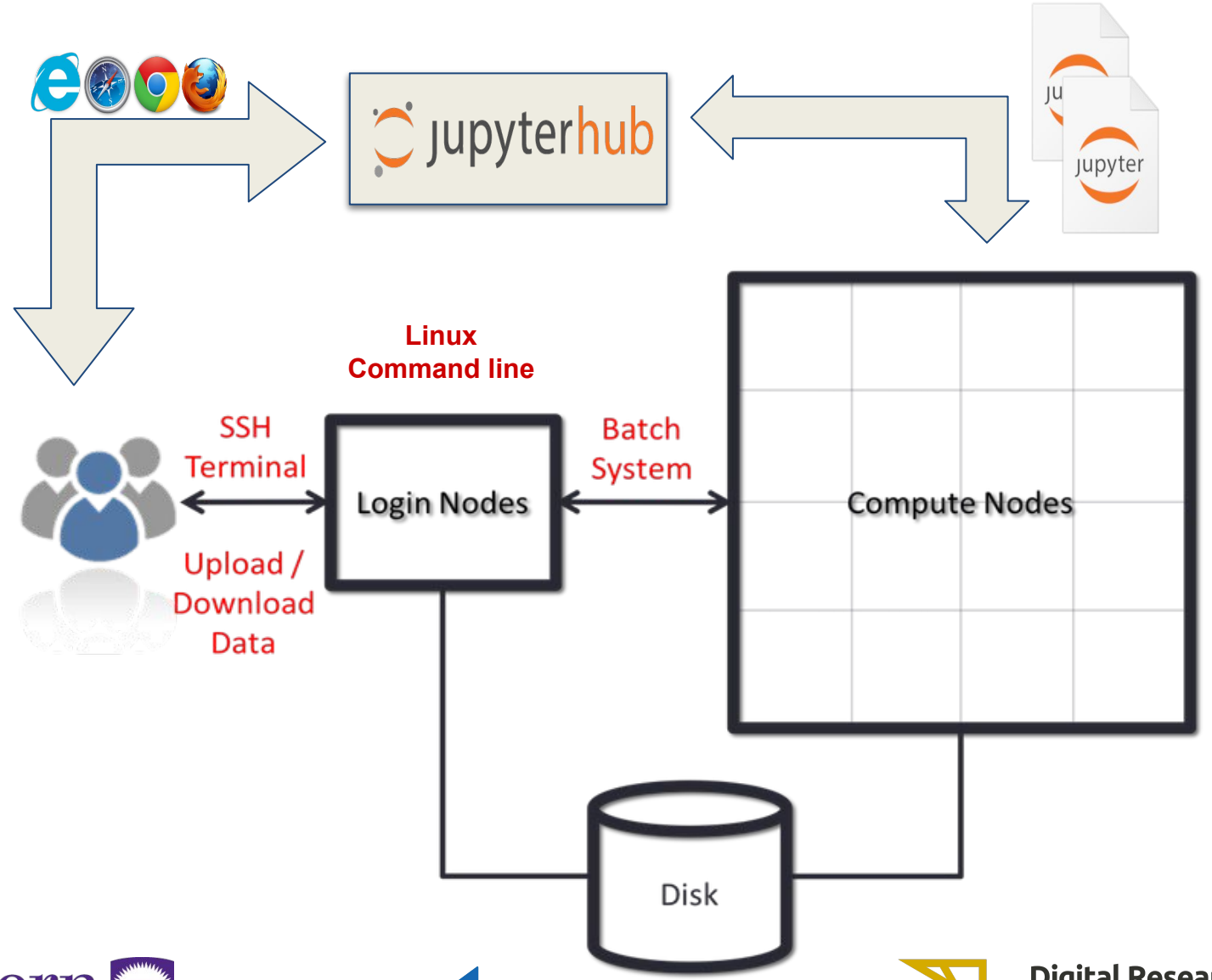
Supercomputing: *Resources*

System refresh by 2025[1]

Resource	Fir	Graham2	Niagara2	Rorqual	Total
CPU cores	165,120	134,400	235,008	131,712	**666,240**
GPUs	640	280	240	324	**1,484**
Storage[2]	49PB	25PB	29PB	TBD	-
Source: https://docs.alliancecan.ca/wiki/National_systems					

1. By July, 2025, these new systems will be in operation.
2. These numbers represent the major storage.

Supercomputing: *Working on a cluster*



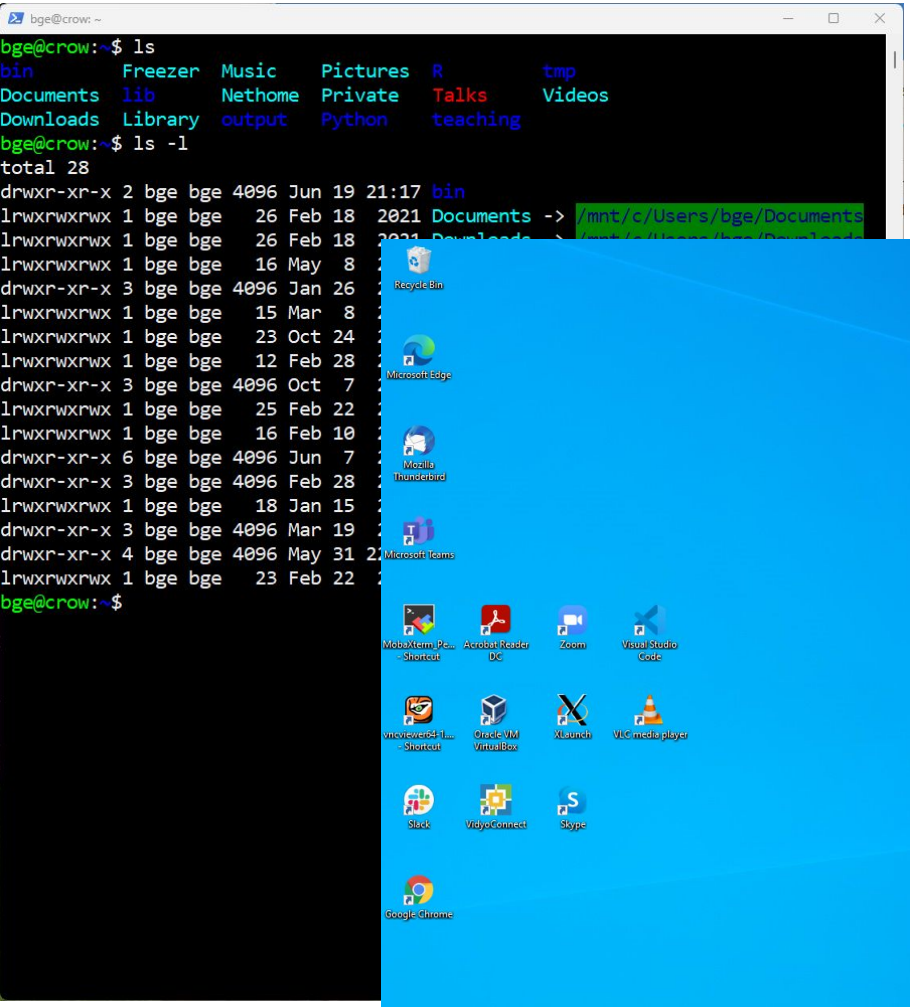
Western UNIVERSITY · CANADA

SHARCNET™

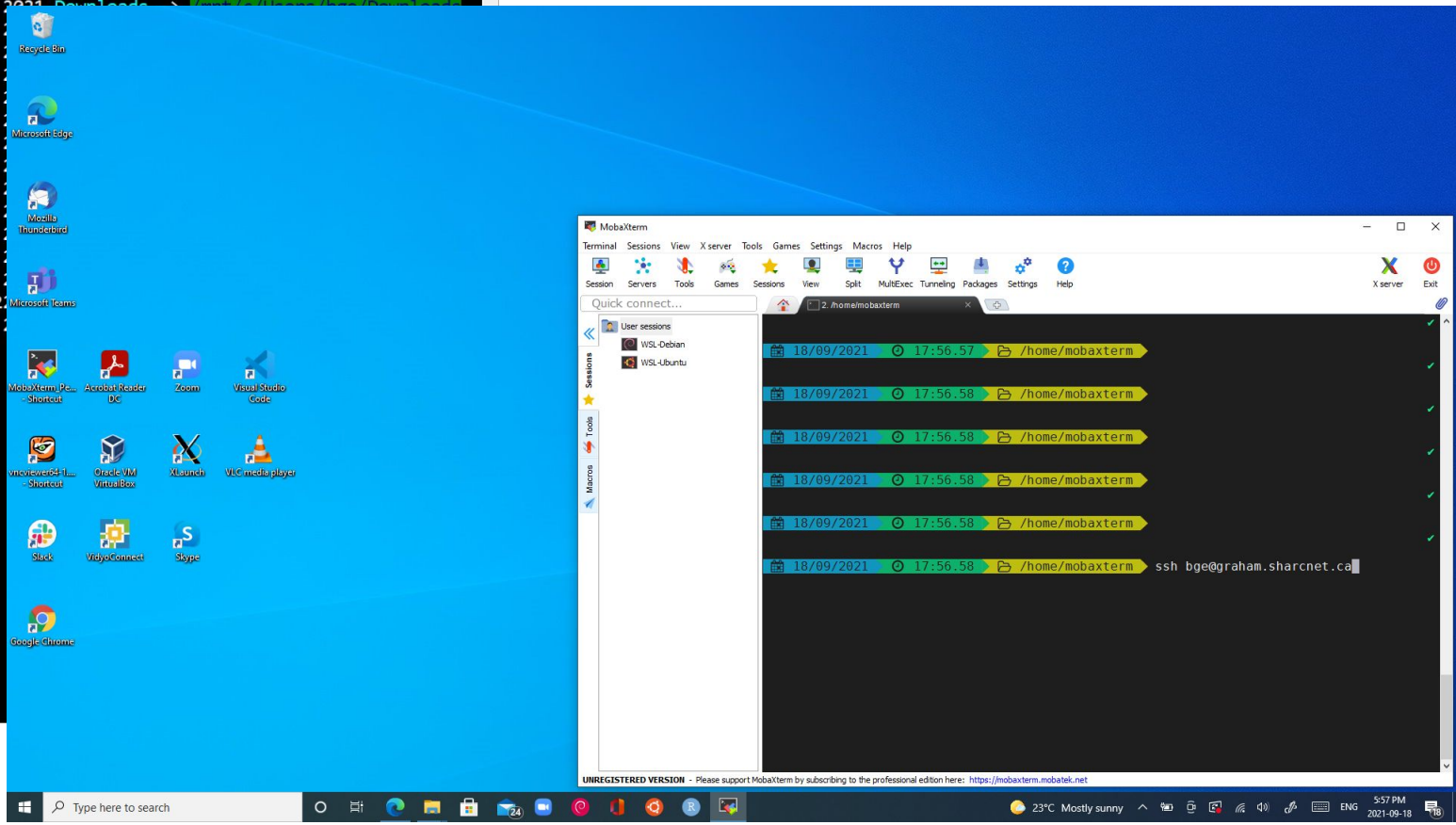
Digital Research Alliance of Canada

Supercomputing: *Connecting to a supercomputer*

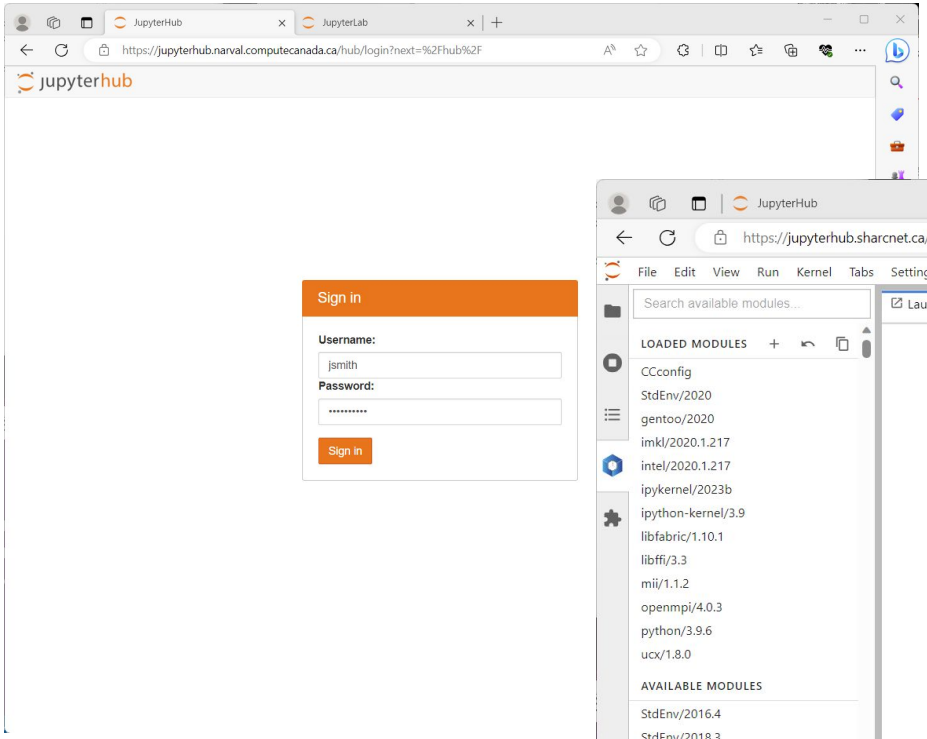
Use ssh to connect and scp to transfer file between your computer and clusters



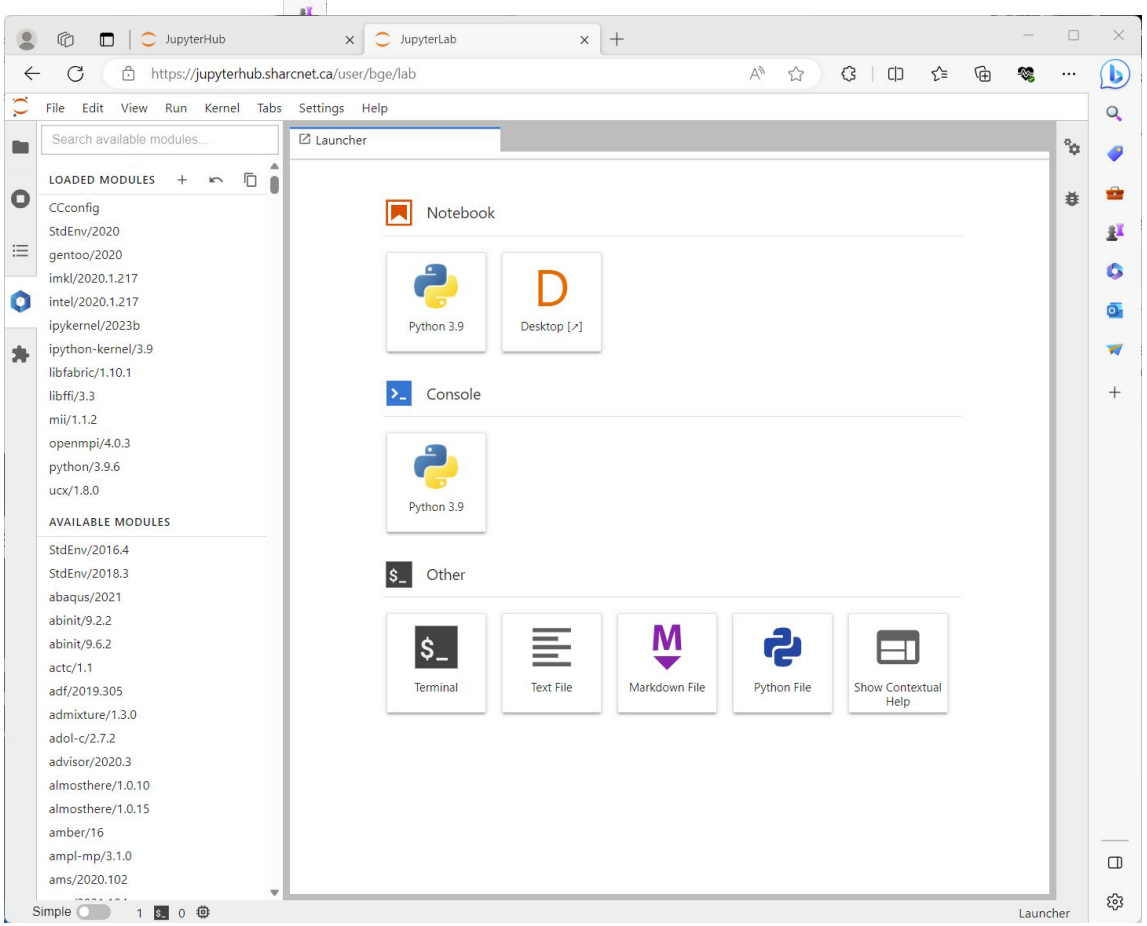
via Linux terminal



Supercomputing: *Connecting to jupyterhub*

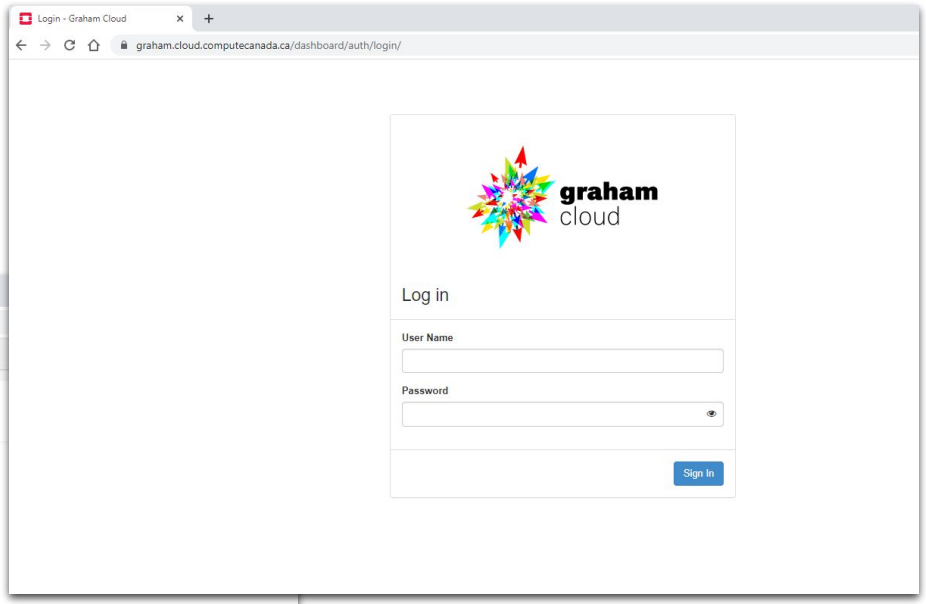


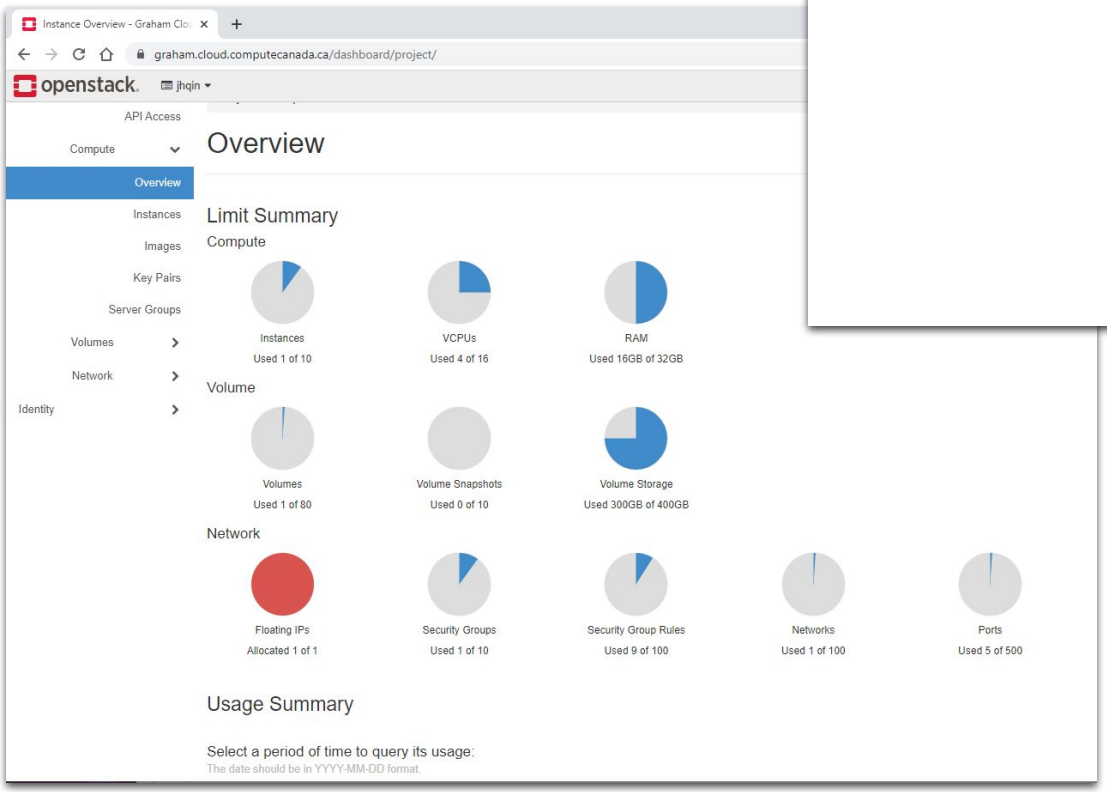
Access to the cluster via jupyter notebook, terminal and desktop in a browser.



Supercomputing: *Connecting to a cloud*

- arbutus.cloud.alliancecan.ca
- cedar.cloud.alliancecan.ca
- graham.cloud.alliancecan.ca
- beluga.cloud.alliancecan.ca





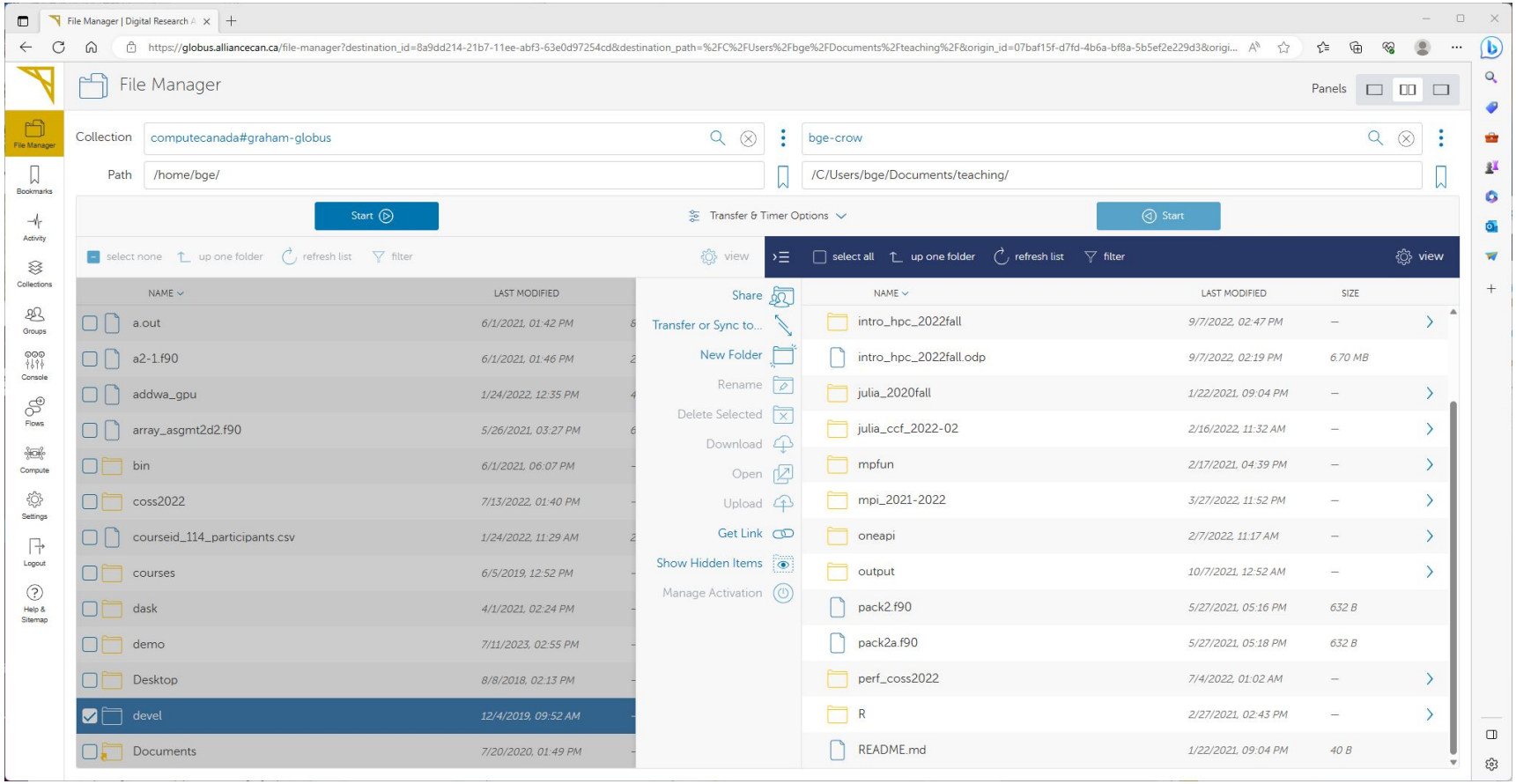
- One needs a project
- No Windows images, all VMs are Linux.
- Up to 80 virtual CPU cores.
- May have virtual GPUs.
- Up to 10 TB storage.
- Up to 25 persistent CPU cores.

Western University · Canada

SHARCNET™

Digital Research Alliance of Canada

Supercomputing: *Transferring files via Globus*[1]



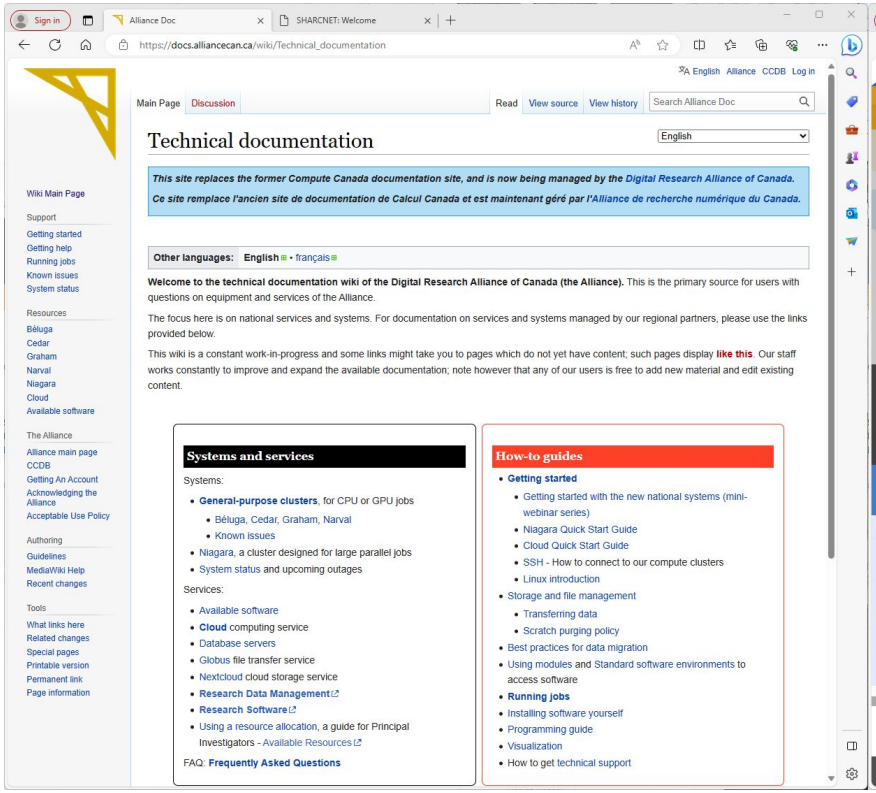
1. Check https://docs.alliancecan.ca/wiki/Globus. Go to https://globus.alliancecan.ca/ and follow the instructions

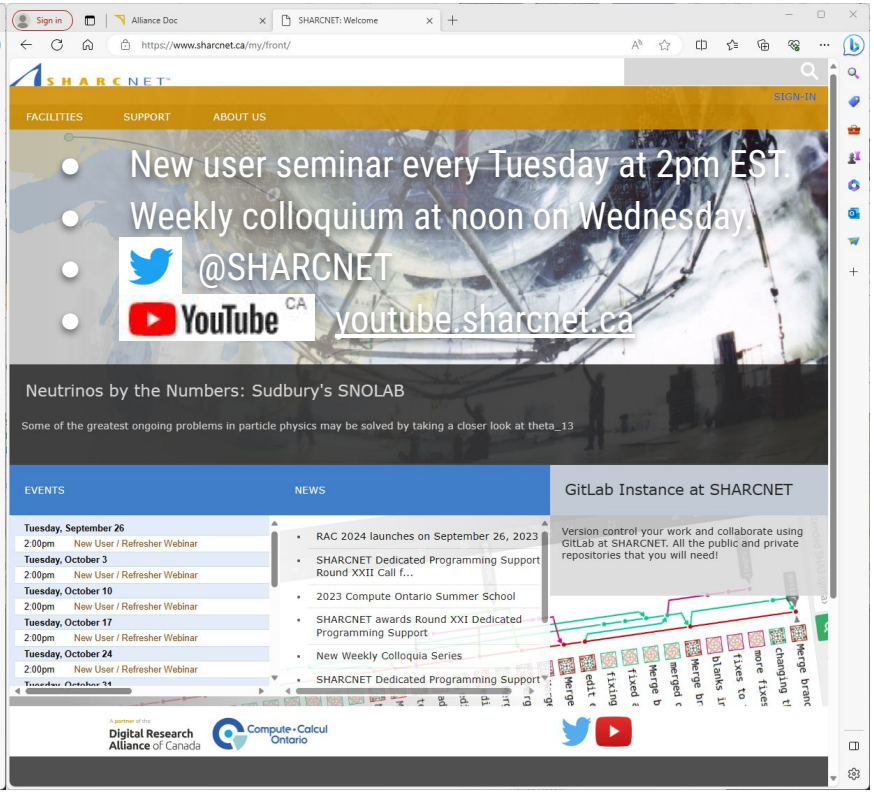
Western University · Canada

SHARCNET™

Digital Research Alliance of Canada

Supercomputing: *Getting help*





https://docs.alliancecan.ca/

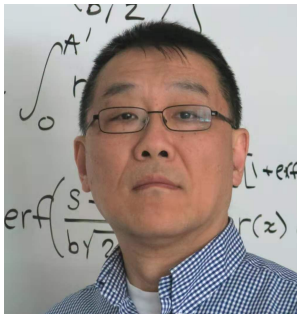
https://www.sharcnet.ca/

Supercomputing: *Getting help*

Local staff - we are in Western Science Centre, first floor

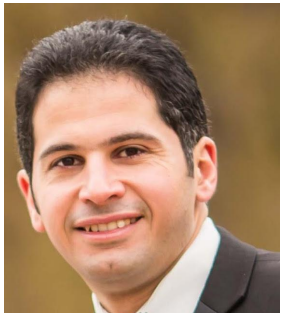
Tyson Whitehead, *HPC*, Math, Stats, EE

Jinhui Qin, *HPC*, Big Data, CS

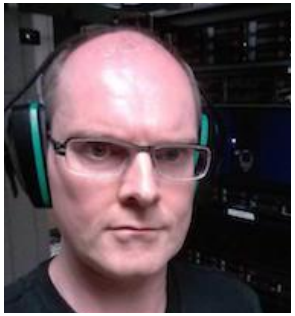


Ge Baolai, *HPC*, Applied Math

Doug Roberts (WLU), *HPC*, CFD, Commercial Software



Mohamed Elsakhawy, *Sysadmin*, Cloud, CS



Fraser McCrossan, *Sysadmin*, CS

Use of systems
Installation of software
Access to commercial software and site licence
Debugging and optimizing code
Programming
Consultation on various research problems
Grant application for compute hardware
… …

Supercomputing: *Getting help*

Help from beyond Western

- Weekly new user seminar: https://www.sharcnet.ca/my/news/calendar
- Ticketing system (most recommended): support@tech.alliancecan.ca
- Staff contact info to email or phone: https://www.sharcnet.ca/
- Arrange an office visit

Use of systems
Installation of software
Access to commercial software and site licence
Programming
Debugging and optimizing code
Consultation on various research problems
Grant application for compute hardware
...

Supercomputing: *Training events*

Local training events, workshops

- Local workshops
- Annual summer school - week long, multi-streams, many courses, mostly hands-on.
- Online, in-person/self-paced learning training course

Supercomputing: *Programming support*

Dedicated programming support

- Staff spending 50% of time working with the PI on specifically defined programming tasks.
- The DP programme runs for about 4 months.
- There are two to three calls a year for PIs for apply.
- The applications are reviewed based on the scientific merits and the feasibility of the proposed programming project.

Supercomputing at Western, SHARCNET and beyond

- Supercomputing at Western, SHARCNET and beyond
- PIs applying for compute, storage and cloud resources
- What every graduate student should know
- Introduction to advanced research computing courses
- Q & A

Resource allocation competition (RAC) background

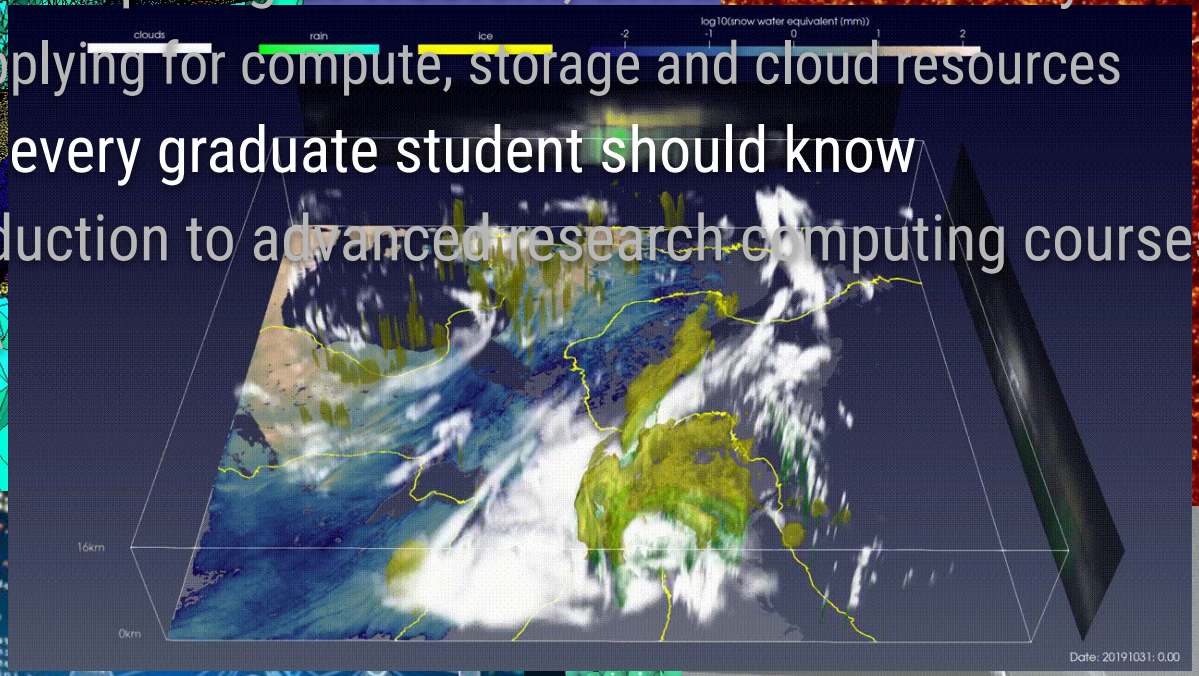
- Every user with an Alliance account may use the supercomputers (clusters), clouds and storage any time through what's termed "opportunistic access". The amount of compute and storage capacity a research group may have access to has limits.
- The majority of the resources are allocated through resource allocation competition (RAC) process, the remaining portion is for *opportunistic access* aka *default allocation* or RAS.
- With RAC allocations, research groups get their queued jobs start sooner, access to more storage and cloud resources beyond the default limits.
- The RAC applications are peer reviewed (scientific and technical reviews).

RAC (cont'd)

- First time applicants must consult with Digital Research Alliance of Canada technical staff for assessment. Send an e-mail to help@sharcnet.ca
- Info session dates:
  - Oct. 1 (English), Oct. 2 (French) on RAC general.
  - Oct. 3 (English), Oct. 4 (French) on GPUs.
  - Oct. 7 (English) on cloud.

Supercomputing at Western, SHARCNET and beyond

- Supercomputing at Western, SHARCNET and beyond
- PIs applying for compute, storage and cloud resources
- What every graduate student should know
- Introduction to advanced research computing courses
- Q & A



What every graduate student should know

Who uses the ARC resources across the country?

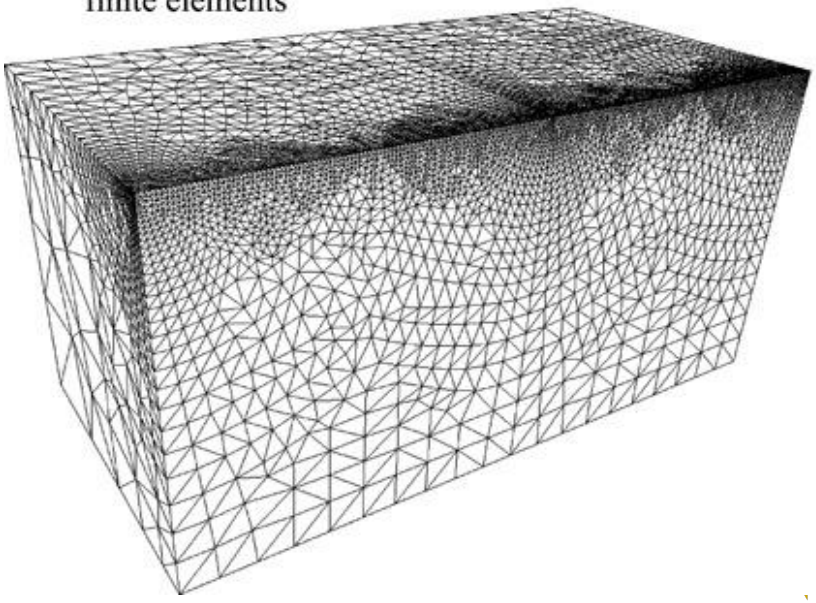
- Your peers.
- Graduate students and postdocs.
- Faculty and researchers.

What to know: *Applications*

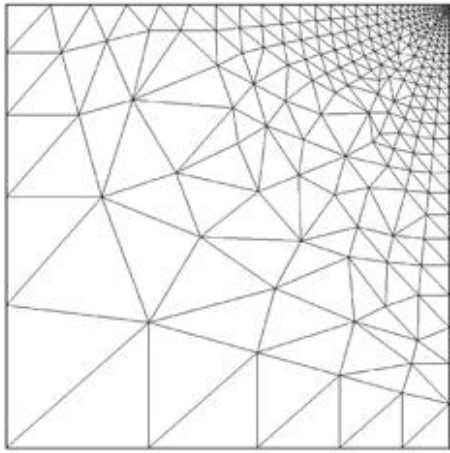
What people do on supercomputers

- Astrophysics simulations
- CFD, Environmental simulations, etc
- Coupled simulations
- Material science
- Applications of AI and machine/deep learning
- Economics, finance studies, etc.

3D mesh of linear tetrahedral finite elements



2D mesh of linear triangular finite elements



Western UNIVERSITY · CANADA

What to know: *Knowledge*

A **supercomputer** is a lot of computers, not a super fast computer

- Thousands of computers, CPU cores, GPUs, and disks
- Requires **concurrent processing** to get work done faster

Research supercomputers run **Linux** and not Windows

- Software has to support Linux to run on the supercomputer
- Linux is a publicly developed Operating System freely available

Background theory

- **Computer architecture**
- **Algorithms and numerical methods**

High performance programming is hard

- Efficient programs, algorithms, and **libraries** take decades – use them

What to know: *Linux and tools*

Using Linux

- Moving around.
- Using shell, automating tasks.
- Remember, the core utilities of Linux are very fast.
- Connecting to other computers via Secure Shell (SSH).
- Access to file systems.
- Running programs.

Using git

- To manage your research project files.
- To access, create git repositories.

What to know: *Popular and lesser known items*

Programming languages

- Higher level: Python, R, Matlab
- Lower level: C/C++, Fortran
- New takes: Julia, Chapel

Libraries

- Parallel programming: OpenMPI, OpenMP, OpenCL, CUDA/HIP
- Numerics classic: BLAS, LAPACK, ScaLAPACK, FFTW
- Numerics exascale/accelerators: magma, slate, heFFTe

Tools

- Editors and integrated development environments
- Compilers (gcc, Intel), interpreters (python, R), and just in time compilers
- Debuggers and profilers (gdb, DDT, MAP, etc.)

What to know: *Popular and lesser known items*

Data science/data wrangling:

- python: numpy, pandas, matplotlib, plotnine, sklearn , dask, rapids
- R: data frames, tidyverse (dplyr, simplr, ggplot, etc.)
- SQL

Machine learning and AI

- TensorFlow
- Keras and PyTorch

Visualization

- ParaView and VisIt

… and more …

Learning is a lifelong journey…

Supercomputing at Western, SHARCNET and beyond

- Supercomputing at Western, SHARCNET and beyond
- PIs applying for compute, storage and cloud resources
- What every graduate student should know
- Introduction to advanced research computing courses
- Q & A

Training courses

SHARCNET offers a variety courses, e.g.

- Introduction to supercomputing
- Introduction to Shell
- Introduction to machine learning
- Introduction to SQL
- Python for high performance computing
- Parallel programming with C++
- Parallel programming with Fortran
- Parallel programming with GPUs
- Programming distributed systems with message passing interface (MPI)
- Introduction to scalable and accelerated data science with Python
- Visualization of scientific data
- More… (https://training.sharcnet.ca/)

Training courses

For example, the ***introduction to supercomputing*** course contains

- Introduction to advanced research computing
- Skill sets and knowledge one should have
- Cluster computing
- Shell programming
- Modern computer architecture
- Fundamentals of scientific computing
- Parallel computing
- Programming languages
- Programming shared memory systems
- Programming distributed memory systems
- Programming GPUs
- High performance computing
- Dealing with data with SQL
- Visualization of data
- Etc.

Self-paced study
Office hours
Quizzes
Homework assignments
Simple, straightforward to bring one up to speed

Training courses

How to enroll

- Sign up for the courses at https://training.sharcnet.ca/

Format and schedule

- Live lectures, workshops fall and winter semesters.
- Self-driven, grading offline courses with course materials, recorded videos, quizzes and assignments.
- Summer schools.
- Forums on different topics are open for attendees.

Outcome

- Certificates of completion.

Q & A

- SHARCNET https://www.sharcnet.ca/
- Compute Ontario https://www.computeontario.ca/
- Digital Research Alliance of Canada https://alliancecan.ca/
- Help: help@sharcnet.ca