Supercomputing in Canada:
An Introduction to WestGrid & Compute Canada

Patrick Mann, Director of Operations, WestGrid

Monday, September 17, 2018
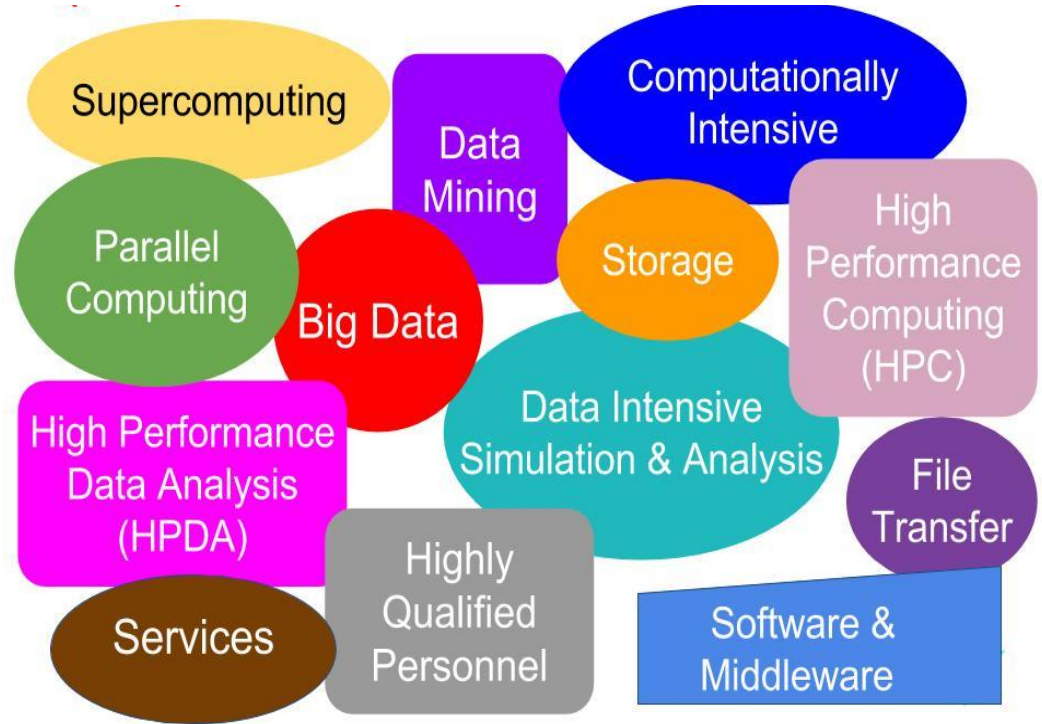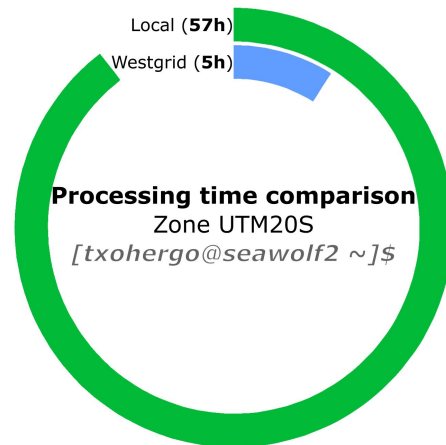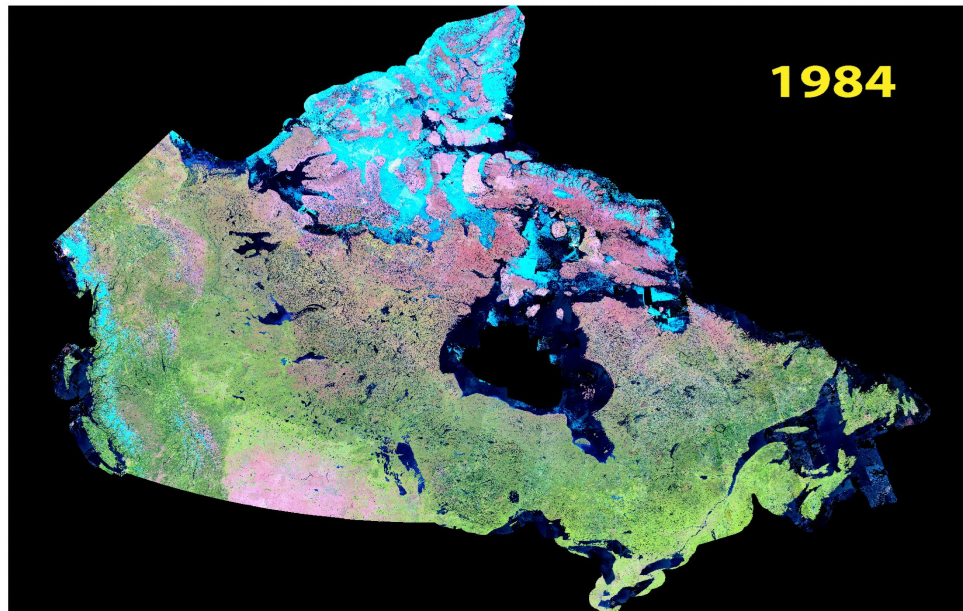
Outline

- What is Advanced Research Computing?

- Why Use ARC? Who Uses ARC?

- Overview of Compute Canada and WestGrid

- Resources & Services Overview

- Account Eligibility

- Industry Support

- Training Opportunities

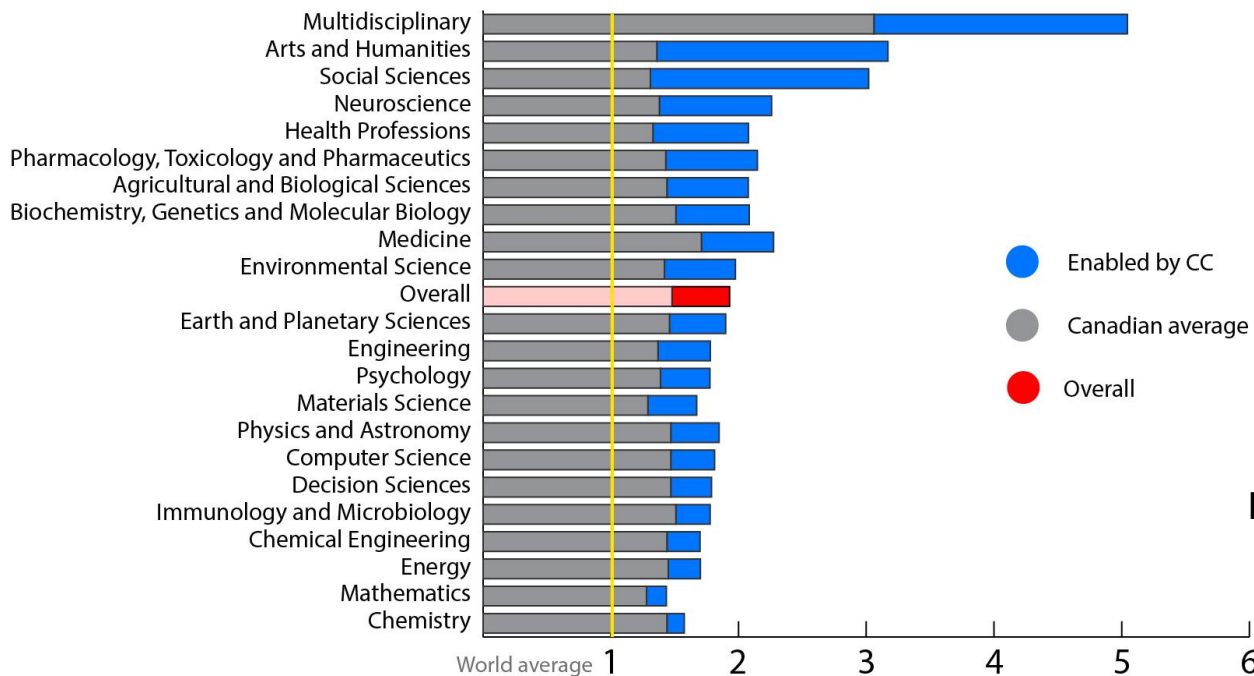# Advanced Research Computing?

Advanced Research Computing (ARC) is **everything beyond a standard desktop workstation**.

Why use ARC?

1984

Local (**57h**)
Westgrid (**5h**)

Processing time comparison
Zone UTM20S
[txohergo@seawolf2 ~]$

Bibliometrics

Impact of publications enabled by Compute Canada compared to the average Canadian impact

- Multidisciplinary
- Arts and Humanities
- Social Sciences
- Neuroscience
- Health Professions
- Pharmacology, Toxicology and Pharmaceutics
- Agricultural and Biological Sciences
- Biochemistry, Genetics and Molecular Biology
- Medicine
- Environmental Science
- Overall
- Earth and Planetary Sciences
- Engineering
- Psychology
- Materials Science
- Physics and Astronomy
- Computer Science
- Decision Sciences
- Immunology and Microbiology
- Chemical Engineering
- Energy
- Mathematics
- Chemistry

World average: 1, 2, 3, 4, 5, 6

Legend:
- 🔵 Enabled by CC
- ⚫ Canadian average
- 🔴 Overall

Field-Weighted Citation Impact (FWCI) of CC-enabled papers

Who is using WestGrid / Compute Canada?

- **Drs. Michael Bowling, Richard Sutton, Patrick Pilarski**

  Working with artificial intelligence powerhouse Deep Mind to launch a research lab in Edmonton, AB -- the company's first outside the United Kingdom.

- **Dr. Susan Brown & Dr. Geoffrey Rockwell**

  Developed Voyant, a digital humanities application made for investigating texts.

- **Dr. Victoria Kaspi** - 2016 Gerhard Herzberg Gold Medal Winner

- **Dr. Arthur McDonald** - 2015 Nobel Prize in Physics Winner

- **As well as:**
  - **46%** of the Canada Excellence Research Chairs
  - **32%** of NSERC's Canada Research Chairs
  - **25%** of Canada's highly cited researchers

Compute Canada Federation (CCF)

Canada's National Provider of Shared Essential Digital Research Infrastructure (DRI)

- CC is a not-for-profit corporation. The membership includes 35 of Canada's major research institutions and hospitals.
- Funding is through a federal grant with matching funds from provincial and institutional partners (40% federal / 60% provinces and institutions), which is the basis of the federated Canadian model.
- Provide shared services to over 12,000 researchers across Canada. No fees. Large requests based on a merit-based access system.
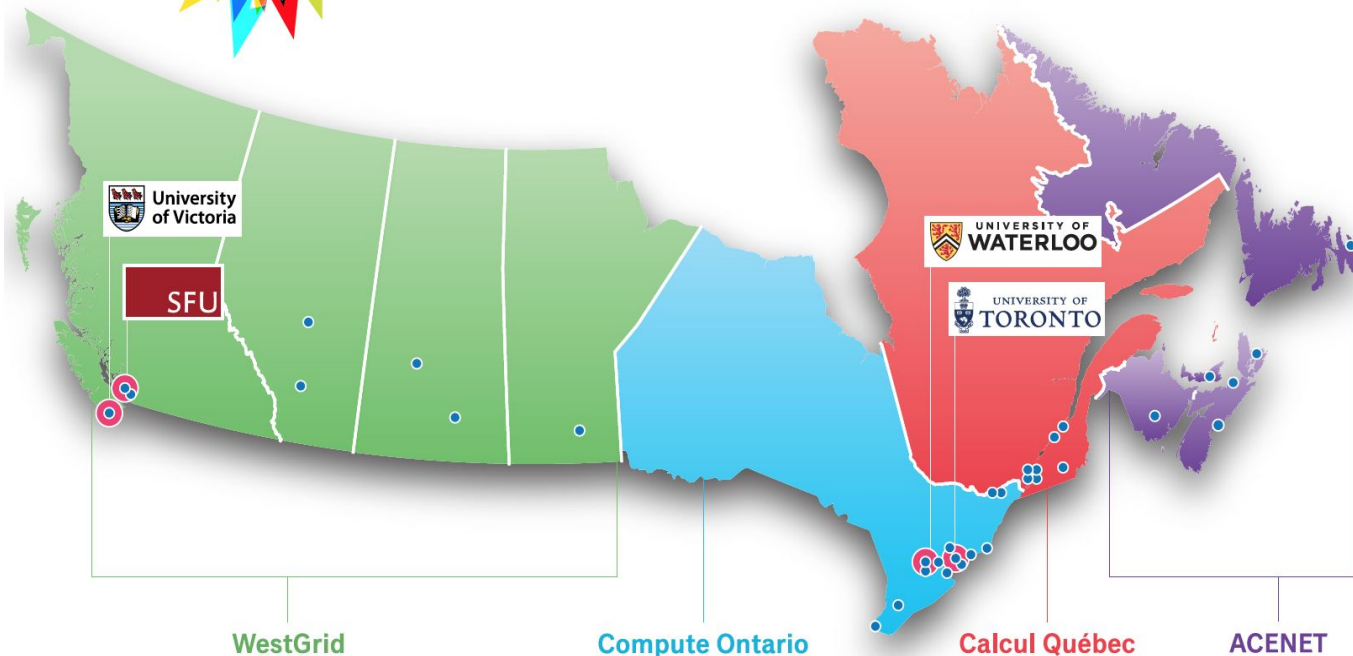
Member locations and new national hosting sites

Regional Consortia

Compute Canada:

> Leadership role and national scale initiatives: Resource Allocation Competition (RAC), procurement, coordinating research data management and other national services, securing funding, advocacy, national/international partnerships

Regions:

> Coordinate local activities, such as training, staff management and ensuring needs of the institutions and provinces are being met.

Brief History of WestGrid

WEST GRID

MACI Formed

A collaboration among Alberta universities, the Multimedia Advanced Computational Infrastructure (MACI) forms in 1997 to address the need for shared high performance computing.

Compute Canada Formed & Incorporated

WestGrid and six other regional consortia join in 2006 to form Compute Canada. Receives $60 million from CFI (National Platforms Fund) and $2 million/year from NSERC. WestGrid's 14 partner institutions receive money for infrastructure. Compute Canada incorporates in 2012 as a not-for-profit with 27 inaugural members. Receives $30 million from CFI for maintenance and operating funds.

1997 | **2001-2** | **2003-5** | **2006-12** | **2016-17** | **2018-**

WestGrid Formed

Seven Alberta and British Columbia research institutions collaborate in 2001 to create WestGrid. Funding is granted through the Canada Foundation for Innovation (CFI).

Facilities Launched

Computing, storage, and collaboration facilities launch in 2003 at the seven founding partner. A high-speed optical link is set up between WestGrid and SHARCNET (Ontario) in 2005, representing a first step towards a pan-Canadian network of research computing facilities.

WestGrid Incorporated

WestGrid incorporates as a not-for-profit. Arbutus (UVic) and Cedar (SFU) launch in 2016 and 2017 as national systems serving researchers across Canada.

Board Appointed

By March 30, 2018, WestGrid appoints its inaugural Board of Directors, its final legacy systems are decommissioned, and Arbutus and Cedar are expanded.
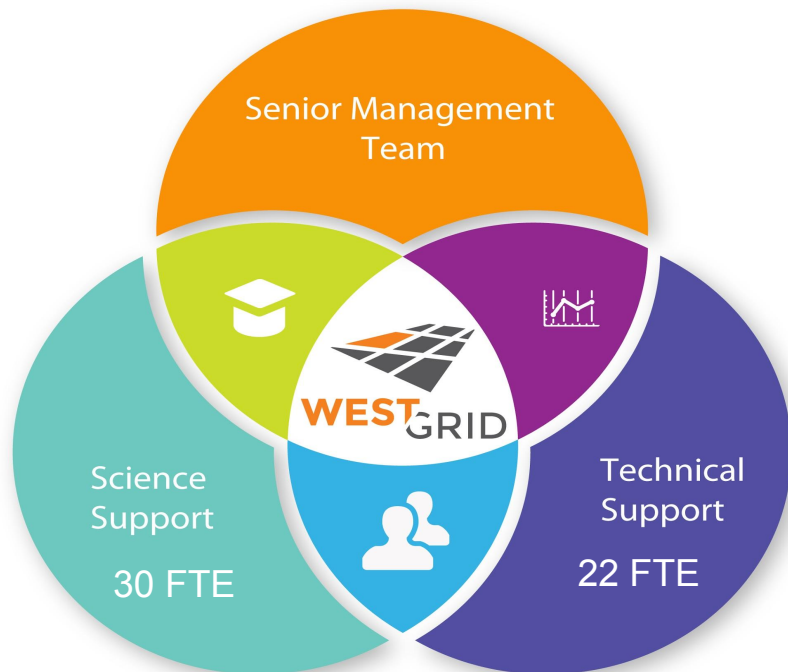
WestGrid Today

VISION: *To ensure researchers in Western Canada are not impeded by any barriers accessing the ARC tools and services required to lead world class discoveries.*

Goals:
1. Develop ARC HQP
2. Support national, regional and provincial collaborations
3. Advocate for ARC needs in Western Canada
4. Create positive user experience

Senior Management Team

Science Support

30 FTE

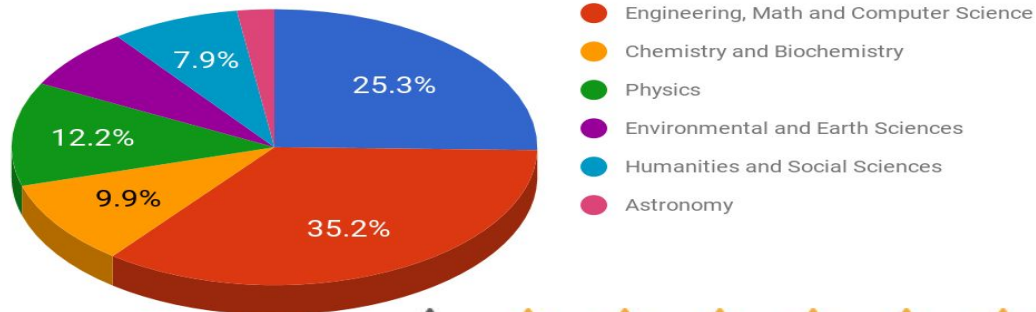Technical Support

22 FTE

WestGrid Members & Partners

Members pay a membership fee to WestGrid and receives funding through the Canada Foundation for Innovation (CFI) Major Science Initiatives (MSI) program. WestGrid Members also provide matching funds for all CFI grants.
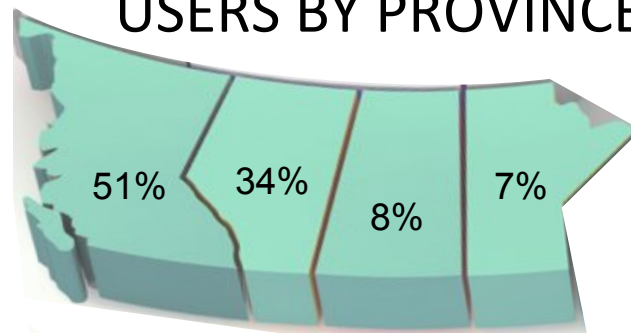
Institutional Partners are the home institutions of any WestGrid user or any research organization that collaborates with WestGrid to lead the acceleration of research and innovation.
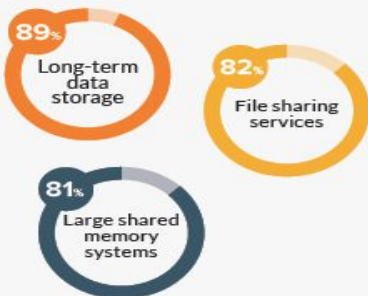
Our Users

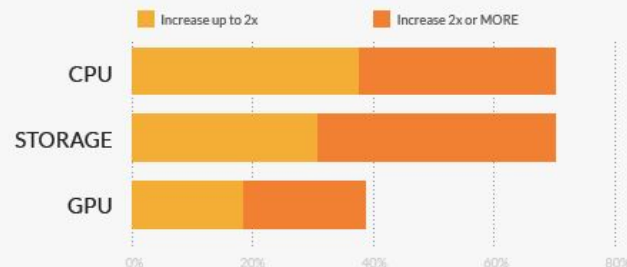## WestGrid Faculty Accounts by Research Area

As of January 2018

- Medical, Biological and Life Sciences
- Engineering, Math and Computer Science
- Chemistry and Biochemistry
- Physics
- Environmental and Earth Sciences
- Humanities and Social Sciences
- Astronomy

25.3%
35.2%
9.9%
12.2%
7.9%

USERS BY PROVINCE

51%   34%   7%
8%

91% are satisfied with WestGrid

MOST IMPORTANT SERVICES & RESOURCES:
(rated somewhat/very important)

- 89% Long-term data storage
- 82% File sharing services
- 81% Large shared memory systems

HOW WILL YOUR NEEDS FOR COMPUTE AND STORAGE RESOURCES CHANGE FROM 2018-2022:

Increase up to 2x   Increase 2x or MORE

CPU

STORAGE

GPU

0%   20%   40%   60%   80%

compute | calcul
canada | canada

From our users…

"The services provided by Westgrid are invaluable and our organization would be substantially less productive without WestGrid's assistance. Some of our research would have been impossible without WestGrid." - University of Victoria, Environmental & Earth Sciences Research Staff

"I strongly believe that the usefulness of Compute Canada in general, and of Westgrid in particular, is a direct function of the strong local support. In that sense, we are really happy to have the team of Dr. Shamov in Winnipeg; they provide excellent service to me and my students." - Manitoba Principal Investigator

What we do: National Host Sites

Arbutus (**University of Victoria)**
- OpenStack cloud computing
- 9,000 CPU cores
- 3.5 PB Ceph storage
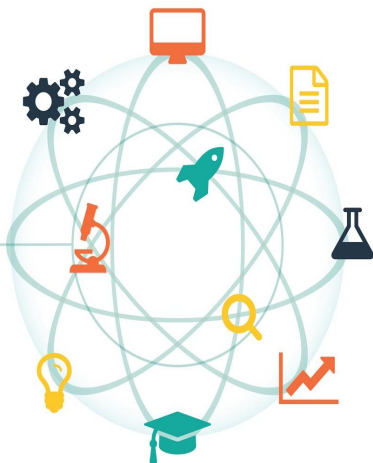- 87% vCPU usage

Cedar (Simon Fraser University)
- General purpose computing, multiple node types
- Over 58,000 CPU cores
- 3.6 petaflops peak performance
- 584 GPU NVidia P100's
- 10 PB /project storage

What we do: Training & Support

SUPPORT
MORE THAN

3,300

RESEARCHERS
ACROSS FOUR
PROVINCES

SINCE 2012
WESTGRID EXPERTS DELIVERED

24,000+
HOURS OF TRAINING TO

6,000+
RESEARCHERS AT OVER

200
OUTREACH & TRAINING EVENTS

What we do: Training & Support

In **2017-18** we delivered:

- **287 hrs** of training at **42** events
- **800+** RSVPs, **42%** new users
- **2** regional summer schools
- National *Visualize This!* challenge
- **24** Software Carpentry events
- **11** Community Town Halls
- Answered **34%** of national support tickets

How Can WestGrid Help You?

Resources/Services

ownCloud

Globus file transfers

Managed cloud space

Stable and secure data storage and backup

High performance, big data and GPU computing and storage

Videoconferencing

Research Data Management

Software distribution

Expertise

Consultation - Help determining required resources

Designing, optimizing and troubleshooting code

Customizing tools

Installing, operating and maintaining advanced research computing equipment

Dedicated domain-specific specialists

Cybersecurity

Training

Group and individual training and ongoing support from novice to advanced

Standard and discipline specific customized training

Livestreaming of national seminar series including VanBUG and Coast to Coast

Support, training videos and other upcoming online workshops
www.westgrid.ca

compute | calcul
canada | canada

Resource Allocation Competition (RAC)

- Competitive
- Proposals in Autumn, with science and technical review in Winter.
- Current RAC starting Sep.27. Best practices this afternoon at UManitoba.

Competitions

1. **Resources for Research Groups (RRG)**:
   a. Focussed on Compute, GPU and storage resources on the big clusters
   b. 1 year allocations: April 1 to Mar 31 allocation year
2. **Research Platforms and Portals (RPP)**:
   a. Scientific gateways generally hosted in the cloud.
   b. Up to 3 year allocations.

https://www.computecanada.ca/research-portal/accessing-resources/resource-allocation-competitions/

RAC 2019 Resources

System	Cores	GPUs	Storage	
Cedar	58,416	584 (NVidia P100)	11 PB	General Purpose, tape backup
Graham	33,472	320 (NVidia P100)	12 PB	General Purpose, tape backup
Béluga (GP4)	~30,000	~600 (??)	TBD	Phase 2 funding: delivery 2018/2019. Production for RAC 2019!
Niagara	60,000	0	2 PB	Large Parallel, tape backup
Arbutus (cloud)	~9,000 (7,640)	0	~3.5 PB	Openstack cloud Infrastructure-as-a-Service (IAAS)
Legacy	TBD	TBD		Some older systems may be available.

Prediction: **we'll have about 44% of CPU ask and only 28% of GPU asks!** **Very competitive!**

Rapid Access Service (Default allocations)

Available to all users with basic account (no competition required)

Cluster

- 50 Core Years*
- 50 GB Home
- 20 TB to 100 TB Scratch
- 5 TB Tape

ownCloud (storage)

- 50 GB

Cloud

- 10 VCPUs (persistent)
- 45 GB RAM
- 2 IP Address
- 1 TB Storage

Globus Transfers

- Unlimited

https://docs.computecanada.ca/wiki/Getting_Started
https://www.computecanada.ca

- Research projects that are supported by a recognized funding agency
- Research projects that are eligible for funding from such an agency
- Industrial research projects
- Research under contract

Industry Use

1. University appointment (e.g., adjunct professor at a university)
2. Collaboration with faculty member (sponsored)
3. As a not-for-profit / for-profit principal investigator ***COST-RECOVERY***

Entry level service block: $2000 (not-for-profit) or $3000 (for-profit) provides:
- 10,000 core hours
- up to 5 accounts
- access to a single CC system
- default storage allocation on that system (home space backed up)
- up to 10 hours of support time

WestGrid Summer Schools

Last summer University of Manitoba and UBC

The courses include:
- Introduction to high performance computing (HPC)
- Using the Compute Canada OpenStack Cloud
- Introduction to Molecular Dynamics
- Data Analysis with MATLAB
- Introduction to Parallel Programming using OpenMP
- Basics of scientific visualization with ParaView

Usually about $35-$55 for 4 days of courses

Researcher Consultation

2016-2017: Leadership Council on Digital Research Infrastructure (LCDRI)	2017-2018: WestGrid Incorporation and planning
● **August 2017**: Report sent to cabinet. ● **March 2018**: Cabinet announced $572M over 5 years for DRI (ARC, RDM, Research Network) ● **Summer 2018:** ISED planning and consultation process.	● **Spring 2018:** new WestGrid board of directors. ● **Summer 2018**: intensive internal strategic planning

This afternoon: informal meeting with researchers and anyone interested.

- RAC: issues, revisions, etc.
- New systems: what did we get right? Wrong?
- Surveys: what should we be asking about?
- Future: what should WG be focussing on? Compute Canada?
- What should we be emphasizing to gov't?
- Anything Else?

About an hour depending on interest and questions

ISED DRI Discussion Paper

Innovation, Science and Economic Development Canada (ISED)
"Canada's DIGITAL RESEARCH INFRASTRUCTURE STRATEGY Discussion Paper July 2018"

In the 2018 federal budget, the Government of Canada committed to greatly strengthen support for Canadian scientists and researchers in conducting world-leading research. As part of this vision, the 2018 budget announced an investment of $572.5 million over five years, with $52 million ongoing, to implement a Digital Research Infrastructure (DRI) Strategy for Canada.

1. **Digital network for Research:** CANARIE
2. **Advanced Research Computing (ARC)**: Compute Canada
3. **Data Management**: research libraries, CARL, Research Data Canada
4. **Research Software**

Lack of Resources

The major issue is lack of resources.

The overwhelming impression from a detailed re-reading of the RAC 2018 WestGrid proposals is that of well-written, well-justified asks from competent and generally experienced teams. The prevalent complaint from the survey, RAC proposals and informal discussion is that of very long queue times and inability to acquire sufficient resources to carry out a research program.

So a strong recommendation is that WestGrid should consider acquiring additional resources.

- Should WestGrid consider buying hardware?
- Where would we find funds? Who should we approach?
- Can we obtain provincial funds? Currently AB, SK and MN do not have sites/systems.
- Where? How do we decide?
- What should we aim for?
  - GPUs and machine learning?
  - ARM-based clusters?
  - ???

Contact us anytime:

info@westgrid.ca
support@westgrid.ca
www.westgrid.ca
docs.computecanada.ca

Any issues or problems? We can advocate for WG member and user concerns within Compute Canada.

And the Manitoba Site Team at the University of Manitoba
Site lead: Grigory Shamov grigory.shamov@westgrid.ca

More information / extra slides

Getting started with the systems

- **Documentation Wiki:**
  - https://docs.computecanada.ca
  - User guides, system details, best practices, etc.
- **Mini-Webinar Tutorials:**
  - Youtube > Compute Canada -- Introductory videos covering the basics of how to use Cedar & Graham
- **WestGrid Training:**
  - www.westgrid.ca/training
  - Webinars & in-person training sessions
- **Questions / Need Help?**
  - support@computecanada.ca

Leading and Renewing Canada's Advanced Research Computing

Four National Data Centres (Stage 1)

Arbutus (University of Victoria)
- OpenStack cloud computing
- Over 6,750 CPU cores
- In production September 2016
- 30% expansion in 2017

WESTGRID

compute | calcul
canada | canada

Leading and Renewing Canada's Advanced Research Computing

Four National Data Centres (Stage 1)

Cedar (Simon Fraser University)
- General-purpose computing, with multiple node types
- Over 26,000 CPU cores
- Batch-oriented parallel/serial
- Over 500 NVIDIA Pascal GPUs

WEST GRID

compute | calcul
canada | canada

Leading and Renewing Canada's Advanced Research Computing

Four National Data Centres (Stage 1)

Graham (University of Waterloo)
- General-purpose computing, with multiple node types
- Over 30,000 CPU cores
- Workload portability & resiliency
- Over 300 NVIDIA Pascal GPUs

Leading and Renewing Canada's Advanced Research Computing

Four National Data Centres (Stage 1)

Niagara (University of Toronto)
- 60,000-70,000 CPU cores
- Purchase target mid-2017
- Designed for large parallel workloads

WESTGRID

compute | calcul
canada | canada

Infrastructure Highlights 2017-19

- Four new national systems connected by CANARIE 100Gb backbone
- National storage infrastructure including:
- More disk storage than before (20PB => 60PB+)
- Considerable tape storage capacity in two of the sites
- Object Store
- Cloud services on three of the sites
- Redundant services between sites
- Heightened security posture at the new sites
- More uniformity between sites
- National services deployed at all four sites (supported by common login, national issue tracking system, etc.)

Compute Canada Cloud

For researchers who need:

- long running jobs or services, rather than batch processing
- a virtual machine (VM)or specialized virtual clusters for big data
- to control their own operating system

  the ability to customize software stacks
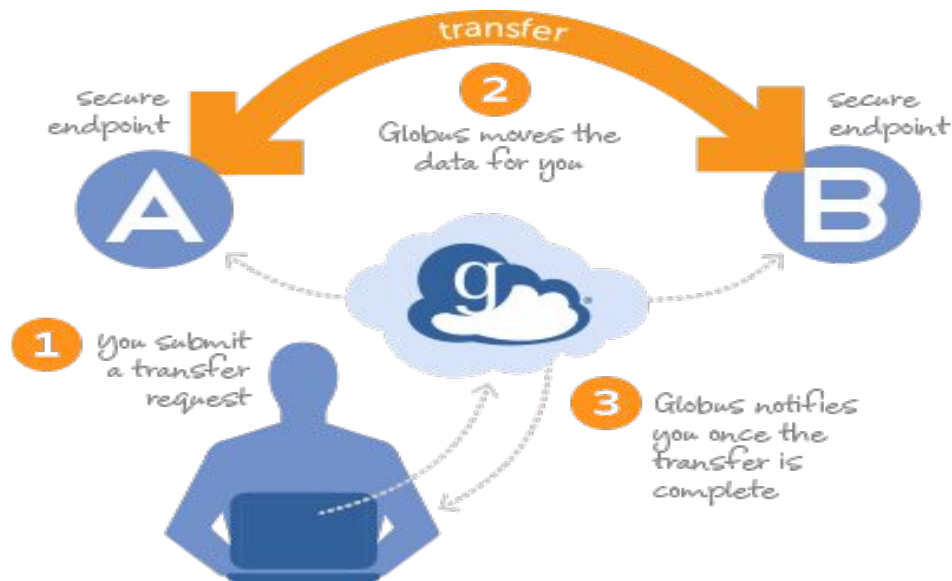
ownCloud - online storage

A safe home
for all your data

Version 9.1
out now!

ownCloud

- Dropbox-like storage resource
- <u>Free</u> for all Compute Canada account holders
- Easy to use, drag-and-drop
- Secure – data stays in Canada
- up to 50 GB per user

WESTGRID

compute | calcul
canada | canada

Globus - Data Transfer Service

Data transfer that is:
- Fast
- Secure
- Easy

Connection to 24 of our sites

Research Data Management

https://portagenetwork.ca/frdr-dfdr/

Portage DMP Assistant

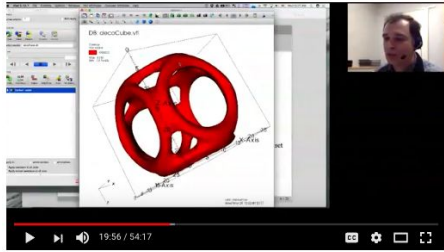**DMP Assistant** is a bilingual tool for preparing data management plans (DMPs).

The tool follows best practices in data stewardship and walks researchers step-by-step through key questions about data management.

Vidyo - videoconferencing technology

- Easy-to-use HD quality videoconferencing that supports video collaboration between [desktops](#) (Windows, Mac, Linux), devices ([IOS](#), [Android](#)) and [h.323 systems](#).

- Vidyo can support:
  - distance collaboration between research teams
  - livestreaming & recording of research seminars
  - remote PhD defenses

ARC Training and Education

WestGrid offers hands-on training & information sessions:

- Online webinars
- In-person workshops
- Summer Schools

… in introductory & advanced topics:

- Intros to HPC, cloud, Linux, etc.
- Programming languages & tools
- Data visualization
- Research data management
- Software Carpentry and Data C
- Domain-specific ARC training

What we do: Training & Support

SUPPORT MORE THAN
3,300
RESEARCHERS ACROSS FOUR PROVINCES

10,000 HOURS
DEVOTED ANNUALLY, IN DIRECT SUPPORT OF RESEARCHERS

SINCE 2012
WESTGRID EXPERTS DELIVERED
24,000+ HOURS OF TRAINING TO
6,000+ RESEARCHERS AT OVER
200 OUTREACH & TRAINING EVENTS

In **2017-18** we delivered:

- **287 hrs** of training at **42** events
- **800+** RSVPs, **42%** new users
- **2** regional summer schools
- National *Visualize This!* challenge
- **24** Software Carpentry events
- **11** Community Town Halls
- Answered **34%** of national support tickets

compute | calcul
canada | canada

43