
Compute Canada's Resource Allocation Competition 2021 (RAC 2021): BEST PRACTICES



Patrick Mann, Director of Operations
WestGrid

<https://westgrid.github.io/trainingMaterials/getting-started>

1. Intro to WestGrid and Compute Canada
2. What is the RAC?
3. RAC Stats
4. RAC Overview and admin details
5. RAC Updates and changes
6. RAC Tips and best practices
7. RRG (first hour)
8. RPP (second hour)
9. Questions and discussion

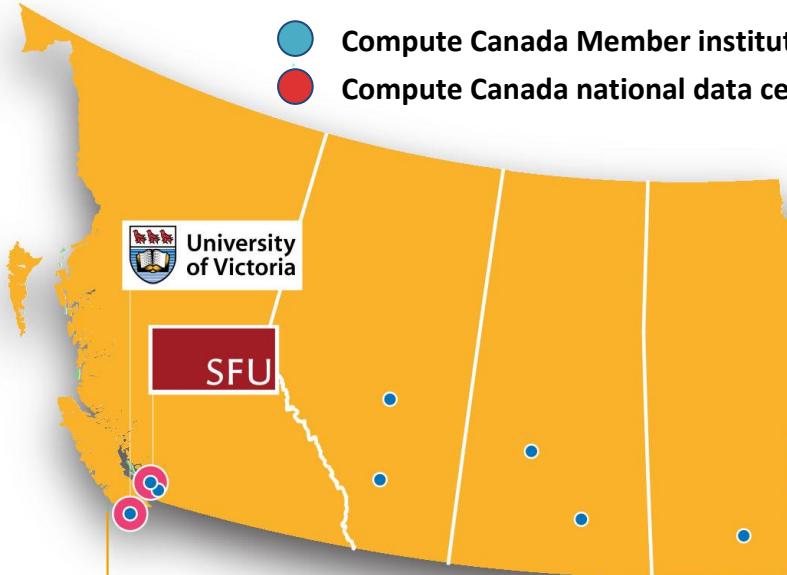


compute
canada | calcul
canada

Canada's National ARC Platform

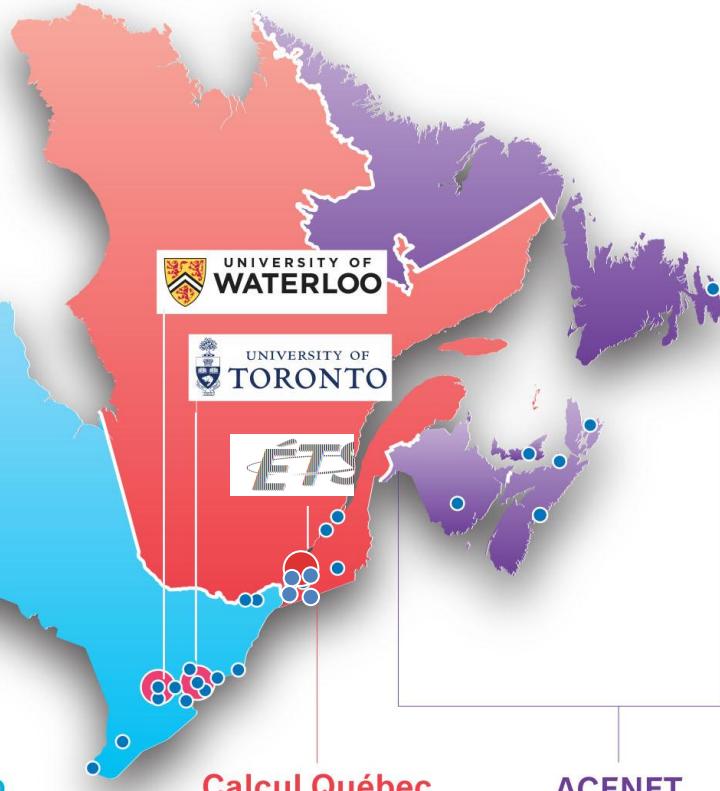
- Compute Canada Member institution
- Compute Canada national data centre host

Total users
15,994



WestGrid

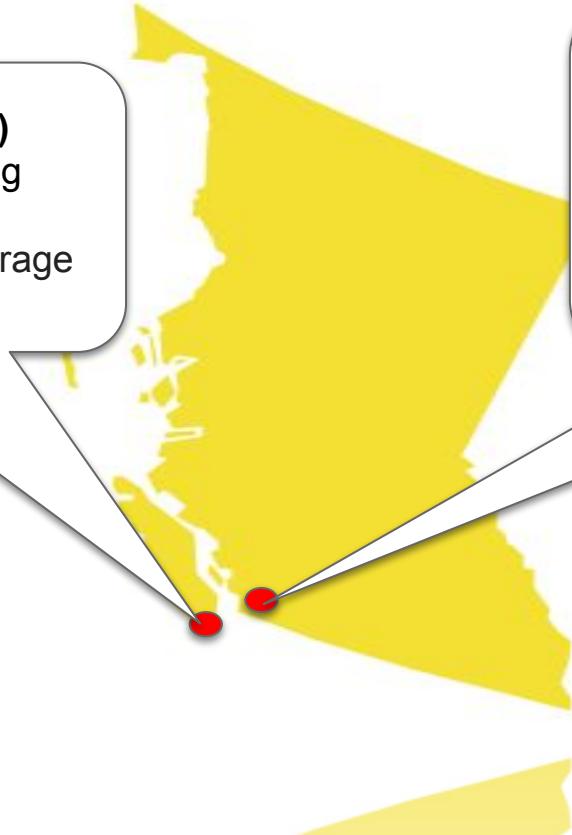
Compute Ontario



National Host Sites

Arbutus (University of Victoria)

- OpenStack cloud computing
- 14,968 (456 nodes)
- 5.7 PB persistent Ceph storage
- 146,944 GB RAM

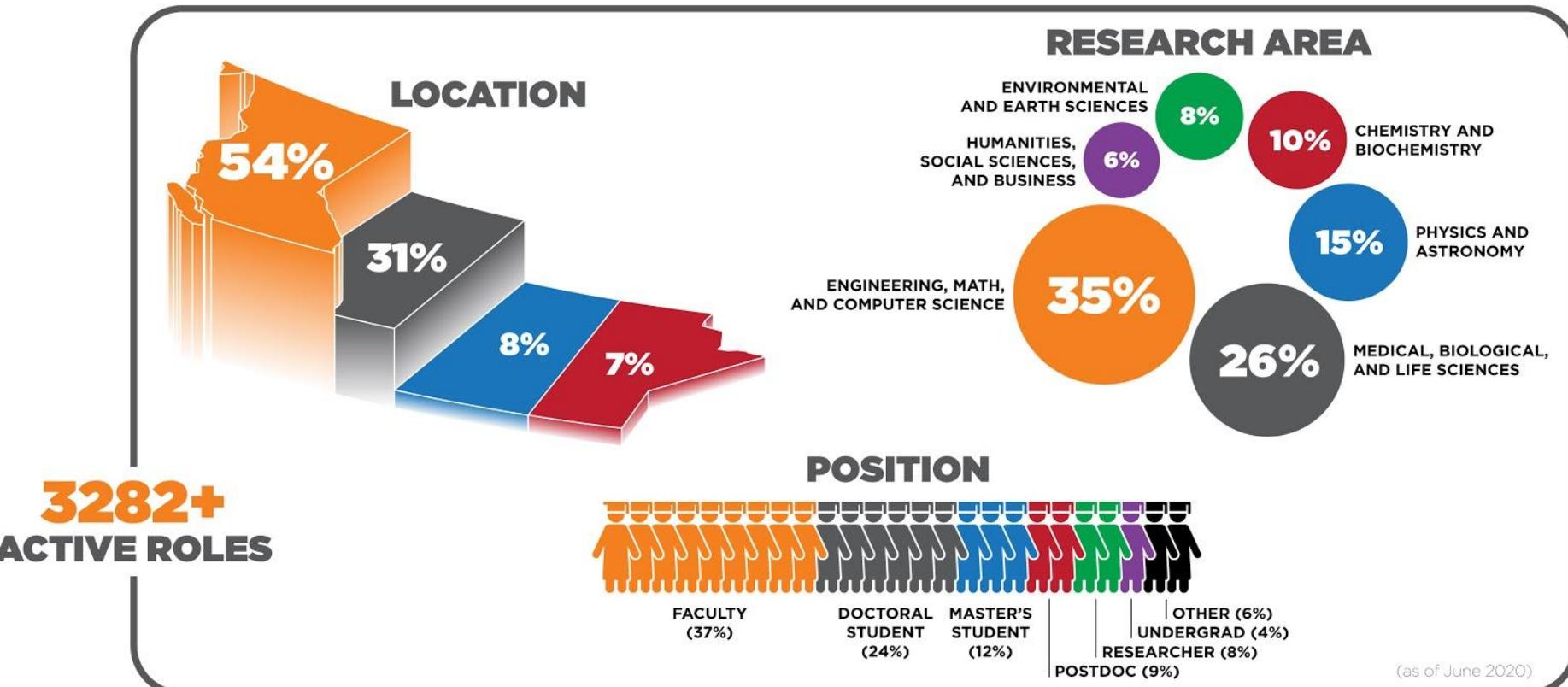


Cedar (Simon Fraser University)

- General purpose computing, multiple node types
- 94,528 CPU cores
- 14 petaflops peak performance
- 1,352 GPU devices
- 23 PB /project storage



WestGrid User Community: A snapshot





WestGrid Training, Outreach & Support

Co-hosting 2021 SciVis Contest w/ IEEE Vis (Launching Wednesday, Oct 28!)	International data visualization competition, seeking novel visualizations of a contributed dataset. https://scivis2021.netlify.app
2020 Online Autumn School (Starts next week - Monday, Nov 2!)	8 days, 8 different courses - focused on getting started / introductory topics. Registration still open! https://autumnschool.netlify.app
2020 Online Summer School (35 days, 16 courses + repeats)	May-July 2020: 300 registrants, 321 people on the waiting list, three repeat courses were open to everyone on the waitlist
Regional workshops (4-day series)	SFU (28ppl) in March 2020, UVic (35ppl) in December 2019
Community Town Halls	4 since January 2020
National Support	WestGrid: 19,434 tickets (Sep 1/2019 to Sep 1/2020), 35% of CC

Questions?

Want to host a training event at your institution?

training@westgrid.ca

<https://westgrid.github.io/trainingMaterials/courses>

2019 Training	Number	Attendees	Attendee-hours
In-person events	37	537	2,104
Online events	52	1775	4,550

Resource Allocation Competition (RAC)

Overview Changes from 2020 Resources for 2021

This presentation is a WestGrid-provided best practices guide for the Compute Canada Resource Allocation Competition (RAC). It does not replace the official RAC documents and PIs should carefully review the Compute Canada official documentation. In cases of conflict between these slides and the official RAC documents, the official RAC documents take precedence.

Resource Allocation Competitions (RAC): Compute, storage and cloud allocations on Compute Canada national systems.

~80% allocated through RAC competitive process.

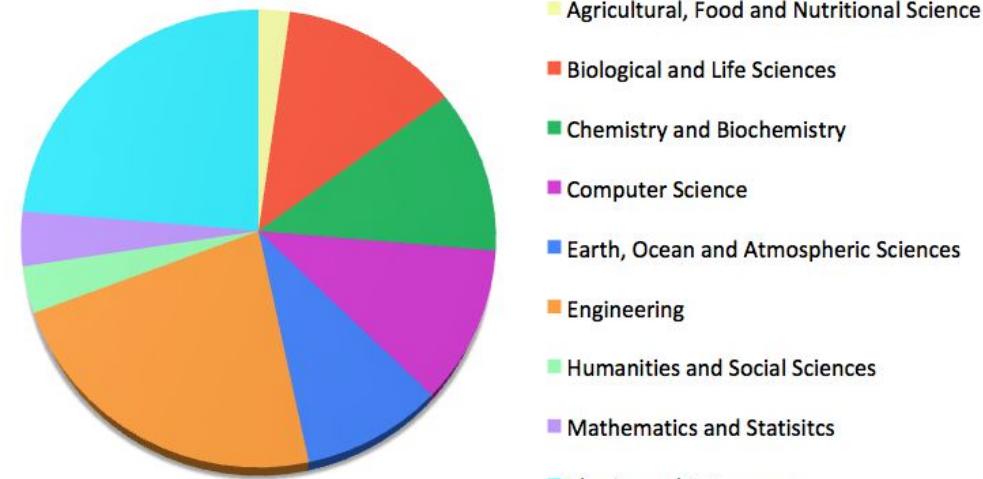
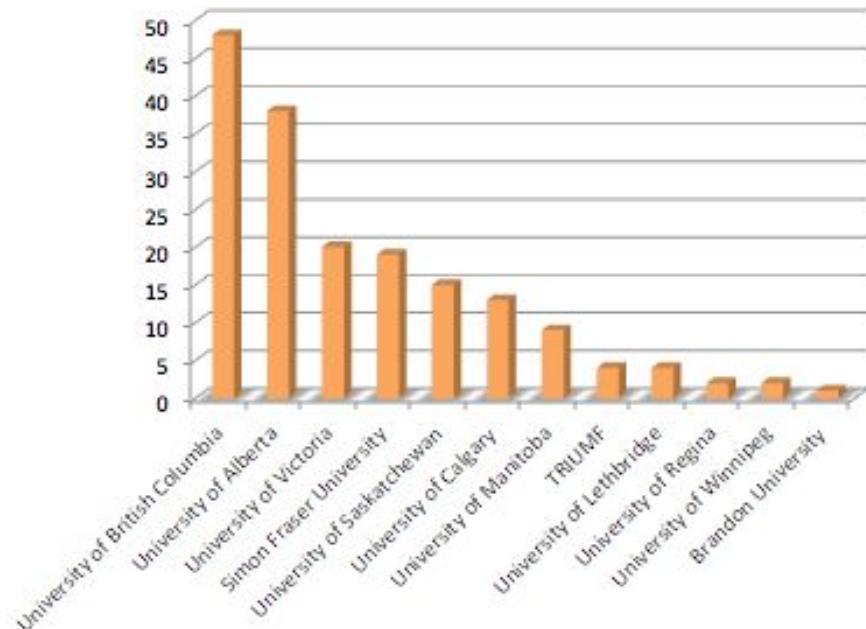
- Applicants must be eligible to hold a grant from a Canadian granting agency.

~20% of Compute Canada national resources are reserved for opportunistic use.

- Available to all CC users - Rapid Access Service (RAS).
- Any researcher / student at a Canadian institution can become a Compute Canada user.
- No application required beyond request for an account:
https://ccdb.computecanada.ca/account_application

RAC 2020 allocations in Western Canada

175 recipients from institutions based in Western Canada
(15% increase from 2019)

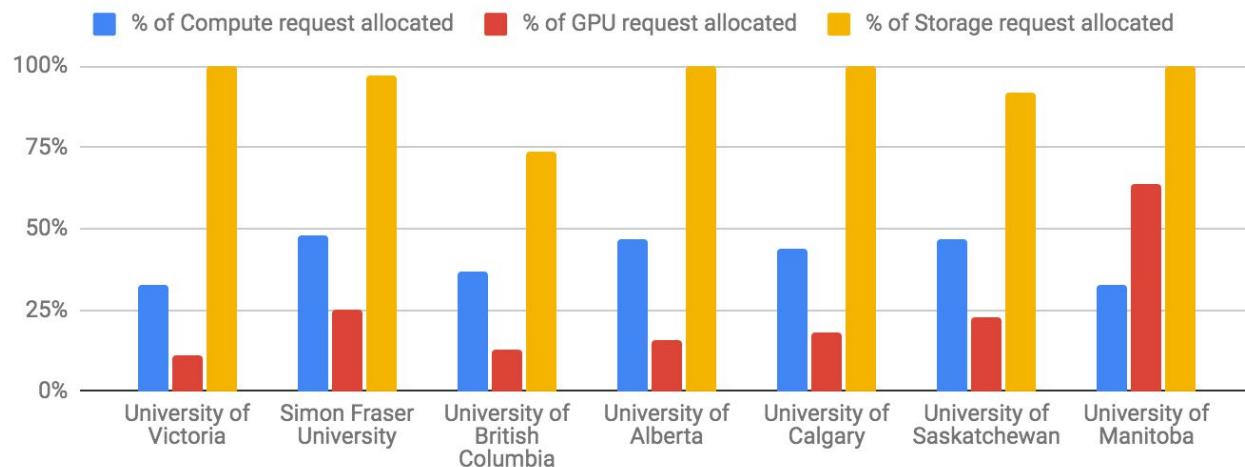


- Agricultural, Food and Nutritional Science
- Biological and Life Sciences
- Chemistry and Biochemistry
- Computer Science
- Earth, Ocean and Atmospheric Sciences
- Engineering
- Humanities and Social Sciences
- Mathematics and Statistics
- Physics and Astronomy

RAC 2020 requests - allocation %

On average, in 2020 RAC recipients at WestGrid Member institutions received 41% of total Compute requested, 24% of GPUs requested and 95% of total storage requested.

Percentage of request amounts allocated in RAC 2020



*On average across Canada,
WG RAC projects received:*

-  **40% of total Compute ask**
-  **26% of total GPUs ask**
-  **86% of total Storage ask**

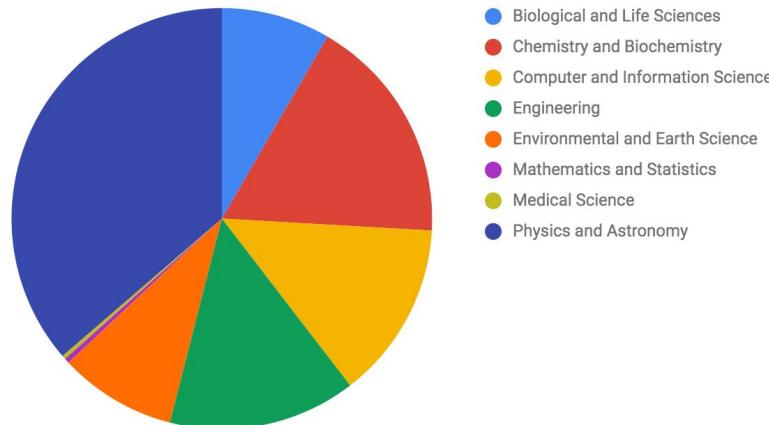
RAC 2020 allocation values

Total value of WG allocations:
\$11.8 million
(32% of total value)

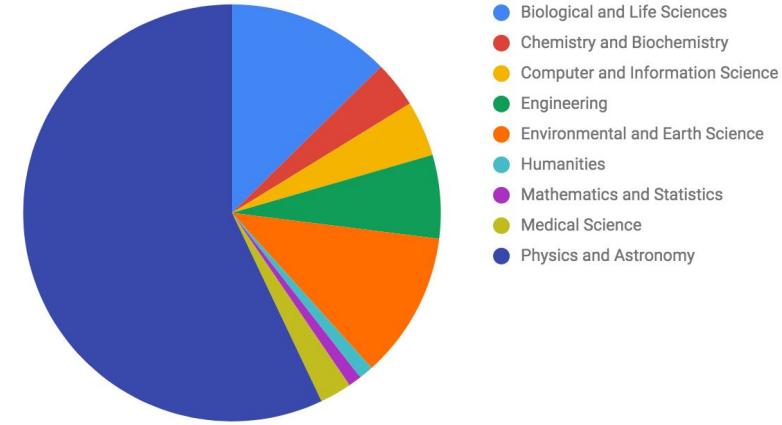
Total value of
 all allocations:
\$32.6 million

1 Core-year: \$121.34
 1 GPU-year: \$2,435.89
 1 TB project storage per year: \$54.96
 1 TB of nearline per year: \$25.66
 1 vCPU year: \$80.93
 1 TB Ceph (cloud) storage per year: \$117.70

Total Compute allocation (core years) by research area



Total Storage allocation (TB) by research area



WestGrid RAC 2020 Results

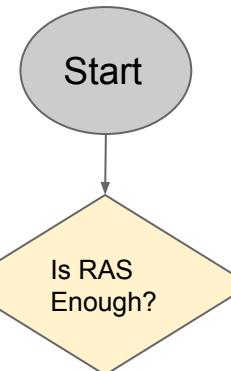
WG Applications received	186	32% of CCF total. CCF=590
Number of successful WG applications	181	97% success rate
WG RRG applications received	87	
WG RRG Fast Track	75	
RPP applications received or continuing	48	RPP 2018, RPP 2019, RPP 2020
Average science score out of 5	3.5	Range: 1.0->4.7. CCF average=3.5
Average CPU scaling	45%	CCF=45%
CPU allocations	62,366 CY	CCF=181,687
GPU allocation	669 GY	CCF=1,938
Storage Allocation	43 PB	CCF=100 PB

RAS vs RAC - What do you need?

RAS: Rapid Access Service

Default resources available to **any user** with a Compute Canada account

Start using RAS: Apply for a CC account



RAC: Resource Allocation Competition

Annual competitive process to allocate resources based on scientific excellence

Compute Services	Max RAS
Compute	50 Core Years
Storage	10 TB
GPU	10 GPU years

Use the RAS for tests and prototypes

- Learn about the systems (early-stage / new users)
- Run prototypes/test jobs, and if possible production jobs
- Acquire performance statistics
- Predict future requirements

Cloud Services	Max RAS
Compute	80 vCPUs
Persistent	25 vCPUs
General (block) storage	10 TB
Shared filesystem storage	10 TB
Object storage	10 TB

RAS and Default Details

Cluster Resource	Default (Automatic)	RAS (by request)	Comment
CPU		Opportunistic	At each site
GPU		Opportunistic	At each site
/project storage	1 TB	Up to 10 TB per group	At each site
/nearline storage	None	Up to 10 TB per group	At each site

See [Cloud RAS Allocations](#)

Cloud Resource	Compute Cloud	Persistent Cloud
VCPUS	80	25
Persistent Storage		10 TB
Public IPs		2

Competitions: RRG and RPP

<u>Research Platforms and Portals Competition</u>	RPP	Scientific gateways <ul style="list-style-type: none">• Provide service to a community of users• Datasets and Toolsets• Generally in the cloud (may include compute)• Multi-year (up to 3 years)
<u>Resources for Research Groups Competition</u>	RRG	Classic HPC with jobs submitted to the big clusters. <ul style="list-style-type: none">• Jobs submitted through a scheduler.• May also ask for cloud resources.• Special purpose only.• Annual - Fast-track for year-to-year continuity.

Documentation & competition information:

<https://www.computecanada.ca/research-portal/>

Go to “Accessing Resources” then “Resource Allocation Competitions”

Recommendation: Consult with support staff.

- support@westgrid.ca
- rac@computecanada.ca
- support@computecanada.ca

Priorities and Allocations

Our usual warning: we translate a *compute allocation* into a **scheduling priority**. This priority is defined such that on average with continuous job submission a user will be able to use their allocation over the year.

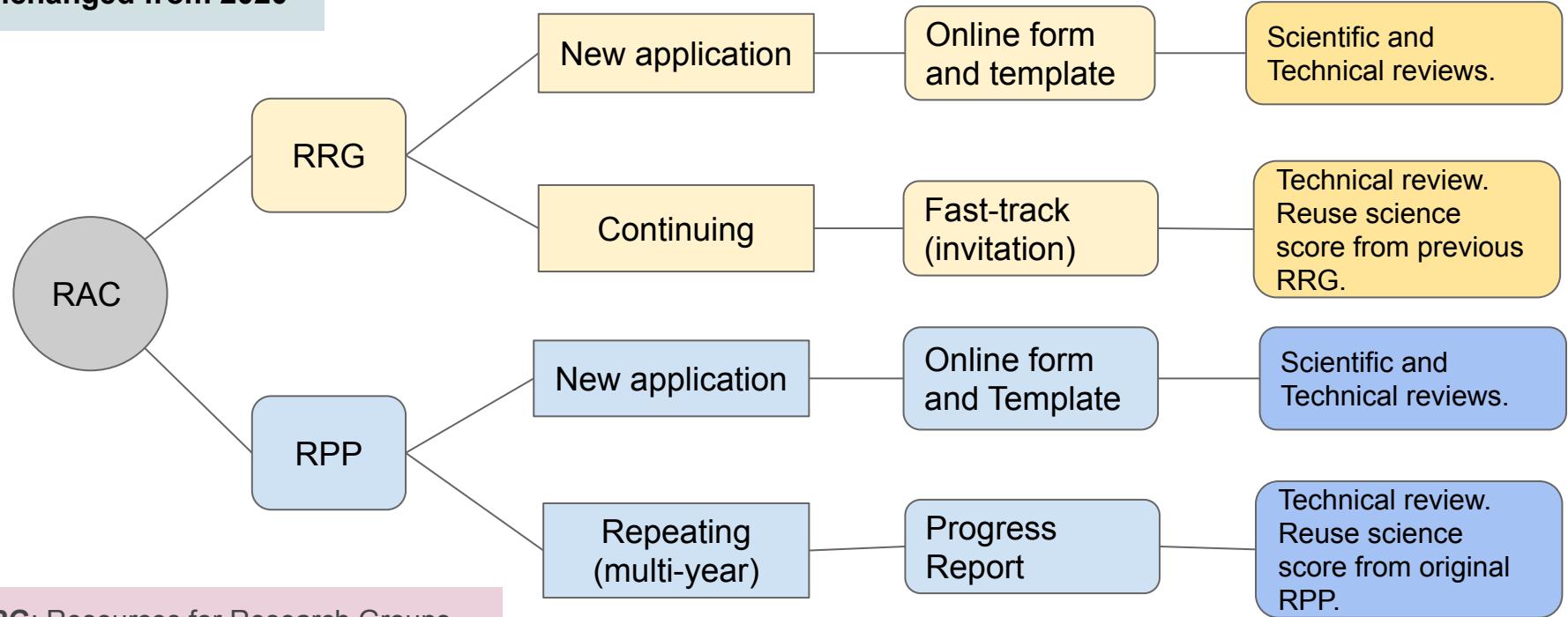
- Intermittent use will result in approximately pro-rated total usage.
- Allocations do not accumulate. **If you don't use it you lose it!**

RAS (default) compute is not an allocation. Priorities are evenly distributed across the resources remaining after RAC, so use is opportunistic.

- Jobs will run in the remaining resources after RAC.
- Small, short jobs!
- Long or large jobs may wait in the queue for significant periods.

Note:
Cloud (RPP) is different

Unchanged from 2020



RRG: Resources for Research Groups.

RPP: Research Platforms and Portals.



2021 RAC Key Dates

	Start	Finish
Fast Track submission (Invitations sent before Sept.24)	Sep 23, 2020	Oct 28, 2020
RRG & RPP full application submission	Sep 23, 2020	Nov 5, 2020* 11:59 PM EST
RPP Progress Report submission*	Nov 10 2020	Dec 11, 2020
Award letters sent*	Mid Mar 2021	
RAC 2020 Allocations implemented*	Mid Apr 2021	

*No extensions or appeal process!

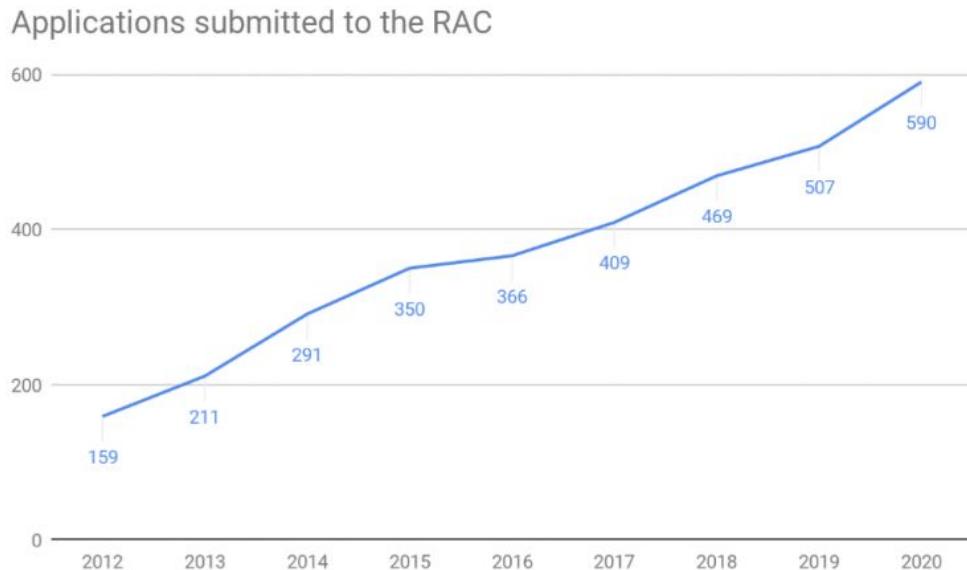
IMPORTANT: The program for special requests of ARC resources for **COVID-19 research** ends on **March 31, 2021**. If you received one of these allocations and you expect your need for resources will continue beyond April 2021, **you must submit an application to the 2021 Resource Allocation Competitions**.

Review of RAC 2020

2020 RAC had the highest number of applications in history: 590 projects 16% more than 2019!

For 2020, the RAC was able to award:

- 40% of total CPUs requested
- 86% of total storage requested
- 26% of total GPUs requested
- 99% of total virtual CPUs requested



2021 RAC will be another competitive year!

System	Cores	GPUs	/project Storage	
Cedar (GP)	94,528	1,352	23 PB	NVidia P100, V100 GPUs
Graham (GP)	41,548	536	12 PB	NVidia P100, V100 and T4 GPUs
Béluga (GP)	34,880	688	2.6 PB	NVidia V100 GPUs
Niagara (LP)	80,640	0	2 PB	RAS requires a request.
Arbutus (cloud)	17,272	0	5.7 PB	Default requires a request. Physical cores. Generally hyperthreaded.
GP Cloud partitions	Cloud partitions are available on GP systems for special purposes.			

Prediction: we'll have roughly 60% of CPU asks and 25% of GPU asks!
Very competitive!

All applications submitted to the RAC are peer-reviewed and scored. The final RAC score is based on the following:

- *the scientific excellence of the specific research project for which computational resources are being requested;*
- *the scientific and technical feasibility of the proposed research project;*
- *the appropriateness of the resources requested to achieve the project's objectives;*
- *the likelihood that the resources requested will be efficiently used.*

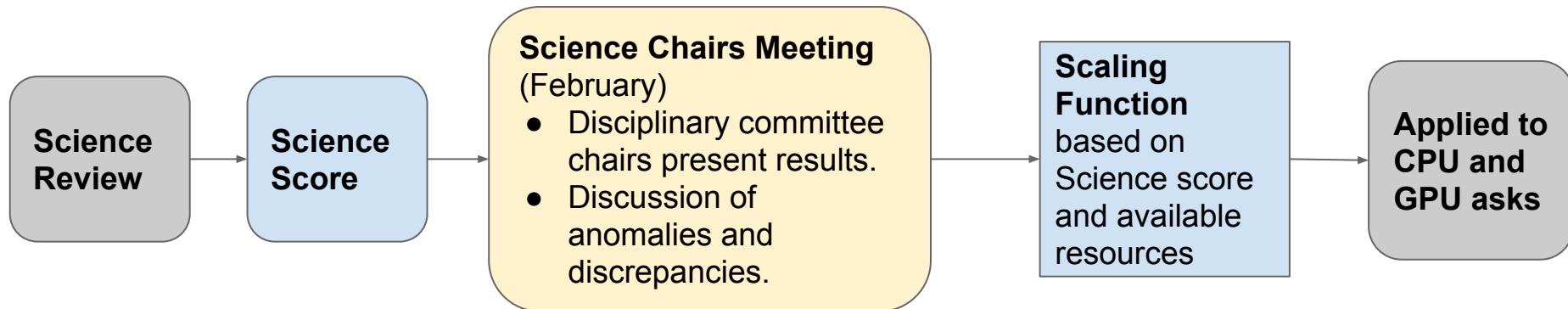
Every year a cutoff score is determined, below which no allocation is granted.

For RAC 2020, the minimum overall score required to receive an allocation was 2.5/5

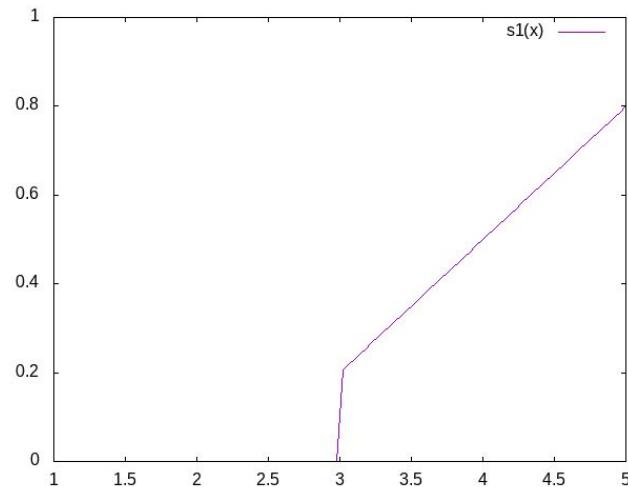
Excellent	4-5
Good	3-3.9
Fair	2-2.9

Scaling is applied to CPU and GPU requests.

- Not applied to persistent Cloud requests (web servers, database servers, ...)
- Not applied to storage



RAC 2019	Cutoff was a science score of 2.3/5
RAC 2020	Cutoff was a science score of 2.5/5
RAC 2021	Expect a slightly higher cutoff



*** Scaling is required due to insufficient resources, which makes the RAC
a very competitive process.***

Additional resources	Mix of GPUs: V100, T4, P100 Cloud services: Object storage and shared filesystem storage
New science review committees	Some committees have been split due to increased workload.
RPP and RRG Evaluation Criteria	Stronger focus on feasibility <ul style="list-style-type: none">Efficient computational methods, methodology, expertise of team, resource justification, .. Increased emphasis on funding sources <ul style="list-style-type: none">Section in the templates for funding details.
Decreased page limits	Generally 8 pages (decreased from 10)
No out-of-round	RAS is the only avenue for non-RAC requests in 2021

RAC 2021 Committees

2021 committees	2020 committees
<ol style="list-style-type: none">1. Astronomy, Astrophysics and Cosmology2. Bioinformatics3. Chemistry, Biochemistry and Biophysics4. Computer Sciences and Mathematics5. Engineering6. Environmental and Earth Sciences7. Humanities and Social Sciences8. Nano, Materials and Condensed Matter9. Neurosciences, Medical Imaging and Medical Physics10. Subatomic Physics, Nuclear Physics and Space Physics	<ol style="list-style-type: none">1. Astronomy and Subatomic Physics2. Bioinformatics, Neuroscience and Medical Imaging3. Engineering, Mathematical and Computer Sciences4. Chemistry, Biochemistry, Biophysics5. Environmental and Earth Sciences6. Humanities and Social Sciences7. Nano, Materials & Condensed Matter

About 90 reviewers!

Resource Allocation Competitions (RAC)

General Tips and Best Practices

This presentation is a WestGrid-provided best practices guide for the Compute Canada Resource Allocation Competition (RAC). It does not replace the official RAC documents and PIs should carefully review the Compute Canada official documentation. In cases of conflict between these slides and the official RAC documents, the official RAC documents take precedence.

Page Limits, PDF, 2nd Contact

Page limits are enforced

PDF only

Secondary Contact Person

PI's are encouraged to appoint a secondary contact person so that important communications don't go astray.

“I don’t understand why, but it seems necessary to continuously emphasize to applicants to closely follow the instructions, including page limits (people ignore this!) and the necessary components.”

No Appeal Process

There is **no appeal process**.

If the reviewers misunderstand a proposal (which does happen) there is no way of correcting or amplifying the proposal.

Need to get it right the first time.

The Reviewers

- Science reviewers are experts in their discipline, BUT - not necessarily experts in the specific sub-discipline or area of any particular project.
 - For details see list of committees and chairs on [CC RAC pages](#)
- Very wide range of RAC proposals do not allow for area experts. **You cannot assume that your proposal is going to be read by someone who works directly in the area or field of the project.**

Avoid lots of detailed jargon.

Write for a discipline expert, but not for the specific area.

Use the Templates!

We strongly recommend that PIs use the templates.

- [RRG Application Template](#)
- [RPP Application Template](#)

Summaries and details are in the following sections.

Always mentioned by reviewers.

Certainly feel free to customize within the templates, but follow the overall structure.

Justify your Request

- Address the **evaluation criteria** (details to follow)
- Provide details about the **research goals and objectives**.
- Justify **how the resources will be used** to accomplish or those goals & objectives.
- Citation rates of recent work help justify science.
 - There is a “References” section in the templates which does not count against the page limit.
- Poor proposals generally do not provide sufficient information, or have mixed technical requirements into the research justification.
 - Very difficult to decipher for both science and technical reviewers, and results in poor scores.

“Please improve motivation of why the proposed calculations are important, and what is to be learned and/or what other science depends on the results.”

Avoid ‘over asks’

*“People **should not over ask** to compensate for scaling as the committee is smart enough to know what is a reasonable ask for various types of calculations and applicants do not want to be penalized. The ask should fit with both the calculations to be done and the size of the research group.”*

*“Requests that were **obviously too large** - especially for new groups, or those trying to run new models, etc, or with newly starting students - in all cases, we know there is significant overhead in getting things working, starting students will take courses, etc., - so don't make it seem like you will be doing full production runs from day 1”*

- Not much to add to the comments!
- There have been very large asks which upon examination are really not very well justified. Committees have made recommendations for drastic cuts.
- Committees are aware of issues with applicants trying to game the system.

Provide Adequate Details

Clearly explain WHAT is planned, using specific details.

- **Use tables** to provide resource details by project. There are lots of recommended examples in the templates!
- If it is difficult to predict usage then emphasize the areas of uncertainty.

RRG template: Section
5.1 Resource Request
Summary

Be consistent with the
web forms!

Resource	Project 1	Project 2	Totals
Team Members	2	2	4
Core years	200	180	380
GPU years	10	10	20
....

“The technical justification did not show a calculation of the computing and storage needs. Providing a table with this information, as suggested in the guidelines, would have made this section stronger.”

Provide ALL the information asked for, but ONLY what is asked for.

- **Use the templates!**
- Answer the specific questions and provide only those details requested.
- Do not re-use submissions from other proposals or competitions.
 - **RAC focusses on computing.** Most other proposals are looking for funding!
 - Students and Postdocs can write sections (they are the experts) but the overall proposal needs an experienced, guiding hand.
- Be clear, avoid jargon - reviewers are not experts in all sub-domains.
- Take time to edit and review the application before submitting it.

"The proposal is very long. The very detailed scientific justification is more reminiscent of a NSERC proposal. For next year, I recommend to shorten the science part, and to work out more clearly the purpose of the CC usage, and the results that will be enabled by the CC Allocation."

Canadian Common CV (CCV)

Keep your CCV up to date!

[CCV Submission Guide](#)

1. CCV is used by the review committees - quality of the PI and his/her research.
Keep your bibliography up-to-date to ensure that reviewers are aware of the latest (and greatest!) work.
2. CC relies on the CCV for bibliographic analyses used both for annual reporting, and for funding proposals.
The Field Weighted Citation Index (FWCI) is a key component of the metrics presented by CC.

- Required for RRG and RPP full applications (update not required for Fast Tracks or progress reports).
- *“Committee members must not impart, refer to or consider information about the applicants that does not appear in the application and the provided Canadian Common CV (CCV).”*

“It is critical also to update the information on the CCV. (...) emphasize that this shows their contributions and gives weight to their proposal and that the expertise exist in order to succeed. Several times this year we saw CCVs that were inconsistent with what we know about applicants, but we have to go with what is submitted. This includes up to date publications and grants. Sometimes we have seen CCVs that are over a year old. This also simply does not look good to the reviewers.”

“Progress over the past year is missing. Based on the CCV, most of the group's publications were enabled through Compute Canada resources, so that information should have been accessible and would have been relevant to include.”

“We often missed full CVs of PIs and co-PIs, making it difficult to evaluate the PIs past contributions”

There should be a thread or narrative used to present a well-connected and justified story.

This has been emphasized by reviewers!



The team (HQP, RPP management, ..) supports the science, the technologies and the ask.

Discrepancies in the Asks

Very important: Ensure there are NO discrepancies between the resources requested in your technical justification document and the online application.

The application form takes precedence.

- Quite a few of these last year!
- Forms and descriptions have been updated to (hopefully) make things more clear
- **The form values are inserted automatically in the master spreadsheet.**
 - The master spreadsheet is the basis for the rest of the process!
- *Be careful - we have missed or mistaken the resource requests in the past.*

Resources for Research Groups (RRG)

Best Practices

This presentation is a WestGrid-provided best practices guide for the Compute Canada Resource Allocation Competition (RAC). It does not replace the official RAC documents and PIs should carefully review the Compute Canada official documentation. In cases of conflict between these slides and the official RAC documents, the official RAC documents take precedence.

Resources for Research Groups

- Aimed at job-based use of the big clusters.
- Simple cloud asks for computation are also available.

Criteria (Changed from RAC 2020)

1. **Research Methods (70%)**
 - a. Outline, Outcomes, Methodology, & technical justification
2. **Resource Management (30%)**
 - a. Funding, Capacity and expertise of the team, Management

Research Methods

70%

Research Outline

- Well-defined and clear research problem, with importance and benefits clearly evident.
- Well-defined and clear overall goals and objectives
- Well-defined and clear deliverables: aligned with the goals, and feasible.

Expected Outcomes

- Presents anticipated outcomes and the means by which they will be measured.
- Clearly describes the relevance and impact of the outputs and results.
- Likely to lead to advances in the research area.

Methodology

- Appropriate tools, methods and approaches.

Progress over Past Year

- Evidence of progress with emphasis on utilization of CCF resources

Resource Justification

- Well-justified resource requests.
- Reasonable for achieving the project objectives.
- Estimates of number of runs (“experiments”) and approx request per run

Resource Management	30%	Funding <ul style="list-style-type: none">Funding is available for the research project to justify the request for computational resources. Computational expertise of the Team <ul style="list-style-type: none">Includes PI, Co-PI(s) and HQP. Research and Computational Expertise <ul style="list-style-type: none">Expertise is sufficient to deliver the project. Management Strategy <ul style="list-style-type: none">A “solid” management structureRoles and responsibilities are clearly identified.
----------------------------	-----	---

RRG Template - Outline

	Section	Description
1	Research Outline (1 page)	Outline the research problem for which Compute Canada resources are being requested. Its importance/relevance as well as your general objectives.
2	Expected Outcomes	Outcomes and results, how they will support the objectives and lead to advances in the research area.
3	Methodology	Describe the tools, methods, and approaches.
4	Progress over past year	Outcomes and impacts. If possible, also include performance or success metrics. Highlight any notable RAC-enabled research. <i>In addition to CCV.</i> <ul style="list-style-type: none"> • opportunity to highlight any particular contributions, and also to add anything new
5	Resource Request Justification	Technical details of your computational and/or storage needs.
6	Funding	Table listing funding sources.
7	Computational Expertise	A table of required expertises and the team members bringing suitable expertise.
8	Management strategy	A table showing the team members with their positions and resource requirements. Extra information as applicable showing the relationship between sub-projects, team leads, resource use, etc.

1. Research Outline

Description of the research problem.

1 page

- Previously this was a more comprehensive “Research Plan” but is now focussed on goals, objectives and deliverables

“Made clear the objectives and big picture science questions, even for non-specialists in their sub-genre”

“Both the science and the details (at appropriate level) are important - make sure they are both there and balanced”

2. Expected Outcomes

Outcomes and Outputs.

As usual table(s) are a nice clean way of providing this kind of information.

½ Page

Actual measurable performance or success metrics are useful, but probably more germane to platforms and portals.

Outcome	Comments
Prototype sample analysis pipeline	Sample ingestion to mass spec and chromatography. Results stored and translated to standard format. Application of analysis module Identification of biomarkers.
ML analysis	Develop ML analysis module from above results.
Thesis defence	Jane Doe thesis built on the ML analysis.
...	

3. Methodology

The usual table(s) describing the methods and tools/applications/..

½ page

Hard to find the right balance: not too technical, but enough for the reviewers to understand the basics.

½ page recommended, so not much room for details.

- A single table would be useful.

Comments about the suitability of the methods are always good. We want to ensure you know what you're doing, and have put some thought into choosing good, efficient approaches.

4. Progress over the last year

Most projects are continuing. So important to emphasize progress:

½ Page

- Conferences
- Papers, theses
- Breakthroughs
- Solution of particularly difficult programming or operational problems.
- Nobel prizes
- ...

This is a great opportunity to highlight important progress during the previous year.

- Criteria specifically mention use of CCF resources.

“Progress over the past year was not mentioned. This is key for sure and goes along with what I said above about having an up-to-date CV. It is critical to show the group has the expertise and status in the field to achieve the research goals.”

5. Resource Request Justification

The critical section!

Carefully justify the resource ask.

3 ½ pages

5.1 Resource Request Summary	Table summarizing each resource (CPU, GPU, Storage, ..) Ensure it's consistent with the CCDB forms. Nice to include a column with short comments explaining the need.
5.2 Compute Requests	5.2.1 System Selection 5.2.2 Size of Compute Request 5.2.3 Memory requirements 5.2.4 Code Details 5.2.5 Code Performance and Utilization
5.3 Storage requests	5.3.1 Details 5.3.2 Performance and Utilization 5.3.3 Size of request

Some room for customization to highlight the particular requirements of your project.

5.2.2. Compute Request

Explain the projects being worked on and the required resources.

- Tabulate the number and size of the needed runs/jobs/virtual machines
 - Again **tables** and breakdowns really add to the proposal.
 - Sub-tasks or sub-projects, phases, team members, ..
- If it is difficult to predict usage then emphasize the areas of uncertainty.
 - early research issues where details are still to be worked out
 - administrative issues like onboarding graduate students or postdocs who have not yet worked out a detailed research plan.

“Clear use of tables, that linked people/projects and computational needs”

“Clear explanations of computational needs and costs, with details, yet not getting bogged down in the very (too) technical”

5.2.2 Compute Request

Recommended table in the template (there's an example spreadsheet)

Type	Number of jobs	Duration of job (hrs)	Cores per job	Memory per core (GB)	Total Core-Year Requested
Development	10,000	22	8	2	200.78
Production	2	768	1024	8	179.43
				Total	380

- **Tough to estimate**, but hopefully you have previous experience or have run tests.
- Breakdown into small/medium/large can be useful.
 - Maybe development requires many small jobs? With a few major production jobs afterwards?
- We realize these are sometimes extreme guesstimates.
 - *We don't use these for anything during the actual RAC year.*
- But does show the reviewers that there is some understanding of the resources required.

5.2.3 Memory Requirements

- **Memory is a limited resource.**
- We're seeing increased numbers of big mem jobs.
- We have a few “fat” nodes but not a lot.
- Note that the scheduler uses “core equivalents” in “charging” for jobs.
 - ie) if a 1 core job requires all the memory on a 40 core node then the scheduler charges for 40 cores and your priority is decreased accordingly for the build-up period (usually a week).
 - https://docs.computecanada.ca/wiki/Allocations_and_resource_scheduling

So please remember to include a table which summarizes your memory use.

- And as usual this emphasizes your understanding of the methods and apps you use. Good for the reviewers!

5.2.4 Code Details

Provide details about the codes.

- name
- key reference publication
- essential numerical methods used
- serial/parallel and the type of parallelism

Again shows familiarity and experience with your applications and codes.

Note: separate section (next) for performance description.

5.2.5 Performance Estimates

Provide performance estimates.

- RAS - default resources are available for testing and prototyping.
- **Performance tables or charts** really help the reviewers, and make the proposal look good!
- Particularly important for applications which are scaling up.
 - Parallel performance? Scaling graphs for test runs? ...
- Don't forget storage
 - **Lots of i/o is always an issue in HPC:**
 - the shared filesystems are high performance but it's never enough.
 - Scratch is fast, but local disks are faster (and more difficult to distribute)

5.3 Resource Justification: Storage

Table of storage types and justifications.

Projects can be very unique so feel free to customize, but keep the table format if all possible. Can add amplifying comments afterwards as necessary.

Use	Type	Amount	Need
Database	/project	100 TB	Needs to be online for access by continuing analysis jobs
Generated results	/project	10 TB	Online cache for application of machine learning training. Training jobs cycle through as data is produced.
Long-term (Tape)	/nearline	200 TB	Offline for archived storage of all generated results.

Performance details show expertise and familiarity:

- Bandwidth, IOPS.
- Very large numbers of files or particularly impressive i/o requirements (*talk to CC support!*)

Special considerations can also be helpful:

- Bursty or particularly variable use.
- Special backup, security or availability requirements (*talk to CC support!*)

6. Funding Available

New this year. Requested by reviewers.

- Gives some idea that the science has already been reviewed.
- And that there are resources available for the project.

If no funding then provide some explanation of the plan

- ie) early-stage researchers may be applying for funds.

Project	Funding Source	Grant	Grant Expiry	Award	Portion allocated to this project
Project 1	NSERC	Discovery grant	07-30-2023	C\$250,000	C\$50,000
...					

7. Expertise of the Team

Expertise	Description of Providers
Machine Learning	New Postdoc position currently being advertised
Parallel programming for CFD finite element methods	3rd year PhD student developing FEM codes.
Job submission optimization for large groups of jobs	Provided by regional experts
...	...

The idea is to show a reasonable, mid-level breakdown of the **expertise required**, showing a clear understanding of the **project requirements and a plan for sufficient staffing**.

- Expertise providers may include external resources like the local IT services, or departmental staff.

Note: The Management Plan (next section), should include a more detailed breakdown that describes each team member, their contribution to the project, and the related resource requirements.

8. Management Strategy

Aimed at big projects with big teams: “*Required when the Team includes more than the PI*”

Name	Position	Core-years	GPU years	/project
Jane Smith	Post-doc, Machine Learning and Tensorflow	1,000	10,000	100 TB
John Smith	Ph.D Student. CFD simulations	2,000	0	0
...				

For **complex projects**, it can be useful to include additional details, such as:

- Change Management approach
- Overview of a Governance model

“proposals with multiple PIs that didn’t link the group members were a problem. Both the individual parts and the integration between parts of a project should be explained.”

Resource Links

- **Ask a question:**
 - About RAC: rac@computeCanada.ca
 - About CCV: ccv@computeCanada.ca
- **Find an answer:**
 - [General RAC information](#)
 - [RRG Guide](#)
 - [RPP Guide](#)
 - [Technical Glossary](#)
 - [Frequently Asked Questions](#)
 - [CCV Guide](#)

Contact us anytime:

support@westgrid.ca

www.westgrid.ca

docs.computeCanada.ca

www.computeCanada.ca

Questions?

GOOD LUCK!

Questions?

Discussion?

Research Platforms and Portals (RPP)

Competition Best Practices

This presentation is a WestGrid-provided best practices guide for the Compute Canada Resource Allocation Competition (RAC). It does not replace the official RAC documents and PIs should carefully review the Compute Canada official documentation. In cases of conflict between these slides and the official RAC documents, the official RAC documents take precedence.

RPP Focus on Gateways

The Research Platforms and Portals (RPP) Competition enables communities to develop research projects that improve access to shared datasets, enhance existing online research tools and facilities, or advance national or international research collaborations.

Emphasis is on **creating scientific gateways which provide services to a community** of users.

Pure compute and storage requests should be in the RRG competition.
Continuing **multi-year asks should make use of the fast-track process.**

Generally we expect RPP's to primarily require cloud resources.

"A Research Platform or Portal is a set of community-developed tools, applications, and data that are integrated via a gateway or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a specific community. Platforms and portals enable entire communities of users associated with a common discipline to use national resources through a common interface."

Process

1. Notices sent in November, with reports due in January.
2. Progress report is reviewed by CCF staff.
3. Science score of the original RPP application is inserted into the normal allocation process, together with any requested updates to the original resource request.
4. **All new and continuing platforms and portals are then allocated through the normal process.**
 - a. The current year's scaling is applied!
 - b. **So no guarantee of an allocation identical to the previous year.**
 - CC strives to keep the allocations consistent from year-to-year

<https://www.computecanada.ca/research-portal/progressreport/>

If the RPP annual report is not received by the deadline, *then the allocation for the current year will not be renewed.*

WestGrid RPP Support

Reminder: we have a small dev team who can help with RPP development. Feel free to contact us.



support@westgrid.ca

The Research Platforms and Portals (RPP) Competition .. enables **communities** to develop research projects that improve access to **shared datasets**, enhance existing online **research tools** and facilities, or advance national or international **research collaborations**.

There have been quite enthusiastic ideas with big asks, but in practice the user community is quite specialized and uptake is dependent on the user interfaces and the services offered. Such asks are critically reviewed and the allocation may be decreased.

In addition to being a useful gateway with a user community the project team must have the skills and ability to **develop, operate and manage the gateway**.

Pure compute and storage requests must now be in the RRG competition.
Continuing multi-year asks should make use of the fast-track process.

1. Research community provided with datasets and toolsets.
 - a. who / what is the user community?
 - b. what is being provided?
2. Development and Management
 - a. Is the team capable of developing and managing a major platform/portal system in the cloud?
 - b. Are the requested resources reasonable for the predicted user community?

The current cloud resources are **IaaS** (“Infrastructure as a Service”) resources. So it is completely up to the project team to design and implement a suitable architecture.

Project Justification (50%)	<p>Project Description, objectives and impact</p> <ul style="list-style-type: none">• Problem or need is clearly presented• Objectives and goals clearly described <p>Use</p> <ul style="list-style-type: none">• Value of the platform for targeted communities <p>Outcomes</p> <ul style="list-style-type: none">• Clear timeline for delivery of outcomes <p>Progress over last year</p> <ul style="list-style-type: none">• Achievements, outcomes or evidence of progress particularly in previous use of CCF resources <p>Resource Request Justification</p>
Resource Management (50%)	<p>Funding is available to support the team and the project.</p> <p>Team has sufficient expertise.</p> <p>Management strategy for the portal.</p> <ul style="list-style-type: none">• Provide access to a community and can manage users, technologies, data sets, etc.

The [RPP Application Template \(Word\)](#) is quite comprehensive. Have a look!

1 Project Justification 50%, 6 pages	<ol style="list-style-type: none">Project description, objectives and goalsUse of the Platform/PortalExpected OutcomesProgress over the Past YearResource Justification (with subsections)
2 Resource Management 50%, 6 pages	<ol style="list-style-type: none">FundingTeam Configuration and ExpertiseManagement Strategy

1. Project Description, Objectives & Goals

Basic description of the platform or portal and what it will do.

- The problem or need that each RPP will address.
- The general objectives and specific goal(s) of the RPP.

2. Use of the platform/portal

1. Identify the **audience/community**.
 - a. Who would be interested in using the proposed platform or portal?
 - b. Why are they interested?
 - c. Is the community national in scope? International?
2. What is the **size of the community**?
 - a. justify any such estimates.
3. What kinds of research would you expect the community to carry out?
 - a. examples of exciting projects that would make use of the portal/platform.
4. Are there any **agreements** in place that would put conditions on the request?
 - a. For instance the Atlas High-Energy physics project is part of the Large Hadron Collider collaboration, and must satisfy international agreements.

3. Expected Outcomes

As usual table(s) are a nice clean way of providing this kind of information.

- *Timelines for delivery over the 3 years of a typical RPP*
- *List of outcomes/deliverables/milestones*
- *Metrics or performance indicators*

Year 1 (2021-2022)	Year 2 (2022-2023)	Year 3 (2023-2024)
Assemble team	Dataset ingestion layer	Analytics engine
Architecture and Proof of concept	Basic searchable interface	Community development
KPI development		

Gantt chart always looks good!

RPP Phased Plans

We encourage phased plans, with for instance performance indicators showing a well-justified and strong understanding of the issues involved with not only technical development, but also governance, marketing and communications.

If possible define performance and success metrics. It's always useful to have a nice table defining your Key Performance Indicators (KPIs)

Especially startups - lots of large, enthusiastic but unjustified estimates.

- Start small.
- Briefly address marketing and communications.

Development plan

- Assembling the team, Development practices, Gantt chart, ...
- Architecture, including monitoring and reporting.

4. Progress over the last year

Most projects are continuing. So important to emphasize progress:

- Is there an existing user community?
 - How big is it?
 - What have they accomplished?
- Conferences, papers, theses, breakthroughs
- Solution of particularly difficult development or operational problems.
- ...

This is a great opportunity to highlight important progress during the previous year.

- Criteria specifically mention use of CCF resources.

"Progress over the past year was not mentioned. This is key for sure and goes along with what I said above about having an up-to-date CV. It is critical to show the group has the expertise and status in the field to achieve the research goals."

5.1 RPP Resource Justification

Usual table but generally focussed on cloud resources.

Resource Type	Year 1	Year 2	Year 3
Compute Cloud VCPU	100	200	500
Persistent Cloud VCPU	10	15	20
Volume and snapshot storage	10 TB	10 TB	5 TB
Object storage	Not used	5 TB	50 TB
Shared filesystem storage	Not used	Not user	
Additional cluster resources (if required)			

Consistent with the forms!

Note: total requirements, not additional on top of existing usage.

5.2 Cloud Needs

Details of the cloud ask in the summary. Tables are good!

- Feel free to customize the descriptions and table formats depending on the specific requirements

Use	Instances	VMs	Storage	Network
Web front-end	2	2-core VM's with 4 GB	100 GB persistent for temporary data	2 x external IPs
Database server	4	16 core VM's with 60 GB	1 TB local	internal

5.3 Storage and Compute Resources

This is aimed more at additional cluster resources

- ie) if jobs are to be hived off to clusters.

Type	Number of jobs	Duration of job (hrs)	Cores per job	Memory per core (GB)	Total Core-Year Requested
Development	10,000	22	8	2	200.78
Production	2	768	1024	8	179.43
				Total	380

- **Tough to estimate**, but hopefully you have previous experience or have run tests.

5.4 Choice of Specific Resources

5.5 Description of high-demand period

5.4 Choice of specific resources

- Any preferred locations?

5.5 High demand periods

- Do you expect any particular periods of high demand?
- How will the portal react?
- ie) load balancing, kubernetes, ...

Not required: may not be relevant to the particular project.

6. Funding Available

New this year. Requested by reviewers.

- Gives some idea that the science has already been reviewed.
- And that there are resources available for the project.

If no funding then provide some explanation of the plan

- ie) early-stage researchers may be applying for funds.

Funding Source	Grant	Grant Expiry	Award	Portion allocated to this project
NSERC	Discovery grant	07-30-2023	C\$250,000	C\$50,000
...				

7. Team Configuration and Expertise

Can the team do the job?

- Development, Operations and Management

Name	Position	Allocation to Project	Description
Jane Doe	Sr Architect	100%	Dev team lead and architecture
Name 2	Asst Prof	50%	Science Lead
Name 3	Jr Developer	50%	Front-end Javascript developer Provided by CANARIE-funded institutional resources.
...

8. Management Strategy

How will the project be managed?

- Project lead?
- Development lead?
- Operations manager?
- Communications lead? (outreach to the user community)
- Science advisory council?
- Change management procedure?
- User support?

Smaller projects may manage this through a single PI, but large projects may have significant management requirements.

Resource Links

- **Ask a question:**
 - About RAC: rac@computeCanada.ca
 - About CCV: ccv@computeCanada.ca
- **Find an answer:**
 - [General RAC information](#)
 - [RPP Guide](#)
 - [RRG Guide](#)
 - [Technical Glossary](#)
 - [Frequently Asked Questions](#)
 - [CCV Guide](#)

Contact us anytime:

support@westgrid.ca

www.westgrid.ca

docs.computeCanada.ca

www.computeCanada.ca

Questions?

GOOD LUCK!

Questions?

Discussion?



Extra Slides

COVID Requests

Summer 2020: special program for COVID requests

- This will end with the RAC 2020 year (March 30, 2021)

Continuing and new COVID requests will now be considered in the normal RAC 2021 process.

- Include a sentence or two in the description if the project involves COVID research

If you are reasonably happy with your 2020 allocation, then Fast Track preferred. Very straightforward and requires no further work on your part!

- Invitations sent Sept.23. Response required by Oct. 28.
- **Fast Track applications cannot be delegated:**
 - Pls must complete the application themselves.
- Proposals with **score < 3.0** from previous year have not been invited.
 - That was really too low to get anything reasonable so we recommend a proposal re-write.

Fast Track Process: Score from previous year is inserted into this year's scaling and allocation decisions.

- **No guarantee that your allocation will remain the same.**
(But generally it's pretty similar)

No Out-of-Round for 2021

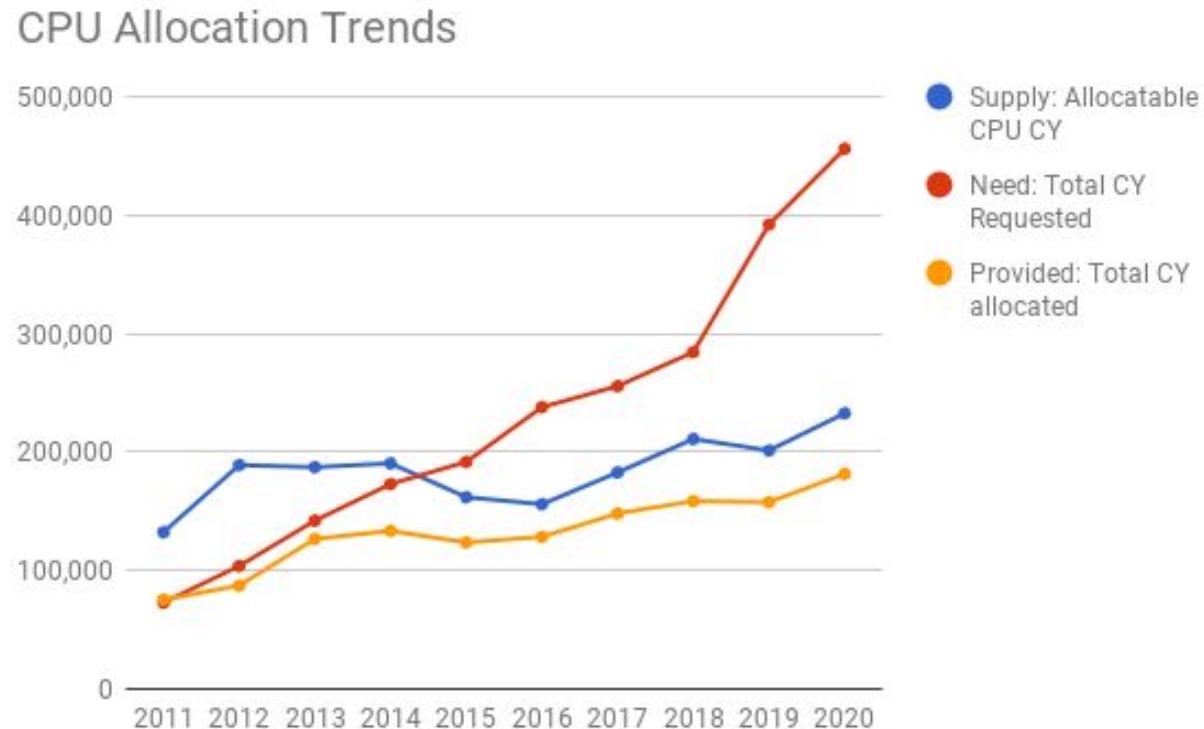
- RAS is the only avenue for non-RAC
- CC did create a COVID process summer 2020.

Use the RAS for tests and prototypes

1. Early-stage or new users get an account and learn about the systems.
2. Run prototypes or test jobs, and if possible production jobs
3. Acquire performance statistics
4. Predict future requirements
5. Create a well-justified RAC application. (out-of-round or regular)

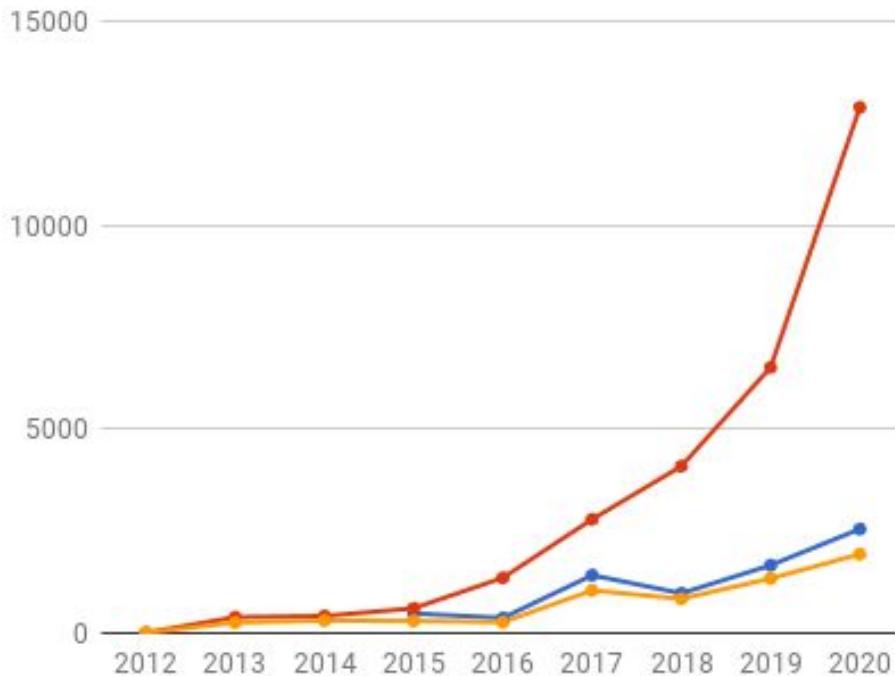
CPU Allocation Trends

As usual, we are not able to meet demand by a very wide margin.



GPU Allocation Trends

GPU Allocation Trends



- Supply: Allocatable GPUs
- Need: GPUs requested
- Provided: Total GPUs allocated

GPUs are
in high
demand!

Year	Supply	Need	Allocated	% of need awarded
2020	2,552	12,885	1,936	26%*
2019	1,644	6,555	1,331	20.3%
2018	976	4,092	840	20.5%
2017	1,420	2,790	1,047	37.5%

GPUs and Machine Learning

- **Major increase in GPU asks!**
 - Machine Learning is the major use
 - Increasing demand for production training
 - Current exponential increase in demand - expected to continue!
- **A mix of GPU resources**
 - Some users are not specifying which ones. The default is generally the V100's so P100's can go unused.

Graham	V100 (72) P100 (320) T4 (144) deep learning
Cedar	P100 (584) V100 (768)
Béluga	V100 (688)
Cloud	Virtual GPU (preliminary service)

GPU requests will require very strong justifications.

Storage Allocations

Second year that we have not had enough storage!

- But allocations are significantly underutilized and we have added storage so not too bad.
 - Cedar: 84% of allocation used but this year capacity increased from 14 PB to 21 PB.
 - Usage: currently ~50% of total available

Type	2020 Supply (TB)	2020 Need (TB)	2020 Allocated (TB)	Allocated / Need	
Project	57,030	56,867	40,732	72%	Project-based, backed up
dCache	14,804	10,200	10,200	100%	Special projects (LHC, ..)
Cloud	4,280	3,920	3,062	78%	Platforms and Portals
Nearline	67,800	43,347	44,650	103%	Tape - 2 replicas
Total	143,914	114,334	98,645	86%	

<https://autumnschool.netlify.app>

Hands-on introductions to essential tools & skills:

- four consecutive Mondays and Tuesdays: November 2-3, 9-10, 16-17, 23-24
- 8 days = 8 courses: Bash command line, Version control with Git, HPC, Basics of Python, Deep learning with PyTorch, Parallel programming in Chapel, Scientific programming in Julia, 3D scientific visualization
- more introductory/streamlined materials compared to the summer school
- open to participants from all institutions
- register once, attend any course(s) you like - registration now open

\$20-\$30 / person
Registration now open!



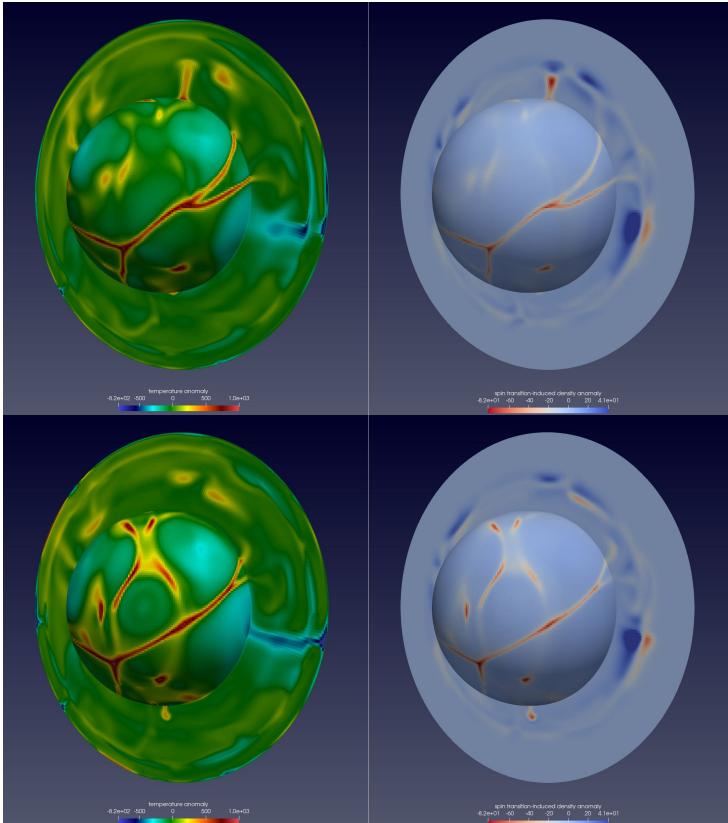
<https://autumnschool.netlify.app>

Visualization Competition

Co-hosting 2021
SciVis Contest with
IEEE Vis

Earth's Mantle Convection

- **Announcement:** late October 2020
- **Submission deadline:** July 31, 2021
- **Winning team notification:** September 10, 2021



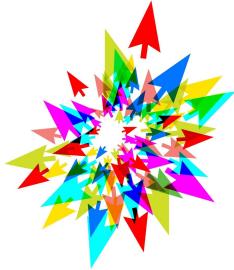
<https://scivis2021.netlify.app>

- Dataset from the U of Toronto
- Simulation conducted on Niagara
- Covers 500 Myrs geological time

In this Contest, participants will create novel approaches and state-of-the-art visualizations to assist domain scientists to better understand the Earth's mantle convection processes:

- Visualize stagnated or diverted cold slabs (descending mantle material) and hot plumes (rising mantle material) at 660 km and 1600 km depths
- Visualize correlations between the physical variables and the flow patterns

National vs. regional roles



compute | calcul
canada | canada



Compute Canada:

- National-scale initiatives:
 - Resource Allocation Competition (RAC)
 - Account renewals, security program
 - Procurement, technology planning (community consultations)
 - Securing funding, advocacy, nat'l/int'l partnerships
- National services
 - The large infrastructure systems: GP, LP, Cloud
 - Research Data Management, Globus, etc.

Regions:

- Local support & infrastructure operations
- User training and support
- Local / regional partnerships
- Addressing institutional & provincial priorities

WestGrid Members

WestGrid is one of Compute Canada's largest regions, spanning four provinces (BC, AB, SK, MB).

Currently, WestGrid has seven Member Institutions.



University
of Victoria



UNIVERSITY OF
ALBERTA



UNIVERSITY OF
CALGARY



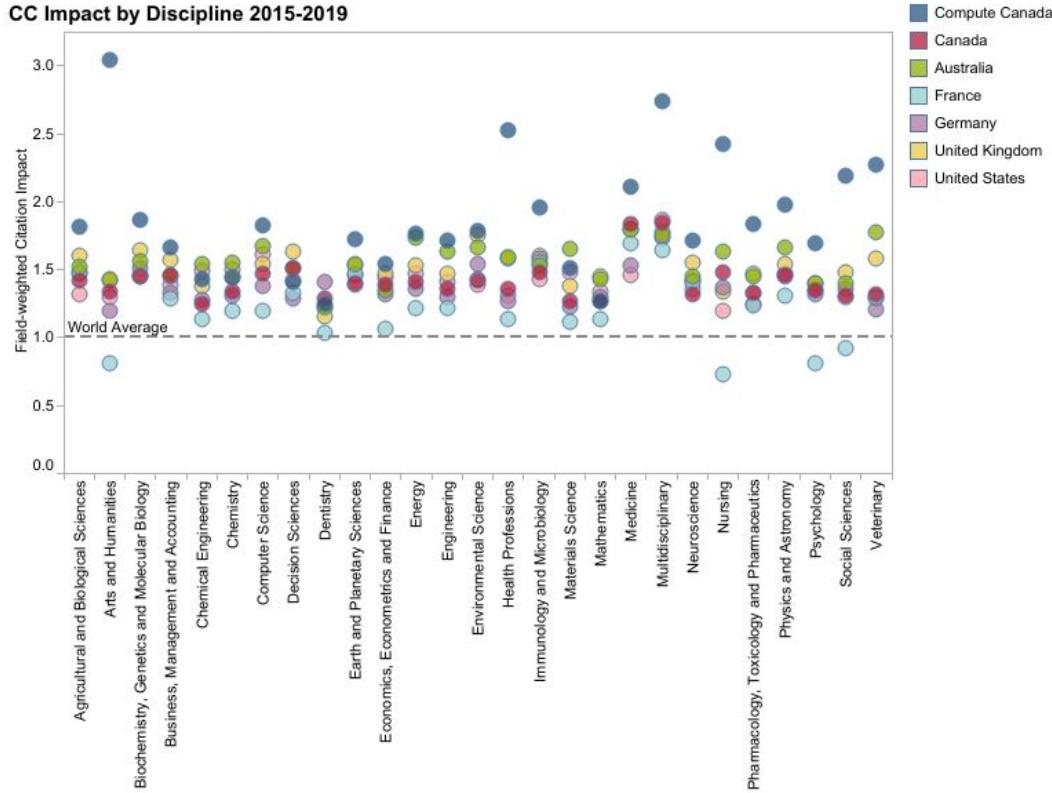
UNIVERSITY OF
SASKATCHEWAN



UNIVERSITY
OF MANITOBA

Impact - FWCI

CC Impact by Discipline 2015-2019



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

- Papers which use CC resources are in blue.
- Canada (red) is well above the world average, and overall comparable to a sample of western countries.
- **CC-enabled papers (blue) are well above in almost all areas!**

Compute Canada Federation

Canada's National provider of shared essential Digital Research Infrastructure (DRI)

- Not-for-profit corporation. 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.
- Provides shared services to More than 17,000 researchers across Canada. No fees.
Large requests based on a merit-based access system.



2. Use of the Platform/Portal

- **What is the user community?**
 - a. Useful to provide estimates of the size and activity.
- **What is the value of the platform/portal?**
 - a. What value does the RPP provide to the community?
 - b. What's new or not available elsewhere?

Only ½ page

If applicable - Describe the level of interaction between Canadian and international research groups and platforms.

1. Carefully describe the server architecture.
 - a. Feeds into the resources ask.
2. Describe any security/privacy concerns or requirements
 - a. steps that will be taken to satisfy such requirements.
3. Is the architecture scalable?
 - a. Refer to the size of the community to justify any scalability issues or plans.
4. Demonstrate the expertise and capability of your staff.
 - a. A portal or platform requires system management so identify the staff resources necessary to both develop and operationally manage your proposed system.
5. Special requirements?
 - a. Cloud or tape based backups
 - b. Particularly large memory requirements
 - c. Performance or response requirements
 - d. Redundancy and Reliability requirements. For instance multi-site architectures.
 - e. ...

NOTES (not for presentation)

- Oct 14 copied from [RAC 2020 Best Practices - UVic - FINAL](#)
- Shared with Jana and Erika with various notes

Lots of questions.

- Probably 1 generic one rather than individual presentations for each site. Just not enough time to do 4 or 5 of them. ([Agreed that 1- maybe recorded? That we can get to a wide range of people, but we are certainly in timing crunch](#))
- Should this be a couple of 1 hr sessions, or a bigger 2 hour session? ([I think one session](#))
- Interaction has always been a key part of the previous f2f sessions, but I don't really know how to do that in this kind of virtual setting. I suspect I will just stop and ask for questions at intervals. ([You might have someone monitoring whatever interaction point \(Chat / Slack\) and pick out some of the key questions but maybe answer the questions and attach it is a FAQ to the recorded presentation?](#))
- Morning or afternoon? I'm usually at my best in the morning. ([Always feel its best to be best :-\)](#))
- Advertising as usual.
- Lots of content, maybe a bit heavy up front where details people are looking for are near the end of the presentation?
- Some of the stats are interesting as per our recent discussion and some of the slides are useful for other purposes - so thank you.

This Presentation:

- <https://westgrid.github.io/trainingMaterials>
 - (Getting Started section)
- <https://westgrid.github.io/trainingMaterials/getting-started/#rac-2021>
- Detailed RAC 2021 Best Practices guide
 - (with everything from this presentation and more!)

Contact us anytime:

support@westgrid.ca

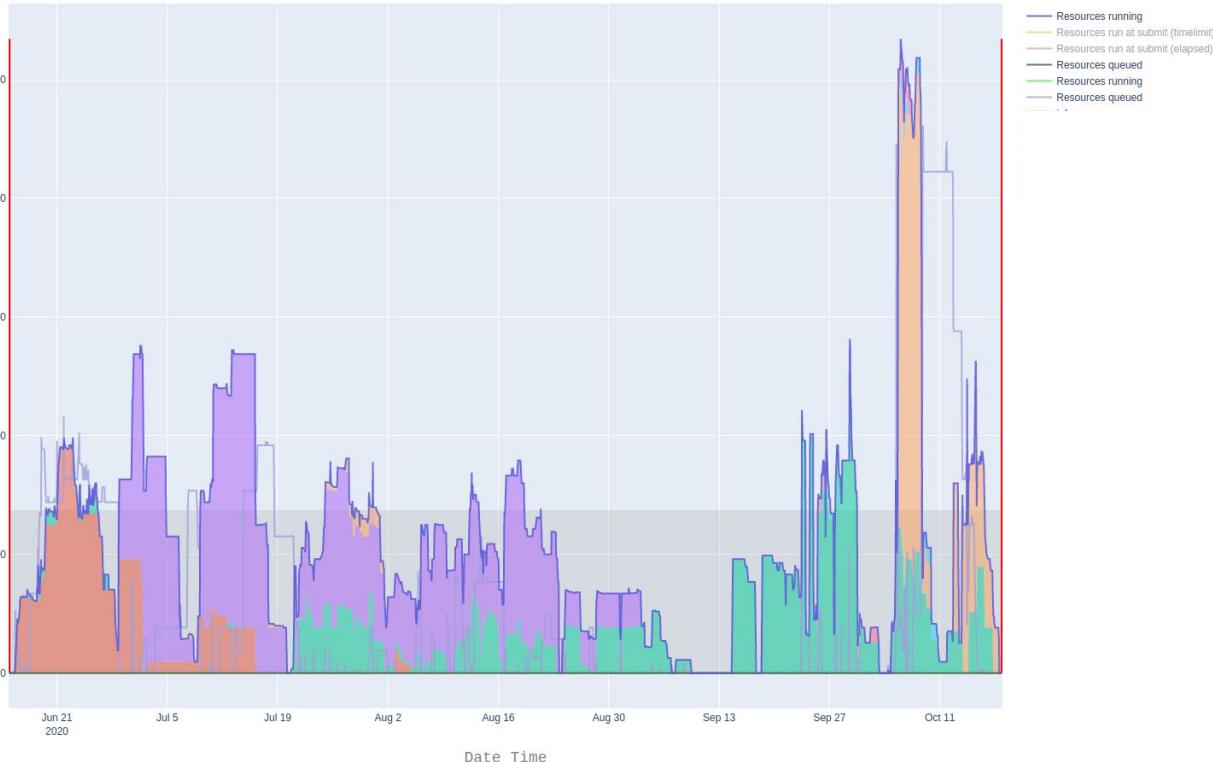
www.westgrid.ca

docs.computecanada.ca

www.computecanada.ca

Bursty use

Resource usage:



Example WG project on Cedar

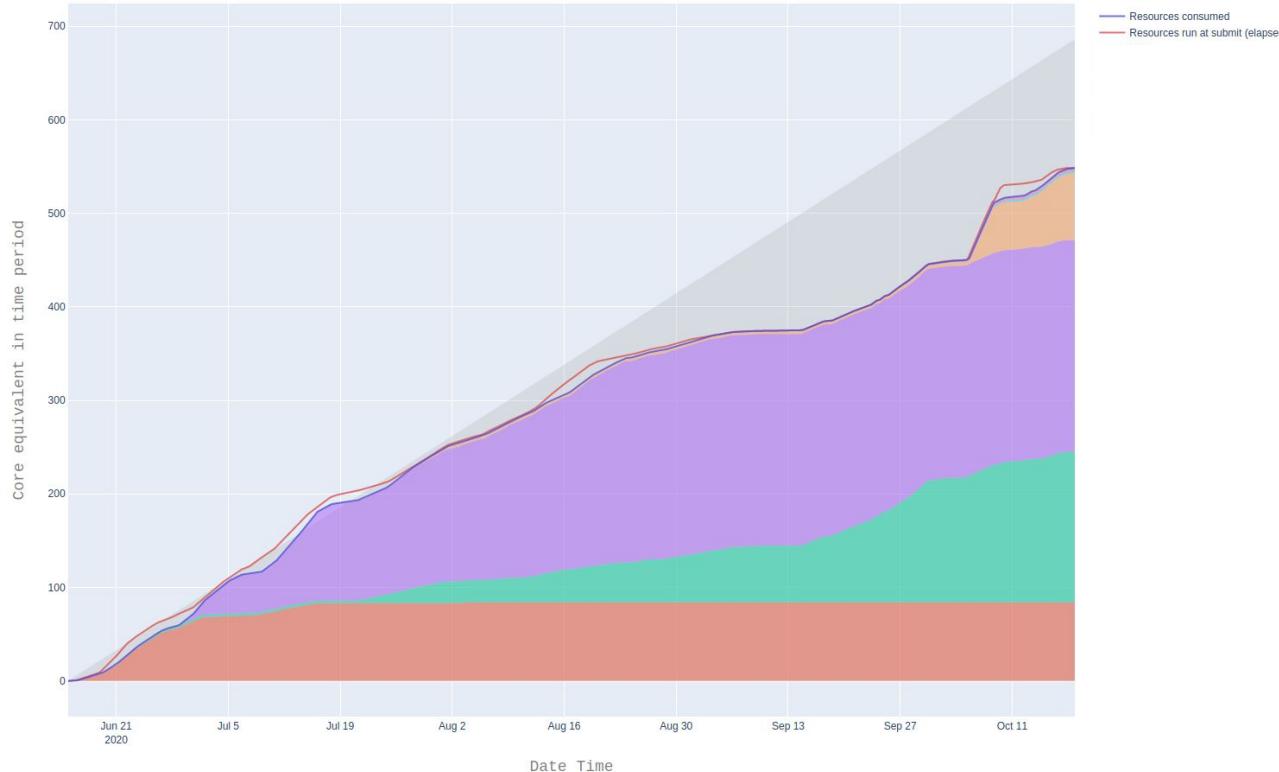
- Usage per day.

Grey = allocation
 Block colors = users within the project
 Purple Line = resources running

- Users start a set of runs and then stop for a while.
- Times when no jobs are running
- Extreme bursts well above the allocation.

Typical User - cumulative use

Cumulative resource usage:



Grey = allocation

Colours = project users

Clearly the project does not use it's allocation.

- This particular project is pretty good with reasonably consistent use.
- Other projects are very bursty and end up well below their allocation.



Intro to WG and CCF

First, an intro to....



and



compute canada | **calcul** canada

compute canada
regional partner

CCF = Compute Canada Federation
CC = Compute Canada the Corporation