



*An important message from Compute Canada
to Canadian Researchers and Research Communities*

Compute Canada's Call for Whitepapers for Sustainable Planning for Advanced Research Computing (SPARC)

Further to the recent announcement for SPARC, Compute Canada is issuing a call for whitepapers addressing the future needs of specific research disciplines. The announcement is attached with this message.

This year, CFI is planning to announce the first major competition to fund shared cyber infrastructure since the National Platforms Fund (2006). The competition is foreseen to include both a renewal of the existing Compute Canada national platform and funding for domain-specific projects built in cooperation with Compute Canada. The infrastructure that will be deployed as a result of this competition will provide the primary advanced research computing resource for thousands of Canadian researchers between 2016-2021. As such, Compute Canada is facilitating your research community to influence the planning for this infrastructure. This is a critical exercise that will have a long-term impact on your research area.

Through our SPARC initiative, Compute Canada will be calling for input throughout the summer of 2014. One element of this engagement is this call for discipline-specific whitepapers addressing the future needs of our research communities.

Different disciplines have different advanced research computing needs which may include:

- Analysis of "big data"
- Running software with a large memory footprint
- Specialized hardware (e.g. GPU accelerators)
- Analysis of sensitive private data
- Access to a very large number of computational "cores" in a single system
- Dedicated software platforms, scientific gateways, virtual machines in a cloud environment

We need to understand your needs. Keeping in mind the 2016 to 2021 timeframe, please address the following basic questions in your response.



- What kinds of problems are you trying to solve?
- What kind of infrastructure is best suited to solve these problems today?
- How much of that infrastructure would be needed to meet the needs of your community?

For your convenience please find a sample requirement below.

Available Support:

Compute Canada can offer support in a number of ways, for example:

- Translating a scientific need into a list of infrastructure requirements.
- Organizing cyber infrastructure focused meetings or workshops among researchers in your discipline.

Deadline:

Please submit your whitepaper by July 31, 2014. Submissions should be sent to sparc@computeCanada.ca.

This will not be Compute Canada's only call for input or discussion. We need to have an overview of your requirements by the end of July 2014 in order to incorporate your discipline specific needs in our own preparation for the CFI's expected request to Compute Canada for a renewal proposal.

If you need further information or have any other questions please feel free to contact us at sparc@computeCanada.ca.

Dugan O'Neil
Chief Science Officer
Compute Canada



Whitepaper Guidelines:

There is no required format for whitepapers submitted in response to this call. However, we provide an example below to help guide disciplinary groups in providing the most useful information for Compute Canada.

Researcher or Discipline-Based Group Sample Whitepaper Requirements:

Identify who wrote the whitepaper, their contact information and provide an overview of the scope. Please make it clear whether the needs reflect only those of the authors or represent the broader discipline and how widely the represented community has been consulted.

Science Description:

What are you studying and why? This can be a single paragraph. Please note: If you wish to paste-in a page from a grant request or other document, that is also welcome.

Current Use of Advanced Research Computing:

There are several relevant questions to answer here:

- Do you make use of existing Compute Canada (CC) systems? If so:
 - Which ones?
 - Do you have a special resource allocation?
- Do you currently make use of non-CC resources? Which ones?
- What is your current level of usage of storage, compute, cloud, portals/gateways, accelerators, etc.? (Note: If you use only CC resources please provide your CCDB IDs — CC staff will calculate the current usage of your group)
- Do you currently rely on any software packages, middleware platforms or other sort of shared software infrastructure?

Future Growth:

Do you expect your needs to grow? What drives this growth? For example, if improvements in sensors, detectors, imaging and equipment are driving progress in your field, you might expect commensurate growth in advanced research computing needs.



What is the estimated scale of growth in the next five years (2016-2021) compared to today? For example, “we expect datasets to grow by a factor of three and our computing needs to expand by a factor of two.”

Technical Details of Future Needs:

Not all groups will be able to translate their scientific needs into technical needs. CC experts can help. Any information you can already provide on future needs would be welcomed. All questions refer to the 2016-2021 period.

- **Data:**
 - How much archival storage do you anticipate needing? Do you have any special requirements around data preservation, curation, access, etc.? What is the time profile of this need?
 - Do you need to be able to serve your data to a compute facility (thereby requiring fast disks connected to a significant source of CPU)? How much fast-connected storage do you envision needing and with what time profile?
 - Does your data need to be shared by many researchers? If so, will it be served to the international community? If so, how will it be served to the community (e.g. web-based portal/gateway)?
- **Computation:**
 - Are your calculations serial or parallel? If you use parallel processing, what is the largest number of cores you use (or would like to use) in a single job?
 - Is the nature of your computation constant or does it have peaks and valleys? If there are peaks, please provide both average and peak amounts.
 - Based on today’s CPU cores, how many cores do you need and what is the time profile of this need? What is the reason for the increase? Please see the example below:

Now	2016	2017	2018 (new sensor comes online)	2019	2020
1000 cores (peak) 200 cores (avg)	1500 cores (etc.)	2000 cores	5000 cores	6000 cores	7000 cores



- What are the memory requirements associated with your calculations? Please specify if the number is a total or is per core. Please specify if you require large shared memory systems and try to quantify the amount of memory needed.
- Do you benefit from accelerator systems (e.g. GPU, FPGA)? What gain do you expect if CC provides accelerators compared to performing your calculation on regular CPU cores? Please provide details on how many GPUs (or otherwise) you would use and a time profile for that use.
- Do you expect to need what are being commonly labeled as “Big Data” clusters and services? This may include hadoop clusters, no-sql and distributed databases. Please quantify the level of need (and provide a time profile) if possible.
- **Software:**
 - Do you foresee a change in your software/middleware needs?
 - Would you benefit from the work of a software development team to bring new software to your field? What do you need?
- **Networking:**
 - Do you envision using multiple sites requiring strong network connectivity between them? What volumes of data do you expect to move from site to site and what is the required transfer rate?