Computation and Data Committee Report to the CASCA Board, May 2017

Current Committee membership:

James Wadsley (McMaster) (Chair)	Term ends: 30 June 2018
Hugo Martel (Laval)	Term ends: 30 June 2017
J. J. Kavelaars (HIA/NRC/CADC):	Term ends: 30 June 2017
Erik Rosolowsky (Alberta)	Term ends: 30 June 2018

Current members are willing to continue (e.g. Martel, Kavelaars). However, if there is a keen potential new member they are also happy to rotate off. If the CASCA board wants to raise this issue at the annual meeting or otherwise bring up the issue with potential candidates please go ahead.

Computing in Canada

Compute Canada

Compute Canada (CC: http://computecanada.ca) is the main source of cycles and storage for researchers at Canadian universities. It is led by CEO, Mark Dietrich, CSO (Chief Scientific Officer) Dugan O'Neil (Physics, Simon Fraser) and CTO (Chief Technical Officer) Greg Newby. Dugan will shortly step down to take on the deputy VPR role at SFU in 2017. The governing board has been reorganized and the past chair has been replaced.

The CC Advisory Council on Research (ACOR) includes James Wadsley (McMaster), of this committee and Robert Thacker (SMU) but has not met for some time. There are serious ongoing concerns about researcher oversight/engagement. The members of CC are the universities and colleges in Canada. Some member and provincial representatives (e.g. Compute Ontario) have raised related concerns about the CC governance structure and have been pushing the members to rethink it. It is unclear on what timescale this might occur.

Hardware Refresh

The Stage 1 hardware refresh settled on 4 sites: SFU (CEDAR: GPU focused ~ 25000 cores, ~ 3000 GPUs), U-Vic (ARBUTUS: cloud focused ~15000 cores, 1.6 PB), Waterloo (GRAHAM: ~25,000 cores) and Toronto (NIAGARA: Large parallel system ~ 66,000 cores). The fastest moving sites are Arbutus (available now) and Cedar which opens to users this month (May-June 2017). Graham's hardware is nearly installed and will be turned over to local staff soon. Niagara is at the RFP stage with installation expected to occur throughout 2017 and becoming available to users in early 2018. Part of the delay for Niagara was to try to get the newest Intel chips so exact core numbers aren't clear.

Systems predating the refresh are old and will be shut down in a staggered way over 2017 but there is still no clear plan for how this will happen. Some machines will be decomissioned this summer (e.g. Colosse). The biggest issue with this is that Quebec and Eastern Canada may be without new systems for an extended period.

The initial expected overall storage looks quite modest 40-60 PB. CC plan to expand this from 2018-2020 to over 200 PB. There are discussions with CFI over further refresh ("Stage 2" submitted in 2016) that would likely allow for current major sites (e.g. Sherbrooke) to be refreshed as well as expanding storage.

CC renamed its user resource competitions (**https://www.computecanada.ca/research-portal/accessing-resources/resource-allocation-competitions/**) as Research Platforms and Portals (RPP – focused on programming and software infrastructure such as web portals) and Resources for Research Groups (RRG – specific hardware resources such as core years for computing and storage). The latter was previously known as RAC and currently allocates up to 80% of most systems. Overall the demand is high compared to the supply of cycles with a high degree of over-subscription and relatively low awards in many cases. Decisions were given in March 2017 and often have delayed starts due to the ongoing installs. The current systems are not well suited to very large parallel (>1000 cores) which may have to wait for Niagara (2018).

CANFAR, CADC and CANARIE

The situation for CADC/CANFAR/NRC is a bit uncertain at this time. The CANFAR project has brought significant effort into the process of providing observational astronomers with access to cloud storage and computing. NRC has agreed to fund CADC to sustain this system, but there are many improvements needed if CANFAR is to be a research platform for all of astronomy. Typically the development of those improvements would be performed under a development grant. Previously these had been awarded by CANARIE but there is now significant funding available from CFI under their cyber-infrastructure program. How astronomy computing can take advantage of the CFI CI opportunity is not clear.

The CFI and NRC are discussing eligibility, regarding the CFI Cyber Infrastructure call, of Astronomers to fund activity at NRC to develop services that would support science exploitation of data acquired by observational facilities that NRC operates. CFI sees the CADC mandate as requiring NRC to fund such activity while NRC sees the CFI mandate as responsible. Clarifying conversations are needed.

Internally, the CADC has been working with Compute Canada to provide archive storage on compute Canada hardware. This has the advantage of moving the data to the processing that Compute Canada has available for the research community, lowering network latency and load. The project to make this transition is moving along. The significant work will be run major parts of the current CADC archive infrastructure on Compute Canada provided services. Work is scheduled to complete in Fall 2018. In addition, NRC has recently approved the expenditures required to purchase refresh of CADC's internal storage capacity. The details of how this refresh will proceed are still uncertain.

CANFAR CFI Cyberinfrastructure proposal

CANFAR submitted an NOI on 12 May 2017 (through U-Vic, PI: Falk Herwig), with engagement from astronomy groups all across Canada) to CFI's Cyberinfrastructure program. Full proposals will be invited in June 2017. CFI would offer 40% of up to $5M (including matches) for 3 years if funded. There are many restrictions on how this money can be spent and it must mostly be programmer salaries. There is also a prospect for partial FTE data specialists who might also be researchers for the other part.

The proposal outlined in the NOI would be to build upon the established CANFAR/CADC low level services to provide higher level services for all Canadian astronomers. This includes frameworks which can readily produce useable data reduction pipelines and other services for end users. It also included expansion in new directions in support of simulations and simulated observations/data

analysis. This proposed work is to help Canadian astronomers deal with the large data volume coming from newer instruments. Numbers based on upcoming needs (VLASS, CHIME) that can be quantified are in the 10 PB range. SKA will be very large but is hard to quantify right now.

CADC is very interested in doing this and have many people with the right expertise. As noted above, CFI might decide CANFAR is not eligible because of the CADC connection, which occurred last time. As a result the NOI allowed for programmers at both universities and CADC, partly as an attempt to manage the uncertainty coming from CFI. Recent communications from CFI are positive in the sense that the NOI will not be automatically disqualified but what kind of spending is eligible remains unclear.

It is important to note that CFI sees this as a one time project. CFI does not plan to renew funding. Instead, it expects this software infrastructure investment to smoothly transition into a long term platform in support of research. For this CADC and its staff will be essential. As a result it is critical that CADC continue to be funded to work on CANFAR services. This might be funded via discussions within NRC (with support from the Astronomy community) or via CANARIE. If the CFI Cyberinfrastructure proposal for CANFAR is allowed to go forward with the expected restrictions on funds going to NRC, it will make the need for supporting CADC in some other way much more urgent. Hopefully this would make it easier to make the case to other funders such as NRC itself or CANARIE.