Submission to
NDRIO's Call for White Papers on Canada's Future DRI Ecosystem from the Compute Canada Federation Subatomic Physics National Team

December 13, 2020

Leslie Groer, PhD[1] - SPNT lead, University of Toronto, SciNet Consortium
Sergey Chelsky, PhD - Simon Fraser University, WestGrid Consortium
Lixin Liu, PhD - Simon Fraser University, WestGrid Consortium
Simon Nderitu, PhD - McGill University, Calcul Quebec Consortium
Erming Pei, PhD - University of Alberta, WestGrid Consortium
Ryan Taylor, MsC - University of Victoria, WestGrid Consortium

Introduction

The Subatomic Physics National Team (SPNT) was formed in 2016 under the auspices of the Compute Canada Federation (CCF) Science Leadership Council. However, there was a long-standing tradition in Canada, stretching back over a decade, of the personnel supporting grid and distributed computing for high-energy physics (HEP) experiments at the five hosting sites[2] to coordinate efforts and meet weekly, along with personnel supporting the TRIUMF Tier-1 ATLAS Data Centre. Face-to-face meetings occurred at least once a year as well.

In many ways, the SPNT was the first truly *national* and *federated* team at the computing sites that eventually were incorporated into the CCF. This team also occupies a fairly unique position within the CCF as team members straddle the research and computing support communities and many of our activities require significant coordination within the federation for hardware purchases (such as for storage and network equipment), integration with scheduling, authentication and monitoring systems and support of the middleware and software layers of the experiments. Large experiments such as those in subatomic physics, but increasingly in other fields, have to focus on data-intensive high-throughput computing (HTC) rather than traditional high-performance computing (HPC), and it can be challenging to design and support large clusters and storage systems that do both well. Given these challenges and the unique position of this team, a white paper submission to the NDRIO call is warranted by the technical and operational members of the SPNT to ensure certain experiences and perspectives from within the current organization are maintained as NDRIO and the new Canadian organization around DRI is formed. We do not purport to address larger structural, organizational or funding issues that are beyond our horizon or purview. This paper is endorsed by the Principal Investigators (PI's) responsible for the CCF computational resource allocations for some of the main experimental communities we support, as listed in Appendix A.

Success Stories in the Current CCF Organization

Below we identify some of the key aspects of the successes of the SPNT within the CCF and liaising with the research communities that we support, such as the ATLAS experiment based at CERN, Switzerland, the largest experiment in the world. Over 150 Canadian researchers (43 PI's) participate in this experiment and investments from CFI and NSERC in ATLAS total over $35 million in the last few years. Similar investments have also gone into other subatomic physics experiments that are hosted at Canadian research facilities, such as SNOLAB (SNO+, DEAP-3600, PICO and soon SuperCDMS) and TRIUMF, where CCF provides the majority of the compute resources

[1] Leslie Groer <leslie.groer@computeontario.ca> is the contact for this white paper
[2] The original five Tier-2 sites were hosted at University of Victoria, Simon Fraser University, University of Alberta, University of Toronto and McGill University/University of Montreal. Past experience supporting experiments out of Fermi National Laboratory such as CDF and Dzero paved the way for the current distributed analysis centres.

for these world-wide collaborations. Also, large computing resource commitments have been made to the Belle II, T2K, IceCube and GlueX experiments, which have significant Canadian contributions. These experiments share many similar operational infrastructure and software frameworks, lending themselves to group support by the SPNT. The SPNT facilitates the use of over 10,000 computing cores throughout the year and over 10 Petabytes of storage on the various CCF platforms.

The key drivers of success for the SPNT are:

- **Integration of Research and Support Personnel**: The Canadian ATLAS Tier-2 analysis facilities have worked very well for over a decade because of the embedding of key personnel with their domain-specific subject-matter expertise within the CCF and at the hosting sites. We have also benefited greatly from the support of the TRIUMF Tier-1 ATLAS Data Facility. This knowledge base and experience is now extended to supporting additional experiments on the CCF platforms which also rely on distributed data models across national borders, involving many institutions in other countries, and require large scale high-throughput computing for simulation production or data reprocessing. Crucially, SPNT experts were part of the procurement teams for the current facilities or had extensive input into design and purchasing decisions. Given the scale of the resources needed, viz. thousands of processing cores accessing many Petabytes of storage, it is not always feasible to impose these requirements on facilities after the fact without incorporating them from the inception. Team members are also cross-appointed to many other CCF teams, allowing for full integration across the enterprise during deployment and operations.
- **Collaboration with Research Communities:** From the inception, team meetings have included key members of the experimental groups (including PIs, postdocs and graduate students) who are responsible for their computing processing and data production. This collaborative model has worked successfully for many years for the large subatomic physics experiments globally. While there are transient groups of postdocs and graduate students that usually run the experiments' data processing, the core SPNT members are permanent site experts who support the work nationally and also make contributions to the centralised experiment teams, for example by participating in workshops at CERN or at international conferences. This fosters productive working relationships and collaborative problem solving with remote colleagues, short-circuiting long email chains or ticket discussions to quickly remedy issues, and allows us to conduct long-term strategic planning, and advocate effectively for the needs of the Canadian sites and researchers.
- **Providing Reliable Infrastructure and Services**: The technical members of the team each have more than a decade of experience in providing reliable 24/7 services to the experiments. There is extensive collaboration with the experienced personnel at the ATLAS Tier-1 Data Centre, which also leads to quick problem resolution, informs many design decisions at the CCF, and allows extended time-zone monitoring of sites which improves reliability and availability of the systems and platforms. Core team members are also integrated with the site teams which promotes timely and effective support.
- **Collaboration with Network Communities:** The SPNT also collaborates with CANARIE and the provincial providers of research and education networks (NRENs) as well as organizations such as HEPnet Canada[3] and the LHC Open Network Exchange (LHCONE) to ensure that adequate network capacity is deployed at the various CCF sites and the necessary protocols are supported. The SPNT was an early adopter to deploy the perfSONAR[4] federated network monitoring platforms at the ATLAS Tier-2 sites. perfSONAR has since developed into a world-wide standard and is employed by CANARIE and the NREN's, as well as all the CCF sites. The SPNT has also collaborated closely with the GridCanada Certificate Authority which provides the necessary credential services to facilitate secure operations of the Canadian grid platforms, and which was absorbed into the CCF organization.
- **Integration of New Technologies and Platforms:** Several fundamentally crucial components of the CCF enterprise grew out of projects originally spearheaded and supported by the SPNT and the HEP community. These include:

[3] HEPnet Canada is responsible for supporting the Canadian Subatomic Physics community on the national and international research networks.

[4] https://www.perfsonar.net/

- HEP pioneered the use of the Globus toolkit and GridFTP technologies for large scale data transfers (along with X.509 public-key infrastructure authentication) which has since been adopted as a standard at CCF in collaboration with Globus Connect.
- Singularity containers for job encapsulation on HPC platforms which is now becoming common across the enterprise.
- The CERN Virtual Machine Filesystem (CVMFS) for software distribution, which was pioneered by the CERN projects and is now the foundation for the distribution of the CCF software stack to all federation sites and the world-wide community, as well as for distributing genomic data sets.

There is still great potential for technologies that have been developed and in use for years among subatomic physics experiments to be reused, such as for large-scale distributed data management and federating sites for better utilisation of resources.

Possible Areas of Improvement or Anticipated Needs for the Future

Below we list some of the issues that the team is currently dealing with or anticipates challenges with in the future as the organization grows, new experiments and groups are on-boarded, and systems evolve. These are in no specific order of priority.

- **Support Distributed Analysis Workloads at all the CCF GPC Sites**: Given the growth in demand and the reliability and uptime requirements from the experiments, the team believes it is essential to have a grid presence at all CCF sites that support general purpose computing (GPC) or at least a common approach to allowing network access to remote data sources (including database and job control information). This not only ensures effective utilisation but also provides resource redundancy for high availability. There may be issues with supporting experiment-specific meta-schedulers (such as grid-based Compute Elements or HTCondor glide-in factories) in the current support model if these are reliant on administration access by the experiment support staff that are not CCF employees. Long-term planning at all sites should anticipate some level of grid-enabled storage as well (mostly dCache, but the team has supported StoRM[5] on shared file systems in the past at McGill University as well).
- **Provide Reliable Infrastructure and Services**: Since most of the experiments and collaborations we support are world-wide and require services and jobs to run 24/7, the requirements of reliability and availability are prodigious. Current support at the federation's facilities works well but is mostly "best effort" in coverage outside of normal working hours. Moving to a support model that adheres to higher reliability standards will take considerable extra personnel and effort so as not to unreasonably tax the current teams, and must include consideration of university contracts and union requirements.
- **Align the Resource Allocation (RAC) Process and Equipment Procurement:** There is currently no formal connection between the yearly RAC application process and allocation of resources, and the actual purchase of resources at the sites. SPNT has participated (or intervened) in the process to ensure success so far, but the model is not efficient or guaranteed to provide the necessary resources when there can be 2-3 year lags in funding, particularly for dCache storage.
- **Integrate Policy Development at Sites and with National Teams:** Policy development at the sites for data and network access and file system structures has not always been done in consultation with the SPNT, leading to some lengthy and sometimes frustrating development cycles integrating the necessary distributed computing platform infrastructures.
- **Provide support in the CCDB for Large Long-term Projects:** From the team's perspective there is a fundamental design challenge in the Compute Canada Database (CCDB) and RAC process: everything is strictly based on PI's and does not account for multi-decadal projects with dozens or more researchers that need to share data and resources. Every RAC year there are a few projects we have to deal with where the PI has changed or left the country but the project persists; inevitably it takes at least a couple of months to ensure all the file ownership and permissions are set correctly and in a way that does not impact users, backups or other longer term storage, chewing up many person-hours in multiple CCF

[5] https://italiangrid.github.io/storm/

teams. Several times this has impacted accessibility and use of the systems for users for extended periods while this is sorted out internally.

- **Support multi-Domain Collaborative Environments and Tools:** As mentioned above, one key aspect to the success of the SPNT activities has been collaborating closely with the stake-holders in the experiments and other organizations (e.g. ATLAS Tier-1, CERN, HEPnet Canada, etc). Some of the current tools used by CCF for internal collaboration (e.g. Slack, GitLab, staff wikis) have strict identity boundaries between those in CCF and those outside, preventing external collaborators from participating and creating information silos. This has caused difficulties or, at best, inefficiencies in setting up communication channels and document sharing for teams that have non-CCF members.
- **Provide Dedicated Test and Development Platforms**: As all the middleware and other supporting service components evolve continuously, there is a need for dedicated Test and Development systems within the CCF to be able to test upgrades or on-board new experiments without disrupting the 24/7 production systems. This would require a few pieces for dCache in particular; it would be good to minimize some of the expensive pieces such as a dCache data storage pool node but still emulate all the hardware components and network connections.
- **Provide Federated Identity Management:** X.509 certificates will be phased out world-wide in the next few years in favour of token-based authentication systems. The CCF needs to engage with other identity providers through member institutions to implement federated identity vetting systems, likely under the CANARIE Canadian Access Federation project[6].
- **Adapt and Evolve Large Scale Distributed Computing and Services:** Broader discussions are needed as experiments adopt more modern computing technologies and platforms, including cloud infrastructure and cloud-native technologies such as Kubernetes and Object Storage, in conjunction with, or as an alternative to, grid computing. There is also evolution in the traditional grid-services being directed under the WLCG long-term Data Organization, Management and Access projects[7]. Non-subatomic physics groups also have similar needs, notably the Square Kilometre Array astronomy project, discussed below. As the cloud resources develop within the hosting sites and as the number of world-wide collaborations grow and move away from centralized grid services to other distributed analysis tools (e.g. RESTful APIs or token-based authentication), planning for these large projects needs to be cognisant that secure world-wide access and high-bandwidth interfacing to the storage technologies is a key component to the successful use of these platforms. Interfacing these with longer-term tape storage may also be required. As such, there needs to be tighter connection between the SPNT and the teams thinking about security, authorization and access to these shared CCF resources.
- **Support New Analysis Frameworks:** New experiments are also exploiting new data analysis tools and techniques such as python-based Jupyter Hubs and projects coming out of the HEP Software Foundation[8]. Adapting and evolving these within the CCF framework and ensuring a consistent user experience across the federated sites can be a challenge.
- **Provide Integrated Long-term Data Storage**: Many of the longer term projects, especially those where Canadian experiments are producing the data (e.g. from SNOLAB) require long-term data archiving and access. Most of the data archiving that is currently done by these groups on CCF platforms is very labour intensive and occurs outside of the normal workflow at the sites. There is currently no model that is supported in CCF which ties the dCache storage systems to a tape archive backend, although dCache is eminently suitable and designed for this purpose. This would be a long term development and support project with the SPNT and other site and national teams to achieve at the necessary multi-Petabyte scale.
- **Develop and Support Network and Research Data-Management Collaborations:** Large world-wide projects increasingly are building collaborative enterprises to ensure their users can get transparent access to very large datasets and to be able to move and process these efficiently, running the gamut from on-demand software defined network routes to large scalable file cataloguing and intelligent

[6] https://www.canarie.ca/identity/

[7] DOMA and ESCAPE Projects: https://projectescape.eu/news/new-directions-distributed-computing-doma-and-escape

[8] https://hepsoftwarefoundation.org/

transfer services. Canada was an early adopter in the LHC Open Network Exchange (LHCONE) which has proved very effective and efficient for ensuring large data flows do not disrupt other research and education activities and served as a model for the private CCF Compute Canada Network layer. It is crucial that the new organization actively participates in these enterprises to maintain the competitive edge in Canada and for the research groups that we support.

- **Provide Unified Change Management, Notification and User Support**: Currently the site-based dispersed change management, notification and support model is neither sufficient nor agile enough for SPNT projects that span multiple sites domestically or internationally. It is likely we need to develop a uniform dashboard for subatomic and similar projects to collectively receive notification of system and other changes. Likewise, integrating performance and other monitoring across the federation sites would be of benefit for the large projects.
- **Prototype and Develop Canadian SKA Regional Centre**: The team anticipates an interesting opportunity as the national Square Kilometer Array Regional Centres develop and take shape as these likely will exist in between a Tier-1 and Tier-2 model that we have for ATLAS now. The current target is that such a facility would be based at a CCF site but there is likely not enough manpower internal to the current astrophysical community to provide and manage the necessary underlying services. Expertise at the CCF sites from the subatomic physics realm may prove crucial here to the success of such a Canadian centre, leveraging the many years of experience in such large distributed projects. Prototyping such centres is possible now and desirable with existing Canadian astronomy and astrophysics experiments such as the Canadian Hydrogen Intensity Mapping Experiment (CHIME), the Legacy Survey of Space and Time (LSST) and other burgeoning experiments.

Appendix A: Endorsements from Members of the Canadian Research Community

Prof. Kenneth Clark	Queen's University	PICO Collaboration (SNOLAB) and IceCube South Pole Neutrino Observatory
Prof. Miriam Diamond	University of Toronto, Faculty, Arthur B. McDonald Canadian Astroparticle Physics Research Institute	SuperCDMS Collaboration Simulations Working Group Chair & Software Infrastructure Support Operations Manager
Prof. Douglas Gingrich	University of Alberta and TRIUMF	ATLAS-Canada Computing Coordinator
Prof. Blair Jamieson	University of Winnipeg	Tokai to Kamioka (T2K) and Hyper-Kamiokande (Hyper-K) experiments
Prof. Chris Jillings	Laurentian University and SNOLAB	DEAP-3600 experiment
Prof. Carsten Krauss	University of Alberta	Director Centre for Particle Physics P-ONE, PICO & SNO+ experiments
Prof. Zisis Papandreou	University of Regina	Head, Department of Physics GlueX experiment (Jefferson Labs) Collaboration Board
Dr. Reda Tafirout	TRIUMF	ATLAS-Canada Tier-1 Manager and Group Leader