September 2022

Digital Research Alliance of Canada HPC Strategy

HPC Strategy Working Group

Scott Northrup (Chair), Bruno Blais, Roy Chartier, Rebecca Davis, Catherine Lovekin, Patrick Mann, John Morton, Florent Parent, Seppo Sahrakorpi

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Table of Contents

1 Executive Summary

1.1 Objective

The intent of this report is to outline a strategy to maintain and expand the High Performance Computing (HPC) national platform infrastructure currently available to Canadian researchers. It provides an overview of the system's current state and capacities, investigates forecasted demands and growth, and proposes scenarios to maintain, improve and support the HPC platform.

1.2 Definition

For the purposes of this document we define High Performance Computing (HPC) as the use of clustered systems in which compute jobs are submitted to the system through the use of a scheduler. This includes attached active storage resources and may provide a variety of compute resources, including accelerators of various kinds, and excludes cloud computing and archival storage.

1.3 Researcher Strategic Needs and Vision

The Researcher Council's Priority Needs report,[1] Researcher Needs Assessment consultations and report,[2] and Advanced Research Computing,[3] Research Data Management[4], and Research Software Current State[5] reports provided the Alliance with an extensive understanding of researcher's needs for the Canadian DRI ecosystem going forward. These reports and additional

[1] Alliance Researcher Council: Meeting the Digital Research Infrastructure Needs of the Canadian Research Community https://alliancecan.ca/sites/default/files/2022-03/researcher-council-prioirities-september-28-2021-final_en.pdf (September 2021).

[2] Alliance: Researcher Needs Assessment: summary of what we heard https://alliancecan.ca/sites/default/files/2022-03/needsassessment_alliance_20220126.pdf (September 2021).

[3] Alliance Advanced Research Computing Working Group: Current State of Advanced Research Computing in Canada https://alliancecan.ca/sites/default/files/2022-03/arc_current_state_report_0.pdf (May 2021).

[4] Alliance Research Data Management Working Group: The Current State of Research Data Management in Canada https://alliancecan.ca/sites/default/files/2022-03/rdm_current_state_report-1_1.pdf (November 2020).

[5] Alliance Advanced Research Software Working Group: Research Software

Current State Assessment https://alliancecan.ca/sites/default/files/2022-03/rs_current_state_report_1.pdf (September 2021).

stakeholder consultations provided the foundation for Alliance's Strategic Plan 2022-2025,[6] the strategic vision and priorities document that was published in February 2022.

While the researcher needs span the whole ecosystem covering advanced research computing to research software to research data management, and ranging from infrastructure to services, operations, training, funding, and personnel concerns, this report focuses on Advanced Research Computing (ARC) and HPC infrastructure needs, excluding cloud computing and sensitive data solutions that will be covered by separate dedicated strategies. Within that scope, the Strategic Plan and researcher community clearly articulated the need for greatly expanding the national advanced research computing capacity by building on Compute Canada Federation's ARC and HPC infrastructure and services, including improving accessibility, cybersecurity, and expanding storage solutions, including long-term storage and preservation solutions. Notably the Researcher Council encouraged the Alliance to double the ARC/HPC compute capacity within its first mandate to reach the average of G7 countries in gross domestic product (GDP) weighted compute capacity. In addition to expanding capacity, inefficiencies in system usage need to be addressed to improve availability and maximize investment value. The usability of ARC/HPC systems needs to be improved via training, and customized and integrated software and workflows for discipline specific needs, e.g., in digital humanities. The investments in ARC/HPC and storage infrastructure need to happen through a financially sustainable plan that considers the infrastructure's maintenance and updating requirements. The future state also needs to include designs and implementation of risk mitigation measures to prevent service disruptions.

1.4 HPC Architecture Recommendations

Researcher demand for HPC resources is growing every year and the need for maintaining, expanding, and supporting capacity growth is a topmost priority. Based on the detailed analysis provided in this report the following key recommendations are summarized here:

Critical Infrastructure Replacement

- Provide immediate funding to replace aging infrastructure to avoid reduction of service and maintain current baseline capacity. By the end of 2024, 234,000 CPU cores and 2,200 GPUs across the federation systems will be greater than 5 years old and need to be replaced to maintain 2021 level capacities. (See Scenario I in Section 4.2 for details).

Expand HPC Capacity

- Adopt a multi-year coupled capital and operating investment approach to continuously grow the baseline HPC capacities. Continuously review deployment planning to respond

[6] Alliance: Strategic Plan 2022-2025 https://alliancecan.ca/sites/default/files/2022-03/Alliance_Strategic_Plan_2022_2025_ENGLISH_SINGLES.pdf (February 2022).

to changing workload characteristics and research project requirements, assess demand, and consider state-of-the-art technologies.

- Increase Canada's research competitiveness, and support growing demand.
    - Double current capacity to move Canada from last to mid-range of G7 countries in FLOPS/$GDP. That is, target growth of an additional 100k CPU cores and 1,000 GPUs per year for the years 2023-2025 to double 2021 HPC capacity by 2025. (See Scenario IV in Section 4.2 for details)
    - Invest to have at least 1 system in the top50[7] of the TOP500, designed for researchers who require massively parallel capability class computing. Maintain at least 3 systems in the top 250 of the TOP500.

Invest in Storage & Data Management

- Increase HPC active storage in step with plans to double HPC capacity by 2025.
- Develop a Long Term Storage plan that integrates research data management and archival storage.
- Deploy a storage solution for common data-sets and critical user data that will be available on all systems.

Design for Greener HPC

- Emphasize efficient (W/Flop) compute in system designs.
- Target infrastructure and datacenter investments which maximize power and cooling efficiency while lowering operating costs and carbon footprint.
- Develop life-cycle plans for HPC equipment to include repurposing equipment where possible extending its service lifetime.

Increase Platform Resiliency

- Invest in geo-replicated backups of critical data.
- Develop site disaster recovery and mitigation plans that align with agreed upon Service Level Objectives (SLO).

Improve Support, Efficiency, and Usability

- Increase research support staff to support expanded HPC compute resources.
- Promote initiatives that increase system and user efficiency.
- Expand access options beyond traditional command line interfaces (CLIs) to support a more diverse population of researchers using HPC resources.

[7] TOP500 https://www.top500.org/lists/top500/ (November 2021).

2 Current State

2.1 Current HPC Resources

The current national HPC ecosystem was refreshed starting in 2016, funded through Canada Foundation for Innovation (CFI) investments in Canadian cyberinfrastructure along with a follow-on investment in 2019 by Innovation, Science, and Economic Development Canada (ISED). The consolidated resources are hosted at five main Compute Canada Federation (CCF) sites. The national systems and their affiliated regional CCF member organizations are as follows, from West to East:

- University of Victoria, Arbutus (BC DRI Group, formerly WestGrid),
- Simon Fraser University, Cedar (BC DRI Group, formerly WestGrid),
- University of Waterloo, Graham (Compute Ontario),
- University of Toronto, Niagara (Compute Ontario),
- McGill University / Calcul Québec, Béluga & Narval (Calcul Québec).

Cedar, Graham, Béluga, and Narval are general purpose heterogeneous HPC clusters to support a wide variety of HPC workloads commonly referred to as General Purpose (GP) Clusters. Niagara is a massively-parallel homogeneous cluster primarily for large-scale scalable HPC workloads commonly referred to as a Large Parallel (LP) Cluster. Table 1 lists each of the clusters, initial install date, compute capacity and allocatable storage. The resources listed do not include additional contributions added onto the base systems.

System	Commission Date	CPU cores	GPUs	Project Storage (TB)
Béluga	Sept 2019	32,080	688	31,000
Cedar	March 2017	94,528	1,352	42,000
Graham	June 2017	34,784	498	20,100
Narval	Sept 2021	61,760	524	29,000
Niagara	March 2018	80,960	216	12,300
Total		304,112	3,278	134,400

Table 1 - Current national HPC resources available to Canadian researchers (March 2022).

Along with the HPC systems, Arbutus is a Cloud system for hosting (mostly Linux based) virtual machines and other cloud workloads. Portions of the general purpose HPC clusters Cedar, Graham, and Béluga are also used for cloud workloads.

System	Commission Date	CPU Cores	GPUs	Storage (TB)
Arbutus	Sept 2016	16,008	108	17,000
Béluga	Sept 2019	3,072		2,000
Cedar	March 2017	1,216		1,300
Graham	June 2017	1,368		84
Total		21664	108	20,384

Table 2 - Current national Cloud resources available to Canadian researchers (March 2022).

2.2 Identified HPC System Challenges

The 2021 Alliance ARC position Paper provides a thorough overview of the key challenges facing the Canadian ARC ecosystem. In this section the challenges specific to HPC are highlighted.

2.2.1 Continuous Investment

To maintain capacity and provide a robust and modern infrastructure to researchers continuous investment in HPC systems is required. Sustained and predictable funding that encompasses capital and operating costs will allow for proper lifecycle planning of each system and the overall HPC architecture.

HPC systems typically have a usable lifetime of 3-5 years, 4.2 years[8] is the average according to Hyperion in 2021. The technological improvements in performance after 5 years especially

[8] HPCWire: Hyperion SC21 Market Update: 2021 Looks Strong (Surprise!); Big Systems, Cloud and AI Are Drivers https://www.hpcwire.com/2021/11/15/hyperion-sc21-market-update-2021 (retrieved May 2022).

in the core compute elements of CPUs and GPUs, makes them no longer competitive or cost effective for HPC applications. Components may be repurposed after this time for less demanding tasks which highlights the need for longer term lifecycle planning, however continuous ongoing investment is necessary to maintain HPC capacity and provide a modern and reliable resource.

With this planned obsolescence in mind it is imperative that the HPC ecosystem of multiple

systems are continuously being upgraded to avoid situations, such as we find ourselves in now, where large portions of the equipment is approaching or over the 5 years. The current state of HPC systems in Canada, as shown by the dates of the systems shown in Table 1 above, is that a large majority of the capacity was installed in 2017 & 2018 and thus needs to be replaced in 2022/2023 to just maintain capacity. Beyond capacity considerations, if planning for the replacement of specialized systems, such as the Niagara system for large scale parallel computation, is not considered, Canadian researchers will lose that capability completely.

2.2.2 Insufficient HPC Supply & Research Competitiveness

As outlined in the 2017 LCDRI ARC Position Paper, the 2021 Current State of Advanced Research Computing in Canada report, and the 2021 Researcher Needs assessment report there is currently an insufficient supply of HPC resources to fulfill demand. Each year the number of users and research groups increases as does the demand for HPC computing resources.

Among its G7 peers Canada is last when one considers aggregate total compute power in the TOP500. Looking at compute performance relative to gross domestic product (Tflops/GDP) Canada is again last within the G7. "To outcompute is to outcompete."[9] In order to give Canadian researchers the tools they need to conduct leading-edge research (ISED Digital Research Infrastructure Strategy),[10] the HPC capacity would need to be at least doubled in order to place Canada in the middle of the pack of G7.

2.2.3 Data Management & Resilience

Currently each of the systems and host sites are mostly independent in terms of operation, except for a few shared services like authentication and centralized software which are designed with High Availability (HA) in case of local failure. This means that in case of a major failure at one site, the other sites can continue to operate and researchers could move any critical workloads to one of the other systems. However, all data is currently local to a system, so if a system goes completely offline, researchers would not be able to access that data. This is a major issue preventing researchers from moving between systems.

---

[9] 1st Annual High Performance Computing Users Conference, Washington, D.C.July 13, 2004, Conference Report https://compete.org/2004/03/16/supercharging-u-s-innovation-competition/

[10] ISED: Digital Research Infrastructurehttps://ised-isde.canada.ca/site/digital-research-infrastructure/en (retrieved May 2022).

A national data strategy that not only considers traditional high performance storage targeted for short and mid-term data storage, but also long-term archival storage, off-site or cross-site critical data backups as well incorporating research data management considerations would go a long way to allowing researchers to more easily move between resources and increase the overall resiliency. However, it should be noted that HPC systems often deal with very large data-sets which are required to be local for performance and where it may not be feasible or even desirable to have replication as such HPC data workflows need to be considered carefully when determining storage designs and policies.

2.2.4 Researcher Adoption and Usability

There are roughly 33,000 full and associate level university professors in Canada, while currently CCF lists roughly 5,500 principal investigator (PI) accounts, i.e. roughly 17% of full and associate professors have registered to use CCF infrastructure. Moreover, the humanities, social sciences, business, and psychology faculty account for roughly 10% of the faculty level user accounts at CCF while these disciplines represent roughly 46% of all the full-time academic faculty in Canada.

As a wider research community requires HPC resources, there is increased demand to support alternative workflows and interfaces. The predominant way HPC resources are accessed is command-line driven (CLI), however to improve usability, solutions such as Virtual Desktop Infrastructure (VDI) or web browser based access, such as using notebooks in Jupyter or web portals are becoming increasingly in demand.

3 Forecasted Demand for HPC Resources

3.1 Compute Capacity Demand

One of the primary challenges identified with the current HPC ecosystem is that despite recent substantial capital expenditure for HPC resources, demand continues to outstrip supply.

Quantifying the demand, however, can be challenging due to the large number of researchers and the wide range of computing needs.

3.1.1 RAC Compute Demand

Currently 80% of the available national platform resources are allocated through an annual Resource Allocation Competition (RAC). Statistics from the 2022 RAC call are shown in Table 3. Requested demand substantially outstripped supply, especially in the areas of CPU and GPU compute.

Resource	Allocated Capacity	Demand	Ratio
CPU cores	238,950	435,672	**1.8x**
GPUs	2,450	9,622	**3.9x**

Table 3 - 2022 Resource Allocation Competition Demand vs Supply.

CPU resources are allocated in CPU core years defined as the equivalent of executing a program on a single CPU core for one full year. GPU resources are allocated in GPU years defined as the equivalent of executing a program on one GPU for one full year. Notably these metrics do not consider the differences in CPU & GPU architectures which vary across the various systems to simplify allocation processes. CPU core performance across generations has not been changing drastically, however differing generations and types of GPUs can offer an order of magnitude

difference in performance. This variation in GPUs makes it very difficult for researchers to estimate their actual needs as well as predict future demand.

If one looks at the RAC request data over the past 10 years from 2012-2022 the number of applications has risen steadily and the average CPU core-year request per application has remained relatively constant. Using this trend and the average request per application it is possible to extrapolate future demands. These projected needs estimates are described in Appendix A and shown in Figure A.4 for CPU demand, and in Figure A.5 for GPU demand. For example the estimated RAC demand in 2024 would be 520,000 CPU core-years and 14,400 GPU-years.

3.1.2 HPC Workload Analysis - Queued Demand

As investigated in the previous section, the RAC allocation data can be used as one measure of demand, however the data is based on the request and not actual system usage. The RAC allocation data also does not consider the demand for a large number of users that use the HPC resources, however do not apply for a dedicated allocation each year. A number of users also do not even submit RAC applications due to known resource constraints, instead e.g. use internationally available resources via their research collaborations. As the HPC systems all use a batch scheduling system, it is possible to examine the historical job characteristics; such as size, runtime, wait times, and memory to investigate trends as well as assess demand.

One such analysis is to investigate demand as a ratio of system capacity to queued workload.

This analysis has been performed on the GP systems (Graham, Cedar, and Beluga) and is described in detail in Appendix B.1. The results of the analysis show a consistent 4-5 times oversubscription of resources even during significant expansions which greatly increased the supply of resources. This observation can be interpreted to show that there is a large demand backlog that is not near to fulfillment, so that the constancy of the queue oversubscription even during capacity enhancements is primarily an indicator of researcher's tolerance for wait times, capped at the said level. With that a potential future indicator for measuring the fulfillment of researcher's compute capacity needs would be to see this indicator start dropping as new compute capacity is introduced.In summary, the consistent high backlog of HPC workload further highlights the need for significant resource expansion as demands are not close to being satisfied.

3.2 Research Competitiveness

Canada currently has 4 federation systems on the most recent (November 2021) TOP500 rankings of world's most powerful supercomputers - Narval (#83), Niagara (#127), Cedar (#137) and Beluga (#288). The current #1 entry, Japan's Fugaku, is 76 times faster than the top Canadian entry. There are currently three systems under construction in the US that are expected

to achieve greater than 1 ExaFlop (10^{18} FLOPs) this year, and two already operating in China.[11] The Exascale era in computing has arrived, but Canada has only barely made it to Petascale.

Admittedly these exascale class systems are very expensive resources to purchase and operate and not really within Canada's reach at this time. However, without continuous investment to keep updating resources Canada will very shortly have no systems in the TOP500, putting researchers at a significant disadvantage in international competitiveness. It is even an explicit performance indicator laid out by the ISED for the DRI program to maintain the number of ARC machines in the top250.[12] Following the performance growth rate which has been fairly consistent over the past 20 years, see Figure 1, it is expected that #500 on the list will be approximately 5 PFlops in 2025. If no new systems were installed before that time, Canada would only have 1 system, Narval, on the list and at 5.88 PFlops would be very near the bottom.

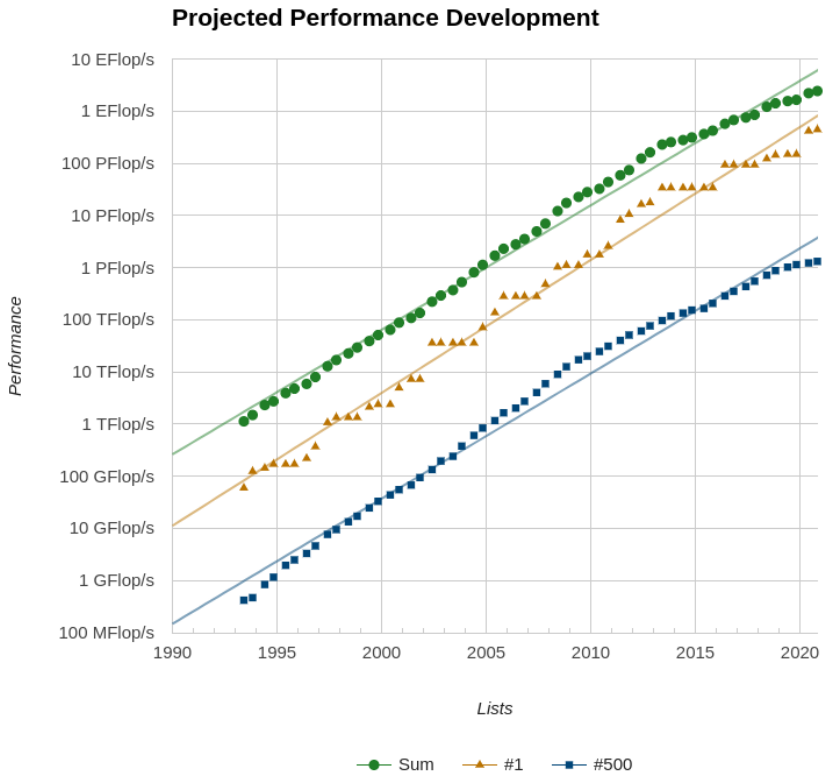

Figure 1 - Top500.org Historical Performance.

[11] The Next Platform: China Has Already Reached Exascale – On Two Separate Systems https://www.nextplatform.com/2021/10/26/china-has-already-reached-exascale-on-two-separate-systems/ (retrieved May 2022).

[12] ISED: Digital Research Infrastructure Contribution Program: Program guide https://ised-isde.canada.ca/site/digital-research-infrastructure/sites/default/files/attachments/DRIContributionProgram_ProgramGuide.pdf (April 2019).

Alternatively to looking at positions and number of systems in the TOP500, a likely more useful metric is to look at the aggregate compute performance provided per Country, normalized by GDP. This data is shown in Table 4 below comparing the November 2021 TOP500 entries to other countries, G7 highlighted in blue, on an aggregate basis. It should be noted that these statistics are based on the TOP500 results as is, without trying to parse multiple entries for the research or academic only systems. The table is ordered based on a country's compute performance per GDP. Canada comes in last in the G7 and 17th overall with an 18.0 aggregate Rmax TFLOPs per GDP. In order to keep up with our G7 peers and provide similar levels of resources to its researchers Canada would need to target a value of approximately 40 TFLOPs per GDP. To achieve this, at least double the currently available HPC capacity would need to be made available. This is another indicator of the insufficient resources available to researchers in Canada, putting them at a disadvantage in terms of research competitiveness compared to researchers in other countries.

Flop/GDP Rank	TOP500 Rank	Country	TOP500 Entries	Aggregate Rmax PFlops[13]	%GDP spend on R&D[14]	TFlops/ 10^9 USD $GDP
1	2	Japan	32	628.2	3.19	128.9
2	9	Saudi Arabia	6	55.3	-	80.5
3	6	South Korea	7	82.2	4.64	53.7
4	15	Finland	3	13.4	2.795	53.1
5	1	United States	149	986.5	3.067	50.6
6	4	Germany	26	181.4	3.19	49.1
7	5	France	19	117.0	2.196	45.3
8	21	Czechia	2	9.6	1.942	44.4

[13] TOP500 https://www.top500.org/statistics/list/ (November 2021).

[14] Organization for Economic Co-operation and Development (OECD): Data on 2019 Gross domestic spending on R&D https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm#indicator-chart (retrieved May 2022).

9	3	China	173	530.1	2.235	43.3
10	11	Netherlands	11	35.9	2.184	43.2
11	7	Italy	6	78.5	1.466	40.4
12	13	Switzerland	3	26.2	2.476	38.6
13	8	Russia	7	73.7	1.039	28.5
14	23	United Arab Emirates	2	9.0	-	23.6
15	18	Sweden	4	12.3	3.388	23.0
16	10	United Kingdom	11	54.9	1.756	20.8
17	**12**	**Canada**	**11**	**29.6**	**1.592**	**18.0**

Table 4 - Comparison of November 2021 TOP500 ranked by TFLOPS/GDP of Aggregate Rmax countries with G7 highlighted in green.

3.3 Storage Demand

HPC storage typically consists of many tiers utilizing differing technologies and policies. The faster more performant storage is considerably more expensive per TB but has the highest bandwidth and IOPS capability, whereas the most economical cost/TB (usually tape) is really only useful as long-term storage with minimal read/writes.

The current federation systems use the following storage tiering:

- **Scratch** - high performance parallel filesystem meant to be used as temporary storage. during calculations, fixed large quota per system, not backup up, purging policy in place
- **Home** - medium performance parallel filesystem primarily targeted for critical data only, fixed small quota per system, backed up locally.
- **Project** - medium performance parallel filesystem for group research data, RAC. allocatable, backed up locally.
- **dCache** - specialized storage for preliminary High Energy Physics, RAC allocatable.
- **Nearline** - low performance (primarily tape) longer term non-active data storage, RAC allocatable.

3.3.1 RAC HPC Storage

As with compute resources, three storage types are allocated during the RAC process; project, dCache, and Nearline. Storage results from the 2022 RAC are shown in Table 5. Unlike compute demand, in general the storage demand is mostly being satisfied, however it does continue to grow at approximately 20% per year; see Appendix A for an analysis of RAC storage demand over time. Like compute allocations, currently RAC allocations of storage are local to each of the systems.

Storage Resource	Allocated Capacity	Demand	Ratio
Project	52.1 PB	62.3 PB	**1.2x**
dCache	13.1 PB	13.1 PB	**1.0x**
Nearline	74.3 PB	74.3 PB	**1.0x**

Table 5 - 2022 Resource Allocation Competition HPC Storage.

3.3.2 Long Term & Archival storage

The current RAC process primarily allocates storage on a yearly basis, with some provisions for upto 3 years for certain projects. If a researcher fails to renew their allocation then there is a danger that the data could be removed. A solution for long-term and archival storage with RDM integration is requested by the research community, however is beyond the scope of the HPC Working Group, and will be considered as part of a future Storage Working Group.

3.3.3 Cloud Storage

Cloud storage is not specifically within the scope of HPC Working Group, however it is considered here as there may be future opportunities to use object storage as another or supplemental storage option directly on future HPC systems.

Supply is meeting demand for the various cloud storage options for RAC 2022, Table 6, however without capacity increases this will not be true in the future. Also the majority of the cloud

equipment including storage, as noted in Table 2, is over 5 years old and will soon need to be replaced. The popularity of Object storage is expected to increase especially if it is made available directly on the HPC clusters, which would make it a viable option to serve as an option for shared data storage.

Cloud Storage Resource	Allocated Capacity	Demand	Ratio
Volume/snapshot	3.0 PB	3.6 PB	**1.2x**
Object storage	7.3 PB	7.3 PB	**1.0x**
Shared Cloud	1.6PB	1.6 PB	**1.0x**

Table 6 - 2022 Resource Allocation Competition Cloud Storage.

3.4 Researcher Specific Demands for HPC Resources Provided by the Alliance

The Researcher Council's Priority Needs report, Researcher Needs Assessment consultations and report, and Current State reports provided the Alliance with an extensive understanding of researcher's needs for the Canadian DRI ecosystem going forward. The researcher community clearly articulated the need for expanding the national advanced research computing capacity by building on the Compute Canada Federation's HPC infrastructure.

As part of the Research Needs Assessment consultation process researchers were asked to submit position papers. Some of these papers provided forecasted HPC resource estimates as well as operational requirements which are summarized here for consideration.

- "Large-Parallel Supercomputer Simulations - Frontiers in Canadian Research"[15] proposes the need for a dedicated ~10k node homogenous supercomputer for large scale simulations. Specifically they identify a need for a homogenous capability system with 5-8x the capability of the existing Niagara system in the 2023 time-frame. This would correspond to a 30-50PFlop/s (390-650k CPU core) class system.

- "DRI Needs Assessment for the Computational Fluid Dynamics (CFD) Research Community"[16] which as a community were allocated 28,777 CPU core years in 2020 emphasize the need for "the renewal of large, parallel, homogeneous cluster resources, with high performance IO servers and associated network facilities, is important for on-going and future CFD research."

- "Digital Research Infrastructure for Canadian Astronomy"[17] provides a general estimate that the astronomical community will need 100 Pflop years of GPU time (5000 GPU-years), 100 Pflop years of CPU time (1.3M CPU-core years, i.e. more than 4x the current total core capacity at Federation's central systems), and 75 PB of online storage by 2025 excluding resources for specific projects, such as the SKA Regional Centre.

- "White Paper on Canada's Future DRI Ecosystem Subatomic Physics in Canada Institute of Particle Physics (IPP) and Canadian Institute of Nuclear Physics (CINP)"[18] describe their requirements for 24x7 operation and support of the HPC systems.

- "Canada's Future DRI Ecosystem: AI Research Needs"[19] outlines the specific needs for the AI research community, including adding specialized capacity for AI research.

---

[15] R. Fernandez et. al.: Large-Parallel Supercomputer Simulations – Frontiers in Canadian Research; Position Paper submitted to Alliance https://alliancecan.ca/sites/default/files/2022-03/ndrio_wp_on_niagara_scale_systems.pdf (December 2020).

[16] CFD Society of Canada: DRI Needs Assessment for the Computational Fluid Dynamics

(CFD) Research Community; Position Paper submitted to Alliance https://alliancecan.ca/sites/default/files/2022-03/dri-needs-assessment-of-the-cfd-community.pdf (December 2020).

[17] C. Lovekin et. al.: Digital Research Infrastructure for Canadian Astronomy; Position Paper submitted to Alliance https://alliancecan.ca/sites/default/files/2022-03/white-paper-on-dri-for-astronomy.pdf (December 2020).

[18] Institute of Particle Physics (IPP) and Canadian Institute of Nuclear Physics (CINP): White Paper on Canada's Future DRI Ecosystem - Subatomic Physics in Canada; Position Paper submitted to Alliance https://alliancecan.ca/sites/default/files/2022-03/sap-white-paper.pdf (December 2020).

[19] Vector Institute: Canada's Future DRI Ecosystem: AI Research Needs; Position Paper submitted to Alliance https://alliancecan.ca/sites/default/files/2022-03/canadas-future-dri-ecosystem_-ai-research-needs.pdf (December 2020).

3.5 External Projects Associated with the Alliance

There are a few major research projects that have their own significant resources for computation funded outside the Alliance. These projects also have components that use or are expected to use Alliance resources so keeping abreast of these projects is important to ensure expectations and planning are aligned. Some of these projects also have significant networking requirements making coordination with NREN/CANARIE also important.

3.5.1 High Energy Physics

TRIUMF is Canada's particle accelerator center, and one of the world's leading laboratories for particle and nuclear physics and accelerator-based science. They operate several dedicated clusters and storage to support major research initiatives such as the ATLAS Tier-1 center. Significant analysis is performed using Alliance resources which require specialized storage (dCache), networking, and services on the systems that support this work. Currently this research is supported on Cedar, Graham, and Arbutus with dCache allocations of ~13PB. The demand for dCache storage is expected to grow at a rate of 20% per year. See the position paper “Role of TRIUMF within the Digital Research Infrastructure Ecosystem” for more details.

Other large computing resource commitments have been made to the Belle II, T2K, IceCube, SNOLAB, and GlueX experiments, which have significant Canadian contributions. The majority of the compute resources are provided by the Alliance resources, currently ~10,000 CPU-cores per year and over 10 Petabytes of storage on the various platforms.

3.5.2 Square Kilometer Array (SKA1)

A SKA Regional center has been identified as one of the top priorities for Canadian astrophysics in the 2020 LRP (Barmby, Gaensler et al. 2020). These data centers will collectively process the 5TB/s produced by the telescope. Canada is committing to provide 6% of the global capacity required. The steady state scenario starting in 2029 is for Canada to provide:

- 9.7 PFlop/year of processing; (126,000 CPU core-years or 485 GPU-years)
- 238 PB/year of online storage which represents one year’s worth of data.
- 654 PB/year of nearline storage growing each year.

As this project is just getting started it is not known how these resources will be provided and how it will integrate with the Alliance National platforms.

3.5.3 Pan-Canadian AI Compute Environment (PAICE)

The three Canadian AI Institutes, Vector, Mila, and Amii, in coordination with CIFAR and the Alliance are in the process of investing in 3 new systems designed specifically for AI research workloads. These systems aim to provide an increase of ~7250 GPUs over the next five years to significantly increase the specialized computing capacity available to researchers. The primary users will be affiliated with the three institutes, however some capacity will be made available for the broader research community. Alliance coordination will be key to ensuring these systems are integrated into the national ecosystem.

3.6 Contributed/Integrated Systems (CFI)

Researchers apply for and get funding for systems. This is mainly through the CFI JELF and IF competitions but also from other funding sources. The Alliance has published guidelines[20] on how these systems should be considered. In some cases systems may have unique hardware or have specific use cases that are incompatible with the national platform and therefore are not considered contributed and instead are operated by the research group or some other local support team outside of the national platform.

Contributed systems that are similar or even identical to the systems in a national platform if integrated properly, can require no additional staff time to operate. Increasing a cluster by a few more nodes does not significantly increase the overall system administrator workload if they are fully integrated into the host system and not significantly specialized. Expanding the available resources benefits the entire user community while providing the contributing group their required resources.

There are increased costs for some software, such as the job scheduler which is licensed on a per node basis but these costs are very minimal. There are power and cooling costs for the hardware which would occur if the hardware is contributed or not. Contributed systems can leverage other existing infrastructure in the national site like networking, login and administration nodes allowing the contributing group to purchase more computing infrastructure than they would otherwise be able.

Some funding agencies are negatively reviewing or even denying applications for compute with the justification that the research can be completed using the national platform. This requires the national platform to be sufficiently large to support those workloads.

The timeline of contributed hardware is important to consider. Adding new hardware to a system that is already over three years old causes an imbalance in capabilities and operational lifetime across the mix of hardware. Ideally contributions occur early in the main system's operational window. This could require later contributions to be added to a different, newer system.

[20] Alliance: Renewed Policy Guidance on Integrated Hardware (Contributed Systems) https://alliancecan.ca/sites/default/files/2022-03/Renewed-Policy-Guidance-on-Integrated-Hardware-Contributed-Systems.pdf (October 2021).

The funding for contributed systems frequently comes with a provincial matching component. This makes it hard if not impossible to add the contribution to an out of province national system even if that would be the best location to receive the contribution.

3.7 International Collaborations

The Alliance should investigate possible collaborations that may allow Canadian researchers access to use of specialized and/or very large systems such as in the US (e.g. National Science Foundation's ACCES systems or Department of Energy's supercomputers) or in Europe (e.g. European Union's PRACE or EuroHPC systems).

3.8 Alignment with Funding Calls, Tri-Council & other Initiatives

The Alliance should ensure its investments into HPC and other DRI resources align with any Tri-Council inititivaties for 22-27 that may require significant resources.

3.9 Cloud Resources

"Cloud" workloads and strategy are not thoroughly investigated within this document as there is a dedicated Alliance working group developing a Cloud -specific strategy. Considerations for various workloads, including using Cloud resources for HPC compute workloads will be part of that strategy. Options such as using a framework like Magic Castle to provide users a similar HPC environment for burst loads or handling emergencies are likely worth investigating. The hardware components of the existing federation cloud resources are highlighted in this report for completeness.

4 Future HPC Architecture

4.1 Maintaining baseline Capabilities - Immediate needs

As described in Section 2.1, the current ecosystem has a total capacity of approximately 300,000 compute cores and 3,000 GPUs for a combined compute capability of 40 PFlops. The current infrastructure however varies in age from equipment that was put into service over 5 years ago to recently installed, as shown in Table 1. Staggering equipment purchases over time is typically a good strategy as it allows for incremental replacement and newer technologies to be deployed, however it is based around the assumption that after a system's typical 5 year lifespan, that equipment will then be replaced with new equipment.

Currently both the Cedar and Graham systems have significant resources that are now beyond due for replacement. Specifically, 60,000 CPU cores, 900 GPUs, and the majority of the primary storage will have passed 5 years in 2022. By the end of 2024 all the existing deployed federation resources, with the exception of the recently deployed Narval system, will be over 5 years and due for replacement. Without immediate investment to replace these aging resources, the currently available capacity will soon start to shrink significantly.

It is possible to operate older systems past the typical 5-year lifespan, however it can be quite expensive as extended warranties are required, maintenance and failures increase, and the performance provided per Watt is much lower than for newer more efficient equipment.

This is especially true for rapidly advancing technologies like GPUs. For example a 5 year old NVIDIA P100 has the same 300W power draw as the latest generation A100, however the performance difference is 4.3 TFlop/s vs 19.5 TFLops, ie a ~4.5 times better Flops/W. It also has double the amount of memory, higher memory bandwidth, and a significant number of features not available on the older hardware.

4.2 HPC Capacity Scenarios

Considering the forecasted demand for HPC resources outlined in Section 3 a series of scenarios are proposed. Summarized in Table 7 and costed in Table 8 the four scenarios considered are:

I. Critical Replacement Needs to Maintain Capacity.
II. Additional capacity requirements to fulfill 100% of forecasted CPU and GPU RAC demand in RAC2024.
III. Additional capacity requirements to fulfill 70% of forecasted CPU and 50% of GPU RAC demand in RAC2024.

IV. Additional capacity requirements to move Canada closer to the middle of the pack in TOP500 aggregate Flops/GDP among G7, i.e. doubling of the current capacity.

4.2.1 Replacement and Expansion Scenarios

	Current total capacity (March 2022)	**I. Critical replacement** needs in 2022-2024	**II. Additional** capacity needed to completely fulfill projected RAC2024 demand	**III. Additional** capacity needed to 'partially fulfill' projected RAC2024 demand	**IV. One-time additional** capacity boost needed in order to move Canada closer to the middle-of-the-pack in TOP500 aggregate Flops per GDP rankings among G7 countries
CPU (cores)	304,112	234,000	224,000 (total of 520,000 in 2024)	68,000 (70% fulfillment of the demand)	Doubling of compute power needed, very roughly equivalent of adding e.g. roughly. 300,000 CPU cores
GPU (units)	3,278	2,200	14,400 (total of 17,500 in 2024)	5,650 (50% fulfillment of the demand)	Doubling of compute power needed, very roughly equivalent of adding e.g. roughly 3,000 GPUs
Active Storage (TB) (project, dCache)	100,000	80,000	48,000 (assuming 20% annual growth, totaling 148,000 in 2024)	48,000	48,000

Nearline (TB)	88,000				
Networking	100Gbps (each site)		200Gbps[21, 22]		

Table 7 - Summary of HPC Capacity Replacement & Expansion Scenarios.

Table 7 above provides the current baseline and estimates for multiple future hardware (first column) requirement scenarios (first row). CPU core compute needs (including CPUs, servers, memory, internal high-speed network interconnects, and scratch storage) are measured using CPU core count as a proxy. GPU accelerated compute needs are measured using a single GPU unit as a proxy. Active storage needs include 'project' and 'dCache' level storage, but do not include scratch, or nearline storage capacity. Nearline storage for repository and archival storage needs, provided by a combination of disk and tape storage systems, are listed separately. The last architectural component is external networking bandwidth requirements for connecting to national and international networks.

The second column indicates for reference the current aggregate compute capacity in Alliance Federation's (former Compute Canada Federation) main compute systems as of March 2022. The data is based on Table 1 above that discusses the details of compute resources per system, including the most recent addition, Narval at Calcul Quebec.

The third column presents the future projected Scenario I, presenting the short-to-mid-term critical replacement needs for the current infrastructure listed in the previous column. Scenario I considers all infrastructure that is past its five-year lifetime as of the end of December 2024, and thus needs to be replaced. Due to the commissioning timing as discussed in Table 1 above, this translates into having to replace the capacity provided by all systems except Narval in the next two and a half years. In this critical needs Scenario (which only maintains the current baseline) no new compute capacity per se is added, the improved compute efficiency thanks to technological advances notwithstanding.

The fourth column presents the future projected Scenario II, the additional new capacity needed in order to fulfill 100% of the projected needs in RAC2024 allocation competition. The projected needs are based on estimates as discussed in Appendix A and shown in Figure A.4 for CPU

[21] According to predicted LHC and BELLE-II requirements, 100Gbps will be required by 2025, and 200Gbps by 2027. This is **in addition** to the existing site networking. Extra capacity above 100Gbps is already required at Cedar (connecting the Tier-1 site).

[22] Extra networking capacity must be coordinated with the respective NREN and CANARIE.

demand, and in Figure A.5 for GPU demand. The infrastructure needs in Scenario II are notably in addition to Scenario I, which would only maintain the current baseline capacity.

The fifth column presents the future projected Scenario III, the additional new capacity needed in order to 'partially fulfill' the projected needs in RAC2024 allocation competition. For CPU demand, 'partially fulfill' assumes that 70% of the total RAC2024 request would be satisfied (in comparison, at RAC2022 the fulfillment was at 54%, see Fig A.4). For GPUs, the 'partial fulfill' assumption is that 50% of the demand would be provided (in comparison, at RAC 2022 the GPU fulfillment was at 24%, see Figure A.5). The infrastructure needs in Scenario III are a scaled back version of Scenario II, and again would be in addition to the critical needs of Scenario I. Interestingly the need for additional new CPUs in this scenario III is 'only' roughly 68,000 cores due to the fact that the recent introduction of Narval lifted the fulfillment rate in 2022 from roughly 54% to 68% (not included in Figure A.4) so that a relatively modest increase in number of CPU cores over the next two years is enough to keep up and lift the rate to 70%.

The sixth and final column presents the future projected Scenario IV, the additional new capacity needed for lifting Canada's ARC and HPC compute infrastructure capacity closer to the middle of G7 from the current bottom among G7 position, as measured by aggregate Flops per GDP (See Table 4 for details). In practice this would amount to (at least) the bare minimum of doubling the current ARC infrastructure capacity (in 2022), e.g. for CPUs growing the capacity from roughly 300k CPU cores to roughly 600k CPU cores as an immediate priority investment. As with Scenarios II and III, the infrastructure needs in Scenario IV are in addition to the critical needs Scenario I.

4.2.2 Cost Estimates for Scenarios

	Current total capacity (March 2022)	**I. Critical replacement**	**II. Additional** capacity needed to completely fulfill projected RAC2024 demand	**III. Additional** capacity needed to 'partially fulfill' projected RAC2024 demand	**IV. One-time additional** capacity boost needed in order to move Canada to the middle-of-the-pack in TOP500 aggregate Flops per GDP rankings among G7 countries
Cap$ CPU	-	$69M	$65M	$20M	$87M

Cap$ GPU	-	$46M	$303M	$119M	$63M
Cap$ Storage	-	$15M	$9M	$9M	$9M
Total Cap$[23]	**-**	**$130M**	**$378M**	**$148M**	**$159M**
PFlops(Rpeak)	40	58	298	115	82
Space (#racks)		162	500	192	220
Power (MW)	5.0	5.0	14.3	5.4	6.0
Util $	$5.0M	$5.0M	$14.3M	$5.4M	$6.0M
Ops $/year	$29M	$29M	$7.2M	$7.2M	$7.2M
Total Ops$[24]	**$34M**	**$34M**	**$21.5M**	**$12.6M**	**$13.2M**

Table 8 - Costing estimates for scenarios outlined in Table 7.

Table 8 summarizes the approximate costs to procure the four scenarios outlined in Table 7. The capital costs of CPU and GPU systems are estimated based on working out a standardized node cost that includes a portion for CPU, memory, network, and scratch space and then scaling by the number of CPU-core and GPUs. It should be noted that technology prices are highly fluid and can change rapidly with the prices shown here being illustrative rather than prescriptive. The guiding principle is to maximize capacity for researchers at the time of RFP with the predetermined budget available. Detailed descriptions of the methodology and pricing used for cost estimations are presented in Appendix E.

Operating costs are estimated using the 2022/23 operating budget for the existing 5 host sites and ~200 staff of ~$34M. Breaking out the utility costing of ~$5M for ~5MW of power of the existing systems, it is possible to estimate the costs of the 4 Scenarios. The same approach can be done for estimating staffing cost increase incurred from the 4 scenarios. Current staffing levels across the federation are 114 research support and 69 technical. Increasing system sizes would require only a modest increase in technical support staff, 1-2 technical staff per system, due to

[23] Appendix E - Reference Design Costing

[24] Total federation operations costs based on the 2022/23 budget including personnel.

the replicated nature of the hardware. Significant resource increases however would lead to proportionally larger support requests and thus the number of research support staff required would need to be increased accordingly. For the expansions considered in Scenarios II-IV, it is estimated that increases of approximately 10 technical and 50 research support staff would be needed based on the 2022 budget with an average salary cost of $120k/person, for a total additional cost of $7.2M each year.

In summary the critical baseline costs to update systems to maintain existing capacity by the end of 2024 would be $136M capital and then $34M/year in operations. To expand the systems based on the demand scenarios II to IV, the additional capital costs would be $378M, $148M, and $159M respectively, with additional operating costs of $21.5M, $12.6M, and $13.2M respectively.

For example the total estimated cost to provide double the current capacity in a fully modernized system by the end of 2024 would cost $295M Capital with an operating cost of $47.2M/year.

4.2.3 Cloud Replacement Costs

	Critical replacement needs in 2022-2024
Cap$ CPU	$5M
Cap$ GPU	$2M
Cap$ Storage	$2M
Total Cap$[25]	**$9M**

Table 9 - Costing estimates for replacing federation cloud resources.

Table 9 summarizes the approximate costs to procure replacements for the existing federation Cloud infrastructure, outlined in Table 2 in Section 2.1. All of the Cloud equipment will be over 5 years old by the end of December 2024 and in critical need of replacement. The capital costs are estimated using the same methodology as the HPC systems, using a standardized node cost that includes a portion for CPU, memory, and network and then scaling by the number of CPU-core

[25] Appendix E - Reference Design Costing

and GPUs. Descriptions of the methodology and pricing used for cost estimations are presented in Appendix E. Discussion of demand and expansion of cloud resources is left to be considered by the Cloud Working Group.

4.3 HPC Storage Architecture

Storage and data management extends into all facets of DRI and as such a forthcoming Data Architecture Strategy Working Group will consider it holistically and in greater detail. As such only the HPC storage specific requirements are considered here, acknowledging that they will need to interact with a larger storage strategy.

4.3.1 Storage Hierarchy

The demands for allocatable storage (active and nearline) have been outlined in Section 4.2's capacity scenario. Storage however is much more complex than just total capacity as performance, permissions, policies, and need for backups and data security also come into play. As briefly mentioned in Section 3.3, the existing federation systems use a tiering approach which is fairly common on most HPC systems. The storage tiers are revisited below with some considerations for the next generation systems.

- **Scratch** is sized based on system, ~10x system memory size, and needs to be sufficiently performant, i.e. both bandwidth and IOPS, with a parallel file system available to all compute resources. The use of all solid state (SSD or NVME) drives for scratch space is becoming more popular and affordable. Purging and quota policies need to be enforced to ensure users are using correctly. Scratch space is considered temporary and not backed up.

- **Home** quotas are small so capacity is not usually a primary concern, however the number of small files can be very high. This data is backed up and possibly could be a good target for cross-site backups due to small size but important files.

- **Project** space needs to be a balance of performance and capacity so also needs to be a parallel file system, however as $/TB typically is the driving factor it can largely be composed of mechanical spinning disks. This space is the primary active storage that is allocated to the user so proper quota and management policies need to be in place to avoid misuse and long term "parked" data which should be migrated (preferably automatically) to a nearline storage layer. Project space has been typically allocated on a yearly basis and is backed up locally, so backup capacity needs to be costed as part of

any project storage planning. In its current design as a single multi-PB space it is likely not feasible to consider full project cross-site backups.

- **Nearline** is an allocated storage tier that is primarily for long term storage and is primarily designed for capacity. Currently users are required to explicitly control the migration of data to nearline, however it is probably preferable that this migration be automated based on policies considering how long data has been static. This could greatly reduce the amount of parked data on spinning disk, however it doesn't force users to curate their own data so could lead to higher storage requirements. Within the federation there are currently multiple nearline solutions in operation, IBM HPSS, Spectra Blackpearl, and a homegrown robinhood based solution.

4.3.2 Specialized Storage

Beyond the common storage tiers described above, there are also more specialized storage components that will need to be considered or find alternatives for in next generations systems.

- **Burst Buffer** - Currently Niagara has what is essentially an even higher performance scratch specifically for High IOPS workload. It is made up of NVME drives and uses an NVMESH technology to provide a very high performance shared storage tier. At the time this was a relatively common approach on large HPC systems as solid state drives were very expensive, however now it is probably not necessary as the full scratch space can be built out of sufficiently performant storage.

- **dCache -** Used by ATLAS, T2K, SNO+, DEEP, (a few other small HEP projects).

- **CVMFS -** Currently used for software delivery across the federation and being piloted for common shared data sets.

- **Cloud Object Store -** A cloud storage technology with standardized non-posix interfaces, such as Swift/S3, that could serve as a common location for datasets and possibly shared user data available across systems.

4.3.3 Other Storage Design Considerations

Once a storage system is deployed it commonly has a longer lifetime than the aforementioned 5 years and may serve multiple generations of compute systems. This is especially true for capacity focused tiers and tape based systems. Unlike compute changes, completely replacing a storage system, especially large capacity storage systems like project can take a long time and can cause

a great deal of user inconvenience. For these reasons storage designs and lifecycles need to be considered carefully. Considerations for in place expansion and upgrades can reduce costs, avoid data migration, and allow capacity to be purchased as required and not all at the outset.

4.3.4 Other Data Management Considerations

The following broader storage considerations are only briefly mentioned here and should be more fully investigated as part of a national data management strategy.

- One of the biggest stumbling blocks right now that keeps users from easily moving workloads and computing between systems is that the data (active, nearline, and backups) storage is all local. There are tools to move data between sites, but nothing automatic or high-availability (HA), and the transfer of or replication of the data is left to the user.
- Currently there is a significant amount of data not accessed in 6 months or more left "parked" on local spinning disk. Also as there is currently no standardized way to centralize common use data such as large data sets, significant duplication can result even within a local site.
- HPC requires local high performance storage, however a coordinated strategy to provide at least some (non-local) HA project type storage would be beneficial as well as increase the resiliency of the HPC ecosystem.
- Long term (>1year) storage planning and allocation. This is also an area where RDM and ARC should be coordinated and where storage policies, planning, and investment would significantly improve the options for researchers.
- Options for secure storage and policies to govern its use and access.
- Planning for Backups (local & offsite)

4.4 Next-Generation HPC Systems Design Considerations

Section 4.2 described the demand requirements based in general capacity terms, i.e. CPU and GPU core-years and generic storage. HPC systems however are typically built to target specific workloads from very specialized capability class machines (such as ORNL's Frontier ExaScale system) for massively parallel workloads to modest commodity hardware based systems that primarily provide capacity computing targeting a wider range of mixed smaller scale workloads.

To serve the wide variety of researchers and workflows, the current ecosystem in Canada contains three classes of systems. Four "GP" or general purpose heterogeneous clusters which serve the majority of small to medium HPC workloads as well as provide the GPU resources, the Niagara "LP" large parallel homogeneous system which is designed for the large scale massively parallel workloads, and the Arbutus Cloud system.

4.4.1 Current HPC Workload Trends

To inform future HPC systems designs and determine the appropriate balance of general and specialized HPC resources, investigating job characteristics from the existing HPC infrastructure can provide insight into workload trends. The Data Analytics National Team (DANT) records and analyzes job scheduler information making it possible to examine job characteristics. When characterizing workloads based on CPU and GPU job sizes as well as different job types on each of the HPC systems, fairly consistent trends emerge. Below is a summary of the key observations with the data provided in Appendix B.

Small Workloads

When looking at job size distributions on the General Purpose Systems, approximately 50% of the CPU job usage is for single node or less, where single nodes are 32 to 48 CPU cores. The number of CPU cores in a single system is continuing to rise so the demand for single and small node jobs is expected to continue, indicating less need for messaging related high-speed internal networking fabrics. Such high-speed internal networking could still potentially be needed on HPC systems supporting smaller workloads due to high-speed storage needs. With greater workload per node, the amount of storage traffic will increase which will require significant networking. Also with fewer larger nodes the cost of adding high-speed networking (being a function of the number of networking cards, and switch connections) drops significantly as a fraction of the node cost.

Memory Requirements

Approximately 85% of jobs requested 4GB/core memory or less across all systems. There exists some demand for large memory per core nodes, however it is relatively small so resources for these workloads should be scaled accordingly. With the trend of increased numbers of CPU cores per node, to maintain 4GB/core memory will require the total memory per node to be increased as well, thus also providing larger memory nodes for those jobs that might require it. Such increase in total memory per node due to increasing core counts must also include architectural considerations supporting increased memory bandwidth needs.

GPU Workloads

While the demand for GPU resources is high, the workloads appear to still be very much dominated by smaller 1-4 GPU jobs. Over 50% of GPU jobs used only 1 GPU and 95% used 4 or less GPUs. i.e. single node. With the rapid increases in per GPU performance and still somewhat limited workloads and codebases suited to use them efficiently (even at the level of leveraging a single GPU) this trend is expected to continue at least in the short term. Moreover, since preliminary data indicates that single GPU utilization is not as high as it could be, new systems should consider investigating GPU subdividing to allow more efficient use of GPUs by smaller workloads sharing a single GPU device. Subdividing the GPU administratively to smaller

functional parts would provide end-users smaller GPUs that could be leveraged more efficiently without a need for major programming efforts by the end-users. Technologies like NVIDIA's Multi-Instance GPU (MIG) allows its latest generation of GPU's to support this.

Large Parallel Workloads

Unlike on the General Purpose Systems, on Niagara 60% of total usage is for jobs requiring 512 cores or more. As this system was designed for large parallel jobs this is to be expected, also clearly indicating the need to provide at least one large system to support such massively parallel workloads. This need has been identified in position paper submissions from domains such as Astrophysics & Computational Fluid Dynamics which are still heavily reliant on massively-parallel homogenous CPU based supercomputers.

> "We recommend an ambitious renewal and expansion path for the large-parallel simulation capability in the Canadian DRI ecosystem, to build on existing progress, and to expand the associated specialized human resources required for scientific discovery."[26]

The research community's need for a next-generation homogenous CPU capability is clear, however, going forward there should be consideration for also having a portion of a future large parallel system also have accelerators. This hybrid approach is being adopted in systems such as the new LUMI system being deployed by EuroHPC to ensure support for a wider range of existing and developing codebases.[27]

Resource Allocation Competition

The RAC process typically allocates 80% of resources available each year. With the exception of the Niagara system, only about 60% of those resources are being used by those RAC accounts. The resources are not being wasted, but instead are being used mostly by the default users who typically do not have dedicated allocations. This is potentially an indication of the demand for flexible, smaller scale resources and should influence future allocation policies and strategies.

---

[26] R. Fernandez et. al.: Large-Parallel Supercomputer Simulations – Frontiers in Canadian Research; Position Paper submitted to Alliance https://alliancecan.ca/sites/default/files/2022-03/ndrio_wp_on_niagara_scale_systems.pdf (December 2020).

[27] LUMI: LUMI's full system architecture revealed https://www.lumi-supercomputer.eu/lumis-full-system-architecture-revealed/ (retrieved May 2022).

4.4.2 Balance of Accelerator and CPU Computing

The use and demand for accelerators (most commonly GPUs) in HPC systems has been increasing as the performance and efficiency in terms of FLOPS/$ or FLOPS/W of accelerators is advancing at a higher rate than traditional CPUs. For some workloads GPUs can be extremely efficient and newer designs and programming environments are expanding their suitability to a wider range of applications. However they can also be very poorly used so should not and can not be seen as complete replacement to CPU computing. Many problems and applications can not be easily (or at all) recast or ported to GPUs. For this reason TACC's Frontera, an NSF-funded system for general academic research, is primarily a pure CPU system with additional GPU capability.

It should be noted that in coordination with the Alliance, the three Canadian AI institutes will be deploying PAICE which consists of 3 new specialized systems designed for AI workloads, so likely GPU centric systems. There will be some availability to the wider research community to use these systems as well, which can only help with the increasing demand for GPU computing resources.

4.4.3 Next-Generation Systems Configuration

In analyzing the current systems usage it is clear that there is still a need to have GP systems for small CPU and GPU jobs, and at least one LP system to support the large scalable parallel workloads. Demand for GPU's continues to grow, however it should be noted that the latest systems installed, Beluga and Narval, are both GPU heavy resources so monitoring their usage can help determine the future balance of CPU to GPU computing resources. The current ratio of GP to LP resources seems to be appropriate based on current demands. With multiple GP system deployments, replacement/upgrades can be planned so that there is continuous access to the latest technologies across the National Platform while maintaining and preferably expanding capacity. For the LP system it should be sufficiently large enough to serve for multiple years as there will likely only be one of these systems deployed in the ecosystem every 4-5 years.

4.5 Networking

HPC networking typically consists of two components, an internal high-speed, low-latency network used for job-to-job communication and file IO, and an external network used for data movement and internet connectivity into the site.

4.5.1 High Speed Networking

The internal high performance network (e.g. Infiniband) is commonly a lower latency non-ethernet fabric with a wide range of network topologies and usually is the single primary component that differentiates an HPC system from typical enterprise and cloud based systems.

The use of high speed networking was originally required primarily for node to node parallel job communication, however with the trend towards much larger high core count single nodes it has also become important for file IO performance. The trend toward larger nodes has also reduced the cost of the networking as a percentage of the total HPC system due the reduced number of networking ports required. As such even GP systems which target smaller jobs should be deployed with some level of high speed networking. See Appendix C for more detailed considerations around high speed networking.

4.5.2 External Networking

External WAN, Internet-protocol based, networking commonly has a high bandwidth to handle the data movement requirements. All the current host sites can connect to each other using 100 Gbps links, however the commercial connectivity, i.e. internet link capacities vary at each site. Discussions around external networking also have a direct bearing on CyberSecurity planning as these connections are how users access and work with the systems remotely.

- 200 Gbps to CANARIE required today for sites hosting data intensive workloads (Atlas)
- Plan 200 Gbps upgrade to support sites that will implement the “Data Management & Resilience” plan.
- Coordinate WAN network upgrades with the NREN and CANARIE

See Appendix C for more detailed discussions around external networking considerations.

4.6 Next-Generation Procurement/Deployment

4.6.1 Multi-Year Capital & Operating Investment & Planning

Modern HPC infrastructure involves substantial capital and operational costs, and staffing needs that should all be considered and secured at the time of the infrastructure Investment. Capital funding for DRI infrastructure has often come in sporadic fashion, commonly addressing an immediate need every 4-5 years, resulting in large purchases with multi-year gaps in between. Moreover, there has been no long-term, sustained funding commitment from the funding agencies, preventing long-term multi-generational and systematic planning. Systems were built to address a specific need and not as part of a planned, sustained ecosystem which considers the lifecycle and refresh of that system. Also the operating costs were not considered as part of the funding call, but instead funded through a different program. As such the procurement of new HPC systems and infrastructure has been disconnected from the funding of corresponding operational costs and concerns, even though these two are closely connected.

The staggered times for commissioning and decommissioning HPC systems, while beneficial from continuity of service delivery and technical updates point of view, can result in difficulties in aligning required capital and operating budgets. The Government of Canada and ISED have recognized this problem and part of the Alliance's mandate going forward is to replace CFI's MSI operational funding envelope for HPC operational costs so that going forward major DRI capital and operational funding decisions will be done holistically. The operational funding planning also needs to also consider costs related to contributed systems.

This planning/procurement process should include regular reviews to make sure investments are serving the needs of the research community, identifying deficiencies or shortfalls, as well as adapting to technological changes which are frequent and to be expected in HPC.

The advantages of a continuous investment approach across multiple sites is that it allows for the ecosystem as a whole to provide continuity of services while allowing the resources to be refreshed integrating new technologies and capabilities.

4.6.2 HPC Deployment & Equipment Lifecycle Considerations

4.6.2.1 Single Large Deployment

In previous generations "new" HPC systems were typically purchased all at once in a single large deployment as a result of the capital funding model. Purchasing a system all at once does have advantages in that it provides a larger homogenous system that has a greater initial resource and commonly better value for the money as large deployments can incentivize vendors to provide better pricing. As such, this type of deployment would likely suit when deploying a cluster for large parallel workloads. The risks in this approach are that if the initial design is not well suited to the researcher or those needs and or technology changes/improved over the typical 5 years a system is in operation then there is little opportunity to adapt the system. Also the systems can be less flexible to upgrade as that was never considered during the design phase.

4.6.2.2 Staged Deployment

An alternative to one off system deployment would be to purchase a system in stages with continuous expansion / upgrades planned. This type of system would be designed with expansion in mind allowing compute and storage to be expanded in time and ensuring core components like the core networking can accommodate the future growth. This type of system

can be more nimble to researcher needs as well as quicker to incorporate newer technology as the stages are typically much shorter. This approach has been used successfully for systems like the NASA Pleiades Supercomputer. A challenge with this approach is that it results in a heterogeneous system which can contain many generations of different hardware which can be harder to schedule and operate as well as can be more complicated for the end user. This has been an issue with the GP systems currently in operation, especially with the contributed

components, however this could be somewhat mitigated by ensuring the upgrade stages are significantly large to not create too much heterogeneity. Another issue that can arise in the staged approach is that later stages may be constrained by the initial systems physical and datacenter design, such as conforming to a maximum power/cooling envelope which could limit future choice.

4.6.2.3 Staggered Deployments

With a national platform consisting of multiple sites/systems the optimum scenario is likely one in which you balance the two approaches by deploying new full systems staggered over time and sites. This allows the platform to be continuously refreshed, accommodating changes in technology and research while avoiding the issue of high heterogeneity which can be difficult to manage and increases complexity for users. From a financial perspective this approach can be advantageous as the yearly expenditure on new capital is more consistent and can provide more opportunities for match as purchases are distributed more regularly over a longer time period.

4.6.2.4 Lifecycle

In each of the scenarios discussed the full lifecycle of a system's components should be considered during the design phase. The typical lifetime of a HPC compute component is 4-5 years as benefits of obtaining newer technology surpasses the benefit of continuing to operate the older technology, however that does not mean that the hardware no longer has value. Many smaller scale and grid-computing type workloads, and even more cloud type workloads are not performance limited and as such can utilize slightly older less performant hardware. Keeping this lifecycle plan in mind can potentially help to reduce the environmental impact by using the equipment longer as well as can provide more resources without additional capital expenditure.

4.6.3 Flexibility in Design

When serving a wide research community with varying needs, such as highlighted in the workload analysis, and deploying rapidly changing technology purchased at different times it is important to take a flexible approach to systems design. Common interfaces and operations for users are important, but should not be so rigid as to constrain offering the best technological solution or to attempt to solve all workflows using the same one size fits all approach. A benefit to considering a national HPC ecosystem is that not all components of the platform need to be the same, allowing for specialization where needed while providing access to all Canadian researchers.

4.7 Datacenter & Environmental Considerations

HPC datacenters generally have more demanding requirements for power, cooling, and floor-space than enterprise IT datacenters, and as such each of the current Federation HPC systems are located in specialized HPC datacenters at each of the host sites. Current host sites have

varying capacities for expansion, which are outlined in detail Appendix D. Overall physical space is not the limiting factor, however power availability and cooling infrastructure as

HPC equipment is becoming more power-dense.

4.7.1 Datacenter Support for New & Expanded Systems

Depending on the timing for retirement of existing systems and sizing of new systems, such as to support the expansion Scenarios outlined in Tables 7 and 8, there will likely be need for datacenter modifications and possibly new host sites. Even without expansion, newer HPC compute technologies, such as upcoming GPUs and large core count CPUs, are very compute dense resulting in higher power and cooling requirements that the current HPC datacenters may not be able to support without some modifications and/or expansion.

Traditionally these costs have not been eligible for federal funding, however this can lead to potential inefficiencies at the national level, and inequalities between host site institutions which incur substantial costs while the majority of institutions are not sharing the infrastructure costs. A more thorough discussion of funding of HPC data center construction, maintenance, and operations is provided in the ARC Current State Report, see pages 108-109.

4.7.2 Environmental Impact

Requirements for datacenter modifications and expansions should be seen as an opportunity to continue to modernize HPC datacenters and invest in more efficient and environmentally sustainable operations. Options such as direct water cooling using warm/hot water can significantly reduce operations cost, lower the carbon footprint, and support the higher thermal loads of new HPC compute servers. Depending on the location and configuration of each data center there may be opportunities to reuse the excess waste heat which if done correctly can significantly lower the carbon footprint of the datacenter.

In all situations future procurements should adopt energy efficient designs and best practices. The Energy Efficient HPC working group has documented the “Energy Efficient Considerations for HPC Procurement Document”[28] (2021) which describes requirements to consider in systems features and capabilities in order to allow reliable and accurate energy efficiency measurements (benchmarking, power and cooling design and measurements, interfacing with facilities).

[28] Lawrence Livermore National Labs’ (LLNL) Energy Efficient HPC Working Group: Energy Efficiency Considerations for HPC Procurement Document: 2021 https://drive.google.com/file/d/1aB7uv47anaHUcHzw140tJLUSe9xVkQfQ/view (October 2021).

4.7.3 Redundancy, Resiliency, and Uptime Considerations

HPC datacenters have commonly accepted the risk of running with known single points of failure, spending less money on the physical infrastructure (e.g. no redundant cooling or full UPS protection), instead spending more of the capital budget on the core HPC compute capacity infrastructure. Costs of fully redundant power, cooling, storage, and networking at the scale needed for HPC systems would be very expensive and would significantly reduce the amount of compute resources available to researchers.

In practice, resiliency may be the more relevant consideration. Designs that ensure a percentage of the infrastructure and critical components like storage can be operational in case of power failure using UPS + generator may be possible for substantially less cost than full redundancy. Networking redundancy can take the form of a lower capacity link to sustain connectivity even if large data transfers need to be suspended.

It should also be considered that even if each individual system is not very locally redundant, the benefit of having multiple systems distributed across separate geographically diverse datacenters provides system resilience as well. Currently the systems are mostly independent in terms of operations, except for a few key services which are supported by HA infrastructure. Total capacity is an issue though as the current systems are fully utilized so there is no significant excess capacity available to absorb increased demand. Another major hurdle preventing each system from providing redundancy for other systems is the data locality and lack of data redundancy. On the current systems all the data is local, with no replication to other sites, as even the backups are currently only stored locally, except for the Niagara system that currently replicates its backups and nearline to the Center for Advanced Computing (CAC).

Future systems should consider cross site backups of at least core critical data to provide at least partial data redundancy. Also providing cluster access to non-local storage such as a decentralized object store or some other replicated HA storage would go a long way to providing a more resilient service and significantly reduce risk in the case of a single site failure.

Other considerations like 24/7 operations support compared to the current business hours + best effort can be investigated as part of setting appropriate Service Level Objectives (SLO) for host sites.

4.8 Future Technologies and Test Environments

HPC Architectures are constantly evolving to adapt and leverage the latest technologies to drive performance. As such it is important to continuously evaluate and provide resources to allow staff and researchers to explore emerging and alternative technologies before investing in large deployments. This also applies for testing of possible new features and services such as software

and storage offerings. As such the Alliance should consider funding small scale test and development systems. Some examples of these include:

- Alliance Hub, centralized location for hardware evaluation and benchmarking
- Test & Development Systems for beta testing before production rollouts
- Continuous Integration/Continuous Delivery (CI/CD) resources to support Research Software and HPC platform developments
- Small scale flexible test environments to investigate emerging HPC architectures
    - Orchestration (Kubernetes/Fuzzball/etc)
    - Composable Infrastructures

4.9 Support, Efficiency & Usability

4.9.1 Support & Training

Providing funding for expanded resources is an important step in fulfilling researcher demand for resources, however there also needs to be funding to grow the staff that operate and support the use of the expanded resources. Research support personnel are critical in making sure systems are maintained and operated properly. Similarly as with resource capacity not being able to adequately support the research demand, the existing federation workforce is not adequately staffed to properly support the ever growing number of users. The federation currently has an ~200 FTE strong Alliance DRI Professional (ADP) workforce which translates to roughly 6 FTE per contributing institution, and to roughly 1:80 ratio between the number of ADP FTEs and registered users. In comparison, the Texas Advanced Computing Center (TACC), the host site for Frontera, that serves ‘several thousand users with a staffing of roughly 190 people, has a staff-to-user ratio in the range of 1:16 to 1:55, i.e. a 2-5 times more favorable ratio.

With increased personnel it will be possible to support increased user training as well as provide more discipline specific workload assistance. With the growing number of HPC users and regular cadence of graduate student turnover, providing regular training is key to ensuring researcher success in using the resources as well as reducing the amount of systems problems related to user mistakes and/or misuse. Having sufficient staff to provide detailed assistance to users in areas such as workload and program analysis can result in significant efficiency gains, leading to increased research productivity and better use of the HPC resource. This is especially true for workflows that suffer from poor parallel scalability, inefficient GPU utilization, and non-optimal IO access patterns.

4.9.2 System and Workload Efficiency

Next generation systems should place an emphasis on tools and operations that improve efficiency, making sure the systems and workloads are making the best use of the available resources. The Alliance should support initiatives such as developing tools to better instrument,

automate, and provide utilization feedback to users and support staff. These tools can also be used to monitor and provide system performance metrics as well, ensuring the resources are used efficiently.

4.9.3 User Experience

As a wider research community requires HPC resources, there is increased demand to support alternative workflows and interfaces. To improve usability the Alliance should investigate broader support of solutions such as Virtual Desktop Infrastructure (VDI) or web based workflows. Dedicated resources to run interactive services like Jupyter notebooks and/or web portals should also be made available at scale.

Appendix A - Resource Allocation Competition (RAC) Demand

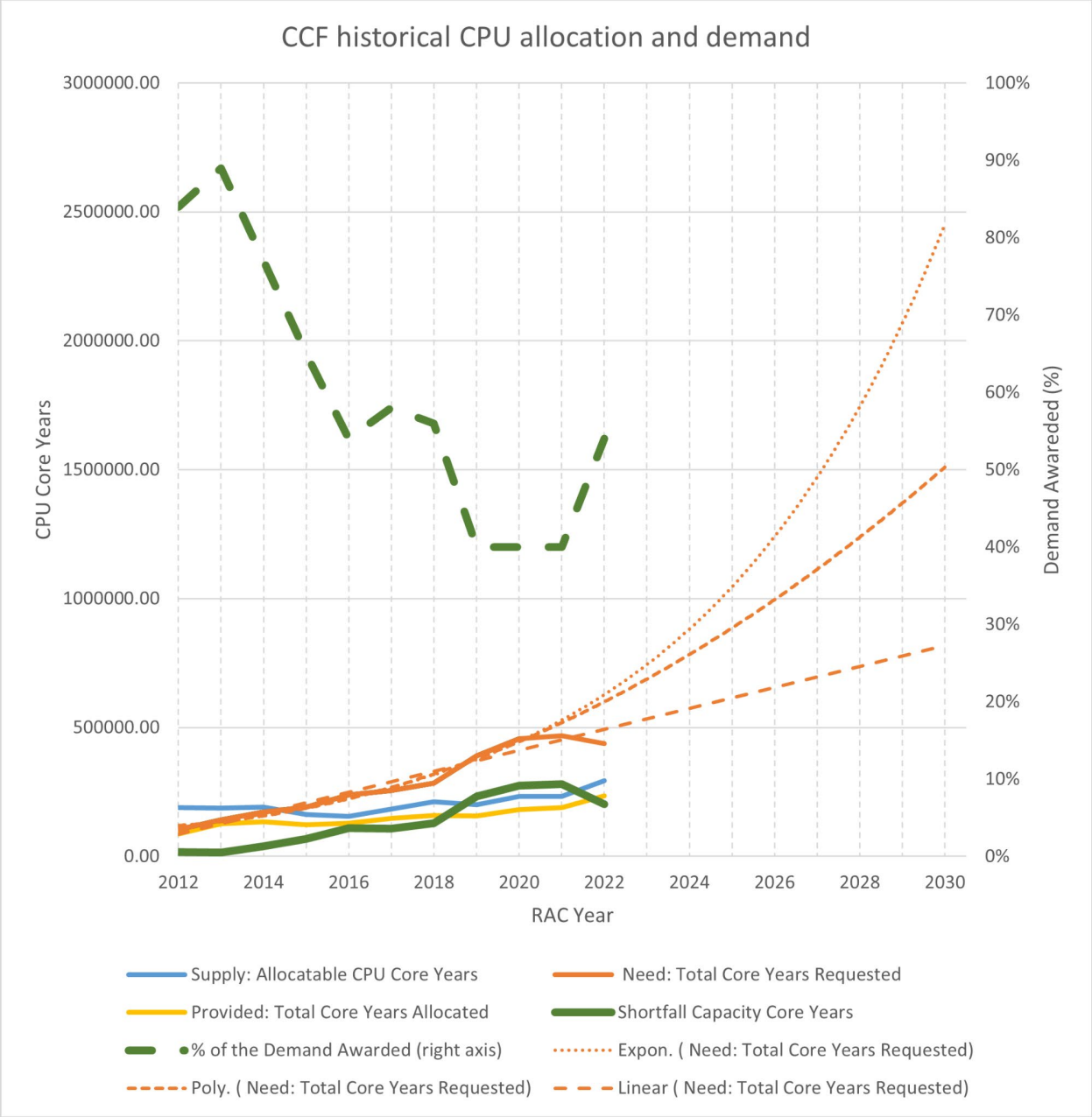

Figure A.1 - CPU RAC demand w/ projected future trendlines.

Figure A.1 above compares CPU compute resource supply and demand in the CCF consortium from 2012 to 2022. The supply is based on actual available CPU capacity in Compute Canada Federation (CCF) systems, and the demand is based on the submitted resource allocation requests in CCF's annual Resource Allocations Competition (RAC) process. The horizontal axis corresponds to RAC allocation years, while the units on the vertical axis are in CPU core-years.

The blue line is the total annual raw compute capacity available in the main CCF systems. The total available capacity initially fluctuated within a relatively narrow range, and then grew more recently to 293k CPU core-years in RAC2022, i.e., roughly doubling the supply from ten years ago. Notably the CPU core-year metric does not consider the actual compute power of each cycle; that is, the increase in processor compute capability thanks to architecture developments.

The orange solid line is the total capacity requested in RAC competitions. In the last ten years the demand has grown from ca 100k CPU core-years to peak of 470k CPU core-years in RAC2021, and then dropped to roughly 440k CPU core-years in RAC2022. The slow down and drop in demand in the last two years is not fully understood, but the bulk of this change can most likely be attributed to natural fluctuations, and (temporary) disruptions in academic research activity and focus due to COVID-19. This WG expects the growth in demand to quickly resume as the research community is able to get back to normal. This conclusion is further supported by international projections for HPC needs: the need for HPC will be strongly increasing for the foreseeable future. The overall growth in demand has been very rapid and semi-linear but does not seem to be exponential. In compounded annual growth rate (CAGR) terms the growth in the demand for CPU computing cycles between RAC2012 and RAC2020 was roughly 21% per year.

The thin orange lines show three trendlines for future projected demand. These projections are based on RAC2012-20 data, excluding RAC2021-22 due to the anomalous and temporary drop in the last two years. The three different trendlines correspond to three different functional assumptions for the future growth: the dotted line assumes exponential growth, the small-dashed line assumes 2nd degree polynomial growth, while the large-dashed line assumes linear growth. These three projections provide insights and ranges for future demand: in RAC2030 the projected demand could be estimated to range between 800k and 2400k CPU-core years.

The yellow line indicates the actual RAC allocated capacity. It should be noted that part of the total capacity is provided to end-users via the RAC application process (deliberately capped at about 80% of the capacity), while the remaining capacity (roughly 20%) is available for use by any user on as needed basis without a need for a formal application. In RAC2022 the available total resource of roughly 294k CPU core years was distributed between the RAC (roughly 230k CPU core years), and unallocated/non-competitive use. The unallocated 64k CPU year capacity is utilized by the rapid access RAS users, i.e. 'opportunistic' use.

The allocated capacity (yellow line) has closely followed the available capacity (blue line), leaving the above-mentioned roughly 20% margin for the rapid access services. Comparing the supply (blue line) and demand (orange line) we see that until RAC2020 the CPU computing capacity had been falling behind with the rapidly growing need and the ARC infrastructure development had not kept up with the demand. In the last two years the situation has improved thanks to the

temporary drop in demand (as discussed above), and recent strong increase in supply (thanks to commissioning of Narval in Fall 2021).

The thick green solid line shows this unmet demand in absolute terms. In RAC2022 this was roughly 200k CPU years. The green dashed thick line highlights the scale of the situation by showing what percentage of the computing demand was actually allocated. Historically this can be seen to decrease from roughly 80% of the demand being satisfied in 2012 to only 40% in 2020, and then climbing to 54% in RAC2022. Even with the recent temporary reprieve, in the last decade the CPU compute capacity shortfall has increased significantly in both absolute and relative terms.

It should be kept in mind that ARC resources by their nature are always in short supply due to the constantly growing number of disciplines leveraging DRI, increasing resolution of experimental or observatory instruments, and the need for higher resolution and accuracy simulations, often adjusted to the available compute allocation. If the available compute resource increases, the researchers quickly switch to leveraging that capacity for new science.

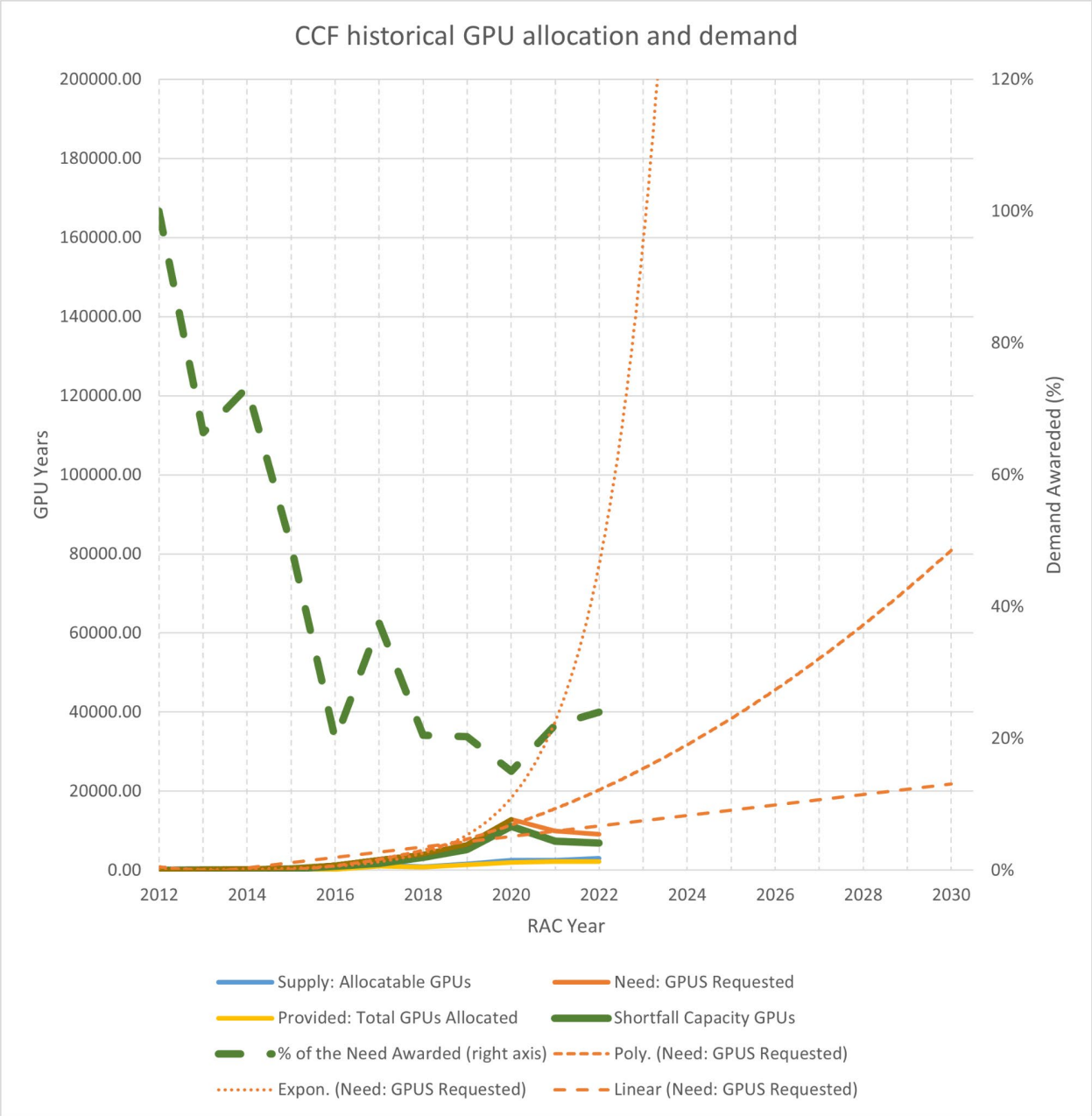

Figure A.2 - GPU RAC demand w/ projected future trendlines.

Figure A.2 above compares GPU accelerator supply and demand in the CCF consortium from 2012 to 2022. The supply is based on actual available GPU capacity in CCF systems, and the demand is based on annually submitted RAC resource allocation requests. The horizontal axis corresponds to RAC allocation years, while the units on the vertical axis correspond to GPU years. The blue line is the total annual allocatable capacity while the orange line is the RAC requested capacity. The yellow line is the capacity allocated to end-users via the RAC mechanism.

The demand (see the orange solid line) for GPU computing resources has grown strongly between 2012 and 2020. In 2012 the requests were minimal, at 10 GPU years, while in 2020 RAC round the total request was nearly 13 000 GPU years. The growth has been non-linear, growing exponentially year over year. In CAGR terms the growth was ca 67% since 2017 (when the demand was ca 2800 GPU years), indicating the very rapidly increasing demand for this resource. More recently in RAC2021-22 the GPU demand has dropped, most likely due to combination of two effects: 1) Similar to drop in CPU demand, a Covid-19 related reduction and refocusing of academic research, and 2) the RAC GPU requests are inflated by the complexities of new technology, new compute methods and paradigms and difficulties related to estimating the actual need. With rapid emergence and adoption of GPUs researchers do not necessarily have good baselines for code performance, or for the number of training runs required or the human-time needed to manage and interpret results. Even with the introduction of new artificial intelligence (AI) workload specific capacity at the leading Canadian AI institutions, this WG expects the growth in GPU demand to quickly resume as the research community is able to get back to normal. This conclusion is further supported by international projections for GPU computing needs: the need for GPUs will be strongly increasing for the foreseeable future.

Putting the recent temporary dip aside, GPU computing capability is falling behind with the rapidly growing needs. On one hand this is positive, showing significant increasing interest for accelerator technologies, while on the other hand the ARC infrastructure has not kept up with the demand. In absolute terms the unmet GPU capacity in 2022 was roughly 7000 GPU years, as shown by the solid green thick line in the Figure above. In relative terms the demand that has been fulfilled (the dashed thick green line) has dropped from roughly 100% in 2012 to roughly 20% in 2020, and rising to 24% in RAC2022.

The thin orange lines show three trendlines for future projected demand. These projections are based on RAC2012-20 data, excluding RAC2021-22 due to the anomalous and (very likely temporary) drop in the last two years. The three different trendlines correspond to three different functional assumptions for the future growth: the dotted line assumes exponential growth, the small-dashed line assumes 2nd degree polynomial growth, while the large-dashed line assumes linear growth. The exponential growth estimate is out of bounds and not shown fully in the graph. Any longer term strong exponential growth will slow down substantially due to technology maturation as discussed above. The linear and polynomial projections provide insights and ranges for future demand: in RAC2030 the projected demand could be estimated to range between 20k and 80k GPU years.

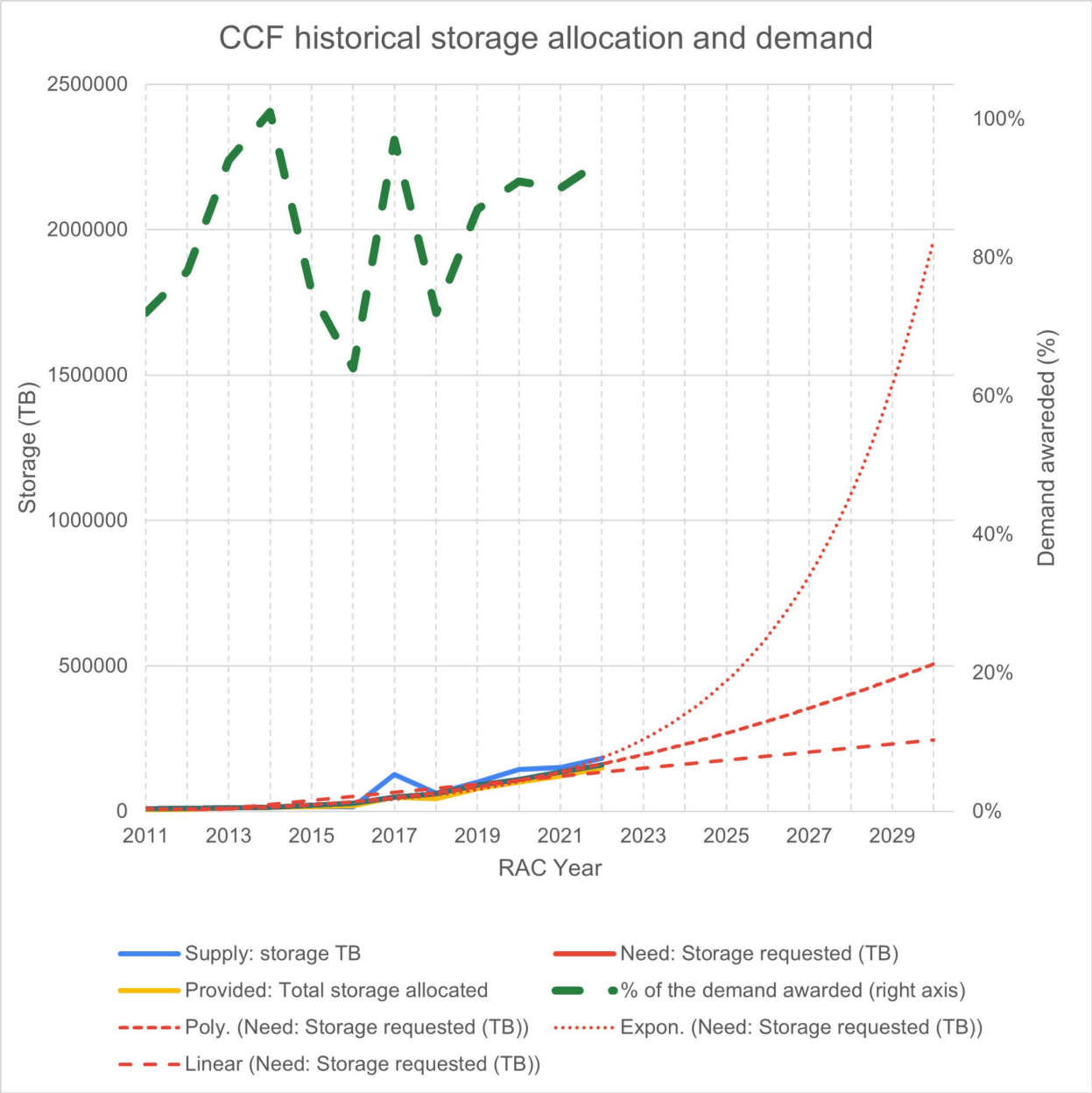

Figure A.3 - Storage RAC demand w/ projected future trendlines.

Figure A.3 above shows the historical storage supply and demand at CCF. The aggregate available supply of the various available storage types is shown by the solid blue line. In RAC2015 the total capacity was roughly 15 PB, increasing more than tenfold to roughly 180 PB in 2022 with substantial annual variations as old systems have been retired and new ones brought online. This total storage is distributed among functionally different storage types including Project, dCache, Cloud, and Nearline storage systems.

On the demand side the red line indicates the historical aggregate storage demand, while the yellow line indicates the storage provided. The demand has increased roughly eight-fold from roughly 21 PB in 2015 to roughly 160 PB in 2022. The roughly eight-fold growth in demand over seven years since 2015 corresponds to roughly 34% CAGR. Notably in 2022 the total storage capacity (blue line) was roughly 20 PB larger than the total request (red line). The green dashed line indicates the awarded/allocated demand as a percentage of the annual storage requests. This fulfilled demand has ranged from 72% in 2011 to 94% in 2022, while dipping to 64% in 2016. Thanks to increases in supply, the storage system on aggregate has been able to keep up with the demand.

The thin orange lines show three trendlines for future projected demand. The three different trendlines correspond to three different functional assumptions for the future growth: the dotted line assumes exponential growth, the small-dashed line assumes 2nd degree polynomial growth, while the large-dashed line assumes linear growth. These three different functional projections provide insights and ranges for future demand: in RAC2030 the projected active storage demand could be estimated to range between 250PB and 2000PB.

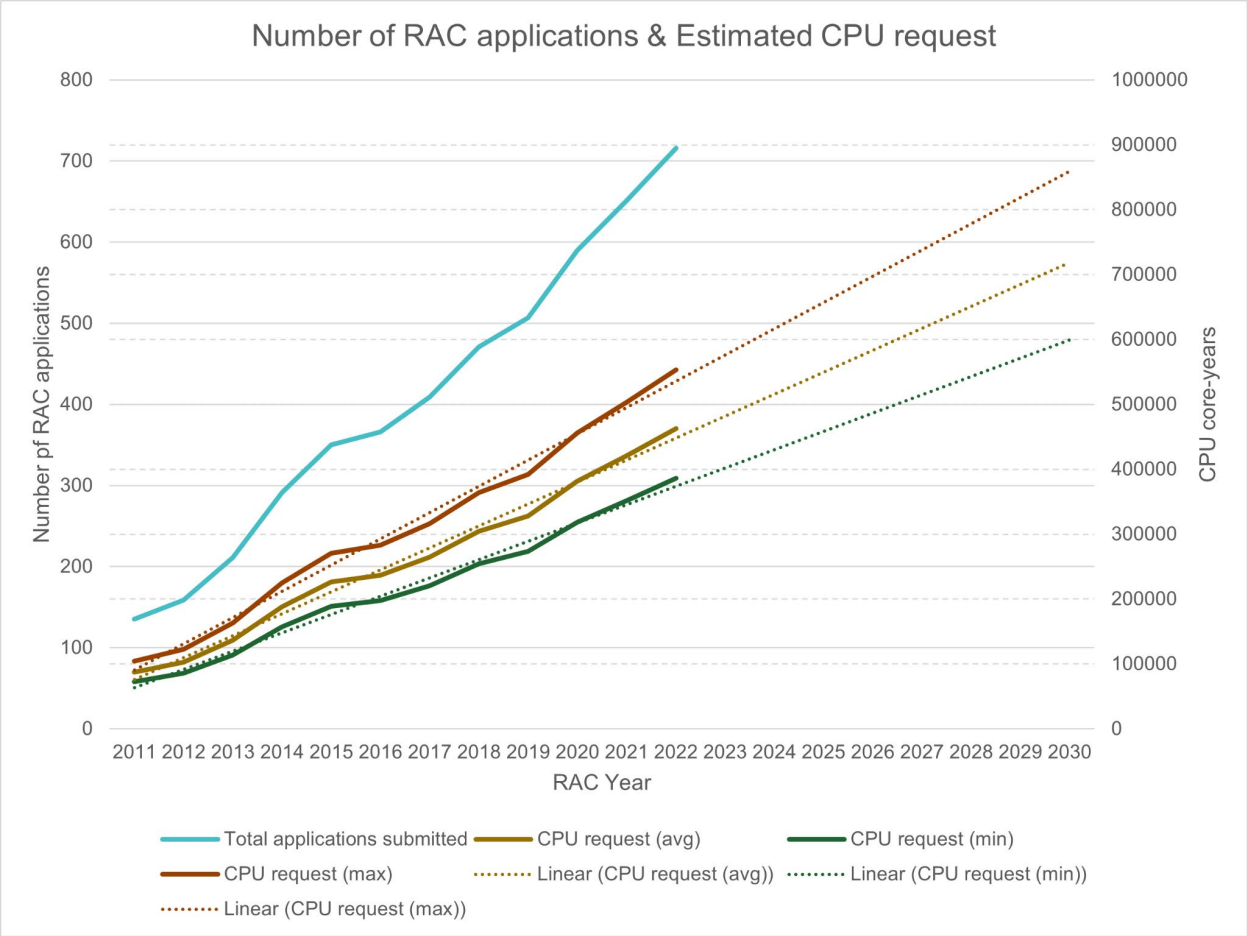

Figure A.4 - Historical number of RAC applications, and estimated CPU core-year demand.

Figure A.4 above shows the historical number of RAC applications (blue solid line, y-axis on the left), and estimated CPU core year demand based on RAC CPU core-year request data (dotted lines, y-axis on the right). The number of RAC applications per year has grown steadily since RAC 2011, growing from 135 in 2011 to 716 in 2022. The nearly linear growth is likely to continue going forward, at least when considering the traditional HPC users of the systems.

The average CPU core-year request per application (not shown) has remained relatively constant over the last decade with some variation between years. The overall average request, from RAC 2011 to RAC 2022, has been 647 CPU core-years per application. The variation has been from 540 to 773 CPU core-years since RAC 2011.

The three solid graphs and extrapolated dotted lines (colored in red, brown and green, y-axis on the right) show the estimated CPU core-year demand, based on linearly growing number of applications per year, and relatively constant CPU core-year request per application.

More specifically, the graphs are formed by multiplying the average/minimum/maximum CPU demand per application with the number of applications for each year, and then extrapolating

linearly to 2030. Based on the historical long-term **average** CPU demand per application the estimated demand would be roughly **700,000 CPU core-years in 2030** (brown dotted line), while the minimum (green dotted line) and maximum (red dotted line) historical CPU demands indicate **600k and 870k CPU core year lower and upper bounds in 2030,** respectively. The spread of 600k to 870k with 700k average can be used as a first-approximation baseline estimate for CPU core count need in 2030, assuming that usage patterns and profiles stay roughly constant. Notably this range for 2030 is roughly 2.5x the current available CPU core count (of roughly 296k CPU cores).

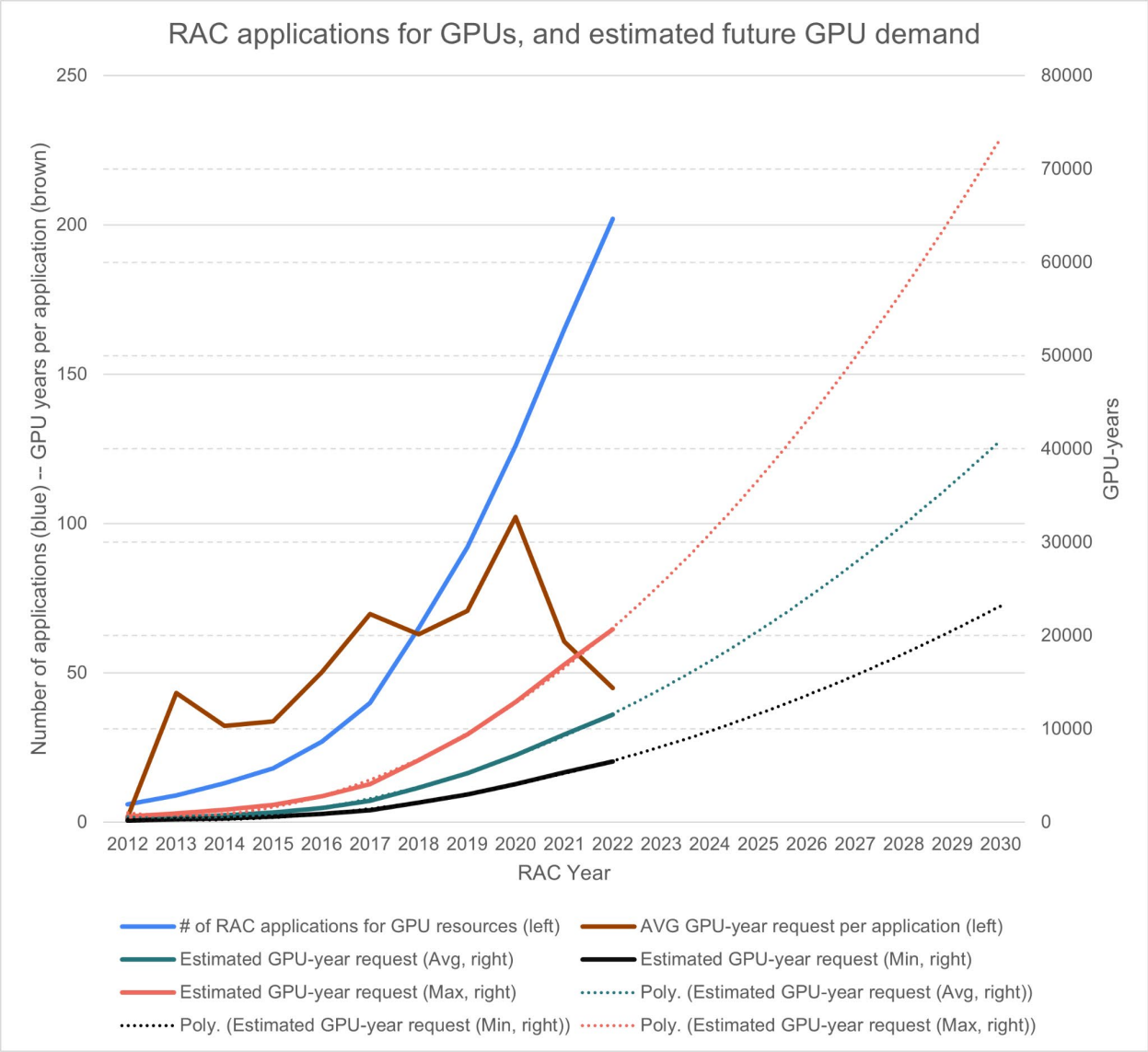

Figure A.5 - Historical number of RAC applications with GPU requests, and estimated GPU demand.

Figure A.5 above shows the historical number of RAC applications that included a request for GPU resources (blue solid line, y-axis on the left), and estimated GPU core year demand based on RAC GPU core-year request data (dotted lines, y-axis on the right). The yearly average GPU request per GPU application is also shown (brown solid line, y-axis on the left). Looking at the blue solid line we see that the number of GPU-related RAC applications per year has grown steadily since RAC 2012, growing from 6 in 2012 to 202 in 2022. The non-linear early rapid growth is an indication of how an increasing ratio of RAC applicants have requested GPU resources over time. In RAC 2012 4% of applicants requested GPU resources while in RAC 2022 the percentage was 28%. With that, the early non-linear growth is not likely to continue going forward, rather will

likely continue as the nearly linear growth in combination with the steady linear growth of all RAC applications (Fig NN3).

The average GPU-year request per application (brown solid line) has fluctuated over the last decade, indicating the difficulty applicants have had in estimating their needs of this emerging technology. The overall average request, from RAC 2013 to RAC 2022, has been 57 GPU core-years per application (per GPU specific applications, not per all applications). The variation has been from 32 to 102 GPU-years since RAC 2013.

The three solid graphs and extrapolated dotted lines (colored in red, green and black, y-axis on the right) show the estimated GPU core-year demand, based on growing number of GPU applications per year, and assuming relatively constant GPU core-year request per application (using similar methodology to the CPU projections above).

More specifically, the graphs are formed by multiplying the average/minimum/maximum GPU demand per GPU application with the number of GPU applications for each year, and then extrapolating linearly to 2030. Based on the historical long-term **average** GPU demand per GPU application the estimated demand would be roughly **40,000 GPU-years in 2030** (green dotted line), while the minimum (black dotted line) and maximum (red dotted line) historical GPU demands indicate **23k and 63k GPU-year lower and upper bounds in 2030,** respectively. This spread can be used as a first-approximation baseline estimate for GPU need in 2030, assuming that usage patterns and profiles stay roughly constant. Notably this range for 2030 is roughly 10-20x the current available GPU count (of roughly 3000 GPUs).

It should be noted that GPU performance is advancing at a much faster rate than other technologies making researchers' abilities to estimate their expected needs often inaccurate. A GPU-year is at best a crude estimate as an NVIDIA P100 from 2017 has a peak IEEE FP64 of 5TFLOPS whereas an H100 from 2022 has a peak of 60 TFLOPS. That is a 12x increase in 5 years where a CPU-core in the same time frame only changed 2-3x in FP64 improvement.

Appendix B - HPC Workload Analysis

Historical scheduling data from existing federation HPC systems has been analyzed and summarized using a series of tools developed by the Data Analytics National Team (DANT). The following is a small sample of the available job data which can be used to provide insight into job trends, scheduling policies and for ensuring efficient usage of the systems.

B.1 Workload Demand

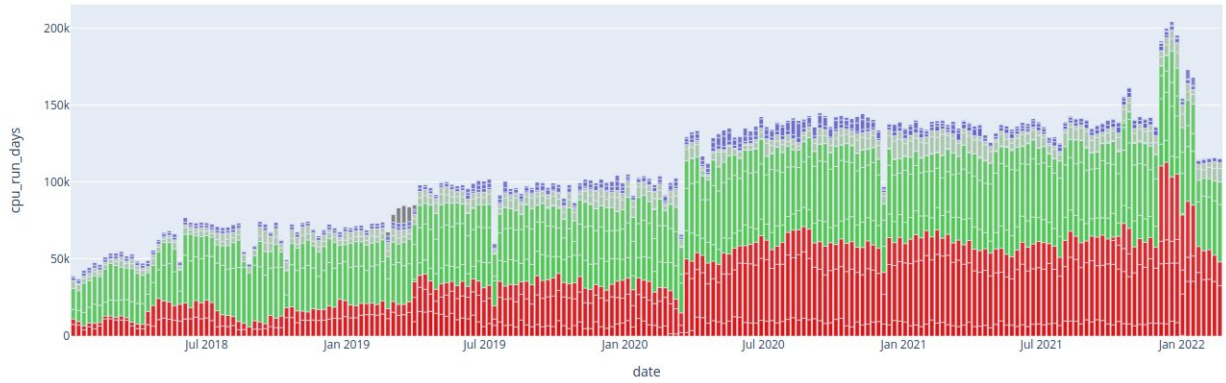

Figure B.1 - CPU workloads for GP systems from 2018-2022.

Figure B.1 shows the total amount of CPU workload executed on the general purpose (GP) systems (Graham, Cedar and Beluga) combined, ranging from 2018 through 2022. The utilization is shown in cpu run days, which is equivalent to using 1 cpu-core for 24 hours. The colors correspond to job account type; red default, green RAC, blue contributed, and grey other, which are discussed in more detail in section B.3. The increase in utilization over time is the result of two expansions to Cedar and the addition of Beluga. Using the job scheduler data we can investigate demand for resources by comparing the executed workload to the amount queued. Specifically if we investigate the demand during the periods of time when the capacity was increased, we can see how the amount of queued workload responds.

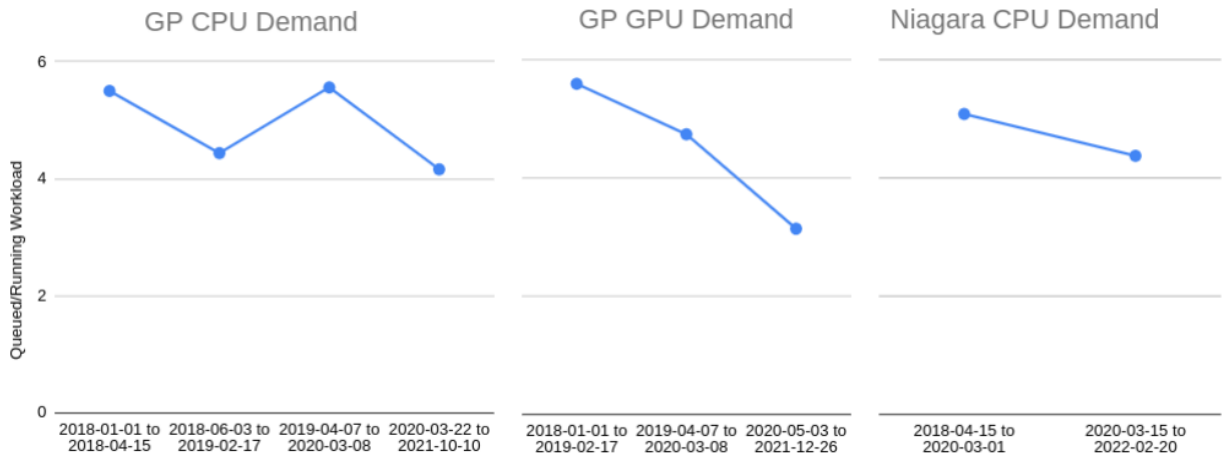

Figure B.2 - Workload demand during capacity expansions of HPC resources.

Figure B.2 shows the mean queued demand normalized by the system's running capacity for (before and after) periods of expansion of GP CPU, GP GPU and Niagara CPU resources. Even though there are fluctuations, the queued demand for CPU resources consistently averages around 4-5 times the running capacity even after significant capacity expansions. The only slight reduction in queued demand is for GPU resources, which appears to have dropped somewhat as more GPU resources were brought online such as the GPU heavy Beluga system, and as users have become more familiar with this emerging technology. Even though there has been some reduction in GPU demand, the queued workload demand is still over 3 times the available capacity.

B.2 Workload Resource Characteristics

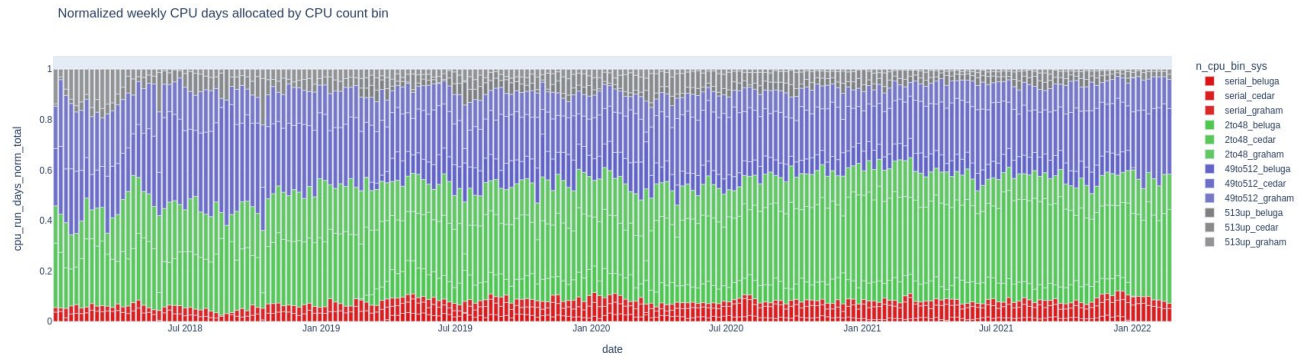

Figure B.3 - GP CPU Job Size Distribution.

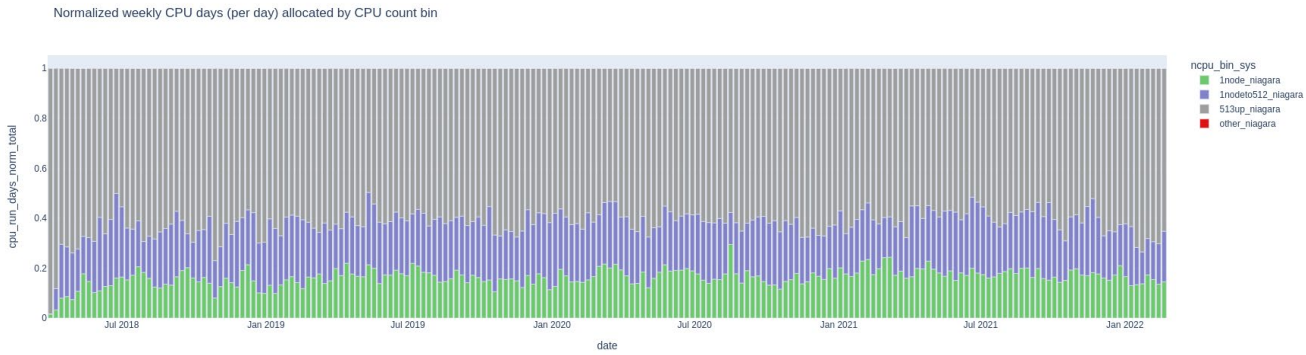

Figure B.4 - Niagara CPU Job Size Distribution.

Figure B.3 shows the normalized utilization of the CPU only workloads over time, from 2018 to 2022, for the Cedar, Graham, and Beluga GP systems subdivided by job size. Figure B.4 shows the same information for the Niagara LP system. The red bars correspond to serial jobs, the green for jobs using 2 to 48 cores, blue for jobs using 49 to 512 cores, and grey for jobs over 512 cores. That data shows that approximately 50% of the GP systems CPU usage is for workloads using a single node or less and the trend has not changed that much over the 4 years of data. For the Niagara LP system approximately 60% of system usage has been used by jobs requiring 512 cores or more. This should not be surprising as this is in keeping with its design and allocation as a specialized system for massively parallel workloads. Notably, there has been no observable changes in job size distribution, indicating that the current allocation of infrastructure between GP and LP type systems remains appropriate.

Figure B.5 - GP GPU Job Size Distribution.

Figure B.5 shows the normalized utilization of GPU workloads over time, from 2018 to 2022, for the Cedar, Graham, and Beluga GP systems subdivided by job size. The red bars correspond to single GPU jobs, the light green jobs using 2 to 4 GPUs, dark green jobs using 5 to 8 GPUs, blue jobs using 9 to 16 GPUs, and grey for jobs using over 17 GPUs. Over 50% of GPU jobs used only 1 GPU and 95% used 4 or less GPUs. i.e. single node. There was very little use of these systems

for larger scale parallel heterogeneous GPU jobs. Again, there have been no major changes in job size distribution over time.

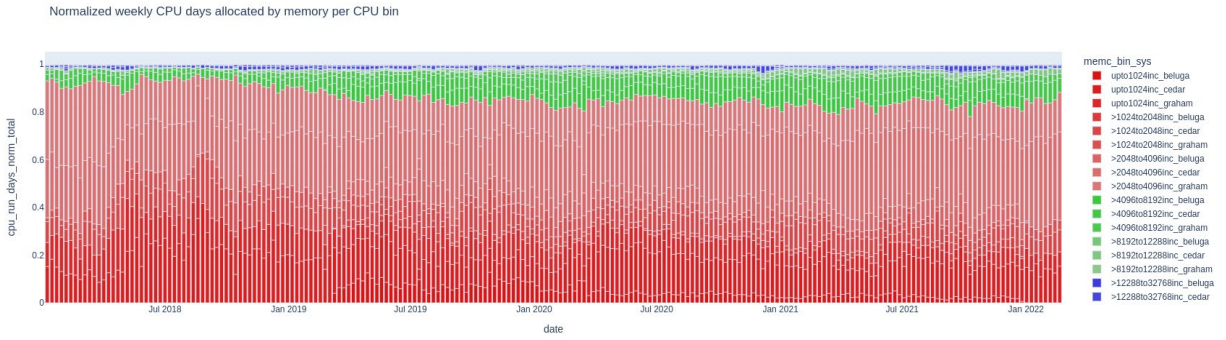

Figure B.6 - GP CPU Job Memory Usage.

Figure B.6 shows the normalized utilization of CPU workloads over time for the Cedar, Graham, and Beluga GP systems subdivided by job memory request. The dark red bars correspond to jobs requesting up to 1GB/core, the light red requesting up to 4GB/core, the green requesting up to 12GB/core, and the blue requesting 12-32GB. Approximately 85% of jobs requested 4GB/core memory or less across all systems, which is not surprising as for the majority of the systems approximately 4GB/core is the design point of the hardware, i.e. a 32 core node with 128GB RAM. Again, there has been no major changes in job size distribution over time. The Niagara system is scheduled only by full node so all jobs have approximately 4.5GB/core.

B.3 Workloads by RAC Allocation Types

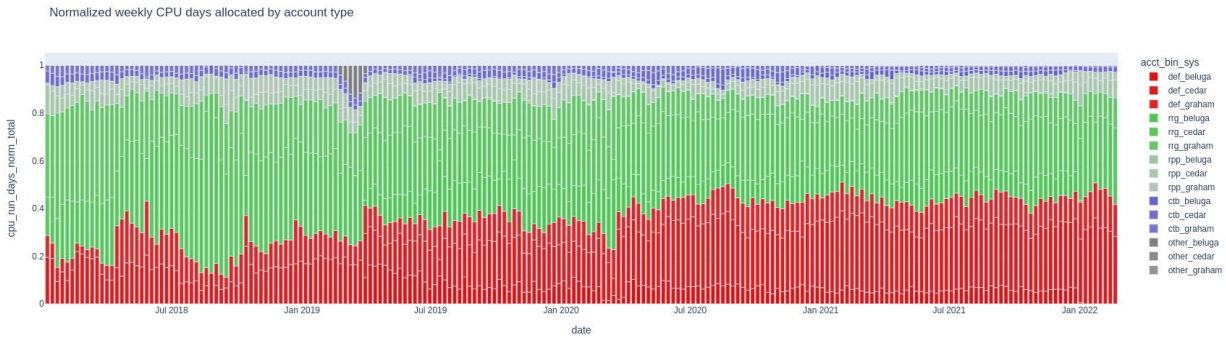

Figure B.7 - GP CPU Usage by Allocation Type.

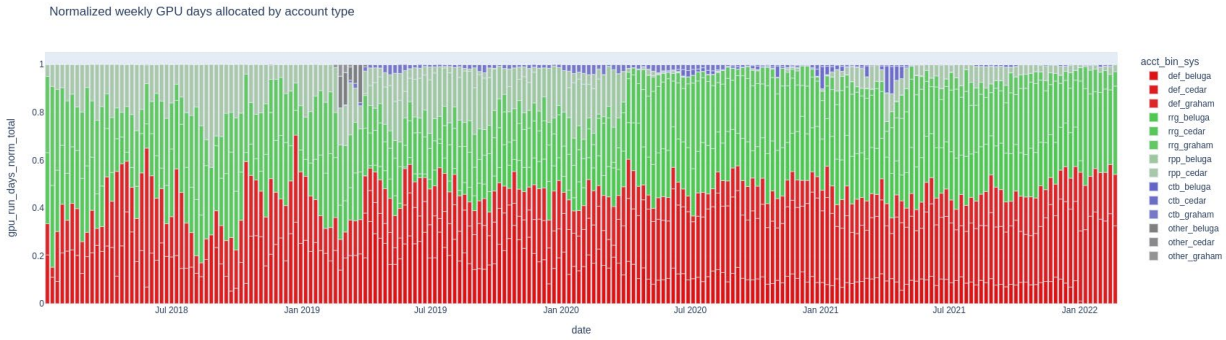

Figure B.8 - GP GPU Usage by Allocation Type.

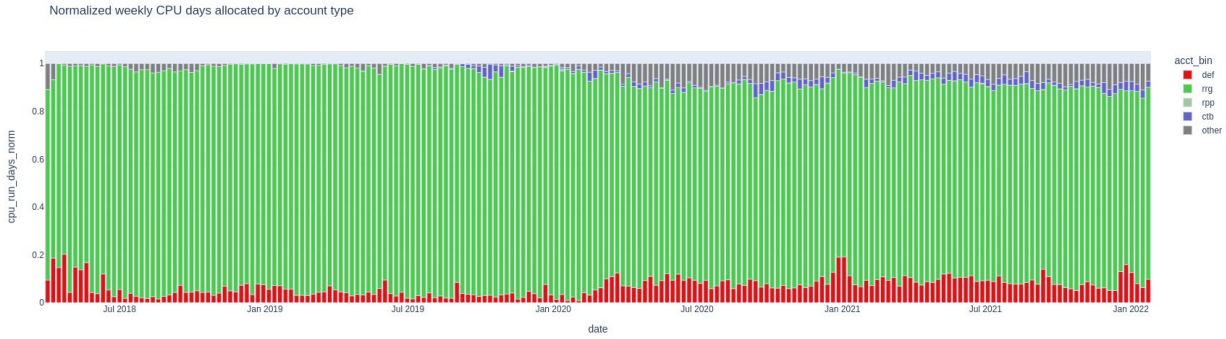

Figure B.9 - Niagara CPU Usage by Allocation Type.

Figures B.7, B.8, and B.9 show the workload usage by allocation type for GP CPU jobs, GP GPU jobs, and Niagara jobs respectively. The red bars correspond to jobs run under default allocations, green to RAC Resources for Research Groups (RRG's) allocations, light green to RAC Research Platforms and Portals (RPP's) allocations, blue to contributed, and grey to other. The RAC process typically allocates 80% of resources available each year, however with the exception of the Niagara system, only about 60% of those resources are actually being used by those RAC accounts. On the GP systems default allocations are using as much as 40% of the CPU resources and over 50% of the GPU resources. There has been no material changes in the distribution of usage of different allocation types, except the increase of 'default' CPU allocation from roughly 20% in 2018 to roughly 40% in 2022 on GP systems.

Appendix C - Networking

C.1 Internal High Speed Interconnects

HPC systems require a high performance network to allow the system to function together and not simply as a collection of individual nodes. The design of the network and the capabilities it provides depends on the targeted use of the system.

The network technology and design needs to be carefully selected. The deployed network needs to have the capability to support the workloads for the cluster without over designing the fabric. Interconnect costs can make up a significant amount of the total cluster budget so needs to be well thought out.

The national platform needs to support all dimensions of HPC computing. This includes both scale out and scale up. The deployed systems should provide a range of capabilities to support the range of targeted use cases. We can use the job size data from above to help design the network. The job data will show the sizes of jobs that users are running and the national platform as whole needs to support the various job sizes.

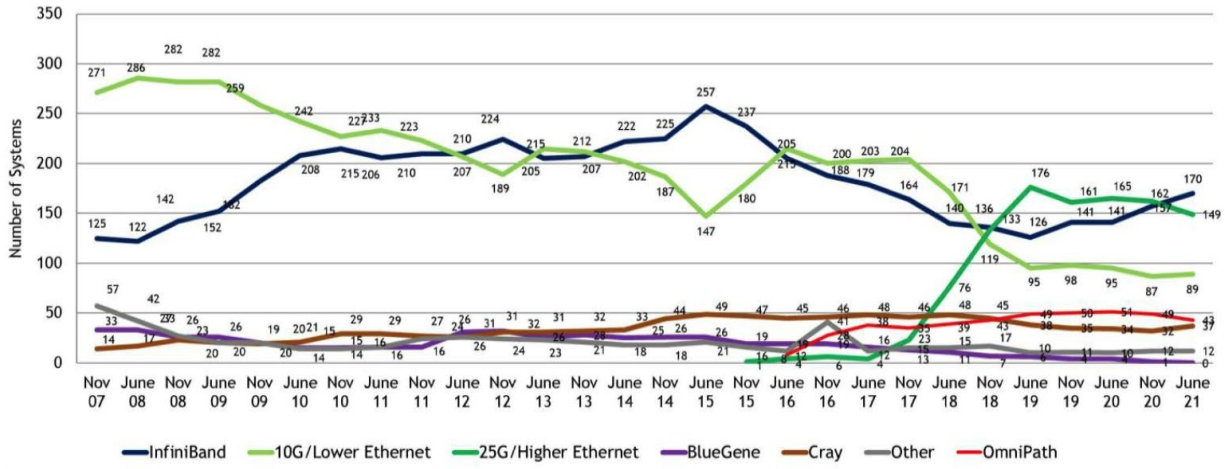

Figure C.1 - Interconnect share on TOP500 systems over time.[29]

Infiniband has been a market leader for years and remains the predominant non-ethernet HPC interconnect fabric for many years, as shown in Figure C.1, but there are other technologies. High

[29] TOP500 http://top500.org/statistics/list/ (November 2021)

speed ethernet is a common and typically a less expensive option but doesn't have the overall performance of Infiniband and has less topology and communication protocol options. This could be a reasonable compromise depending on the workloads being targeted.

There are also other, less common options which may be suitable. Care needs to be taken to mitigate risk and not have multiple sites all depending on a new technology. Intel's OPA had promise but now it is not a common solution making support and future expansion difficult. New fabrics like HPE Slingshot and Rockport are now available and should be considered for limited or trial installations.

The fabric is used for job communication but also for file system access. With more data focused computing the fabric needs to be able to quickly transport files from central storage to the compute nodes. With larger nodes there is less demand for inter-node communication since more jobs will fit within a node but there is a larger demand for bandwidth into the node to get data to all the cores.

Considerations at the time of procurement for the fabric:

1. Support the targeted job size and mix, including job communication bandwidth and latency needs, and job data I/O needs
2. Diversity of fabric across the national platform
3. Cost, scalability, and availability of each option
4. Staff experience operating the technology

C.2 External Networking

Wide area network (WAN) connectivity at national sites can be summarized into 3 types of services

1. CANARIE and National Research and Education Network (NREN) IP service
2. Commercial Internet (commodity)
3. Research project networks

All sites have connectivity to the CANARIE/NREN and commercial Internet IP services (1,2). These services are described in the following subsections.

C.2.1 Canarie/NREN IP Service (Research & Education)

Each national site leverages the national Canadian research backbone operated by CANARIE, in partnership with regional NREN, to enable high speed connectivity between national sites and to any institution connected to CANARIE. The CANARIE network also provides connectivity to international research and education networks.

CANARIE does not provide direct connectivity to a site, so the connectivity details for each site are dependent on the regional NREN which provides access to the CANARIE backbone. There is no commercial Internet traffic over the CANARIE IP network.

Table C.1 lists the national sites R&E connectivity to the NREN and CANARIE.

Example of traffic flows using this service:

- Inter-site traffic (e.g. data transfers between sites)
- Traffic from a research campus (Institutions connected to their provincial NREN)

C.2.2 Commercial Internet (commodity)

A commercial Internet IP service is required to provide access to the HPC and cloud services for users connecting from the Internet, and to support normal systems operations (DNS, system patches, mail, etc.). The regional NREN or the local campus are the providers for this service, and the bandwidth capacity is determined by the local site and the provider.

Although transfers for research datasets are expected to use the high speed CANARIE/NREN IP service, commercial cloud is increasingly being used to share datasets required by researchers, and these transfers typically use the lower speed commercial Internet service. When a peering agreement between CANARIE/NREN and the commercial cloud provider is in place,[30] then a higher speed network path to the cloud provider is used for these transfers.

As the usage of commercial cloud resources increases (e.g. dataset transfers between hosting site and commercial cloud), the network capacity to reach these resources will need to be improved (increased Internet bandwidth or CANARIE/NREN peering).

C.2.3 Research Project Networks

Although the CANARIE/NREN IP backbone enables high speed connectivity for research and education, data-intensive large scale and international science projects require end-to-end network capacity (bandwidth, latency, “friction-free” path) in order to meet science requirements.

One such example is LHCONE connectivity for the ATLAS initiative (3.5.1 High Energy Physics). In Canada, the Cedar, Arbutus and Graham sites are interconnected to LHCONE and participate in the ATLAS initiative. Cedar is the host site for the ATLAS Tier-1 center.

The Square Kilometer Array (SKA) is another large-scale science project that is expected to generate and distribute massive amounts of research data (3.5.2 Square Kilometer Array (SKA1)). The architecture for data movement is not yet defined, but it is expected to bring important networking requirements for participating sites in Canada.

[30] CANARIE: Content Delivery Service https://www.canarie.ca/network/services/cds/ (retrieved May 2022).

Site	NREN	R&E link	Commodity link provider	Commodity link capacity*	Research project network (as of 2022)
Narval/Beluga	RISQ	100Gbps	RISQ	400 Mbps	
Graham	ORION	100Gbps	ORION	2000 Mbps	ATLAS/LHCONE
Niagara	GTANet/ ORION	100Gbps	UofT		
Cedar	BCNET	100Gbps	BCNET	200 Mbps	ATLAS/LHCONE (Tier-1)
Arbutus	BCNET	100Gbps	BCNET	600 Mbps	ATLAS/LHCONE

Table C.1 - Host Sites Internet Connectivity. (*) Capacity from content providers peering agreements not included.

C.2.4 Traffic Estimates and Forecast

Using the data available from local network monitoring systems at each site, a summary of site WAN network usage can be analyzed to estimate the current traffic volumes. Since the data is summarized over a long period of time, traffic bursts are not recorded.

- Site hosting Atlas/LHCONE workloads (Cedar, Arbutus and Graham) measure higher WAN traffic compared to other sites (Beluga, Niagara).
- Cedar is the ATLAS Tier-1 hosting site, and shows the highest WAN network usage. Measurements from 2021 show WAN reaching its current 100Gbps link capacity.

The estimated Tier-1 bandwidth requirements from LHC for the next ATLAS run 3 are listed in the table C.2 below. These network bandwidth requirements are on top of the existing network bandwidth required for other non-LHC traffic to the site.

Year	Bandwidth
2023	60 Gbps
2025	100 Gbps
2027	200 Gbps

Table C.2 - Tier-1 bandwidth requirements for ATLAS run 3.

In addition to bandwidth requirements, the Worldwide LHC Computing Grid (WLCG) requires all participating sites to support IPv6 protocol. Although networking equipment has supported IPv6 for many years, procurement of new equipment will need to consider support for dual-stack (IPv4 and IPv6) operations.

The BELLE-II project (data center hosted at UVic) is not expected to generate much traffic for the next few years, but could grow in the same ballpark as the ATLAS traffic to Arbutus (Tier-2).

The Square Kilometer Array (SKA) is expected to bring important networking requirements both to the participating sites and to CANARIE to plan for the international link capacity that will be required to transport the data from the remote sites. This project is in the architecture and prototyping stages, so there is no immediate need for networking.

Summary

- Short term increased bandwidth capacity is required to support LHC Tier-1 operations at the Cedar site.
- Any increased network capacity must be coordinated with CANARIE, NREN, and in some sites with the local campus.
- Material, staff, and equipment resources need to be allocated and deployed to operate next-gen DTNs and network measurement tools (ex: perfSonar, link usage).

Appendix D - Host Site DataCenter Capacities

The current 5 national host sites have varying levels of available space, power and cooling capacities as outlined in Table D.1 In general physical space is not the limiting factor, but power availability and cooling infrastructure. HPC systems are becoming more power-dense and as such may require expanded or different electrical and cooling infrastructure to support new equipment. All sites, with the exception of UVic which has only air cooling, currently use predominantly Rear Door Heat Exchangers (RDHx) fed by chilled water for cooling. Standard air cooled racks typically can not handle power densities higher than 10-15kW per rack and RDHx's typically cannot handle more than 30-40kW/rack efficiently.

		SFU	Waterloo	UofT	McGill	UVic	Total
Racks	InUse	75	60	72	77	105	389
	Max	120	200+	120	230	140	810+
Power (MW)	InUse	1.45	0.6	1.45	1.75	0.496	5.746
	Max	10.0	1.4	3.75	2.0	3.0	21.15
Cooling (T)	InUse	400	200	475	400	120	1595
	Max	950	260	735	515	266	2726

Table D.1 - Current host site datacenter capacities.

Appendix E - Costing Resources

To estimate capacity and expansion costs two reference systems were designed, one with 100,000 CPU cores and another with 1000 GPUs. Both systems use an NDR/HDR based Infiniband network with 3:2 blocking. A high performance “scratch” parallel filesystem based on solid state storage is included and sized at ~10 times the system's memory. The reference compute node configuration and systems are shown in Table E.1 and E.2 respectively. Costs were estimated based on recent large system purchase prices and vendor provided typical educational pricing including 5-year warranties. These reference designs were then scaled to provide the cost estimates for the Scenarios listed in Section 4.2.

Compute Node (plus scratch & network fraction)	Power	Cost
2x Intel 32C, 256GB Ram, HDR200	800 W	$19,000
2x AMD 24C, 512GB Ram, 4x Nvidia A100 80GB, HDR200	3000 W	$84,000

Table E.1 - Reference Compute Nodes.

Systems	#Nodes	Scratch	Power	Racks	Cost
100k core CPU	1562	4PB	1250 kW	40	$29.5M
1000 GPU	250	1.3PB	750 kW	28	$21M

Table E.2 - Reference HPC Systems.

It should be noted that technology prices are highly fluid and can change rapidly. The guiding principle is to maximize capacity for researchers at the time of RFP with the predetermined budget available. This principle also applies to technology selection, which in this document has been simplified for ease of discussion and comparison. In the GPU space this is especially true as there are a growing number of GPU options available across multiple performance and price points. The most full featured and performant, as well as quite expensive, device was chosen for the GPU representative system, however it is unlikely that using only this one type of high-end GPU exclusively on all systems would provide the best value for researchers.

A similar approach was used for storage using three reference storage tiers, based on differing storage technologies. These are outlined in Table E.3. Choice of file system design, software licenses, and technology can greatly affect the storage pricing, however for these generalized cost estimates typical federation storage configurations are used. Project space is estimated including its dual-copy tape based backup, so works out to $150k/PB. Nearline also uses dual-copy tape so is estimated at $50k/PB.

Storage Type	Technology	Storage Tier	Cost/PB
High Performance	Solid State	Scratch	$500k/PB
Mid Range	Mechanical Disk	Project, dCache, Cloud	$100k/PB
Tape	Tape	Backup, Nearline, Archival	$25k/PB

Table E.3 - Reference Storage Systems.