A Cloud Ecosystem for Canada's Digital Research Infrastructure

Prepared for NDRIO's Call for White Papers on Canada's Future
Digital Research Infrastructure Ecosystem
Date: December 14, 2020

Contact: Wency Lum, Associate Vice-President University Systems & Chief Information Officer, University of Victoria wencylum@uvic.ca

Co-authors: Jeff Albert, Senior Advanced Research Computing Systems Administrator jralbert@uvic.ca
Ryan Enge, Manager Research Computing Services, University of Victoria renge@uvic.ca

INTRODUCTION

The government of Canada defines Digital Research Infrastructure[1] (DRI) as having four key elements: network, data management, research software and advanced research computing (ARC). Within the ARC domain, this has traditionally focused on high-performance computing (HPC); however, cloud computing is also a vital component of ARC and DRI.

Our team has been operating Arbutus, Canada's largest research cloud computing service, for over four years under the Compute Canada Federation (CCF). Through our experience as members of the ARC community working with cloud Infrastructure as a service (IaaS), we have observed how cloud computing enables Canada's research community.

In this paper, we propose a cloud ecosystem that has community cloud platform services, commercial cloud services, and cloud-specific scientific support, to serve a spectrum of needs for researchers. We begin by describing our observations of the current state of cloud computing and research, which has enormous opportunities and challenges. We then make the case for creation of platform services, a process for commercial cloud adoption, and the creation of a scientific support team with specialization in cloud for both community and commercial cloud services. The recommendations in this paper are not specific to the University of Victoria, and we envision moving forward as a community across multiple institutions and regions.

CURRENT STATE - CLOUD COMPUTING IN RESEARCH

Cloud computing provides on-demand availability of compute, storage, networking, and other technology resources. Cloud computing helps researchers to enable portals, platforms, and web visualization, and offers access to ephemeral resources for prototyping and medium-scale HPC workloads.

The current CCF cloud systems are implemented as an IaaS community cloud (i.e., private cloud within the CCF) that is provided in large part by Arbutus. There are other distributed cloud systems within the

[1] https://www.ic.gc.ca/eic/site/136.nsf/eng/home

CCF, such as at Beluga, Graham, and Cedar host sites. Examples of the ways that researchers are using cloud include:

Enabling access to cloud-based, container-orchestrated work-flow systems. Cloud computing plays an important role in helping artificial intelligence and machine learning researchers access cloud-based, container-orchestrated workflow systems, such as Kubeflow and Polyaxon. These play a critical role in managing the complexity of scheduling for huge numbers of individual job repetitions necessary to train machine learning pipelines.

Supporting shared scheduling of HPC-like workloads. Bioinformatics and ‘omics researchers are using community cloud for projects like iReceptor[2] and GenAP[3] to interactively schedule HPC-like workloads against centrally held datasets and collaborate on the results through portal interfaces.

Enabling interactive and collaborative workloads. Interactive and collaborative workloads, such as textual and media analysis, are common in the humanities and social sciences. Cloud infrastructure makes it possible to provide portal access to essential tools, such as Voyant and Islandora, which allow for them to work collaboratively across projects and institutions.

Providing alternative OS tools. Researchers in areas such as varied as social sciences and humanities, and geospatial analysis often have tools (e.g., NVIVO, ArcGIS) that require specific operating environments (e.g., Windows). Community cloud can be a critical support in enabling these environments on demand, without being limited to physical workstations held by individual projects or institutions.

Supporting rapid access to virtual HPC. Cloud environments can provide HPC capabilities that support key research activities. For example, the Magic Castle platform provides a cloud-hosted virtual HPC cluster that perfectly mirrors other HPC environments and can be used for training and prototyping.

Offering collaborative portals and publishing. Cloud computing contributes critical access to collaborative analysis portals like GenAP (for bioinformaticians), and web-specific platforms like Érudit.org (a Canadian scholarly publication platform), which allows open access to Canadian journal publications. These collaborative platforms provide access to research data and improve visibility and reproducibility in many fields.

CHALLENGES WITH THE CURRENT STATE

Although cloud has many uses for research, there are challenges:

Lack of IT skills and staffing. Researchers are currently provisioned with virtual machines and are responsible for building, securing, and maintaining the technology platforms themselves. To do this, research project teams must develop capabilities in information technology (IT) skill areas including

[2] http://ireceptor.irmacs.sfu.ca/
[3] https://genap.ca/p/help/introduction

system and network administration, storage and backup, infrastructure design, and cloud architecture. Each IT skill area is specialized and can be hard to recruit and retain. There is also the increasing importance of cybersecurity and data privacy, which adds additional complexity and responsibilities to IT roles.

Lack of scientific computing support and staffing. In addition to IT skills, cloud-allocated research projects must also develop their own scientific computing knowledge and staffing. Scientific computing knowledge is even more highly specialized and is required to make determinations about the best software, algorithms, tools, and practices particular to their fields of research and investigative techniques.

Duplication of technology effort between projects. Researchers are often trying to build and run the same technologies, such as Magic Castle, JupyterHub, Galaxy, etc. Each research project must bridge IT and scientific support gaps in isolation, resulting in duplication of effort between projects.

Risk of inequity. Large projects with generous technical staffing, projects attached to international collaborations with external technical resources, and projects in fields with a history of ARC have produced remarkable results. Smaller projects, those in emerging fields, and those in fields without a long history of ARC have fared less well.

Insufficient capacity. Demand for community cloud resources is currently governed by a resource allocation process where demand exceeds available capacity. Some researchers have turned to commercial cloud solutions to address this and other gaps, such as scalability and specialized services.

RECOMMENDATIONS

To address the opportunities for cloud computing in research and the challenges described above, we recommend that NDRIO shift community cloud services from solely IaaS to creation of platforms, and provide better scientific support for cloud, including support for researchers to choose and manage the right cloud solution, whether community cloud or commercial. These recommendations are not specific to any institution and are intended to be a national collaboration across multiple institutions and regions. They are described further below:

1) Create community cloud platforms for research

We can improve the challenges related to the lack of IT skills and staff, duplication of technology effort, and risk of inequity between projects by creating community cloud platform services (e.g., Magic Castle, JupyterHub, Kubeflow, R Shiny, etc.) that are frequently used in research.

Standardizing and centralizing the provision of common platforms is a more effective use of resources. Rather than multiple teams running multiple similar services with multiple sets of administrators, policy, support, and security outcomes, a single central team develops a single platform that serves the needs of all of them, with a single manageable set of outcomes and a reduced total cost. This allows research projects to focus on their investigative outcomes, while IT specialists focus on running the platforms. This approach also reduces single-expert risk in individual projects, ensures a consistent application of nationally aligned policies in areas like cybersecurity and research data management, and helps to overcome the technological barriers that can limit the success of smaller projects or those in emerging fields.

This work can be achieved through a national team, working with researchers to prioritize and build the platforms services that would most benefit researchers. The team is intended to be distributed, instead of being in a single institution. This distributed model can help grow regional and local expertise in cloud computing and ARC, as well as support NDRIO's vision of inter-provincial and inter-institutional collaboration.

Using a community cloud model for research platforms has the following benefits:

- **Cost Effective.** Community cloud is a cost-effective solution because it provides storage access for large and fast-changing datasets without the metered data egress and metered storage services of commercial clouds.
- **Experimental and exploratory workloads.** A community cloud environment enables experimental and exploratory workloads without the risk of cost overruns from metered compute resource access.
- **Network.** Community clouds at host site institutions are directly connected to Canada's National Research and Education Network (NREN), providing the performance necessary to access and transfer the huge footprint of research data without traversing the public Internet.
- **Tailored for research.** Community clouds have been purpose-built to serve research needs.

Governments and research funding agencies in other jurisdictions are also investing in community cloud capabilities, such as:

- **United States** - Jetstream[4]
- **Australia** - Nectar[5]
- **European Union** - European Open Science Cloud[6]

2) Enable commercial cloud's role in the cloud ecosystem

Researchers are also using commercial cloud services and this usage will grow due to demand for capacity, scalability, and new capabilities available. Commercial cloud works well for areas such as burst capacity and high specialization, such as artificial intelligence (AI) pipelines that are available with commercial cloud providers but cost-prohibitive to develop on community clouds. There should be a national strategy, training and education, and financial models to help adoption. The implications for cybersecurity, data privacy, long-term data storage, and research data management needs must also be assessed. Commercial cloud requires expertise to optimize for cost and performance and may not work well for all use cases.

3) Improve scientific computing support for cloud

Regardless of what type of cloud resources are used, whether community cloud or commercial cloud, a common difficulty that researchers face stems from a lack of scientific support resources with specialization in these tools.

[4] https://itnews.iu.edu/articles/2020/NSF-awards-IU-10M-to-build-Jetstream-2-cloud-computing-system-.php
[5] https://ardc.edu.au/news/our-platforms-investment/
[6] https://www.eosc-portal.eu/services-resources

A crucial element in the cloud ecosystem is the creation of a Cloud Scientific Support National Team to address this gap in scientific support. This could be a team of staff from multiple institutions and regions to provide on-demand support to cloud ARC projects as well as consulting on projects at or before allocation requests to provide awareness on types of resources, platforms, and services that can enable their work. As commercial cloud options become available in the national ecosystem, these cloud scientific support personnel could also be trained to help researchers optimize the use of available commercial solutions.

CONCLUSION

Cloud computing has enormous potential in ARC, and NDRIO has a unique opportunity to move Canada forward by investing in a cloud ecosystem model that serves researchers. There is a spectrum of need that is challenging to fill with a single model, whether that is community cloud or commercial cloud. We believe that a cloud ecosystem with community cloud platforms, commercial cloud services, and scientific support for cloud will enable better research outcomes for projects and NDRIO's vision of a world-class, collaborative, and competitive Digital Research Infrastructure community in Canada.