Spring 2026 HPC and Cloud computing Workshop

May 20-22, 2026

Advanced research computing Team /

Alliance Distrib. Workforce Support Site for Manitoba

Digital Research Alliance of Canada

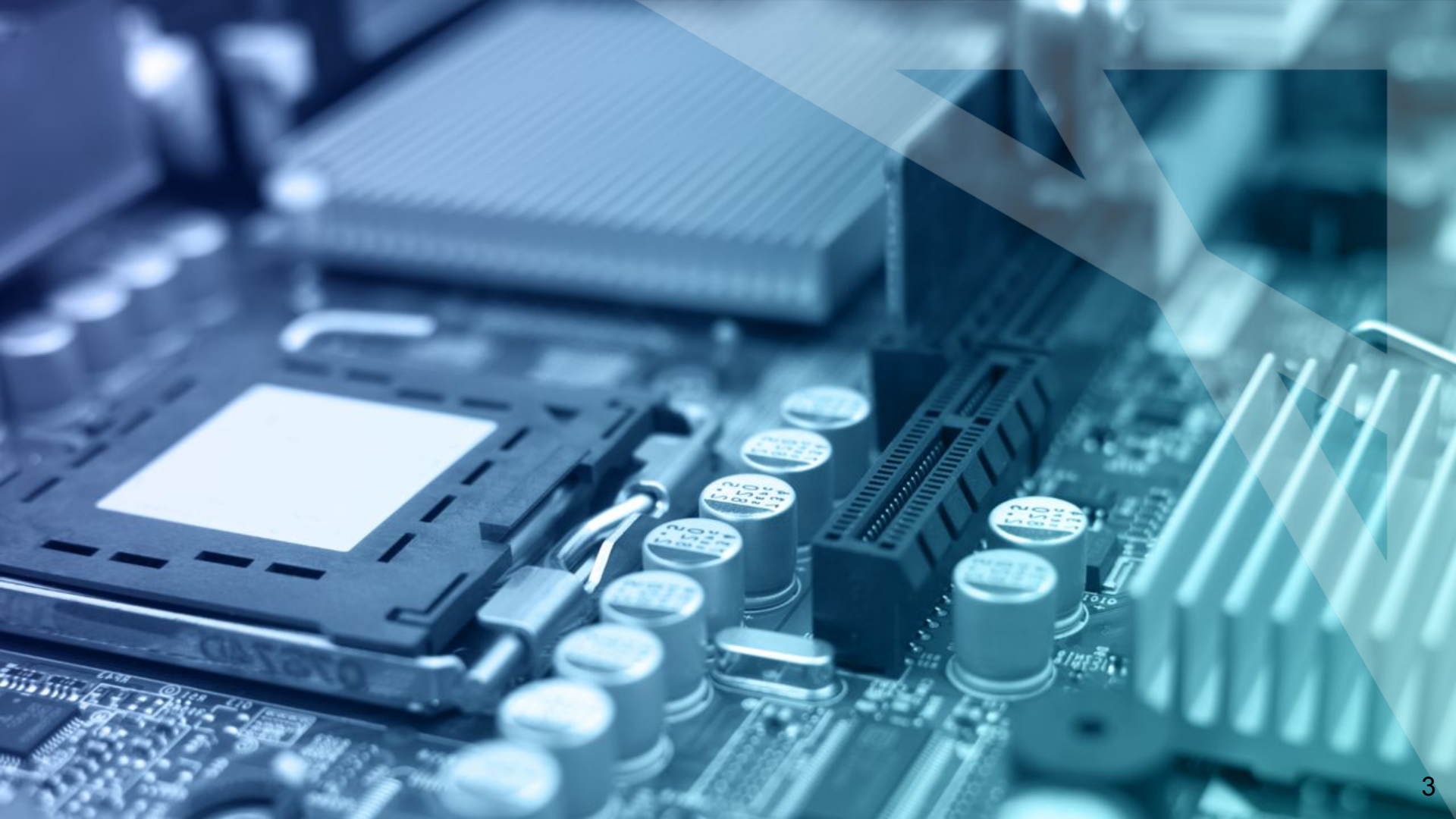
University of Manitoba

TRADITIONAL TERRITORIES ACKNOWLEDGEMENT

University of Manitoba

The University of Manitoba campuses and research spaces are located on original lands of Anishinaabeg, Ininiwak, Anisininewuk, Dakota Oyate, Dene and Inuit, and on the National Homeland of the Red River Métis.

UM recognizes that the Treaties signed on these lands are a lifelong, enduring relationship, and we are dedicated to upholding their spirit and intent. We acknowledge the harms and mistakes of the past and the present. With this understanding, we commit to supporting Indigenous excellence through active Reconciliation, meaningful change, and the creation of an environment where everyone can thrive. Our collaboration with Indigenous communities is grounded in respect and reciprocity and this guides how we move forward as an institution.



What to expect of this Workshop?

Overview of DRI, in particular Research Computing resources

- What is available nationally and locally?
- How it can be useful for your research?
- How to get access and support

Better understanding of the resources kinds and capabilities

- ARC, RDM, RS overview (focus on ARC)
- HPC and cloud computing, various kinds of storage, data movement
- How so these sit with the emerging AI technologies?

Hopefully, improving your skills at getting most of the resources!

- Batch computing, getting and using Software, using IaaS cloud

Who we are

“Advanced Research Computing Team”

- Running UManitoba’s own HPC system Grex

Digital Research Alliance Support Site

- Supporting users from all of Manitoba on National DRI
  - Training and outreach!
- Contributing to DRACs National and Regional Teams
- Grant consultations for CFI and other contributed hardware

Grigory Shamov; Ali Kerrache; Stefano Ansaloni; Kamrul Shaker

Program of the Workshop

Start	End	May 20 — Theme: Available Resources, Software, and Data Movement
10:00	10:45	Introduction to Alliance and Manitoba HPC, AI and Cloud Computing resources
10:45	11:00	Housekeeping: how to connect to and use the training systems
11:00	12:30	Linux Command Line Interface: BASH and SSH (hands-on)
12:30	13:00	*break*
13:00	14:20	Intro to HPC software, Lmod modules tool (hands-on)
14:20	15:00	Data transfers and storage with Globus, NextCloud

Start	End	May 21 — Theme: SLURM Jobs, OpenOnDemand Web Portal, and Containers
10:00	10:10	Housekeeping: how to connect to and use the training systems
10:10	11:00	OpenOnDemand HPC Web portal: File Transfer , Remote Desktop and interactive GUI applications
11:00	12:00	Containers in HPC: using Singularity/Apptainer (hands-on)
12:00	12:30	Containers in HPC: using Podman and Pixis
12:30	13:00	*break*
13:00	13:30	OpenOnDemand HPC Web portal: Running Jobs with Job Composer
13:30	14:30	Running HPC jobs with SLURM scheduler (hands-on)
14:30	15:00	Using Jupyter notebooks on HPC systems

Start	End	May 22 — Theme: AI , OpenStack Cloud and Object Storage
10:00	10:10	Housekeeping: how to connect to and use the training systems
10:10	11:30	AI on HPC systems topics
11:30	12:30	Introduction into Using OpenStack Cloud: dashboard (hands-on)
12:30	13:00	*break*
13:00	13:45	Using OpenStack Cloud, deploying Web applications (hands-on)
13:45	14:45	Using OpenStack CLI and ObjectStorage
14:45	15:00	Closing Remarks

An overview of Canadian Digital Research Infrastructure ecosystem

Grigory Shamov

May 20, 2025

Digital Research Alliance of Canada

University of Manitoba

- What is ARC and DRI, and why should you care
- National Digital Research Infrastructure
- Local HPC offerings

- Technology : HPC, CLoud, AI
- How to access and use these resources

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

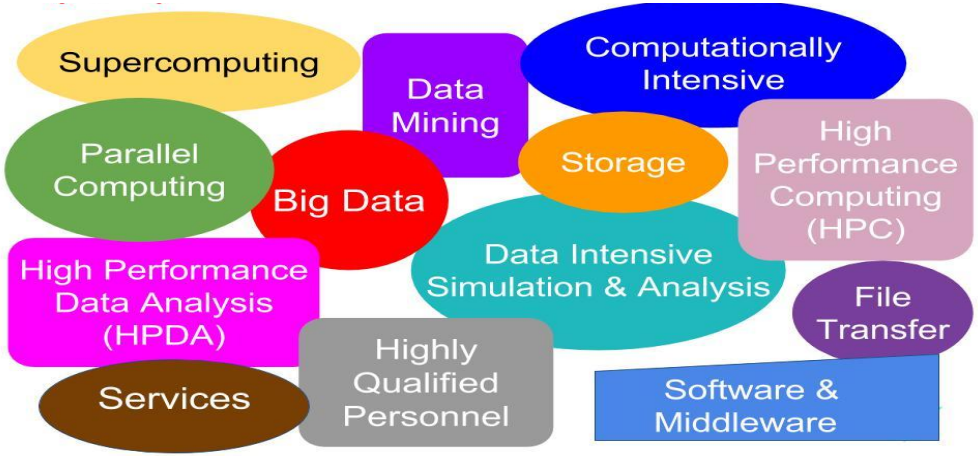
University of Manitoba

What is Advanced Research Computing?

Advanced Research Computing (ARC) is **everything beyond a standard desktop workstation**. This includes:

- Cloud computing for research
- High Performance Computing
- Data movement
- Data storage
- Services and Support
- highly qualified personnel

… **all in support of research.**



Slide by Alex Razoumov, SFU Training coordinator

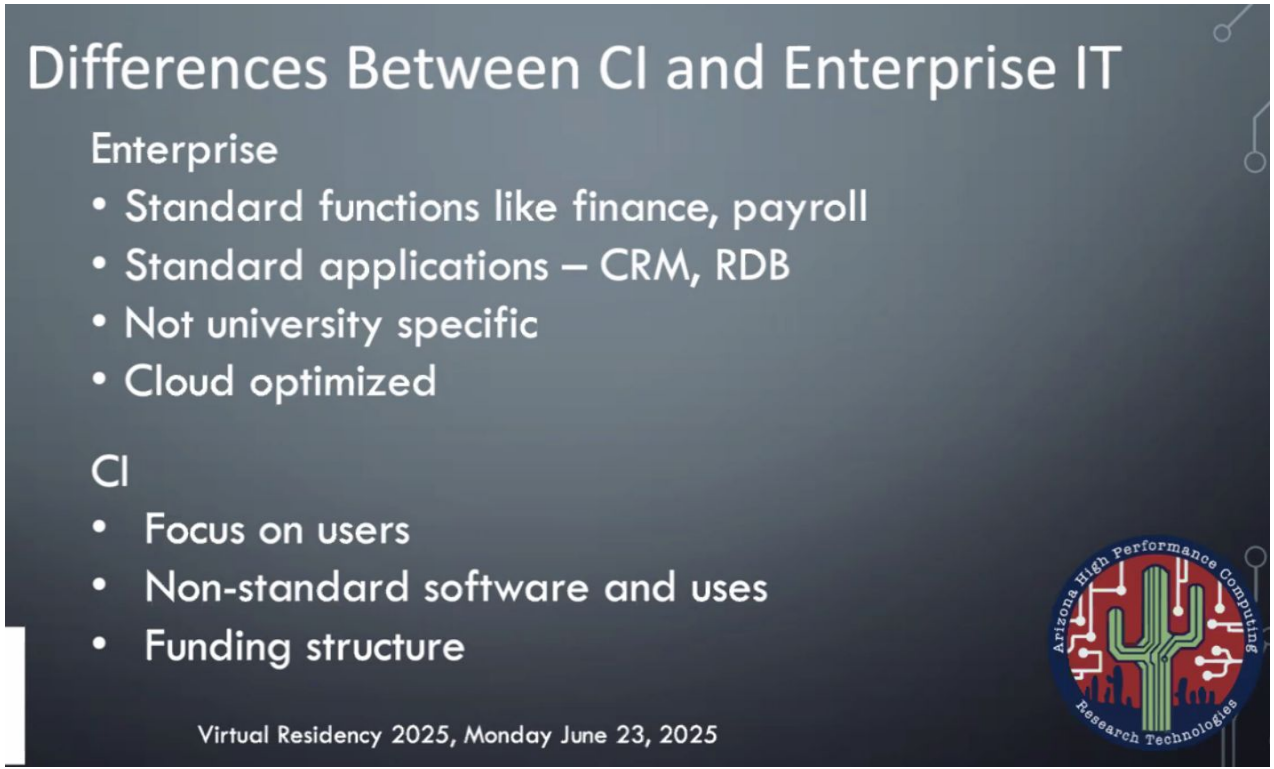
DRI (Digital Research Infrastructure)

“Advanced Research Computing” (ARC), HPC, Cloud etc.

- Focuses on enabling computational research / users
- Provides capabilities that are not available with common (desktop, enterprise server) computing environment:
    - CPU time and memory, fast interconnect
    - Accelerators (GPU, TPU hardware)
    - Fast and scalable storage capacity
    - Network/data transfer resources optimized for research
- A specialised set of hardware and software
- Managed differently from Enterprise IT, more research-specific

The two most popular modes of delivery are “Traditional **HPC**” and “**Cloud** computing”

HPC/ARC/CI as an form of IT ?



Slide by Henry Neeman, OSU RC Director

High Performance Computing history

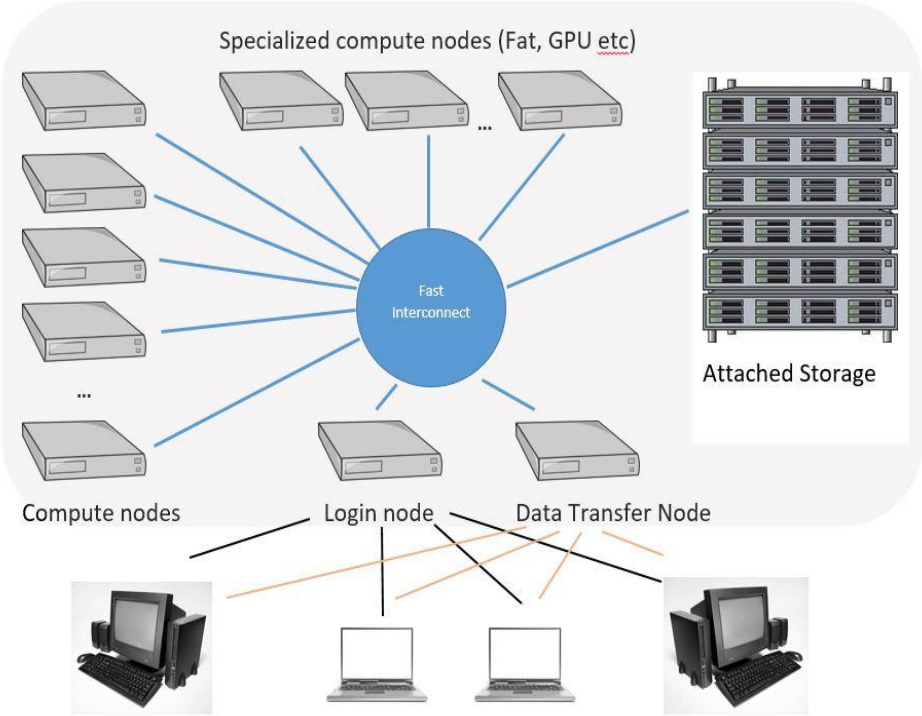
Quite old by now! Started in late 1990s-early 2000s.

- Beowulf clusters, 10s of PCs. – “compute nodes”
- Single shared system under command of a “Batch Job Scheduler”
    - Quickly evolved into 100s and 1000s of nodes
    - Low-latency interconnect and parallel programming
    - Fast and scalable storage was needed
    - Later, accelerators like GPUs become popular

- ARC systems tend to be large and thus expensive (but efficient)
    - Triggered formation of Consortia, National HPC

The two most popular modes of delivery are “Traditional **HPC**” and “**Cloud** computing”

HPC as a technology, architecture

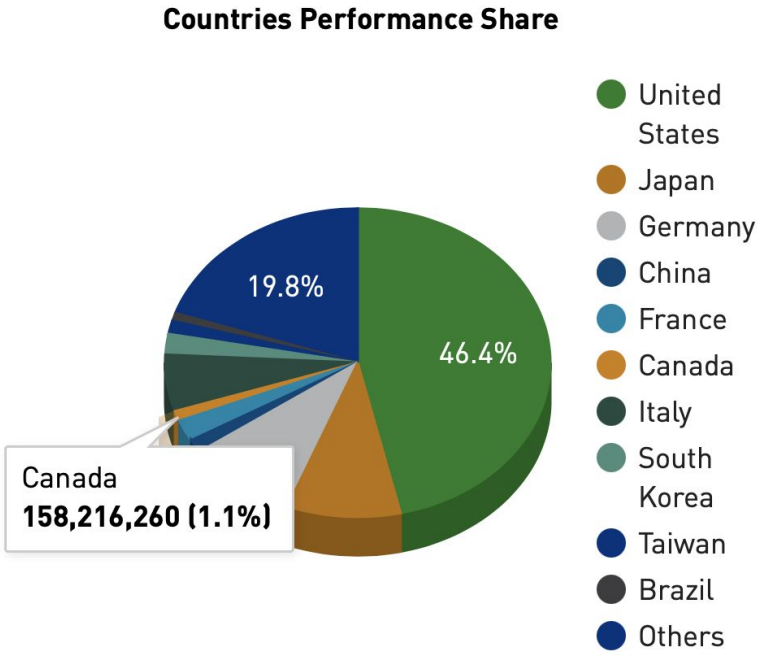


High Performance Computing

Top500 rating in 2025 : Exascale HPC

https://www.top500.org/

- Traditionally led by US DoE systems. El Capitan #1, Frontier #2, Aurora #3
- European first Exacale System JupyterBooster #4, then LUMI, Leonardo, ..
- Japan's Fugaku
- …



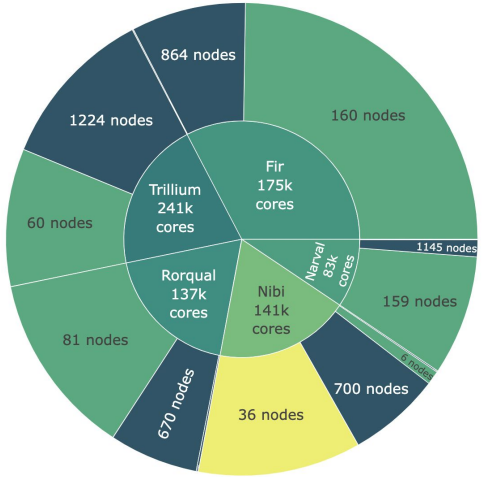
- Canadian HPC renewal in 2025 (~220M):
    - systems (Fir, Trillium) mid- Top-100

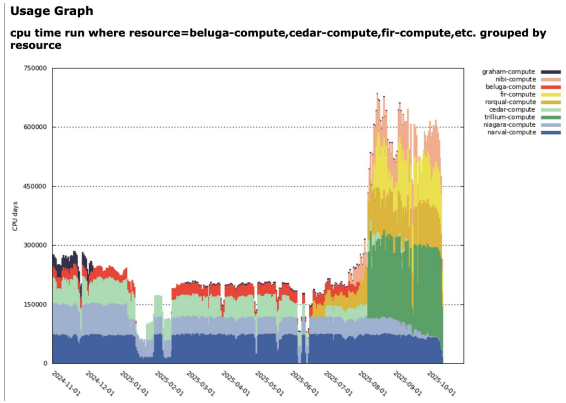
Size of National HPC systems 2025

Digital Research Alliance of Canada

Infrastructure renewal (**~220M**) for HPC and Cloud in 2022-2025.

- Fir (SFU)
- Nibi (UWaterloo)
- Rorqual (ETS Montreal)
  - Narval , same DC
- Trillium (UToronto)
- Arbutus Cloud (UVictoria)
  - Rolling update





https://wgpages.netlify.app/files/clusters.html , *an infographics by Alex Razoumov @ SFU*

Cloud Computing for ARC

A newer way to deliver Started in 2010s.

- From Web-hosting and Virtual Machines technologies
- Hosting services (websites, databases) – can also compute
- “Tenants” isolated using Software defined tech: SDNs, multi tenant storage
    - Quickly evolved into 100s and 1000s of nodes
    - Low-latency interconnect was lacking for a long time
    - Developed “Object Storage” like Amazon S3
    - Public cloud providers , commercial
    - “*Cloud is a business model*”
- Started to shift into providing compute too, GPUs, etc.
- More costly, but much more advanced infra for services (load balancers, etc.)
- Community clouds developed : OpenStack project

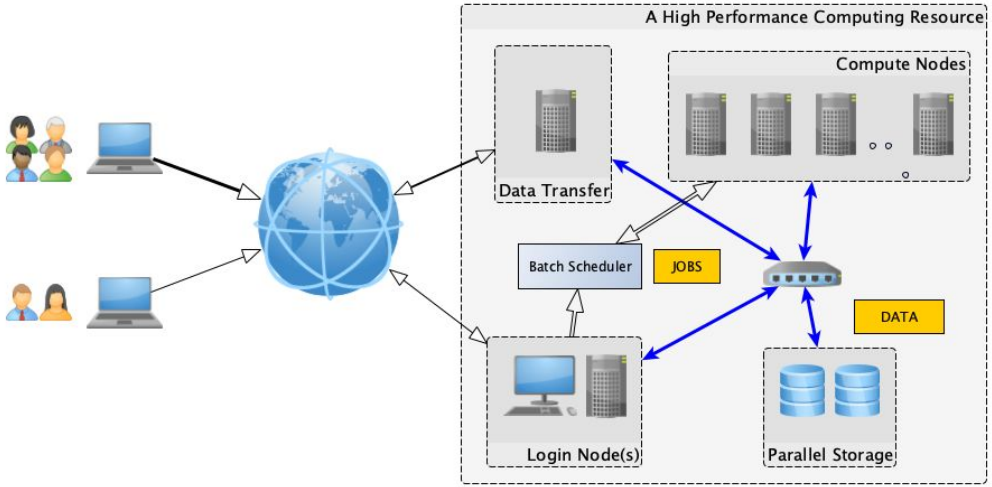
The two most popular modes of delivery are “Traditional **HPC**” and “**Cloud** computing”

DRI (Digital Research Infrastructure)

- **HPC** is about efficiently organizing shared, bare-metal resources for top performance:
    - CPU, accelerators, scalable storage,
    - High-speed Interconnect.
    - High-density datacentres
    - Exascale HPC and High-Throughput computing / streaming

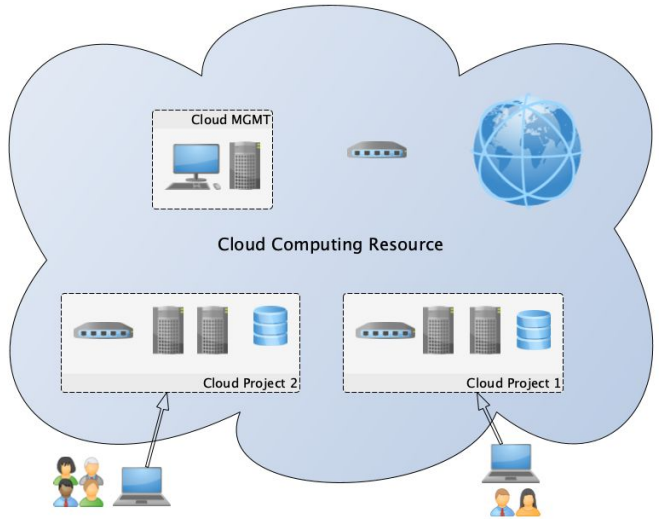
- **Cloud** computing originally focused on Enterprise computing
    - Business flexibility : virtualization, software-defined services
    - Standard datacentre,
    - Running persistent services.
    - Container orchestration platforms (Kubernetes)
    - But can (and has to be in some cases) be used for ARC!

HPC vs Cloud computing



- HPC clusters are *shared* systems with r*emote access*
- Batch mode of usage
- A central Software delivery on HPC
- Dealing with Data (storage, transfer etc.)

- Public and Community Cloud computing
- Flexible, elastic, Provides isolation of tenants
- SDN, SDS, Virtualized compute
- Self-service




Artificial Intelligence and Machine Learning revolution
https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/35179.pdf
"The Unreasonable Efficiency of Data"
http://www.incompleteideas.net/IncIdeas/BitterLesson.html "The Bitter Lesson"
https://dl.acm.org/doi/10.5555/3295222.3295349 "Attention is all you need""

DRI for AI
(last year's Slide….. What have changed?)

IN MANY WAYS, AI VINDICATES THE "HPC WAY"

- ***AI needs fast interconnects***. We had them, the cloud and the enterprise did not.
  - Microsoft deployed 40,000 KM of *Infiniband*, in 2023, built for the HPC market ~1999,
- ***AI needs message passing***. MPI, the message passing interface, was built Open Source in the HPC community, ~1993
  - Now the standard library for transformer-based generative AI (e.g. ChatGPT, DeepSpeed, OpenAI etc.).
- ***AI needs heterogeneity*** – GPUs for general purpose computing – the hardware building block for AI - came out of the HPC world ("GPGPU" ~2004).
- ***AI needs fast, large scale filesystems –*** not object stores
- ***AI needs liquid cooling*** – even 5 years ago, many datacenter providers were convinced they could just use air, now none are. HPC systems switched to liquid cooling a long time ago.
- This means AI needs HPC hardware (probably good) and HPC programmers (good if you are one, bad if you need to hire one).

Slide by Dan Stazione, Director of TACC, US

***tamIA**, an AI supercomputer of LavalU*

What changed since?

AI workloads meant model Training and Inference

- Training shops did ok in HPC settings (either SLURM or Kubernetes)
- By 2026, a large shift to inference!
- RAGs, Vibe Coding using public Inference services/APIs
- Agentic AI

- Seems that the scene shifts to Inference workloads, and many of them DevOps/Cloud-like etc. rather than HPC

The “Sovereign AI” idea, to decrease the risks of using US Big Tech

- Canada has issued 80M stop-gap funding in 2025
- AI SCIP program call, 800+M$ is open now => will likely change DRAC
    - https://ised-isde.canada.ca/site/ised/en/ai-sovereign-compute-infrastructure-program

DRI requires a large capital investment

DRI requires an operation to run it:

- staff, expertise, real estate (datacentre), electricity

In Canada, there are two levels of DRI above “under your desk”:

- Institutional or “local” system.
  - Funded by Universities and / or projects (Faculties, Departments, PI grants)
- National DRI Organization / Consortium / Federation.
  - Funded by the Government of Canada with Provincial and/or Institutional matching
  - Currently, DRAC https://alliancecan.ca

About the Digital Research Alliance

National Scope and Mandate

Member-Based

Non-Profit

Federally-Funded

The Digital Research Alliance of Canada (the Alliance) integrates, supports and improves access to digital research infrastructure (DRI) to deliver world-class digital tools and services for Canada's researchers.

By enhancing and transforming how Canadian research data is organized, managed and shared, we empower researchers to lead discovery and innovation with global impact.

What *is* Digital Research Infrastructure?

Technology, tools and services for global impact.

Digital research infrastructure (DRI) is the collection of tools and services—including AI-enabling infrastructure—that allow researchers to turn big data into new discoveries and breakthroughs.

Three Pillars of DRI:

1. **Advanced Research Computing (ARC)** | High-performance computing and storage, underpinned high-speed networks, systems software, standards and services
2. **Research Data Management (RDM)** | Services supporting the full project lifecycle including data collection, management, storage, preservation, sharing, discovery and access.
3. **Research Software (RS)** | Cutting-edge, integrated research platforms, tools and services to increase access to DRI and data.

Digital Research Alliance of Canada

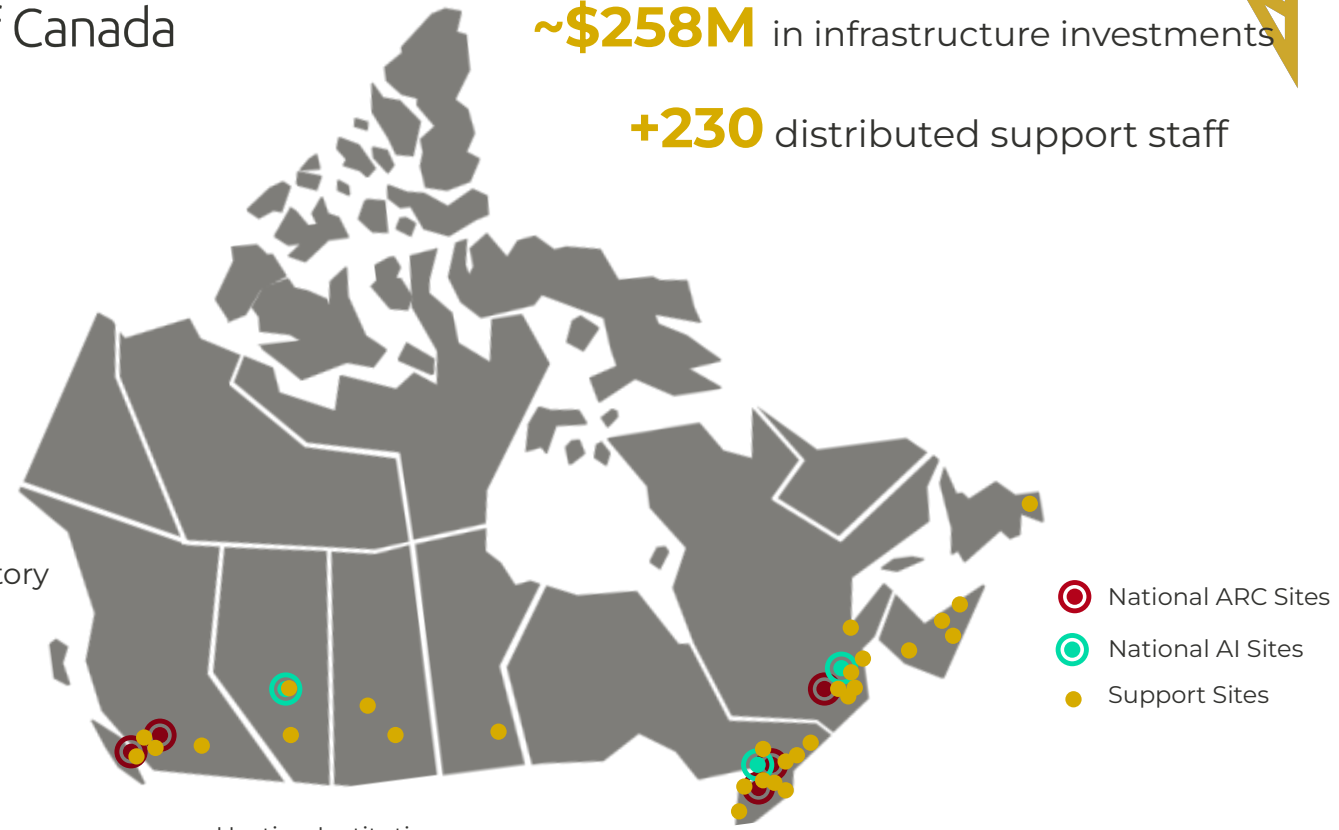
+26,000 users

~$258M in infrastructure investments

+230 distributed support staff

Alliance National Services:

- National Identity Management
- Merit-based resource allocation
- Cybersecurity Frameworks
- National Help Desk
- Explora Training Portal
- Service Catalogue
- Federated Research Data Repository
- Data Management Planning Assistant
- Lunaris Data Discovery Service



Regional service delivery partners:

Calcul Québec · Compute Ontario · ACENET accelerate discovery

Hosting Institutions:

McGill University · SFU Simon Fraser University · University of Alberta · Université Laval · University of Toronto · University of Victoria · University of Waterloo

DRI (Digital Research Infrastructure)

“Advanced Research Computing” (ARC), HPC

- Focuses on enabling computational research / users
- DRAC Provides compute and storage capabilities (HPC, cloud)
- Data transfer capabilities (Globus)
- Active research phase. This Workshop is about ARC!

“Research Data Management”

- Archive data storage, Data curation, DMPs, etc.
- Libraries-focused. DRAC provides Lunaris, Borealis
- New pilots in National Data Spaces, Canadian Research Data Platforms.

“Research Software”

“Research Network”

- Provided by another org, CANARIE
- Optics cable networks for Research across Canada, 100 to 400 Gbps.

Digital Research Alliance of Canada

Digital Research Alliance of Canada

- Infrastructure renewal (**~220M**) for HPC and Cloud in 2022-2025.
  - Completed in September 2025!
  - Total: **5,175** nodes, **776,912 CPU** cores, **2,152** GPUs (A100, H100)
  - ..and counting

- PAICE systems: AMII, Vector Institute and MILA (**~40M**)
- Stop-gap 80M AI expansion in 2025-2026
  - More B200, H200 GPUs for HPC and PAICE systems. Being deployed now!

- ISED AI SCIP call. Sovereign AI , to close in a week, Jun 1
  - https://ised-isde.canada.ca/site/ised/en/ai-sovereign-compute-infrastructure-program
  - Expected to be a private-public partnership led by a public org (University, Consortium or National Org.)

Example of a the HPC system Fir

Digital Research Alliance of Canada

https://docs.alliancecan.ca/wiki/Fir

- **175,104** CPU cores, **640** GPUs – **#78** in July's Top500 HPC for GPU-only
- 864 base CPU nodes:
  - 192 cores, 750 GB memory, 2x AMD EPYC 9655 Turin CPUs
- 8 large-memory CPU nodes:
  - 192 cores, 6000 GB memory, 2x AMD EPYC 9654 Turin CPUs
- 160 GPU nodes:
  - 48 cores, 1125 GB memory, 1 x AMD EPYC 9454 Genoa CPU
  - 4x NVidia H100 GPUs 80GB VRAM, SXM5
- **51 PB** high-performance Storage (DDN Lustre FS, all-SSD)
  - */home* , */scratch* , */project* partitions
- InfiniBand NDR interconnect **400Gb/s**

Size of PAICE systems for AI

Digital Research Alliance of Canada

System, kind (2025)	# GPU nodes	GPUs per node layout	Interconnect	Storage, PB
TamIA , HPC (Laval)	42 (H100)	**4 x** NVIDIA HGX H100 SXM	**4 x** HDR200 Infiniband, **non-blocking**	0.7 PB
Vulcan, HPC (UofA)	205 (L40s)	4 x NVIDIA L40s	1x100Gbps Ethernet	5.0 PB
Killarney, HPC (UofT)	168 (L40s) 10(H100)	4 x NVIDIA L40s, 8 x NVIDIA H100 SXM	1x HDR100, 2x HDR200	1.5 PB
HPC systems, GPU				
Fir, HPC	160	4 x NVidia H100 SXM	1x NDR200 Infiniband, blocking	51PB
Nibi, HPC	36	8 x Nvidia H100 SXM	1x Nokia 200/400G Ethernet	25PB
Trillium, HPC	60	4 x NVidia H100 SXM	1x NDR200/NDR400 Infiniband	29PB

Grex and HPCC

A step-up machine for UM users to DRAC resources

Grex used to be a Westgrid / National HPC machine kept and maintained by the University of Manitoba as a local system. ~10K CPUs, ~2PB of storage

- Authentication through DRAC CCDB systems
- Managed by the same local DRAC Federation team.
- Provides both a local and the DRAC software stacks
- *Hosts CFI and other user-contributed hardware*

Renewed with help of the IST and VPRI funding and SISF 2024

- Serves ~160 research groups, ~500 users

System, kind (2024-2025)	CPU cores	GPU devices
Intel CascadeLake	2820	36 V100
AMD Genoa	7368	4 L40s
AMD rome	112	12 A30
	~10K	**52**

Access to and use of ARC resources

Access, Costs of Alliance Resources?

- Free for Eligible PI's : https://ccdb.alliancecan.ca
- Access through CCDB (ComputeCanada user DB)
  - The unit of resource allocation is "Research Group".
    - Not a university, not individual users
  - Thus, a Principal Investigator has to apply first, then "sponsor":
    - Undergraduate students research projects
    - External collaborators
    - Grad students, postdocs, support staff
- Eligible use: for academic research
- A user can belong to more than one research group.

What is useful for me in ARC?

Service	Limits?	Notes
CPU and GPU Compute	Default allocation 200 CPU core-years;Default allocation of ~1 year of H100. more with RAC. H100, A100, H200 and B200. More with RAC!	CPUs with fast interconnect, next to a scalable storage. Can be useful for Large Parallel or High-Throughput workloads *Note that it would not make serial code any "faster"*
Cloud Compute	80 VCPUs/year; 10TB of ObjectStorage, IPs, 10TB of local volumes, vGPUs, etc. More with RAC.	OpenStack cloud(s) for both Persistent and Ephemeral Computing workloads.
Active, Scalable HPC Storage	40TB/group ; more with RAC	Directly attached HPC storage; tape backup and "Nearline" *"Moving compute closer to the data"*
Globus Data Transfer and Data sharing	n/a	Allows for moving data in, out, and between Alliance systems. Allows for data sharing and publishing.
NextCloud	100Gb	A Dropbox-like service. Several instances exist in the ecosystem.
CVMFS Software stack	n/a; thousands of software items	A curated, central software stack distributed across Alliance and other systems (like Grex)
Visualization, interactive compute	n/a	Support for Jupyter, OpenOnDemand, and remote viz ParaView

Access, Costs of Alliance Resources?

- Free for Eligible PI's : https://ccdb.alliancecan.ca

- Default allocation (CPU, storage)
  - Immediately available for active accounts
- RAS (Rapid Access Service), on request any time of the year.
  - Mainly for Storage and Cloud resources.
- RAC (Resource allocation call) for increased requests, annually (Nov.)
  - Application process through CCDB
  - Proposal needed for CPU and GPU years, Storage TBs, cloud resources
- AI systems follow their own Tiered allocation process
  - Tier-4 for non members, needs an application.

What ARC-specific skills do I need?

- Working with remote systems
- Working with Command Line Interfaces (CLI)
  - Automation and AI ♥ CLI

- HPC works in batch mode
  - Interactive GUI / Jupyter as jobs, OOD and JH portals
- Centralized, optimized software stacks
  - Or bring your own software via containers.

- Persistent/Server workloads can be served by OpenStack Cloud (UVic)
  - Familiarity with Infrastructure as a Service (Iaas); Object Storage


University
of Manitoba