Policy Considerations on the Convergence of High Performance Computing & Artificial Intelligence

Moderator: Nizar Ladak

Panelists: Chris Loken, Alison Paprica, Suzanne Talon & Alain Veilleux

Government of Canada

2018 Federal Budget – “Harnessing Big Data”

“The Government proposes to provide $572.5 million over five years, with $52 million per year ongoing, to implement a Digital Research Infrastructure Strategy that will deliver more open and equitable access to advanced computing and big data resources to researchers across Canada.

The (Federal) Minister of Science will work with interested stakeholders, including provinces, territories and universities, to develop the strategy including how to incorporate the roles currently played by the Canada Foundation for Innovation, Compute Canada and CANARIE, to provide for more streamlined access for Canadian researchers”.

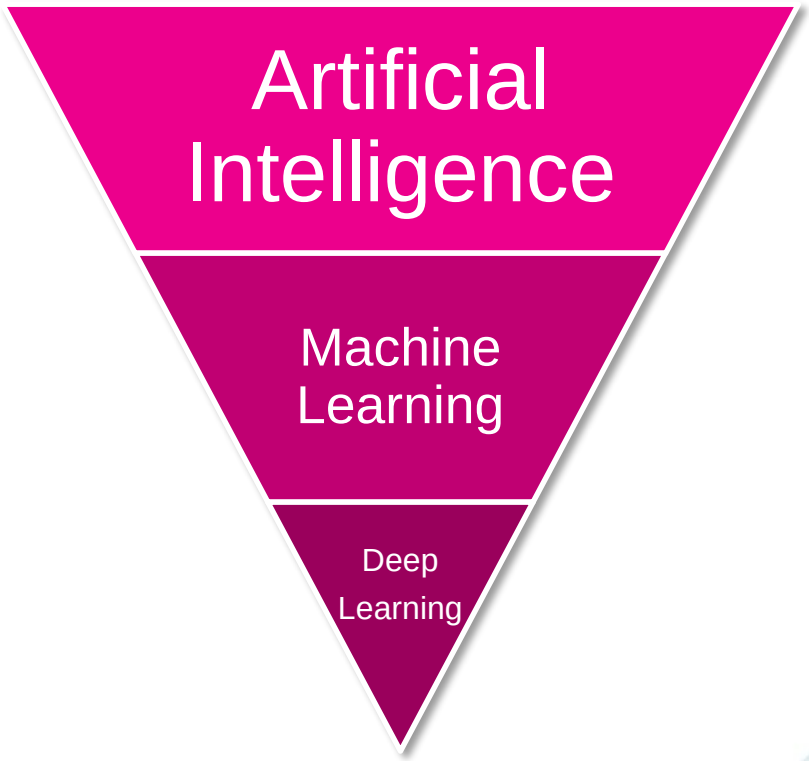
Compute • Calcul Ontario

What is HPC?

High Performance Computing is the use of high-end computing resources (computers, storage, networking and visualization) to help solve highly complex problems, perform business critical analyses, or to run computationally intensive workloads that are in scale far beyond the tasks that could be achieved on today's leading desktop systems.

Compute • Calcul Ontario

AI, ML & DL



Compute • Calcul Ontario

Artificial Intelligence

- Generally speaking, artificial intelligence refers to computers that can learn about the world flexibly, make inferences about what they see and hear, and achieve human-like understanding of information. Artificial intelligence involves capacities like visual and auditory perception, the ability to read and to make sensible decisions, and the power to make accurate predictions based on existing knowledge.

Machine Learning

- Machine learning algorithms give computers the power to learn and recognize patterns the way people do, without requiring specific and repetitive instructions for each new piece of data.

Deep Learning

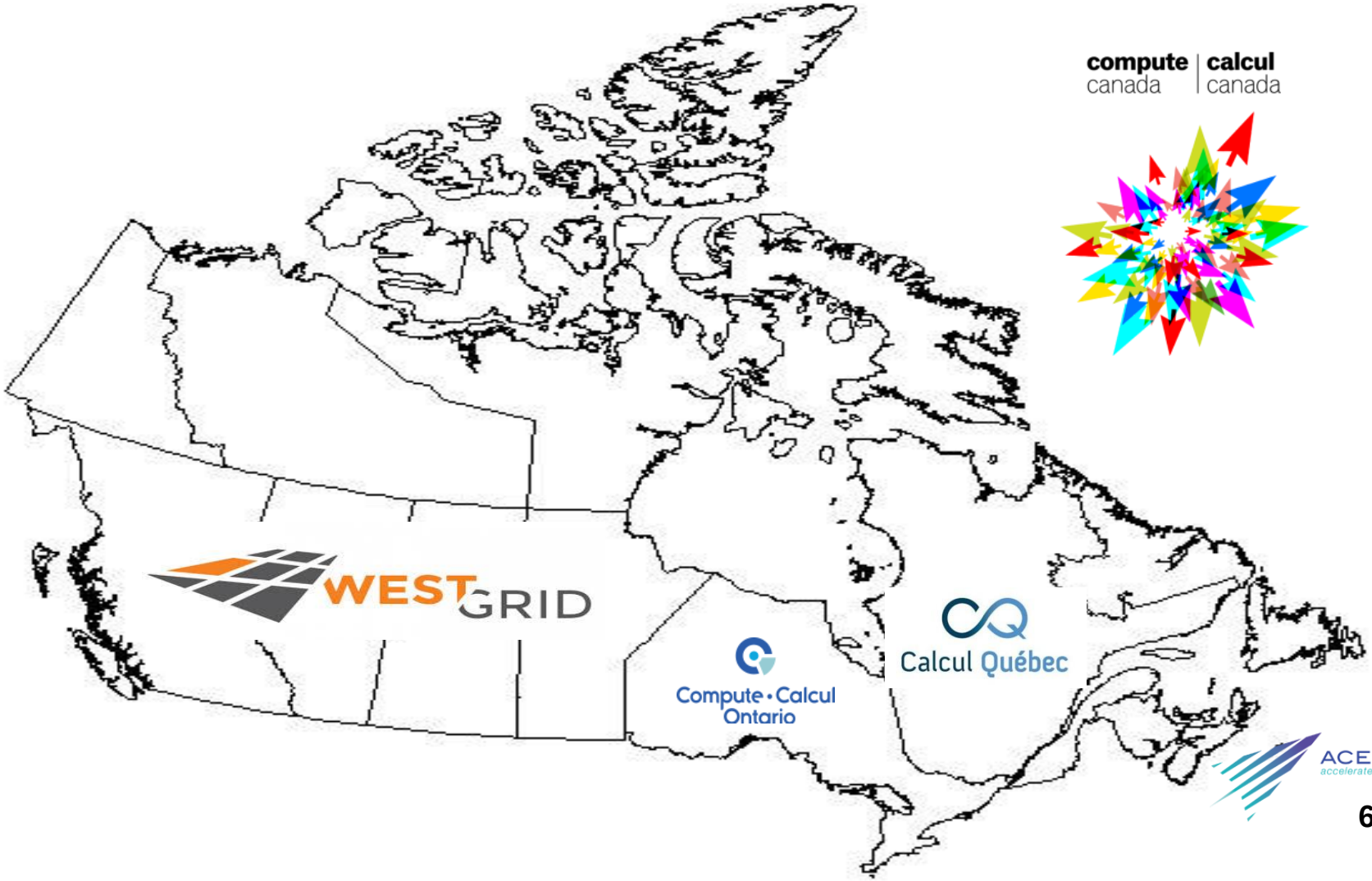
- Deep learning is a subfield of machine learning. Many other types of machine learning work well with limited data sets, but deep learning alone appears to continuously improve as the machine absorbs more and more data.

What is Digital Research Infrastructure?

DRI is a general term that includes a robust network, advanced research computing (ARC), data management, research software, and highly qualified personnel (HQP) to maintain Canada's DRI resources and train researchers in advanced computing techniques.

Compute • Calcul Ontario

The National DRI Landscape



How Advanced Research Computing (ARC) is funded in Canada

40% Federal

40% Provincial

20% Institution

Compute • Calcul Ontario

Emerging Policy Topics…

- skilled labour (Highly Qualified Personnel),
- technology investment,
- big data,
- public-private collaborations,
- multi-level governance,
- funding,
- evaluation and impact,
- national DRI policy (HPC, AI)
- Others…

Compute • Calcul Ontario

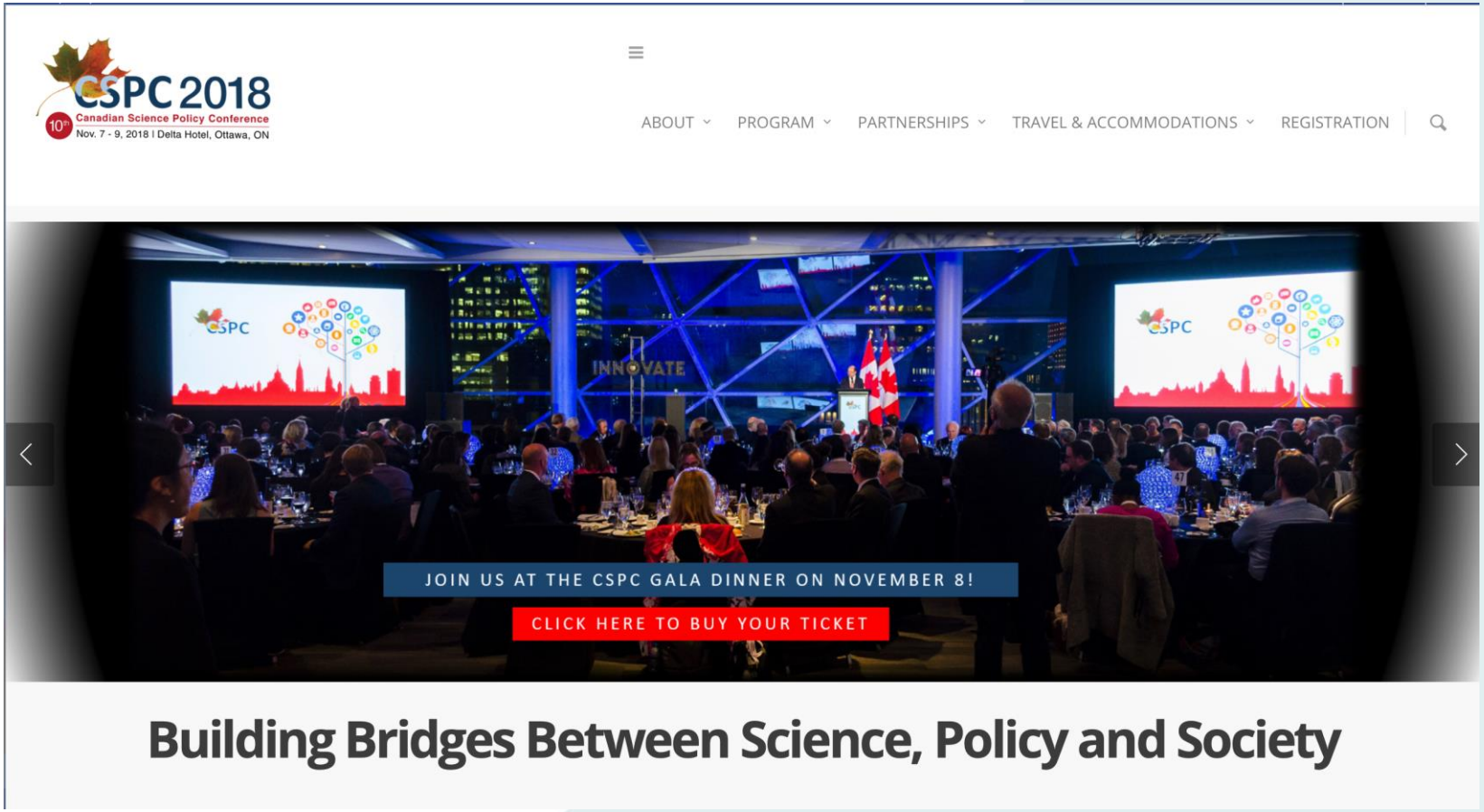
The way we have organized and collaborated speaks to our very identity as Canadians

Melting Pot

Canadian Mosaic

Compute • Calcul Ontario

CSPC 2018 Theme



Compute • Calcul Ontario

Our Panelists:

Suzanne Talon, CEO Calcul Quebec

Alain Veilleux, CTO Calcul Quebec

Chris Loken, CTO Compute Ontario

Alison Paprica, VP Health Strategy & Partnerships Vector Institute

Compute • Calcul Ontario

Calcul Québec

Canadian Science Policy Conference 2018

Suzanne Talon, Directrice générale

Alain Veilleux, Directeur de la technologie

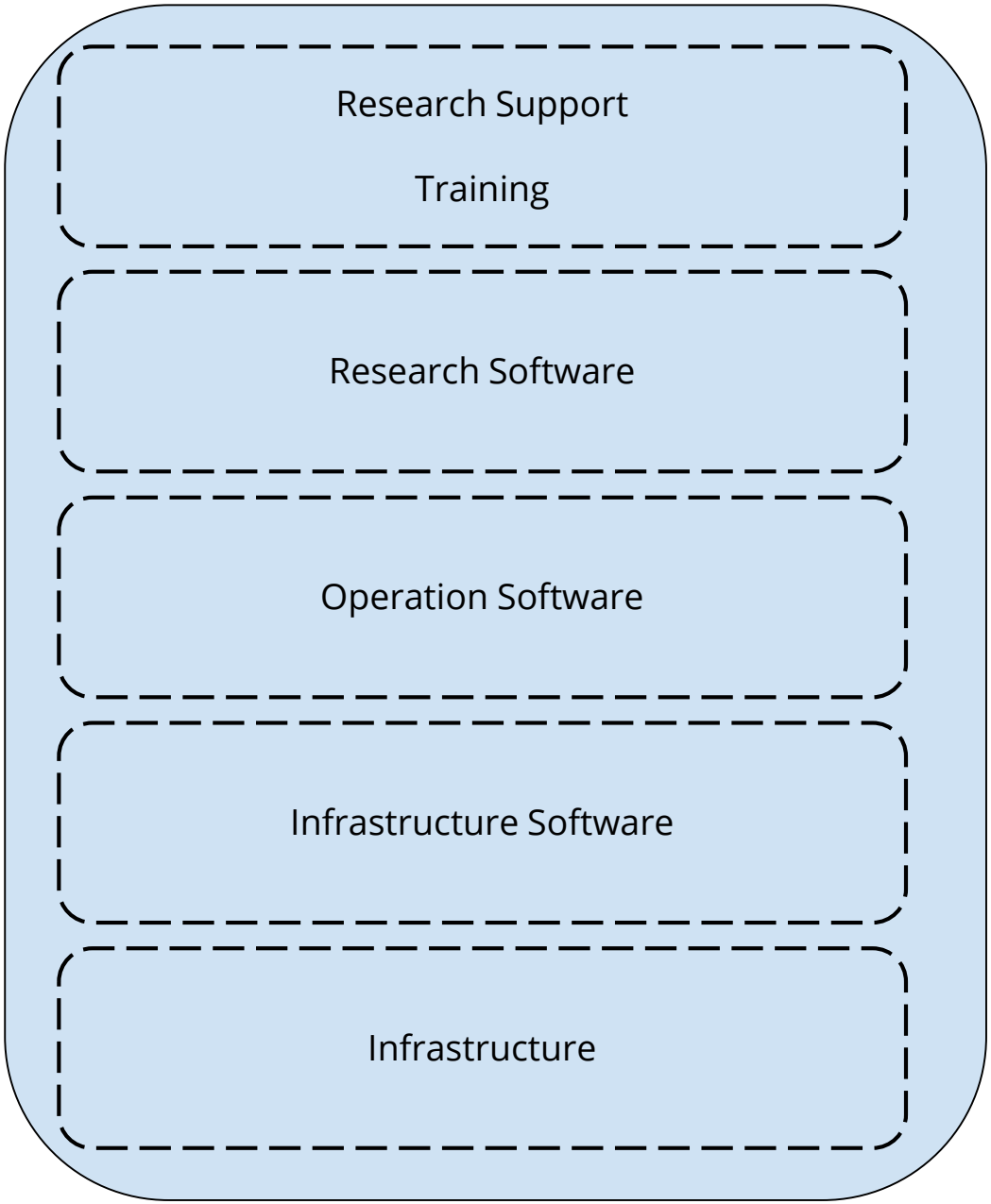
Calcul Québec

computecanada regional partner | calculcanada partenaire régional

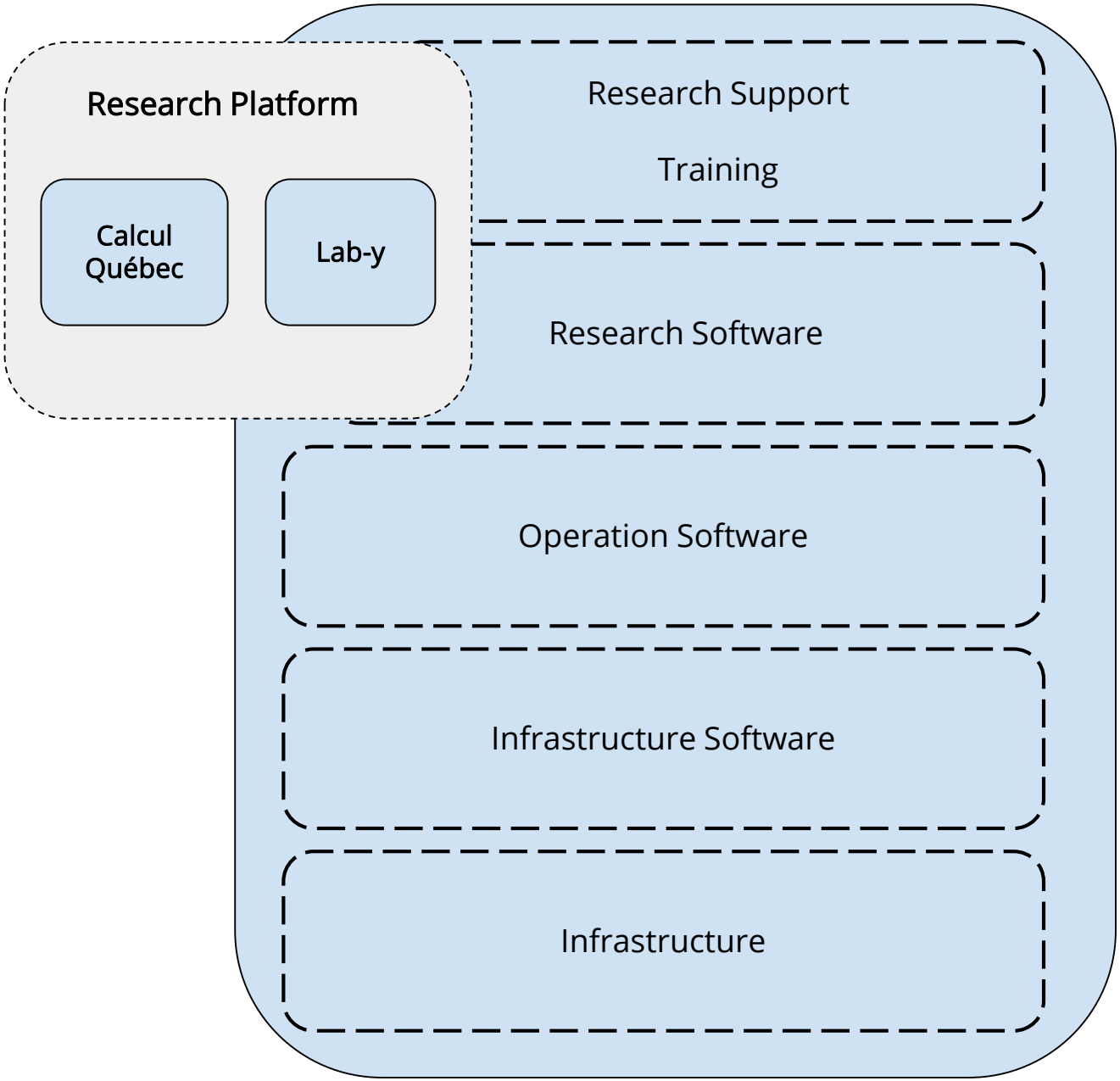
What is Calcul Québec?

Calcul Québec is an umbrella organization composed of Québec Universities brought together by advanced research computing (ARC). These member Universities, committed to better serve the needs of the research community through collaboration, pooled together their human and material ARC resources under the Calcul Québec banner. To meet the demands of today's researchers, Calcul Québec relies on three main components:

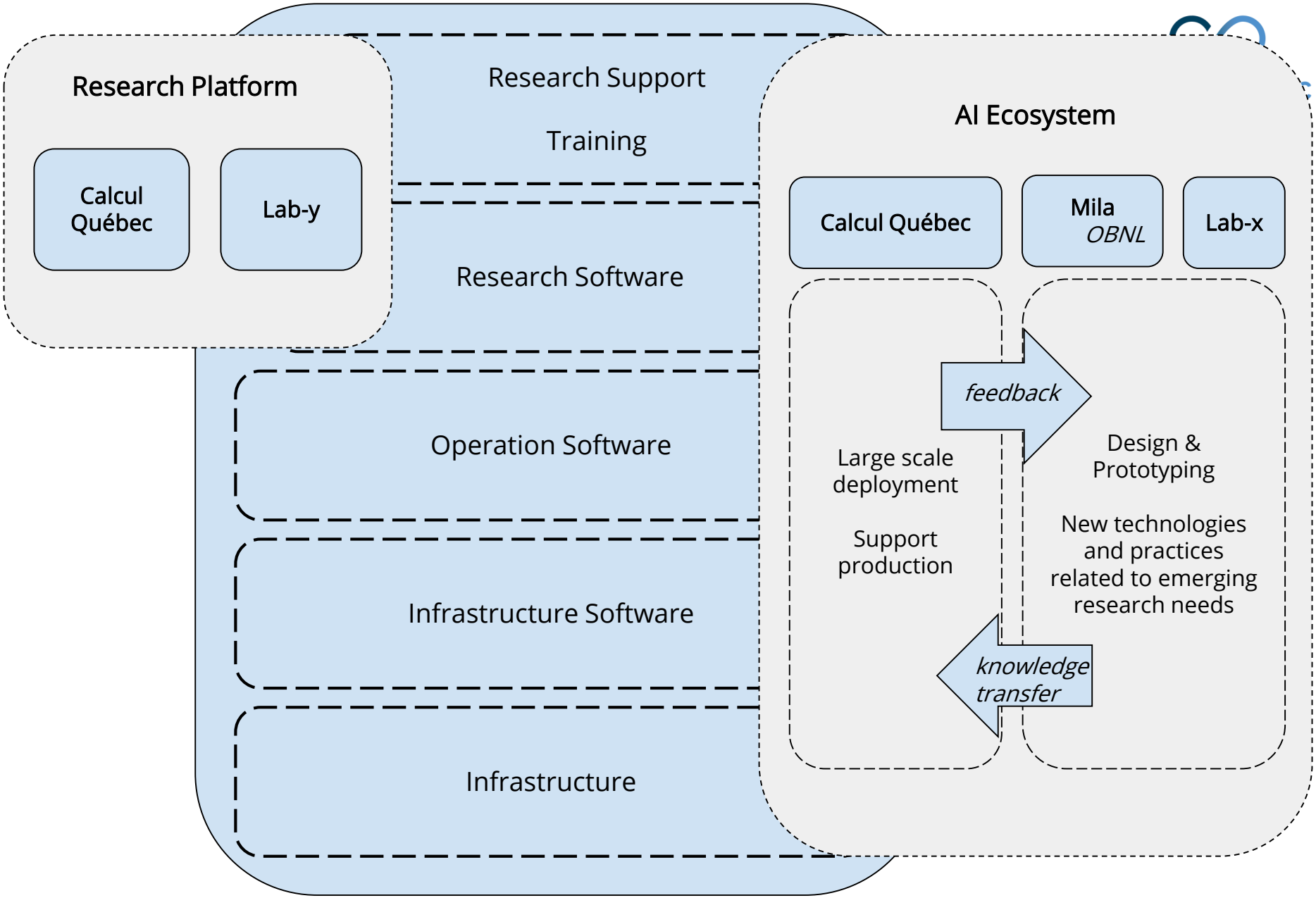
- ⇨ Data centers hosting supercomputers at the leading edge of technology
- ⇨ Teams of highly qualified advanced research computing specialists
- ⇨ The excellence and expertise of Québec researchers in all spheres of knowledge



Calcul Québec



Calcul Québec


Research Platform
Calcul Québec
Lab-y
Research Support
Training
Research Software
Operation Software
Infrastructure Software
Infrastructure
AI Ecosystem
Calcul Québec
Mila
OBNL
Lab-x
feedback
Large scale deployment
Support production
Design & Prototyping
New technologies and practices related to emerging research needs
knowledge transfer

The Team

Calcul Québec

Staff count

- ~40 people
- ⅓ Systems' Support (Sysadmins)
- ⅔ User and Application Support (Analysts) + Admin

Staff is collaborating from different locations

- Montréal, Sherbrooke, Québec, Texas (!)

Expertise → HQP

- Sysadmins: Lots of experience with big systems in Canada, many ranked on the TOP500 list, some were #1 in Canada.
- Analysts: physics, bioinformatics/genomics, biology, mathematics, engineering, etc.
- It takes time to build up such a team, and little time to lose it. My experience: 5-7 years for building a performing team of 11.

Data centers

Common data centers

- Located at ETS (École de Technologie Supérieure), Montréal
- ETS-1 : 1 MW computer room
- ETS-2 : 5 MW computer room (being built)

CQ has other data centers capable of hosting systems as of today

- Montréal, Québec, Sherbrooke
- Summing ~6 MW

Future plans

- 15+ MW data center (site not selected yet)
- Make use of Quebec's low priced hydropower
  - As the system becomes bigger, power prevails as the prime operating cost
- Economies of scale on the operating side (e.g. staff)

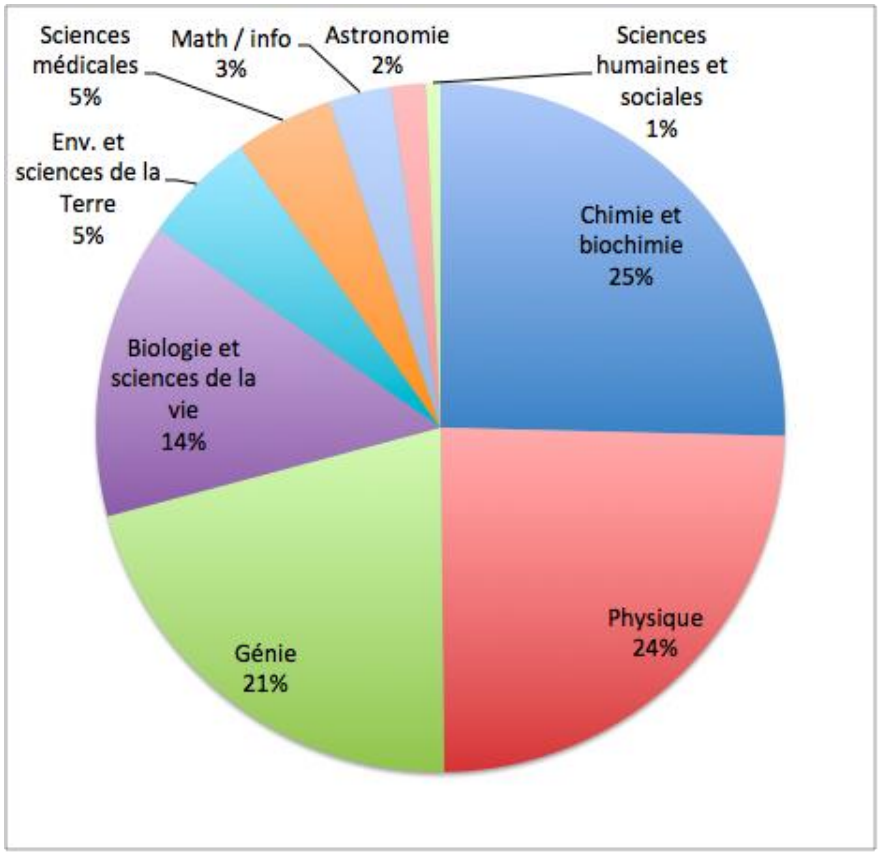
The Béluga system

Hardware is currently being installed at ETS

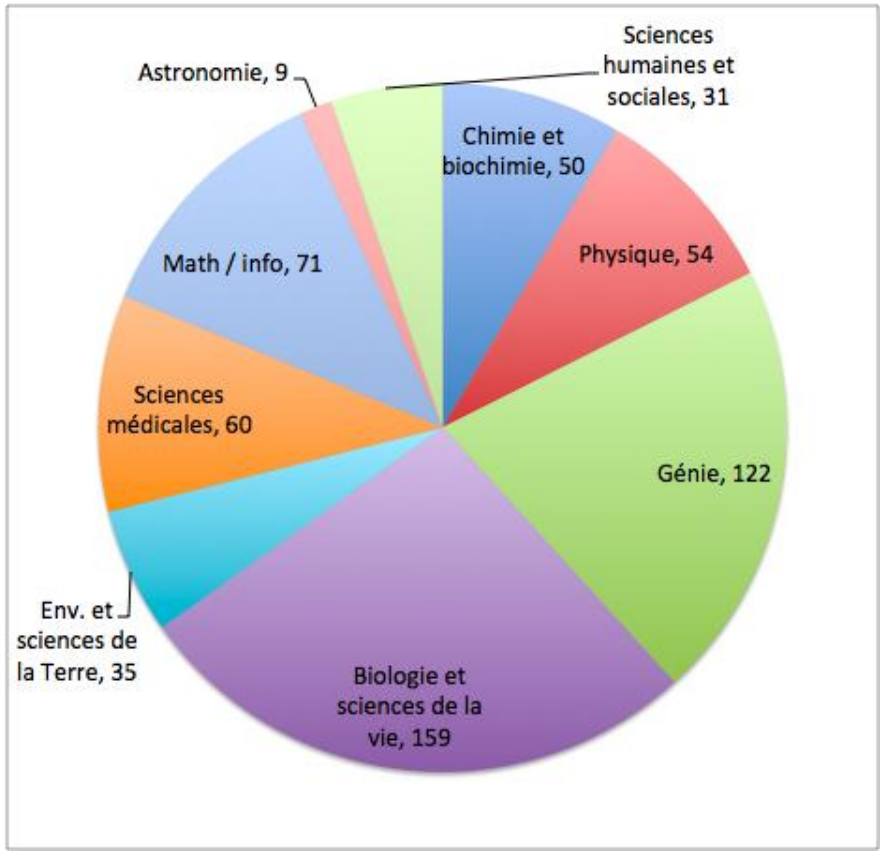
- Compute power
  - 872 compute nodes
    - 700 CPU based → 28 000 Intel Skylake 2.4 GHz cores
    - 172 GPU based → 688 NVIDIA Volta 16 GB RAM GPUs
    - 168.8 TB RAM
- Storage
  - Disk storage: 13 PB, more to come next year
  - Deep Storage: Disks/Tapes 50 PB, more to come next year
- Networking
  - 25 GBit local area network
  - 100 Gbit external connectivity
- Other stuff: Cloud partition, Experimental systems, etc.
- 25 M$ total capital budget

CPU utilisation
Per research field

Calcul Québec



CPU utilization

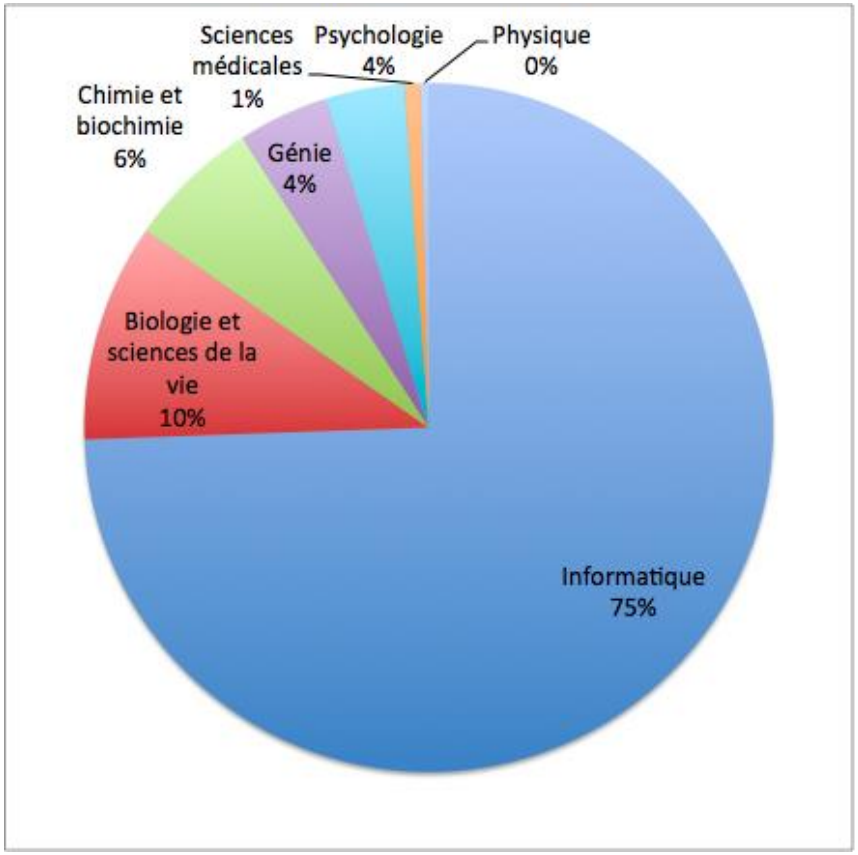


Research groups

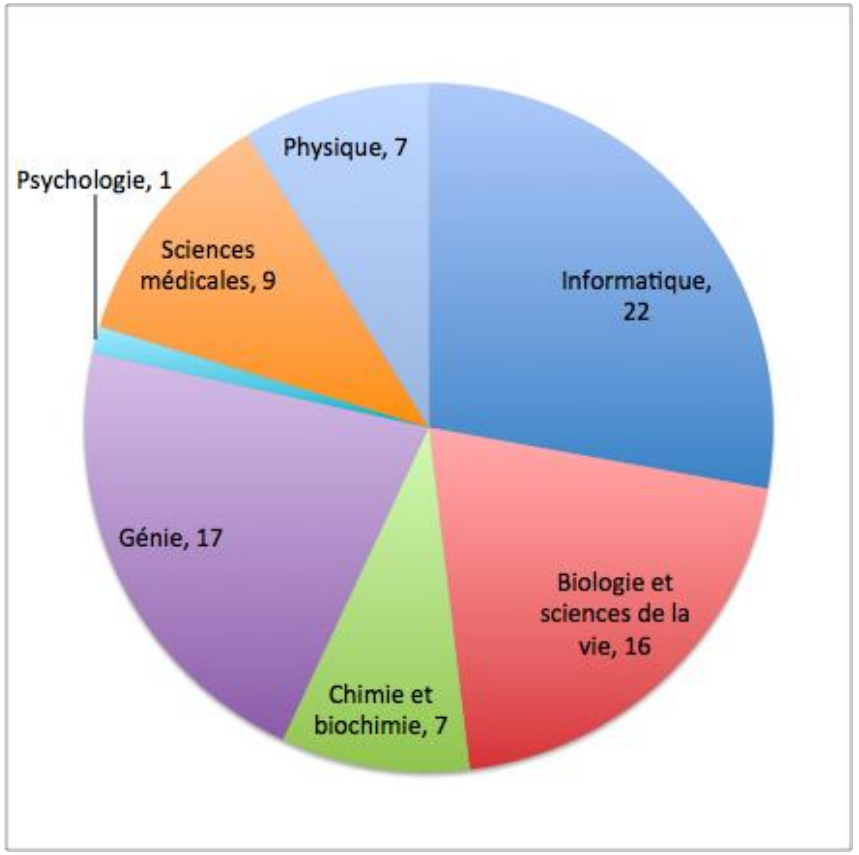
Total of 43 292 core-years from April 1st, 2016 - March 31st, 2017

GPU utilisation
Per research field

Calcul Québec



GPU utilization



Research groups

Total of 92 GPU-years from April 1st, 2016 - March 31st, 2017

Take-home message

HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 ,
HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 , HPC + AI = 🤍 .

:)

Convergence of High Performance Computing & Artificial Intelligence

Chris Loken
Chief Technology Officer
Compute Ontario

Bridges Supercomputer Used to Build AI Model for Beating Humans at Poker

Michael Feldman | January 12, 2017 23:54 CET

Supercomputing speeds up deep learning training

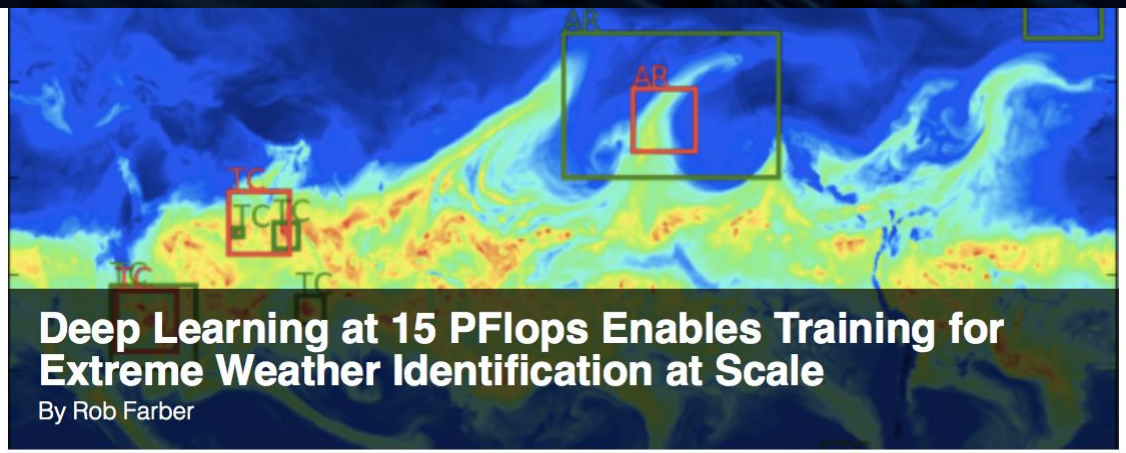
New algorithm enables researchers to efficiently use Stampede2 supercomputer to train ImageNet in 11 minutes, faster than ever before

Deep Learning Hits 15 Petaflops on Cori Supercomputer

Michael Feldman | August 30, 2017 14:34 CEST

AI Uses Titan Supercomputer to Create Deep Neural Nets in Less Than a Day

By **Peter Rejcek** - Jan 03, 2018 14,330



Deep Learning at 15 PFlops Enables Training for Extreme Weather Identification at Scale

By Rob Farber

Compute • Calcul Ontario

Mar 2018

Some Recent Canadian Research Papers….

Article

Machine Learning Techniques for Modelling Short Term Land-Use Change

Enabling Deep Learning of Emotion With First-Person Seed Expressions

Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks

Automated analysis of high−content microscopy data with deep learning

Genome-wide prediction of cis-regulatory regions using supervised deep learning methods

Lunar Crater Identification via Deep Learning

Compute • Calcul Ontario

Editors' Suggestion

Deep learning and the Schrödinger equation

Ontario Snapshot

- Strong HPC “infrastructure” and history: technical staff (HQP), researchers, datacenters and systems at multiple sites
- Hosts 3 of 6 national HPC systems including only health/PHI node
- 700 (and growing) Ontario-based PIs and 2,500 users account for (35-42)% of all national system usage
- Major emphasis on training of graduate students, PDFs and PIs to fully exploit compute power and capabilities
- Vector Institute concentrates AI expertise

Compute • Calcul Ontario

Compute Ontario Perspective

- Compute Ontario commissioned two reports in 2017-2018 which highlight strengths and gaps in provincial HQP and computing-related resources. Recommendations include:
  - Increased access to accelerators and other ARC resources designed to support emerging applications involving Big Data and artificial intelligence
  - Consider a new initiative dedicated to AI and high performance data analysis
  - Broaden industry access to HQP and systems
  - Predictable and sustained funding

Compute • Calcul Ontario

Why HPC Matters

- "computational science now constitutes the 'third pillar' of scientific inquiry, enabling researchers to build and test models of complex phenomena." 2005 U.S. Presidential Information Technology Advisory Committee report
- Complexity, size and scope of problems we can tackle computationally (the boundary of computable “universe”) is defined by size and capability of HPC systems
- “to out-compete we must out-compute”. US Council on Competitiveness

Compute • Calcul Ontario

Why HPC + AI?

- AI *requires* large datasets for accuracy as well as compute power for training and inference
- HPC users *already* incorporating AI techniques to improve and speed-up research as well as tackle harder questions (genomics, weather forecasting, drug discovery etc)
- HPC community continually *innovates* in terms of both hardware and software in order to maximize research capabilities
- It's already happening….

Compute • Calcul Ontario

HPC + AI

“This is why around 2008 my group at Stanford started advocating shifting deep learning to GPUs (this was really controversial at that time; but now everyone does it); and **I'm now advocating shifting to HPC (High Performance Computing/Supercomputing) tactics for scaling up deep learning. Machine learning should embrace HPC.** These methods will make researchers more efficient and help accelerate the progress of our whole field.”

Andrew Ng, 2016

Co-founder of Google Brain Project and Coursera

Former VP and Chief Scientist at Baidu

Former Director of AI Lab at Stanford

Compute • Calcul Ontario

Adoption of AI in Science – an example

A volumetric deep Convolutional Neural Network for simulation of dark matter halo catalogues

ACKNOWLEDGMENTS

We thank Erik Spence for a great course on "Advanced Neural Networks" where we initially came up with the idea for this project and further developed our skills. We also thank

NVIDIA's Tesla GPU powers Tsubame 2.0 to green supercomputer supremacy

Nov 2011

Baidu built a supercomputer for deep learning

Derrick Harris Jan 14, 2015 - 5:25 PM CST

NVIDIA announces a supercomputer aimed at deep learning and AI

Devin Coldewey @techcrunch / 3 years ago

Browse Conferences > 2016 IEEE 22nd International ...

Using Supercomputer to Speed up Neural Network Training

3 Author(s) Yue Yu ; Jinrong Jiang ; Xuebin Chi View All Authors

Compute • Calcul Ontario

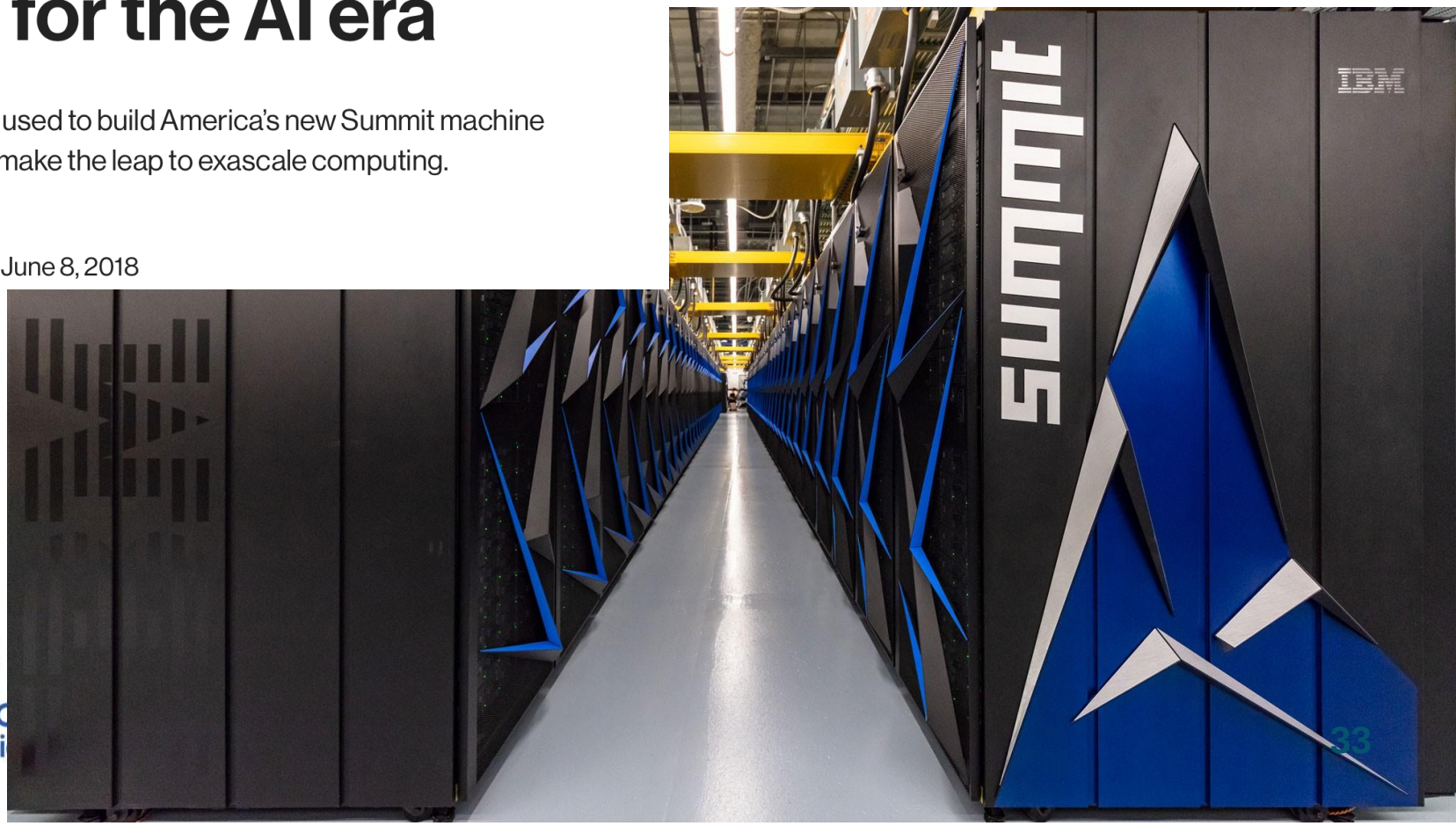
Machine Learning is The Killer App for High Performance Computing

Michael Feldman | April 1, 2017 05:08 CEST

The world's most powerful supercomputer is tailor made for the AI era

The technology used to build America's new Summit machine will also help us make the leap to exascale computing.

by Martin Giles June 8, 2018



Vector Institute and the Health AI Data Analysis Platform (HAIDAP)

November 2018

VECTOR INSTITUTE | INSTITUT VECTEUR

Pan Canadian AI Strategy

Collaborating with fellow AI institutions and Academic partners across Canada

amii

MILA

VECTOR INSTITUTE

UNIVERSITY OF TORONTO

UNIVERSITY OF WATERLOO

CIFAR
CANADIAN INSTITUTE FOR ADVANCED RESEARCH

ICRA
INSTITUT CANADIEN DE RECHERCHES AVANCÉES

Vector: Vision & Mission

Deep Learning & Machine Learning

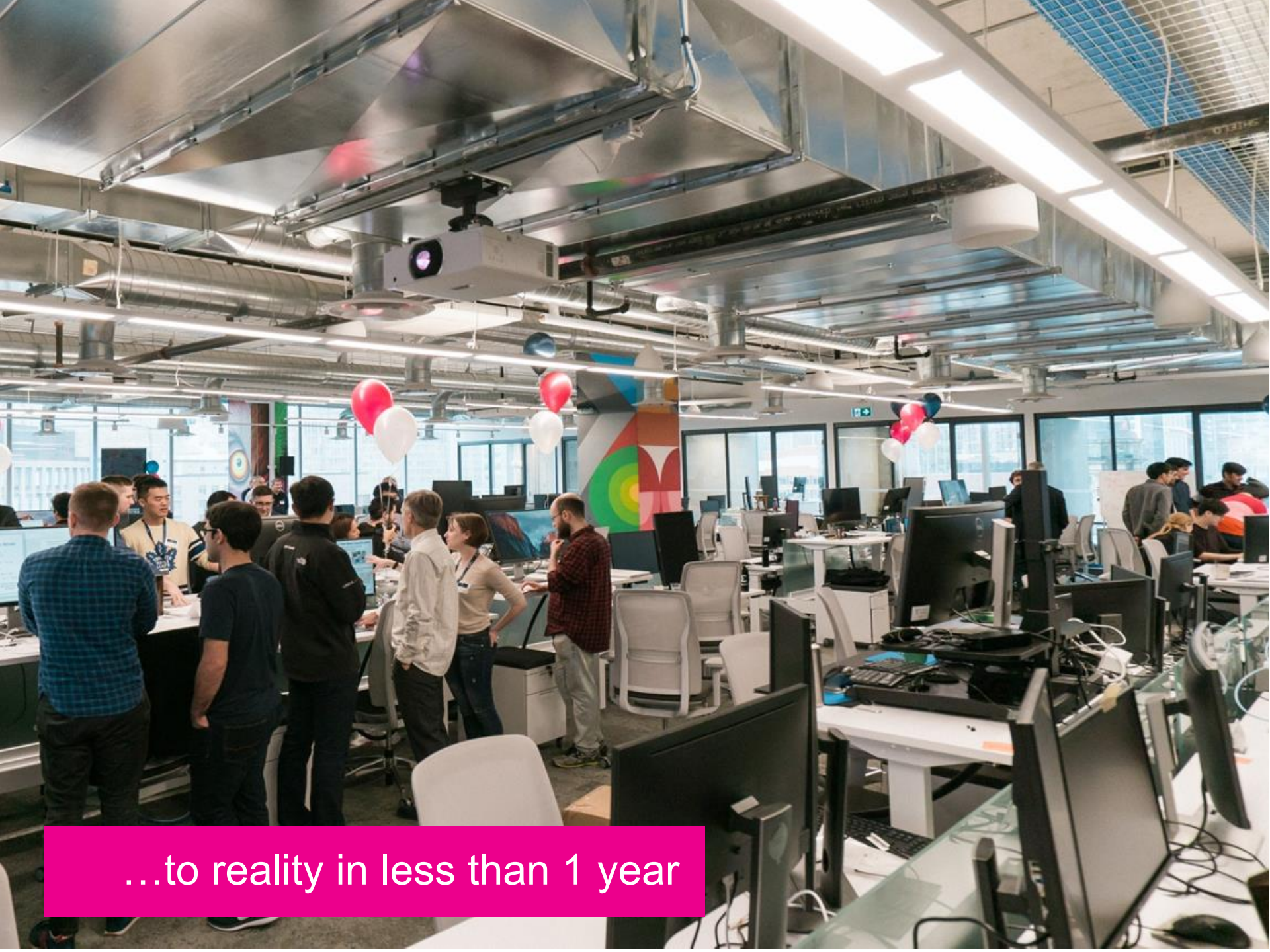
Attract + Retain Best Global Talent

Vector will drive excellence and leadership in Canada's knowledge, creation, and use of artificial intelligence to foster economic growth and improve the lives of Canadians.

Supporting Innovation Clusters & Start-ups

Partnering with Canadian industry and institutions


From concept and initial announcement …



…to reality in less than 1 year

Vector is Growing

A community of over **220 active researchers** with broad expertise in Artificial Intelligence.

- 21 Faculty members
- 63 Faculty Affiliates, 32 Postgraduate Affiliates
- 106 Students (Postdoctoral, PhD & Master Students)

What we do

World-Class Research

- Hiring top AI scientists and giving them flexibility
- Publishing & presenting in top conferences and journals
- Talks & seminars at Vector
- Collaborative office space
- Computing resources
- Growing community of Vector faculty, affiliates and students

Work with industry

Help companies become better AI users via:

- training for technical professionals
- training for executives
- connections between companies and researchers where interests align
- Match-making to help small AI companies find new customers and grow
- Job fairs and postings to help companies fill AI-related roles

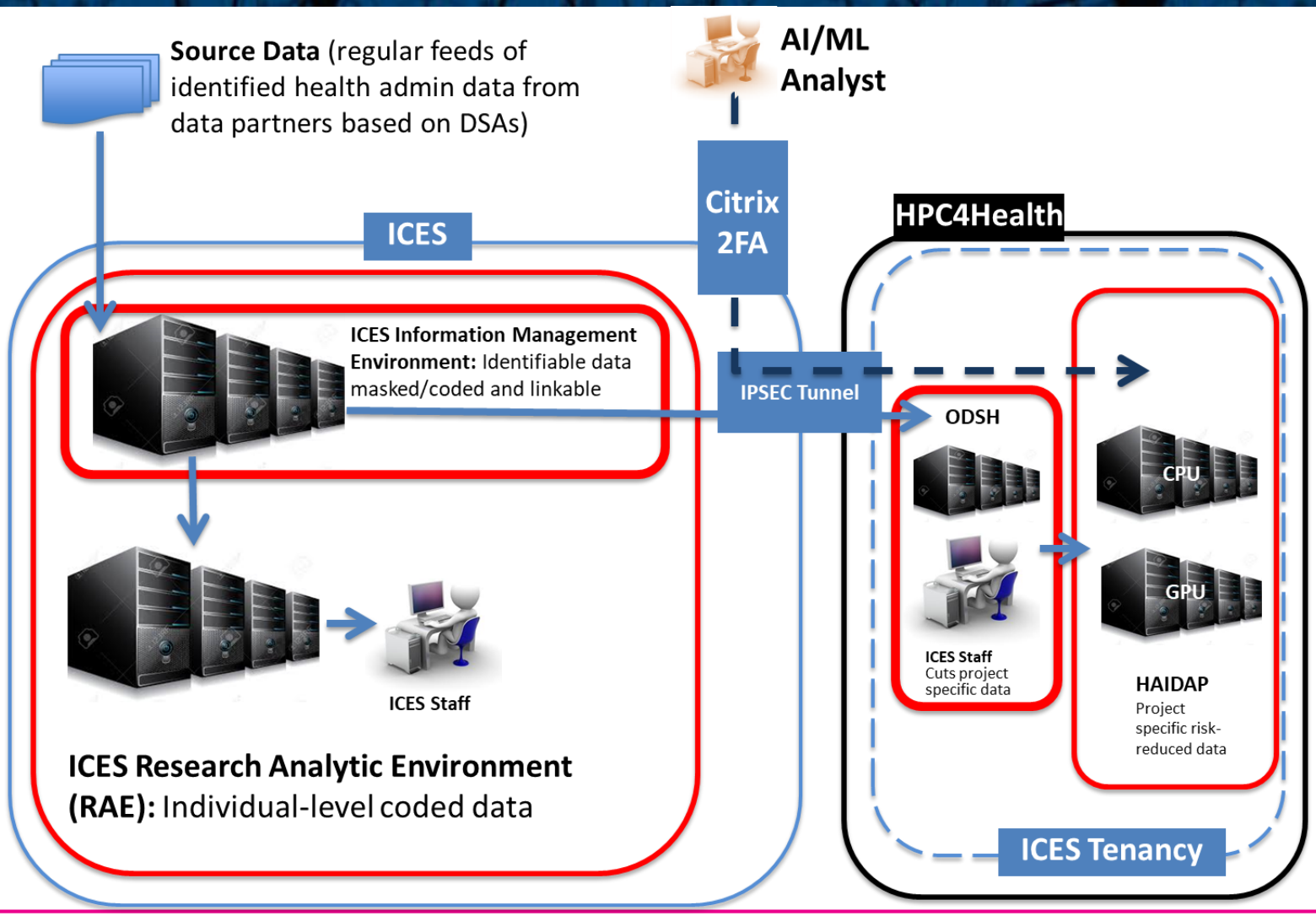
Training & Education

- Hiring more faculty to teach more students
- Working with universities to deliver more AI Master's programs
- Training industry professionals

Work with the health sector

- Creating opportunities for AI research and widespread application
- Focus on multi-site or province-wide data stemming from Ontario's large single-payer health system

Health AI Data Analysis Platform (HAIDAP)



HAIDAP: Province-wide Data + High Performance Computing → World Class AI

With generous funding from Compute Ontario and the Government of Ontario, we are moving population-wide longitudinal health data holdings into a secure environment with the compute power required for modern machine learning research

	ICES Research Analytic Environment	Planned HAIDAP Infrastructure by Spring 2019
Annual Analytic Projects	300-500	TBD
CPU Cores	80	400+
GPU Clusters	1 (<100TFLOPS)	13 (1.26 PFLOPS)
Storage	200 TB	2+ PB (est)

HAIDAP as Team Science

- ICES, HPC4Health (at the Hospital for Sick Children), and Vector will all contribute their expertise to the design, build, and operations of the HAIDAP
- Within that partnership:

- HPC4Health is the primary lead for infrastructure build and maintenance
- ICES is the primary lead for data governance, data, and data access and clinical/subject matter expertise
- Vector brings machine learning expertise, is a key user of the HAIDAP and has a lead role in defining essential functions/specifications and priority datasets to bring into the HAIDAP for machine learning-enabled research

Learn more

https://vectorinstitute.ai

@vectorinst

Vector Institute

VECTOR INSTITUTE | INSTITUT VECTEUR

Panelist Questions

1. Current efforts to create a national Digital Research Infrastructure Strategy have focused primarily on high performance computing, big data and network capacity. Complementary activities such as scaling Artificial Intelligence or promoting academic-industry collaborations do not appear to be addressed in the proposed approach. Based on the key messages from your respective presentations, what advice do you have regarding the inclusion of these components as the national strategy begins to take form?

2. High performance computing and Artificial Intelligence are complementary, but take different methodological approaches and research foci. What advantages or disadvantages do you see from converging the two?

3. HAIDAP was described as an Ontario initiative that brought together experts in research data, high performance computing, AI and networking to create a novel platform to support research.

    a) What lessons learned would you share with others around the benefits of this approach (i.e. is it scalable)?
    b) Given the key messages discussed today, what changes do current funding models need to adopt to support Canada's research community so it can continue to thrive?

Compute • Calcul Ontario