

Trends in Scientific Computing Training by an HPC Center

November 12, 2018

SciNet HPC Consortium

Ramses van Zon, Marcelo Ponce, Erik Spence, and Daniel Gruner

*Increasing number of fields and students need to use HPC.
These people need training in scientific computing.
They do not get this training in their studies.
HPC Centers have the expertise to deliver this training.*

So... what is SciNet?

- An academic HPC center, located at the University of Toronto.
- Has been hosting the largest academic supercomputers in Canada (*GPC, Niagara*).
- Early users were predominantly from physical sciences and engineering.
- User's research fields quickly diversified, in part due to the increase of data availability in many disciplines: bioinformatics, health sciences, social science, digital humanities, commerce, etc.
- ~200 active users, 14 support staff



SciNet's Scientific Computing Training

- From the very start, SciNet offered training events on how to use HPC systems and how to write parallel code.
- To reflect the changes in computing fields and computational knowledge of the users, the training program was adapted to contain a broad range of workshops and courses.
- Tremendous growth in attendance over the years.

Training Repertoire

Initially (~2011)

Programming Clusters with MPI

Programming Shared Memory with OpenMP

Programming GPUs with CUDA

Modern Fortran, Co-array Fortran

Scientific C++

Parallel I/O

Training Repertoire

Initially (~2011)

Programming Clusters with MPI
Programming Shared Memory with OpenMP
Programming GPUs with CUDA

Modern Fortran, Co-array Fortran
Scientific C++
Parallel I/O

In addition, now also:

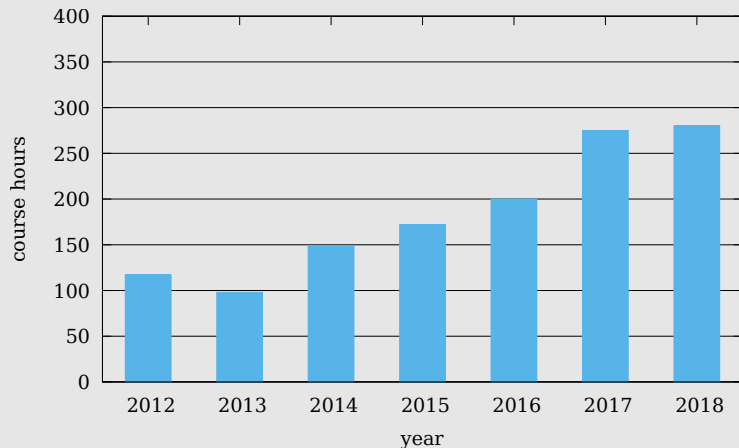
Quantitative Applications for Data Analysis
Computational BioStatistics with R
Scientific Computing for Physicists
Introduction to Programming with Python
Neural Networks
Apache Spark
Machine Learning Workshop
Scalable Data Analysis Workshop
Relational Database Basics

Storage and I/O in Large Scientific Projects
Python for High Performance Computing
Parallel R
Scientific Visualization
Research Computing with Python
Advanced Parallel Scientific Computing
Debugging, Optimization, Best Practices
Linux Shell
Advanced Shell Programming

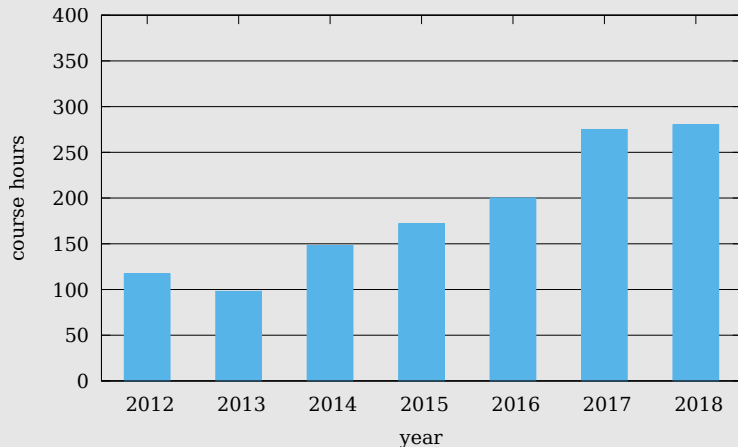
Forms of training

- **Workshops:** Our training initially was mostly in the form of one-day or multi-day workshops.
- **Lecture-style courses:** Most of the training is now spread out over the year in courses.
- **Online Assignments, forums, slide, recording of the lectures, etc.:**
Even the lecture-style courses have a large on-line component.
- **Summer school:** Partly to serve non-local users, we hold an annual one-week intensive workshop with three streams (high-performance computing, data science, and biomedical computing). We collaborate with other HPC centers in Ontario for this.
We're also involved in the IHPCCS.
- All this training is coordinated through SciNet's education site, used by ~1800 users so far.
<https://courses.scinet.utoronto.ca>

Growth in hours taught by SciNet

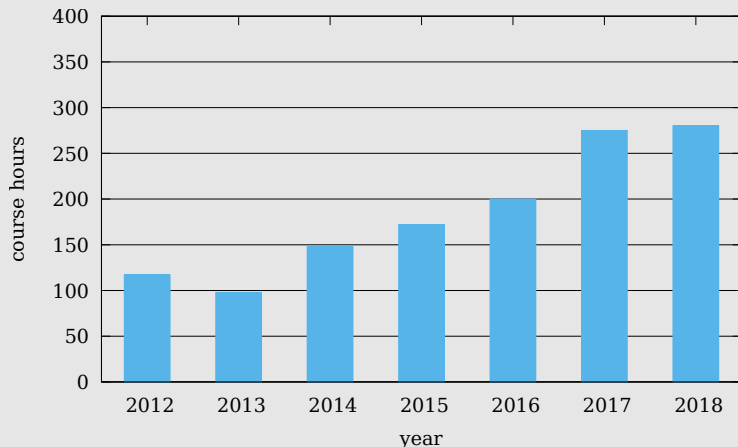


Growth in hours taught by SciNet



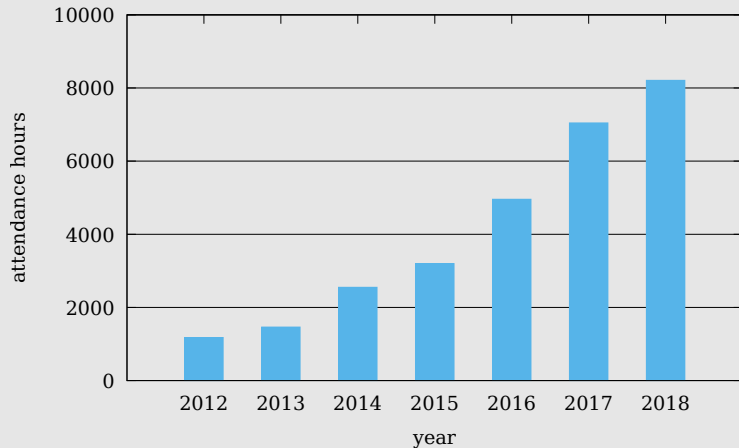
• Σ duration

Growth in hours taught by SciNet

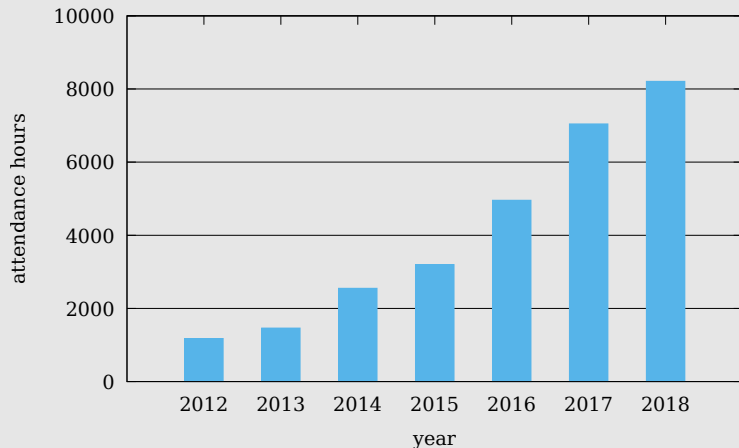


- Σ duration
- Growing over the years.
- Hitting a plateau due to limited human resources.

Growth in attendance

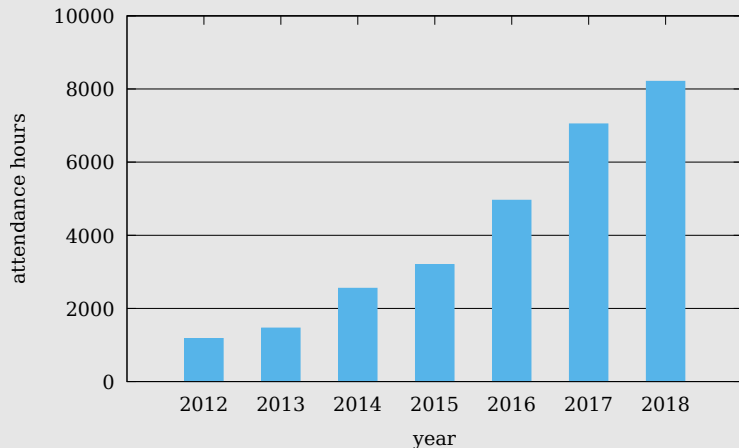


Growth in attendance



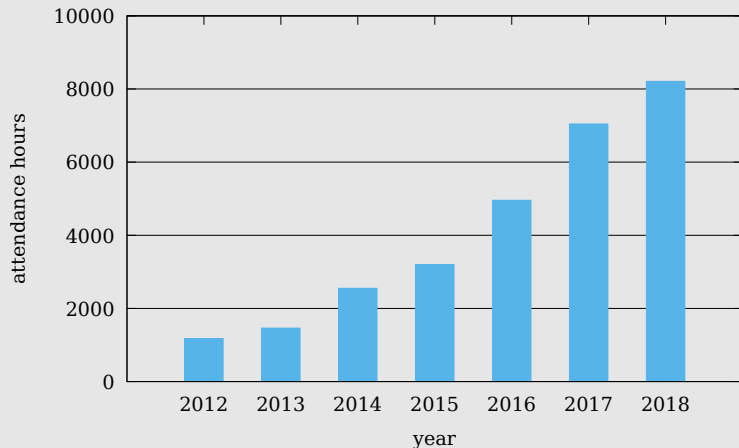
- $\sum \text{duration} \times \# \text{students}$

Growth in attendance



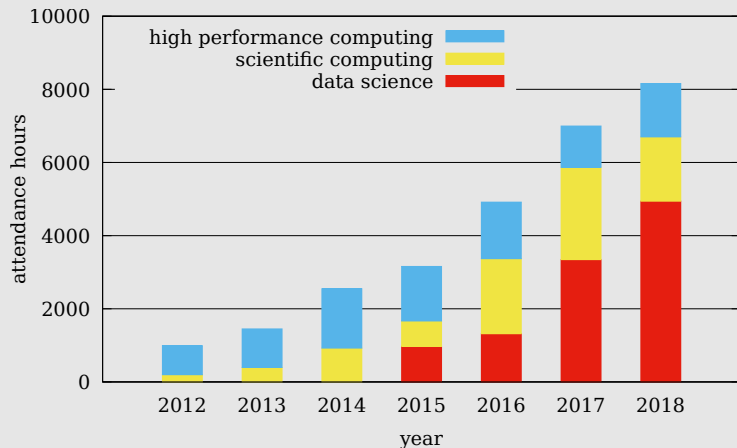
- $\sum \text{duration} \times \# \text{students}$
- For comparison, one term-long course is 24 attendance hours per student.

Growth in attendance

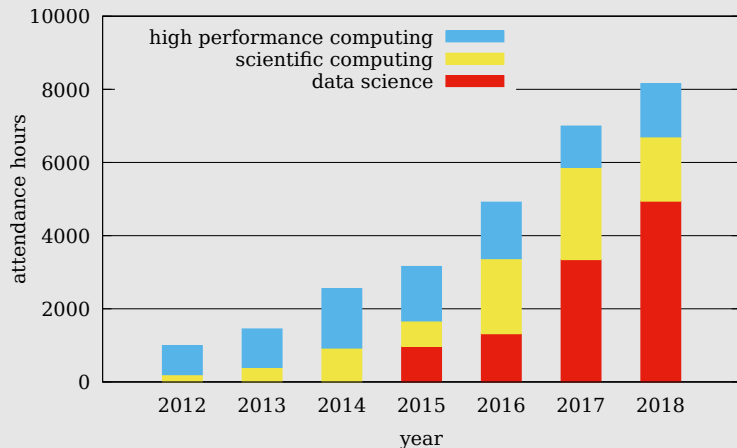


- $\sum \text{duration} \times \# \text{students}$
- For comparison, one term-long courses is 24 attendance hours per student.
- Continuing rapid growth in demand.
(2018 not yet over!)

Evolution of supercomputing training

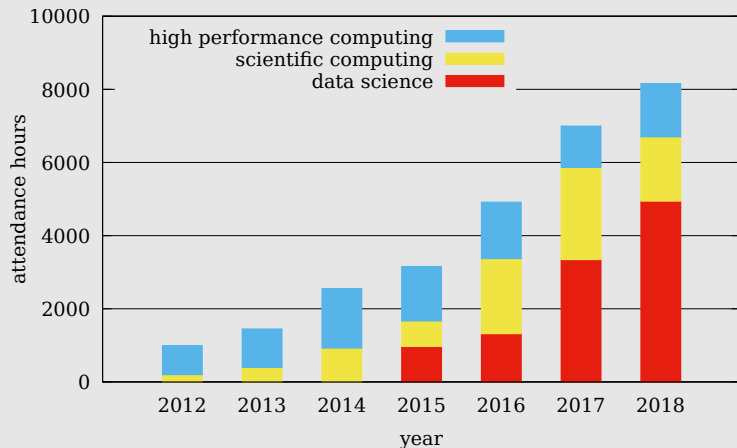


Evolution of supercomputing training



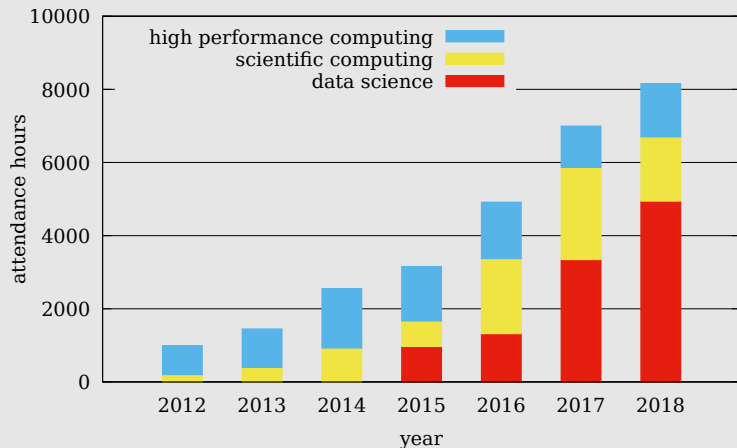
- Enormous uptake of data science.

Evolution of supercomputing training



- Enormous uptake of data science.
- Data science includes machine learning, AI.

Evolution of supercomputing training



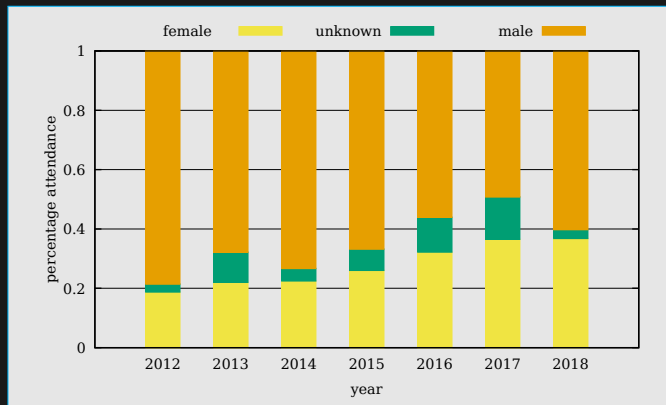
- Enormous uptake of data science.
- Data science includes machine learning, AI.
- Traditional HPC steady, but not growing.

Shifts in participation by discipline

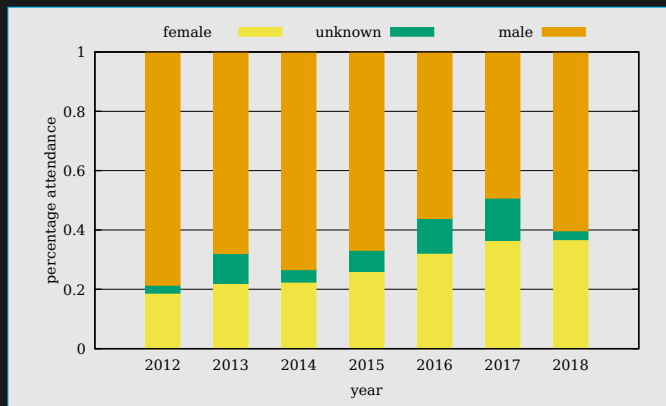
	2013	2014	2015	2016	2017	2018
engineering	34%	14%	15%	20%	12%	17%
physics	29%	36%	30%	24%	18%	12%
earth science	10%	8.1%	8.8%	6.0%	3.4%	2.3%
chemistry	6.7%	16%	2.7%	4.4%	3.4%	4.4%
computer science	1.2%	1.4%	0.3%	1.6%	1.7%	0.8%
mathematics	0%	0%	0%	2.1%	0.8%	2.5%
medicine	10%	17%	20%	25%	35%	40%
biology	8.0%	4.8%	22%	16%	24%	18%
economics	0%	0.9%	0%	0.4%	0.4%	2.2%
social science	0%	0%	0%	0%	0.2%	0.2%
humanities	0%	0.1%	0%	0%	0%	0%

Relative percentage of the attendance in SciNet training by researchers from different fields.

Shifts in participation by gender



Shifts in participation by gender

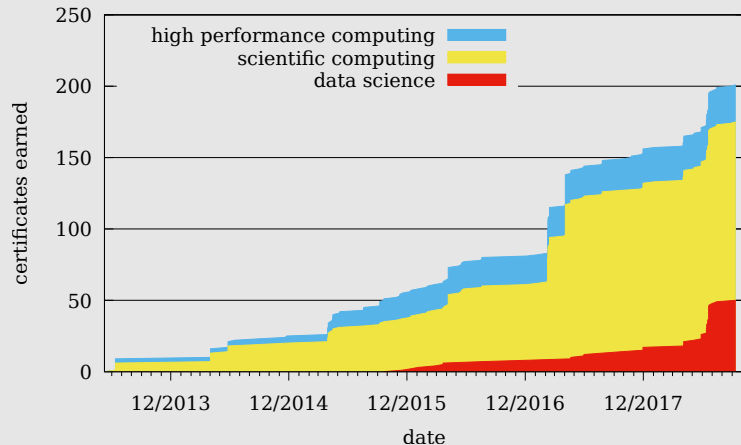


- Growth in female participation from 20% in 2012 to 40% in 2018.
- Quite possibly correlated with shift in participating disciplines.

Credit

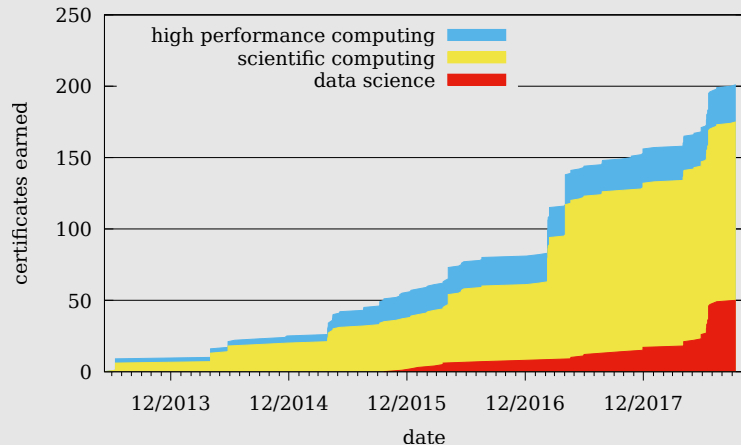
- Many participants of the training are grad students from the University of Toronto, yet, they originally could not get credit for participating in SciNet training on their transcript.
- In 2012, SciNet started a [certificate program](#) that awarded people that took 36 hours worth of training with a certificate.
- There are now three separate certificates for scientific computing, high performance computing, and data science.
- Increasingly, SciNet training courses are recognized and converted as [university graduate courses](#).
- The graduate courses are made possible by partnering with other departments: **Physics, Institute for Medical Science, Physical and Environmental Sciences.**

Certificates trends



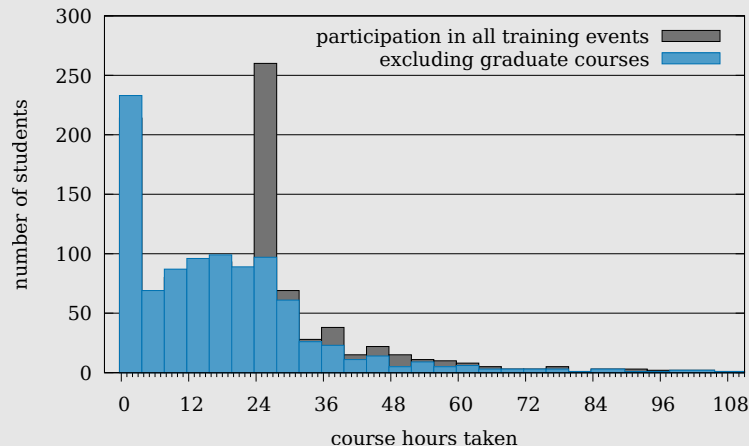
- Over 200 certificates earned.
- Again, strong uptake of data science.
- Evidence of students taking several courses.

Certificates trends



- Over 200 certificates earned.
- Again, strong uptake of data science.
- Evidence of students taking several courses.
- Other evidence of retention. . .

Retention



- Histogram of how many training hours individuals have taken.
- Some attended only one event, but many came back for more.
- This is true even if we take the graduate courses out.

Conclusions

- HPC centers can fill the gap in scientific computing, data science and HPC training.
- Among the results of the analysis of attendance data are:
 - ▶ a steady overall growth,
 - ▶ a sharp and steady increase in the demand for data science training,
 - ▶ a wider participation of ‘non-traditional’ computing disciplines,
 - ▶ and a larger female participation.
- Training initiatives can and have evolved into graduate courses.

Conclusions

- HPC centers can fill the gap in scientific computing, data science and HPC training.
- Among the results of the analysis of attendance data are:
 - ▶ a steady overall growth,
 - ▶ a sharp and steady increase in the demand for data science training,
 - ▶ a wider participation of 'non-traditional' computing disciplines,
 - ▶ and a larger female participation.
- Training initiatives can and have evolved into graduate courses.
- **Lessons learned:**
 - ▶ Listen to demand.
 - ▶ Collaborate with other departments.
 - ▶ Automate evaluation.
 - ▶ Gather participant's information like gender and field of research (voluntarily).