Training for the Next HPC Generation

David Henty
HPC Training and Support
EPCC, The University of Edinburgh

September 7, 2017

The EXDCI project has received funding from the European Unions Horizon 2020 research and innovation programmed under the grant agreement No 671558.

Overview

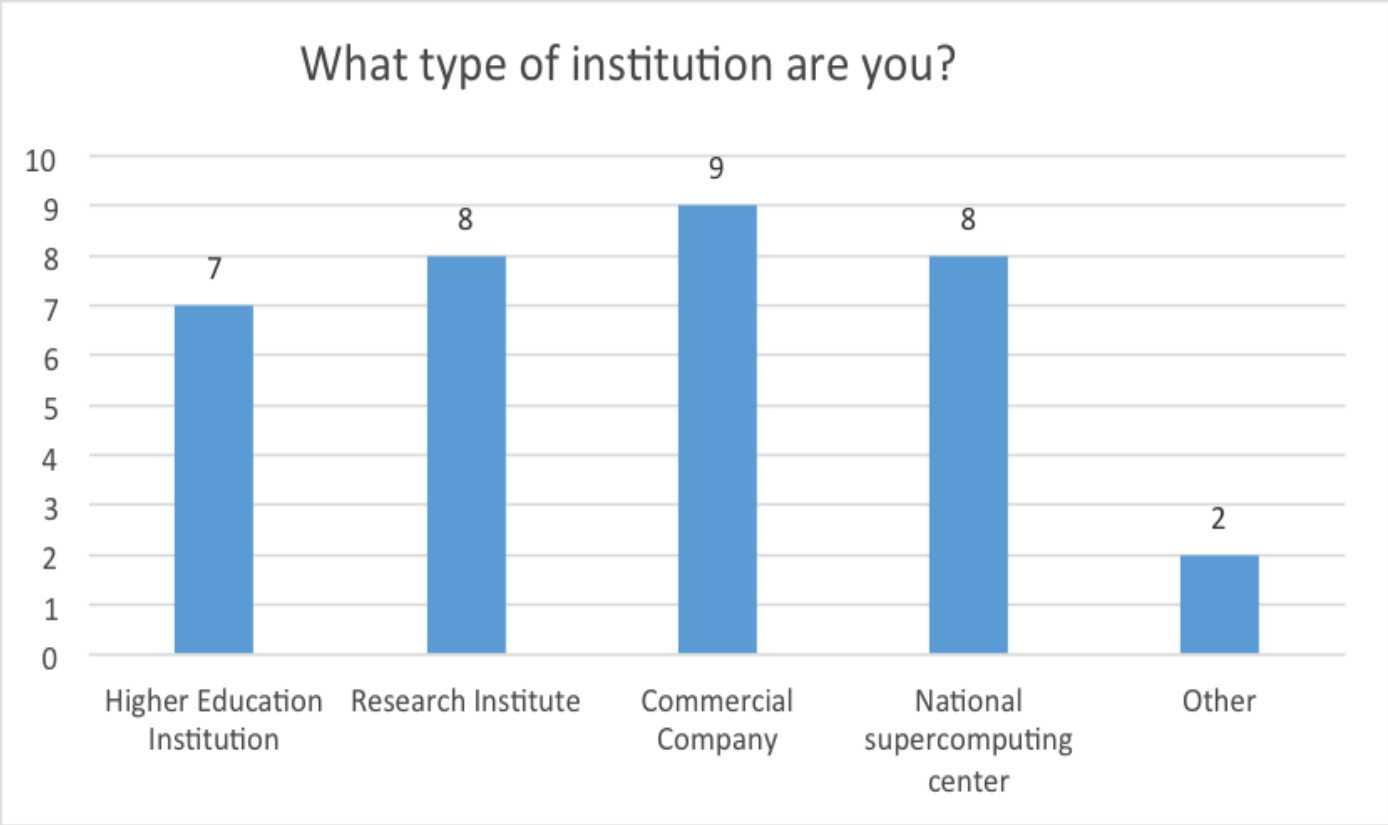
- EXDCI activities
  - HPC Training Roadmap
  - Results and recommendations
- Personal experiences
  - Online
    - MOOCs
    - University-accredited Masters courses
  - Campus-based
    - MSc in HPC
    - HPC Summer School

EXDCI Activities

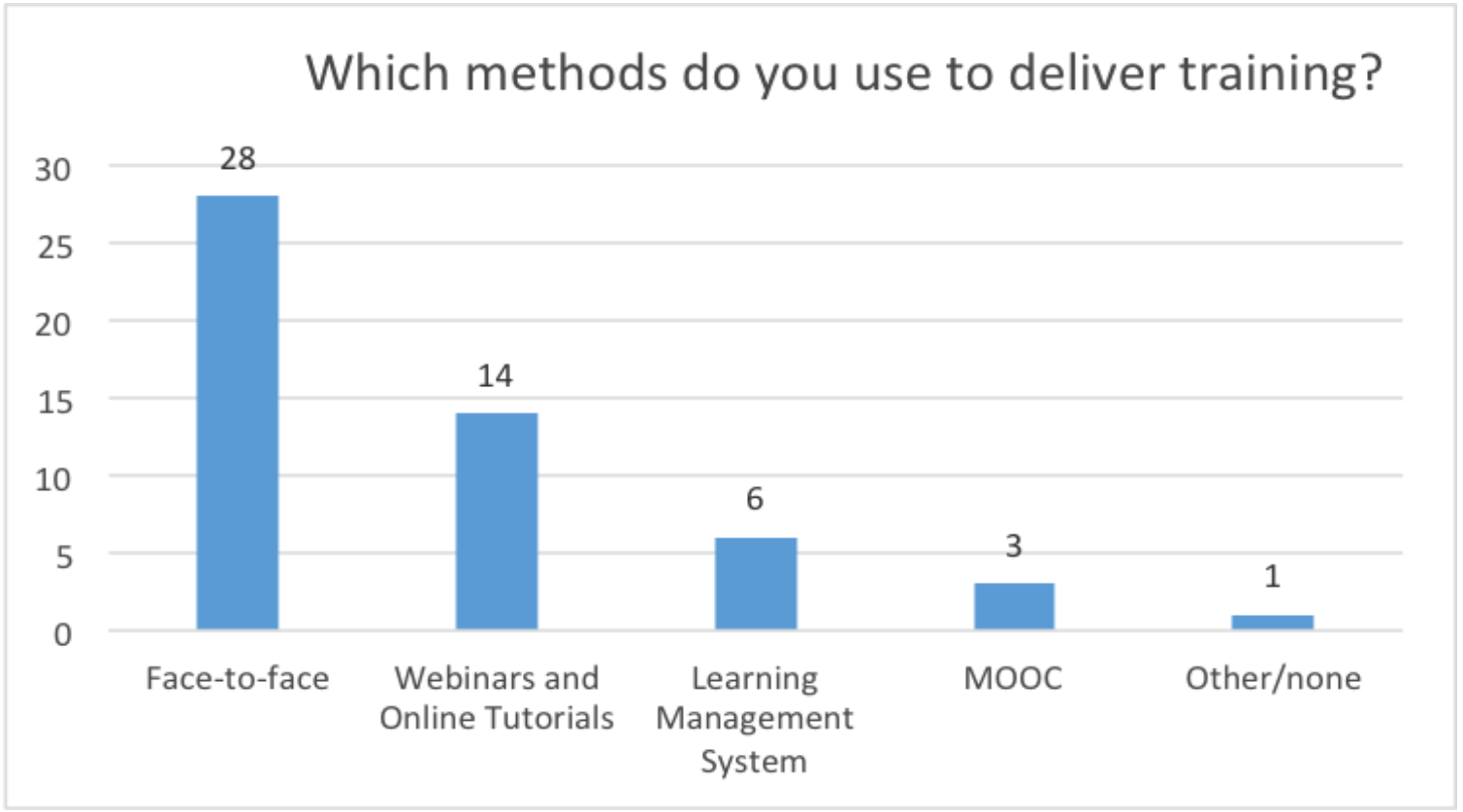
- Promoting:
  - HPC as a career choice
  - availability of HPC jobs
  - availability of HPC training
- Resources:
  - https://exdci.eu/jobs-and-training
    - https://exdci.eu/jobs-and-training/hpc-career-case-studies
      - https://exdci.eu/sites/default/files/public/files/d5.3_promoting_hpc_careers.pdf
    - https://exdci.eu/jobs-and-training/training-portal
      - https://exdci.eu/jobs-and-training/training-portal/training-roadmap
    - https://exdci.eu/jobs-and-training/job-portal

HPC Training Roadmap

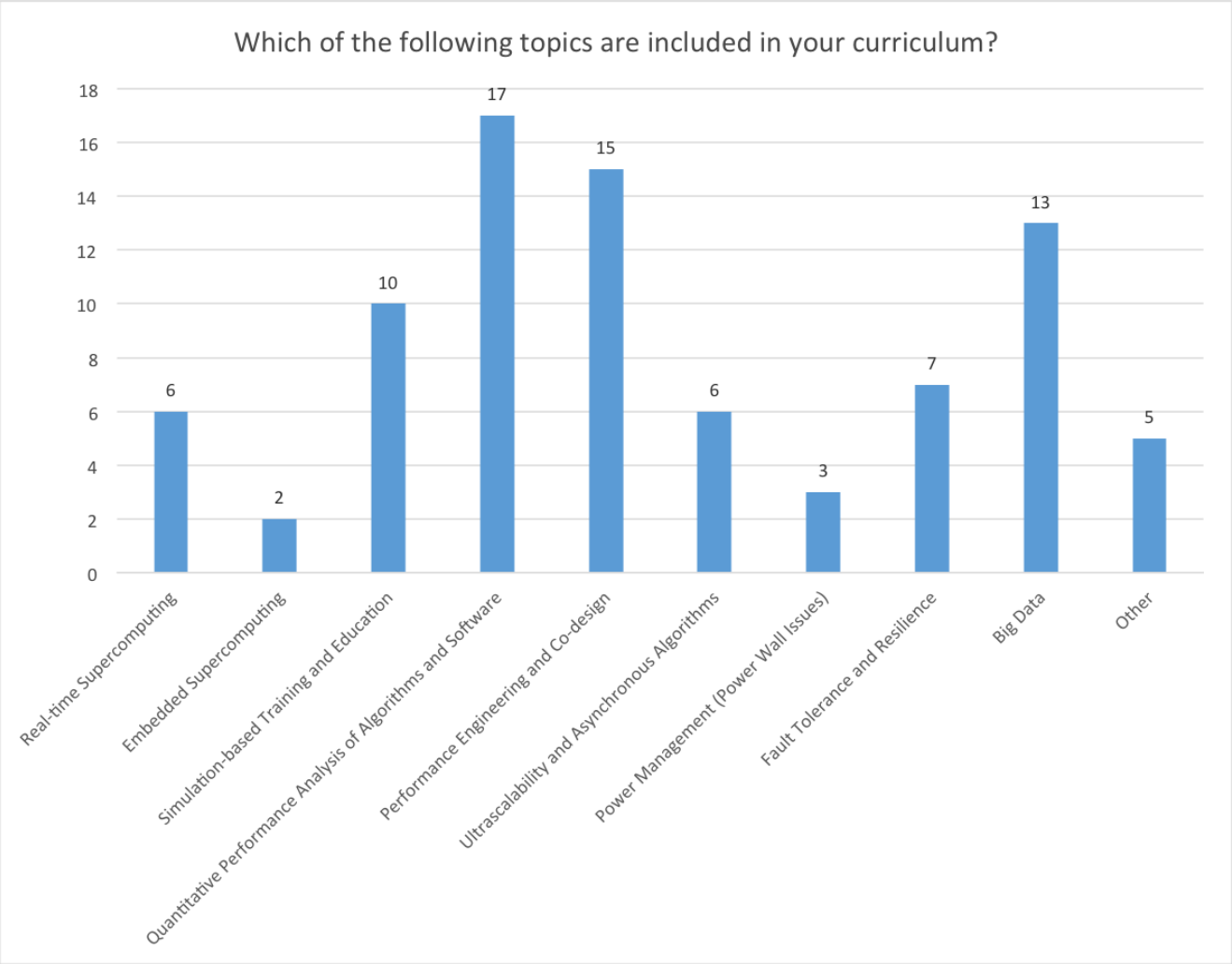
- Survey collected responses from 29 providers



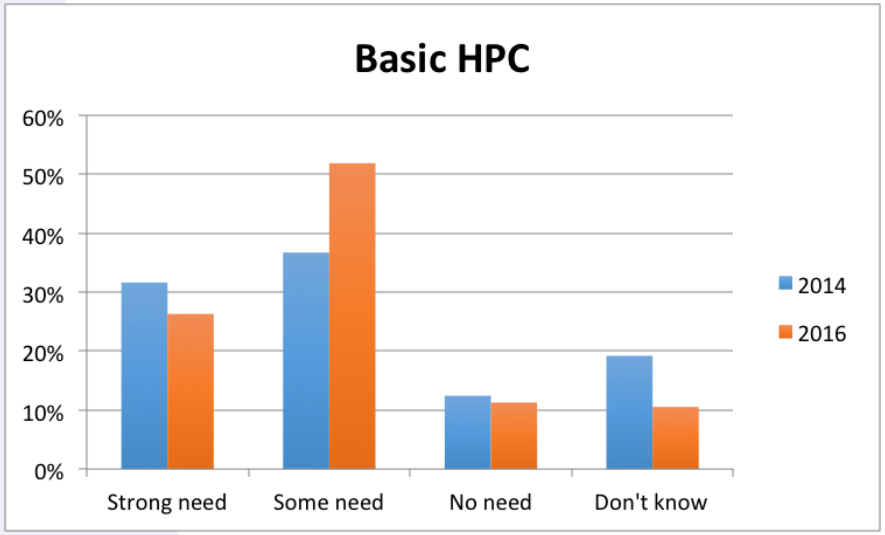
Delivery Mechanism

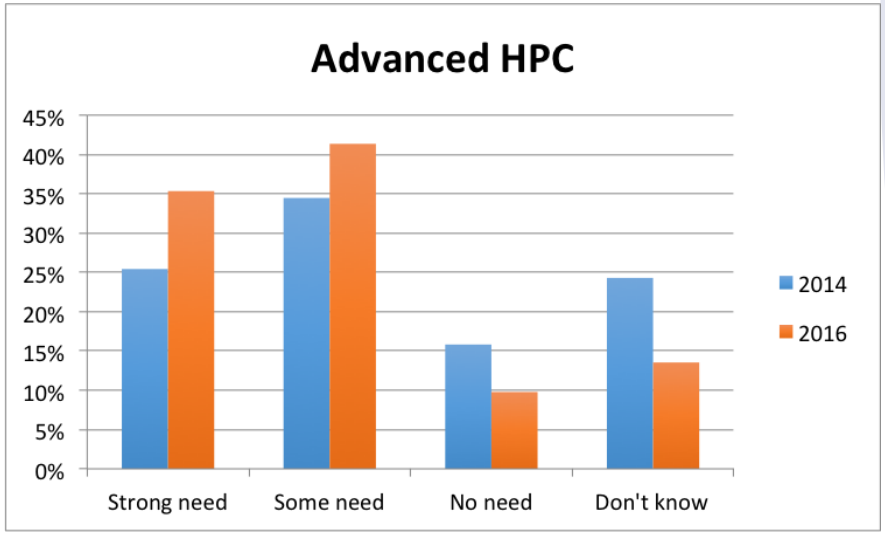


Subject Area

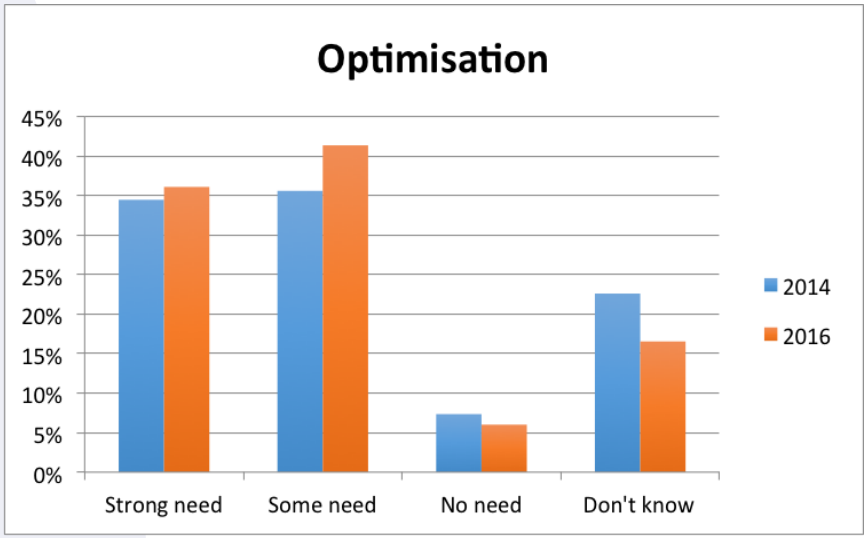


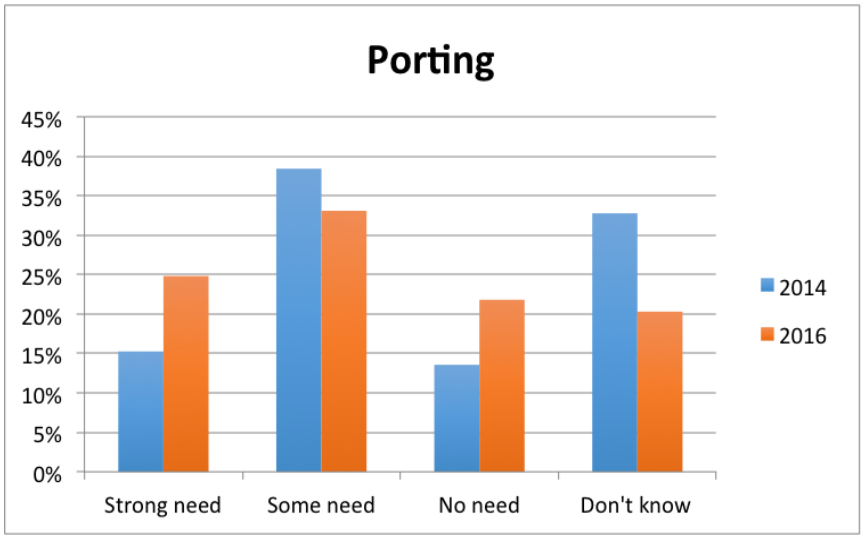
User Needs: PRACE Advance Training Centres





User Needs: PATCs





Recommendations (selected)

- Rising Demand
  - systematic recording and archiving of training
  - more MOOCs and Summer Schools
- Widening Audience
  - target communities and co-locate with other events
  - widen dissemination
  - increased engagement with industry and technology partners
- Integration into Curriculum
  - develop university links and develop one-hour HPC promo for UG/PG
  - online delivery to open up beyond host institute

Training and Education at EPCC

MSc in HPC / HPC with Data Science
Accredited / assessed
Fee paying ~12K €
On-campus
One year

ARCHER / PRACE training
Non-assessed
Free
On-campus
Several days

MOOCs
Non-assessed (?)
Free
Online
Several weeks

Postgraduate Professional Development
Accredited / assessed
Fee paying ~2K €
Online
Several months

Practical Introduction to HPC
Practical Introduction to Data Science
January – June each year

MOOCs

- PRACE funded development of two 5-week MOOCs
  - **Supercomputing**: EPCC and SURFsara
  - **Managing Big Data with R and Hadoop**: University of Ljubljana
- Both run for first time in early 2017
  - Supercomputing 2nd run now ongoing (since 28 August)
  - Big Data runs again on 9 October
- Developed under FutureLearn
  - https://www.futurelearn.com/courses/supercomputing
  - https://www.futurelearn.com/courses/big-data-r-hadoop

Curriculum

- Week 1: Supercomputers
  - motivation, history and performance
- Week 2: Parallel Computers
  - parallelism, distributed vs shared memory, architecture
- Week 3: Parallel Computing
  - message passing vs shared variables, traffic model
- Week 4: Computer Simulation
  - concepts, illustrated by weather forecasting
- Week 5: Case Studies
  - real simulations, benefits, future of supercomputing

General outline

Using supercomputers, we can now conduct virtual experiments that are impossible in the real world – from looking deep inside individual atoms, to studying the future climate of the earth and following the evolution of the entire universe from the big bang.

Discover how supercomputers are powering scientific breakthroughs

This free online course will introduce you to what supercomputers are, how they are used and how we can exploit their full computational potential to make scientific breakthroughs.

Over five weeks, we'll look at:

- **supercomputers**: introducing supercomputing terminology and some of the largest machines in the world.
- **parallel computers**: how they are built from hundreds of thousands of CPUs, each similar to those in a desktop PC.
- **parallel computing**: using parallel processing to harness the power of all of those CPUs for a single calculation.
- **computer simulation**: how we can perform virtual experiments to make real-life predictions.
- **case studies**: how supercomputing is making scientific breakthroughs that were never possible before.

EDUCATORS

David Henty

IN ASSOCIATION WITH

THE UNIVERSITY of EDINBURGH

SURF SARA

Typical activities and steps in week 2

Connecting multiple computers

Here we explain how large supercomputers containing many thousands of CPU-cores are constructed from commodity building blocks.

2.6 **DISTRIBUTED MEMORY ARCHITECTURE** ARTICLE

2.7 **SIMPLE PARALLEL CALCULATION** DISCUSSION

2.8 **CASE STUDY OF A REAL MACHINE** ARTICLE

2.9 **UNDERSTANDING PARALLEL COMPUTERS** QUIZ

Comparing the two approaches

The shared and distributed memory architectures are very different: each has its own pros and cons, with performance governed by different factors.

Support

Typical article around 750 words max

© iStock.com/BrianAJackson

Shared memory v.s. Distributed memory

We've seen how individual CPU-cores can be put together to from large parallel machines in two fundamentally different ways: the shared and distributed memory architectures.

In the shared-memory architecture all the CPU-cores can access the same memory, much like several workers in an office sharing the same whiteboard, and are all controlled by a single operating system. Modern processors are all multicore processors, with many CPU-cores manufactured together on the same physical silicon chip.

There are, however, limitations to the shared-memory approach due to all the CPU-cores competing for access to memory over a shared bus, much like the obvious issues in trying to cram too many workers into the same office. This can be alleviated to some extent by introducing memory caches or putting several processors together in a NUMA architecture, but there is no way we can reach the hundreds of thousands of CPU-cores we need for today's multi-petaflop supercomputers.

© iStock.com/Maxiphoto

Simple Parallel Calculation

Let's return to the income calculation example. This time we'll be a bit more ambitious and try and add up 8000 salaries rather than 800. This list of salaries fills 10 whiteboards (800 on each) all in separate offices.

If we have one worker per office, think about how you could get them all to cooperate to add up all the salaries. Consider two cases:

- only one "boss" worker needs to know the final result;

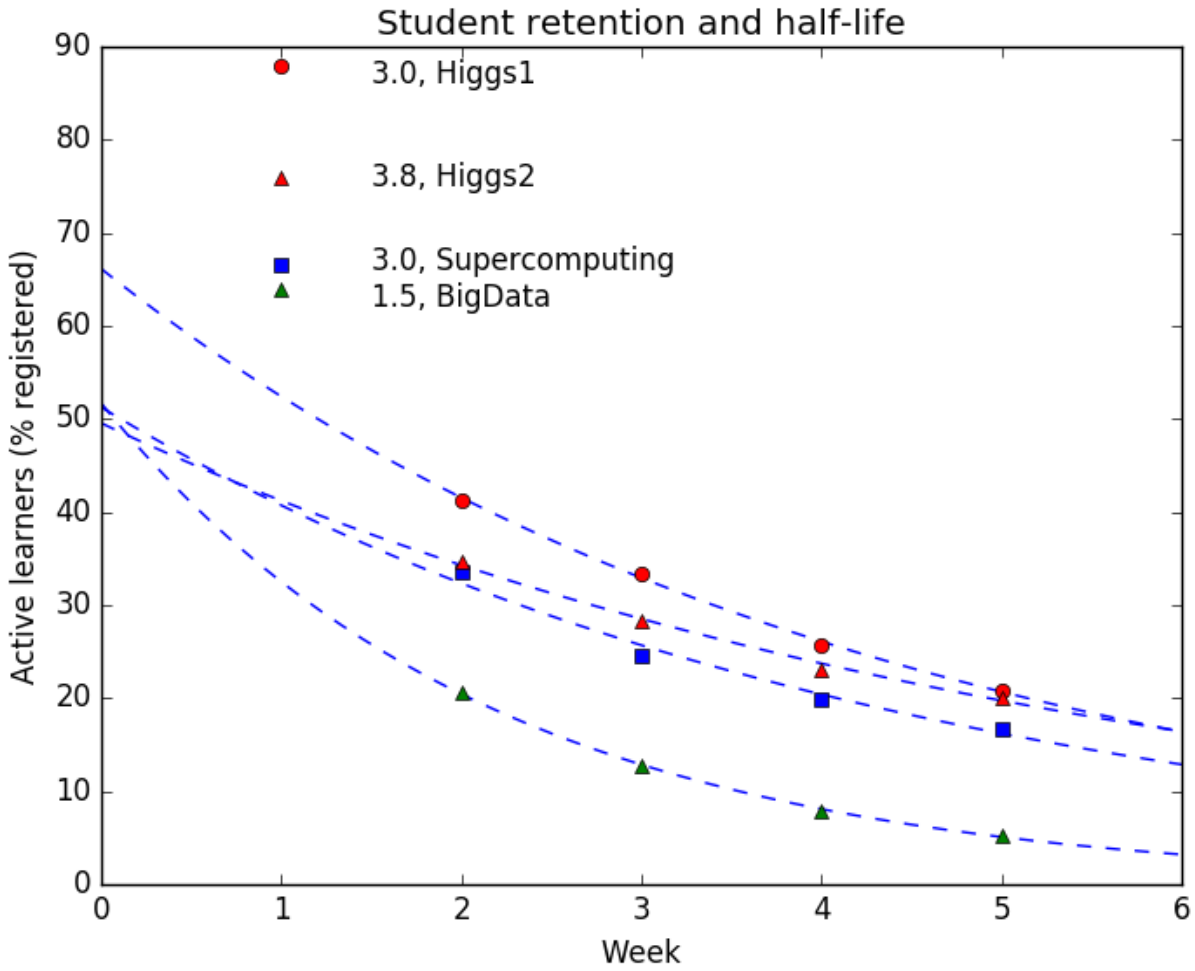
- all the workers need to know the final result.

Simple “blue-screen” videos

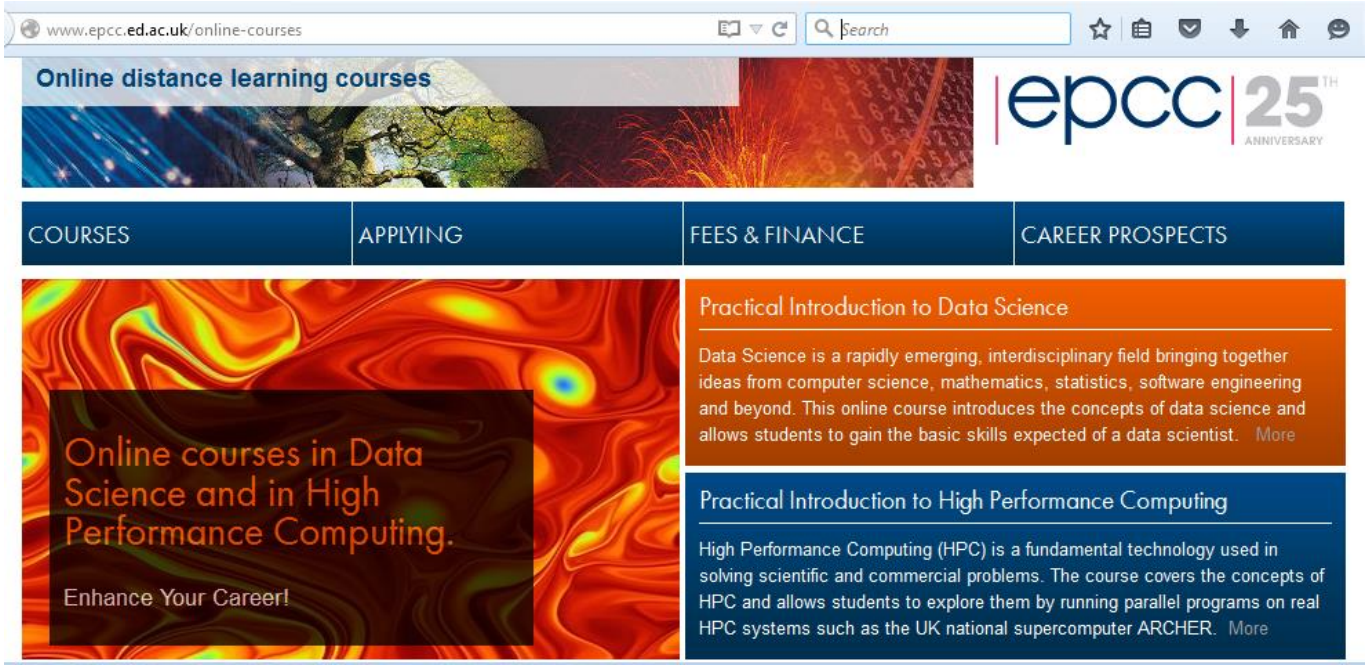
Experiences

- Two styles of MOOC
  - one very introductory with no programming (approx 3K registered)
  - one more technical using a Virtual Machine (4K registered)
- FutureLearn ethos is very collaborative
  - a lot of discussion in the comments sections
- Significant effort required to develop first run
  - not much effort to support an ongoing run
  - not much effort to re-run again
- Dissemination is a big issue
  - how to promote to non-HPC communities?
- Would MOOC format suit advanced MPI programing course?

Comparison with Edinburgh "Higgs" MOOC



Online accredited courses



- Run from January to June each year
  - entirely online: www.epcc.ed.ac.uk/online-courses/.
  - each course is 20 credits (c.f. a 180-credit MSc)

MSc in HPC / HPC with Data Science

- Running since 2001
  - typically 25-30 students each year
  - many dissertations collaborative with local companies

International HPC Summer School

- Started in 2010 in Catania, Sicily, as DEISA/TeraGRID School
  - then PRACE/XSEDE
  - now Europe, US, Japan and Canada
  - Lake Tahoe, Dublin, New York, Budapest, Toronto, Ljubljana & Boulder
- Strong focus on diversity of attendees and hands-on sessions

The Future

- HPC is becoming more complex
  - demand for more advanced training
  - solid grounding in fundamental concepts ***even more important***
- Advanced training is difficult to deliver
  - small pool of potential trainers
  - small audience
  - high maintenance cost

- Seems ideal for online delivery
  - Would MOOC platforms work?
  - or follow XSEDE remote delivery format?

Conclusions

- Need to increase the number of expert users
  - promote awareness raising for beginners
  - promote solid understanding of fundamentals
  - make advanced training easily accessible
- Recognise ***Research Software Engineer*** as career
  - break down training vs education barriers