ISC2017 - PRACE Training Workshop

Ioannis Liabotis
GRNET S.A.

Agenda of PRACE Training Workshop on ISC17 (half day - 22/06)				
sessions	**talk**	**title**	**speaker**	**chair**
014:00 - 16:00	14:00-14:10	Welcome address and review of PRACE training network	PRACE representative	chair: Nia Alexandrov(BSC)
	14:10-14:30	Developing Learning Paths through the PRACE PATC Curricula to Address HPC User Communities	Maria Ribera Sancho (BSC)	
	14:30- 14:50	Reaching out: MOOCs as introduction to HPC	Emmanouil Farsarakis (EPCC)	
	14:50- 15:10	On High performance Big Data	Dhabaleswar K. Panda (OSU)	
	15:10 -15:30	PRACE Internships: Influencing university leavers towards HPC related career choices	Leon Kos (UL)	
	15:30-16:00	SME HPC Adoption Programme in Europe	Paul Graham (University of Edinburgh)	
16:00- 16:30	**coffee break**			
16:30-18:00	16:30-17:00	On the HPC convergence with Data Science and how PRACE and EUDAT training can address it	Shaun de Witt(UK Atomic Energy Authority)	chair and panel moderator: Ioannis Liabotis
	17:00-17:30	Training Portal as teaching resource for a widening community of users	Stelios Erotokritou (CYI)	
	17:30 - 18:00	Panel Discussion: EU HPC Initiatives and HPC Training Dhabaleswar K. Panda(OSU) Paul Graham (University of Edinburgh) Shaun de Witt (UK Atomic Energy Authority)		

ISC 2017 PRACE Training Workshop - Bios

Invited Speakers:

Name: Dhabaleswar K. (DK) Panda
Email: panda@cse.ohio-state.edu
Company/Institution: The Ohio State University
Role in the institution and/or project: Professor and University Distinguished Scholar
Biography: DK Panda is a Professor and University Distinguished Scholar of Computer Science and Engineering at the Ohio State University. He has published over 400 papers in the area of high-end computing and networking. The MVAPICH2 libraries with support for MPI and PGAS on IB, Omni-Path, iWARP, RoCE, GPGPUs, Xeon Phis and virtualization (http://mvapich.cse.ohio-state.edu), are currently being used by more than 2,725 organizations worldwide (in 83 countries). More than 408,000 downloads of these libraries have taken place from the project's site. These libraries are empowering several InfiniBand clusters (including the 1st, 13th, 17th and 40th ranked ones) in the TOP500 list. The RDMA packages for Apache Hadoop, Apache Spark and Memcached together with OSU HiBD benchmarks from his group (http://hibd.cse.ohio-state.edu) are also publicly available. These packages are currently being used by more than 205 organizations from 29 countries. More than 19,000 downloads of these packages have taken place from the project's site. High-performance Deep Learning frameworks like Caffe are available from the newly created High-Performance Deep Leaning (http://hidl.cse.ohio-state.edU) project site. He is an IEEE Fellow. More details about Prof. Panda are available at http://www.cse.ohio-state.edu/~panda.

Name: Paul Graham
Email: p.graham@epcc.ed.ac.uk
Company/Institution: EPCC, University of Edinburgh
Role in the institution and/or project: Software Architect at EPCC, Coordinator of the PRACE SHAPE (SME HPC Adoption Programme in Europe) initiative
Biography: Paul Graham is a Software Architect at EPCC at the University of Edinburgh, and is the coordinator of SHAPE (SME HPC Adoption Programme in Europe), a pan-European programme that promotes High Performance Computing adoption by SMEs (small to medium sized enterprises), supported as part of the PRACE initiative.
Paul graduated in 1995 with a BSc (Hons) in Computational Physics from Edinburgh University, and then went to work at the ICI Wilton Research & Technology Centre in Middlesbrough as a member of their Computer Modelling Team. Paul subsequently joined EPCC in 1998, and in the years since has worked on a broad range of projects, principally with industrial and commercial partners, as both project manager and technical lead. Technical highlights include data mining for a national bank, software performance optimisation for Rolls Royce, parallelisation of electro-magnetic modelling code for the oil industry, modelling virtual musical instruments using GPUs, and many projects with local SMEs, enabling the technology transfer of EPCC's High Performance Computing expertise.

Name: Shaun de Witt
Email: shaun.de-witt@ukaea.uk
Company/Institution: UK Atomic Energy Authority

Role in the institution and/or project: Senior Software Development Engineer at UKAEA / Culham Centre for Fusion Energy; EUDAT training responsible

Biography: Shaun de Witt is an expert in federated data management within the Science and Technology Facilities Council (STFC) in the UK. He has previously worked on a number of federated and distributed data management projects including the Worldwide LGC Computing Grid (WLCG) and the NASA Mission To Planet Earth program. He is currently leading the EUDAT work package investigating Scalability and Data Preservation

PRACE Presenters:

Name: Ioannis Liabotis

Email: iliaboti@grnet.gr

Company/Institution: Greek Research and Technology Network, Greece

Biography: Ioannis Liabotis is a Compute Infrastructures Project Manager at GRNET S.A. During his nine years at GRNET he has been involved in a series of EC and nationally funded e-Infrastructures projects related to HPC, Grids and Cloud. He has been the technical coordinator of three e-Infrastructure projects in the area of South Eastern Europe, namely SEE-GRID-2, SEE-GRID-SCI, HP-SEE. He is currently a member of the PRACE Council and a member of the PRACE projects management boards representing Greece. Previously he has worked in research and commercial organizations doing research and development in the area of computer networking

Name: Nia Alexandrov

Email: nia.alexandrov@bsc.es

Company/Institution: Barcelona Supercomputing Centre, Spain

Biography: Dr. Nia Alexandrov is the HPC Training Coordinator at BSC, Barcelona. She has over 11 years' experience as Postgraduate Studies Coordinator at the School of Systems Engineering and ACET (Advanced Computing and Emerging Technologies) Centre at the University of Reading and now at BSC. She coordinated multidisciplinary, joint and research based MSc Programs and taught on Post-Graduate level. She is co-organiser of education and training related workshops on the SC and ICCS conferences. Her research is in the area of collaborative and e-learning, with focus on curricula development, e-learning methodology and methodology for evaluation of on-line and blended learning. She was a member of the Career Space consortium and was involved with the Innovation Value Institute HE working group. She was researcher on a string of EC learning related projects and is the coordinator of PRACE Advanced Training Centre at BSC.

Name: Emmanouil Farsarakis

Email: e.farsarakis@epcc.ed.ac.uk

Company/Institution: *Edinburgh Parallel Computing Centre,* University of Edinburg(EPCC), UK

Biography: Emmanouil Farsarakis is an Applications Consultant on the Computational Science & Engineering team at EPCC, the University of Edinburgh. He has a background in Physics which he studied at the University of Crete. He moved to Edinburgh to continue his studies with an MSc in High Performance Computing from EPCC after several years of working in science and TEFL education.

Since joining the CSE team at EPCC, he has enjoyed working on numerous projects. He is a part of the IPCC team which works on modernising HPC codes for Intel Xeon Phi coprocessors. He is also involved in the European project NEXTGenIO, doing innovative research in HPC I/O for Exascale.

Last but not least, he is greatly involved in HPC training at EPCC. Emmanouil leads Software Carpentry and Data Carpentry workshops for ARCHER training, he participates in MSc in HPC and HPC with Data Science training and he is also involved in the development of a PRACE Supercomputing MOOC on Futurelearn.

Name: Leon Kos

Email: leon.kos@lecad.fs.uni-lj.si

Company/Institution: University of Ljubljana, Slovenia

PRACE affiliation: PRACE Summer of HPC Internship Program, 2015

Biography: Leon Kos is a 20+ years veteran of using Linux desktop on a daily basis to build digital relationships for research, teaching, and getting the job done by using many programming languages and Unix flavours. He likes embedded systems the same as big machines. Advocates open science and data provenance. Research areas include plasma physics, fusion, CAD kernel and 3D scientific visualisation, where he frequently encounters software development requests by HPC related projects. Leon advises and leads national HPC centre initiative, administers university cluster, organises HPC training courses in Slovenia and coordinates PRACE Summer of HPC 2015.

Name: Stelios Erotokritou

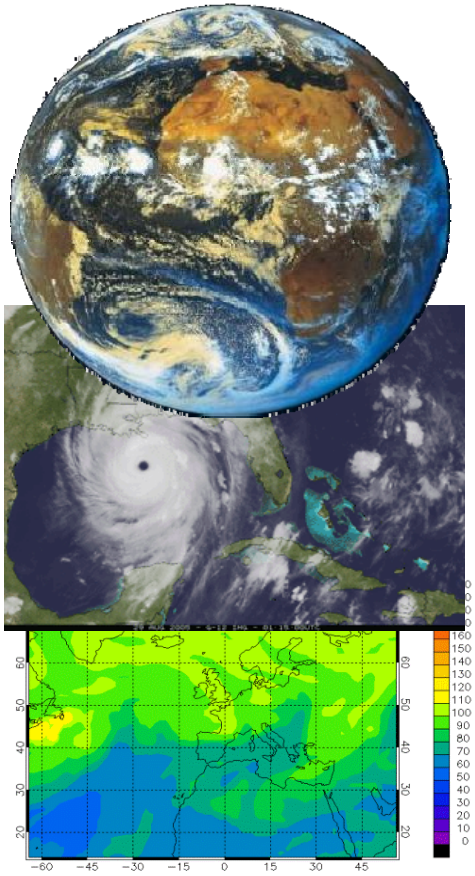
Email: s.erotokritou@cyi.ac.cy

Company/Institution: The Cyprus Institute, Cyprus

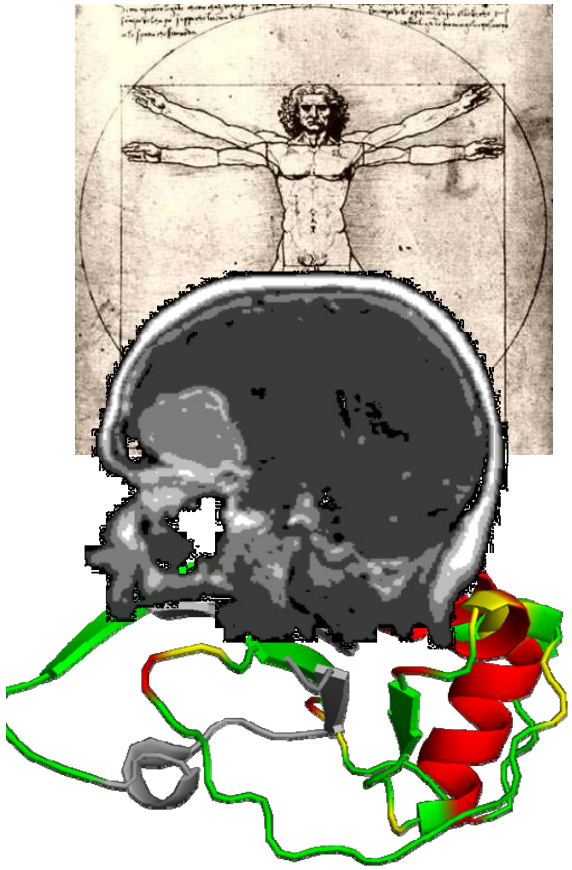
PRACE affiliation: Training Portal Management

Biography: Stelios received his BSc in Computer Science from the University of Cyprus. He continued his studies at University College London (UCL) with MSc studies in Networking and PhD studies in the field of Cryptography. His PhD studies were partly funded by British Telecom (BT) and focused on Perfectly Secure Message Transmission Protocols (PSMT) and their Applications. In September 2011 he joined The Cyprus Institute in a computing support specialist role. Since then he has mainly been working in the PRACE projects (1/2/3/4 IP) with most of his involvement in the Training and Dissemination Work Packages. At The Cyprus Institute he is also a member of the HPC operations team and is also the application chair for the LinkSCEEM/Cy-Tera production calls.

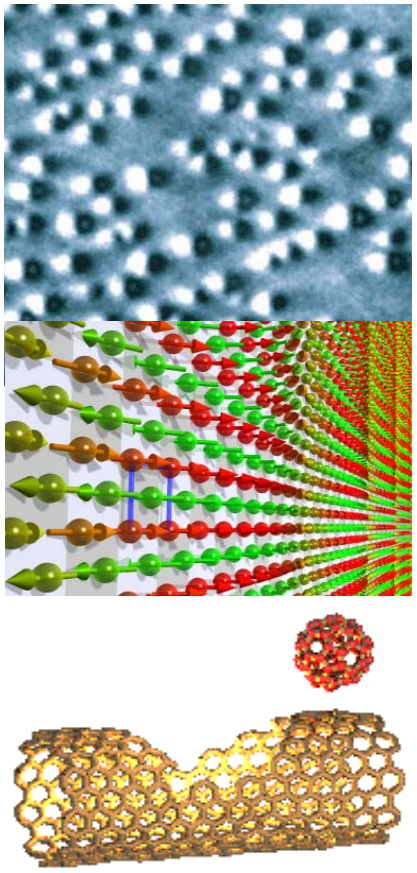
Supercomputing Drives Science through Simulation



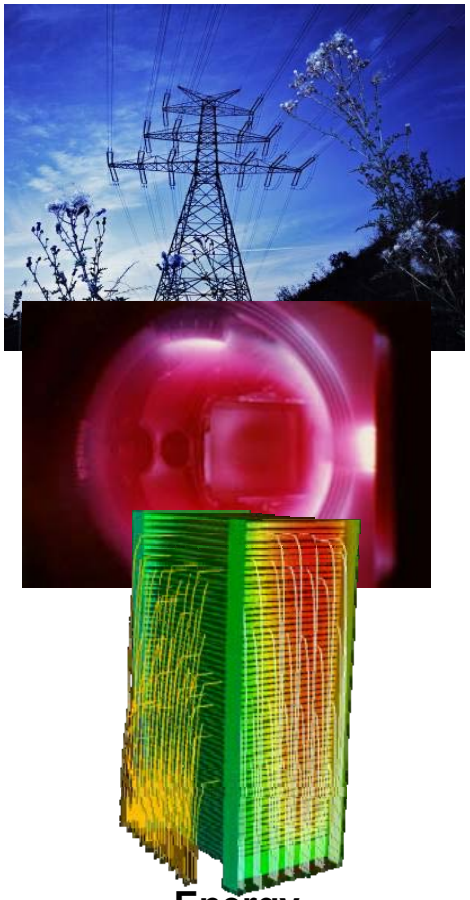
Environment
Weather/ Climatology
Pollution/Sustainability



Ageing Society
Medicine
Biology



Materials/ Inf. Tech
Nano-science
Quantum Computing

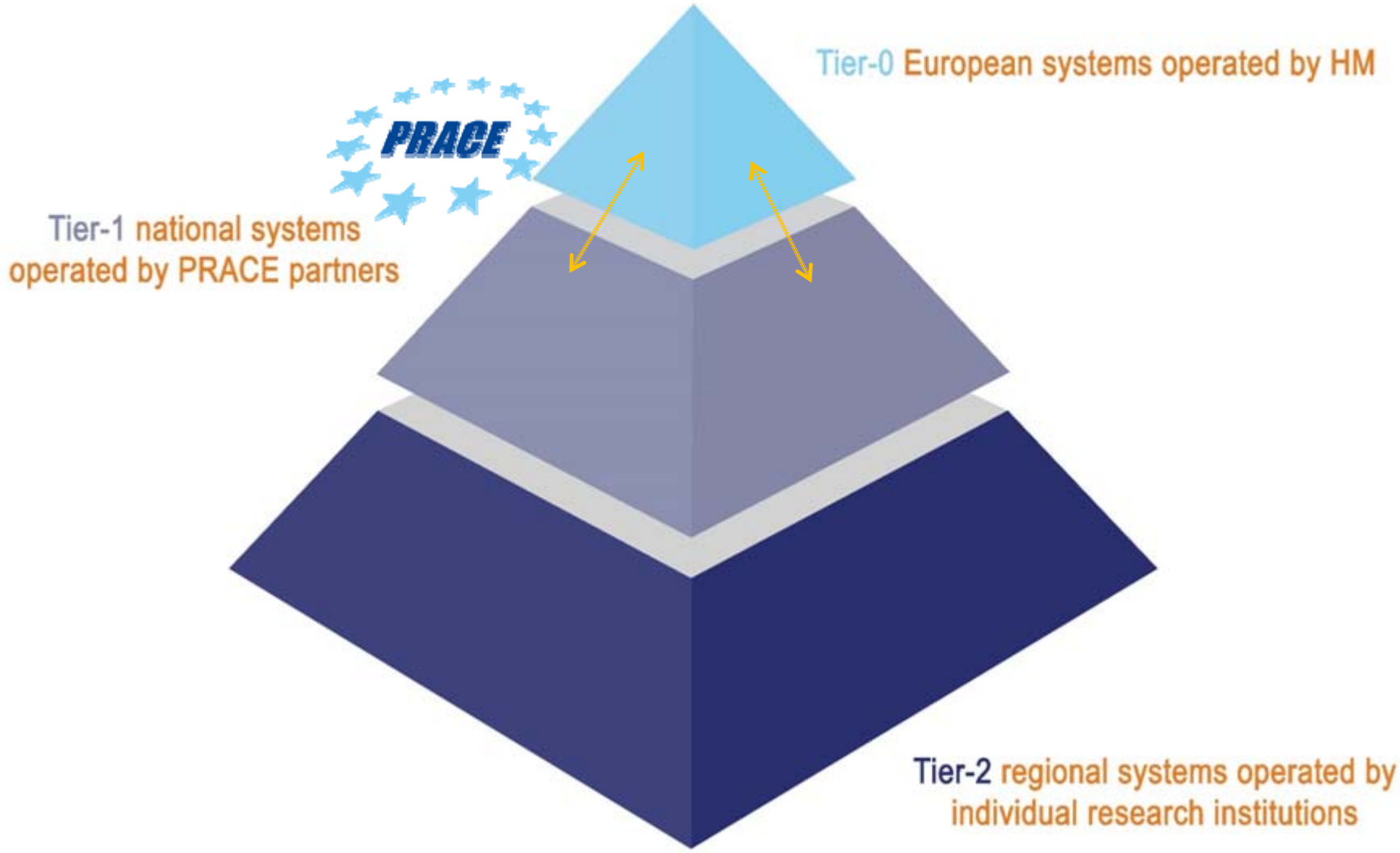


Energy
Plasma Physics
Fuel Cells

Partnership for Advanced Computing in Europe

- Open access to best-of-breed HPC-systems to EU Scientists
- Variety of architectures to support the different scientific communities
- High standards in computational science and engineering
- Peer review on European scale to foster scientific excellence
- Robust and persistent funding scheme for HPC supported by the national governments and the EC
- Support the development of IPR in Europe by working with industry and public services
- Collaborate with European HPC industrial users and suppliers

HPC Pyramid


Tier-0 European systems operated by HM
Tier-1 national systems
operated by PRACE partners
Tier-2 regional systems operated by
individual research institutions

PRACE Tier 0 Systems

MareNostrum: IBM
BSC, Barcelona, Spain

NEW ENTRY 2016

Piz Daint: Cray XC 30
CSCS
Lugano, Switzerland

CURIE: Bull Bullx
GENCI/CEA
Bruyères-le-Châtel, France

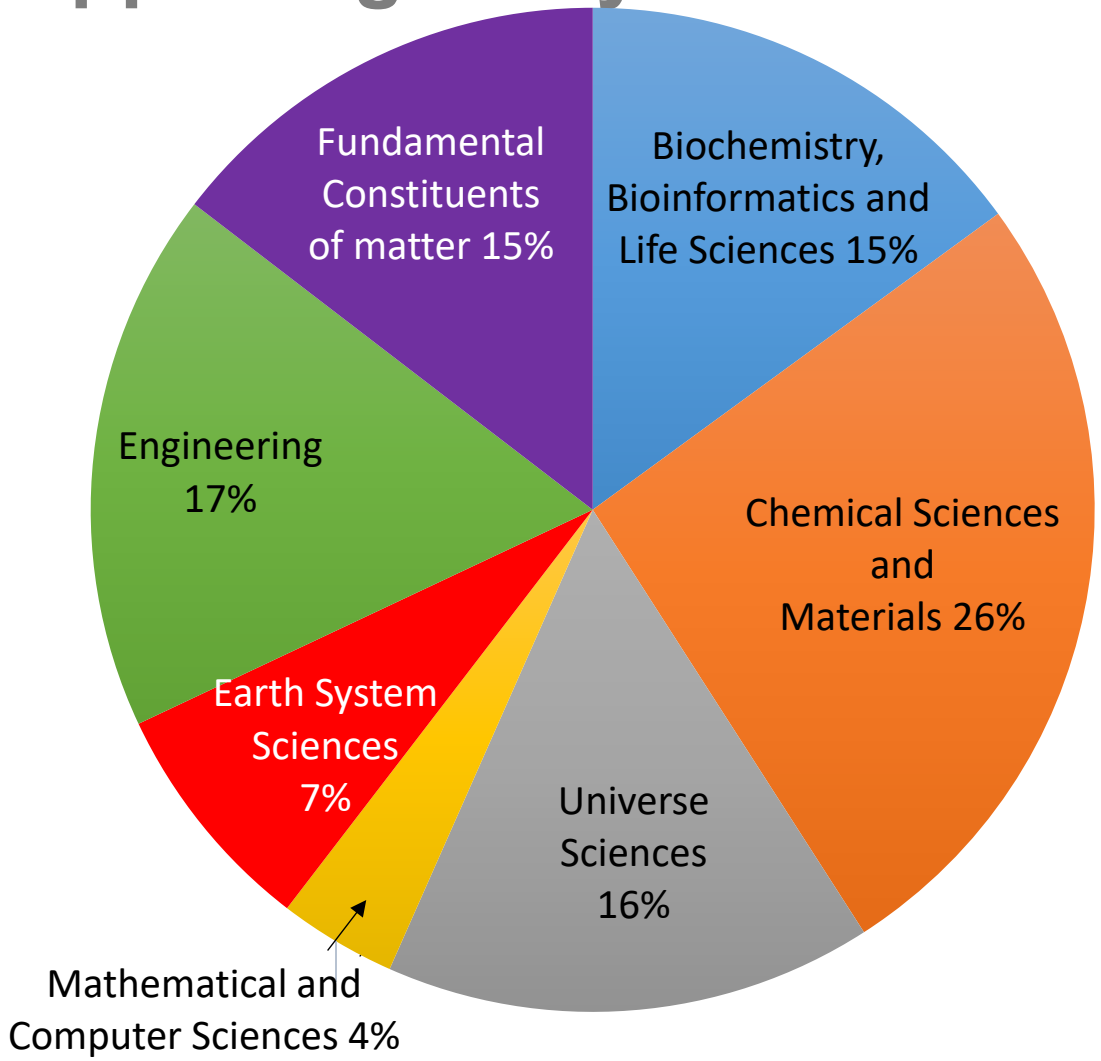
JUQUEEN: IBM
BlueGene/Q
GAUSS/FZJ
Jülich, Germany

SuperMUC: IBM
GAUSS/LRZ
Garching, Germany

Hazel Hen: Cray
GAUSS/HLRS,
Stuttgart, Germany

MARCONI: Lenovo
CINECA
Bologna, Italy

Supporting many scientific domains

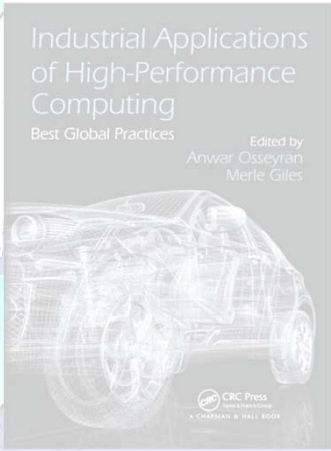


Research Domain Pie Chart up to and including Call 14, % of total core hours awarded

Criterion: Scientific Excellence

PRACE Achievements to-date

- ✓ 524 scientific projects enabled
- ✓ 14 000 000 000 (thousand million) core hours awarded since 2010
- ✓ Of which 63% led by another PI nationality than the HM
- ✓ R&D access to industrial users with >50 companies supported
- ✓ >10 000 people trained by 6 PRACE Advanced Training Centers and others events
- ✓ >60 Petaflops of peak performance on 7 world-class systems
- ✓ 24 PRACE members, including 5 Hosting Members (France, Germany, Italy, Spain and Switzerland)
- ✓ Best Seller book on Industrial Applications of HPC



PRACE Training and Outreach activities

provide a sustained, high-quality training and education service for the European HPC community

- 6 PRACE Advanced Training Centres
  4 PRACE Training Centres
- PRACE training events: Seasonal Schools, International HPC Summer School, On-demand training events
- Summer of HPC (programme for undergraduate and postgraduate students)
- PRACE Training and Events portal
- Code Vault, Massive Open Online Courses (MOOCs)

Training topics

Different levels of training

- Basic, intermediate, advanced

High performance computing

- Parallel programming
- Accelerators
- Performance optimization

Domain-specific topics

- Simulation software
- Visualization
- Data intensive computing

Open access to industry since January 2012

Access to leading-edge resources

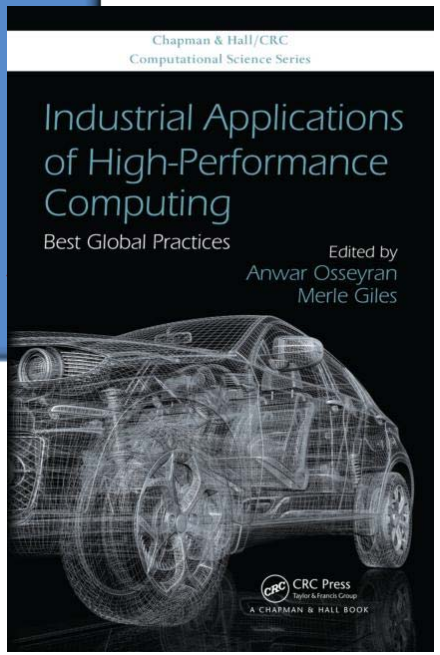
- Calls open to industry twice a year
  - commitment to publish results
- Permanent Preparatory Access call
  - For scalability and porting

Access to high-value services

- HPC knowledge
- Training
- Code enabling
- Information, Promotion and Networking

Since 2012: >50 companies supported

PRACE SHAPE
SME HPC Adoption Programme in Europe



THANK YOU FOR YOUR ATTENTION

www.prace-ri.eu

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Developing BSC Learning Paths through the PATC Curricula to Address HPC User Communities

PRACE Training Workshop at ISC 17

Nia Alexandrov

Maria-Ribera Sancho

22/06/2017

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

BSC Profile

Education and training at BSC

Engagement with the University

- MSc Degree in High Performance Computing FIB-UPC
- Double Diploma Agreement CIC(IPN)-FIB(UPC)
- Bachelor Degree in Bioinformatics
- Doctoral Program Affiliation with Catalan Universities
- PhD and Postdoc (currently 89 PhD Students and 51 Postdoc Researchers)

Training Programme

- PRACE Advanced Training Center (14 courses currently offered with over 1400 attendees up to date)
- RES workshops and meetings(for researchers, technicians and scientists, over 600 attendees)
- CUDA Center of Excellence courses (over 730 CUDA trainees in 2011-15. Now NVIDIA GPU Education Center

In Projects

- Project training initiatives (RISC, EUBrazilOpenBio, ETP4HPC, Montblanc, EESI, BigStorage, HBP …)
- Education Research & development (RISC, PRACE, ITNs, AGAUR,…)
- Severo Ochoa leveraging E&T Program:
  - ✓ Mobility Program (30 incoming and 30 outgoing mobility grants in 2013-2015)
  - ✓ Doctoral Symposium 2014, 15, 16 with an average of 75 attendees

Research workshops and publications

- Workshops (Best Practices for HPC Training on SC14, 15, 16, PRACE Training Workshop on ISC 15 and 16, BRIDGE ICCS 14, 15 and 16….)
- Summer Schools (Computational Simulations on Aircraft Icing 2014 and 16, E2SCMS2 2014, MAPFER 2013, PUMPS 7 editions, Deep Learning for Computer Vision, Machine Learning Tutorial 2016 …)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

Creating International Synergies in Education and Training

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Education and Training at BSC – 2 sides of training provision

Postgraduate Studies and Professional Development of Young Researchers

BSC upholds HRS4R and European Charter for Researchers

- Mentoring programs
- Individual Learning path
- Long term career goals

Researcher Profile:

- Multidisciplinarity of PhD projects
- Non-CS background is frequent

Degree Awarding Institution processes and requirements

- Subject and institution specific
- Project timeframe and monitoring process

Professional Training Advanced Centre

Curricula targeting HPC skills gap

Mixed audiences:

- PhD, MSc students
- Post Doc researchers
- Industrial attendees

Short term engagement with attendees

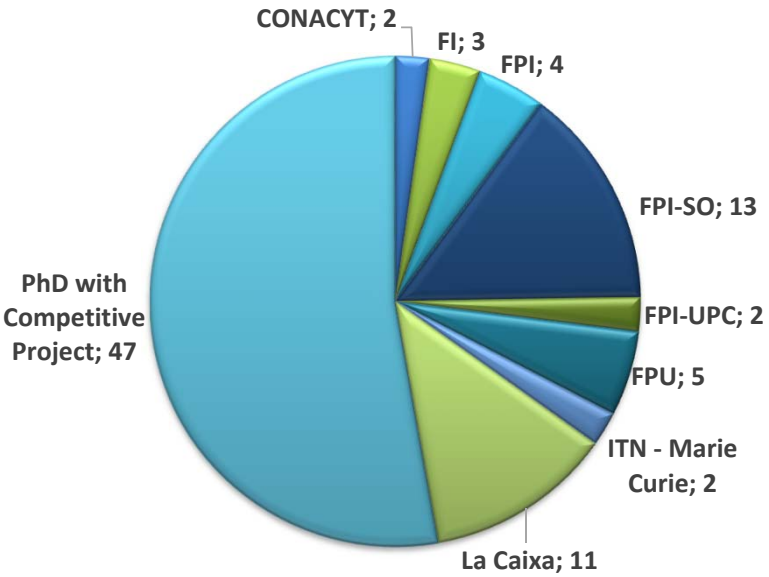
- All courses are maximum of 5 days

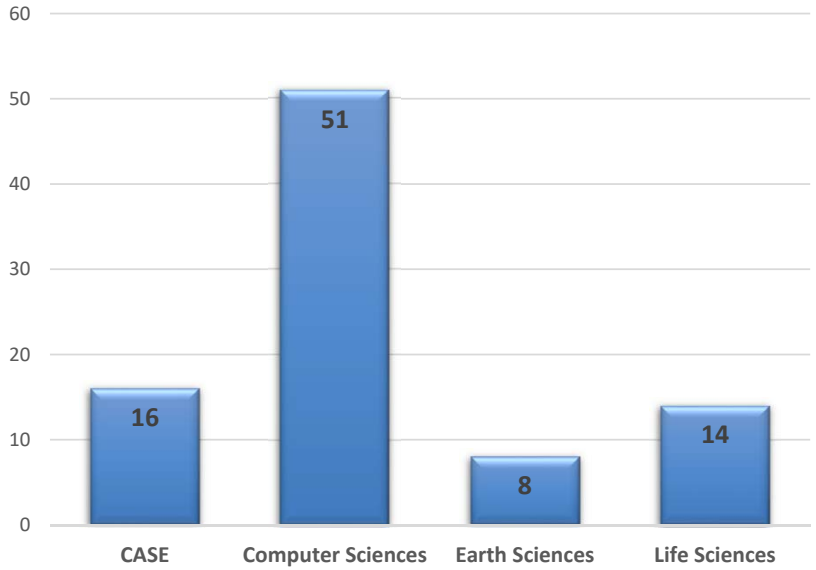
Short term applied goals of the attendees

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

Education: Doctoral Studies

- Current PhD's figures:
  - **89 PhD Students**
  - **21 Nationalities**





43% are from non-computer science background or working in application areas

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

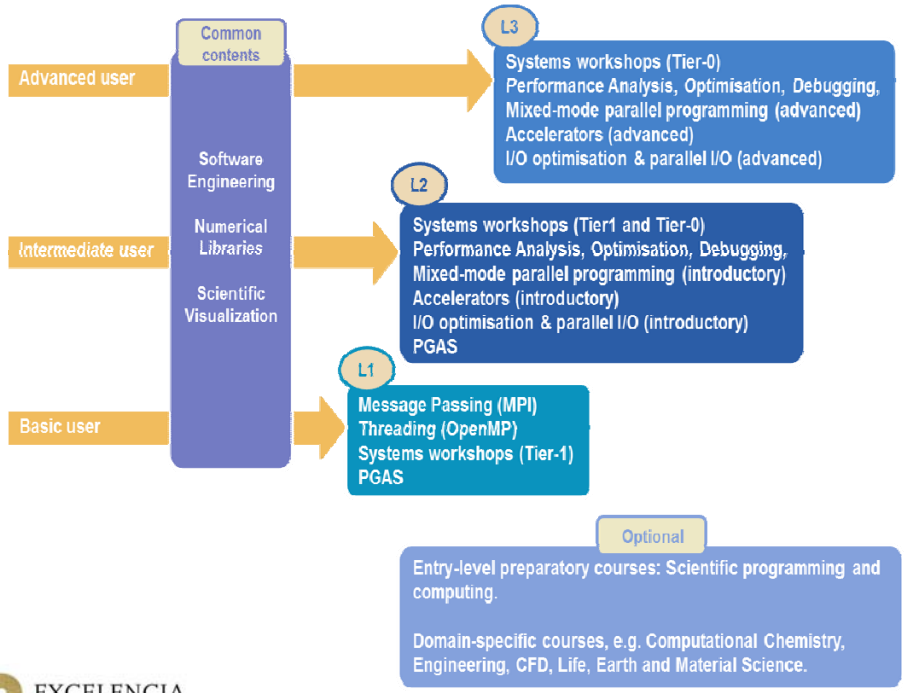
EXCELENCIA SEVERO OCHOA

Designing Learning Paths

Designing a Training Profile specific Learning Path 1/2

Motivation:

Although there are existing curricula(PRACE curricula) and competence based structuring of learning materials (XSEDE Roadmap Training), very often users get lost in the multitude of possibilities.



Our approach – designing a Training Profile specific Learning Path 2/2

- Attendees often under or overestimate their understanding in particular area and get frustrated when do not achieve the desired results

In some of our surveys, computational scientists show unreasonably high expectations for end results of courses clearly marked as beginners and with specific learning outcomes included in the description. For example:

- Beginners course cannot make you expert in parallel programming
- Beginners course for parallel programing requires knowledge in sequential programming
- Computational scientists should consider starting with OpenACC instead of CUDA especially if they are not strong programmers

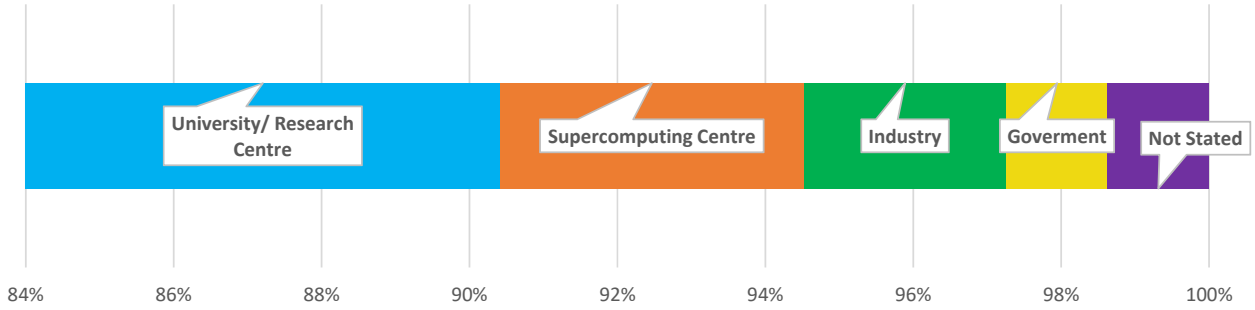
- We believe that “Start here if you are so-and-so profile” information may reduce the issues in the current context

PATC @ BSC program 2017 - 18

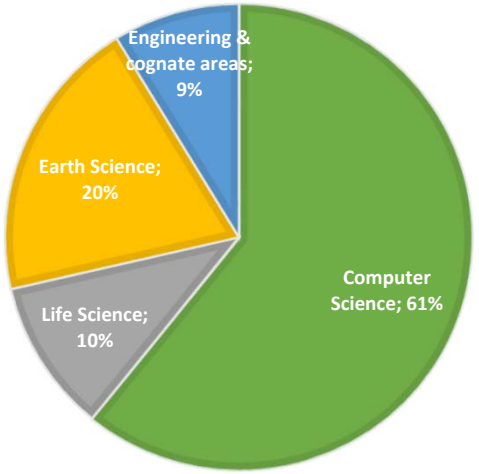
Systems Workshops as Tier 0 site:		
Programming of Petaflop Machine (Presenting MareNostrum IV)	BSC06	25-26 Sept 2017
Administration of Petaflop Machine	BSC11i	31 Jan -1 Feb 2018
Programming of Petaflop Machine	BSC06	11-12 April 2018
Core subject courses providing access to BSC developed technology:		
HPC core training	BSC01	7-11 May 2018
Part 1: Performance Analysis and Tools (Paraver)		
Part 2: Heterogeneous Programming on GPU with MPI(OMPSs)		
Part 3: ARM-based Architectures		
Programming Distributed Systems (COMPSs)	BSC14	02-feb-18
Computational Science:		
HPC simulations for Science and Engineering	BSC07	14 - 16 Feb 2018
Introduction to simulation environments for Life sciences	BSC08	14 -15 March 2018
Earth Sciences Simulation Environments	BSC09	30 Nov -2 Dec 2017
Skills for Exascale Computing:		
Parallel Programming Workshop	BSC10	24 - 20 Oct 2017
Introduction to programming in CUDA	BSC04a	Spring 2018
Introduction to OpenACC	BSC04b	Spring 2018
PUMPS	BSC05	July 2018
Big Data Analytics	BSC15	6-9 Feb 2018

Attendees – respondents profile

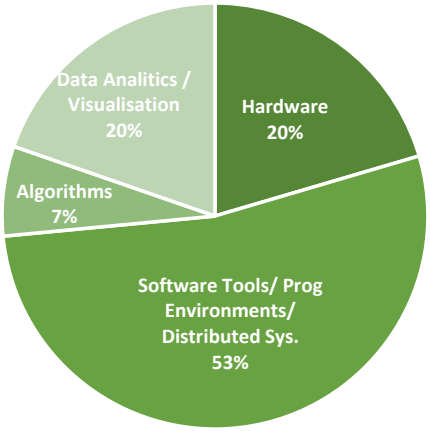
Institution



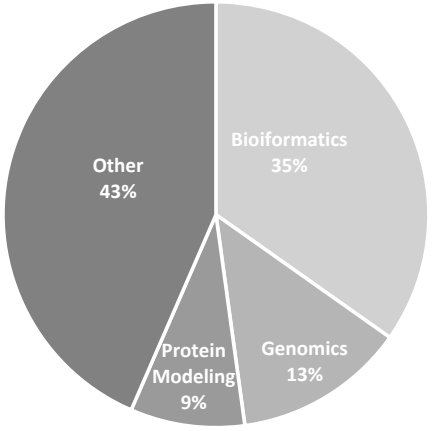
Research Area



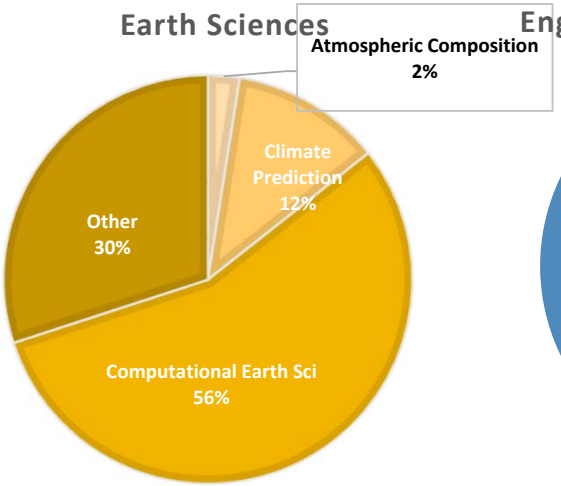
Computer Science



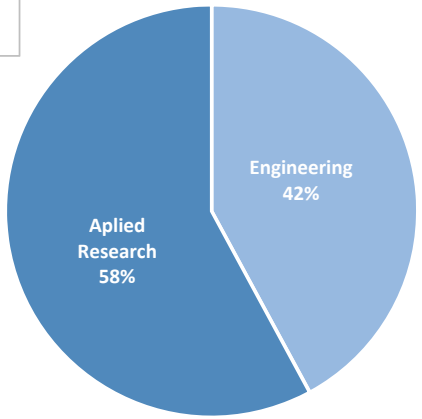
Life Sciences



Earth Sciences

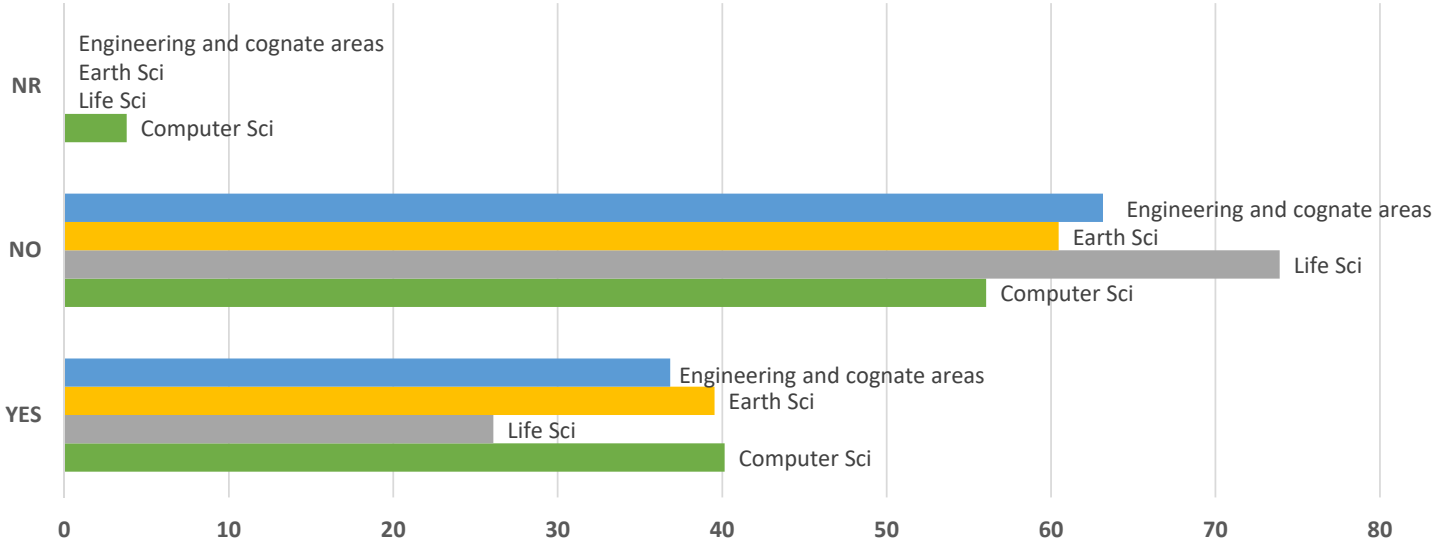


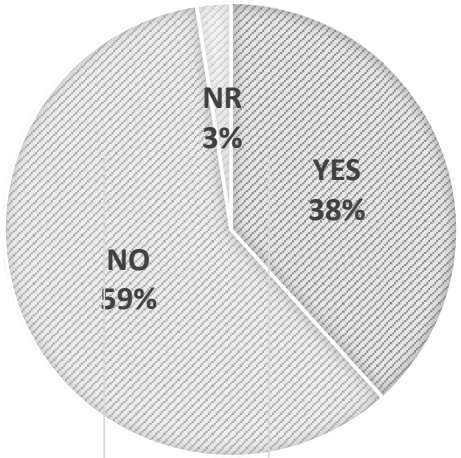
Engineering & Cognate Areas



HPC skills and BSC courses – students responses

- In the start of your work/studies did you get advice on which are the most important HPC skills you should acquire?

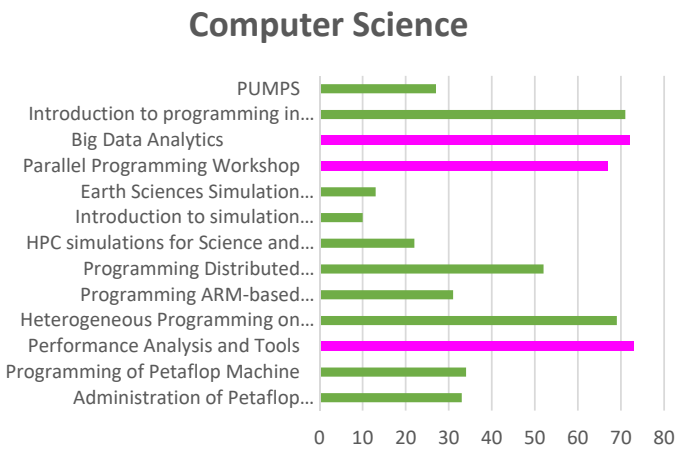


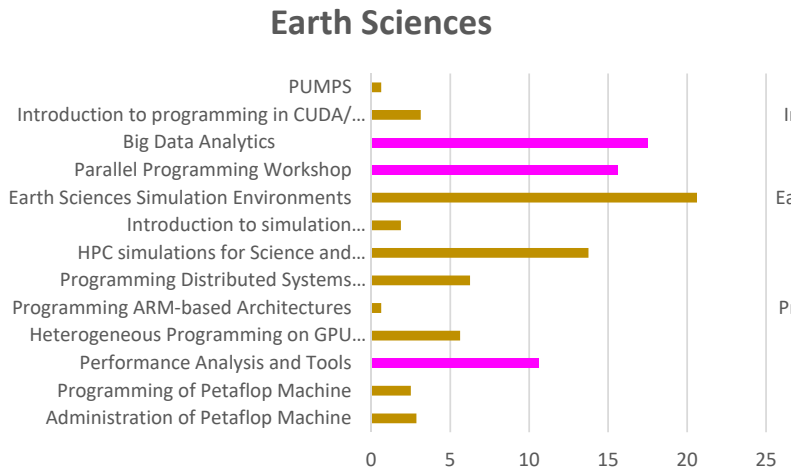


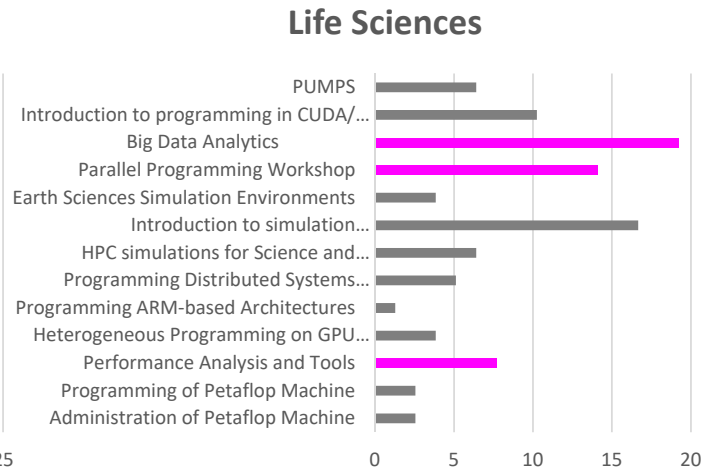
HPC skills and BSC courses – students responses

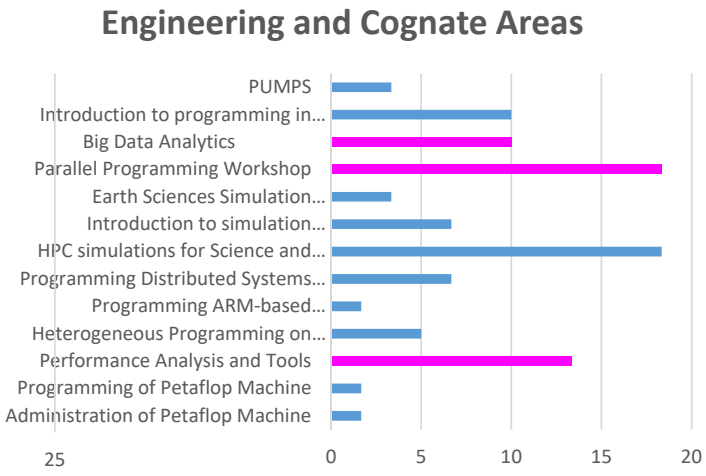
- In your opinion, which are the most useful HPC related courses for people in your line of study/work:

Consistently at the Top – Big Data Analytics, Parallel Programming Workshop, Performance Analysis and Tools

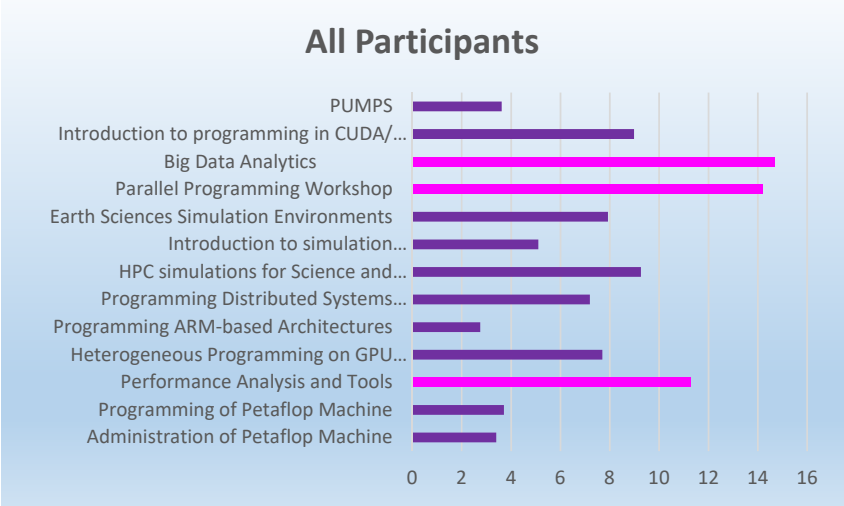


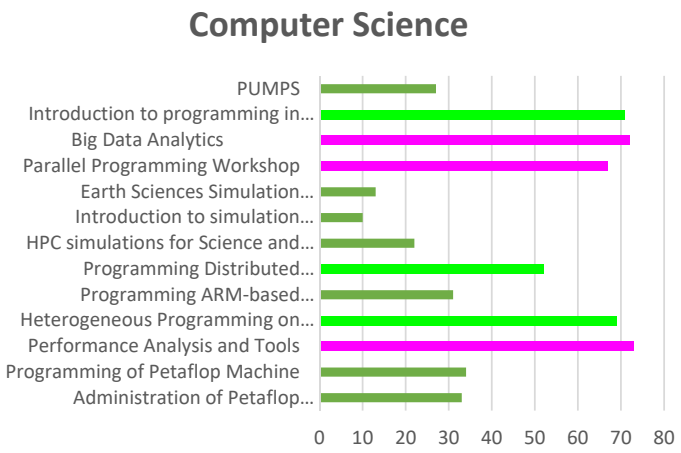


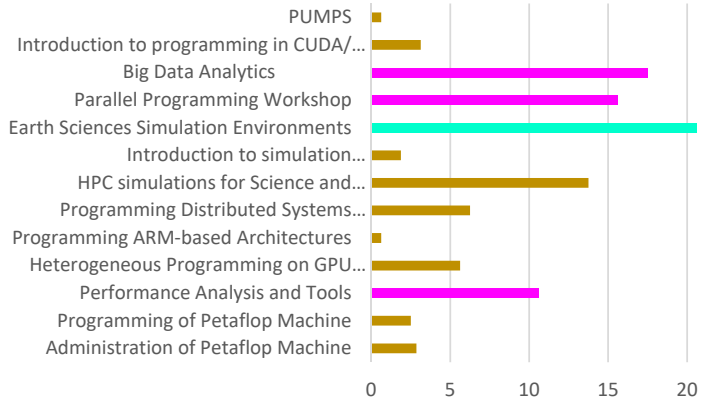


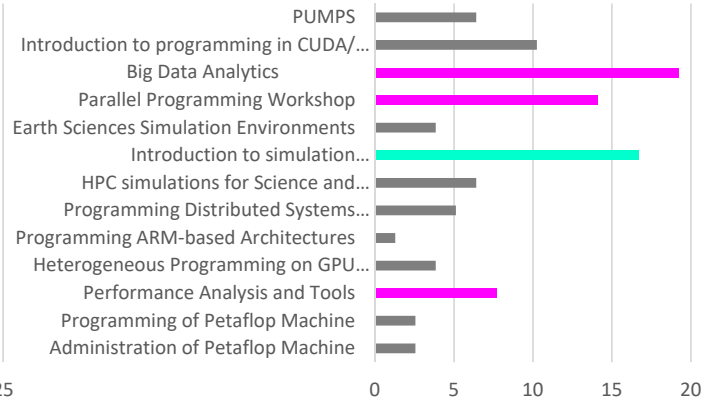


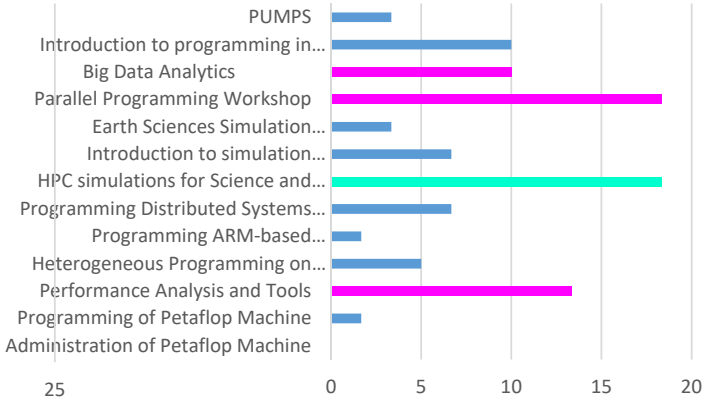
HPC skills and BSC courses – students responses



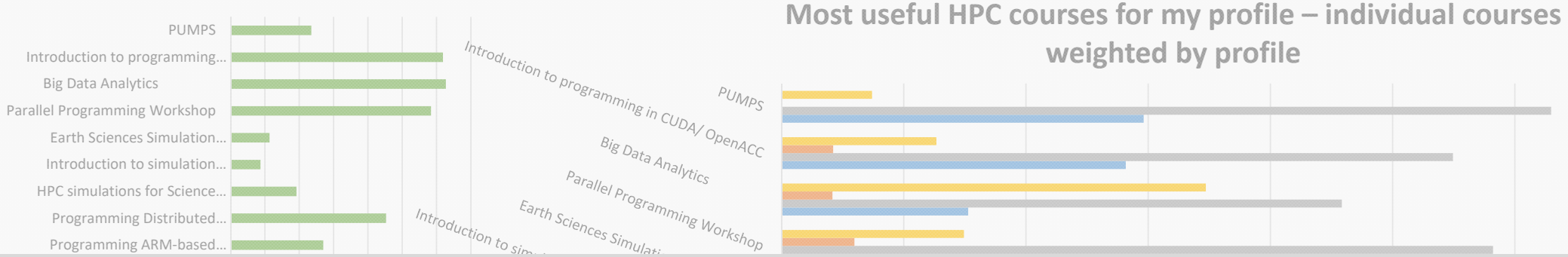






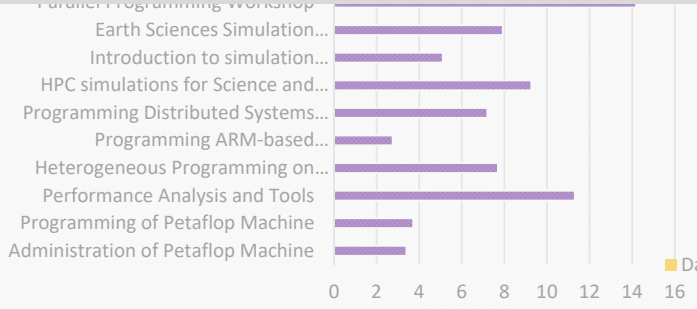


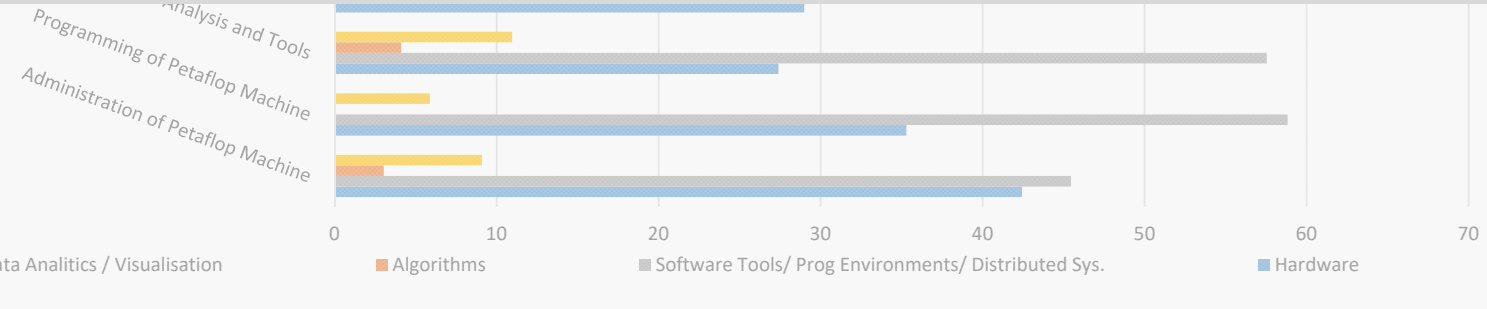
HPC skills and BSC courses: feedback returns by profile – Computer Science (PATC courses 2016-17)



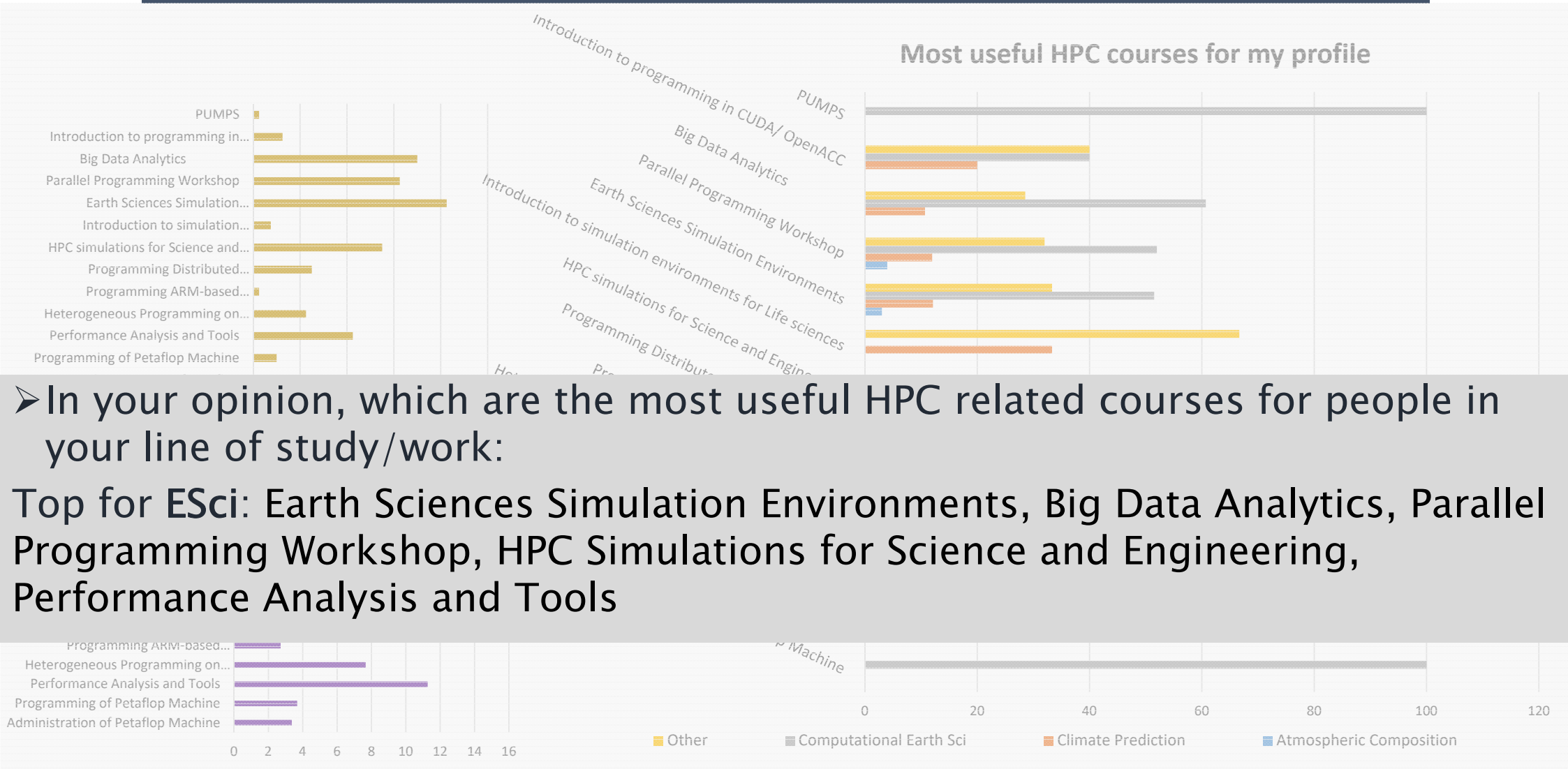
- In your opinion, which are the most useful HPC related courses for people in your line of study/work:

Top for **CSci**: Performance Analysis & Tools, Big Data Analytics, Introduction to CUDA, Heterogeneous Programming, Parallel Programming Workshop and Programming Distributed Systems





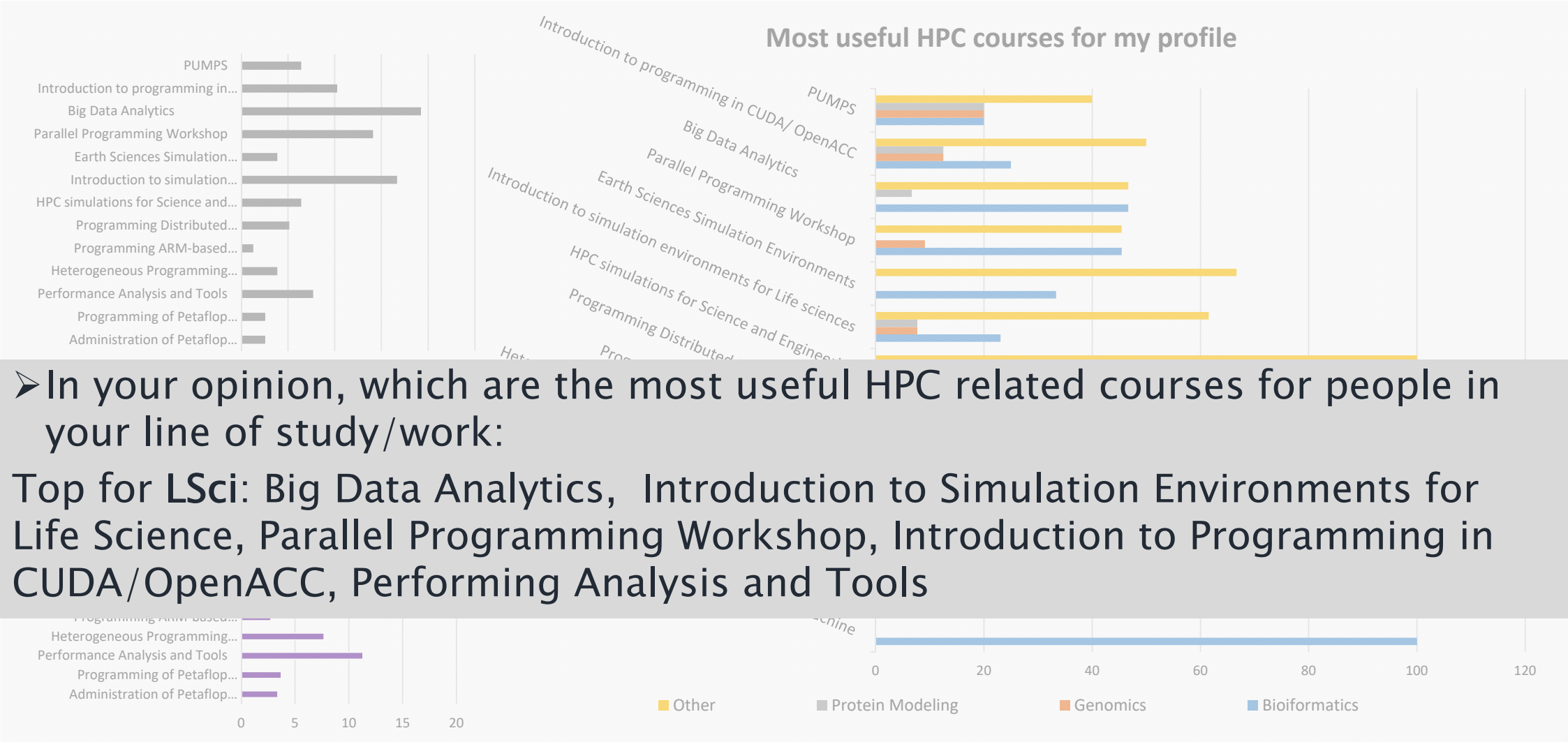
HPC skills and BSC courses – feedback returns by profile Earth Sciences (PATC courses 2016-17)



- In your opinion, which are the most useful HPC related courses for people in your line of study/work:

Top for ESci: Earth Sciences Simulation Environments, Big Data Analytics, Parallel Programming Workshop, HPC Simulations for Science and Engineering, Performance Analysis and Tools

HPC skills and BSC courses: feedback returns by profile – Life Sciences (PATC courses 2016-17)



- In your opinion, which are the most useful HPC related courses for people in your line of study/work:

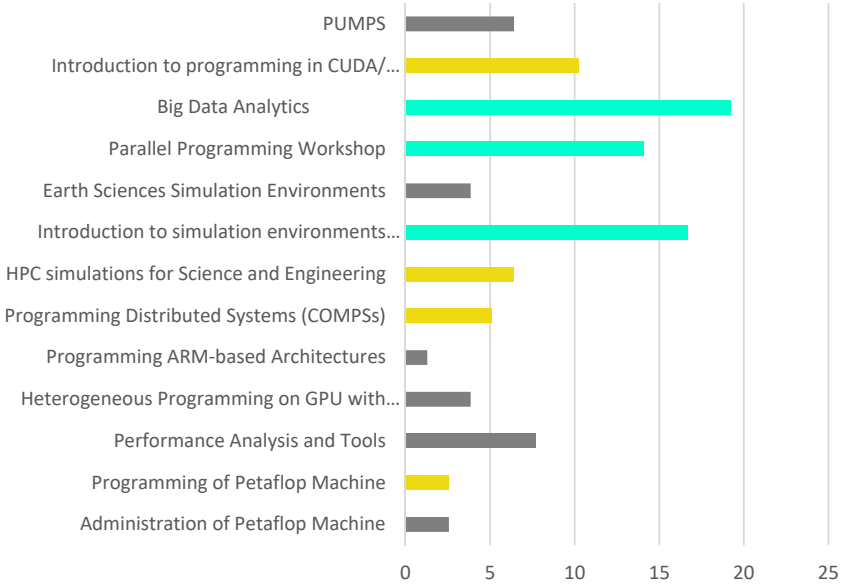
Top for **LSci**: Big Data Analytics, Introduction to Simulation Environments for Life Science, Parallel Programming Workshop, Introduction to Programming in CUDA/OpenACC, Performing Analysis and Tools

Developing Learning Paths as part of BSC Mentoring Program:

- In collaboration with senior researchers supervising students and mentoring junior staff
- Based on required key competences
- Taking into account current trends influencing participant choices
- Using the resulting information to approach research communities outside BSC

Initial sub-set for the Bio Research Community:

Preferred by attendees & suggested by conveners courses:



Sub-set based on BSC conveners' advice:

- If you are interested to understand computer simulations, you should participate on the **"HPC-based simulations, Engineering and Environment"**
- For hands–on experience of life science simulation environments, we recommend: **"Simulation Environments for Life Sciences"**
- There are several programming courses suitable for computational scientists: **"Parallel Programming Wsh", "Programming Distributed Computing Platforms with COMPSs"** as well as the **Introduction to OpenACC**
- Finally, if you would like to learn how to use one of the BSC supercomputers to run your code, you should take the **Systems Workshop: Programming MareNostrum**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Education and Training, BSC

maria.ribera@bsc.es

nia.alexandrov@bsc.es

education@bsc.es

www.bsc.es/education

Reaching Out: MOOCs as an Introduction to HPC

Emmanouil (Manos) Farsarakis

EPCC, The University of Edinburgh

Outline Points

- Initial goals and flexibility
- 1st run facts and experience
- Lessons learned
- Future plans

Development Plans

MOOC 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

Learning Outcomes

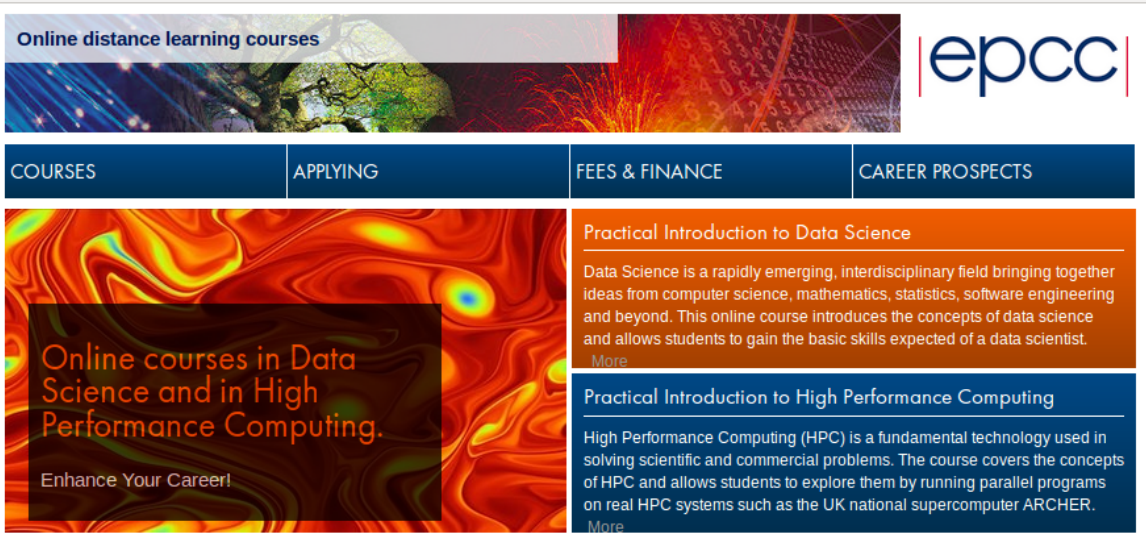
- On completion course, students should be able to:
  - understand how the performance of modern supercomputers is measured and achieved;
  - explain why they are built from thousands of simple processors;
  - understand the differences between shared-memory and distributed-memory computers;
  - compare the architecture of a modern supercomputer with a desktop PC;
  - explain why computer simulation is a fundamental component of modern scientific discovery;
  - work with simple cellular automaton models;
  - analyse simple problems and look for opportunities for parallel processing;
  - explain the limitations of parallel computing.
  - give examples of scientific areas where computer simulation is used.

Planned Delivery Method

- Largely lecture-based
  - re-use many existing lectures from introductory courses
  - record a number of short (10-minute) presentations

Planned Delivery Method

- Largely lecture-based
  - for example, re-packaging lectures from EPCC's University-accredited online programmes



www.epcc.ed.ac.uk/online-courses/

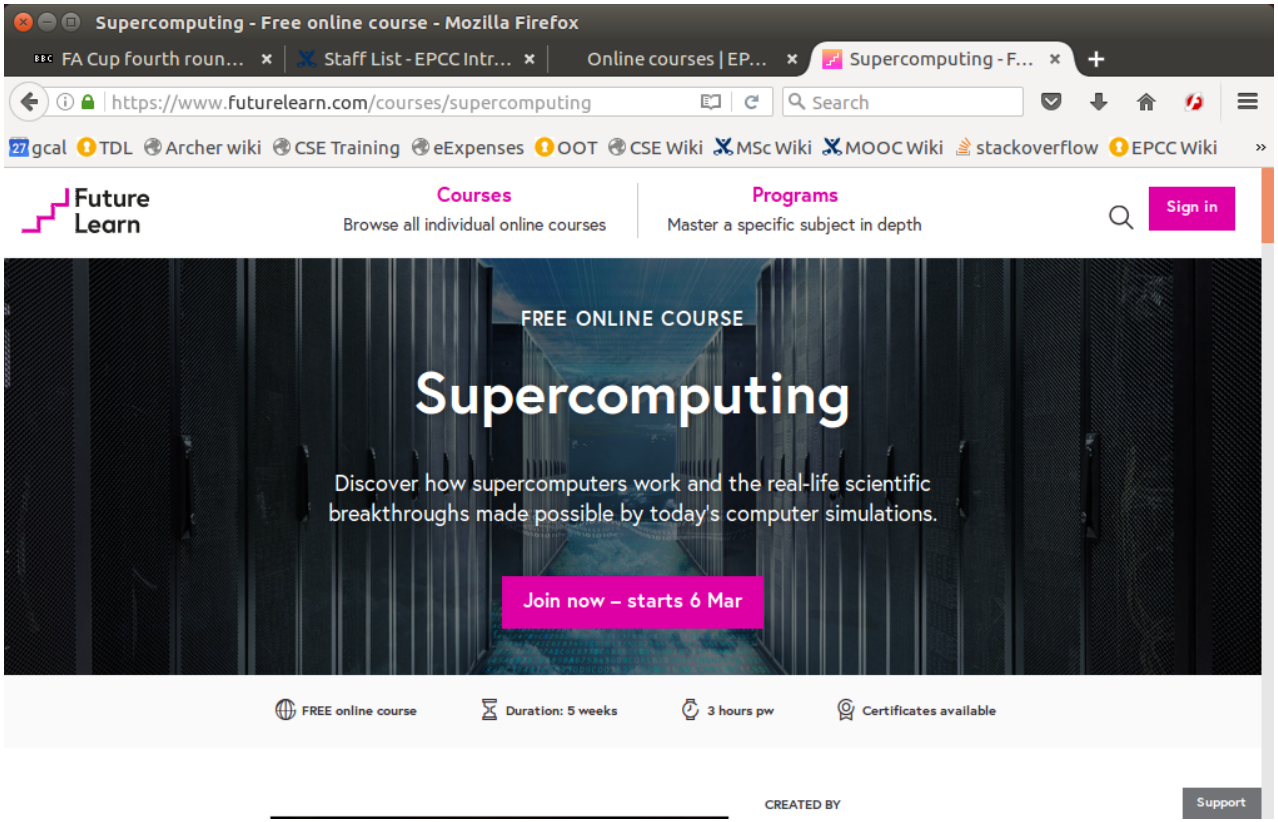
Plans Revisited

“Agile” Course Development

Curriculum and Learning Outcomes unchanged. But...:

- .... delivery not compatible with FutureLearn style:
  - many small “steps”
  - short articles and quizzes, discussion topics, external links
  - very short (max 5 minute) videos to reinforce key topics
  - videos encouraged but not for major content delivery
- FL “house style” well suited in principle to this MOOC
  - many existing FL courses entry level, pre-access to University
- So
  - needed less content overall than initially envisaged
  - but more new material had to be developed

futurelearn.com/courses/supercomputing



Promotional video

Download video: standard or HD

CREATED BY

SHARE

ABOUT THE COURSE

Today's supercomputers are the most powerful calculating machines ever invented, capable of performing more than a thousand million million calculations every second. This gives scientists and engineers a powerful new tool to study the natural world – computer simulation.

START DATES

6 Mar 2017 Join
UPCOMING

TBA Register interest
DATE TO BE ANNOUNCED

Support

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

Activities and steps

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.

Short ”Blue-screen” Videos

View transcript

Download video: standard or HD

Welcome to Week 2

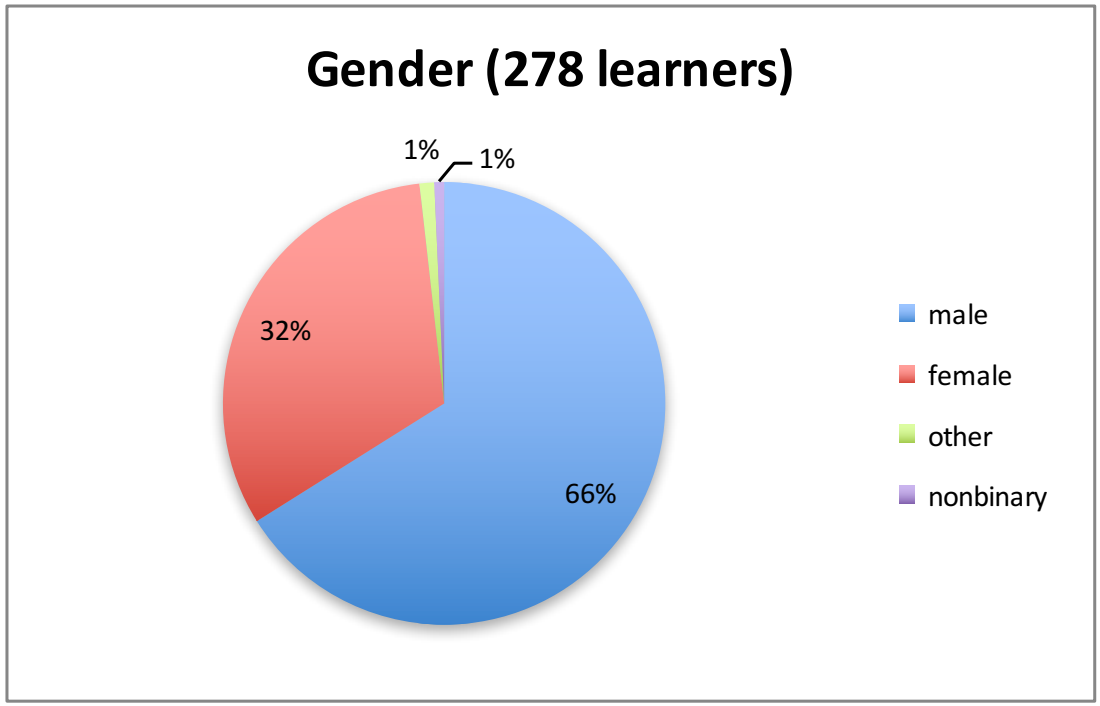
Short Promo Video

First run: Facts

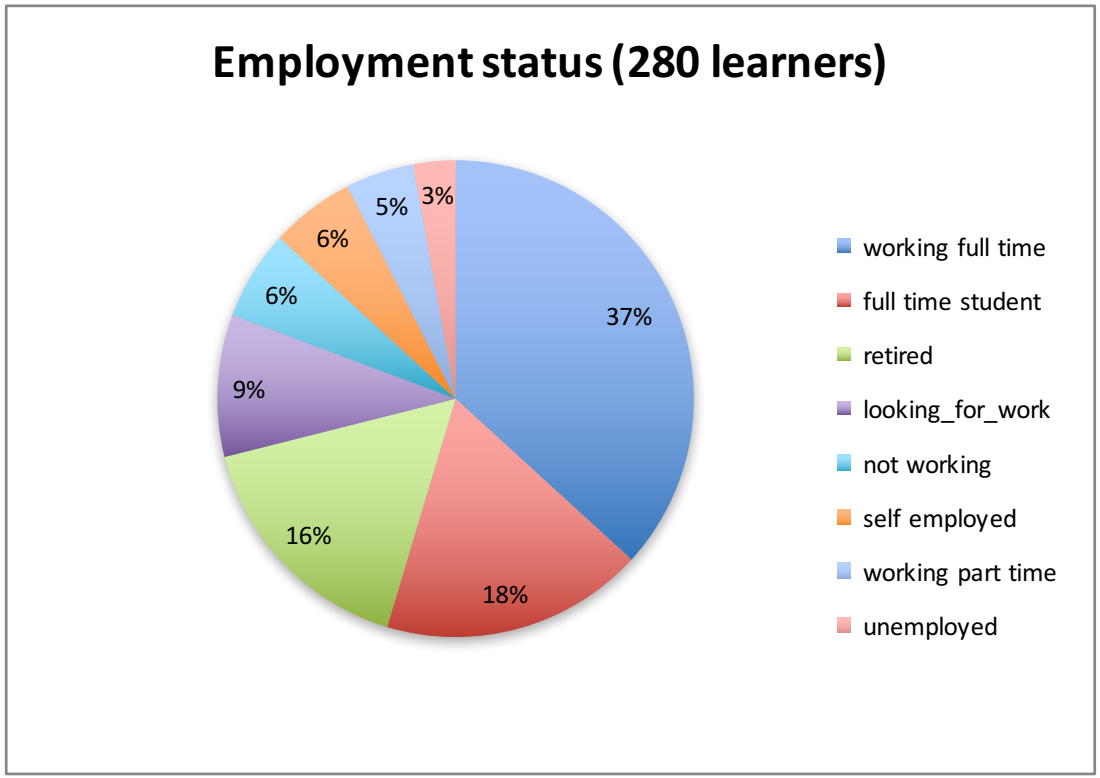
First run: Facts

- Started 6th March 2017
- Ran for 5 weeks
- More than 3000 people enrolled
- Expected small % to actually participate

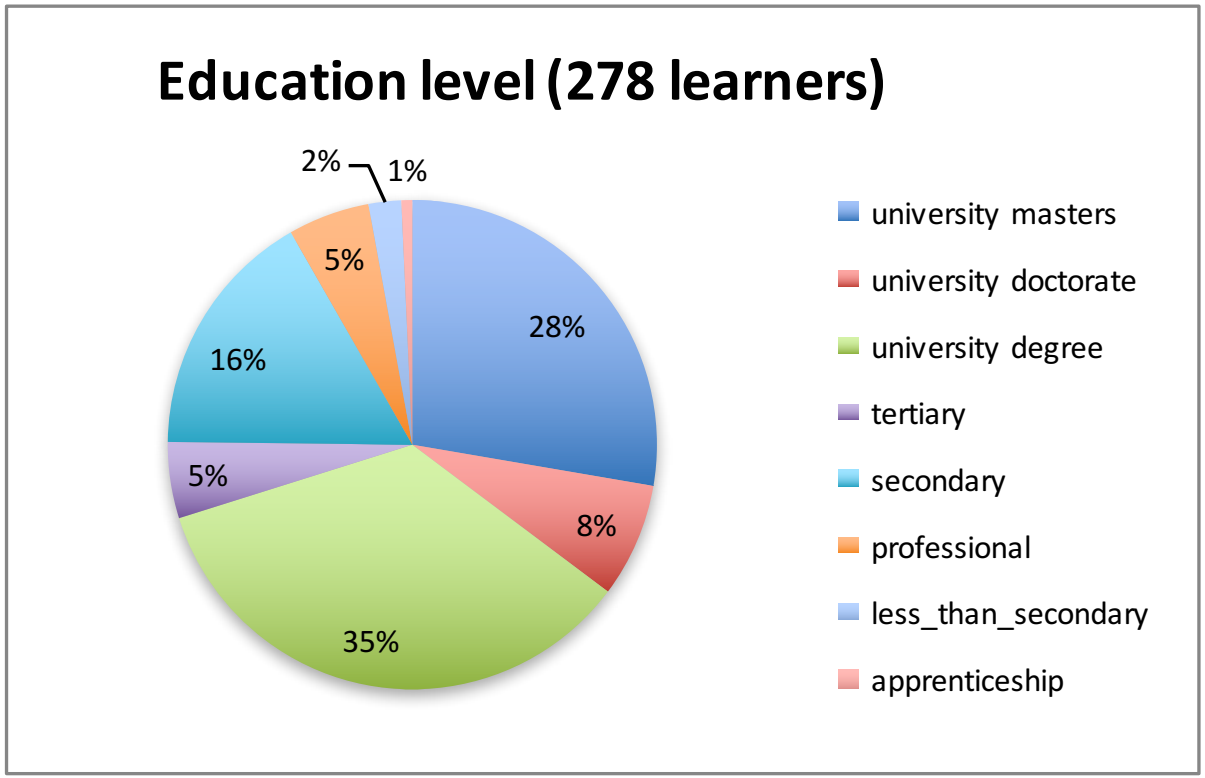
First run: Facts



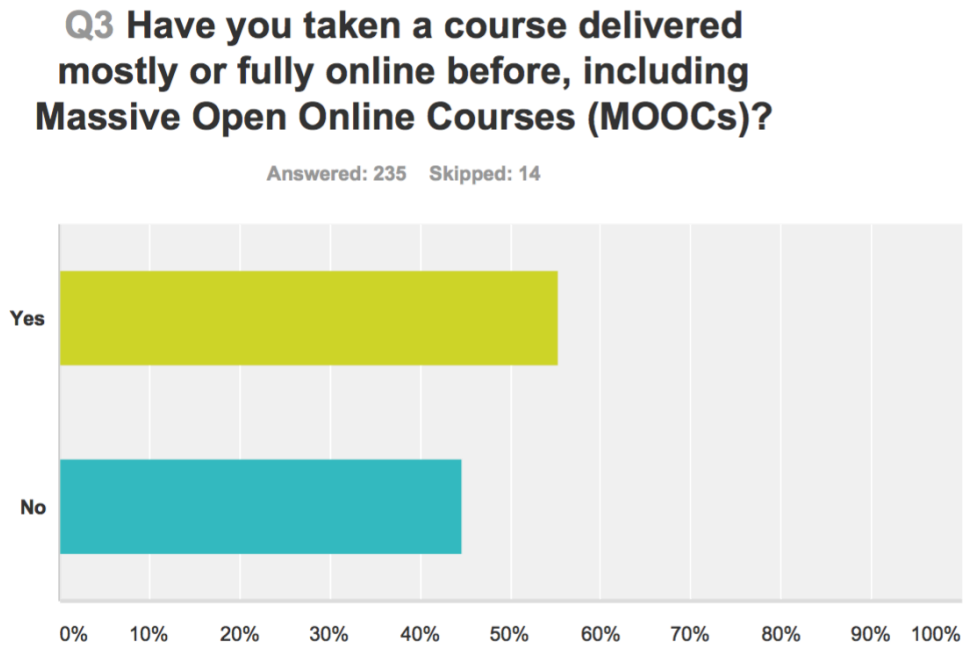
First run: Facts



First run: Facts



First run: Facts



First run: Facts

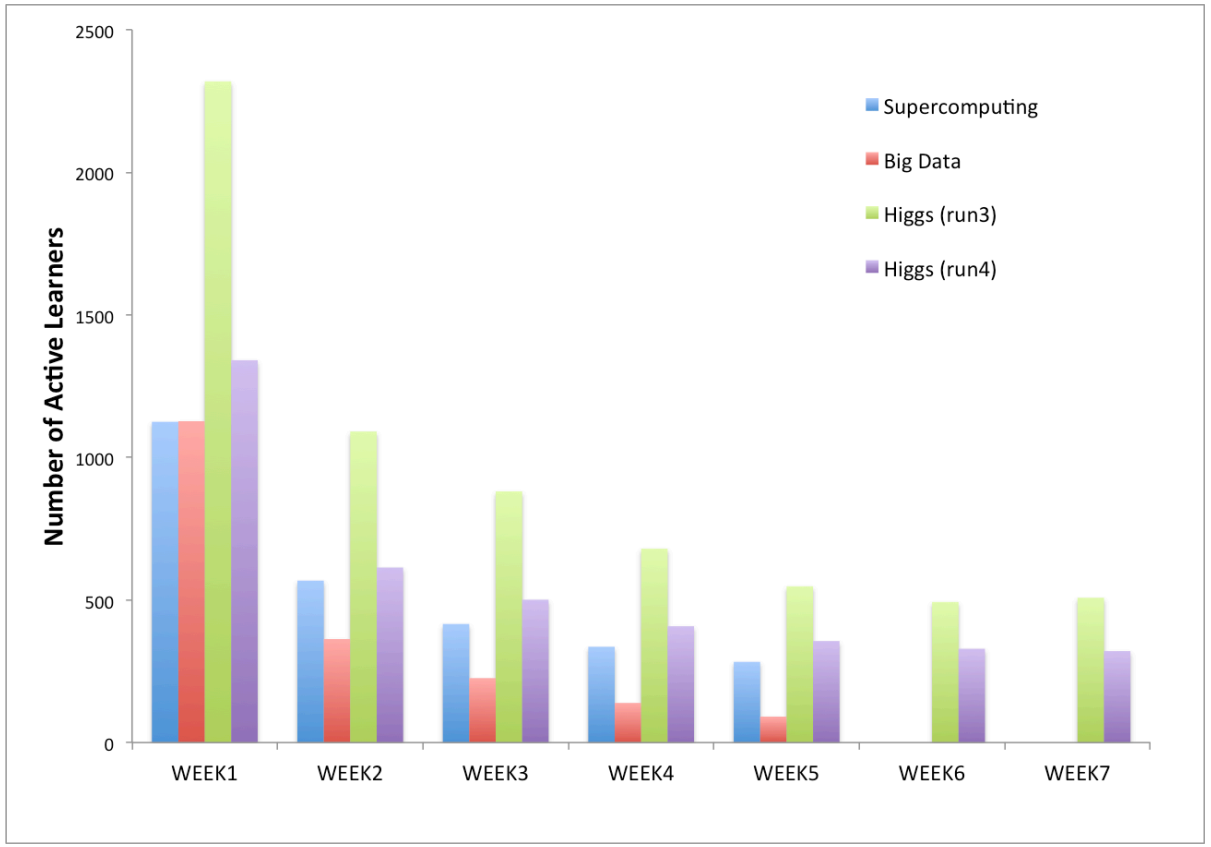
I found it whilst browsing FutureLearn website	**18.05%**	37
I followed a link in a FutureLearn email or newsletter	**18.54%**	38
I read about it in an online article	**4.39%**	9
I read about it in an offline article e.g. newspaper	**1.46%**	3
A friend recommended it to me	**35.61%**	73
I saw a social media message about it (e.g. Twitter, Facebook)	**3.90%**	8
I found it by using a search engine (e.g. Google, Bing)	**1.95%**	4
I followed a link on a blog	**1.95%**	4
I followed a link on a website about online courses	**2.93%**	6
I saw it advertised on the university's website	**8.29%**	17

First run: Results

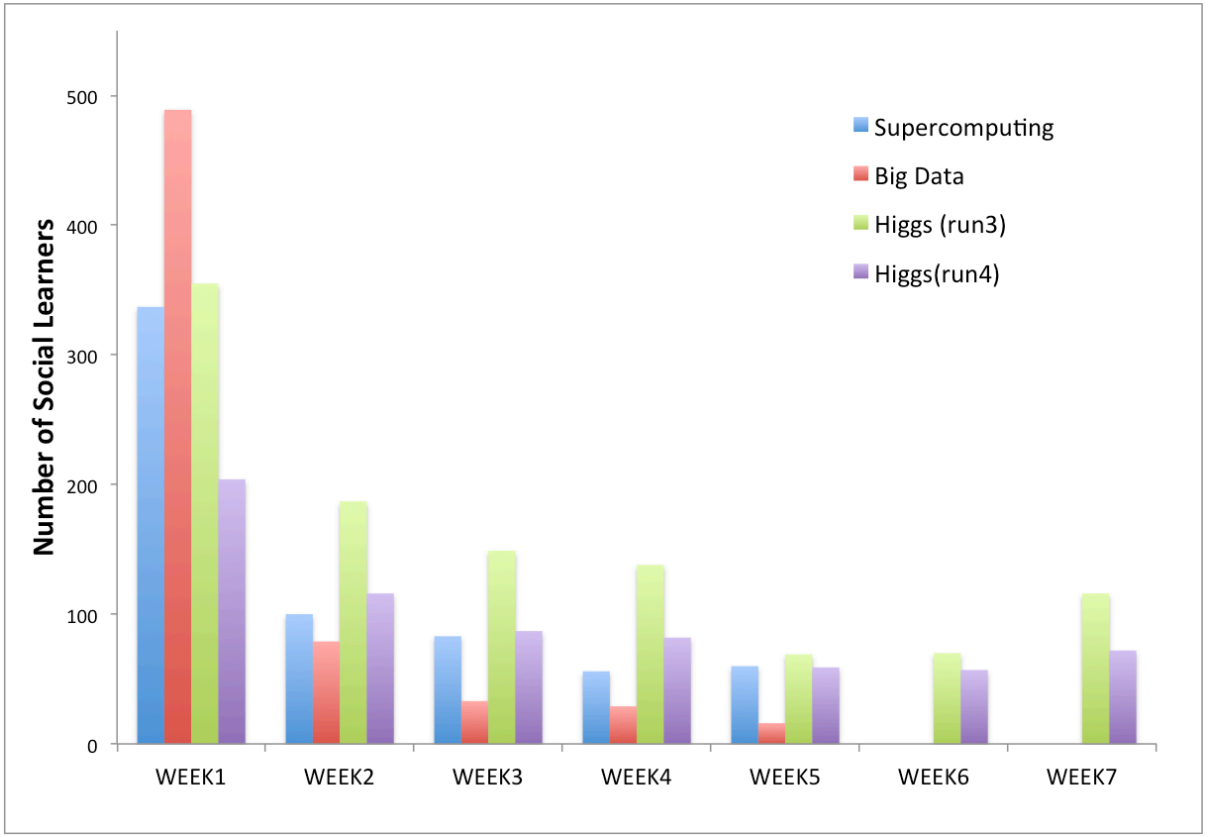
First run: Results

	Total
Steps visited	51,732
Steps completed	44,460
Comments posted	2,791

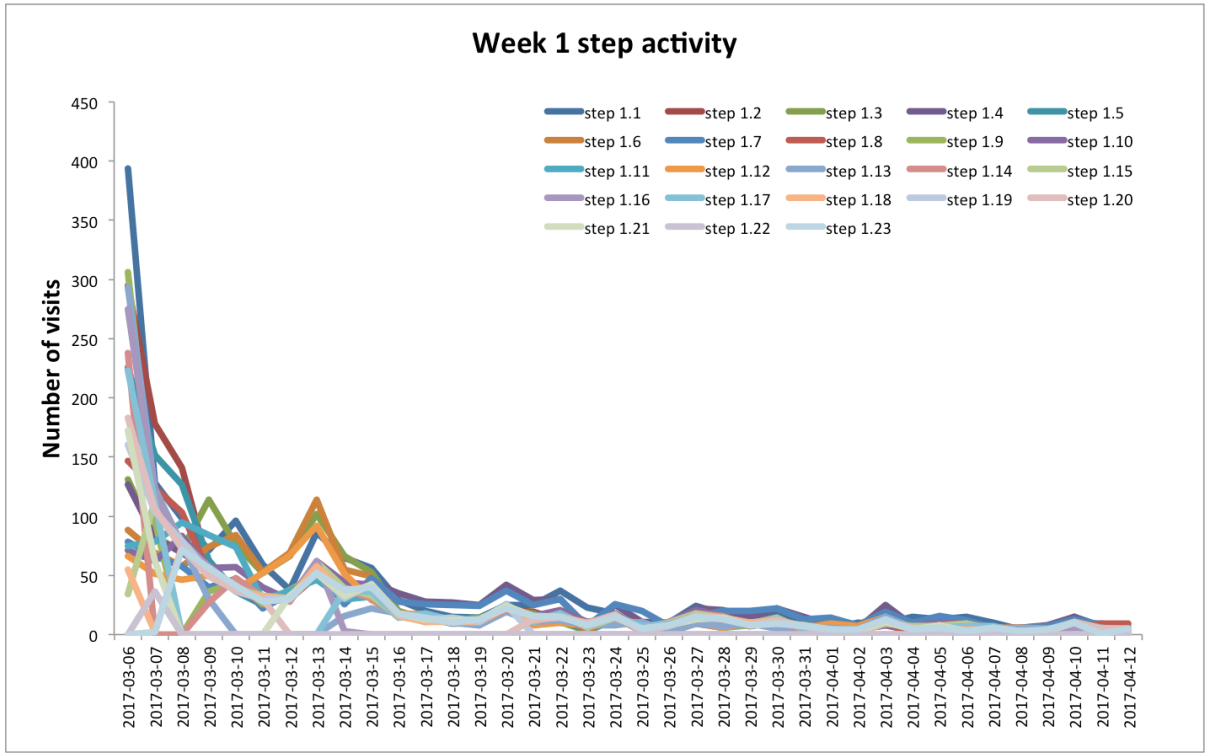
First run: Results



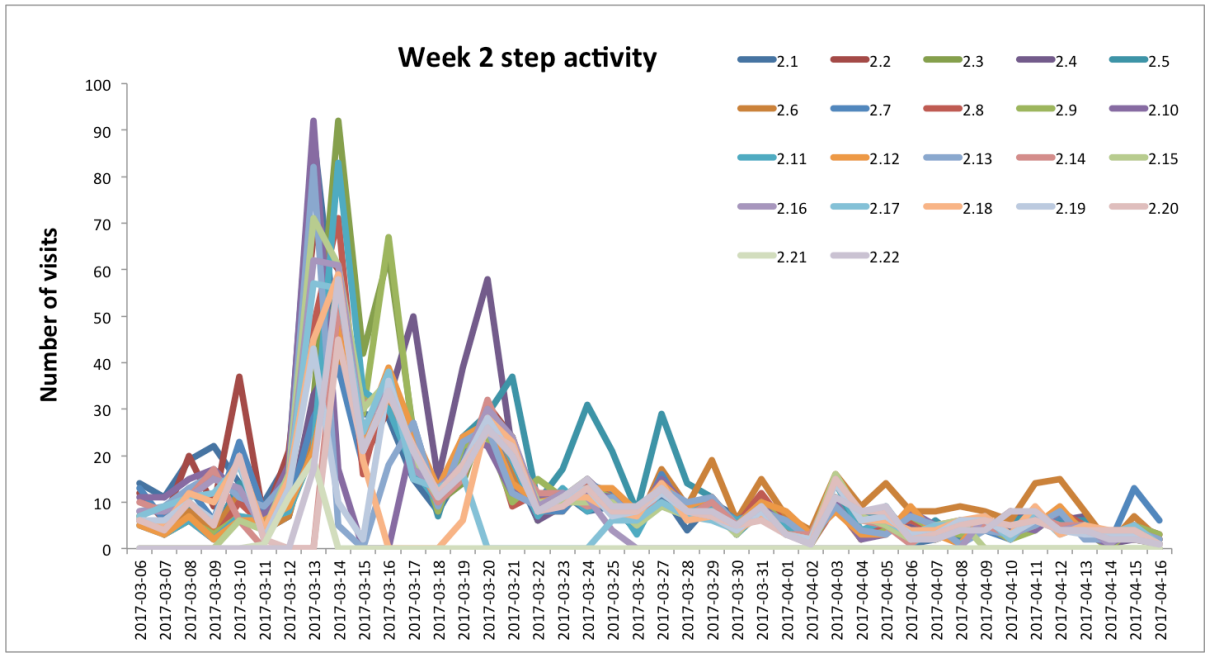
First run: Results



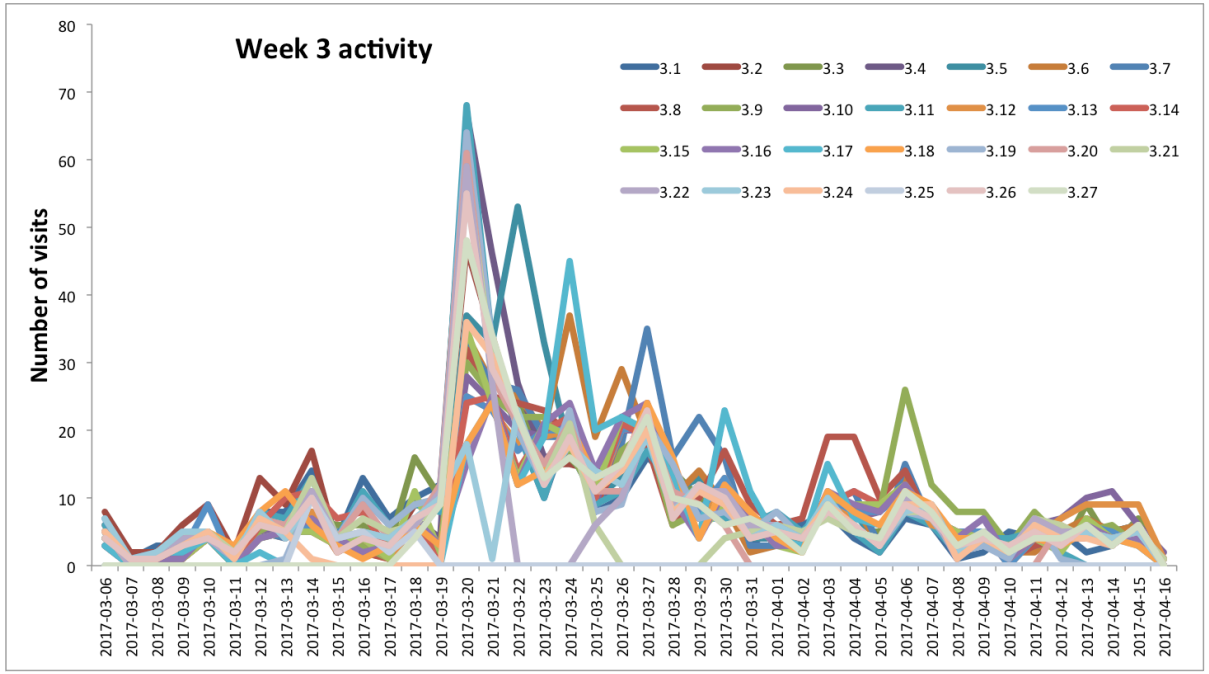
First run: Results



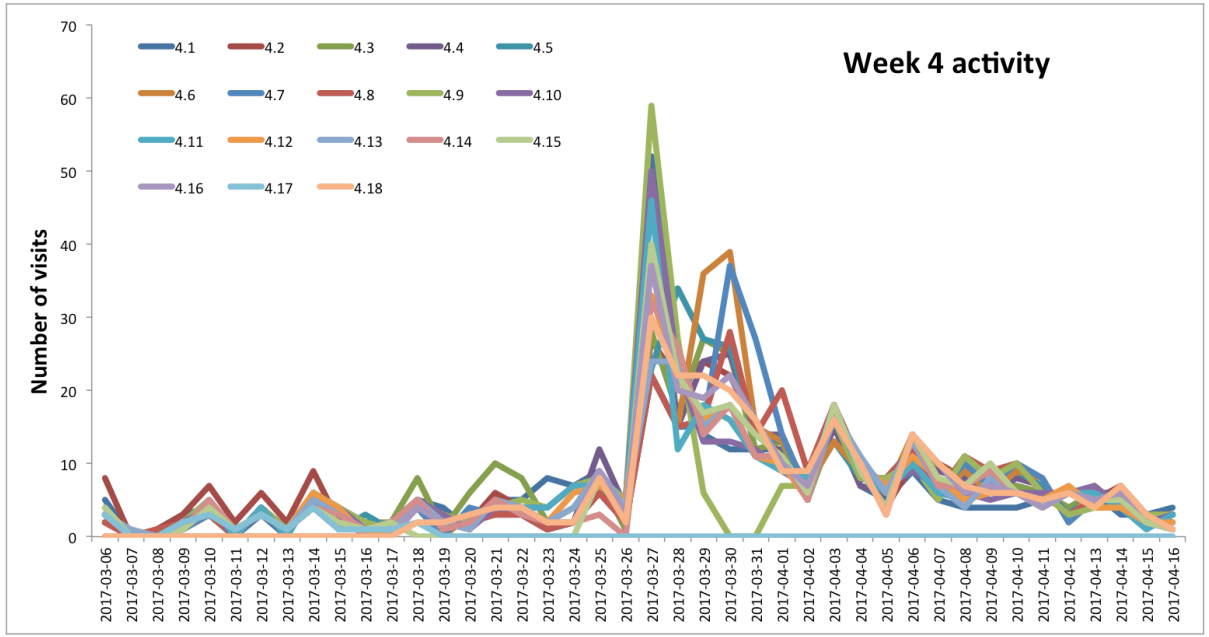
First run: Results



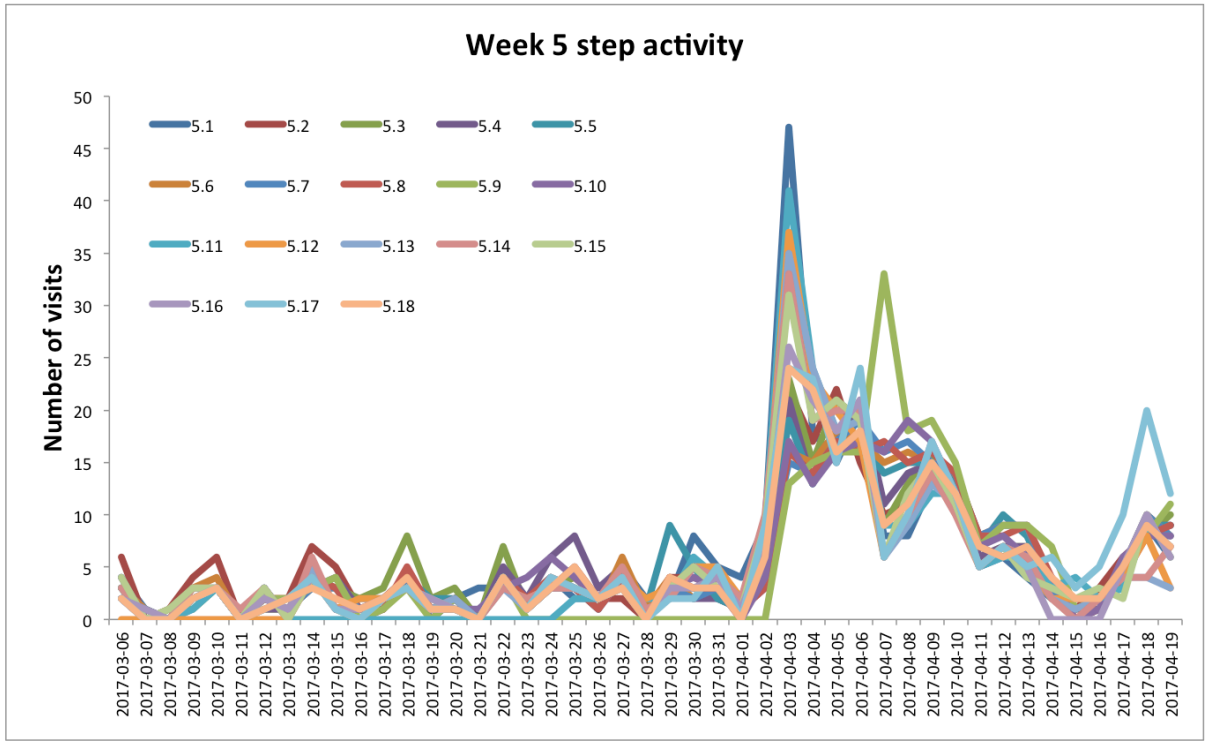
First run: Results



First run: Results



First run: Results



Survey Question 39:

What was your favourite part of the course, and why?

Learner: *Videos*

Learner: *The overall course was highly interesting. Well done*

“

Learner: *Probably the effort the educators put into the course after it started, i.e. the engagement with the learners by the educators. It really helped lift what would have been a good course to an excellent course. The videos and articles were also well prepared.*

First run: Results

- Social vs non-social learners
  - Most liked activity:

Taking part in a small Study Group	**0.00%** 0	**5.88%** 1	**11.76%** 2	**5.88%** 1	**5.88%** 1	**70.59%** 12	17	0.40

  - Most disliked activity:

Reading comments posted by other learners	**5.88%** 1	**0.00%** 0	**23.53%** 4	**35.29%** 6	**29.41%** 5	**5.88%** 1	17	0.88

Lessons Learned

- Worked out well:
  - Automatic subtitles on 3Play
  - Blue-screen
  - Multi-angle videos
  - Course run had minimal overhead
- Difficulties:
  - Audio of (some) videos
  - “Creative-commons, non-commercial”
  - Video editing takes time
  - Backup and reuse on Futurelearn not straightforward

Going Forward

- Next run scheduled:
  - 28 August 2017
  - >200 signed up as of 22 June 2017
  - https://www.futurelearn.com/courses/supercomputing
- Planned changes
  - Re-record some videos with improved audio
  - Reorganise some material
  - Continue to improve in-house AV abilities for agility
  - Continue to look for new use-cases

Going Forward

- Next run scheduled:
  - 28 August 2017
  - >200 signed up as of 22 June 2017
  - https://www.futurelearn.com/courses/supercomputing
- Planned changes
  - Re-record some videos with improved audio
  - Reorganise some material
  - Continue to improve in-house AV abilities for agility
  - **Continue to look for new use-cases**

THANK YOU FOR YOUR ATTENTION

www.prace-ri.eu

About the presenter

- Emmanouil (Manos) Farsarakis

  Applications Consultant

  EPCC, The University of Edinburgh

  e.farsarakis@epcc.ed.ac.uk

- More information:
  - https://www.futurelearn.com/courses/supercomputing
  - https://www.epcc.ed.ac.uk/online-learning
  - https://www.epcc.ed.ac.uk/msc

Exploiting and Enabling HPC Technologies for Big Data, Deep Learning, and Clouds

Talk at PRACE Workshop, ISC '17

by

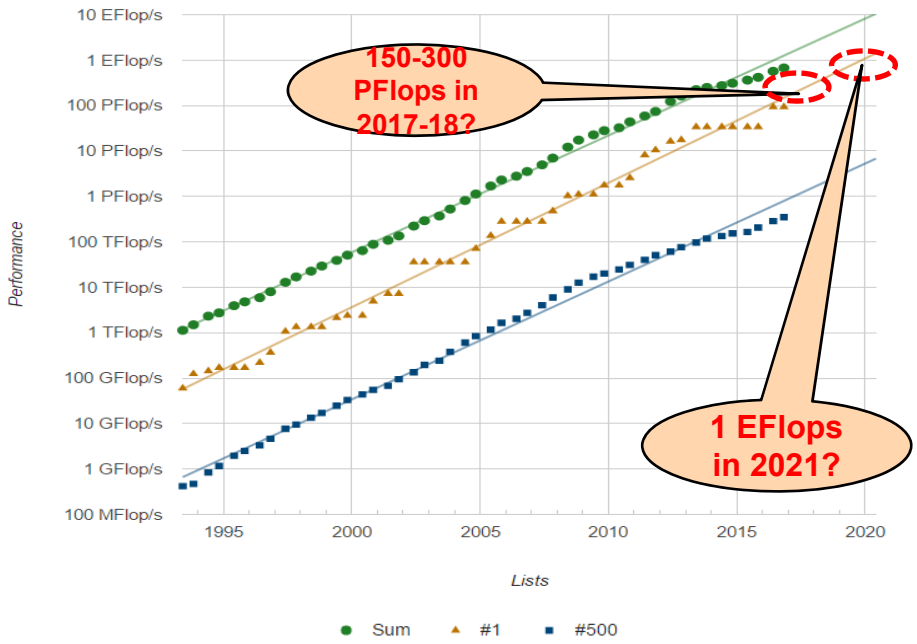
Dhabaleswar K. (DK) Panda

The Ohio State University

E-mail: panda@cse.ohio-state.edu

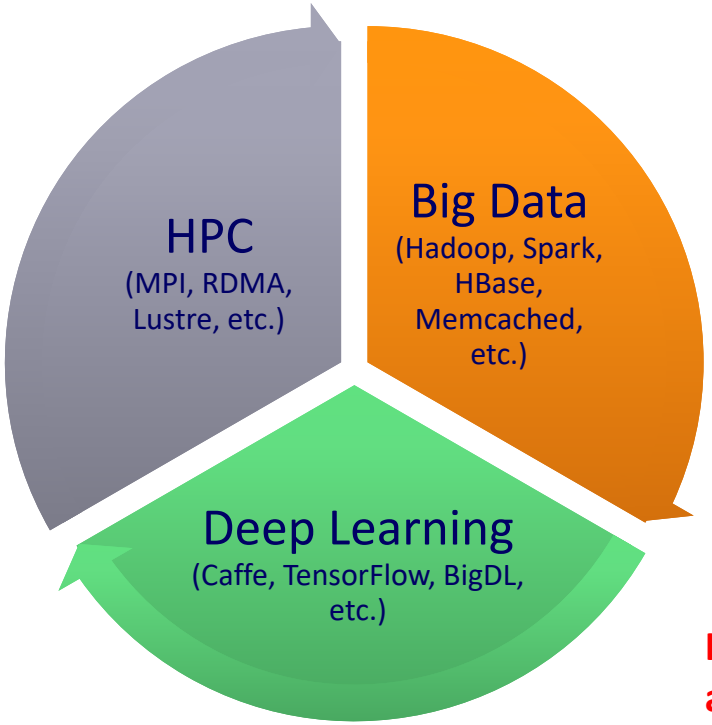
http://www.cse.ohio-state.edu/~panda

High-End Computing (HEC): Towards Exascale-Era



Expected to have an ExaFlop system in 2021!

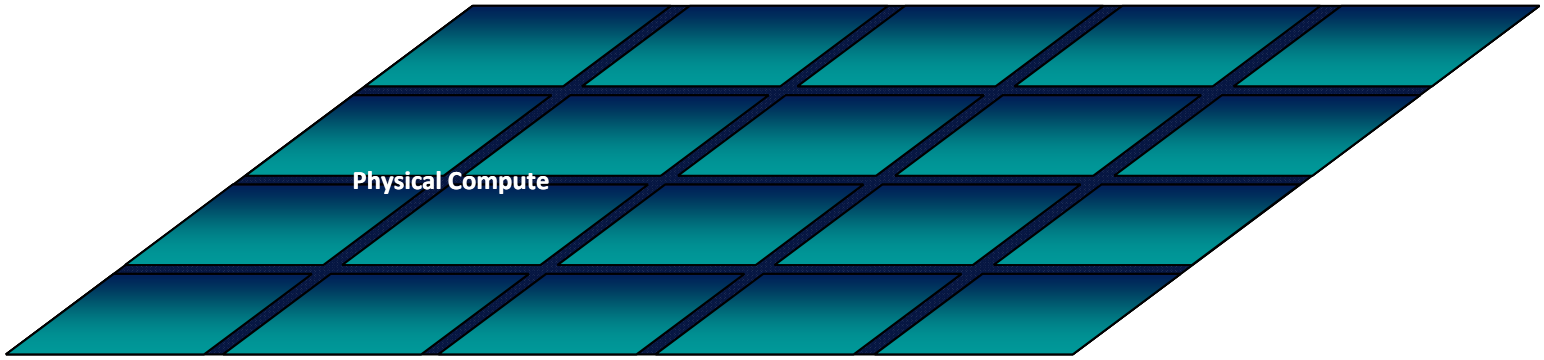
Increasing Usage of HPC, Big Data and Deep Learning



Convergence of HPC, Big Data, and Deep Learning!

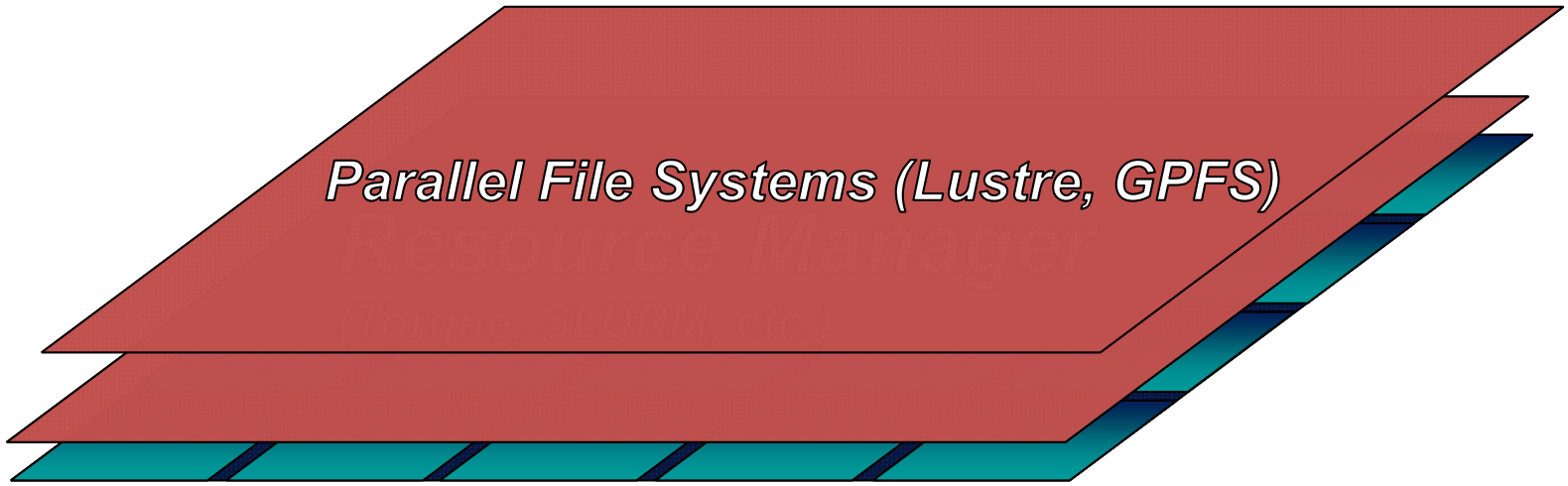
Increasing Need to Run these applications on the Cloud!!

Can We Run Big Data and Deep Learning Jobs on Existing HPC Infrastructure?

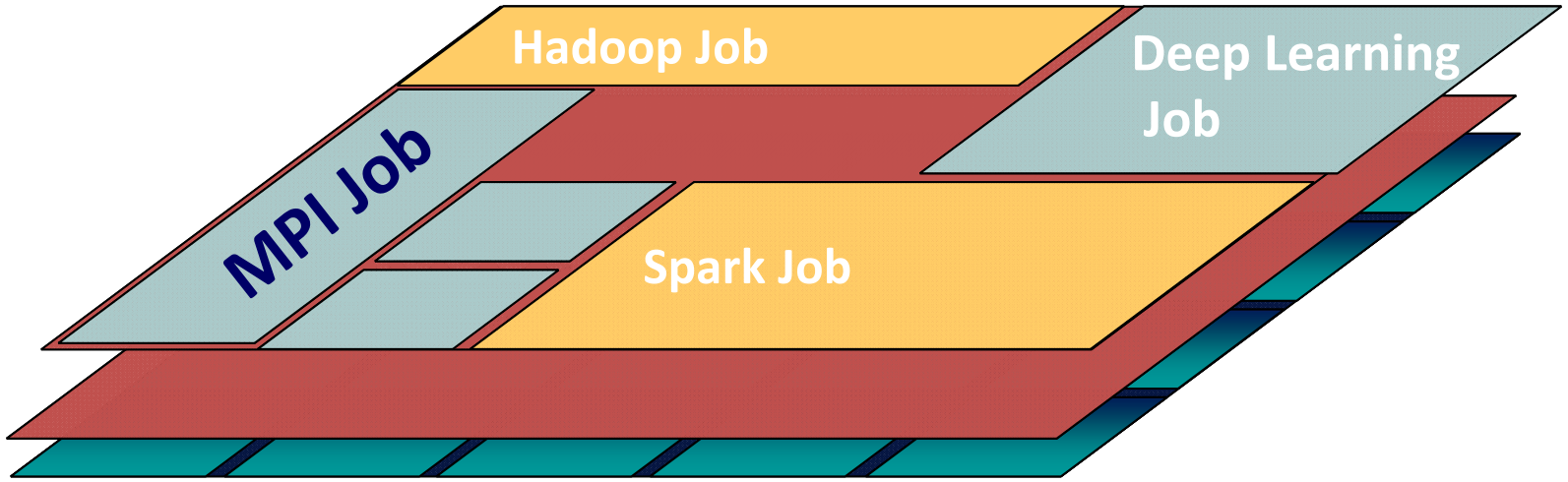


Can We Run Big Data and Deep Learning Jobs on Existing HPC Infrastructure?

Can We Run Big Data and Deep Learning Jobs on Existing HPC Infrastructure?



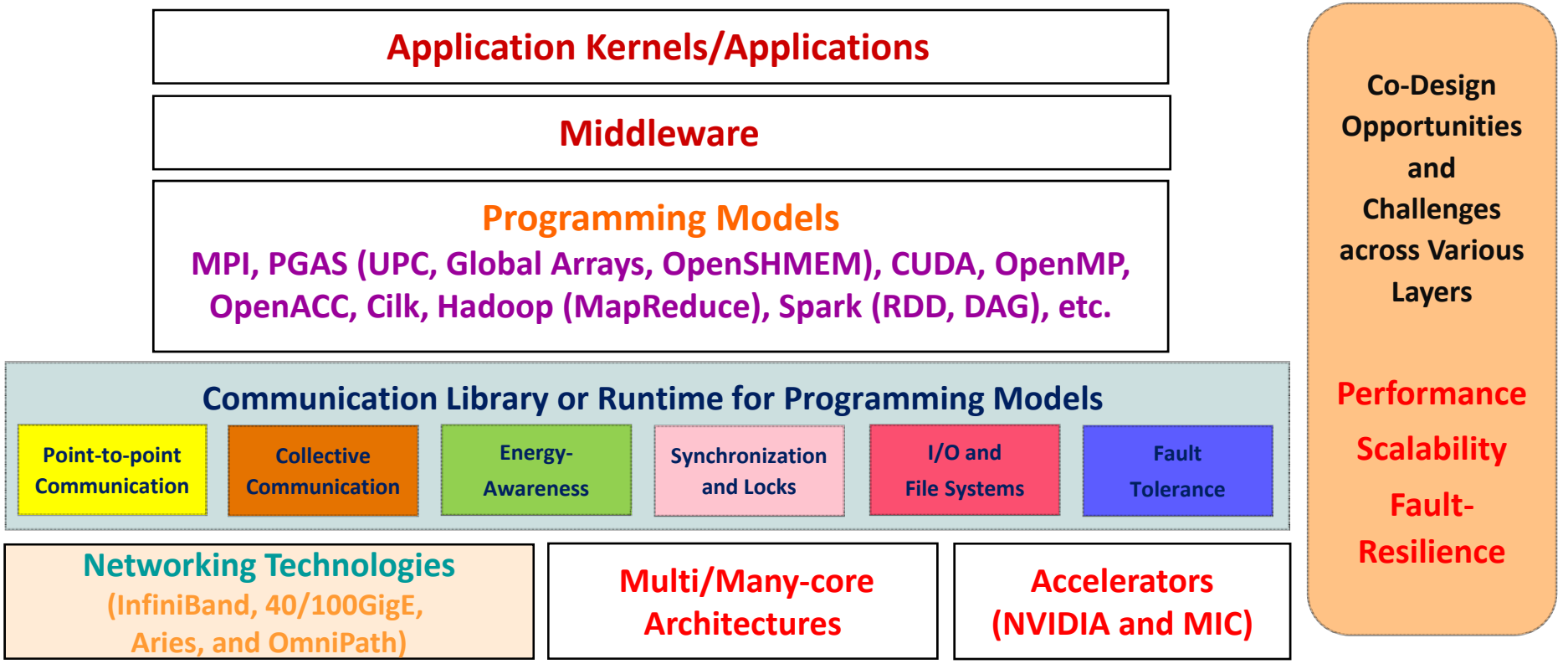
Can We Run Big Data and Deep Learning Jobs on Existing HPC Infrastructure?



HPC, Big Data, Deep Learning, and Cloud

- Traditional HPC
  - Message Passing Interface (MPI), including MPI + OpenMP
  - Support for PGAS and MPI + PGAS (OpenSHMEM, UPC)
  - Exploiting Accelerators
- Big Data/Enterprise/Commercial Computing
  - Spark and Hadoop (HDFS, HBase, MapReduce)
  - Memcached is also used for Web 2.0
- Deep Learning
  - Caffe, CNTK, TensorFlow, and many more
- Cloud for HPC and BigData
  - Virtualization with SR-IOV and Containers

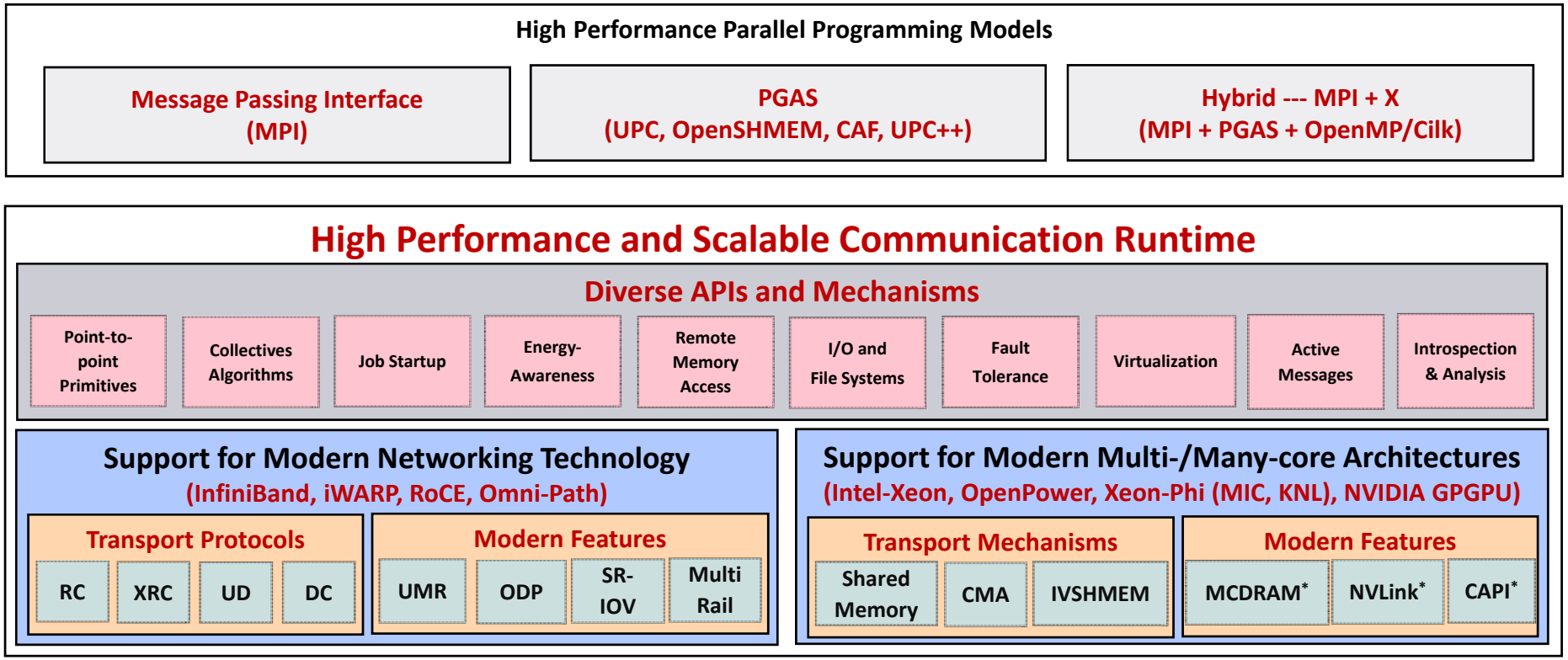
Designing Communication Middleware for Multi-Petaflop and Exaflop Systems: Challenges



Overview of the MVAPICH2 Project

- High Performance open-source MPI Library for InfiniBand, Omni-Path, Ethernet/iWARP, and RDMA over Converged Ethernet (RoCE)
  - MVAPICH (MPI-1), MVAPICH2 (MPI-2.2 and MPI-3.0), Started in 2001, First version available in 2002
  - MVAPICH2-X (MPI + PGAS), Available since 2011
  - Support for GPGPUs (MVAPICH2-GDR) and MIC (MVAPICH2-MIC), Available since 2014
  - Support for Virtualization (MVAPICH2-Virt), Available since 2015
  - Support for Energy-Awareness (MVAPICH2-EA), Available since 2015
  - Support for InfiniBand Network Analysis and Monitoring (OSU INAM) since 2015
  - **Used by more than 2,775 organizations in 85 countries**
  - **More than 420,000 (> 0.4 million) downloads from the OSU site directly**
  - Empowering many TOP500 clusters (June '17 ranking)
    - 1st, 10,649,600-core (Sunway TaihuLight) at National Supercomputing Center in Wuxi, China
    - 15th, 241,108-core (Pleiades) at NASA
    - 20th, 462,462-core (Stampede) at TACC
    - 44th, 74,520-core (Tsubame 2.5) at Tokyo Institute of Technology
  - Available with software stacks of many vendors and Linux Distros (RedHat and SuSE)
  - **http://mvapich.cse.ohio-state.edu**
- Empowering Top500 systems for over a decade
  - System-X from Virginia Tech (3rd in Nov 2003, 2,200 processors, 12.25 TFlops) ->
  - Stampede at TACC (12th in Jun'16, 462,462 cores, 5.168 Plops)

Architecture of MVAPICH2 Software Family

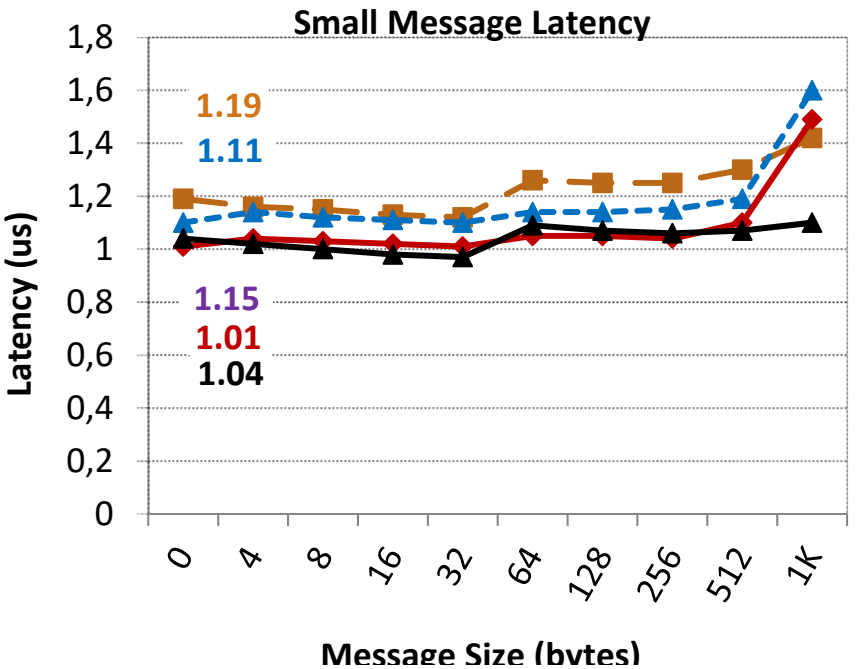


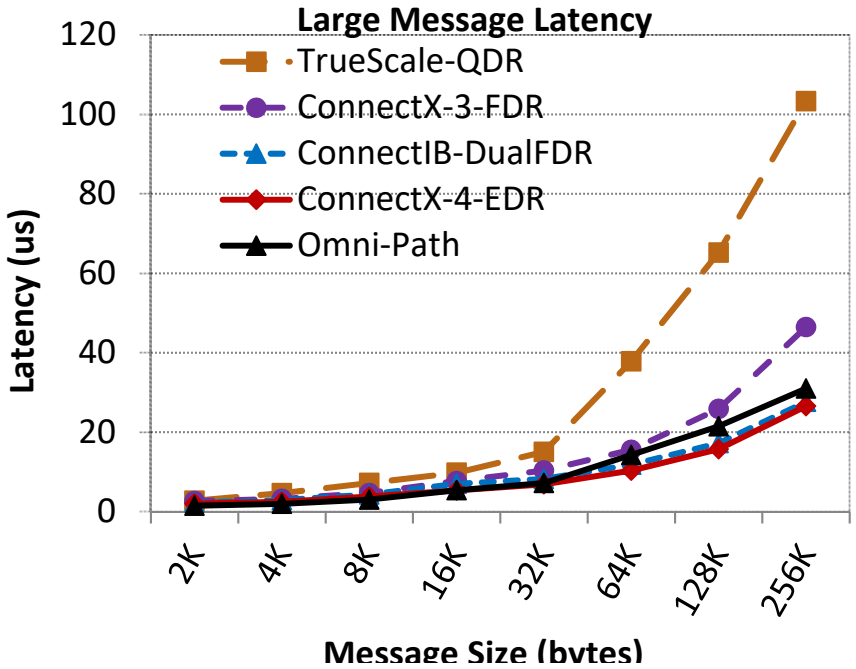
* Upcoming

MVAPICH2 Software Family

High-Performance Parallel Programming Libraries	
MVAPICH2	Support for InfiniBand, Omni-Path, Ethernet/iWARP, and RoCE
MVAPICH2-X	Advanced MPI features, OSU INAM, PGAS (OpenSHMEM, UPC, UPC++, and CAF), and MPI+PGAS programming models with unified communication runtime
MVAPICH2-GDR	Optimized MPI for clusters with NVIDIA GPUs
MVAPICH2-Virt	High-performance and scalable MPI for hypervisor and container based HPC cloud
MVAPICH2-EA	Energy aware and High-performance MPI
MVAPICH2-MIC	Optimized MPI for clusters with Intel KNC
Microbenchmarks	
OMB	Microbenchmarks suite to evaluate MPI and PGAS (OpenSHMEM, UPC, and UPC++) libraries for CPUs and GPUs
Tools	
OSU INAM	Network monitoring, profiling, and analysis for clusters with MPI and scheduler integration
OEMT	Utility to measure the energy consumption of MPI applications

One-way Latency: MPI over IB with MVAPICH2





TrueScale-QDR - 3.1 GHz Deca-core (Haswell) Intel PCI Gen3 with IB switch

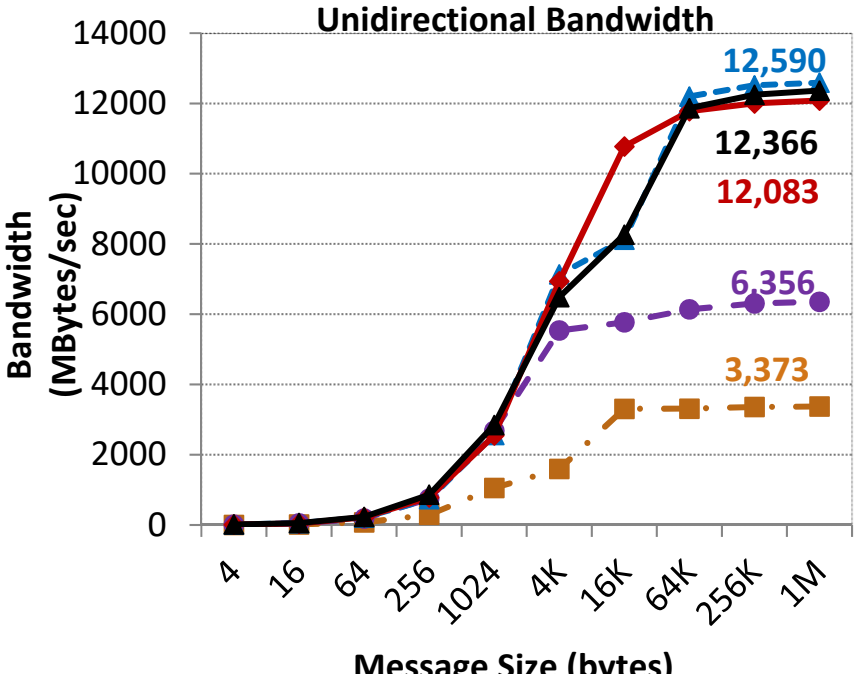
ConnectX-3-FDR - 2.8 GHz Deca-core (IvyBridge) Intel PCI Gen3 with IB switch

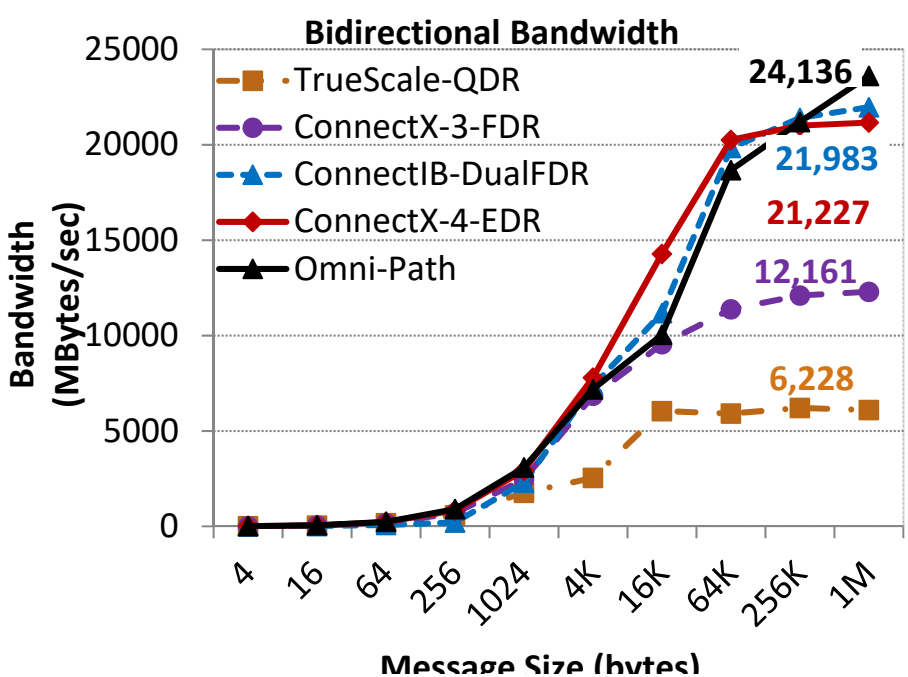
ConnectIB-Dual FDR - 3.1 GHz Deca-core (Haswell) Intel PCI Gen3 with IB switch

ConnectX-4-EDR - 3.1 GHz Deca-core (Haswell) Intel PCI Gen3 with IB Switch

Omni-Path - 3.1 GHz Deca-core (Haswell) Intel PCI Gen3 with Omni-Path switch

Bandwidth: MPI over IB with MVAPICH2





TrueScale-QDR - 3.1 GHz Deca-core (Haswell) Intel PCI Gen3 with IB switch

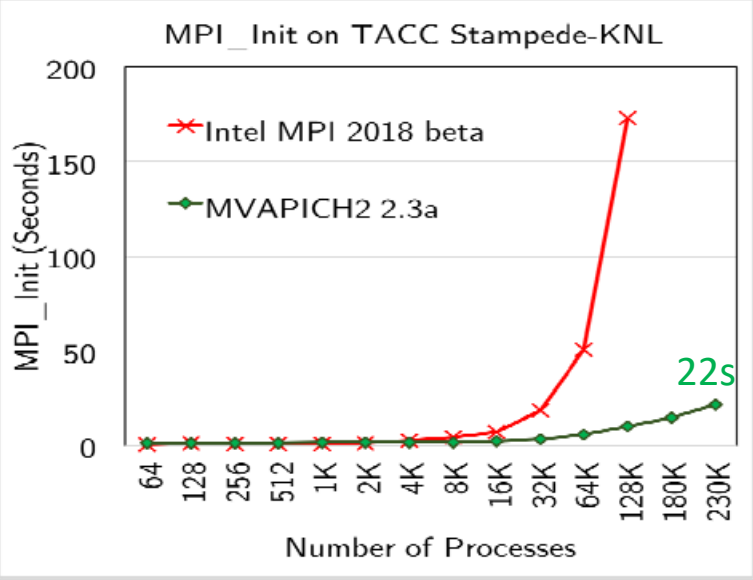
ConnectX-3-FDR - 2.8 GHz Deca-core (IvyBridge) Intel PCI Gen3 with IB switch

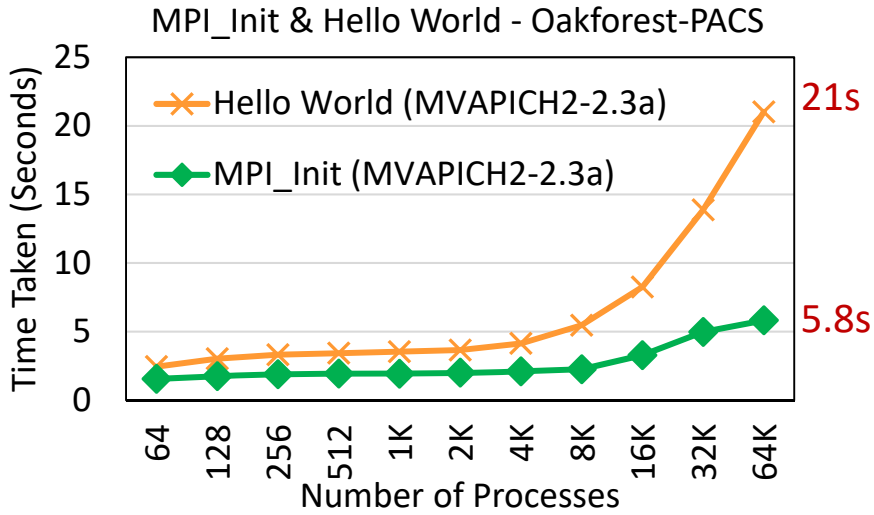
ConnectIB-Dual FDR - 3.1 GHz Deca-core (Haswell) Intel PCI Gen3 with IB switch

ConnectX-4-EDR - 3.1 GHz Deca-core (Haswell) Intel PCI Gen3 IB switch

Omni-Path - 3.1 GHz Deca-core (Haswell) Intel PCI Gen3 with Omni-Path switch

Startup Performance on KNL + Omni-Path

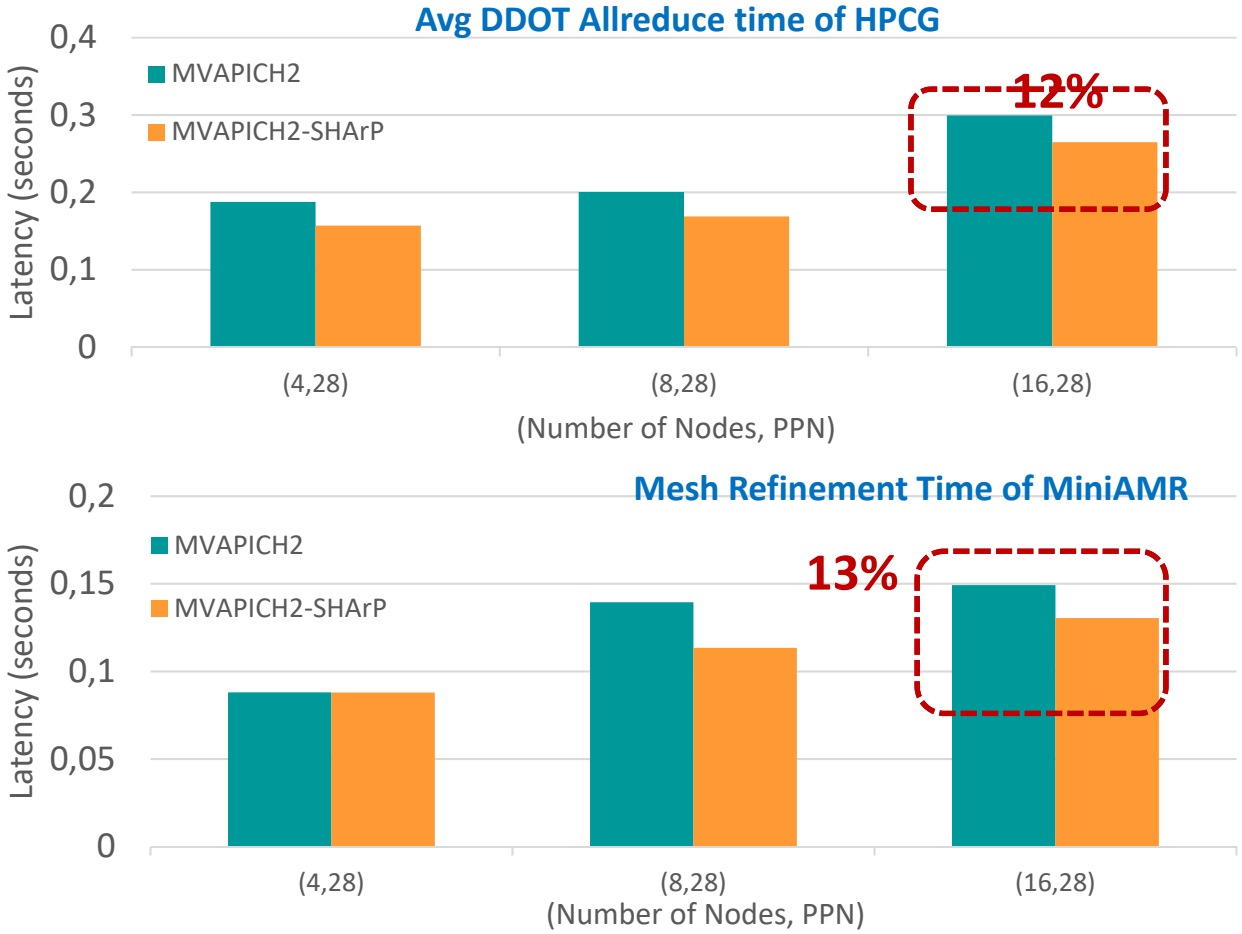




- MPI_Init takes 22 seconds on 229,376 processes on 3,584 KNL nodes (Stampede2 – Full scale)
- 8.8 times faster than Intel MPI at 128K processes (Courtesy: TACC)
- At 64K processes, MPI_Init and Hello World takes 5.8s and 21s respectively (Oakforest-PACS)
- All numbers reported with 64 processes per node

New designs available in MVAPICH2-2.3a and as patch for SLURM-15.08.8 and SLURM-16.05.1

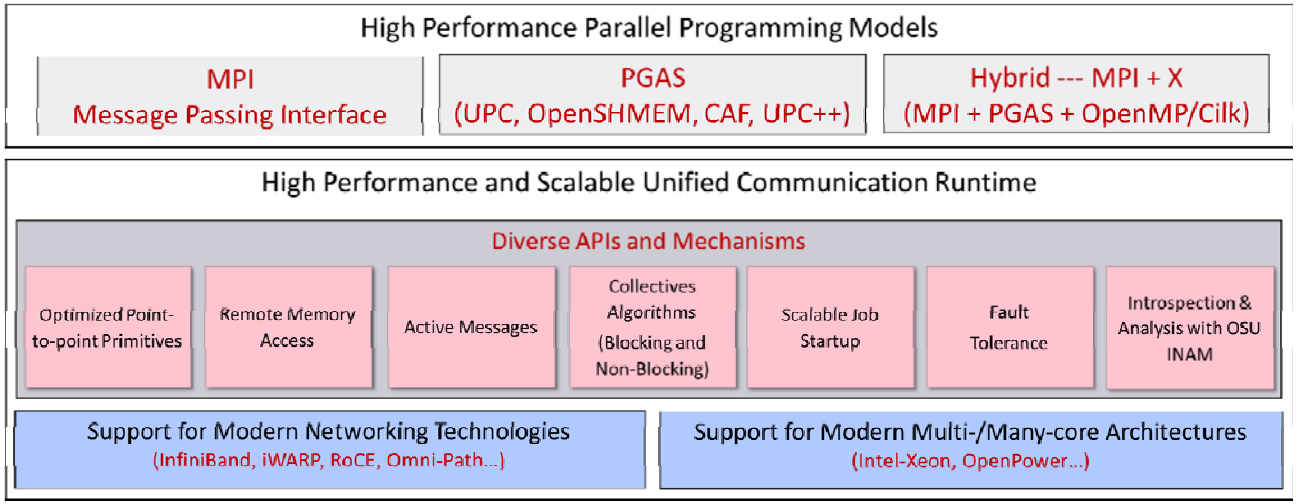
Advanced Allreduce Collective Designs Using SHArP



SHArP Support is available in MVAPICH2 2.3a

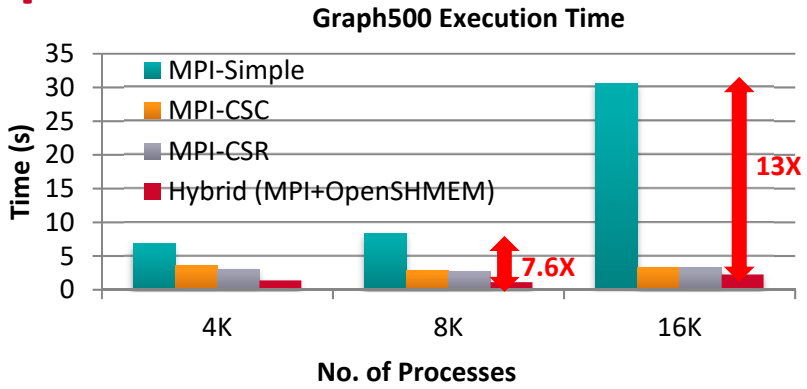
M. Bayatpour, S. Chakraborty, H. Subramoni, X. Lu, and D. K. Panda, Scalable Reduction Collectives with Data Partitioning-based Multi-Leader Design, Supercomputing '16 (Accepted).

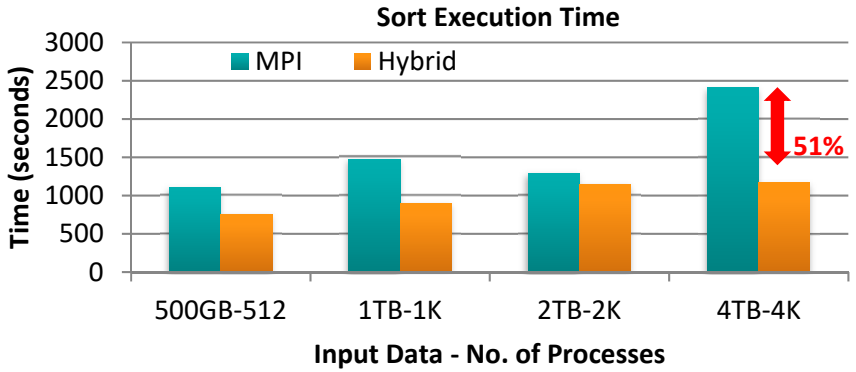
MVAPICH2-X for Hybrid MPI + PGAS Applications



- Current Model – Separate Runtimes for OpenSHMEM/UPC/UPC++/CAF and MPI
  - Possible deadlock if both runtimes are not progressed
  - Consumes more network resource
- Unified communication runtime for MPI, UPC, UPC++, OpenSHMEM, CAF
  - Available with since 2012 (starting with MVAPICH2-X 1.9)
  - http://mvapich.cse.ohio-state.edu

Application Level Performance with Graph500 and Sort





- Performance of Hybrid (MPI+ OpenSHMEM) Graph500 Design
  - 8,192 processes
    - **2.4X** improvement over MPI-CSR
    - **7.6X** improvement over MPI-Simple
  - 16,384 processes
    - **1.5X** improvement over MPI-CSR
    - **13X** improvement over MPI-Simple

- Performance of Hybrid (MPI+OpenSHMEM) Sort Application
  - 4,096 processes, 4 TB Input Size
    - MPI – 2408 sec; 0.16 TB/min
    - Hybrid – 1172 sec; 0.36 TB/min
    - **51%** improvement over MPI-design

J. Jose, K. Kandalla, S. Potluri, J. Zhang and D. K. Panda, Optimizing Collective Communication in OpenSHMEM, Int'l Conference on Partitioned Global Address Space Programming Models (PGAS '13), October 2013.

J. Jose, S. Potluri, K. Tomko and D. K. Panda, Designing Scalable Graph500 Benchmark with Hybrid MPI+OpenSHMEM Programming Models, International Supercomputing Conference (ISC'13), June 2013

J. Jose, K. Kandalla, M. Luo and D. K. Panda, Supporting Hybrid MPI and OpenSHMEM over InfiniBand: Design and Performance Evaluation, Int'l Conference on Parallel Processing (ICPP '12), September 2012

GPU-Aware (CUDA-Aware) MPI Library: MVAPICH2-GPU

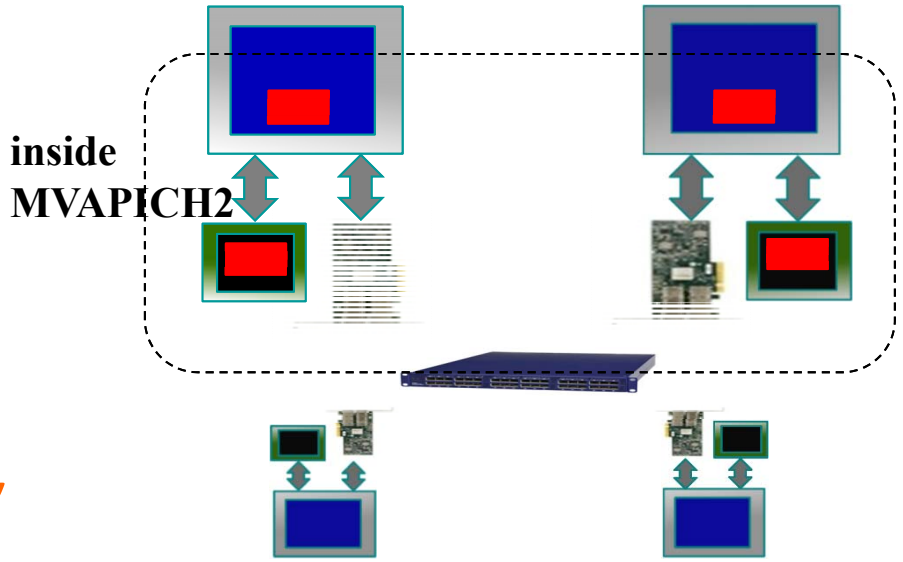
- Standard MPI interfaces used for unified data movement
- Takes advantage of Unified Virtual Addressing (>= CUDA 4.0)
- Overlaps data movement from GPU with RDMA transfers

At Sender:

```
MPI_Send(s_devbuf, size, …);
```

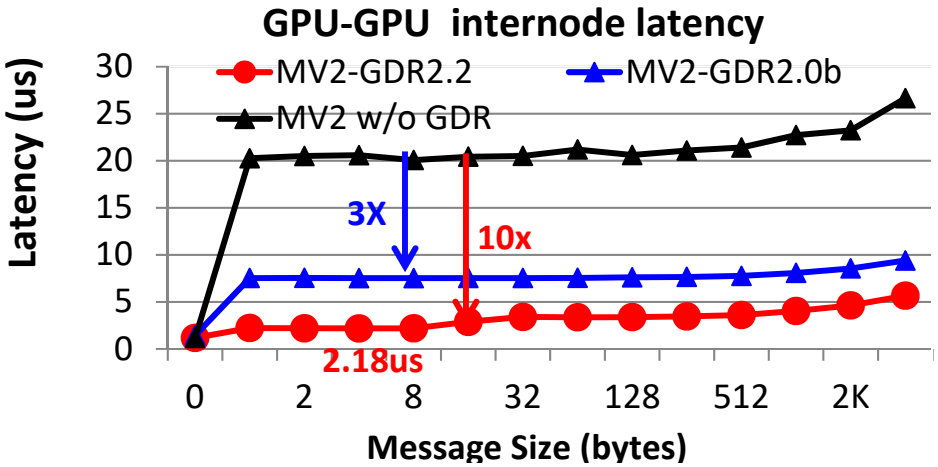
At Receiver:

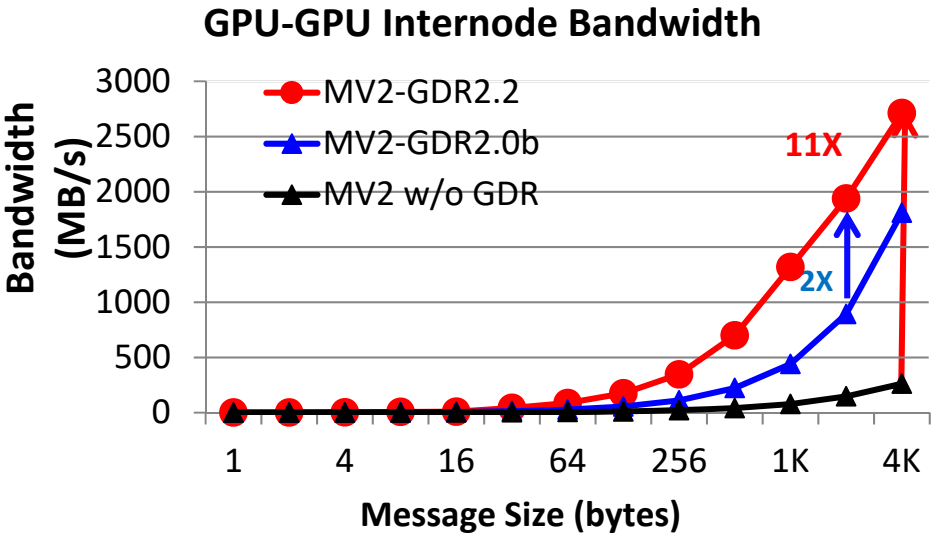
```
MPI_Recv(r_devbuf, size, …);
```

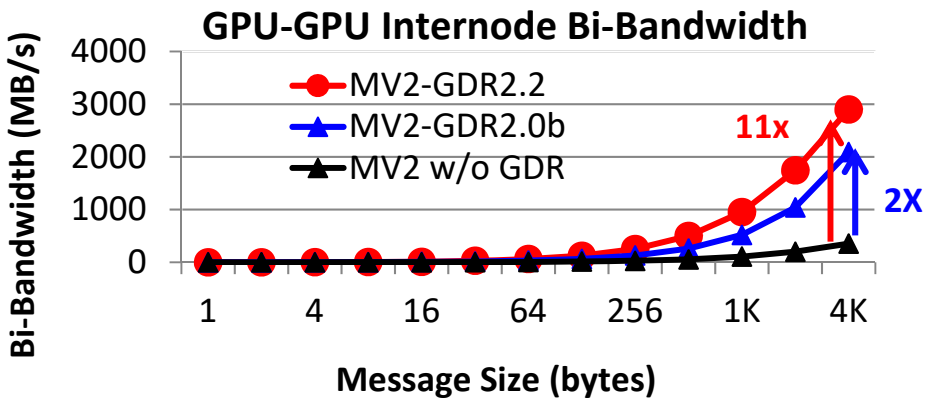


High Performance and High Productivity

Performance of MVAPICH2-GPU with GPU-Direct RDMA (GDR)

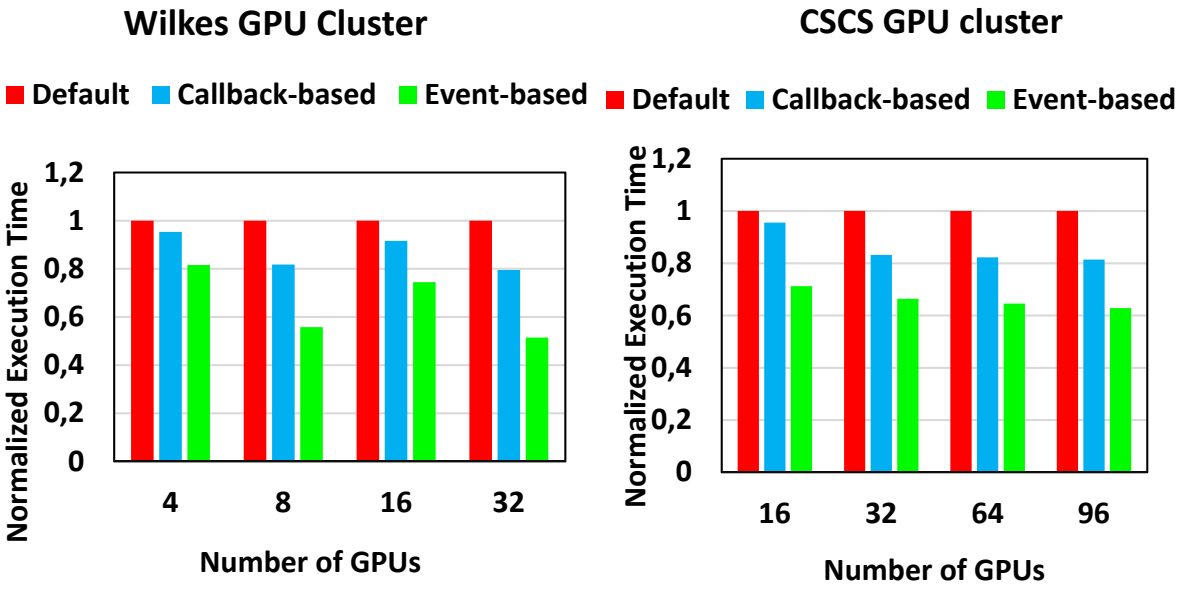


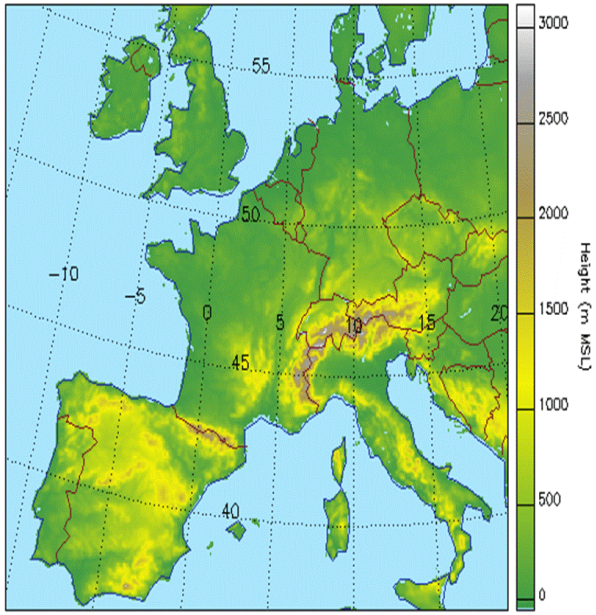




MVAPICH2-GDR-2.2
Intel Ivy Bridge (E5-2680 v2) node - 20 cores
NVIDIA Tesla K40c GPU
Mellanox Connect-X4 EDR HCA
CUDA 8.0
Mellanox OFED 3.0 with GPU-Direct-RDMA

Application-Level Evaluation (Cosmo) and Weather Forecasting in Switzerland





- 2X improvement on 32 GPUs nodes
- 30% improvement on 96 GPU nodes (8 GPUs/node)

Cosmo model: http://www2.cosmo-model.org/content/tasks/operational/meteoSwiss/

On-going collaboration with CSCS and MeteoSwiss (Switzerland) in co-designing MV2-GDR and Cosmo Application

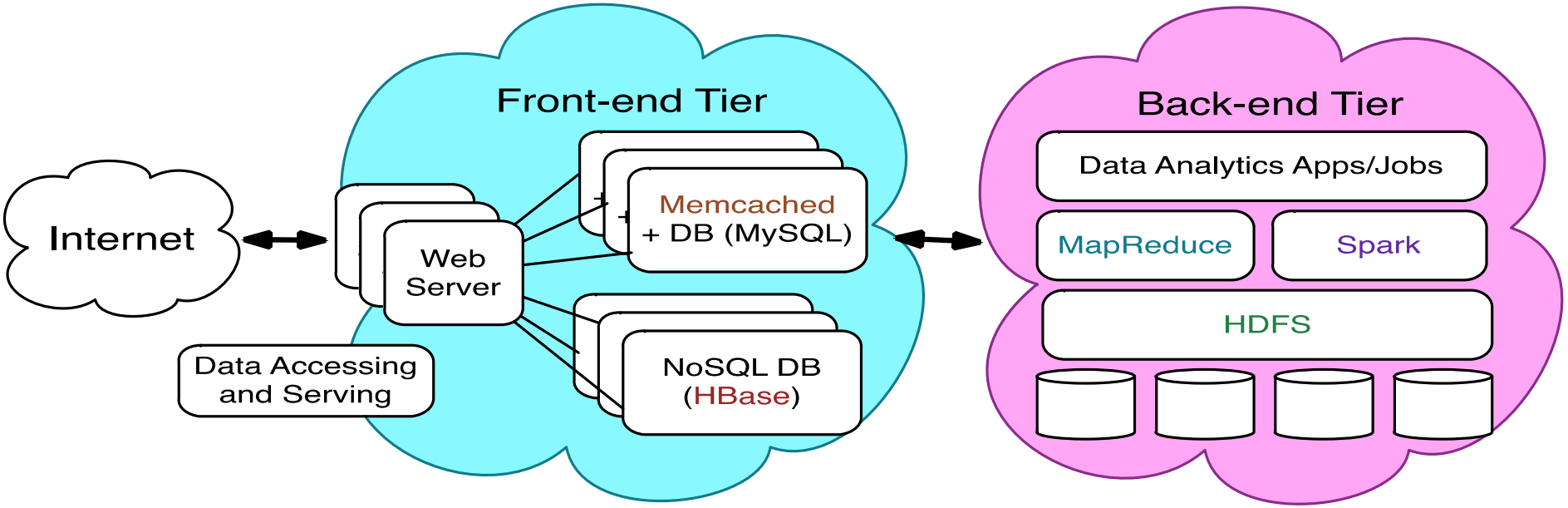
C. Chu, K. Hamidouche, A. Venkatesh, D. Banerjee , H. Subramoni, and D. K. Panda, Exploiting Maximal Overlap for Non-Contiguous Data Movement Processing on Modern GPU-enabled Systems, IPDPS'16

HPC, Big Data, Deep Learning, and Cloud

- Traditional HPC
  - Message Passing Interface (MPI), including MPI + OpenMP
  - Support for PGAS and MPI + PGAS (OpenSHMEM, UPC)
  - Exploiting Accelerators
- Big Data/Enterprise/Commercial Computing
  - Spark and Hadoop (HDFS, HBase, MapReduce)
  - Memcached is also used for Web 2.0
- Deep Learning
  - Caffe, CNTK, TensorFlow, and many more
- Cloud for HPC and BigData
  - Virtualization with SR-IOV and Containers

Big Data Management and Processing on Modern Clusters

- Substantial impact on designing and utilizing data management and processing systems in multiple tiers
  - Front-end data accessing and serving (Online)
    - Memcached + DB (e.g. MySQL), HBase
  - Back-end data analytics (Offline)
    - HDFS, MapReduce, Spark



How Can HPC Clusters with High-Performance Interconnect and Storage Architectures Benefit Big Data Applications?

Can the bottlenecks be alleviated with new designs by taking advantage of HPC technologies?

Can RDMA-enabled high-performance interconnects benefit Big Data processing?

Can HPC Clusters with high-performance storage systems (e.g. SSD, parallel file systems) benefit Big Data applications?

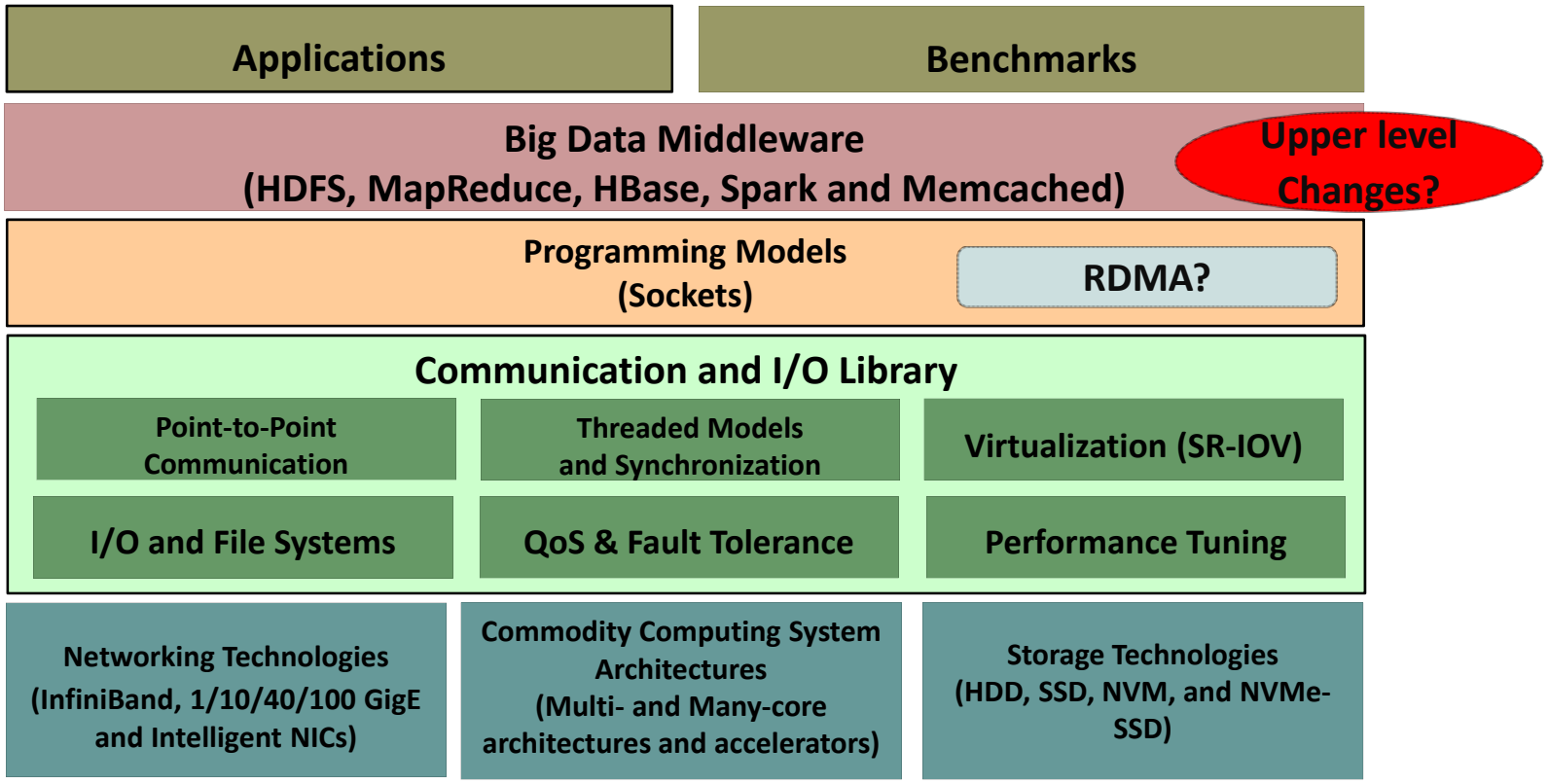
How much performance benefits can be achieved through enhanced designs?

What are the major bottlenecks in current Big Data processing middleware (e.g. Hadoop, Spark, and Memcached)?

How to design benchmarks for evaluating the performance of Big Data middleware on HPC clusters?

Bring HPC and Big Data processing into a "convergent trajectory"!

Designing Communication and I/O Libraries for Big Data Systems: Challenges

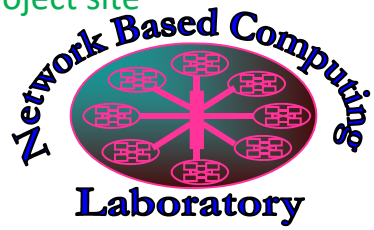


The High-Performance Big Data (HiBD) Project

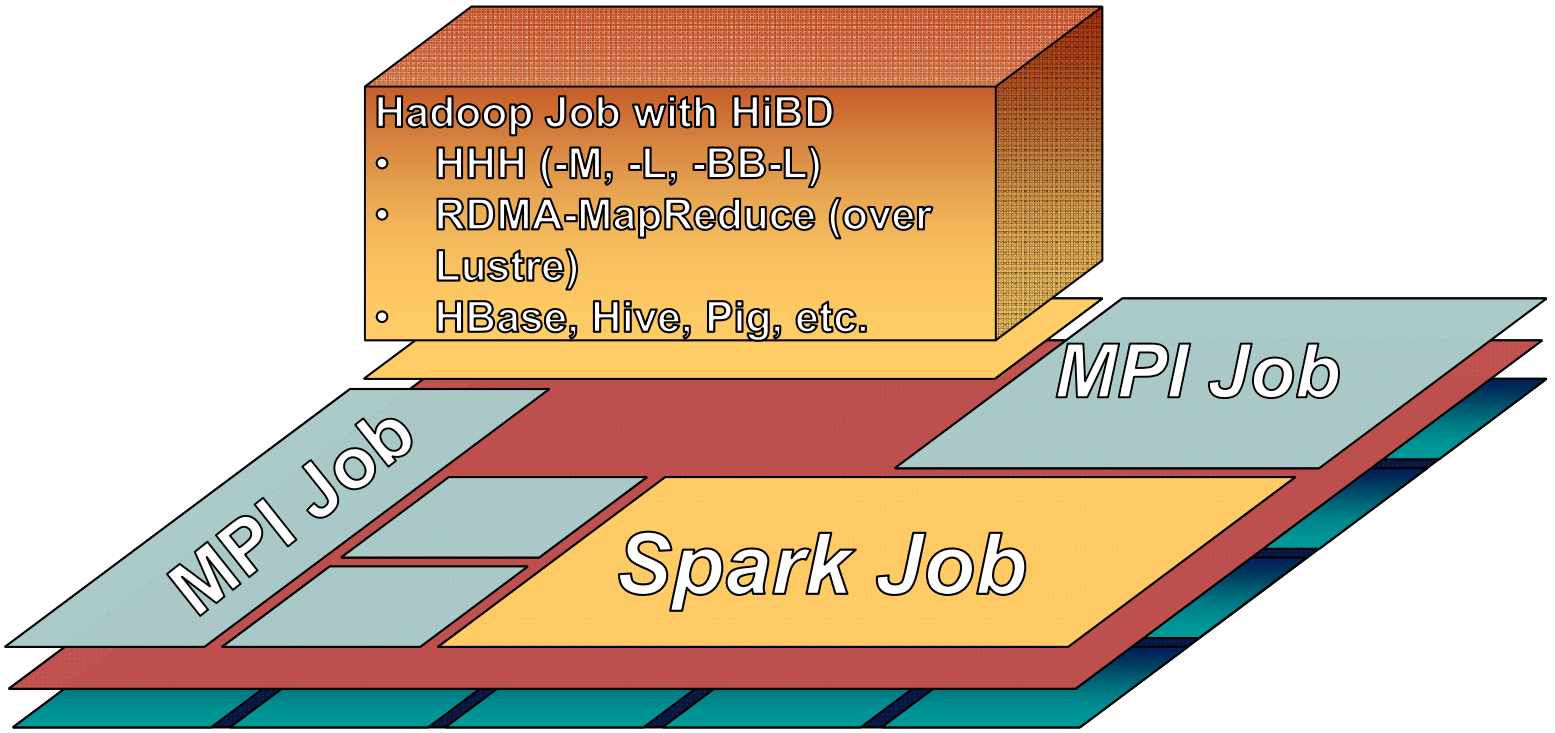
- RDMA for Apache Spark
- RDMA for Apache Hadoop 2.x (RDMA-Hadoop-2.x)
  - Plugins for Apache, Hortonworks (HDP) and Cloudera (CDH) Hadoop distributions
- RDMA for Apache HBase
- RDMA for Memcached (RDMA-Memcached)
- RDMA for Apache Hadoop 1.x (RDMA-Hadoop)
- OSU HiBD-Benchmarks (OHB)
  - HDFS, Memcached, HBase, and Spark Micro-benchmarks
- http://hibd.cse.ohio-state.edu
- Users Base: 235 organizations from 30 countries
- More than 22,000 downloads from the project site

Available for InfiniBand and RoCE

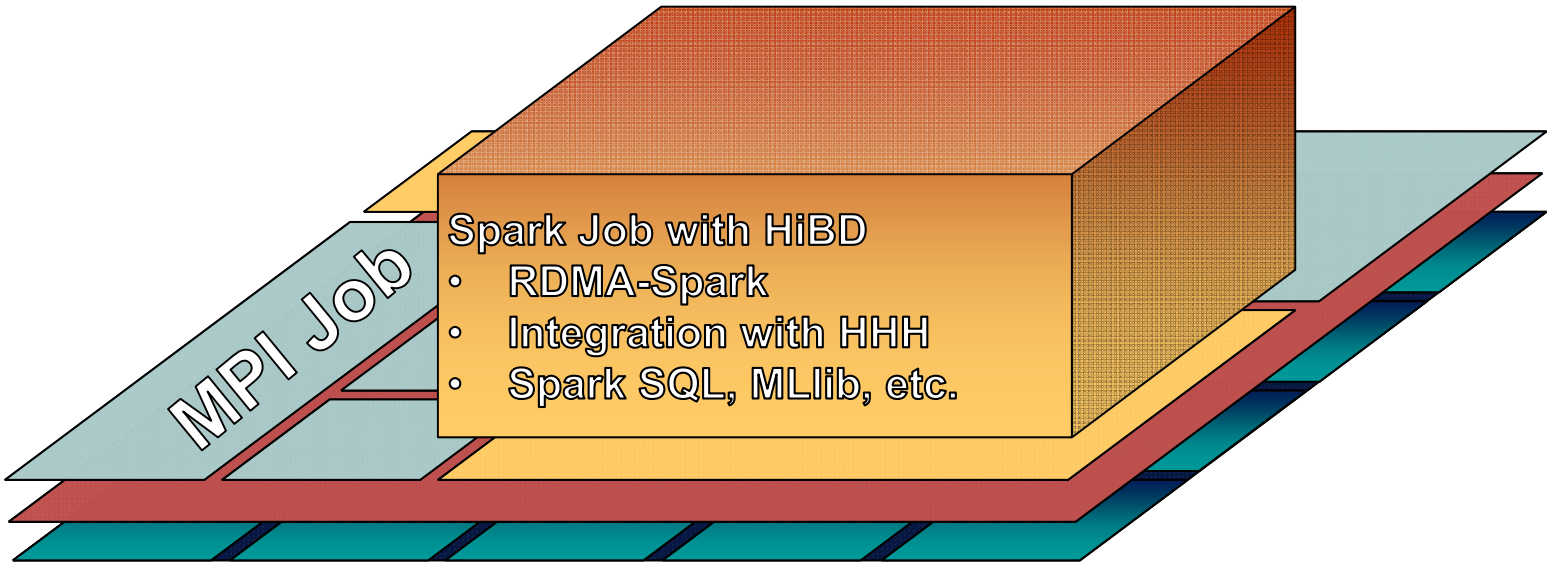
Also run on Ethernet



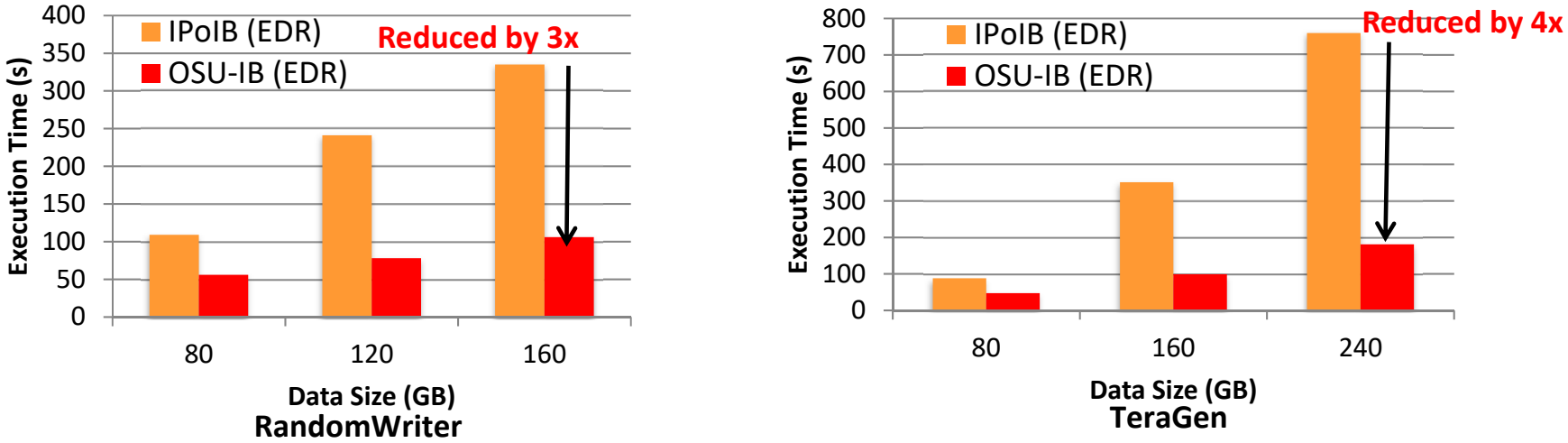
Using HiBD Packages on Existing HPC Infrastructure



Using HiBD Packages on Existing HPC Infrastructure



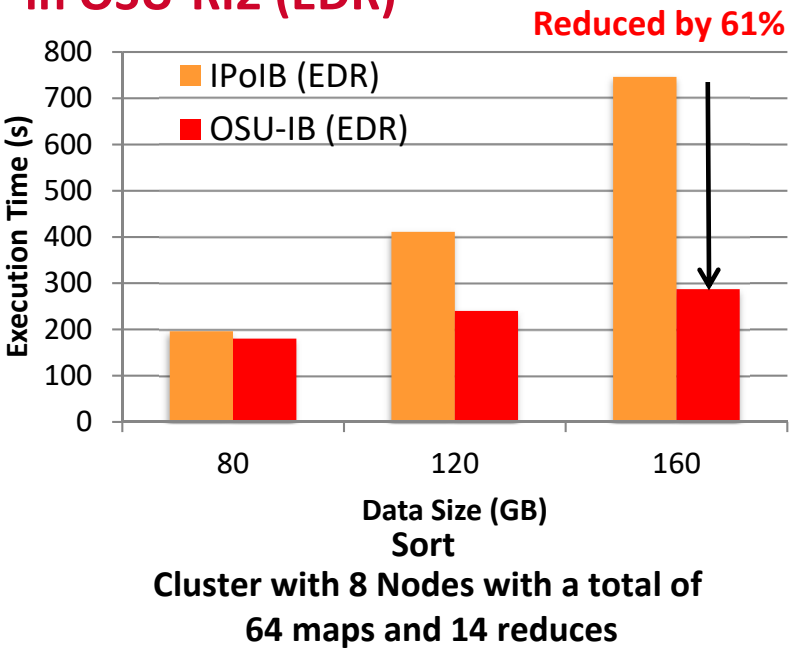
Performance Numbers of RDMA for Apache Hadoop 2.x – RandomWriter & TeraGen in OSU-RI2 (EDR)



Cluster with 8 Nodes with a total of 64 maps

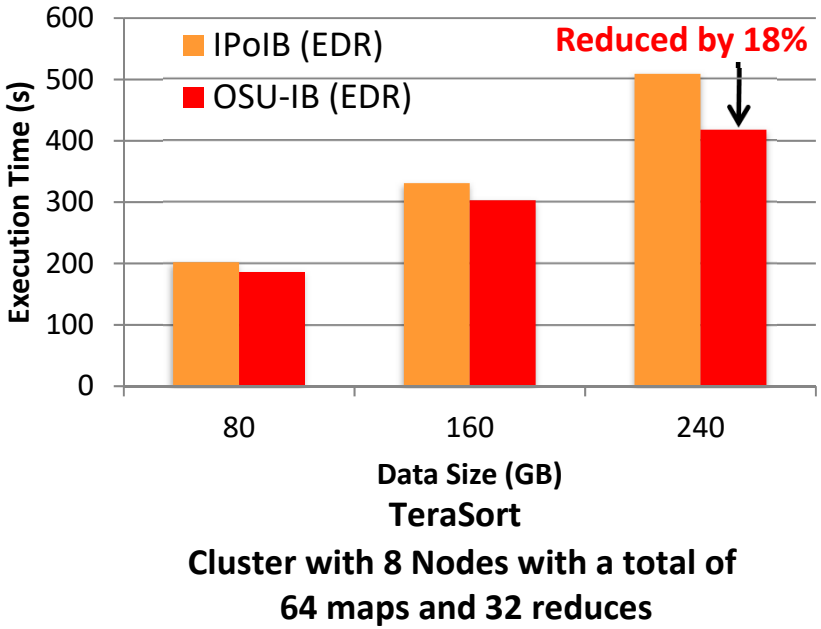
- RandomWriter
  - **3x** improvement over IPoIB for 80-160 GB file size
- TeraGen
  - **4x** improvement over IPoIB for 80-240 GB file size

Performance Numbers of RDMA for Apache Hadoop 2.x – Sort & TeraSort in OSU-RI2 (EDR)



Sort

Cluster with 8 Nodes with a total of 64 maps and 14 reduces

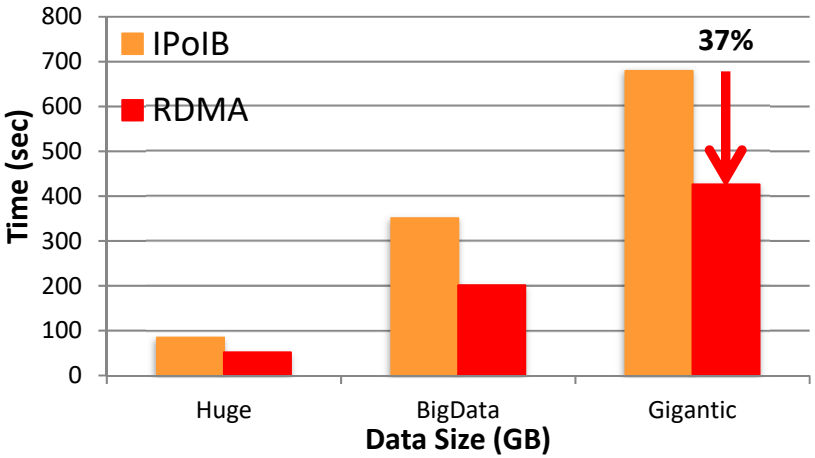


TeraSort

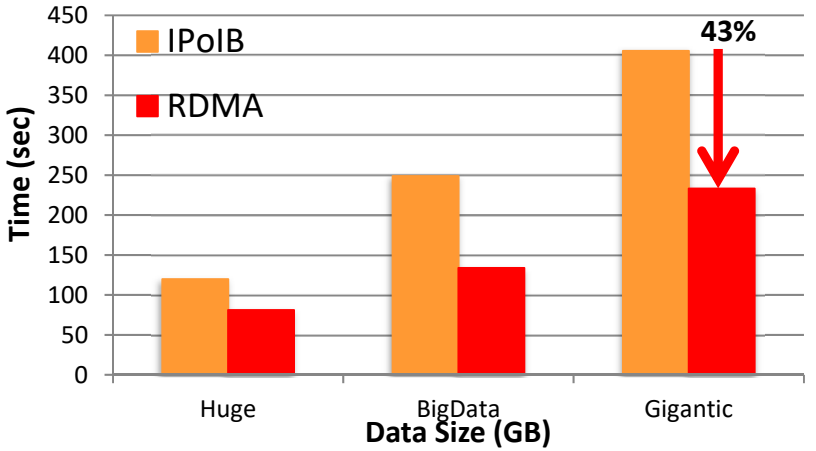
Cluster with 8 Nodes with a total of 64 maps and 32 reduces

- Sort
  - **61%** improvement over IPoIB for 80-160 GB data
- TeraSort
  - **18%** improvement over IPoIB for 80-240 GB data

Performance Evaluation of RDMA-Spark on SDSC Comet – HiBench PageRank



32 Worker Nodes, 768 cores, PageRank Total Time



64 Worker Nodes, 1536 cores, PageRank Total Time

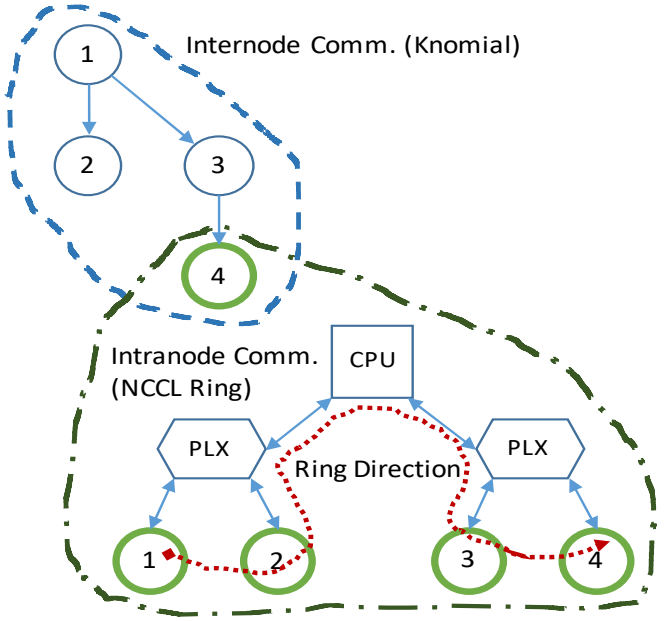
- InfiniBand FDR, SSD, 32/64 Worker Nodes, 768/1536 Cores, (768/1536M 768/1536R)
- RDMA vs. IPoIB with 768/1536 concurrent tasks, single SSD per node.
  - 32 nodes/768 cores: Total time reduced by 37% over IPoIB (56Gbps)
  - 64 nodes/1536 cores: Total time reduced by 43% over IPoIB (56Gbps)

HPC, Big Data, Deep Learning, and Cloud

- Traditional HPC
  - Message Passing Interface (MPI), including MPI + OpenMP
  - Support for PGAS and MPI + PGAS (OpenSHMEM, UPC)
  - Exploiting Accelerators
- Big Data/Enterprise/Commercial Computing
  - Spark and Hadoop (HDFS, HBase, MapReduce)
  - Memcached is also used for Web 2.0
- Deep Learning
  - Caffe, CNTK, TensorFlow, and many more
- Cloud for HPC and BigData
  - Virtualization with SR-IOV and Containers

Deep Learning: New Challenges for MPI Runtimes

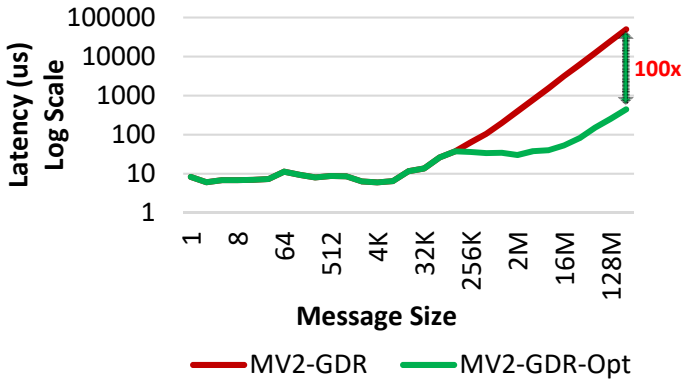
- Deep Learning frameworks are a different game altogether
  - Unusually large message sizes (order of megabytes)
  - Most communication based on GPU buffers
- How to address these newer requirements?
  - GPU-specific Communication Libraries (NCCL)
    - NVidia's NCCL library provides inter-GPU communication
  - CUDA-Aware MPI (MVAPICH2-GDR)
    - Provides support for GPU-based communication
- Can we exploit CUDA-Aware MPI and NCCL to support Deep Learning applications?



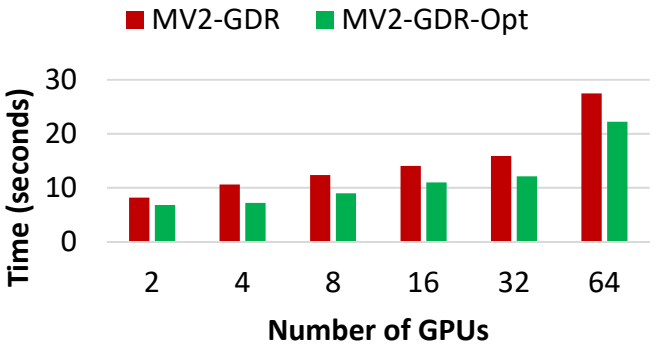
Hierarchical Communication (Knomial + NCCL ring)

Efficient Broadcast: MVAPICH2-GDR and NCCL

- NCCL has some limitations
  - Only works for a single node, thus, no scale-out on multiple nodes
  - Degradation across IOH (socket) for scale-up (within a node)
- We propose optimized MPI_Bcast
  - Communication of very large GPU buffers (order of megabytes)
  - Scale-out on large number of dense multi-GPU nodes
- Hierarchical Communication that efficiently exploits:
  - CUDA-Aware MPI_Bcast in MV2-GDR
  - NCCL Broadcast primitive



Performance Benefits: OSU Micro-benchmarks

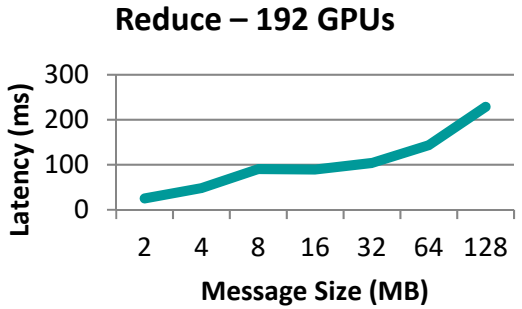


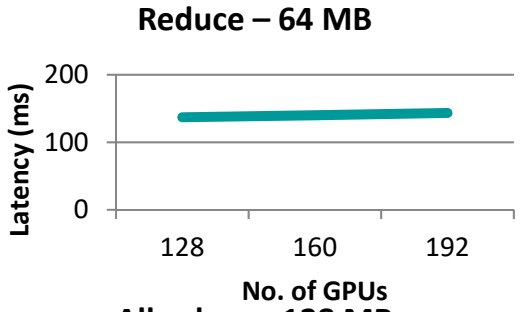
Performance Benefits: Microsoft CNTK DL framework
(25% avg. improvement)

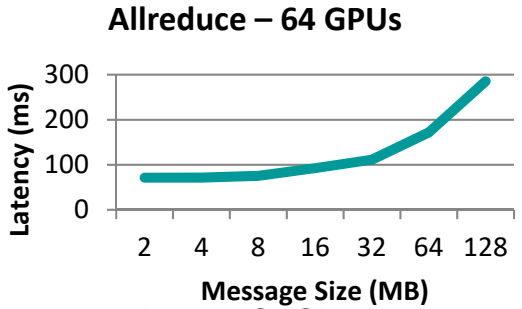
Efficient Large Message Broadcast using NCCL and CUDA-Aware MPI for Deep Learning,
A. Awan , K. Hamidouche , A. Venkatesh , and D. K. Panda,
The 23rd European MPI Users' Group Meeting (EuroMPI 16), Sep 2016 [Best Paper Runner-Up]

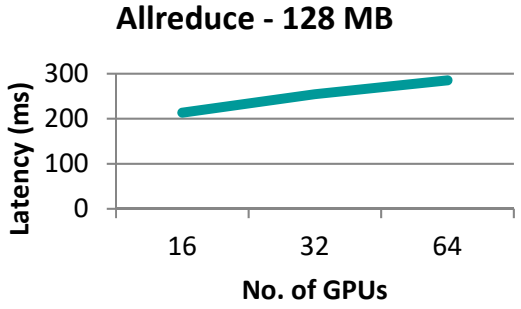
Large Message Optimized Collectives for Deep Learning

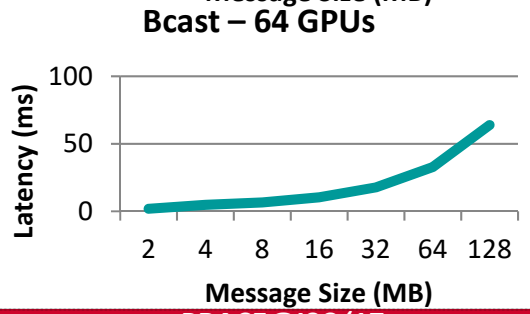
- MV2-GDR provides optimized collectives for **large message sizes**
- Optimized Reduce, Allreduce, and Bcast
- **Good scaling with large number of GPUs**
- **Available with MVAPICH2-GDR 2.2GA**

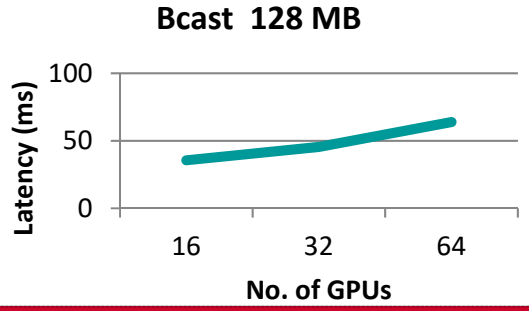










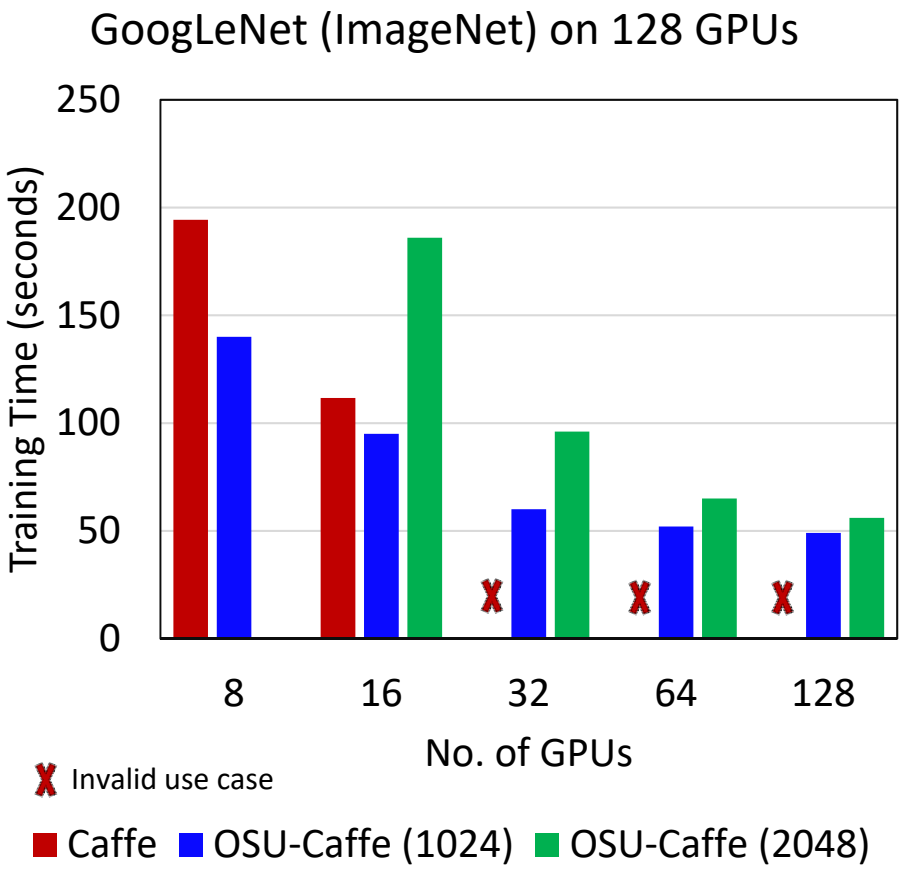


OSU-Caffe: Scalable Deep Learning

- Caffe : A flexible and layered Deep Learning framework.
- Benefits and Weaknesses
  - Multi-GPU Training within a single node
  - Performance degradation for GPUs across different sockets
  - Limited Scale-out
- OSU-Caffe: MPI-based Parallel Training
  - Enable Scale-up (within a node) and Scale-out (across multi-GPU nodes)
  - network on ImageNet dataset

Awan, K. Hamidouche, J. Hashmi, and D. K. Panda, S-Caffe: Co-designing MPI Runtimes and Caffe for Scalable Deep Learning on Modern GPU Clusters, PPoPP, Sep 2017

OSU-Caffe is publicly available from: http://hidl.cse.ohio-state.edu



HPC, Big Data, Deep Learning, and Cloud

- Traditional HPC
  - Message Passing Interface (MPI), including MPI + OpenMP
  - Support for PGAS and MPI + PGAS (OpenSHMEM, UPC)
  - Exploiting Accelerators
- Big Data/Enterprise/Commercial Computing
  - Spark and Hadoop (HDFS, HBase, MapReduce)
  - Memcached is also used for Web 2.0
- Deep Learning
  - Caffe, CNTK, TensorFlow, and many more
- Cloud for HPC and BigData
  - Virtualization with SR-IOV and Containers

Can HPC and Virtualization be Combined?

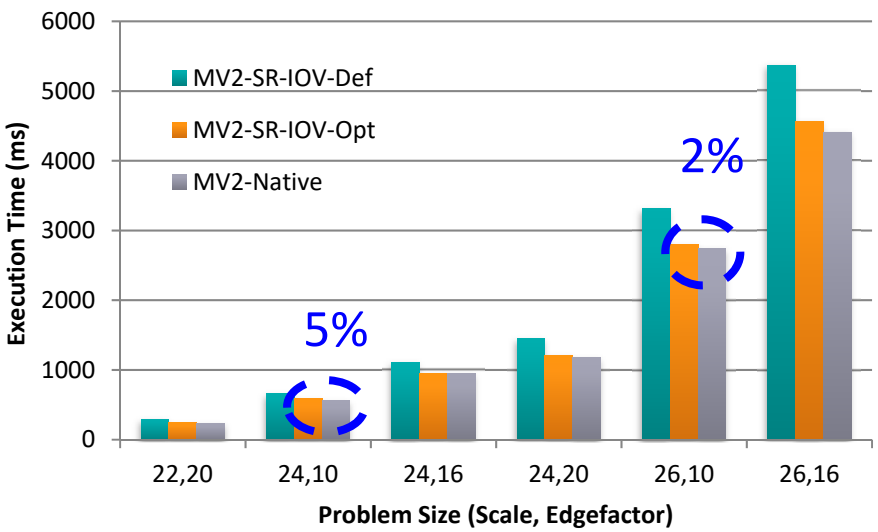
- Virtualization has many benefits
  - Fault-tolerance
  - Job migration
  - Compaction
- Have not been very popular in HPC due to overhead associated with Virtualization
- New SR-IOV (Single Root – IO Virtualization) support available with Mellanox InfiniBand adapters changes the field
- Enhanced MVAPICH2 support for SR-IOV
- MVAPICH2-Virt 2.1 (with and without OpenStack) is publicly available

J. Zhang, X. Lu, J. Jose, R. Shi and D. K. Panda, Can Inter-VM Shmem Benefit MPI Applications on SR-IOV based Virtualized InfiniBand Clusters? EuroPar'14

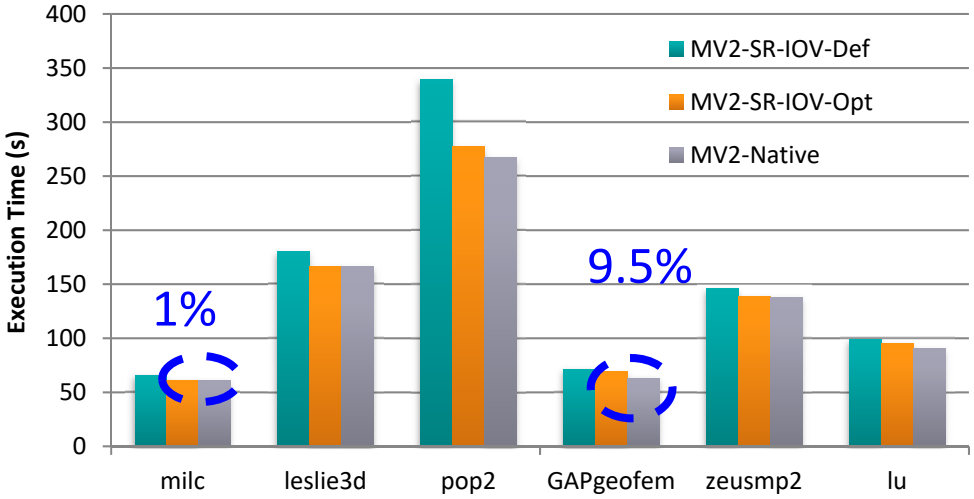
J. Zhang, X. Lu, J. Jose, M. Li, R. Shi and D.K. Panda, High Performance MPI Libray over SR-IOV enabled InfiniBand Clusters, HiPC'14

J. Zhang, X .Lu, M. Arnold and D. K. Panda, MVAPICH2 Over OpenStack with SR-IOV: an Efficient Approach to build HPC Clouds, CCGrid'15

Application-Level Performance on Chameleon



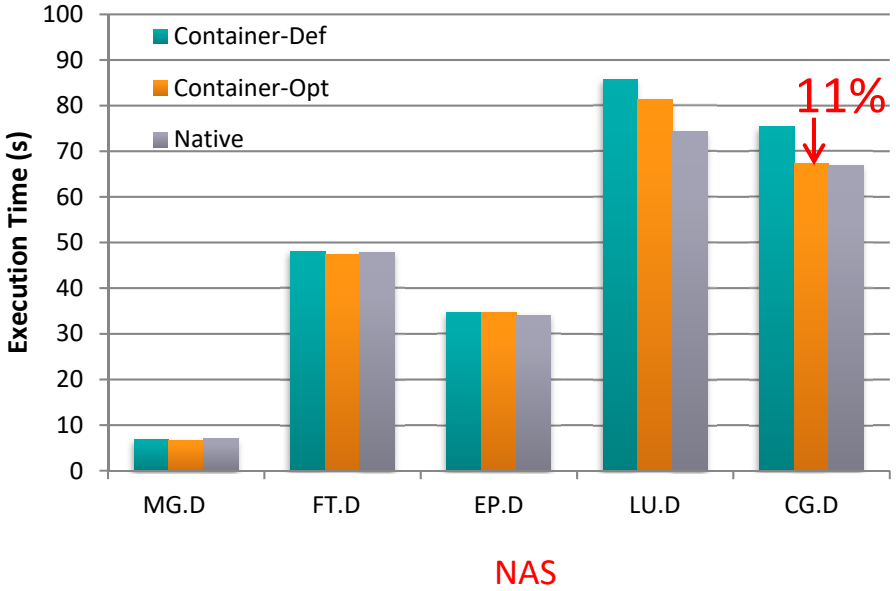
Graph500



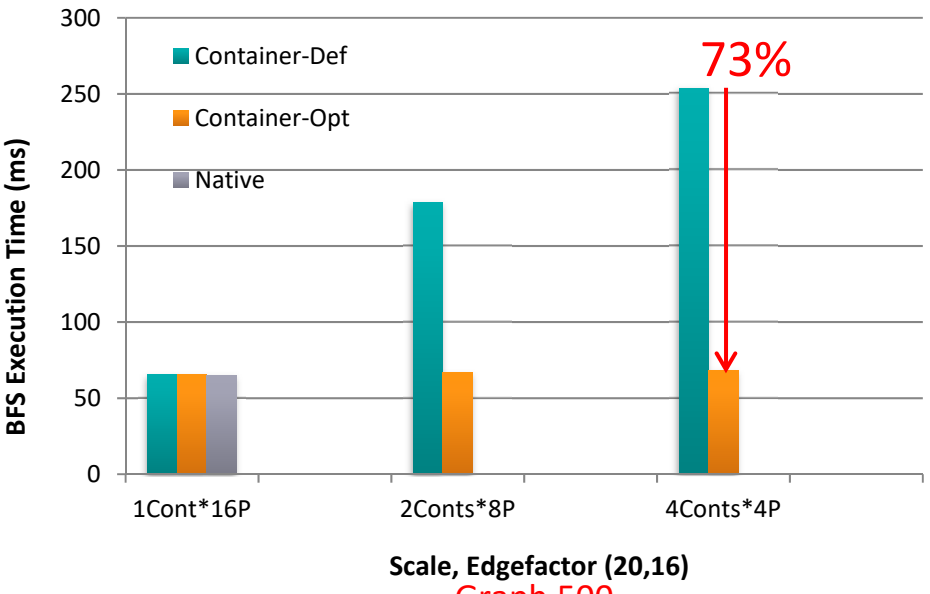
SPEC MPI2007

- 32 VMs, 6 Core/VM
- Compared to Native, 2-5% overhead for Graph500 with 128 Procs
- Compared to Native, 1-9.5% overhead for SPEC MPI2007 with 128 Procs

Application-Level Performance on Docker with MVAPICH2



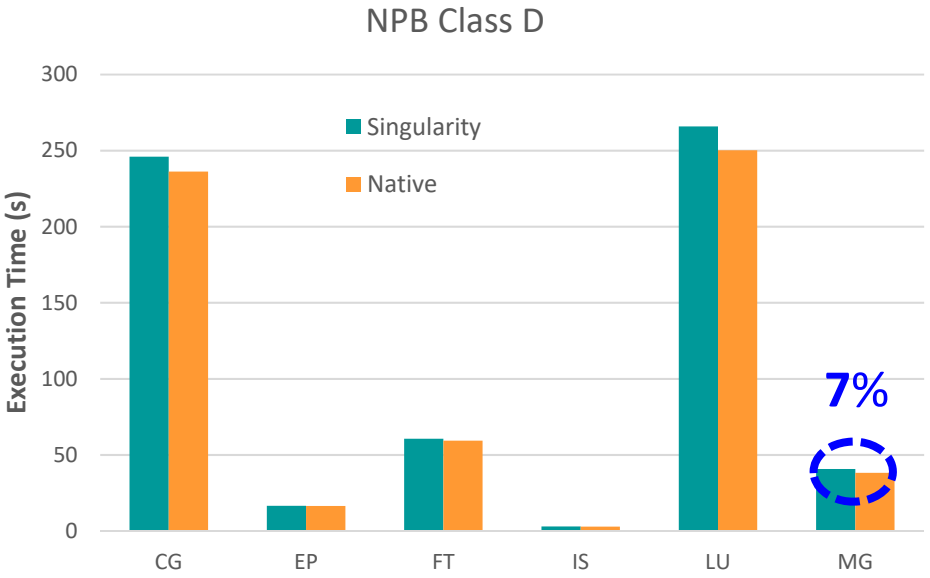
NAS

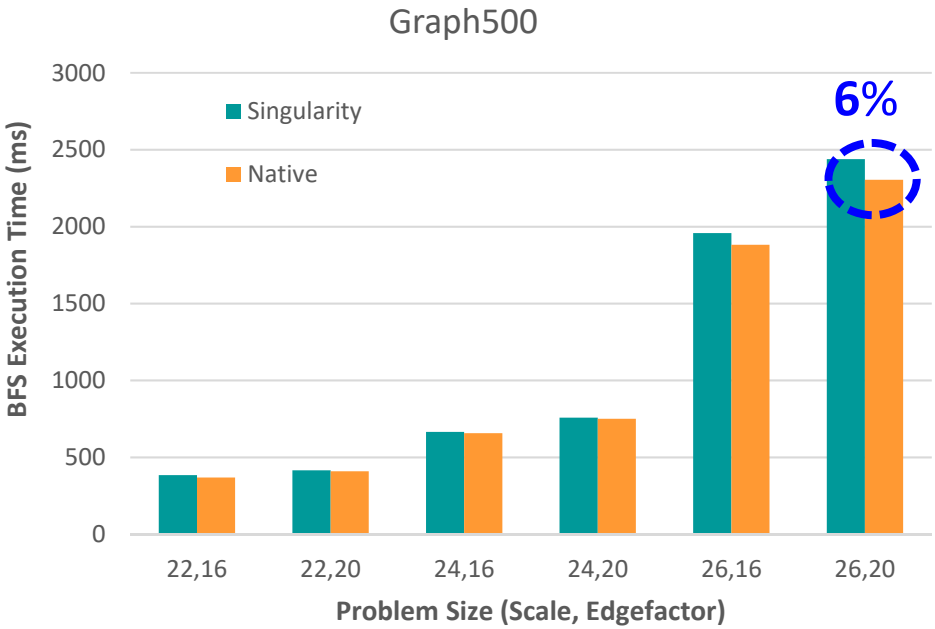


Graph 500

- 64 Containers across 16 nodes, pining 4 Cores per Container
- Compared to Container-Def, up to 11% and 73% of execution time reduction for NAS and Graph 500
- Compared to Native, less than 9 % and 5% overhead for NAS and Graph 500

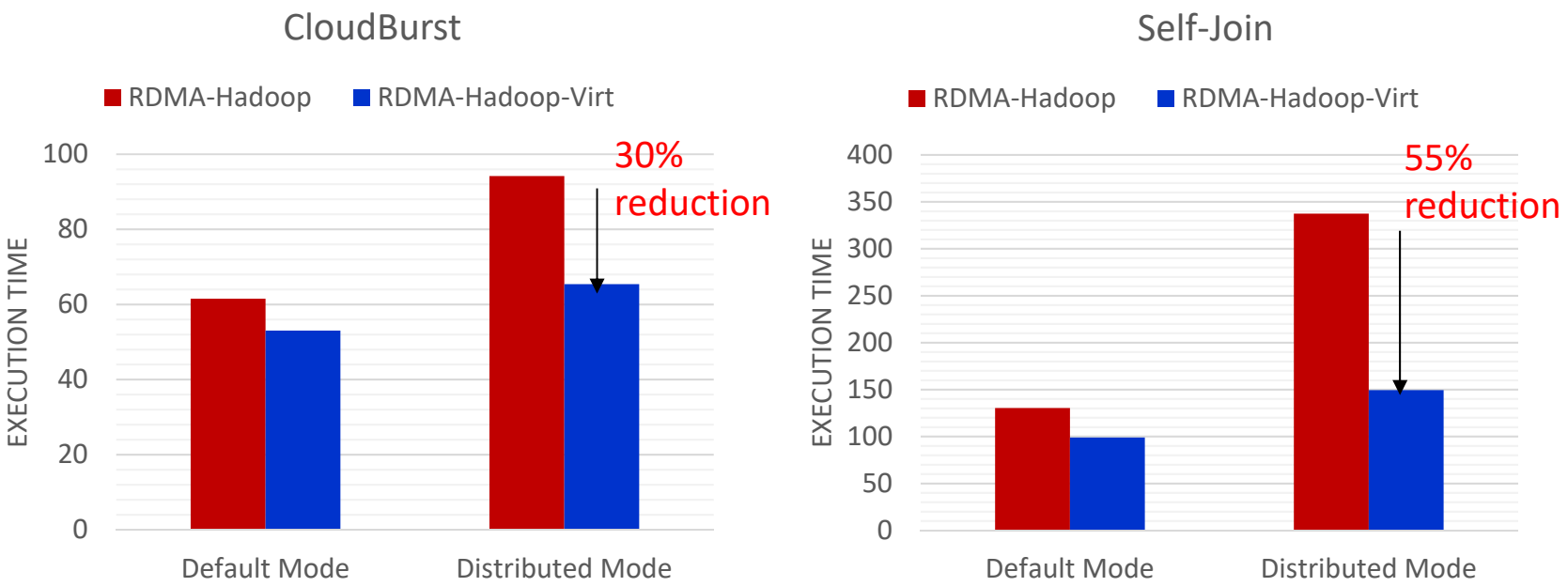
Application-Level Performance on Singularity with MVAPICH2





- 512 Processes across 32 nodes
- Less than 7% and 6% overhead for NPB and Graph500, respectively

RDMA-Hadoop with Virtualization



- 14% and 24% improvement with Default Mode for CloudBurst and Self-Join
- 30% and 55% improvement with Distributed Mode for CloudBurst and Self-Join

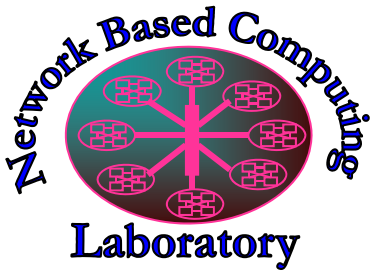
S. Gugnani, X. Lu, D. K. Panda. Designing Virtualization-aware and Automatic Topology Detection Schemes for Accelerating Hadoop on SR-IOV-enabled Clouds. CloudCom, 2016.

Concluding Remarks

- Next generation HPC systems need to be designed with a holistic view of Big Data, Deep Learning and Cloud
- Presented some of the approaches and results along these directions
- Enable Big Data, Deep Learning and Cloud community to take advantage of modern HPC technologies
- Many other open issues need to be solved
- Next generation HPC professionals need to be trained along these directions
- Looking forward to collaboration with the PRACE community

Thank You!

panda@cse.ohio-state.edu



Network-Based Computing Laboratory
http://nowlab.cse.ohio-state.edu/

The High-Performance MPI/PGAS Project
http://mvapich.cse.ohio-state.edu/

The High-Performance Big Data Project
http://hibd.cse.ohio-state.edu/

The High-Performance Deep Learning Project
http://hidl.cse.ohio-state.edu/

PRACE Internships: Influencing university leavers towards HPC related career choices

Leon Kos, PRACE Summer of HPC and MOOC coordinator

University of Ljubljana

ISC High Performance
The HPC Event.

PRACE Training Workshop
22 June 2017, Frankfurt

Vision: PRACE wishes to inspire and encourage the next generation of software engineers, system administrators, and general users of HPC systems in order to keep Europe at the forefront of supercomputing.

Offerings by PRACE to university students

- Summer of HPC – from 2013
- International HPC summer school – from 2010
- Massive Online Open Courses – from 2016
- Outreach to universities – from 2017

PRACE **Summer of HPC** outreach programme

- All paid two months placement in a PRACE partner country for **undergraduate and junior postgraduate** university students.
- Students undertake a HPC project, preferably visualisation, based on the outcomes from PRACE technical work or other work using PRACE resources, under supervision of a **project mentor**.

Goals of the PRACE Summer of HPC programme

- The primary goal of the programme is **outreach**. The programme aims to ensure a positive experience for all students and through that experience to encourage them in their path to become the next generation of HPC users. It is hoped that the programme will also create goodwill around PRACE and HPC both in the home institutions and amongst the peers of all applicants and participants.
- The secondary goal of the programme is the successful completion of the student projects; with the visualisations produced being of benefit to PRACE in future **outreach** and **dissemination** activities.

Projects

- 20+ projects with HPC and visualization relevance (outreach).
- Proposed by mentors that work with students during the summer.
- Ideally one-to-one (student-mentor) match.
- 10+ sites hosts at least two students.

Timeline

- December — Projects and application dates announced
- January — Call for Applications opens
- February — Call for Applications closes
- March — Student selection and travel arrangements
- first week of July — Training week
- July – August — Students at sites with weekly telcos
- end of August — Presentations and reports by students
- October — Awards ceremony

Student Selection Process

- Students apply selecting maximum 3 projects they are particularly interested in doing and why.
- The Selection panel reviews each application evaluated by two reviewers.
- Grading based on CV, reference, code test and communication skills including social media.
- Consensus meeting on ranking by projects.

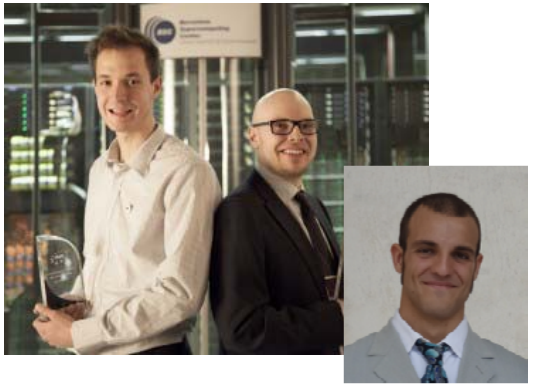
History

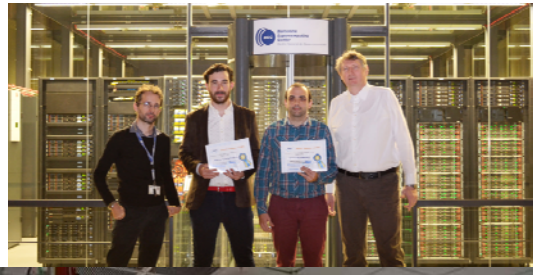
- Programme started in 2013 with 24 students
- 2014 edition with 10 students
- 2015 edition with 20 students at 9 sites
- 2016 with 21 students at 10 sites
- 2017 with 21 students at 10 sites


Training week at Juelich Supercomputing Centre, July 2016

Awards

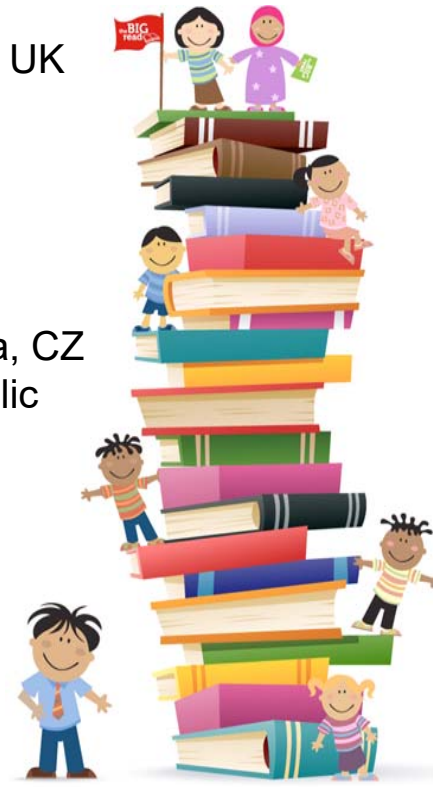
- The Best HPC Ambassador award
  - Presented to the participant who best embodies the outreach spirit of the programme.
- The Best Visualisation Award
  - Presented to the participant who completes the best visualisation during the Summer of HPC Programme.
- The awards consist of a diploma, a trophy and a budget of € 1.500,- per winner to go to a PRACE Training Event or similar.





Projects 2017

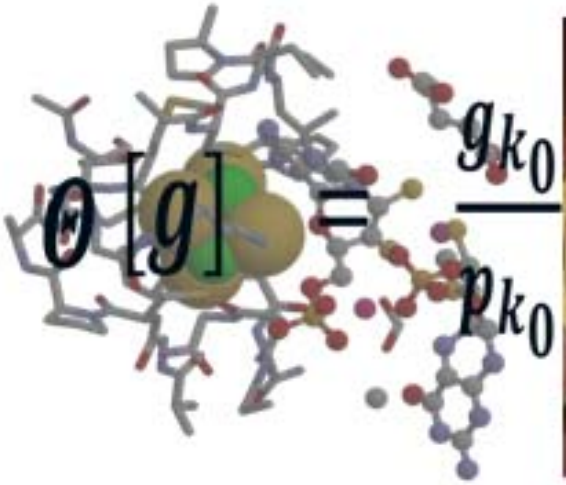
- 1701 Hybrid Monte Carlo Method for Matrix Computation on P100 GPUs, BSC, Barcelona, Spain
- 1702 Monte Carlo and Deep Learning Methods for Enhancing Crowd Simulation, BSC, Barcelona, Spain
- 1703 Apache Spark: Are Big Data tools applicable in HPC?, CC SAS, Bratislava, Slovakia
- 1704 Calculation of nanotubes by utilizing the helical symmetry properties, CC SAS, Bratislava, Slovakia
- 1705 Web visualization of Energy load of an HPC system, CINECA, Casalecchio di Reno, Italy
- 1706 Web visualization of the Mediterranean Sea, CINECA, Casalecchio di Reno, Italy
- 1707 Development and validation of real-time earthquake hazard models, EPCC, Edinburgh, UK
- 1708 Interactive weather forecasting on supercomputers as a tool for education, EPCC, Edinburgh, UK
- 1709 Online visualisation of current and historic supercomputer usage, EPCC, Edinburgh, UK
- 1710 Visualizing European Climate Change, AUTH/GRNET, Athens, Greece
- 1711 European climate model simulations, AUTH/GRNET, Athens, Greece
- 1712 El-Nino: It's periodicity and impact on world weather, ICHEC, Dublin, Ireland
- 1713 Radiosity in Computer Graphics, ICHEC, Dublin, Ireland
- 1714 Visualization of real motion of human body based on motion capture technology, IT4I, Ostrava, CZ
- 1715 Performance visualization for bioinformatics pipelines, IT4Innovations, Ostrava, Czech Republic
- 1716 Cude colors on phine grid, JSC, Jülich, Germany
- 1717 Hip, hip, hooray! Get your 2-for-1 GPU deal now., JSC, Jülich, Germany
- 1718 Accelerating climate kernels, UCPH, Copenhagen, Denmark
- 1719 Tracing in 4D data, UCPH, Copenhagen, Denmark
- 1720 Parallel algorithm for non-negative matrix tri-factorization, UL, Ljubljana, Slovenia
- 1721 CAD data extraction for CFD simulation, UL, Ljubljana, Slovenia

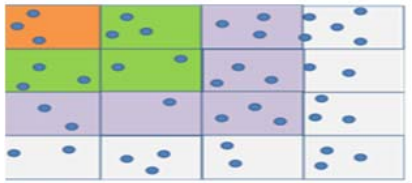


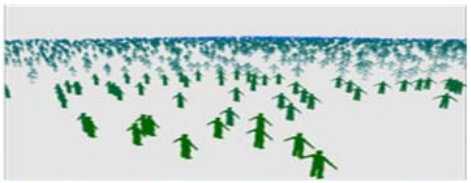
BSC Spain

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

- **Hybrid Monte Carlo Method for Matrix Computation on P100 GPUs**
- **Monte Carlo and Deep Learning Methods for Enhancing Crowd Simulation**



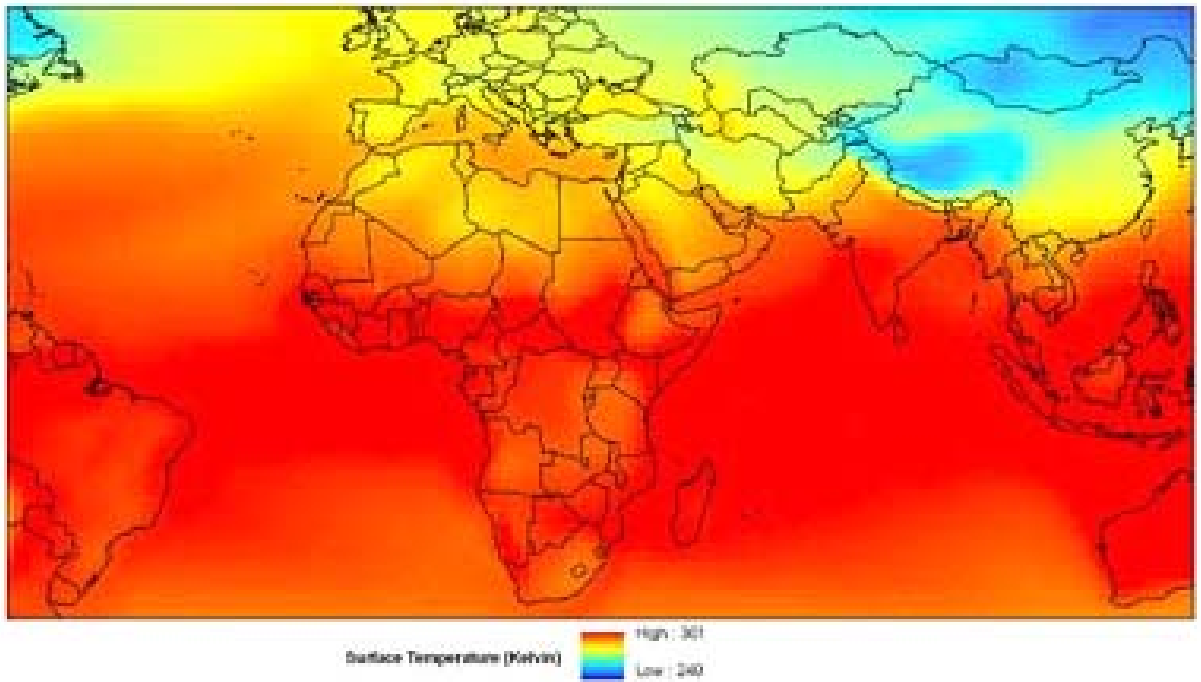


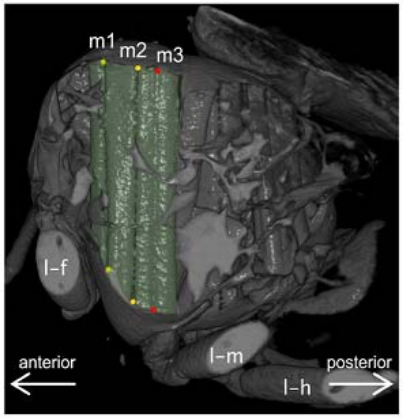


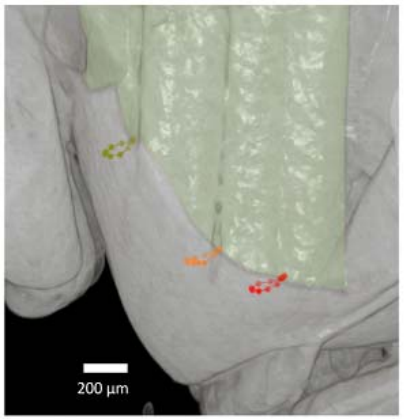
UCPH Denmark

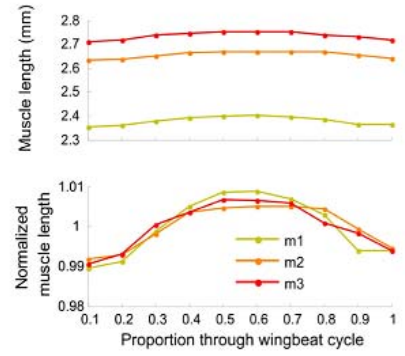
Niels Bohr Institutet

- **Accelerating climate kernels**
- **Tracing in 4D data**



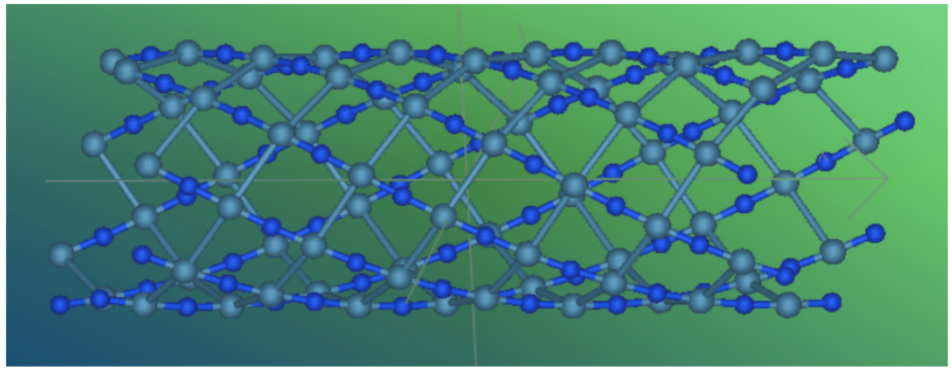


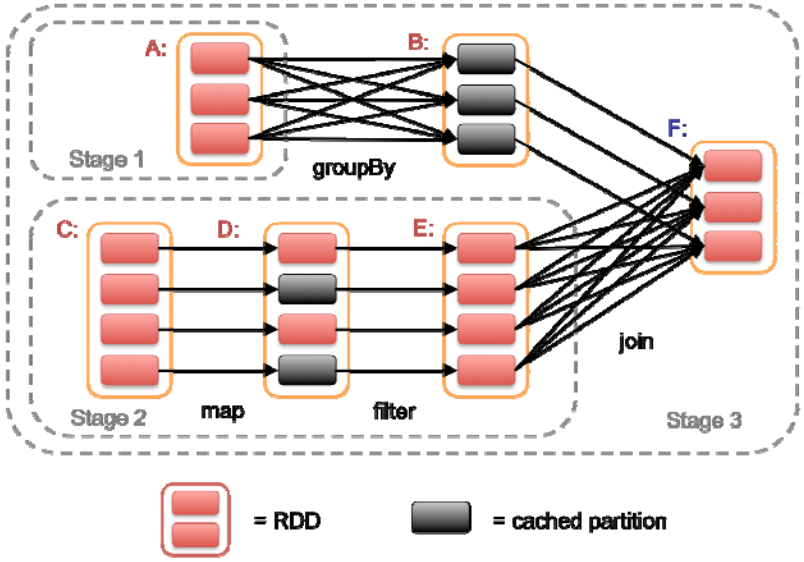




CC SAS Slovakia

- **Apache Spark: Are Big Data tools applicable in HPC?**
- **Calculation of nanotubes by utilizing the helical symmetry properties**

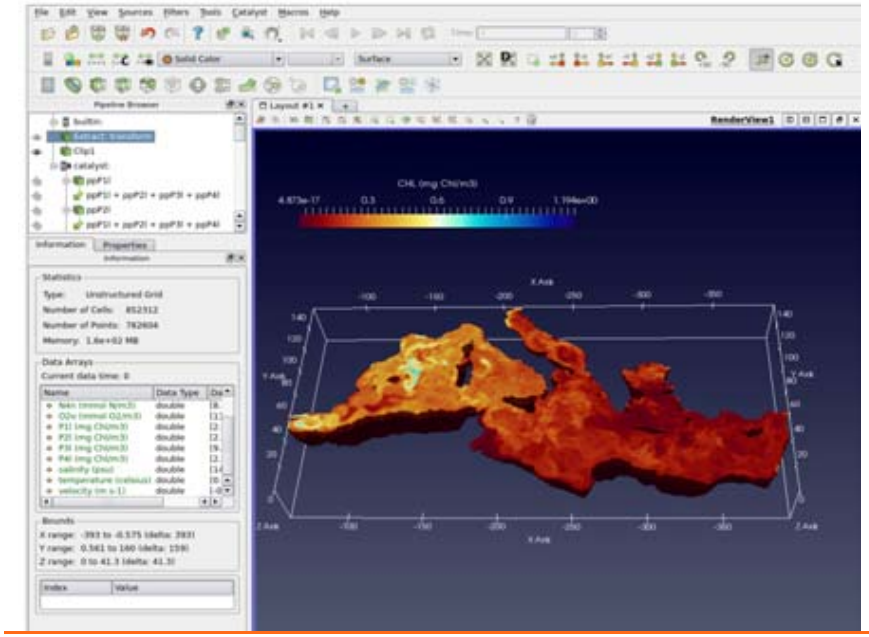


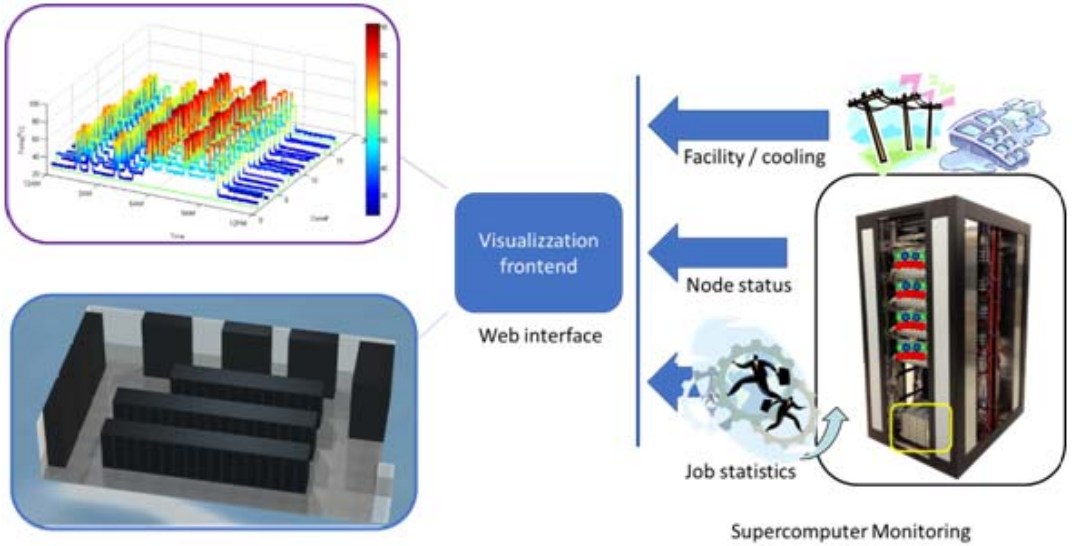


Source: Matei Zaharia slide deck on Spark at Strata conference Feb 2013

CINECA Italy

- **Web visualization of Energy load of an HPC system**
- **Web visualization of the Mediterranean Sea**

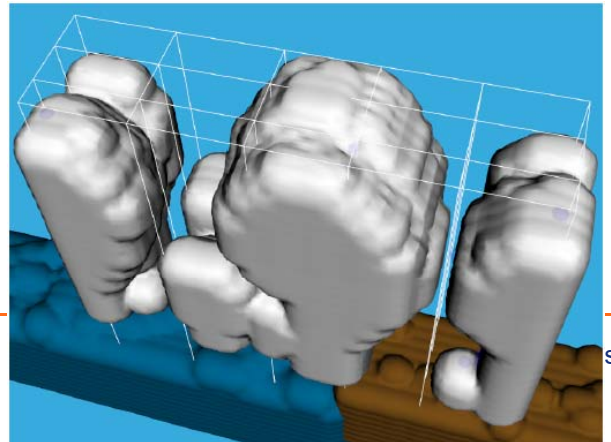


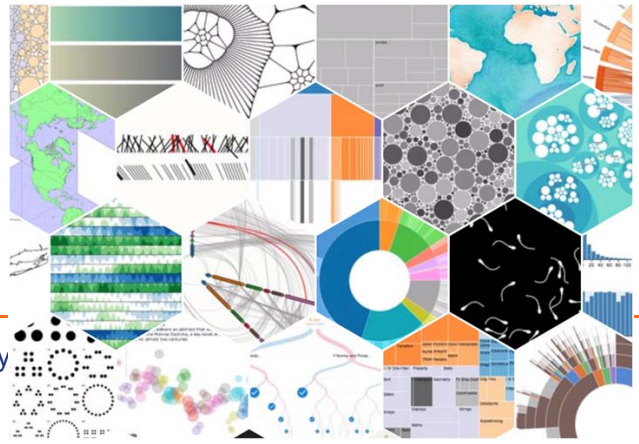


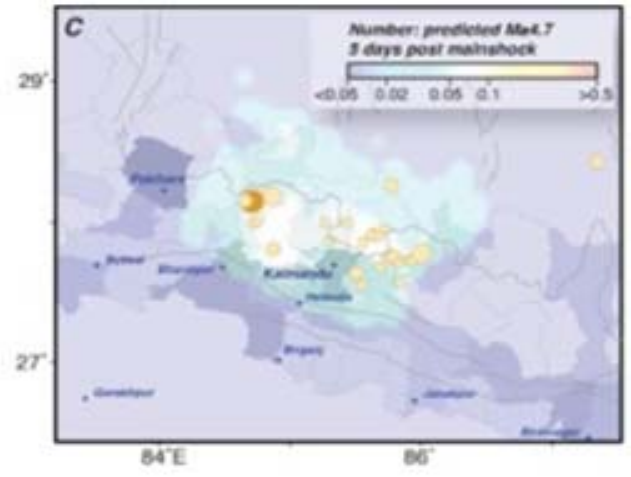
epcc

UK

- Interactive weather forecasting on supercomputers as a tool for education
- Online visualisation of current and historic supercomputer usage
- Development and validation of real-time earthquake

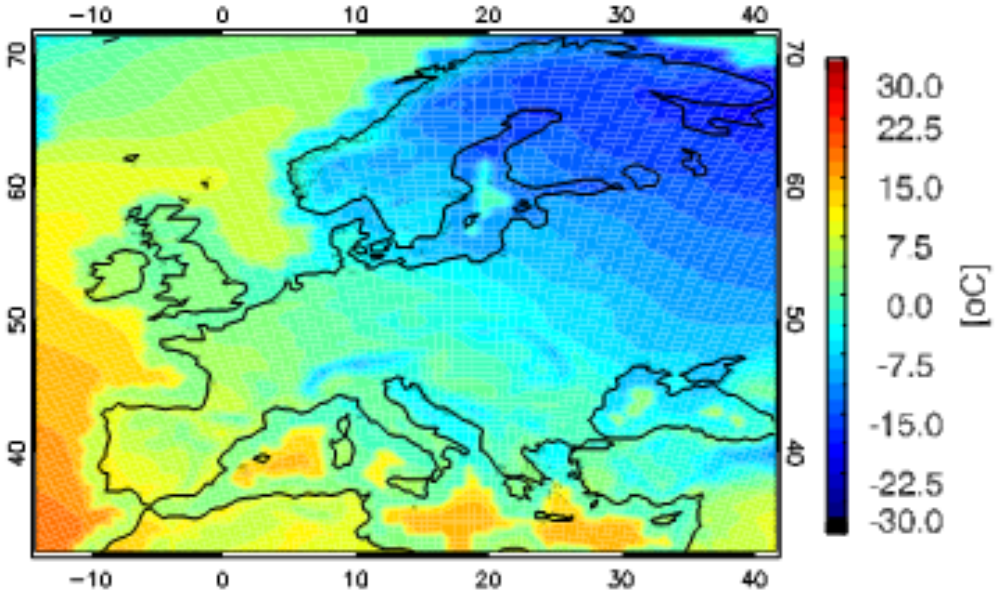


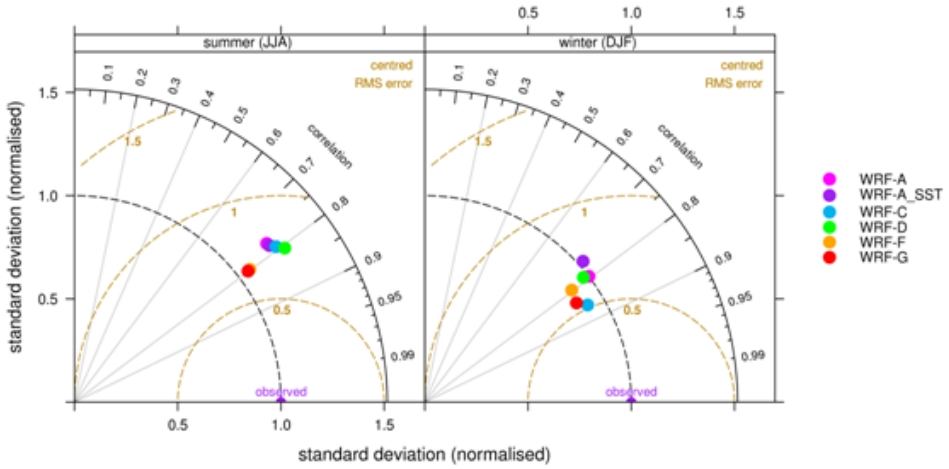




GRNET Greece

- **European climate model simulations**
- **Visualizing European Climate Change**

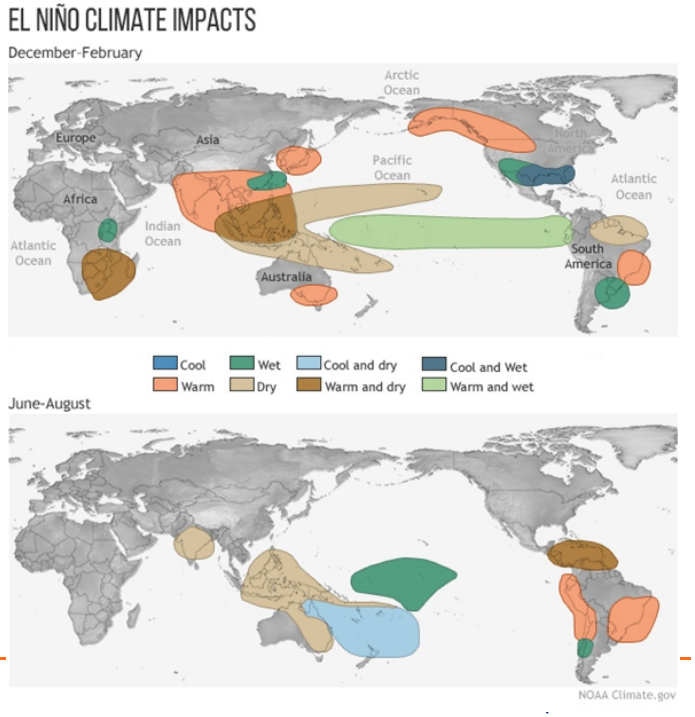




ICHEC Ireland

ICHEC
Irish Centre for High-End Computing

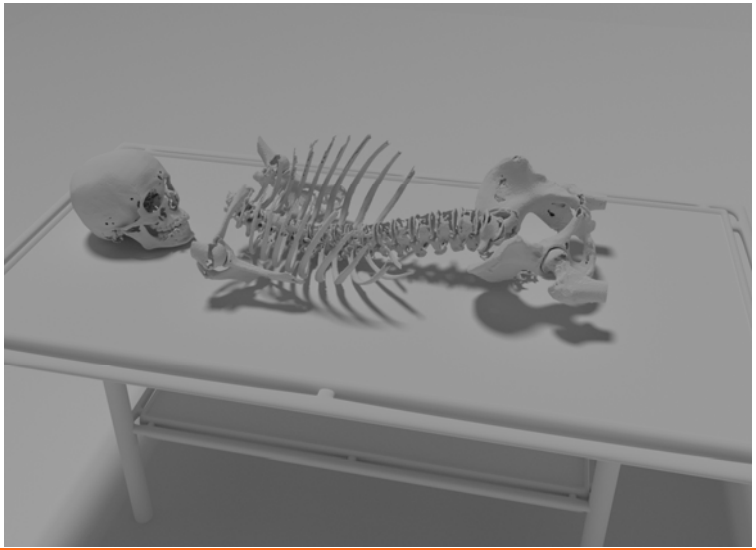
- **“El-Nino: It’s periodicity and impact on world weather”**
- **Radiosity in Computer Graphics**

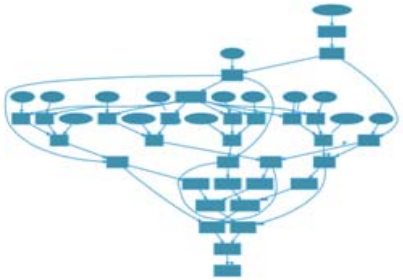


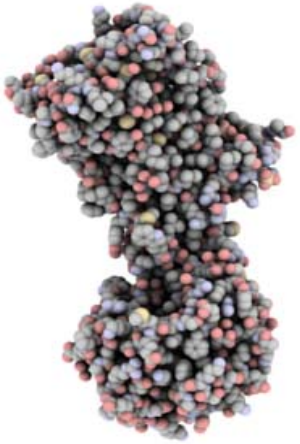
IT4I Czech Republic

IT4Innovations national supercomputing center

- ***Visualization of real motion of human body based on motion capture technology***
- **Performance visualization for bioinformatics pipelines**

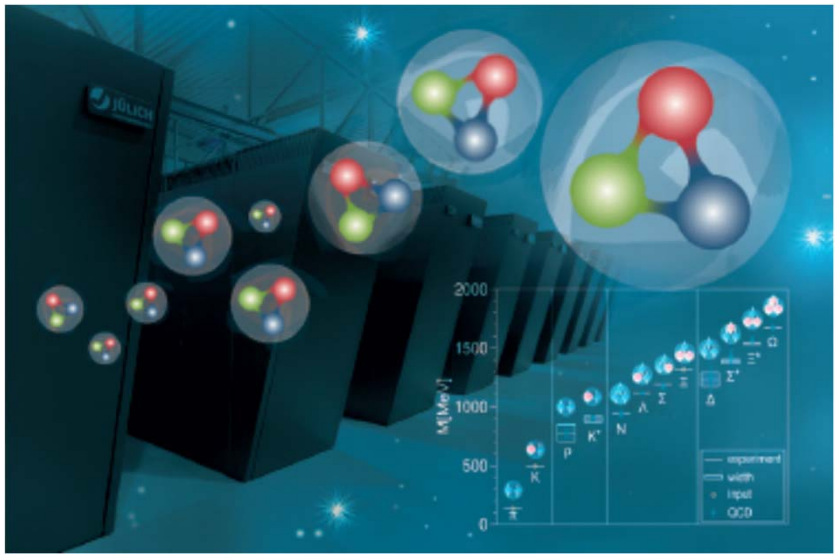






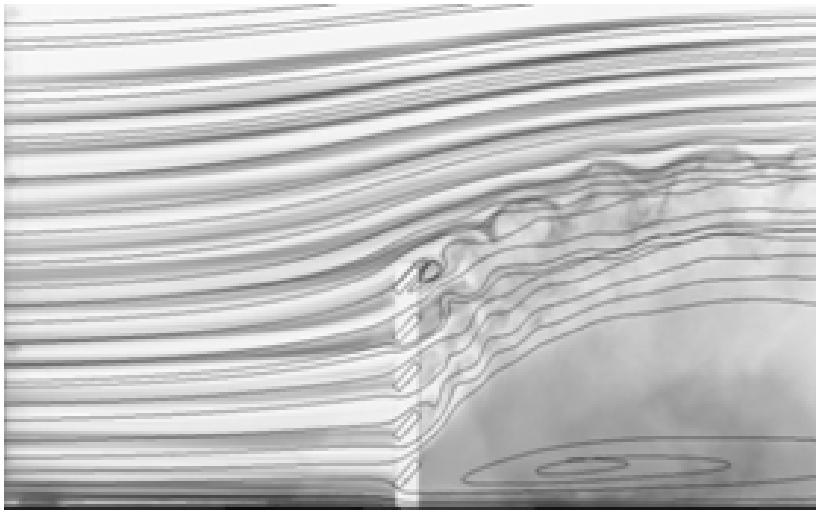
JSC Germany

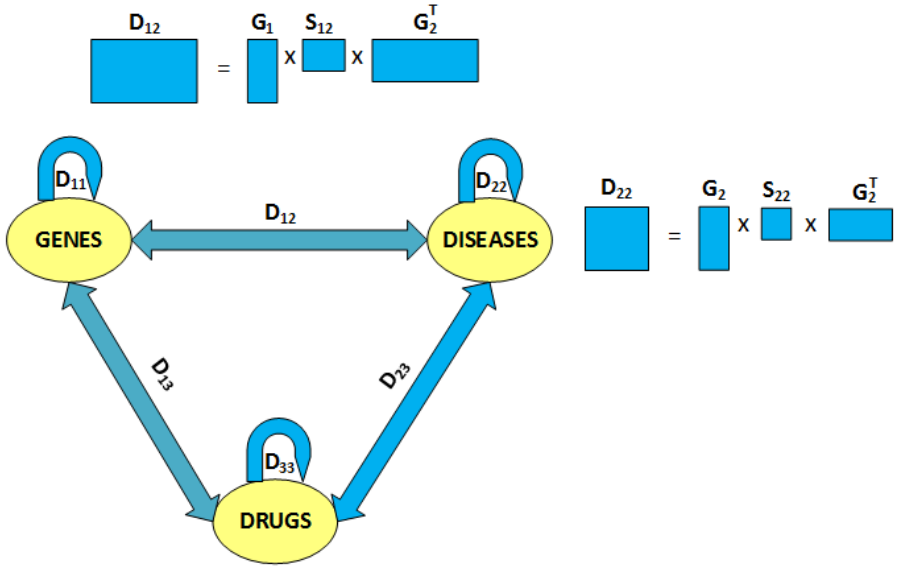
- **Cude colors on a phine grid**
- **Hip, hip, hooray! Get your 2-for-1 GPU deal now.**



University of Ljubljana Slovenia

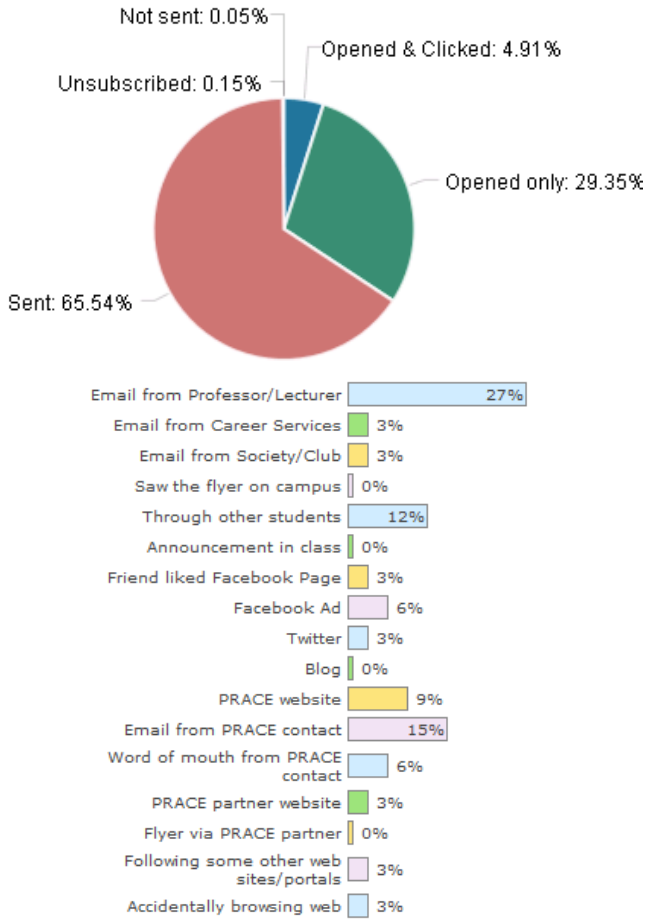
- **Parallel algorithm for non-negative matrix tri-factorization**
- **CAD data extraction for CFD simulation**





KPI and Impact

- Google analytics and Klout
- PRACE website announcements click stats
- E-mail is a prime communicator
- Importance of liaising with university career office
- Surveys for feedback with BSCW
- Excellent examples of careers



http://summerofhpc.prace-ri.eu

THANK YOU FOR YOUR ATTENTION

www.prace-ri.eu

SMES, HPC AND TRAINING

Paul Graham, Software Architect

EPCC, University of Edinburgh

PRACE SHAPE Coordinator

ISC 2017: PRACE Training Workshop

SME HPC Adoption Programme in Europe

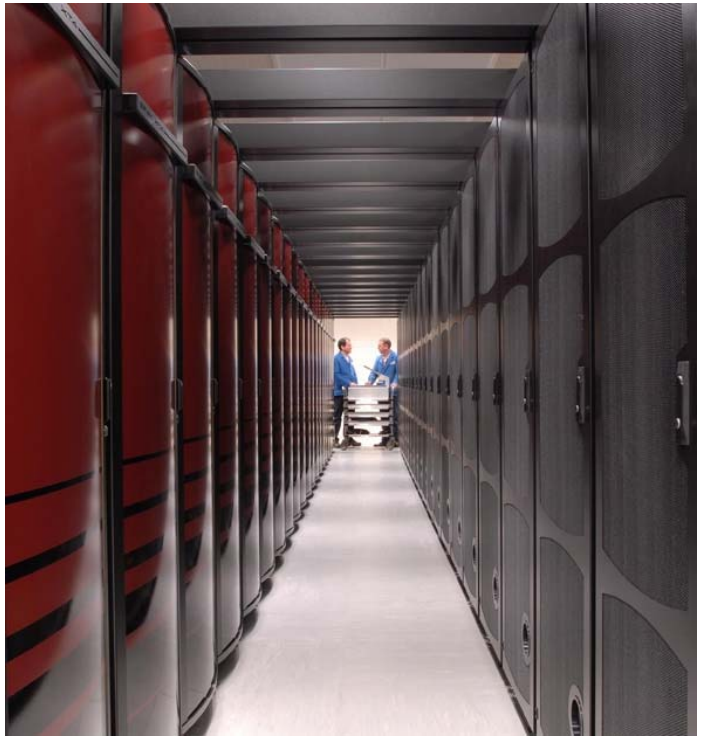
www.prace-ri.eu

Overview of talk

- EPCC and industry
- SMEs in Europe
- European HPC programme for SMEs: SHAPE
- Training needs of SMEs

EPCC

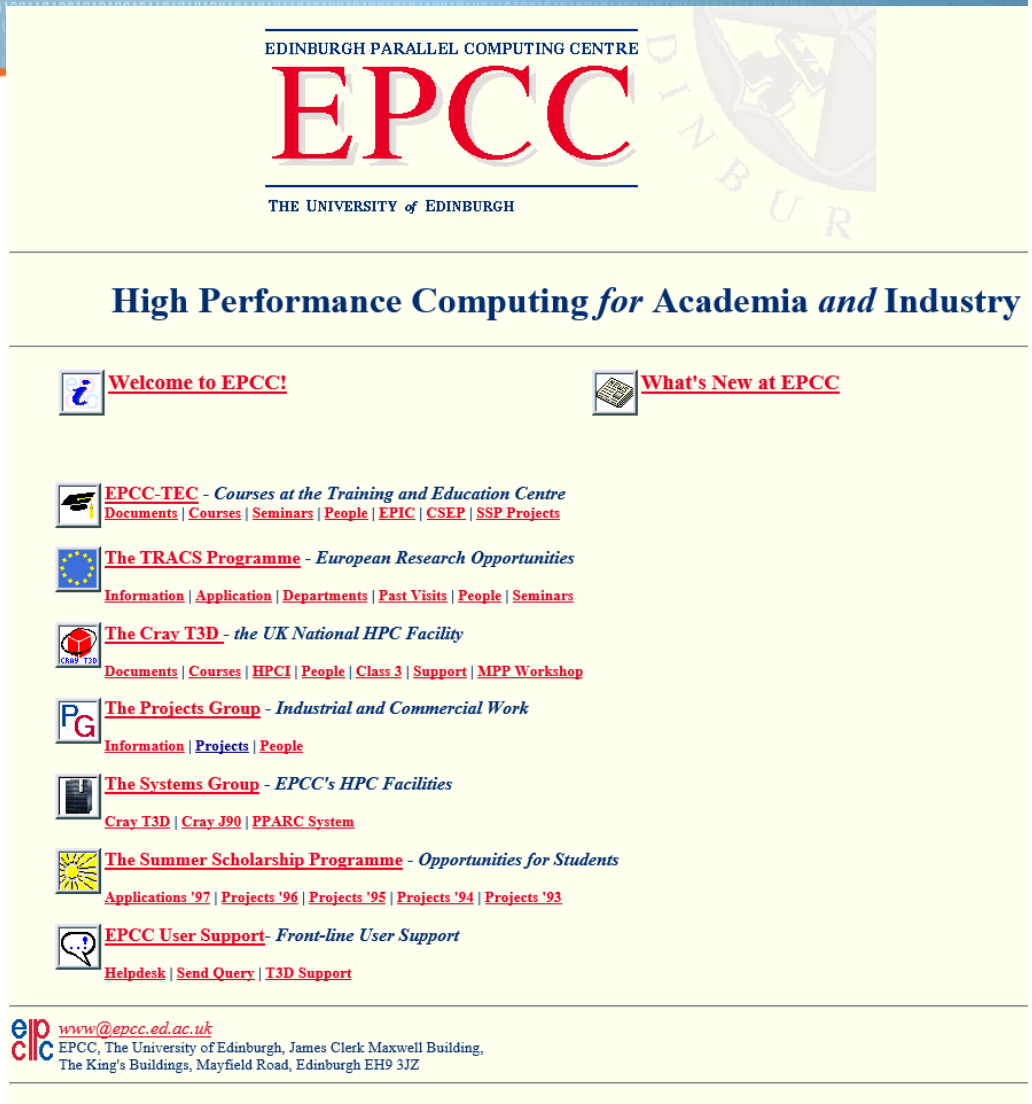
- UK National HPC centre at the University of Edinburgh
- 90 staff - Majority are software developers working with academia and industry
- National HPC and Big Data services - ARCHER & RDF : 3500 users
- Various research & project specific services
- Broad range of activities



SME HPC Adoption Programme in Europe

Industry & EPCC

- Cycle sales and storage
  - Cirrus
    - *10K+cores*
    - *400+TB storage*
- Training
- Consultancy & Expertise
- Projects
  - Directly funded
  - Locally supported (SCS)
  - European



PRACE SHAPE
SME HPC Adoption Programme in Europe

Small and Medium Enterprises

- SMEs are the backbone of the European Economy
  - 99/100 businesses are SMEs
  - In 2015, just under 23 million SMEs generated €3.9 trillion in value added and employed 90 million people (source: EC SME Performance Review 2015/2016)
- How many can realistically benefit from HPC?
  - In 2008 a study in Scotland found around 1,400 companies *could* use HPC, with 300 of these *likely* to
- Scaling across Europe indicates around 142,000 potential SME HPC users and 30,000 that are likely to

SMEs and HPC: opportunities

- Enable development of new products
- Reduce time-to-market and R&D costs
- Increase quality of services and products
- Reduce manufacturing costs
- Innovation
- Next five years will see HPC and Data Analytics systems merging (HPDA)
  - Great opportunities for HPC in new sectors

SMEs and HPC: barriers

- Cost of operation
  - A first project can easily cost €60K-100K
  - This is a lot of money for an SME
- Lack of expertise
  - No in-house experience
  - Cost of consultancy
- Lack of resources
  - No access to appropriate hardware
- Risk!
  - If never used HPC before, high risk, but … potentially high benefit

SHAPE: SME HPC Adoption Programme in Europe

- Supported by PRACE (Partnership for Advanced Computing in Europe)
  - launched in 2013
- Programme to raise awareness and assist SMEs in taking advantage of HPC
  - aimed primarily at HPC “first-timers”

- Regular calls for applications, which are judged on
  - Strength of business case
  - Achievability, commitment from the SME and innovation
  - Social and economic impact for society

SHAPE
SME HPC Adoption Programme in Europe

SHAPE process

- SME applies to a call (6 monthly)
  - Online form to complete
  - Prior detailed knowledge of HPC not required
  - Advice available from PRACE centres at this stage
- Applications reviewed
  - Mixed panel – PRACE board, industry and centre representatives …
- Successful applicants get
  - Effort from a PRACE partner expert (typically 2-6 months)
  - Machine time on a PRACE machine
- In return, the SME
  - Supplies domain expertise and in-kind effort to work with the PRACE expert
  - Produces a white paper, publicity, feedback etc

SHAPE
SME HPC Adoption Programme in Europe

SHAPE projects

- To date, 33 SMEs have been helped
  - From 9 different countries
  - 5th call ongoing, so more in the pipeline

Pharmacelera™

VORTEX Bladeless

CREO DYNAMICS

BAC ENGINEERING CONSULTANCY GROUP

OPTIMAD

airinnova
Engineering Solutions and Innovations

JUAN YACHT DESIGN
Juan Kouyoumdjian · Naval architecture

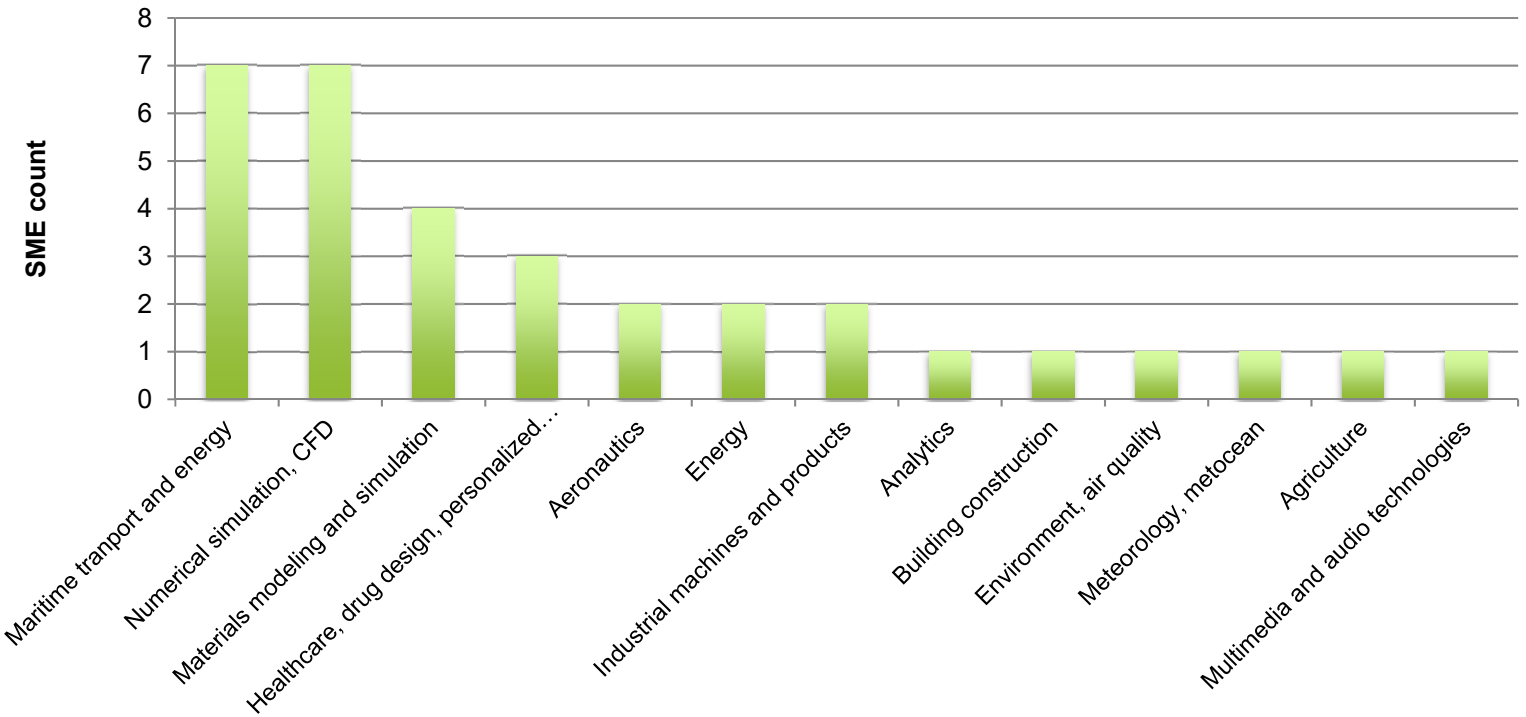
DESIGN METHODS™

Algo'Tech INFORMATIQUE

nsilico
SIMPLIFYING SCIENTIFIC PROCESSES

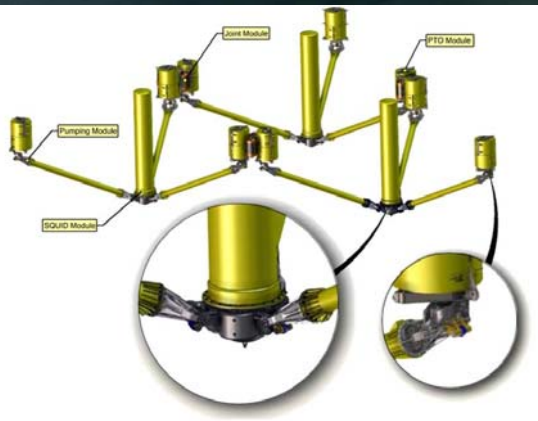
PRACE SHAPE
SME HPC Adoption Programme in Europe

SHAPE SME Industry Domains



SHAPE success stories

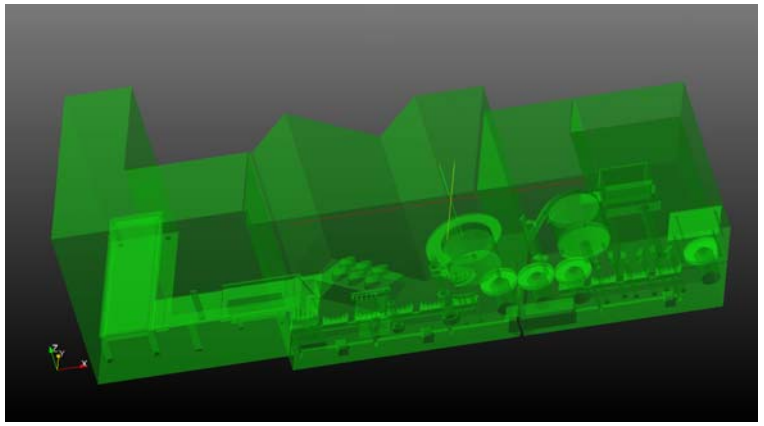
- AlbaTERN: modelling of innovative wave power generation devices
  - Gained expertise to expand their simulation capability to large km-wide device arrays



PRACE SHAPE
SME HPC Adoption Programme in Europe

SHAPE success stories

- Optima PHARMA: airflow simulations
  - Used OpenFOAM to simulate airflow in “clean rooms” used in various industries, particularly for pharmaceutical products



- More at http://www.prace-ri.eu/hpc-access/shape-programme

SME HPC Adoption Programme in Europe

SHAPE measurements

- Follow-up with SMEs post project
- Return on Investment
  - Jobs created
  - Reduced costs
  - Contracts won
  - Commitment to continue working with HPC (in-house and/or via third party)
- Intangible measures
  - Optimism that improved service/software solution will lead to an increase in customers
  - R & D accelerated

SHAPE
SME HPC Adoption Programme in Europe

SMEs: Industry Relevant Training

- Challenges from a PRACE perspective
  - Often require specific packages eg fluent, OpenFOAM
    - *Not usually what HPC centres do*
    - *Anti-competition*
    - *License/access issues*
- Example of a solution
  - As a PRACE training centre, EPCC hosted a CFD packages course
    - *Three different vendors*
    - *Trainers came to site*

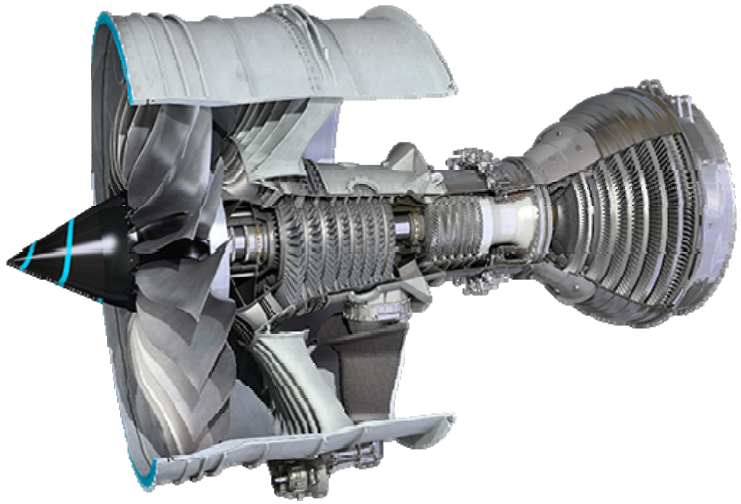
SMEs: training

- SMEs can (and do) attend standard courses run by centres
- @EPCC, courses include (many from our MSc programmes)
  - Intro to HPC, MPI, OpenMP, Software Engineering, Data Science, GPGPU …
- ARCHER training courses, of order £100s/day
- Outreach is a challenge
  - Generally course attendees already known to the centre
- Bespoke offerings
  - Very time-consuming
  - Too specific, limiting re-use opportunities
- Distance learning
  - Great opportunity
  - May fit in better with SMEs requirements i.e. staff availability, costs etc

Training example 1: Codeplay

- Software tools developer – approx 40 staff
- They approached EPCC
- Requested OpenMP course
  - Time of multicore generation of consoles
  - Few changes required from existing course
- Course delivered at their site
  - Suitable facilities
  - Paid for course, travel/accommodation
  - 2 –day course, 2 EPCC staff
- Time built-in for consultancy

Training example 2: Rolls Royce



- Not an SME!
- They approached EPCC – projects
- Small sub-team (effectively an SME …)
  - Developed thermal modelling package
  - Sold externally
- Requested OpenMP course
  - Delivered on site
  - 1 staff member, 10 attendees, 2 day course

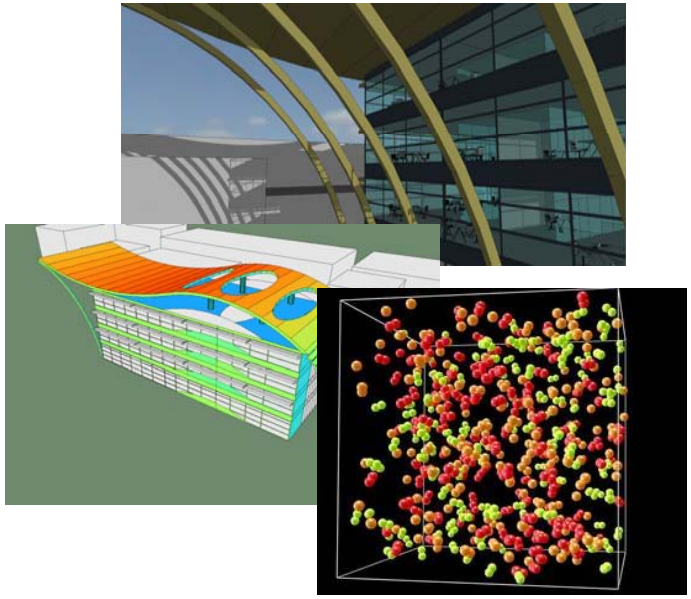
Training example 3: RSPB

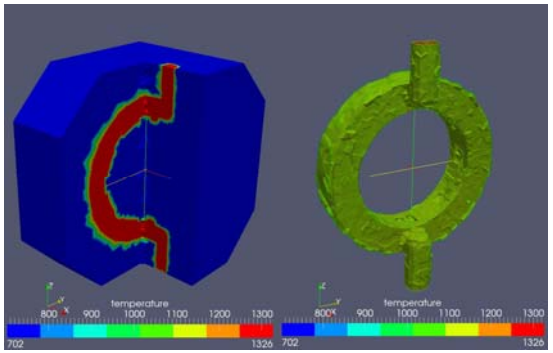
- Charitable organisation
- Heard of EPCC through uni outreach programme
- Bespoke course: Parallel R & general HPC overview
  - Based on existing MSc material
  - Approx 1 week effort to write course
  - Produced 1 day course
- Course delivered twice, at two different RSPB sites
- Not “large-scale” HPC
  - users wanted to exploit multicore PCs for their modelling of environmental factors and habitats

Conclusion

- SMEs are key to a thriving economy

- Many SMEs would benefit from using HPC
  - Reduction in costs, time to market, etc
- There are barriers to this …
  - Initial cost, risk, access…
- The SHAPE programme allows SMEs to overcome these barriers
  - A “proof of concept” exercise can be tried out via SHAPE
- Training is an important aspect
  - Providers need to be flexible
  - Easiest to build on existing material
  - Online/distance learning opportunities



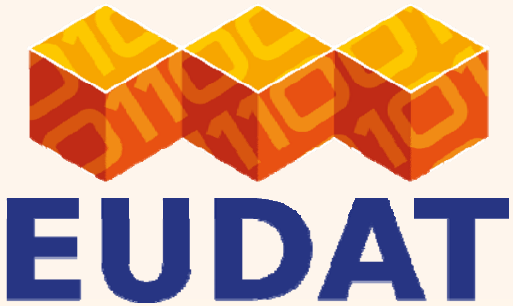


Thanks - Any questions

- SHAPE: http://www.prace-ri.eu/hpc-access/shape-programme/

- Me: Paul Graham p.graham@epcc.ed.ac.uk

PRACE SHAPE
SME HPC Adoption Programme in Europe



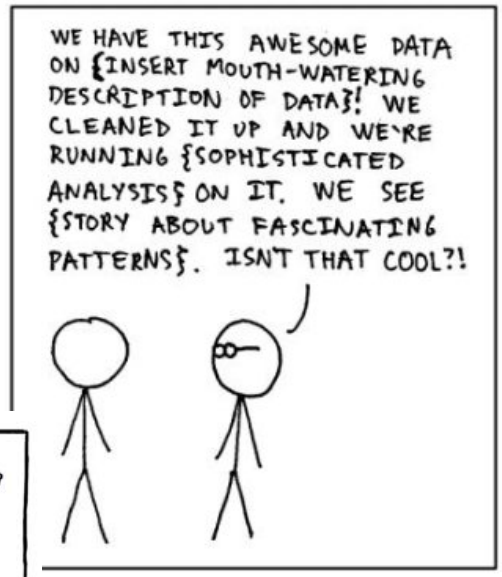
On the HPC convergence with Data Science and how PRACE and EUDAT training can address it

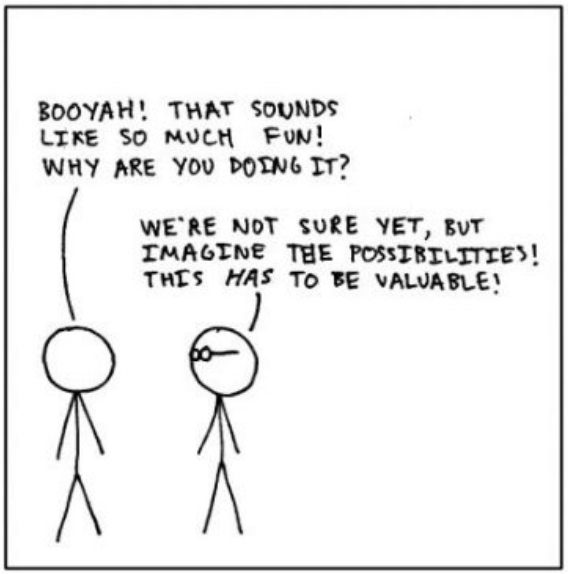
Shaun de Witt (CCFE)
Giuseppe Fiameni (CINECA), Stephane Coutin (CINES)

ISC2017, June 22nd

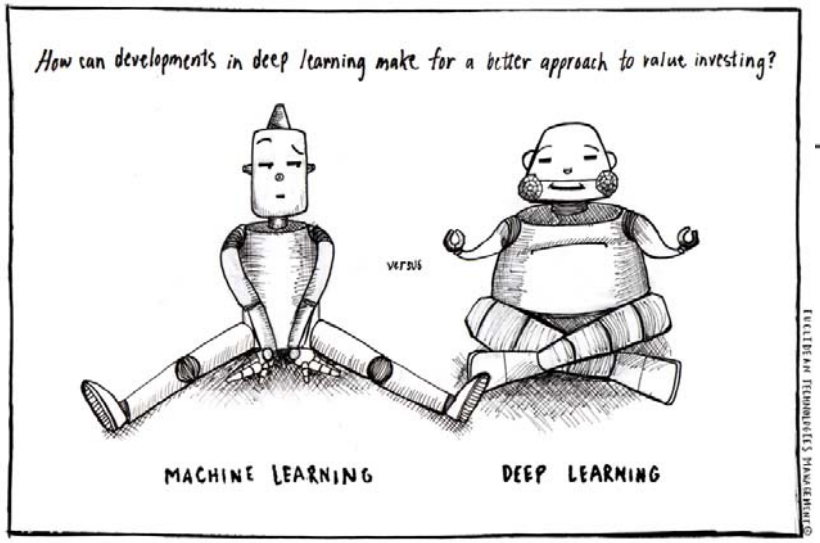
www.eudat.eu

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065



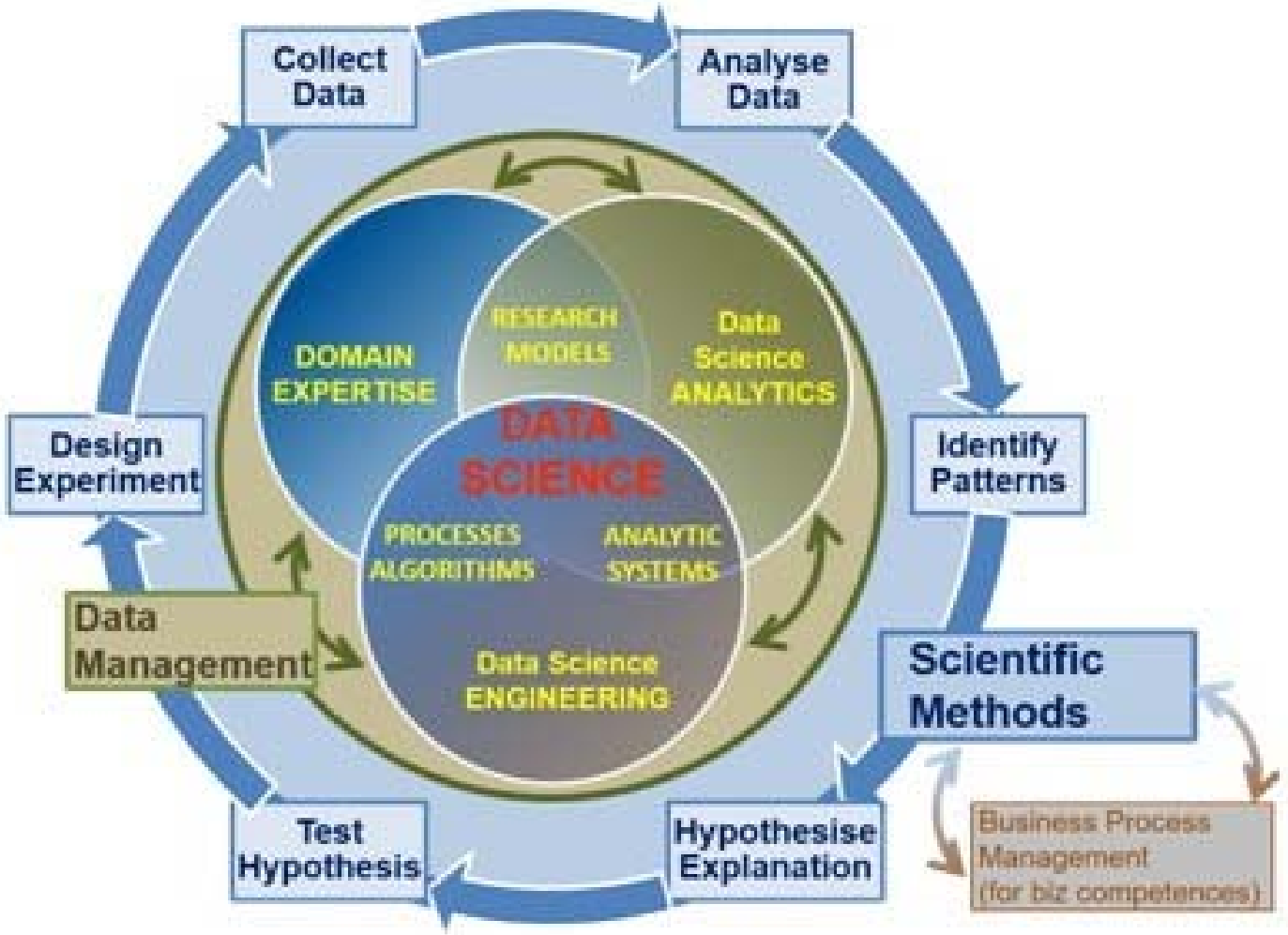


© xkcd



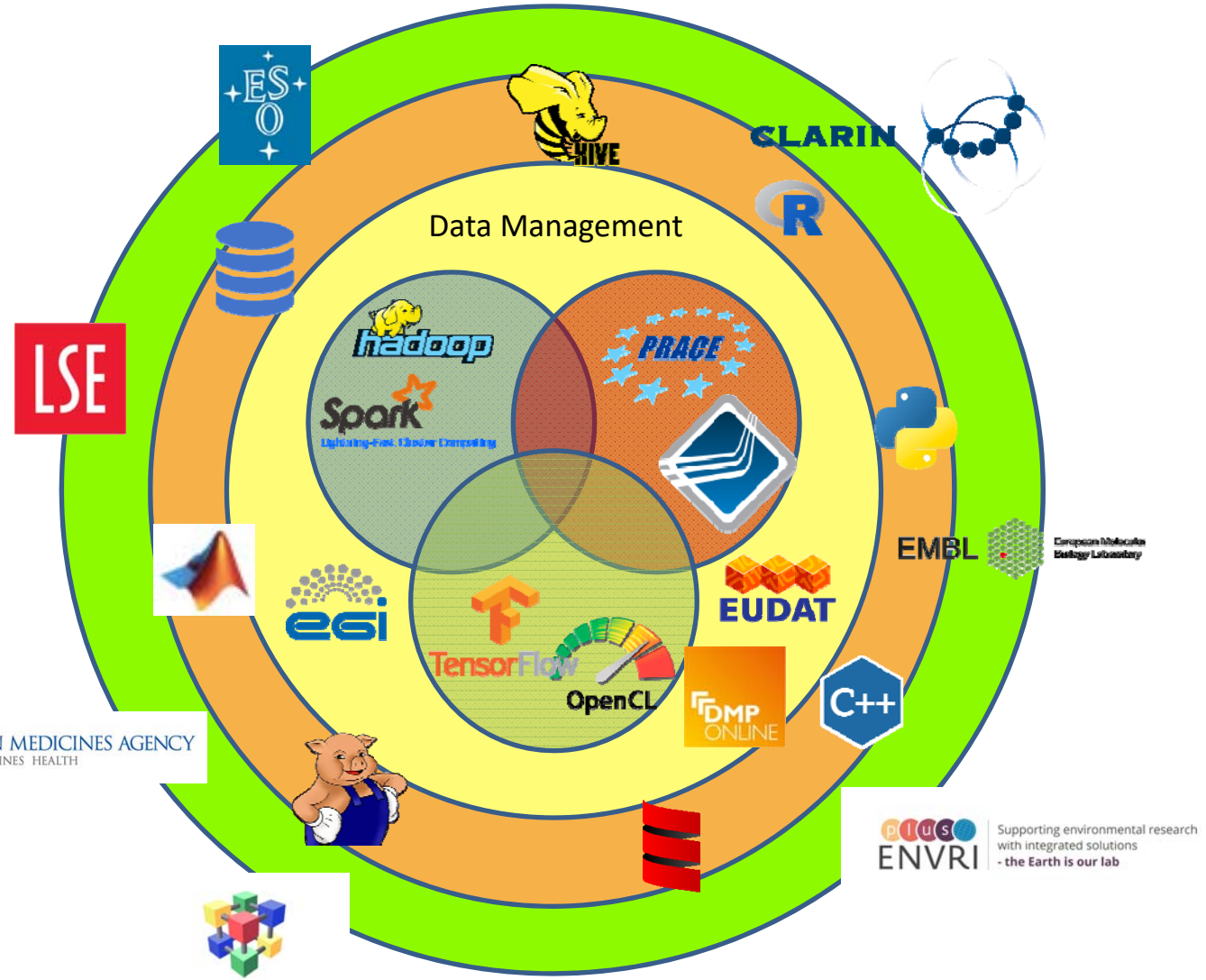
THE DATA SCIENTIST

Skills (Research Oriented)

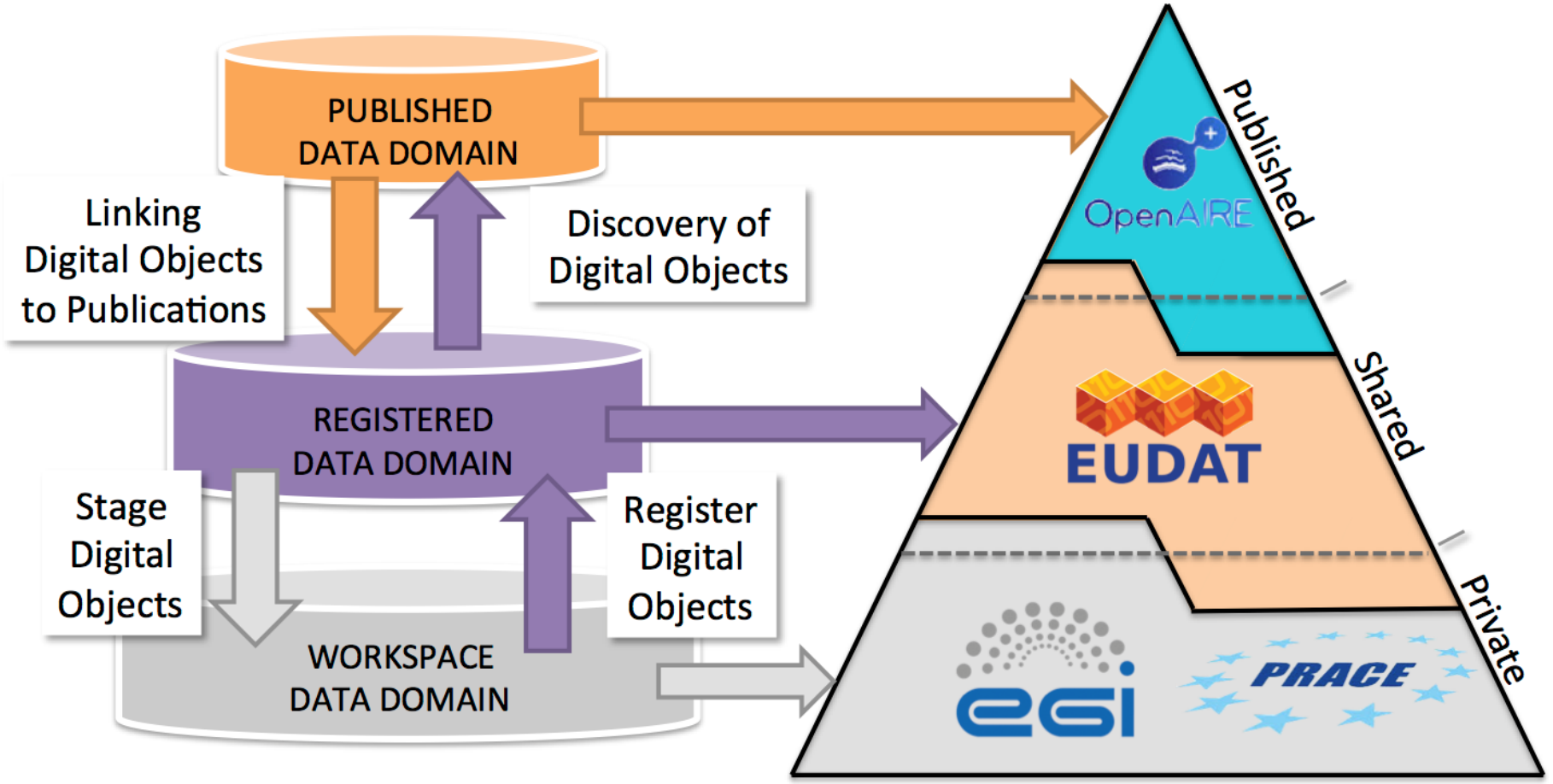


From EDISON Data Science Framework

The Computing Version



Data Domains



EUDAT Data Domain modeled on the ANDS[1] Data Curation Continuum

1. Australian National Data Service organization – www.ands.org.au

EUDAT
The pan-European Data Infrastructure

www.eudat.eu

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065

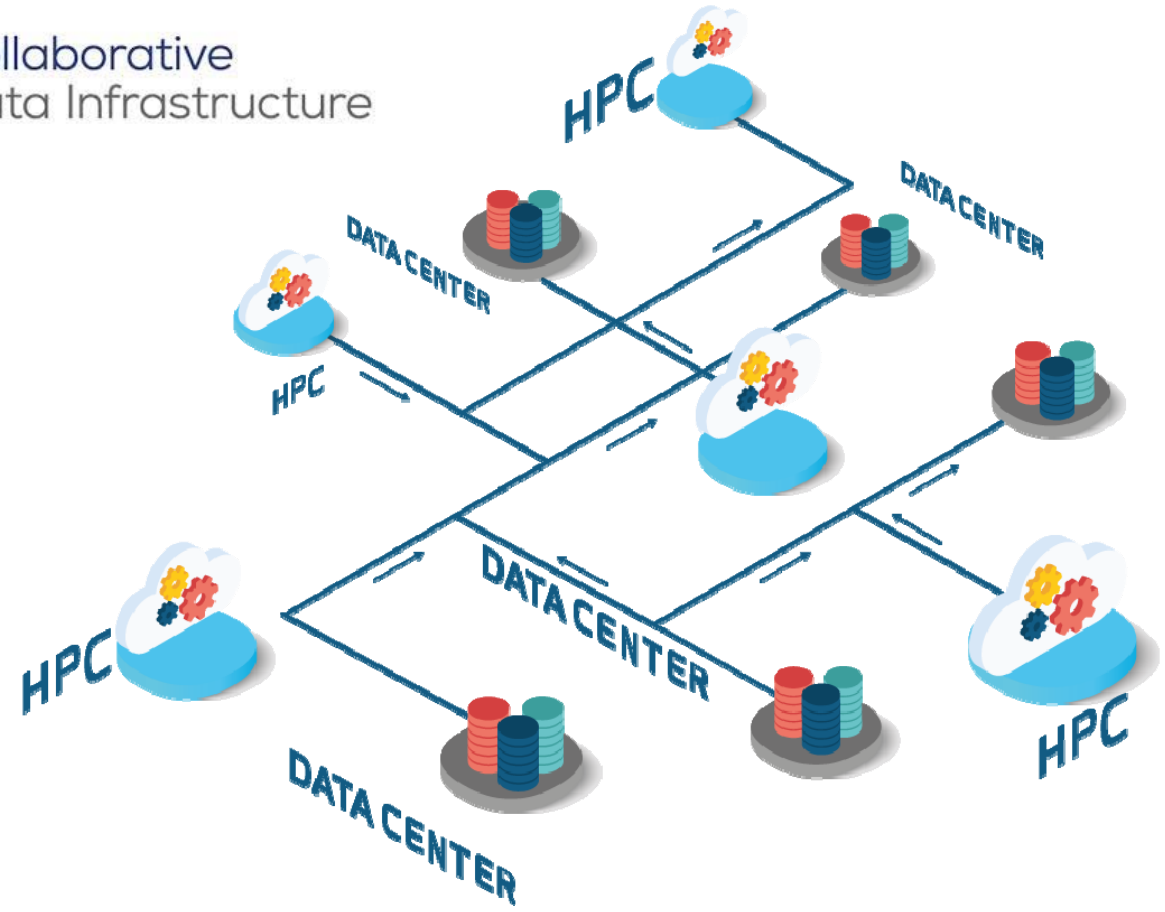
EUDAT2020 - 35 Partners

Community-Driven Solutions

EUDAT services are **designed, built and implemented** based on **user community requirements.**

The CDI – A Service Infrastructure

EUDAT Collaborative Data Infrastructure



10 year initial commitment

A landmark agreement sustaining the pan European Collaborative Data Infrastructure for the next 10 years

3rd October 2016

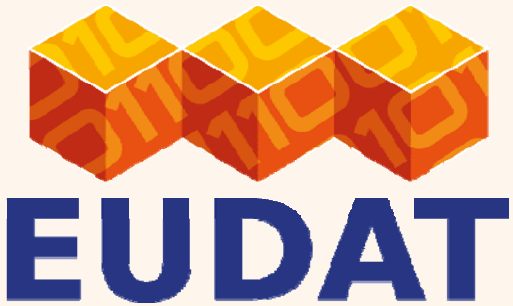
16 major European research organisations, data and computing centres *signed an agreement to sustain the EUDAT – pan European collaborative data infrastructure for the next 10 years. The organisations stand together behind a long term sustainability plan and commit to develop, maintain and deploy pan-European research data services and to promote harmonization of research data management practices across centres.*

What does this mean?

Service providers, both generic and thematic, and research communities have joined forces as part of a common framework for developing an interoperable layer of common data services. Known as the EUDAT Collaborative Data Infrastructure (CDI), this is essentially **a European e-infrastructure of integrated data services and resources** to support research. This infrastructure and its services have been developed in close collaboration with **over 50 research communities** spanning across many different scientific disciplines and involved at all stage of the design process. The establishment of the EUDAT CDI is timely with the imminent realization of the European Open Science Cloud[1] which aims to offer open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines.

What benefits will this bring to researchers?

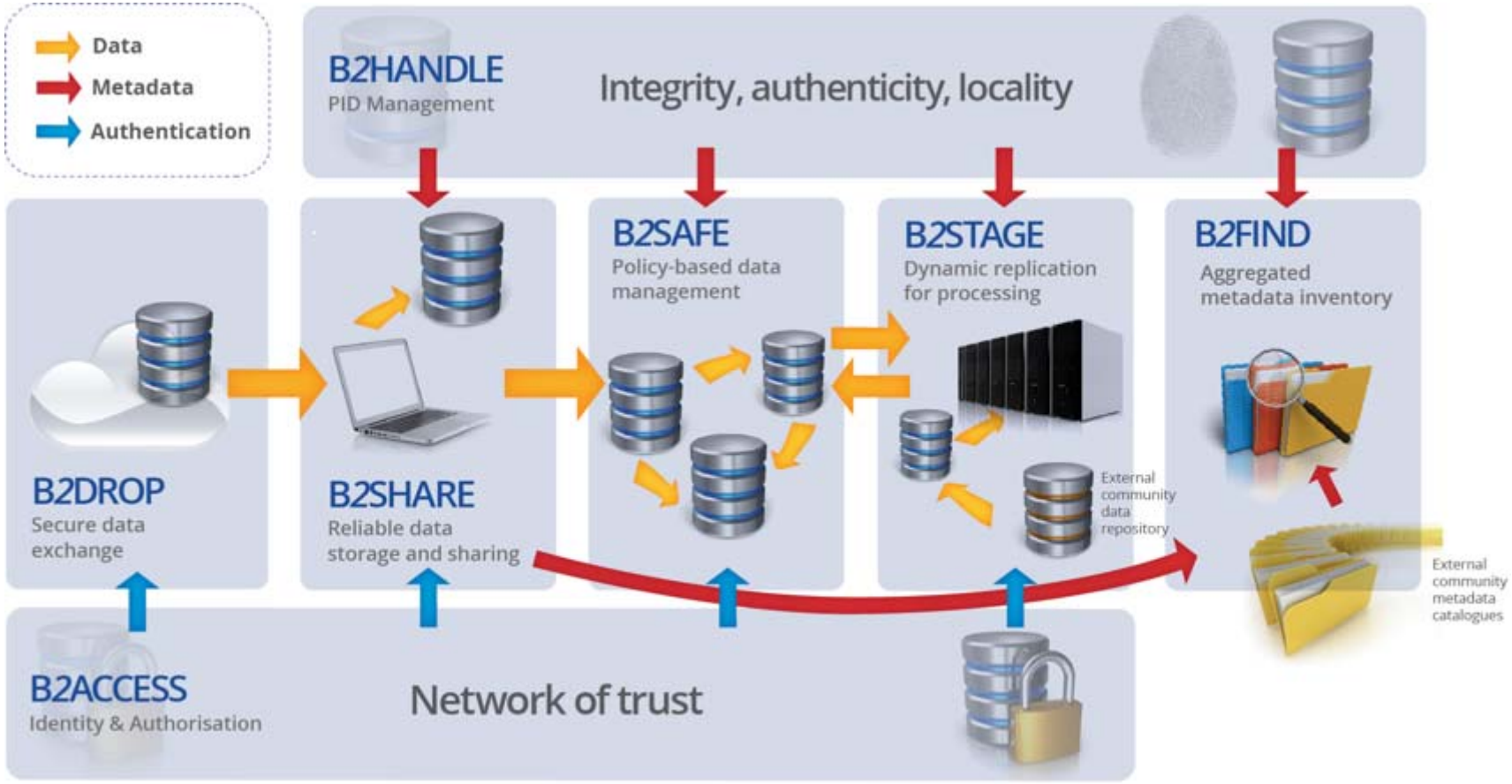
Researchers can rely on **innovative data services** to support their research collaboration and data management. Additionally they benefit from a common service management framework delivered by CDI service providers and the connection between sites. These 16 organizations[2] from 11 countries with a further 4[3] in the process of joining, are European leaders in the deployment of data management and data infrastructure services. By joining the EUDAT CDI they are able to better **serve their local customers across borders and support cross-national collaborations**.



EUDAT B2Service Suite

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065

What is the EUDAT Service offer?



B2FIND — Find Research Data

B2STAGE — Get Data to Computation

B2HANDLE — Register your Research Data

B2SHARE — Store and Share Research Data

B2SAFE — Replicate Research Data Safely

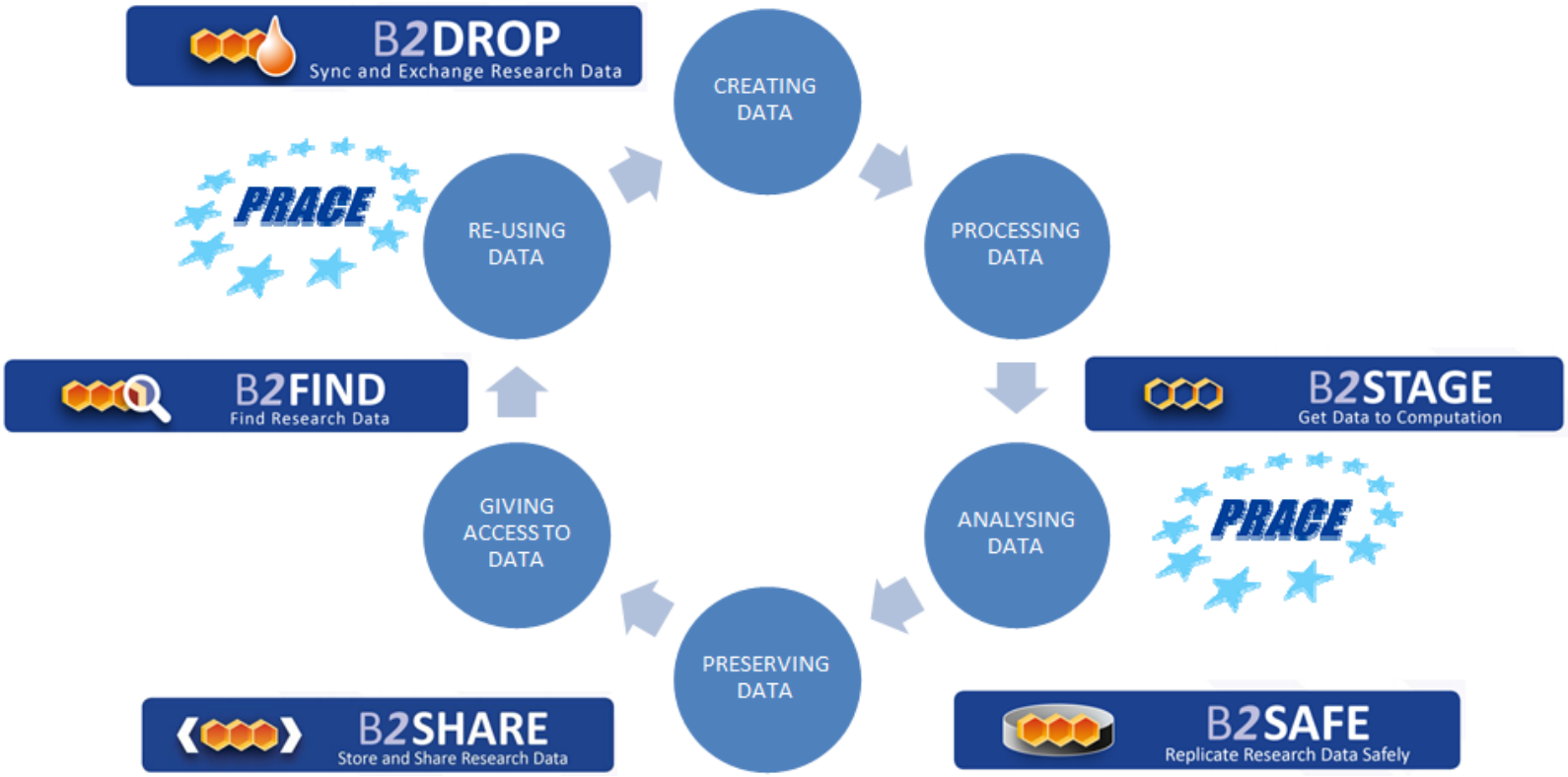
B2DROP — Sync and Exchange Research Data

B2ACCESS — Identity & Authorisation

B2NOTE

DATA TYPE REGISTRY

Addressing the full lifecycle of research data



eudat.eu/training/research-data-management

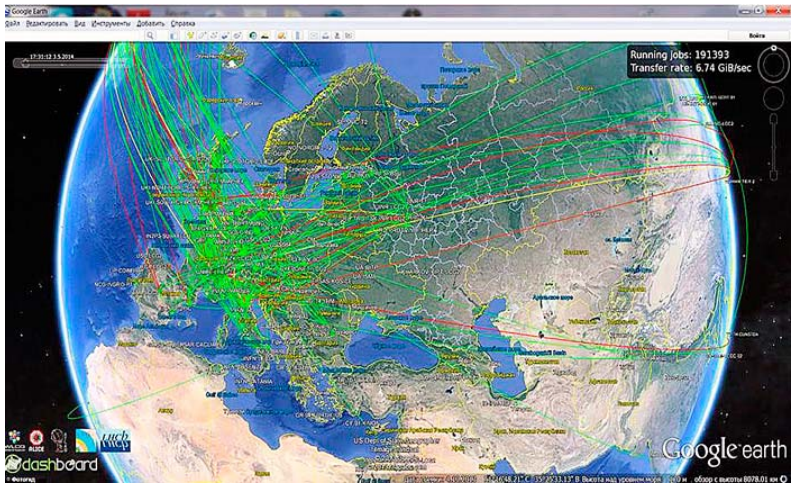
EUDAT – Making things Easier

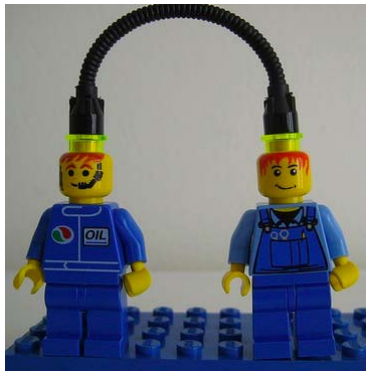
- More than just a GUI
- Rich set of services enabling you to share, protect and preserve your data
- HTTP APIs to most service
- Python library under test
- Fast Data Transfer Service
- Subscription service (under development)

DATA MANAGEMENT PLANNING

Plan & Protect Your Data

- Data Loss
- Integrity Loss
- Knowledge Loss
- Who, What, Where, When



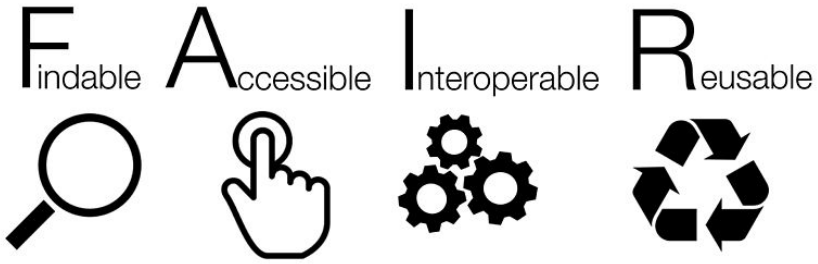


Key Elements

- **Good** Metadata
  - Defined ontologies and data types
  - Human undersrandable
  - ‘Metadata is your gift to the future’
- **Well thought-out** Data Management Plan
  - DMP Online (https://dmponline.dcc.ac.uk/)
  - Data Mangement is not just about preservation
- **Pesistent** Identifier
  - PID, pURL, DOI, URN, …
  - location independence
  - citability

(H2020) Open Research Data (Pilot)

- Already expanded from a select pilot to **all** work areas
- All need to consider which data can be made open
- Mantra = “As open as possible as closed as necessary”
- Underlying driver is good (FAIR) data management



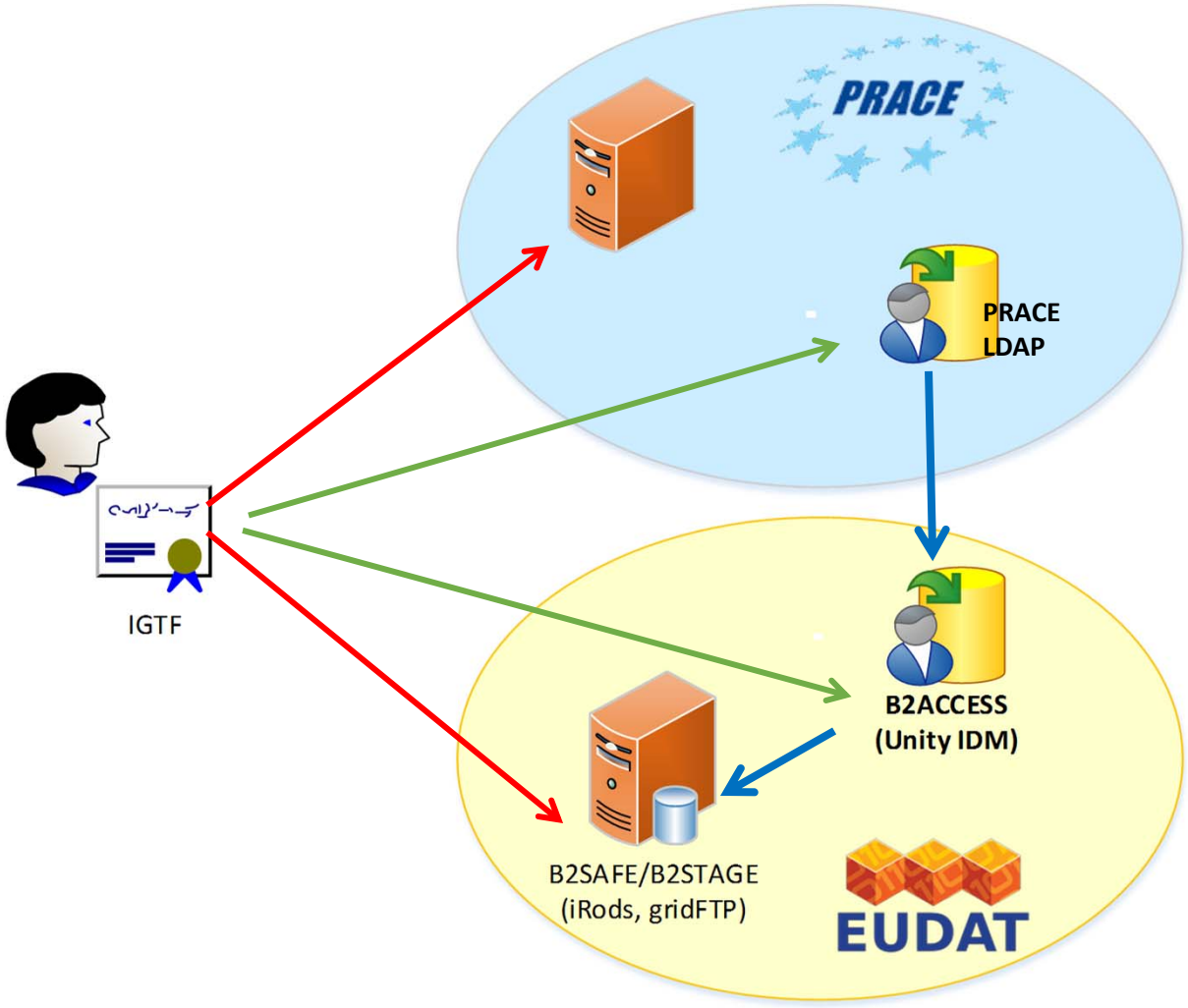
COLLABORATION WITH PRACE

Coupling HPC & Data Resources

- **PRACE** – *coupling PRACE HPC resources to EUDAT storage resources.*
  - PRACE & EUDAT infrastructures are now fully interoperable from the technical point of view
  - EUDAT has contributed to three PRACE Calls for proposals (Tier0 Call 11, Tier1 Call 13, Tier1 Call 14)) by offering data services and storage resources to interested PRACE users.
  - Users now have the possibility to apply for EUDAT services to preserve their data throughout and beyond the lifecycle of their PRACE computational grant!

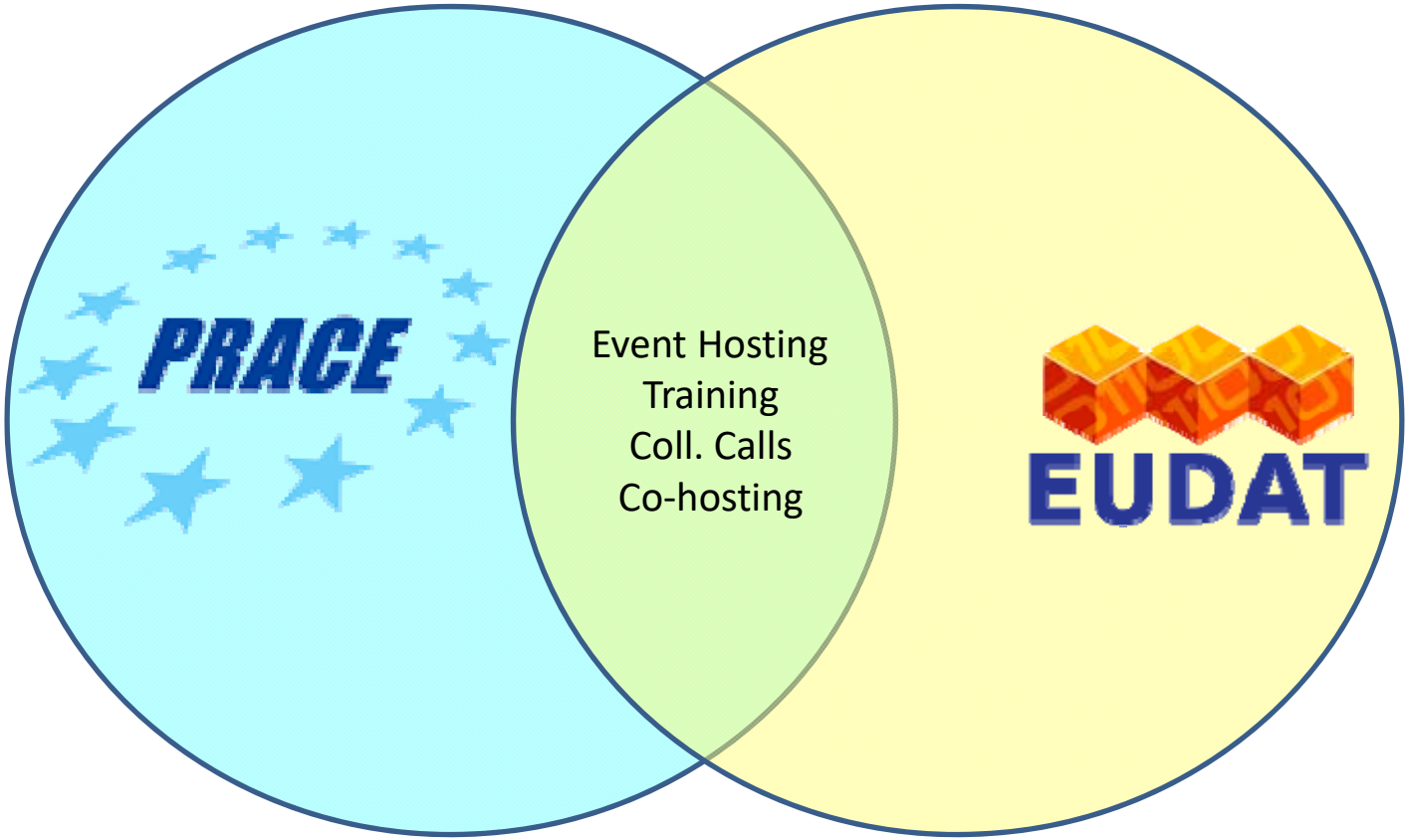
Acronym	Title	Field	Country	Data requirements in TB during the PRACE project	Data requirements in TB after the PRACE project	Duration of the access to this data service after the PRACE allocation	PRACE Site	EUDAT Site
HybTurb3D	Hybrid 3D simulations of turbulence and kinetic instabilities at ion scales in the expanding solar wind	Astro Sciences	IT	140 TB	140 TB	24 months	**SurfSARA**	**CINECA**
MULTINANO	Multiscale simulations of nanoparticle suspensions	Engineering	IT	30 TB	30 TB	24 months	**MPCD**	**CINECA**
~~**HiResClimate**~~	~~High Resolution EC-Earth Simulations~~	~~Earth Sciences~~	~~IE~~	~~150TB~~	~~150TB~~	~~12 months~~	~~**KTH**~~	~~**EPCC**~~
AFiD	Effect of rotation and surface roughness on heat transport in turbulent flow	Engineering	NL	11TB	10TB for 24 months 1TB for long-term storage and publication	10TB for 24 months 1TB for long-term storage and publication	**EPCC**	**SurfSARA**
CHARTERED	Charge transfer dynamics by time dependent density functional theory	Materials Science	SE	30TB	30TB	24 months	**IT4I**	**KTH/PDC**
				Total	**448TB**			

Users credential synchronization



1. PRACE LDAP -> B2ACCESS synchronisation
2. B2ACCESS -> B2STAGE synchronisation
3. Access to web based EUDAT services

Working Together for the Data Scientist



For more info:

https://b2drop.eudat.eu
https://eudat.eu/services/userdoc/b2drop

https://b2share.eudat.eu
https://eudat.eu/services/userdoc/b2share

https://eudat.eu/services/userdoc/b2safe

https://eudat.eu/services/userdoc/b2stage

http://b2find.eudat.eu
https://eudat.eu/services/userdoc/b2find

http://b2handle.eudat.eu

http://b2access.eudat.eu

Credits

- Damien Lecarpentier (CSC) - damien.lecarpentier@csc.fi
- Mark van de Sanden (SURFsara) - sanden@sara.nl
- Johannes Reetz (MPDCF) - johannes.reetz@mpcdf.mpg.de
- Roberto Mucci (CINECA) - r.mucci@CINECA.IT
- Claudio Cacciari (CINECA) - c.cacciari@cineca.it
- Massimo Fares (INGV) - massimo.fares@ingv.it
- Giovanni Morelli (CINECA) - g.morelli@cineca.IT
- Stéphane Coutin (CINES) - coutin@cines.fr
- Paolo D'Onorio De Meo (CINECA) - p.donoriodemeo@cineca.it


EUDAT

PRACE Training

Objectives

- Organise training events supporting the needs of users and the CoEs
- PATC training and improve the network even further (PTCs)
- Expand on-line training opportunities

Outline

- PRACE Training Events
- PRACE Training Portal
- PRACE Events Website
- MOOCs

PRACE Training Events

- PRACE Seasonal Schools
  - Seasonal Schools take place throughout Europe in the Winter, Autumn and Spring.
  - For more information: https://events.prace-ri.eu/category/6/
- International HPC Summer school
  - In collaboration with XSEDE, RIKEN, Compute Canada
- PRACE Training Centers (PATCs) Training Program
  - Around 12 Training events per year
  - 4 PRACE Advanced Training Centers
    - IT4I – National Supercomputing Center VSB-Technical Universitz of Ostrava(Czech Republic)
    - GRNET – Greek Research and Technology Network (Greece)
    - ICHEC – Irish Center for High-End Computing (Ireland)
    - SURFsara (The Netherlands)

  For more information: https://events.prace-ri.eu/category/12/

PRACE Training Events

PRACE Training Events

- PRACE Advanced Training Centers (PATCs) Training Program
  - Around 80 Training events per year
  - 6 PRACE Advanced Training Centers
    - CSC - IT Center for Science (Finland)
    - MdS - Maison de la Simulation (France)
    - CINECA - Consorzio Interuniversitario (Italy)
    - GCS - Gauss Supercomputing Center (Germany)
    - BSC - Barcelona Supercomputing Center (Spain)
    - EPCC at the University of Edinburgh (UK)
- PATC and PTC 2017-2018 Tentative training programme available
  - Please pick up a copy

For more information: https://events.prace-ri.eu/category/2/

PRACE Training Portal

- http://www.training.prace-ri.eu/
- Platform to make publicly available training material from PRACE Training Events
  - Hundreds of training material documents
- Advertises forthcoming PRACE Training Events
- Videos Tutorials
  - http://www.training.prace-ri.eu/tutorials/index.html
  - Short videos on specific subjects such as MPI, OpenMP, PETSc
- Links to PRACE partner portals, “Best Practise Guides”

PRACE Training Portal - Training Material

- http://www.training.prace-ri.eu/material/index.html
  - Can be found towards end of page
- Training material of over 150 PRACE Training Events
  - Includes hands on exercises
  - Some of these have videos of the lectures also.
    - Marked with “Video Lectures” in the title
- Most of material tagged and categorised too

New Training Portal

- Completely new system
  - Clean / Better UX/ integration with Events, Codevault portal
- New Categorisation
  - Update of categories
  - Relevant scientific methods
- Tags
- More material being made available
- Video recordings of lectures too
- More metadata, detailed search

PRACE Events Website

- http://events.prace-ri.eu/
- Advertises forthcoming PRACE Training Events
  - PATC Courses
    - https://events.prace-ri.eu/category/2/
  - PATC Courses
    - https://events.prace-ri.eu/category/12/
  - PRACE Seasonal schools
    - https://events.prace-ri.eu/category/6/
- Each post provides full details of the forthcoming event
  - Allows visitors to register for events

Massive Open Online Courses

- Two available MOOCs
  - Available on FutureLearn
  - Supercomputing
    - https://www.futurelearn.com/courses/supercomputing
  - Managing Big Data with R and Hadoop
    - https://www.futurelearn.com/courses/big-data-r-hadoop
- Further to be available soon

Thanks for your attention

Questions?