Teaching High Throughput Computing

eResearchAustralasia, October, 2017

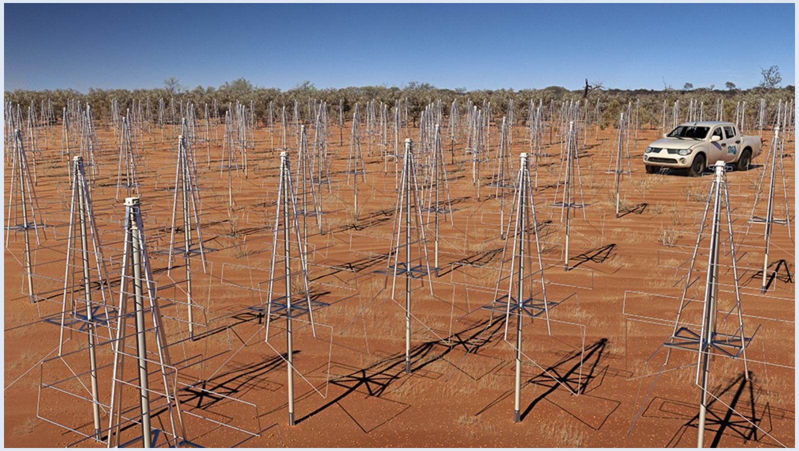
Lev Lafayette, University of Melbourne <llafayette@unimelb.edu.au>
Anja Gerbes, Goethe University Frankfurt <gerbes@csc.uni-frankfurt.de>

1.0 The Importance of High Throughput Computing

1.1 High Performance/Throughput Computing is of increasing importance. Datasets are growing larger than desktop systems can process them. Square Kilometre Array is an extreme example of a general problem.

1.2 Distinction between HPC and HTC exists because performance does not always correlate with throughput due to opportunity costs. Money spent on high-speed interconnect could be used to purchase more compute cores, which should be proportional to the number of single-node jobs being run.

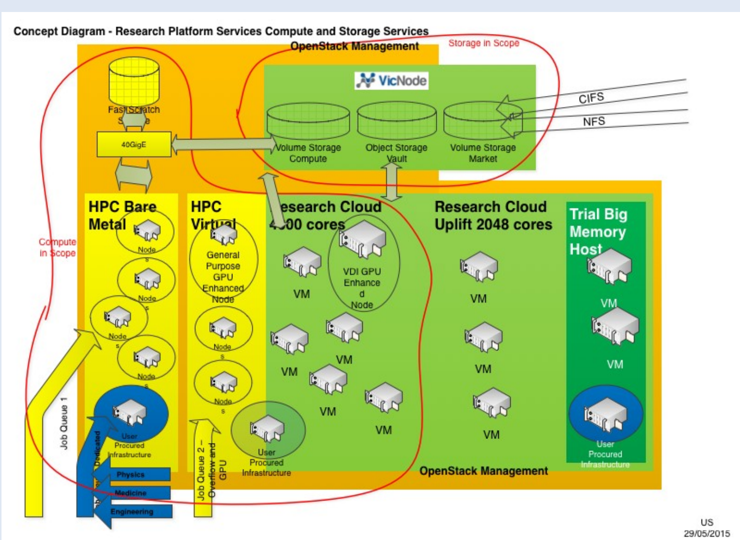
1.3 Nascent research indicates that research output correlates strongly with availability of HPC/HTC resources. Whilst physical provision must come first, user education is also a necessary component. Evidence shows that the provision of training material has a significant effect on HPC usage.

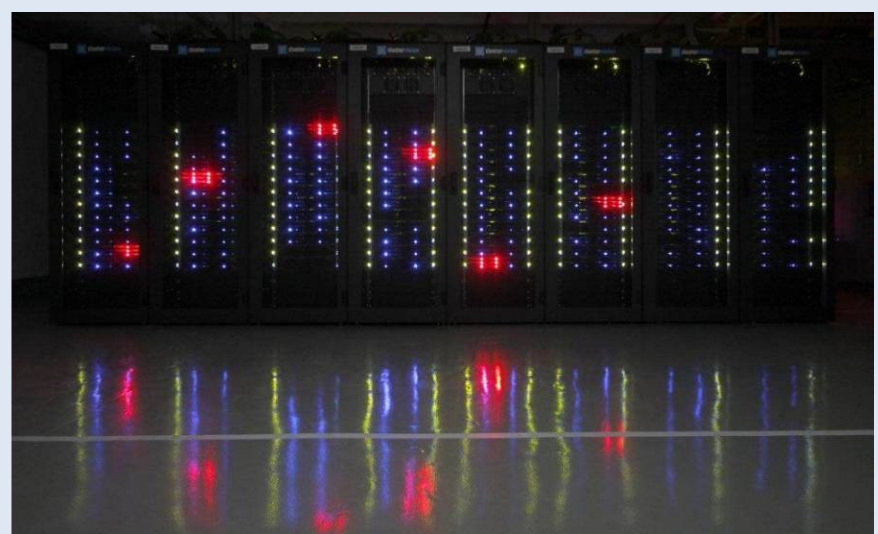


2.0 University of Melbourne and Goethe University Frankfurt

2.1 University of Melbourne has a small HTC system ("Spartan"), with a small traditional HPC system and a larger cloud partition based on Haswell and Broadwell with approximately 4,000 cores in total. Massive GPGPU partition just being introduced which will bring the system into the Top500.

2.2 Goethe University Frankfurt has two Linux-based computer clusters FUCHS, and LOEWE-CSC. These systems with approximately 40,000 cores (FUCHS-CSC), and 170,000 cores (LOWE-CSC) with Opteron, Ivy Bridge, Broadwell, and GPU processors. LOWE-CSC was one the Top 500 public machines in the world from 2010-2016; refresh due in 2017-2018.





3.0 The Situation of HTC Researchers

3.1 Most researchers do not have formal teaching in HTC skills prior to need (Linux command line, HPC job submission). Only a handful of education institutions include HTC systems utilisation or parallel programming in the science curriculum. At the point of need, researchers need also to become information scientists in addition their primary area of research.

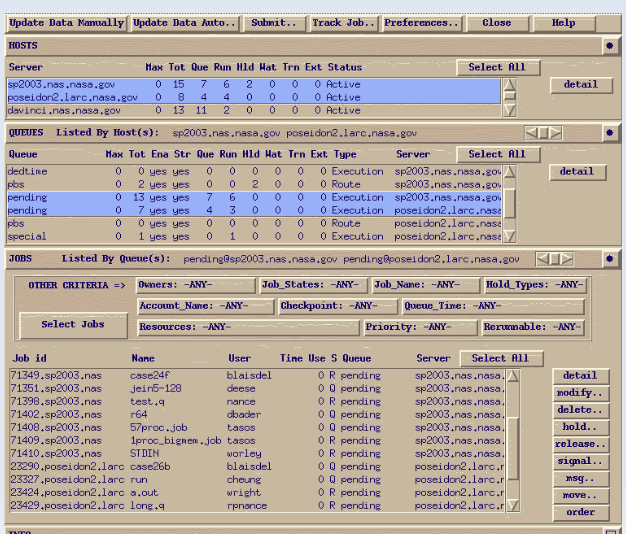
3.2 Fortunately most researchers are competent learners and can pick up new subject-matter quickly if delivered appropriately. A day's training is sufficient to introduce researchers to the concepts and practise of command-line Linux and job submission. Another day for shell scripting for HTC job submission; a day for the core concepts of parallel programming (multithreaded and message passing) and so forth. This follows the proposals of the "software carpentry" response to the skills-gap in scientific computing.

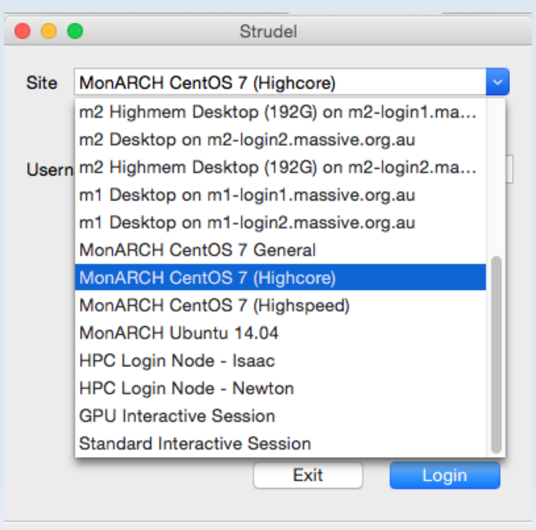
software carpentry

4.0. Interface Improvements or Skill Improvements

4.1 The main methods for improving eResearch computational ability consist of developing the skillset among users to use the existing tools, or modifying the existing tools to fit the existing skillset. Much effort has been invested in the latter (e.g., grid computing interfaces, distributed computing installers, web portals etc), a review suggests that only in a limited number of cases that are orientated towards a non-research use (e.g., folding@home) or with relatively simple submission environments (e.g., Monash eResearch's STRUDEL - ScienTific Remote User Desktop Launcher) shows a combination of usability and uptake.

4.2 The intrinsic level of complexity in the environment and the requirement a grounded understanding of the process limits the capacity for automation and simplification. Without comprehension the eResearcher will be caught in an application relearning cycle. Grounding requires incorporating the core insights of adult and advanced education, including andragogical education.





5.0 The Continuum to Advanced Adult Education

5.1 The adult learner has different characteristics to the child learner; (1) autonomy of direction in learning and (2) the importance of the use of personal experience as a learning resource (3) the emphasis on intrinsic rather than extrinsic motivations. These differences should be considered as a continuum with graduated equilibrium. In the contemporary environment this is supplemented with the notion of lifelong learning.

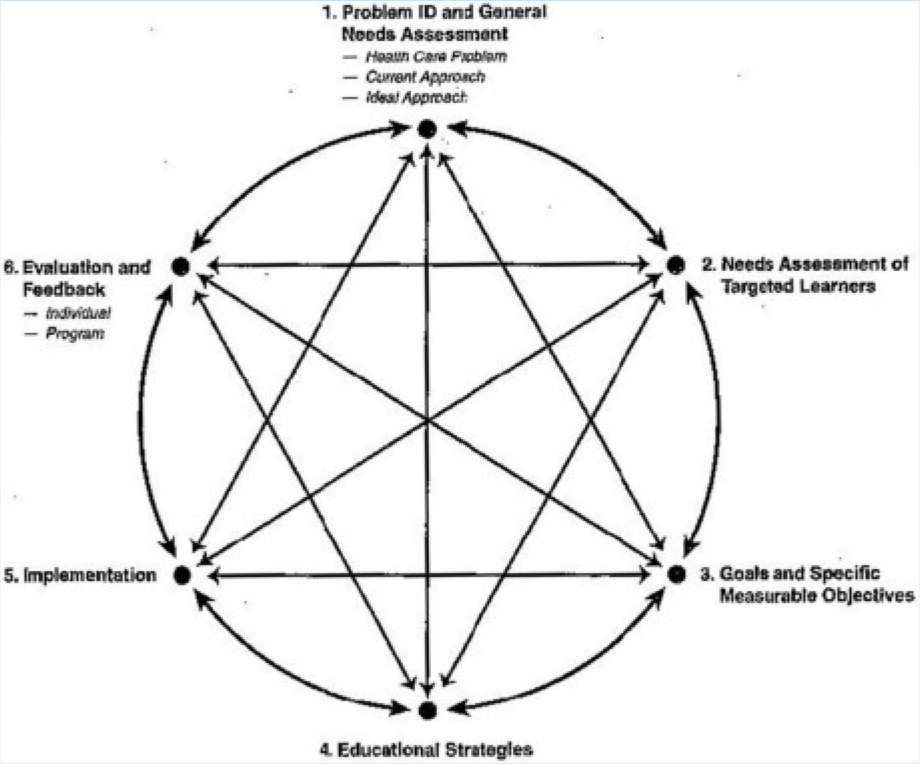
5.2 Content needs to organised in terms of objectives, timed, and revised! Content needs to provided in as modular 'structural knowledge', with narrative, analogies, and humour. Provide grounding to a concept; facts and reasons provides understanding (*Verstehen*) which allows elaboration by the learner.

5.3 Delivery should make use of discipline-based learning styles. For computer use, connectivism (e.g., pair programming) and direct usage ("yield to the hands-on imperative"). Needs to be followed up with anonymous feedback, and proximal learning with a follow-up connectivist mentoring and outreach program.

6.0 Training Programmes: An International Comparison (UoM)

6.1 University of Melbourne runs a number of general and specialist one-day training courses, including Introduction to High Performance Computing (command-line Linux, HPC architecture, job submission options with Slurm), Shell Scripting for High Performance Computing (intermediate CL Linux, regular expressions, bash shell scripts), Introduction to Parallel Programming (multithreading concepts and programming with OpenMP, concepts and programming with core routines for OpenMPI). Courses are designed to have a structured approach in content.

6.2 Specialist courses are typically disciplinary and application specific (e.g., neuroscience research for Orygen Youth Mental Health), but also includes systems administrator's course for the Spartan-GPGPU virtual team. Reviews for general and specialist courses are conducted under Research Platforms based on a simple 1-10 with free-text entries, with courses updated after each session to account for review commentary.



7.0 Training Programmes: An International Comparison (GUF)

7.1 A number of HPC courses are conducted at Goethe University Frankfurt; Introduction to UNIX, Software Tools for UNIX Systems, Introduction to Shell Scripting, Cluster Computing Course, Introduction to C++, Introduction to TotalView Debugger, Introduction to Python. See: `https://csc.uni-frankfurt.de/wiki/doku.php?id=public:workshops`. Courses conducted in English (primarily) and German.

7.2 In addition to the University there are independent institutions like The Frankfurt Institute for Advanced Studies (FIAS), the Max Planck Institute for Brain Research, Max Planck Institute for Biophysics, the Sustainable Architecture for Finance in Europe. Courses are increasingly being orientated towards discipline-specific groups (physics, biology, chemistry, economics, medicine) in both Frankfurt and Hesse. Detailed feedback on design and content for each course.

7.2 Other research centres in Germany (e.g., Dresden, Stuttgart, Munich, Erlangen or Jülich) have HTC courses. The Gauss Allianz centred at Hamburg University conducts an HPC certification program. Further coordination with PRACE (Partnership for Advanced Computing in Europe).

GA Gauß-Allianz

PRACE

8.0 Adult Education Stages and Institutional Opportunities

8.1 An important insight from several years of using andragogical techniques with advanced computer training is the recognition that adult learner components (i.e., autonomy, personal experience, intrinsic motivation) varies significantly within the general status of advanced adult learner. At least part of this can be attributed to age and cultural diversity. Disciplinary diversity is increasingly challenging as researchers may be more familiar with different learning styles.

8.2 Other institutions have the opportunity to make use of the material and experience from the University of Melbourne and Goethe University Frankfurt. University of Melbourne teaching material is all available online at: `https://github.com/UoM-ResPlat-DevOps`.

9.0 Lessons Learned and Future Initiatives

9.1 The independent variation in components suggests that a review of researcher's needs prior to attending classes and bespoke content will have the best possible outcome. Conducting highly granular course content can contribute significantly in this process. Integration of student management systems with learning systems and curriculum content.

9.2 From the University of Melbourne perspective, as significant advantage can be gained by Australian HPC education providers acting in a co-ordinated manner, similar to the PRACE system in Europe, providing coordinated content provision and according to system scale. Tyranny of distance has a significant effect however on fully adopting a European-like model.

9.3 Also adopted form PRACE, the utilisation of MOOC environments for introductory material. Positive components: combination of video and text components, fine-grained content, active involvement by educators in learner forums. Immediate concerns: too high level, not enough hands-on activities, free-text Cloze skills evaluation.

9.4 Need to expand feedback including learners as partners in learning design, teaching governance, support (peer mentoring), evaluation and evidence, learning environment. Due to time constraints, option for longer term summer school programmes (e.g., Le calcul haute performance au CEA

`http://www-hpc.cea.fr/SummerSchools/SummerSchools2017-CS.htm`)

10.0 References and Thanks

Martin Hilbert, Priscila López. "The World's Technological Capacity to Store, Communicate, and Compute Information", Science Vol. 332 no. 6025 1 April 2011, pp. 60-65.

Malcolm Shepherd Knowles, Elwood F. Holton, Richard A. Swanson, "The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development", Gulf Publishing Company, 1998

K Holtzblatt., H.R. Beyer, "Contextual Design" in M. Soegaard, and R.F. Dam, (eds.). "The Encyclopedia of Human-Computer Interaction, 2nd Ed.". Aarhus, Denmark: The Interaction Design Foundation, 2014

George Siemens, "Connectivism: A Learning Theory for the Digital Age", International Journal of Instructional Technology and Distance Learning, January 2005

Thanks to the University of Melbourne and Goethe-University Frankfurt Am Main

SKA image from ICAR, papers from foreignpolicy.com, hammer from yellopages.ca, xpbs from University of Buenos Aires, strudel from Monash University, child to adult from icytales.com, David Kern and Patricia Thomas curriculum development model from US National Library of Medicine.


THANKS FOR WATCHING
& LISTENING PATIENTLY
makeameme.org