

THETA

The Higher Education Technology Agenda

Software Tools Compared To User Education in High Performance Computing

The growth in datasets is well recognised due to the increasing ubiquity of information-gathering. These datasets are providing a challenge for issues such as storage, processing, analysis, and curation. Viewed as a workflow, it is the second issue that is discussed here. In such an environment, uncore desktop applications and traditional file systems are not capable of providing researchers their needs within a reasonable time [1]. Instead, the only viable option is parallel processing on high performance computing clusters and grids with the use of parallel file systems on mass storage. However the necessary skillset – the command line interface, job submission, scripting, parallel programming – is not common among researchers and training is not generally available. It is in this respect the common conflation of high performance computing with scientific computing is quite inaccurate [2]. Scientific computing should be analogous to high performance computing, and indeed must be if an research organisation is to remain a viable entity as existing research shows a very strong correlation between provision and research output [3].

Two broad methods exist for bringing scientific and high performance computing together; (i) modify the HPC environment to suit the existing skillset or (ii) develop the skillset to match the HPC environment. There has been significant development in the former area, especially championed by software developers and management who want to simplify job submission tools. Well-known examples include xpbs, grid computing interfaces such as the former Grisu project, distributed computing installers such as folding@home, web portals such as the Workflow Management System of BeesyCluster [4] or even from the direction of applications developing parallel capacity, such as Matlab's DCS and parallel computing toolbox. An Australian example which is reviewed is the implementation of Monash eResearch's STRUDEL (ScienTific Remote User DEsktop Launcher) which has been shown to usability and uptake of CQUniversity's High Performance Computing (HPC) facilities [5].

However, even the provision of the most user-friendly and modular submission tools remains

challenging because parallelisation and high performance computing requires a degree of understanding of the process. Without the grounded understanding the eResearcher will be learning (and relearning) applications. The alternative is to provide a graduated training that provides both the skillset for HPC utilisation but also implicit learning adaptable for future situations. For past several years the Victorian Partnership for Advanced Computing (VPAC), and the successor organisation, V3 Alliance, have conducted a range of training courses designed to bring the capabilities of postgraduate researchers to a level of competence useful for their research. These courses make use of some of the key insights of andragogical education [6], particularly the use of integrated structured knowledge orientated towards understanding [7] which encourages learner self-efficacy [8] and combining the insights of proximal learning with a follow-up connectivist mentoring and outreach program [9]. This strategy has also resulted in significant increases in use on the 'Trifid' cluster.

A comparison between software tools and user education indicates that the best tools provide a low-level entry from desktop application use to making use of multicore cluster resources, but with the increasing need for grounded understanding achieved via user education as complexity grows. Additional positive externalities are provided through the provision of structured online course material with feedback, as a ongoing variation of a massive open online course which encourages connectivist approaches [10], but with a more targetted and community-oriented approach which leads to many of such endeavours having low completion rates. It is expected that this new approach will deliver ever greater user-numbers, utilisation, and successful research projects.

[1] Adam Jacobs, 'The Pathologies of Big Data', Queue, Association for Computer Machinery, July 6, 2009

[2] Greg Wilson, 'High Performance Computing Considered Harmful', 22nd International Symposium on High Performance Computing Systems and Applications, 2008

[3] Amy Apon, Standley Ahalt, Vijay Dantuluri, et. al., 'High Performance Coputing Instrumentation and Research Productivity in U.S. Universities', Journal of Information Technology Impact, Vol 10, No 2, p87-98, 2010

[4] Paweł Czarnul, 'A Workflow Application for Parallel Processing of Big Data from an Internet Portal', Procedia Computer Science, ICCS 2014. 14th International Conference on Computational Science, Volume 29, 2014, p499-508

[5] Jason Bell, Chris Hines, 'Improved High Performance Computing usability and uptake through the utilisation of Remote Desktops', eResearch Australasia, 2014

[6] Malcolm Knowles, Elwood Holton, Richard A. Swanson, 'Beyond Andragogy' in 'The adult learner: The definitive classic in adult education and human resource development' (5th ed) .Gulf Publishing Company, 1998, p153–179

[7] Marilla Svinicki, 'Helping students understand' in 'Learning and Motivation in the Postsecondary Classroom', Anker, 2004, p39-60

[8] Dale Schunk, Frank Parajes, 'Self-Efficacy in education revisited: Empirical and applied evidence' in D.M. McInerneyt & S.V. Etten (Eds.) "Big Theories Revisited", Information Age Publishing, 2004, p115-118

[9] George Siemens, 'Connectivism: A Learning Theory for the Digital Age', International Journal of Instructional Technology and Distance Learning, Vol. 2 No. 1, Jan 2005

[10] Rita Kop 'The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course', International Review of Research in Open and Distance Learning, Volume 12, Number 3, 2011

Lev Lafayette

V3 Alliance

■ [Full Paper](#)

SHARE THIS:



Loading...

[+ Follow](#)