Training and Education in High Performance Computing for eResarchers

Eresearch Australasia 2014
Melbourne, 28th October
Lev Lafayette
Increasing Data Processing Needs

Datasets are growing in size due to the increasing scale of information-gathering devices and traditional tools for analysis.

Desktop systems and applications are simply not capable of providing research needs within a reasonable time and instead require parallel processing on high performance computing clusters and grids [1].
The Missing Skillset

However the necessary skillset - the command line interface, job submission, scripting, parallel programming - is not common among researchers and training is not generally available.

The general level of HPC skills among researchers is becoming worse, not better.
High Performance versus Scientific Computing

At least part of the cause of this is due to an assumed conflation of high performance computing with scientific computing.

The assumption, concentrating on hardware and software capabilities, rather than the actual practise has led to a situation where critics can argue that high performance computing is considered harmful [2].
Crisis? What Crisis?

There is a growing disparity between needs and competence which reduces research output.

If the eResearch community do not have the opportunity to utilise parallel processing and high performance computing, they will continue to work with their preferred applications on desktop environments.
Mortality of eResearch Organisations

As a result relative research output will decline, as empirically illustrated in studies of research productivity and investments in HPC [3].

This will prove to be a critical issue for such institutions, in the sense that the survival of the research capabilities of the organisation is at stake [4].
Two broad methods exist for achieving such a match; (i) modify the HPC environment to suit the existing skillset or (ii) develop the skillset to match the HPC environment.

There has been significant develop in the former area, especially championed by software developers and management who want to simplify job submission tools.
Interface Improvement Limits

Well-known examples include xpbs, grid computing interfaces such as the former Grisu project, distributed computing installers (folding@home), or even from the direction of applications developing parallel capacity (Matlab PCT).

The provision of user-friendly trivial submission tools remains challenging because high performance computing requires a degree of understanding of the process.
User-centered and Integrated Training

Without the grounded understanding the eResearcher will be learning (and relearning) applications.

The alternative is to provide a structurally integrated training that provides both the skillset for HPC utilisation but also implicit learning adaptable for future situations.

The training needs to be user-centered rather than content-centered.
Andragogy and Advanced Education

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.

More recently these courses include some of the key insights from the discipline of adult and advanced education in the context of the increasing trend towards lifelong learning.
Current course content begins with basic CLI Linux, file transfers, environment variables, job submission and PBS scripts. An follow-up course covers more advanced CLI commands, regular expressions, scripting, and more elaborate job submissions (interactive, dependencies, arrays), basic compiling, with the advanced course providing a deeper theory of parallel computer architecture, and MPI programming fundamentals. Mathematical and statistical programming with R and Octave is taught as well.
The success of these courses correlated with improved cluster usage in terms of the quantity of job submissions and the number of users. RMIT, which makes up approximately 65% of the training group since 2013 has increased CPU-Hours by 1.61m to 9.96m.
References

THANKS FOR WATCHING

& LISTENING PATIENTLY