Open Source Vocational Engineering with High Performance Computing

Lev Lafayette

Open Source Developers Conference

Hobart, October 2015

Some Relevant Technical Definitions

High-performance computing (HPC) is any computer systems whose architecture allows for well-above average performance.

A supercomputer is a nebulous term for a computer system that is at the frontline of current processing capacity, particularly speed of calculation. It *used* to be mainframe systems with vector processors (e.g., Cray-1).

For the past twenty years clusters have been dominant: A cluster consists of a set of connected computers whose architecture allows them to operate as a single unit.

Some Relevant Technical Definitions

Parallel computing refers to the submission of jobs or processes over multiple processors and by splitting up the tasks ("task parallelism") or data ("data parallelism") between the processors.

The clustered HPC that makes use of parallelisation is the most efficient, economical, and scalable method, and for that reason it dominates supercomputing today.

In June 2015 of the Top 500 Supercomputers worldwide, only 0.2% did not use a "UNIX-like" operating system [1]. Linux is 488 of the top 500 systems (97.6%); UNIX/BSD is 10 (2%), MS-Windows is 1 (0.2%) and there is 1 mixed system (0.2%).

And Here's A Picture

Contemporary Success of HPC

High performance computing is increasingly essential for the processing of large datasets. Unicore desktop applications are simply not capable of providing researchers their needs within a reasonable time. The Square Kilometre Array (SKA) is estimated will collect more data per day than what is transferred through the entire Internet does in year [2].

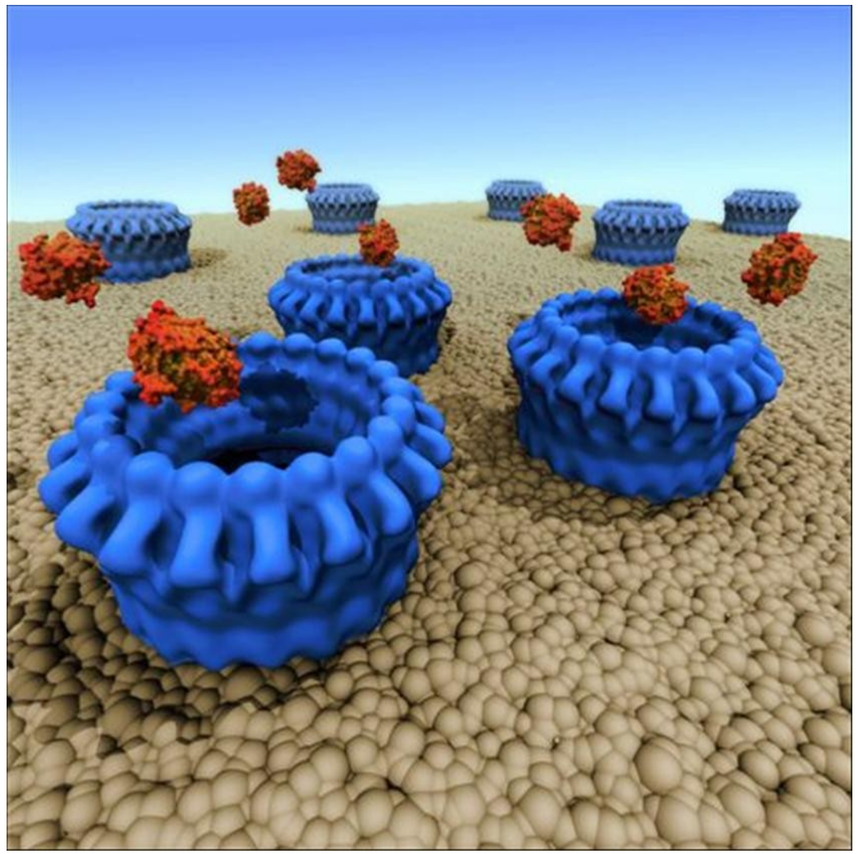
Supercomputers are typically (but not exclusively) used for scientific computing. Some applications have included weather forecasting, aerodynamic design, fluid mechanics, radiation modelling, molecullar dynamics, CGI rendering for popular movies.

Contemporary Success of HPC

For Victorian examples, researchers from Monash University, the Peter MacCallum Cancer Institute in Melbourne, the Birkbeck College in London, and VPAC in 2010 unravelled the structure the protein perforin to determine how white blood cells attack pathogens [3]

More recently Australian scientists discovered the workings antifreeze proteins using HPC systems at the Victorian Life Sciences Initiative [4]. It is hoped in the near future that this could be used to prevent crop frost damage.

And Here's *Two* Pictures!



VPAC's (V3 Alliance's) Teaching of HPC

The Victorian Partnership for Advanced Computing (VPAC) is a not-for-profit registered research agency established in 2000 by a consortium of Victorian Universities. In 2013, VPAC and VeRSI merged to become the V3 Alliance (trading name) with a spin-off commercial company, VPAC Innovations. The V3 Alliance is thus a merger of the Victorian Partnership for Advanced Computing (VPAC) and the Victorian eResearch Strategic Initiative [5].

By the mid-2000s the organisation had introduced a number of training courses both in general HPC usage and in MPI programming. It is fair to describe these courses as being technically correct, but often lacked the integrated education approach that was required.

VPAC's (V3 Alliance's) Teaching of HPC

In the late-2000s the courses were revised to incorporate the insights of adult and tertiary education and disciplinary preferences. Tailored courses cover topics such as HPC usage, mathematical programming, researcher utilities, scientific programming etc. The courses are conducted face-to-face with small interactive classes, real-time presentation, reference material, and learner feedback.

Over 1000 student-days of courses have been conducted with researchers in recent years from RMIT, La Trobe University, Swinburne University, Victoria University, Deakin University, Sydney University, the University of New South Wales, Macquarie University, the University of Western Australia, Australian Institute of Marine Science, ARPANSA, and the Dept. of Environment and Primary Industry.

Contemporary Advanced Learners

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.

Lifelong Learning: The industrial-era lifepath (finish school, studying or training for a particular career or job, and then remaining in that job until retirement), no longer holds true.

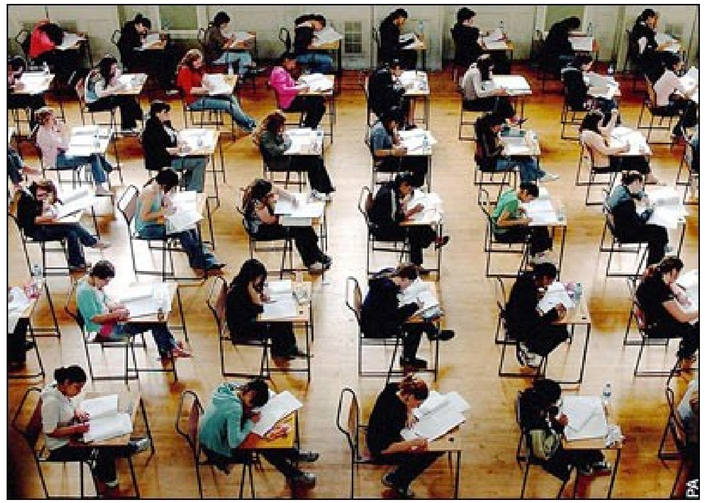
Individual people use different mechanism for immediate goals compared to future goals. An educator needs to discover what motivates a person to engage and persist in a learning experience. In VPACs case there was also the additional issue of a strong multicultural enrolment.

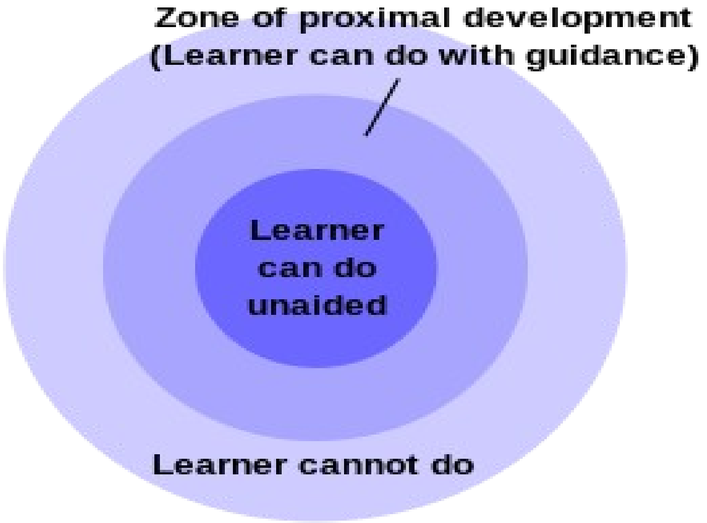
Structured Disciplinary Content

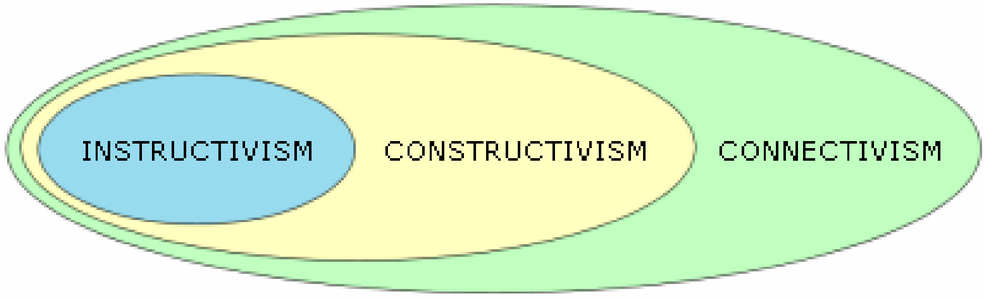
Content needs to organised in terms of objectives, timed, and revised! Content needs to provided in as modular 'structural knowledge', with narrative, analogies, and humour. Provide grounding to a concept (facts *and* reasons). Teach the concepts and the practise as it allows the learner to elaborate their knowledge into new constructions.

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"). The learner-teacher environment is most effective in the "zone of proximal development" (c.f., Vygotsky).

And Here's *Three* Pictures!







The Importance of Feedback

Courses also allow for anonymous feedback on matters of the facilities, content, delivery, relevance and further suggestions. Whilst these indicate an excellent level of "customer satisfaction" they also provide pointers for keeping the material up-to-date and relevant to the user's needs. Also required as part of the ISO standard!

One of the earliest discoveries was that researchers did not have the often-assumed skillset of basic Linux operating system commands or intermediate level scripting. There has also been demand for open-source mathematical languages (e.g., Octave and R) and for interpretative programming languages (e.g., Python).

A Scoreboard Tells A Story

The validity of the approach can be indicated by changes in usage. An excellent starting point for this consideration for V3 Alliance is the end of 2012, when the former cluster (Tango) was in the process of being shut-down (final shutdown was not until April 12 2013), a new cluster has just been introduced (Trifid, November 29, 2014), and most importantly, the course content and material had just undergone a thorough revision.

As an example; VPAC teaches postgraduate researchers how to use Linux and HPC clusters. RMIT and LaTrobe started roughly equal in terms of usage. Change in cluster saw significant increase in usage from both partner institutions... but the one which make use of HPC training improved much more.

A Scoreboard Tells A Story

Trifid Usage (CPU Hours) to December 31st 2014

Year	RMIT	La Trobe	Cluster
2012	1,729,837h	1,719,554h	Tango
2013	8,108,695h	3,301,052h	Trifid
2014	9,760,919h	4,964,297h	Trifid

Trifid Course Enrolments to December 31st 2014

RMIT enrolments 229

La Trobe enrolments 29

From Research to Vocational Engineering

Whilst the utility of HPC in research is well-established and thus also the need for researcher-level training we need to also consider vocational engineering. Higher educational courses in HPC are typically left to computer specialists on a postgraduate level, however the need and demand for such courses occurs far earlier.

The importance of introducing such skills into the vocational engineering space is indicative of the industrial transformation which increasingly decentralises mechanisation into the world of computer simulations for prototyping, testing etc. Technological innovation, more than any other cause, is the driver of improved productivity and wealth (c.f., Solow residual)

Two Proposed Courses

Initial consideration of two proposed units at RMIT's School of Vocational Engineering have been developed for the Advanced Diploma or Associate Degree level. The first is Implementation of High Performance Computer Systems for the Advanced Diploma in Computer System Engineering.

The second High Performance Computers for Engineering would be a more generic course for advanced manufacturing and mechatronics, aerospace engineering, and civil engineering.

Course Content and Delivery

The HPC Implementation course would cover principles and architecture of HPC, project planning and documentation, datacentre preparation, node installation, networks, operating systems, job schedulers and resource management, building compute nodes with DCMS, environment modules, optimising scientific applications, license managers, monitoring and reporting, directory information services, parallel file systems, PBS job submissionm, and disaster recovery.

The course would consist of 80 contact hours spread over a 36 term; continuous assessment would be utilised to provide the best possible feedback for learner progress which improves self-efficiacy. Delivery would follow the principles given in researcher courses, but with a stricter guidance as per the continuum model (learners will be, in general, younger with lower intrinsic motivation).

Strategic Decisions

The increasing use of HPC facilities in the vocational engineering space should not be surprising. It is part of the general trajectory that is required with large datasets, precise engineering simulations, and industrial automation. Increased connectivity, storage, artificial intelligence, the digitisation of matter, driverless cars etc., are all seen as main trends by the World Economic Forum [6].

The question for advanced post-industrial economies is whether they can transform themselves appropriately. Recently in the United States a promising agreement has been generated by Lawrence Livermore National Laboratory and Rensselaer Polytechnic Institute [7] in precisely the areas discussed.

References

[1] "Operating system Family share for 11/2014". Top 500 project. http://www.top500.org/statistics/list/

[2] H. Barwick, "SKA telescope to generate more data than entire Internet in 2020", Computerworld Australia, 2011

[3] Ruby H. P. Law, Natalya Lukoyanova, Ilia Voskoboinik, et al., The structural basis for membrane binding and pore formation by lymphocyte perforin, Nature 468, p447-451, (18 November 2010), doi:10.1038/nature09518 http://www.nature.com/nature/journal/v468/n7322/full/nature09518.html).
See also http://www.abc.net.au/science/articles/2010/11/01/3053386.htm

[4] Michael J Kuiper, Craig J Morton, Sneha E Abraham, Angus Gray-Weale, "The biological function of an insect antifreeze protein simulated by molecular dynamics", eLife 2015;10.7554/eLife.05142
See also: http://www.abc.net.au/news/2015-05-19/australian-scientists-study-antifreeze-proteins/6479602

[5] c.f., http://www.versi.edu.au/about-us/about-versi, https://www.vpac.org/about/, http://www.v3.org.au/about-us http://vpac-innovations.com.au/about-us

[6] "Deep Shift : Technology Tipping Points and Societal Impact", World Economic Forum, 2015

[7] "LLNL & Rensselaer Polytechnic to Promote Industrial HPC" Inside HPC, September 17, 2015 http://insidehpc.com/2015/09/llnl-rensselaer-polytechnic-to-promote-industrial-hpc/


THANKS FOR WATCHING
& LISTENING PATIENTLY
makeameme.org