

Skill Improvements versus Interface Designs for eResearchers

Lev Lafayette

25th March 2015

eResearchNZ



alliance

Accelerating Research



A Backward's Glance: eResearchNZ 2011



The presentation to EResearchNZ 2011 claimed the computational and research resources of high performance clusters and parallel applications is less than optimal from the very research community who would benefit from it. Primarily due to a lack of familiarity on the requisite tools primarily due to changes in tacit knowledge in computer system operation.

At that stage the Victorian Partnership for Advanced Computing had just overhauled their tutorial programme to include the use of multiple learning style preferences, a high trainer-student ratio, and constant feedback and course revision. Initial results were positive.

It was also just before the retirement of the 'Tango' classic (in operation from 2007 to April 2013) and the introduction of the 'Trifid' cluster (started November 29, 2012); 40% allocated to RMIT, 40% to La Trobe University, 20% V3 Alliance (VPAC)

A Backward's Glance: eResearchNZ 2013



The Critical Dataset Challenge



Datasets are increasing in size (Hilbert, López, 2011) faster than processing improvements to personal computer systems with systems like the Square Kilometre Array planned to collect aggregate data on a daily basis greater than the rest of the world on a yearly basis (Barwick, 2011).

The necessary skillset is not common among researchers, leading to a sharp distinction between "scientific computing" with "high performance computing" (Wilson, 2008). Some attempt to bridge the computational skills gap through improving knowledge of basic tools (e.g., Software Carpentry (Wilson, 2006)). However scientific computing *should* be analguous to high performance computing.

Existing research shows a very strong correlation between provision and research output and the provision of HPC systems (Apon et. al, 2010), which can be expected to increase.

HPC Interface Tools



The increased need for HPC can be achieved by either (i) modifying the HPC environment to suit the existing skillset or (ii) develop the skillset to match the HPC environment. Whilst most development has been in the latter attempts to implement a GUI alternative or 'skin' have been less than successful. This is generally because HPC systems require an understanding of the process, rather than an intuitive response, and is not particularly prone to increase simplicity.

An interesting alternative and counter-example is STRUDEL (Bell, Hines, 2014). A default configuration was established that allow for such applications to run in an interactive mode, in addition to significant automation in connectivity. This level of automation allows users to have a transitional stage between the desktop application and the HPC application as structured knowledge and also as an opportunity to witness the improvements in job speed.

HPC Interface Tools

The screenshot shows a terminal-based HPC interface with the following sections:

- HOSTS**: A table listing server information.
- QUEUES**: A table listing queue information, filtered by host(s).
- JOBS**: A table listing job information, filtered by queue(s).
- INFO**: A section for system information.

Server	Max	Tot	Que	Run	Hld	Mat	Trn	Ext	Status	PEsInUse	Select All
miniclick.hrz.tu-chemnitz	0	0	0	0	0	0	0	0	Active	-/-	detail
cllc0a1.hrz.tu-chemnitz	0	7	0	7	0	0	0	0	Active	-/-	Submit..

Queue	Max	Tot	Ena	Str	Que	Run	Hld	Mat	Trn	Ext	Type	Server	Select All
cllcTimeQS	0	0	yes	yes	0	0	0	0	0	0	Route	cllc0a1.hrz.tu-chemnitz	detail
cllcTimeQM	0	0	yes	yes	0	0	0	0	0	0	Route	cllc0a1.hrz.tu-chemnitz	
cllcTimeQL	0	0	yes	yes	0	0	0	0	0	0	Route	cllc0a1.hrz.tu-chemnitz	
cllcTimeQXL	0	0	yes	yes	0	0	0	0	0	0	Route	cllc0a1.hrz.tu-chemnitz	
cllcTimeQXXL	0	0	yes	yes	0	0	0	0	0	0	Route	cllc0a1.hrz.tu-chemnitz	
cllcMainQ	0	0	yes	yes	0	0	0	0	0	0	Route	cllc0a1.hrz.tu-chemnitz	

Job id	Name	User	PEs	CputUse	WalltUse	S	Queue	Select All
2919	cllc0a1.hrz...	rask	100	00:00:24	15:50:08	R	cllcNodeQS7@cllc0a1.hrz.tu-chemnitz	detail
2920	cllc0a1.hrz...	klpa	61	00:00:18	14:48:51	R	cllcNodeQS1@cllc0a1.hrz.tu-chemnitz	modify..
2921	cllc0a1.hrz...	kber	64	00:07:50	04:15:16	R	cllcNodeQS1@cllc0a1.hrz.tu-chemnitz	delete..
2924	cllc0a1.hrz...	kumat	4	00:00:18	01:44:36	R	cllcNodeQS1@cllc0a1.hrz.tu-chemnitz	hold..
2925	cllc0a1.hrz...	tnc	1	01:27:59	01:28:33	R	cllcNodeQS1@cllc0a1.hrz.tu-chemnitz	release..
2926	cllc0a1.hrz...	mibe	4	00:00:01	00:44:05	R	cllcNodeQS1@cllc0a1.hrz.tu-chemnitz	signal..
2928	cllc0a1.hrz...	mibe	200			R	cllcNodeQS7@cllc0a1.hrz.tu-chemnitz	msg..
								move..
								order

The screenshot shows the Strudel Beta login window with the following fields and options:

- Site**: Standard Interactive Session
- Username**: bellj
- Memory (GB)**: 100
- CPUs**: 10
- Show Advanced Options**:
- Exit** and **Login** buttons.

HPC User Education



The eResearch community requires use of andragogical techniques. Postgraduates especially tend towards more voluntaristic engagement, are more self-motivated, and have a greater range of experiential resources. They participate because they have specific relevant problems. The learner is autonomous and usually self-determining with internal rather than external motivations.

Content is provided in a structured manner embedding knowledge in organised context with other knowledge, adding to learner self-efficacy (a type of formative assessment).

The delivery of the material makes use of andragogical and connectivist disciplinary learning style, whilst providing options for preferences. Material is partially provided in interactive lecture form, and with a wide opportunity for connectivist hands-on activity with proximal development, and a hardcopy for reference.

Results 2012-2014



V3 Alliance has concentrated on the HPC User education side path. Since the breakpoint of 2012 there has been very significant increase overall in cluster usage...

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

Results 2012-2014



... but even more so for those who participated in the training courses.

Trifid Course Enrolments to December 31st 2014

RMIT enrolments: 229 (62.74%)
La Trobe enrolments: 29 (7.96%)
University of Melbourne 38 (10.41%)
DEPI/DPI 28 (7.67%)
Swinburne University 16 (4.38%)
Deakin University 15 (4.11%)
Victorian Univeristy 7 (1.92%)
Commercial 3 (0.82%)
Total enrolments: 365

Conclusions



The common path of providing "user friendly" interfaces for eResearchers may be an expensive tangent that does not improve research productivity. Improved user interfaces are helpful when the interface can be reduced to an intuitive action and when a complex process can be automated. Task sequence, software artifacts, network connectivity must be evaluated (Holtzblatt, K., Beyer, H.R. (2014) along with user reviews.

The alternative path of improving the skill-set of users, even in processes which are complex and require deeper understanding, will be successful when applied to groups which are known for 'determined intelligence' and where it is lack of familiarity that is preventing current use. In the case of eResearchers, andragogical principles must be applied which recognise the capacity and status of the learner, along with new learning environments which combine proximal development with connectivism.

References



- Apon A., Ahalt S., Dantuluri V, et. al., (2010) "High Performance Coputing Instrumentation and Research Productivity in U.S. Universities", Journal of Information Technology Impact, Vol 10, No 2, p87-98
- Barwick, H., (2011), "SKA telescope to generate more data than entire Internet in 2020", Computerworld Australia
- Bell J., Hines C., (2014) 'Improved High Performance Computing usability and uptake through the utilisation of Remote Desktops', eResearch Australasia
- Hilbert, Martin., López, Priscila. (2011) "The World's Technological Capacity to Store, Communicate, and Compute Information", Science Vol. 332 no. 6025, pp. 60-65
- Holtzblatt, K., Beyer, H.R. (2014) "Contextual Design" in Soegaard, M and Dam, R.F. (eds.). "The Encyclopedia of Human-Computer Interaction, 2nd Ed.". Aarhus, Denmark: The Interaction Design Foundation.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. Psychological science in the public interest, 9(3), 105-119.
- Schunk, D. H. & Pajares, F. (2004). Self-efficacy in education revisited: Empirical and applied evidence. In D. M. McInerney & S. V. Etten (Eds.), Big theories revisited (pp. 115-138). Greenwich, CT: Information Age Publishing.
- Shabani, K., Khatib., M, Ebadi, S. (2010). Vygotsky's Zone of Proximal Development: Instructional Implications and Teachers' Professional Development, appearing in English Language Teaching Vol. 3, No. 4; December 2010
- Sheldon, K. M. (2009). Changes in goal-striving across the life span: Do people learn to select more self-concordant goals as they age?. In M. C. Smith (Ed.), Handbook of research on adult learning and development (pp. 553-569). New York: Routledge.
- Svinicki, M. D. (2004). Helping students understand. In Learning and motivation in the postsecondary classroom (pp. 39-60). Boston: Anker.
- Wilson, G. (2008), High Performance Computing Considered Harmful, 22nd International Symposium on High Performance Computing Systems and Applications, 2008
- Wilson, G. (2006). Software carpentry. Computing in Science & Engineering, 8, 66.
- Zarestky, J., & Bangerth, W. (2014). Teaching High Performance Computing: Lessons from a flipped classroom, project-based course on finite element methods. In Proceedings of the Workshop on Education for High-Performance Computing (pp. 34-41). IEEE Press.

THANKS FOR WATCHING



& LISTENING PATIENTLY