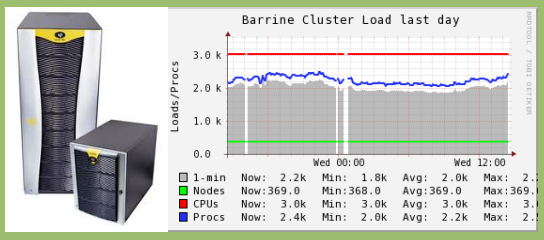
Research Computing @ UQ



3 February 2014 Issue 2

The Research Computing Centre supports collaboration to facilitate discoveries in science and engineering, humanities, and social sciences, through advanced computation, data analysis and other digital research tools.

Message from the Director

Welcome to the first newsletter for 2014. There's a lot happening in RCC, so this newsletter gives me a chance to highlight a few of the more important activities.

Over the past six months I have been working with colleagues, both within RCC and outside, to build a new structure and engagement model. In this issue I will present a high level view of this, and give you the motivation and key features. Of course, I welcome feedback on how you feel it will work for you and your research groups.

UQ now has a data management policy, having been ratified by the Academic Board and Council. This policy outlines the responsibilities for stewardship and protection of research data, and is an important step towards meeting our national and international obligations.

RCC, together with partners, was successful in an ARC LIEF proposal to fund a new computational platform. The new system, codenamed Flashlite, will support data intensive computing, combining parallel computing facilities with high speed storage. Modelled on the Gordon machine at the University of California, San Diego, FlashLite will allow application to manipulate large amounts of data very rapidly, adding capabilities not previously available. Whilst the specifications are not locked down, we give a little information in the newsletter, and will b defining the system capabilities over the next few months.

I am delighted to announce that UQ has chosen an o site data centre for hosting QRIScloud, FlashLite an other significant research computing infrastructure. More details will be announced over coming month The facility will support a substantial expansion ove on campus facilities.

Finally, I am delighted to welcome a number of new staff to RCC, and look forward to working with them over coming months to support research at UQ.

Yours Sincerely,

David Abramson
Director

UQ Launches Research Data Policy

The Research Data Policy aims to ensure that the research data required for validation of research results are properly managed according to recommendations made in the Australian Code for the Responsible Conduct of Research and applicable legislation. This code states that all individuals and institutions engaged in research have a responsibility to manage research data effectively, by addressing ownership, storage, retention and access issues.

The University recognises the significant v alue of the research data generated by its large investment in research. Durable research data are essential to justify, and defend when required, the outcomes of research. Good stewardship can increase the efficiency and maintain the integrity of research results.

The potential cumulative value of research data should also be considered and, where possible, it should be made available for re-use. The university recognises that access to research data can raise the research profile of individuals and institutions, increase returns on public investment, promote open inquiry and debate, and enable innovative uses of data that may not have been foreseen by researchers at the time of its creation. The university is committed to supporting long-term research data management to enable continuing access.

To optimise research outcomes, data must be stored, retained, documented and/or described, made accessible for use and reuse, and/or disposed of, according to legal, statutory, ethical and funding body requirements. Research data management is a shared responsibility. The university expects all researchers, academic units, the Library and central administrative units to work collaboratively to implement good research data management practice.

Research Data Descriptions now live in UQ eSpace

UQ Library's Data Collections Form allows researchers to describe their research data according to good practice, and further supports them by offering either mediated access or open access to their research data. It aids discovery, dissemination and preservation of research data as a first class research output at UQ. It also makes research data visible via search engines such as Google, as well as through national data services including Research Data Australia, and allows researchers to build an index of their research data in eSpace, and link their publications to the underlying data. For more information email data@library.uq.edu.au.

FlashLite - Data Intensive Science comes to UQ

Data is predicted to transform the 21st century, fuelled by an exponential growth in the amount of data captured, generated and archived. Foster predicts a ten-fold increase every five years from 2000 to 2015 alone. Microsoft has identified data intensive science as the fourth scientific paradigm after computation as the third. Not surprisingly, there are numerous projects targeting the challenges that underpin the exploitation and management of this data explosion.

Australia has made significant progress towards addressing some of the opportunities and infrastructure challenges posed by managing such rapid increase in data volumes. Substantial government investment, through the Australian National Data Service (ANDS), the Research Data Strategic Infrastructure (RDSI) and the National eResearch Collaboration Tools and Resources (NeCTAR) programs will advance our ability to search, manage and store large amounts of data. However, whilst important, these investments do not address the escalating scientific imperative to exploit and process data.

FlashLite will be designed explicitly for Australian research to conduct data intensive science and innovation. It will support applications that need very high performance secondary memory as well as large amounts of primary (main) memory, and will optimise data movement within the machine. Data intensive applications are neither well served by traditional supercomputers nor by modern cloud-based data centres. Conventional supercomputers maximise Floating Point Operations per Second (FLOPS) and inter-processor communication rates through high bandwidth and low latency networks. Conversely, modern Cloud systems minimise the cost of ownership through reliance on virtual machines and shared storage; they thus utilize relatively slow processors and networks and, by and large, do not support parallel processing.

FlashLite, on the other hand, will maximise Input Output Operations per Second (IOPS) while achieving competitive FLOPS ratings and high performance networking, producing a balanced system for applications that exploit parallelism, high speed arithmetic, and high performance Input/Output (IO). The machine will also incorporate novel software mechanisms that provide seamless access t data regardless of its location, making it easier to build new data-intensive applications, whilst supporting legacy codes using familiar techniques.

FlashLite is inspired by the US National Science Foundation (NSF) machine (called Gordon) that is already achieving impressive performance on data intensive applications at the San Diego Supercomputer Centre (SDSC). An early prototype Gordon machine accelerated real applications by 2-100 times.

The growth of digital data is leading to the so-called "data tsunami", driven by advances in digital detectors, networks and storage systems. Each of these technologies are following a trajectory known as Moore's Law, which provides improvements in detector resolution (with associated lower costs), network transmission rates and memory densities th

are doubling at least every 18 months. The growth i also driven by increased resolution of computationa models, which leverages improved computational power through multi-core devices, again following Moore's Law. Critically, the gap between the arithmetic processing speeds and transfers between main and disk memory is increasing, making it mor and more difficult to

analyse and process increasingly large volumes of data in a timely way.

Traditional High Performance Computing systems (HPC) are designed to deliver maximum computational performance on scientific codes, but are not typically optimised for maximising memory throughput – either primary (main) memory or secondary (disk) memory. In order to enable HPC applications over large data volumes, it is time to rethink existing computer architectures to redress this imbalance. FlashLite is such a system, and incorporates three key innovations that are not typically embraced by traditional HPC systems:

- high throughput solid state disk (instead of spinning disk);
- large amounts of main memory; and
- software shared memory.

"Gordon is a unique machine in NSF's Advanced Cyberinfrastructure/XSEDE portfolio," said Barry Schneider, NSF program director for advanced cyberinfrastructure". It was designed to handle scientific problems involving the manipulation of very large data. [Using Gordon] researchers identified the hierarchical tree of coherent gene groups and transcription-factor networks that determine the patterns of genes expressed during brain development.

These innovations mean that applications have increased access to a high performance memory system, spanning high speed main memory (usually supported by DRAM technology) and low latency, high throughput, solid state disk (also known as Flash memory). A novel (commercial) software abstraction, called Virtual Shared Memory (vSMP) bridges these components, thus simplifying programming.

This application comprises 18 CIs from 5 collaborating universities, representatively spanning a wide range of research endeavours. The CIs, all of whom have high profile reputations in their respective fields and lead major computational research efforts, are representative of their universities' research strengths/priorities. This list includes current and former ARC Federation Fellows (Burrage), Professorial Fellows (Abramson), Future Fellows (Gu, Zhu), CIs in Centres of Excellence and CRCs (Abramson, Burrage, Tomlinson), and holders of ARC project grants (Discovery and Linkage).

All of the CIs maintain excellent international linkages, and these further enhance the research capability. For example, Abramson actively collaborates with research staff at SDSC, and regularly attends research meetings and workshops there. He is well placed to leverage the SDSC investment in Gordon and to lead FlashLite. Further, Abramson and Burrage collaborate with researchers at Oxford University, bringing a significant international flavour to the cardiac modelling activity at UQ and QUT. Other CIs maintain similar linkages.

The Research Computing Centre

New techniques and technologies are enabling us to both ask, and answer, bold new questions. This is causing a significant shift in the way research is conducted, leveraging new resources such as massive data stores, enormous computational power and new ways of communicating. This emergent infrastructure is called e-Research.

For researchers trained in recent decades, this represents a substantial challenge. In order to even remain competitive, let alone lead their field, they need to embrace the new ways of working, and must feel comfortable and productive in this new environment. In response, governments and private organisations around the world have invested not only in e-Research, but also in support services that allow researchers to be productive.

The Research Computing Centre (RCC) is a University-level Centre that provides coordinated management and support of the University's sustained and substantial investment in e-Research. RCC is an innovative and multidisciplinary environment that supports collaboration to facilitate discoveries in science and engineering, humanities, and social sciences, through advanced computation, data analysis and other digital research tools. The centre enhances the University's e-Research infrastructure, and provides support for interdisciplinary research and education.

RCC has developed a structure and interaction model that matches the way research is conducted at UQ, leveraging expertise in faculties, Research Centres, Institutes and other support groups. This is no mean feat – the UQ research landscape is complex, diverse and distributed, and employs widely different models that are domain specific. RCC has developed a unique, multi-tiered structure that both shrinks the gap between researchers and e-Research infrastructure, but empowers those who already hav the skills to excel.

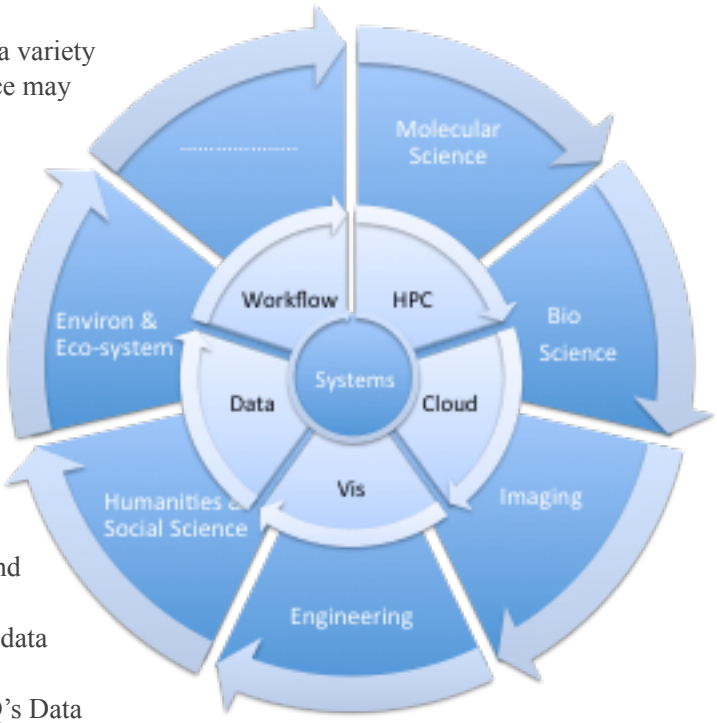
RCC leverages investment by government in initiatives such Queensland Cyber Infrastructure Foundation (QCIF); the National eResearch Collaboration Tools and Resources (NeCTAR); the Research Data Storage Infrastructure project (RDSI and the Australian National Data Service (ANDS). I also builds on key support services in the University namely Information Technology Services (ITS) and the Library.

RCC aggregates expertise in core e-Research technologies, such as Cloud Computing, Data Management, High performance Computing (HPC), Workflow Tools and Visualisation. Over and above this, it builds an expandable layer of domain expertise, initially in Bioinformatics and Genomics Computational Engineering; Environment and Ecology; Humanities and Social Sciences; and Advanced Imaging. These domain layers are the primary interface to researchers, and are collaborati ventures with existing groups.

RCC uses a collaborative funding model, utilising multiple sources, including the Office of the Deputy Vice-Chancellor (Research), QCIF, NeCTAR, RDSI

The RCC model layers domain specific expertise on top of generic computing tools. Researchers can engage at a variety of levels depending on their own skills and needs. The innermost ‘Core’ layers contain generic technologies that can be applied across a wide range of research areas. For example, workflow tools can be applied in everything from digital humanities to molecular science. The outer ‘Theme’ layers are domain-specific, and contain multi-disciplinary teams with both domain knowledge and computing skills.

Agile teams are formed to pull expertise from a variety of sectors. For example, a project in bio-science may require skills in data management, high performance computing and workflows. Another project in imaging may require expertise in data management alone.



Importantly, RCC will fund all activities in the inner two layers, and will call for co-investment in the outer, domain-specific, layers.

RCC maintains a close working relationship with UQ Library around research data management. Specifically, RCC will focus on technical issues associated with data storage and transport, whilst the library will focus on the development of data management plans, meta-data specification and management and linking to publications. Together, these will underpin UQ’s Data Management Policy.

QASMT embraces supercomputing

RCC supported three students from the Queensland Academy of Science, Mathematics and Technology to attend the 2013 IEEE Supercomputing conference in Denver (SC13).

For a quarter of a century, the Supercomputing Conference has served as the crossroads for the entire HPC community. From users and program managers to colleagues and vendors...from government to private industry to academia...SC has provided unparalleled cooperation, unequaled collaboration, and unmatched exposure.

Spotlighting the most advanced scientific and technical applications in the world, SC13 brought together the international supercomputing community for an exceptional program of technical papers, tutorials and timely research posters. The SC13

Exhibition Hall featured exhibits of the latest and greatest technologies from industry, academia and government research organisations; many of these technologies were seen for the first time in Denver.

The conference gave the QASMT students a unique opportunity to learn about supercomputing. Steve Blair, science teacher from QASMT who accompanied the students, said: “SC13 was amazing It has spurred the students into taking a more active interest in how supercomputing interacts with scienc We will certainly be integrating this into our curriculum”

International Conference on Computational Science (ICCS)

10-12 June 2014 - Cairns, Australia

"Big Data meets Computational Science"

The International Conference on Computational Science is an annual conference that brings together researchers from various application areas who are pioneering computational methods in sciences such as physics, chemistry, life sciences, and engineering, as well as in arts and humanitarian fields, along with researchers and scientists from mathematics and computer science as basic computing disciplines. This mixture of computational researchers are able to discuss problems and solutions in their areas,

identify new issues, and to shape future directions for research.

For more information see: http://iccs2014.ivec.org/

ICCS is well known for its excellent line up of keynote speakers. The keynotes for 2014 are:

Professor Vassil Alexandrov, ICREA Research Professor in Computational Science, Barcelona Supercomputing Centre, Spain

Professor Dr Luis Bettencourt, Santa Fe Institute, New Mexico, USA

Professor Professor Peter T. Cummings, Department of Chemical and Biomolecular Engineering, Vanderbilt University, USA

Professor John Mattick, Garvan Institute of Medical Research, Sydney, Australia

Professor Bob Pressey, Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Australia

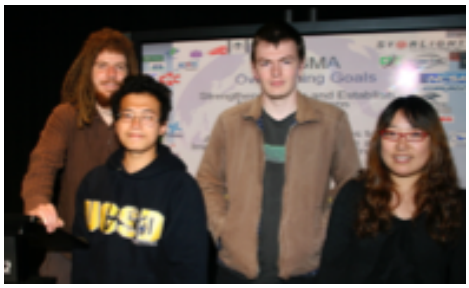
Professor Mark Ragan, Institute for Molecular Bioscience, The University of Queensland, Australia

International Expert comes to RCC

RCC recently hosted Professor Bob Panoff, from the SHODOR organisation at UQ. Bob Panoff is founder and Executive Director of The Shodor Education Foundation, Inc., and has been a consultant at several national laboratories. He is also a frequent presenter at NSF- sponsored workshops on visualisation, supercomputing, and networking, and continues to serve as consultant for the education program at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign.

In October Bob presented a seminar at UQ on "Transforming learning through Computational Thinking". Bob also gave a seminar to school students at QASMT.

International Research Internships come to RCC

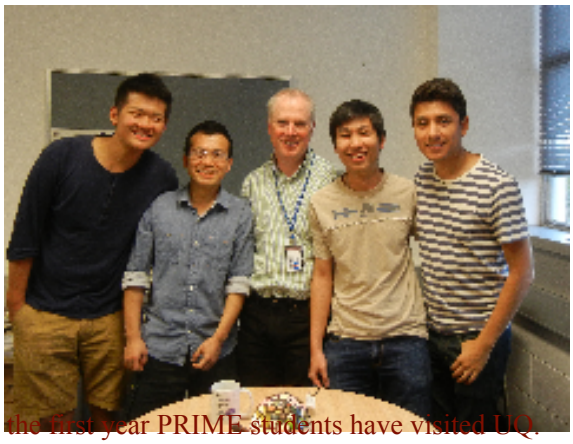


In today's educational arena, universities must provide students with opportunities to work and study abroad to prepare them for global citizenship and professional competence in a multi-cultural workplace. Numerous reports have challenged universities to develop educational programs that provide an integrated academic basis for developing students' cultural/global competencies.

Over the past five years, 26 Monash University students have travelled to international partners at the University of California, San Diego, The National Center for Supercomputing Research in Illinois, The Technion in Israel and the University of Warwick. Students are placed for a period of eight weeks, allowing them to integrate into the research groups as team members. Students have a local mentor in Australia as well as one in the remote site, and often bridge international research projects.

This year RCC has launched QURPA, and UQ students will join Monash students for the first time. This not only allows UQ students to engage in some fascinating advanced computing projects, but opens the door to trans-national student-lead collaboration in undergraduate research. We have also added collaborators at the Institute for Infocomm Research (I2R) in Singapore in addition to projects at UCSD. Projects span a wide range of advanced computing technologies and leading edge applications, and serve as incubators for RCC projects.

MURPA and QURPA involve an advanced seminar scheme, in which students can attend virtual seminars given by world leading experts before they leave. These seminars also allow students to "meet" potential UCSD mentors and get some information about potential projects. In 2013 seminars were sourced from Faculty at UCSD, the University of Indiana, I2R in Singapore, the National Centre for Supercomputing Applications (NCSA) and Rensselaer Institute of Technology. In 2013 seminars were broadcast simultaneously to Monash (in Melbourne) and UQ (in Brisbane), with audiences able to ask questions from either venue. The seminar infrastructure supports a wide range of video conference technologies (both open source and commercial), and is displayed on a 20 MPixel OptiPortal.



In addition, RCC recently hosted 2 UCSD students in an NSF funded program called PRIME. PRIME has been running since 2004, and sends students from UCSD to research partners across the Pacific Rim. This is the first year PRIME students have visited UQ.

Rasellus hendrerit pulvinar nibh