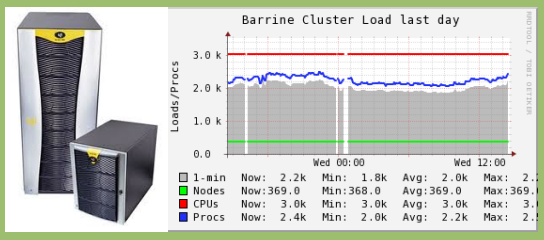
Research Computing @ UQ



5 May 2014 Issue 3

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 May RCC newsletter. It's hard to believe we are almost half way through 2014!

In this newsletter we report on a number of activities around the University. By the time you read this we will have completed deployment of the second stage of the QRIScloud infrastructure I mentioned in the last newsletter, out at the Polaris Data Centre in Springfield. This time, we will give you a few details about this facility. RCC is also well advanced in the next stage of procurement for data storage, and this will boost the cloud storage from its current PB to some 10s of PB.

Our first QURPA international intern, Brian Song, reports back from his trip to our partner, the University of California, San Diego. Brian tells us a little about his trip. We are happy to promote the Winter School in Mathematical and Computational Biology to be held at UQ in July, and give you some details about this exciting event.

We also discuss an enhancement to the UQ Library eSpace repository that allows it to store computer code as well as data. This advance will make it possible to publish code, and if appropriate, link this to specific publications. This is a great advance for researchers involved in computational science.

Speaking of which, [illegible] Conference in Computational Science will be held in Cairns in June. It represents a great opportunity to hear some excellent international speakers, as well as meet up with colleagues.

Last newsletter I mentioned the Genomics Virtual Laboratory (GVL), a NeCTAR funded virtual lab.

In this edition we give you a little more information. We also give you some detail of the AuScope Inversion Lab which is using the National Computing Infrastructure for Large-Scale Geophysical Inversions.

Don't forget that we are always happy to see you at our HPC training sessions to help you get the most out of the HPC infrastructure, both at the University and nationally. Finally, we have the first of what will become a rolling set of introductions to members of the RCC team. David Green, Deputy Director, kicks off with a candid presentation.

Yours Sincerely,

David Abramson
Director

UQ Chooses Polaris

UQ has chosen Polaris to host the new QRISCloud infrastructure discussed in the last newsletter. The Polaris Data Centre offers a very high level of professional management. It is regarded as one of Australia's most modern, highly secure, purpose-built, Tier 3+ data centres. The facility is unique in delivering a 99.99% availability.

Housed in Springfield, Polaris provides a 14,000m^2 purpose-built, Tier 3+, high security facility delivering nearly 7000 m^2 of raised floor over three levels. It uses biometric man-traps on main entries to all raised floor areas, with full digital CCTV surveillance and logging, and provides a scalable, redundant power and cooling capacity. Polaris has been reviewed and independently audited in regards to the security and availability requirements of a major banking / financial institution, an Australian State Government, and other multinational organisations.

Whilst the initial deployment of equipment will be QRIScloud, RCC expects to place further equipment, including the new FlashLite Data Intensive computer out there later this year. Collocating these high-end resources will provide significant performance for research computing at UQ.

First QURPA Student Reports back

Brian Song (Bachelor of Engineering (Software), and now an MPhil candidate and member of the Complex and Intelligent Systems group supervised by Professor Janet Wiles was recently deployed to the University of California, San Diego (UCSD) for four weeks sponsored by the Queensland Undergraduate Research Projects Abroad (QURPA) program.

One of the main purposes of this trip was to attend the Temporal Dynamics of Learning Conference, all-hands meeting at

UCSD. Despite the official title, TDLC attracted a vast array of topics to be presented including gaze tracking, social interaction, robotics, neural networks, dolphin vocalisation and studies in autistic children. Brian's research, Computation Time Scales in iRat-Rat Social Interactions, was presented at the TDLC poster session. (The iRat was originally built by UQ to be used in rat-robot experiments conducted at UCSD).

TDLC was only 3 days, so for the remainder of the trip, Brian spent most of his time assisting in iRat-Rat experiments in Professor Andrea Chiba's Rat Lab (in UCSD's Cognitive Science department). The overall purpose of the rat-robot experiments was to see if they could elicit or verify altruistic motives made by (real) rats. Brian also assisted in a neural recording experiment in Professor Doug Nitz's lab where they raced the iRat and a rat down a straight 2m track. The purpose of this experiment was to observe the neural activity of mirror place cells in the rat as it races against the iRat.

Throughout the trip, Brian also met with key researchers at CatIT2 (Qualcomm Institute), the Salk Institute and Brain Corporation. CalIT2 is a significant partner for the Research Computing Centre.

What is MURPA/QURPA?

MURPA (Monash Undergraduate Research Project Abroad) was launched at Monash University in 2008. QURPA is the local extension of this program. The program supports a unique summer research placement for final year undergraduate students studying computer science, software engineering, or technically oriented IT units, in a leading research group overseas. It not only provides a research experience at the undergraduate level, but does that in an international context.

Students attend advanced seminars, simultaneously transmitted to Monash and UQ in semester 2. Interns then undertake an 8-week summer semester international research project at the prestigious University of California, San Diego (UCSD) and the University of Warwick, UK. This year for the first time, students were placed at the Institute for Infocomm Research (I2R) in Singapore.

This annual program exposes final year undergraduate students to an international research experience within a leading research laboratory.

Winter School attracts White House Champion of Science Keynote

Harvard Professor John Quackenbush will deliver a keynote address on "Taming the Big Data Dragon" at the Winter School in Mathematical and Computational Biology to be held at UQ in July. His address will cover the data challenges facing the biological and computational biological sciences. Professor Quackenbush was recently honoured at a White House ceremony as one of the Obama Administration's Champions of Science for his contributions to Open Science.

Professor Quackenbush's address will be one of several sessions on Big Data at the Winter School, organised by RCC Director Dr David Abramson and Dr Nick Hamilton, a recent co-appointment between the Institute for Molecular Bioscience and the Queensland Cyber Infrastructure Foundation.

The Winter School, now in its tenth year, is designed to introduce mathematical and computational biology and bioinformatics to advanced undergraduate and postgraduate students, postdoctoral researchers and others working in the fields of mathematics, statistics, computer science, information technology, complex systems analysis,

engineering.

The School has been noted for the "compelling orators" it attracts and this year is no different with six international, 14 national and 15 Queensland field-leading researchers presenting. Highlights include sessions on Next Generation Sequencing and Bioinformatics, Modelling from High-Throughput Data, Statistical Applications and Molecular Phylogenetics.

The Winter School will be held at UQ from 7 to 11 July 2014. Early Bird registration is ridiculously cheap at only $205 for students and includes lunches and morning teas across the week.

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 is 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

Research code record type to be featured in UQ eSpace

Scientists who write research code and software engineers will soon be able to record their work in the UQ eSpace repository.

Until now, code has not been recognised as a distinct research output type in UQ eSpace. In the past, anyone wanting to describe research code had to make do with record types such as Data Set or Creative Work, neither of which include the fields needed to describe software adequately.

Some researchers felt this was unfair, as writing and developing code is an important part of their research. In many cases, research would be neither repeatable nor understandable without access to the accompanying code base. In other cases, the code is the research.

Having a record in the UQ eSpace repository helps make research code discoverable.

As one earth sciences researcher put it: 'The existence of releases as individual entities is important in cases like ours where funding is contingent on regular releases'.

It also helps 'legitimise' the code by providing its context. The same researcher went on to say: 'When people are deciding whether to use software, they may wish to know how long the programs have been in development, and whether [the code] is still in active development'.

UQ eSpace staff are now developing a metadata record for research code/software. The record type was designed by library staff in conjunction with staff of the Queensland Cyber Infrastructure Foundation and the Research Computing Centre, after QCIF staff were lobbied on the issue by a number of researchers.

The new record type will allow depositors to describe research code quite finely. Fields include title and description, version, the research project to which it belongs (if any), the programming languages and distribution file formats, and the required software and hardware platforms. Licensing and access conditions can also be recorded. It will be possible to link records to earlier or later versions of the same code.

UQ eSpace staff are still deciding whether to host the code itself. Records can be linked to code lodged in external repositories such as GitHub or SourceForge. Related materials such as documentation or README files can certainly be deposited, as multiple datastreams attached to the code record.

The UQ eSpace repository already provides a permanent identifier for records. Staff are currently investigating minting DOIs, in which case researchers might soon be able to get a DOI for their code.

The Genomics Virtual Laboratory

Through the Genomics Virtual Laboratory (GVL), researchers can instantiate, manage and tailor their own Genomics Analysis platform on the nationally funded Australian Research Cloud (NeCTAR) using national genomics collections (RDSI)

GVL provides Researchers and Developers with a number of resources:

Galaxy: perform reproducible analyses via a friendly user interface

Command Line Notebook: perform reproducible analyses with full command line access and a Windows interface

Tutorials: learn bioinformatics analysis techniques

CloudMan: manage your cluster on the cloud, dynamically add and remove nodes

and several platforms:

Personal: Launch your own instance with your defau Australian Research Cloud allocation

Server: Launch your own analysis cluster, tailored t your needs

Managed: Use the GVL managed resources

GVL's resource aimed at biologists, Galaxy, is an open, web-based platform for accessible, reproducible, and transparent computational biomedical research.

Accessible: Users without programming experience can easily specify parameters and run tools and workflows.

Reproducible: Galaxy captures information so that any user can repeat and understand a complete computational analysis.

Transparent: Users share and publish analyses via th web and create Pages, interactive, web-based documents that describe a complete analysis

For researchers with more IT expertise, GVL's Command Line Notebook is a command line Ubuntu server with:

iPython Notebook, a web-based interactive computational environment where you can combine code execution, text, mathematics, plots and rich media into a single document:

Xfce Desktop Environment, a lightweight desktop environment for UNIX-like operating systems.

RStudio, a powerful and productive user interface for R.

GVL Tutorials are a set of self-paced instructions allowing you to learn analysis techniques on our GVL Galaxy Tutorial server (galaxy-tut.genome.edu.au). You can further explore analysis techniques by using the GVL protocol documents.

The GVL also hosts a mirror of the UCSC Genome Browser on the Australian Research Cloud and a research-scale Galaxy instance.

For more information about the GVL and its facilities, contact Ron Horst

HPC Training

The RCC hosts regular training sessions for its new users with monthly introductory "hands on" sessions in a computer-equipped teaching space. Details of sessions are available on the RCC website.

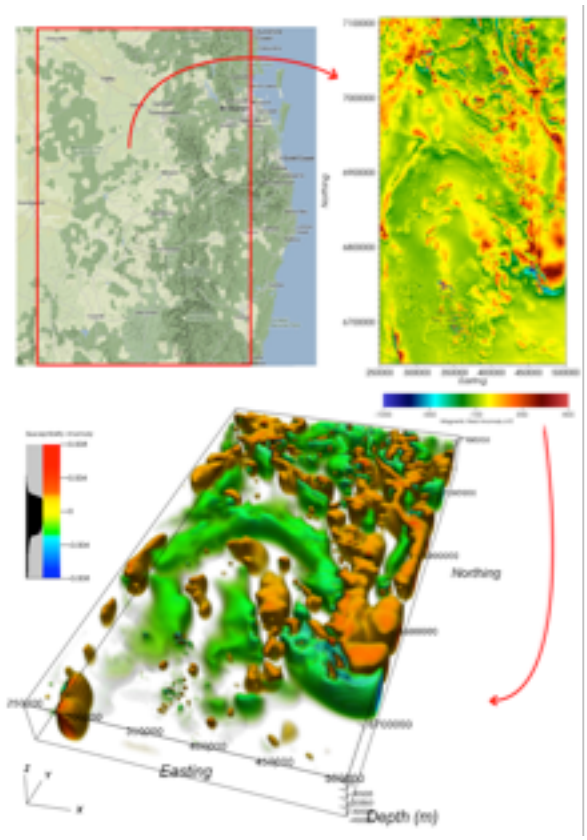
After attending one of these hands on sessions, users are invited to attend a follow up session in which more individualised assistance can be provided. We can revisit how to use the RCC facilities in greater detail or explore how to best utilise the facilities to solve your research problem. Booking in for training is easy- just email us!

AuScope Inversion Lab is using the National Computing Infrastructure for Large-Scale Geophysical Inversions

Geophysical observations are widely used in geological sciences, exploration and, more recently, environmental monitoring in areas such as coal seam gas operations. Through the process of inversion, geophysical data collected above or near the surface are extended into three-dimensional information about the distribution of physical rock properties in the subsurface, which is shown in Illustration 1. Inversion is a computationally intensive task, especially when combining data sets with different physics in joint inversions. As an example of this, the inversion of a magnetic anomaly data set on a provincial scale at a resolution of 250m will require a grid size of over 300 million cells. As it is infeasible to complete an inversion of this size on a desktop computer, compute clusters need to be used. These supercomputers provide a powerful computational environment by linking weaker computational nodes through a fast interconnect. In order to make efficient use of these systems, scalable software with suitable mathematical methods and data structures need to be used.

Under the Australian Geophysical Observation System (AGOS) and with additional funding through the AuScope CRIS and NCRIS2 initiatives, The University of Queensland has developed a new scalable software "Downunder" for the inversion of gravity and magnetic anomaly data on compute clusters. The numerical methods use state of the art optimization algorithms and the finite element technology, and are implemented in a way which is optimal for running on compute clusters with multi-core processors. The software has now been ported to the newest National Computing Infrastructure (NCI) supercomputer and has been scaled up to 8192 CPU cores for a joint inversion of synthetic gravity and magnetic data. The lower portion of Illustration 1 is the susceptibility distribution obtained from the inversion of magnetic anomaly data (top right) for the Clarence-

Moreton Basin in South East Queensland and northern New South Wales. The magnetic data at a resolution of 160m have been obtained from the National Geophysical Airborne Data Base via the NeCTAR funded Virtual Geophysics Laboratory (VGL, see http://vgl.auscope.org.au). VGL is a collaborative effort of Geoscience Australia, CSIRO and NCI. The inversion grid was composed of 140 million cells distributed over 512 cores (16 cores on 32 nodes) and completed in under 80 minutes on the NCI facility. In comparison, the inversion of the same data at an 80m resolution used 697 million cells on 8192 cores (512 nodes) and completed in 70 minutes.



It is a key capability of the AGOS inversion software to allow users with different level of skills to apply and extend the software to their particular problems. Using python as a programming language, the program supports users of varying skill levels for both programming and mathematics. The numerical package escript is the computational back end for the inversion process. It provides a generic computational environment for solving partial differential equations which are at the core of geophysical inversions. Its numerical solver can run in parallel on supercomputers and also offers the ability to work on unstructured 2D and 3D meshes, which enables the inversion of data with spatially variable resolution. The escript software is maintained and distributed through AuScope CRIS and NCRIS2 funding and can be downloaded. For more information contact Lutz Gross, School of Earth Sciences.

Meet RCC: David Green, Deputy Director

Dr David Green majored in physics and applied mathematics at UQ, thankfully after the demise of flared jeans and platform shoes. His first real science job was in environmental modelling for the Qld Electricity Commission.

David completed a PhD in the School of Physics at the University of Sydney on the properties of electrochromic thin film window coatings that control solar heat gain. Even if we could have solved all the materials science challenges, people still did not like the colour! He worked as an academic in Applied Physics and Computer Systems departments at UTS for almost fifteen years. His commitment to fostering student learning was recognised in various forums. He career-changed into research computing support roles at Griffith University in the mid-2000s. David joined The University of Queensland as HPC Manager within ITS in 2007.

David has long been using computers to "figure stuff out" and enjoys working with researchers to help them to use computers to "figure stuff out". He has been Deputy Director of the RCC since its establishment in 2011.