Enabling Research Outcomes Using QCIF HPC Facilities

Gacenga, F.[1], Bell, J.[2], Samalca, I.K.[1], Brieva, D.[1]

[1]University of Southern Queensland, Toowoomba, Australia
[2]Central Queensland University, Rockhampton, Australia

QCIF HPC facilities

QCIF (Queensland Cyber Infrastructure Foundation) partly funds the procurement of High Performance Computing (HPC) facilities across its Queensland member universities. One of these is USQ's Sun X2200 HPC running on an older RHEL 5.7 Linux distribution. The USQ HPC is supported by USQ's ICT services and has been serving the computing needs of USQ researchers since 2008. The use of HPC and research support services are available to staff and students to enhance research outcome and productivity. To facilitate the collaborative use of HPC and other eResearch services, QCIF has an established network of eResearch analysts at each of the member universities.

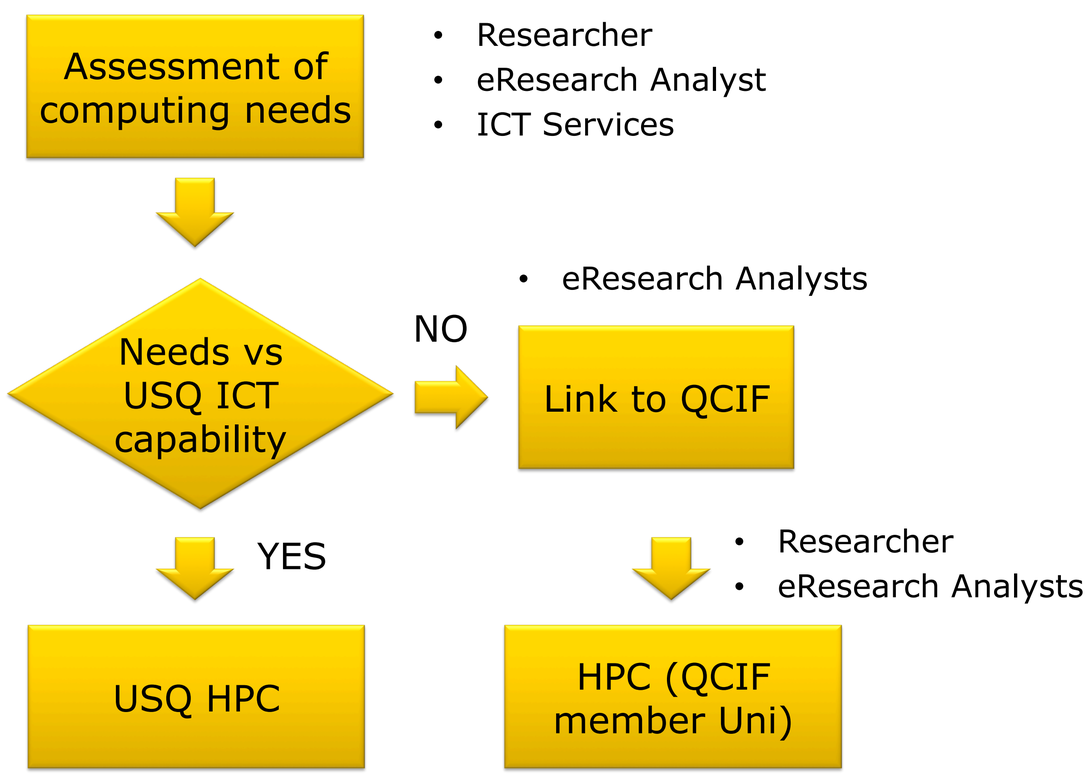

Figure 1. Processes involved in the provision of eResearch services to researchers from QCIF member universities.

Case Studies

Case 1: Maize crop simulation using APSIM in the Northern Grains Region

The research project integrates spatial scale in the point-based Agricultural Production Systems sIMulator (APSIM) in order to better understand the effect of climate variability on crops on a spatio-temporal perspective. The project involves performing crop simulation using APSIM on a 0.05 degree grid (approx. 5 km) covering the cropping areas within GRDC Northern Grains Region, totalling 9,197 grids. A record of 114 years of climate data was used to drive the simulation. The output is a spatio-temporal quantification of potential yield as influenced by changes in climatic variables (mainly rainfall and temperature), and crop management (timing of sowing).

Computing requirements, challenges and solution

- CPU intensive simulation that is not possible to perform using ordinary desktop PCs;
- In-house HPC cannot be used due to incompatibility of libraries used between APSIM and USQ HPC;
- The required software and hardware is available at CQU HPC;
- Through QCIF, USQ researcher was given access and support to the relatively newer and more powerful CQU HPC.

Case 2: Circulation of the Southeast Queensland Coastal Ocean

The research project involves numerical simulations to describe the marine circulation off Southern Queensland using ROMS_AGRIF. The numerical simulations involved more than 5 million points in a 3-dimentional matrix representing the region. Due to the large amount of points, smaller jobs covering fewer points were created through Message Passing Interface (MPI) libraries, and executing them in parallel under the HPC environment. This approach significantly reduced the overall simulation time to describe the marine circulation behaviour (e.g. velocity, salinity and temperature). Results of this study complements the description of the marine circulation based on variables estimated from remote sensing and few measurements that exist along this section of the marine coast.

Computing requirements, challenges and solution

- Setup the simulation in a workstation environment, but more computing power is required;
- Access and support to CQU HPC was granted;
- Migration to CQU HPC was a challenge collaboratively solved by the researcher and HPC administrator;
- Improvement of codes to efficiently implement the simulation.

Conclusion

Collaborative eResearch provides solution to the researchers' computing needs, even beyond the institution level. This scheme improves utilisation of eResearch resources by centrally maintaining high performance cluster and made available to wider users. Technical and administrative support to the use of the computing resources are also in place to help the researchers.

ACKNOWLEDGEMENT

We would like to acknowledge the efforts of **Richard Young**, USQ HPC support officer, and **Peter DeVoil**, APSIM developer, based at DAFF for their efforts in installing APSIM on a HPC running Linux.

The HPC Next Door – Enabling Research Outcomes Using QCIF HPC Facilities

Francis Gacenga[1], Jason Bell[2], Irvin Samalca[1], Daniel Brieva[1]

1University of Southern Queensland, Toowoomba, Australia, gacenga@usq.edu.au
2CQUniversity, Rockhampton, Australia, j.bell@cqu.edu.au

INTRODUCTION

The paper begins with a background of the problem, a description of the research projects requiring use of HPC, followed by the problems with accessing the HPC resources required, and an account of the approach taken to solve the problem. A description of the solutions is provided, followed by a discussion on the research outcomes and a reflection on lessons learnt.

Background:

Researchers at the University of Southern Queensland (USQ) have had access to a High Performance Computing (HPC) centre on the Toowoomba campus since 2008 [1]. The USQ HPC is supported by USQ's ICT services division and is listed on the catalogue of services under research computing and support is provided by HPC support based in the Network Infrastructure team [2]. The USQ HPC is part of a network of HPCs part funded by QCIF (Queensland Cyber Infrastructure Foundation) and available to QCIF member universities. QCIF states that "HPC and research support services are available to all staff and students and are designed to enhance research outcomes and productivity"[3]. To facilitate the collaborative use of HPC and other eResearch resources QCIF has an established network of eResearch Analysts at each of the member universities.

The collaborative use of QCIF HPC facilities has led to two research projects at USQ accessing solutions that have enabled them to progress their research projects. In both projects the researchers began their quest for compute power by using faculty maintained laboratory high-end computers. In the first research project, "Maize crop simulation using Agricultural Production Systems sIMulator (APSIM) [4] in the Northern Grains Region" the research team began its quest for a HPC solution at USQ in September 2013 aiming to access 200 or more CPU cores to reduce the time to run simulations. Meetings with the researchers led to a clearer understanding of the research project requirements and initial attempts at accessing a NeCTAR VM with the required number of VCPUs were unsuccessful leading to an extensive search for a HPC solution. The second research also running at ICACS was at the start of the modelling stage of a PhD research project. The research project uses Regional Ocean Modeling System (ROMS) [5], the French version (ROMS-AGRIF). A few tests had been run on the USQ HPC to see if the model was able to use the libraries, and test a basic configuration. The researchers had limited experience in compiling the code with all the libraries MPICH2, NetCDF and the model written in FORTRAN was suited for HPC and had successfully run test simulations on a smaller scale. The project aimed to improve productivity by running on a larger number of parallel cores.

Maize crop simulation using APSIM in the Northern Grains Region

The research project combines spatial scale in the point-based Agricultural Production Systems sIMulator in order to better understand the effect of climate variability on crops on a spatio-temporal perspective. The research study aims to identify opportunity cropping for maize in the light of changing climate. The project involves performing crop simulation using APSIM on a 0.05 degree grids (approx. 5 km) covering the cropping areas within GRDC Northern Grains Region, totalling 9375 grids. The simulation looks at maize cropping opportunities, using 113 years of climate data, in examining potential yields through changes in management such as timing of planting with respect to occurrence of rainfall. This is part of a PhD project under the Faculty of Health, Engineering and Sciences, USQ.

The majority of APSIM users perform simulation using stand-alone computers on a windows environment, as in the case of the maize crop simulation researcher. In typical applications, this setup is enough. However, in more extensive studies involving thousands of simulations more computing resources are required. DAFF and CSIRO have been using a grid-computing infrastructure using Condor as a solution to this kind of job. Another approach is to set up APSIM in a HPC environment to leverage more on a centralised computing resources and job management. In the above case study, simulation time can take around five days to complete using stand-alone computers with 8 cores and 3.0 GHz CPU speed. Performing the same job using the grid computing infrastructure can potentially cut down the completion time in half depending on the availability of computing resources. The desire to cut down the simulation time further presents the idea of running APSIM under HPC system.

An attempt to install APSIM on the USQ HPC was made, however, after hours of building installation libraries and consultation with APSIM developers, the attempt was unsuccessful due to incompatibility with the version of Linux on the USQ HPC. APSIM is designed to run typically on Windows platform and there were few resources or guides on how to implement it on Linux based HPC. Another attempt was made to install APSIM using CQUni HPC. While the installation and testing was successful, getting APSIM to work in the HPC facility was a challenging endeavour involving a number of complex processes and consultation between CQUni HPC officer and eResearch Analyst, APSIM developer, researcher. Running APSIM simulation on a HPC system may be the first attempt of this kind to support a research project, providing benefits including access to larger and shared computing resources and potentially cutting down simulation time by a factor of 35. This also presents opportunity of scaling up the scope of simulation project as more computing power becomes accessible.

The successful installation, testing and use of APSIM on a HPC running Linux may set a precedent that will impact on facilitating the use of APSIM on HPC as well as provide a model that can be used on eResearch service delivery, HPC collaboration and resource sharing. Researchers at USQ, DAFF and CSIRO stand to benefit from the higher number of cores available from running APSIM on a HPC. It has also been established that there are other researcher using APSIM at CQUni and there may also be others at the QCIF member universities.

Circulation of the Southeast Queensland Coastal Ocean

In this project the PhD program splits the research into two main stages. The core of the second stage involves numerical simulations to describe the marine circulation base on ROMS_AGRIF. In the second year of the PhD program the first numerical tests were ready to be implemented in a High Performance Computer (HPC). The researcher had access to the ICACS research centre workstations enabling the researcher to configure, test and analyse the numerical simulations. The researcher had no experience working in a HPC environment and was in need of extra support to effectively use the HPC.

The first step involved running the model on a workstation PC, which helped with understanding of the complexity, involved in providing a clean link between the numerical model and all the libraries needed. The next stage involved the use of a HPC and it was found that this required interaction with the HPC administration team familiar with the HPC. The researcher relied on assistance from the HPC administration team as the researcher was familiar with using the numerical model but not with aspects of using the HPC. Communication between the HPC administration team and the research project team proved crucial to the progress of the project. A challenge faced was access to the HPC administrator who did not have time to review and assist improve the code that the researcher needed to implement. In most cases the HPC support team faces competing priorities and have preference working with research projects with code that works. In this project it proved important to ensure that the model worked and a high-end local workstation and access to a small cluster allowed for preliminary testing simulations prior to using the HPC. This may however not be the case for all research projects as a model may work on a local workstation but does not work on the HPC. In such as case there may be need for availability of additional support from the HPC administration team to save the researcher crucial time. In this research project a significant breakthrough was achieved following assistance with rewriting of simulation code and implementation at CQU HPC resulting in significant research outcomes that include time saving, efficient code, and model completion.

The success of the second research project on ocean modelling will be useful not only to USQ researchers but also confirms the value in collaborative eResearch service delivery. The liaison provided by USQ eResearch services enabled the communication with extra HPC administration support from the CQUni HPC that provided the required assistance involving model implementation in a new cluster, issues resolution, HPC process guidance, and advise on HPC error handling. The time required by the researcher to learning as well as get familiar with the HPC, was shortened and the model worked in less than four days. The research model is now able to use 195 cores with its highest efficiency at around 110 cores. The research has already achieved some results from the model showing circulation and the calibration process. The task using the local workstation would have taken weeks but with the CQUni HPC was complete in four days.

The experiences for both research projects mirrors the findings in research institutions in North America using HPCs, which listed “another benefit of the system to faculty and students is that faculty does not have to waste research students’ time having them maintain a departmental computer cluster. And finally, faculty and students can scale their research to any size they want because of the size of the HPC platform”[6]. The communication between user and administration is vital for the one without experience in the use of HPC. Take time to explain the new user details that in the future will save time. Working in collaboration across University HPCs in QCIF has led to a solution for researchers.

ACKNOWLEDGEMENT

We would like to acknowledge the efforts of Richard Young, the USQ HPC support officer and Peter Devoil, APSIM developer based at DAFF for their efforts in attempting to install APSIM on a HPC running Linux.

REFERENCES

[1] USQ, HPC Facilities, in: USQ HPC Home, CRICOS: QLD 00244B | NSW 02225M | © University of Southern Queensland | Privacy |, 2010.
[2] USQ, Service Catalogue, in, 2014.
[3] QCIF, HPC, in, 2014.
[4] DAFF, APSIM Initiative, in, 2014.
[5] O.M.G. MCS, Regional Ocean Modeling System (ROMS), in, MCS, Ocean Modeling Group, 2014.
[6] D. Geer, High Performance Computing: How the investment in heavy-duty computer processing power is paying off in research output, in: University Business, 2013.

ABOUT THE AUTHOR(S)

Dr Francis Gacenga is the QCIF (Queensland Cyber Infrastructure Foundation) eResearch Analyst at the University of Southern Queensland (USQ). He has 15 years' experience in IT service management as a researcher and practitioner. Francis has experience working in government agencies in IT service strategy, design, transition, operation and continual service improvement. He offers a service oriented approach to eResearch service delivery beginning with an understanding of the researchers requirements and offering a suite of cloud compute, storage and collaboration tools, platforms and infrastructure as services that enable data-driven, collaborative research that fosters an interdisciplinary research culture. He conducts research at USQ and has presented research at international conferences and published articles in academic and industry journals. In 2002 the Institute for the Management of Information Systems (IMIS) UK awarded him a gold medal. His research interests include IT service management, eResearch, ITIL, service science, performance measurement, content analysis and design science. He has served as a reviewer in a number of international information systems journals and conferences.

Jason Bell Jason is the Senior Research Technologies Officer at CQUniversity Australia and video collaboration champion and eResearch Analyst for the Queensland Cyber Infrastructure Foundation (QCIF). He supports national video collaboration tools, such as the Access Grid and SeeVogh, provides technical feedback to state and national infrastructure service providers on behalf of CQUni, manages the CQUni video collaboration Access Grid facilities, and provides High Performance Computing Support, through Systems Administration of HPC facilities and technical advice and support to researchers. He also manages the CQUniversity dedicated research data storage facilities and assists CQUni researchers in the use of eResearch tools while promoting their use in achieving better research outcomes. He is Chair of the Central Queensland Junior Robotics Competition and the lead of a HPC SIG. Jason holds a Bachelor of Information Technology (Honours) from CQUniversity Australia and his interests include video collaboration/conferencing tools, programming, nonlinear optimization, evolutionary algorithms, HPC, system administration, distributed and parallel computing and robotics.

Irvin Samalca works at the Bureau of Soils and Water Management on various capacities, but mainly on GIS analysis, cartography, and land resource evaluation. He holds a Bachelor's degree in Statistics and a Master of Science degree in Geoinformation and Earth Observation. His research interests include application of GIS and remote sensing technologies for agricultural and resource management applications. Currently, he is working on a PhD project which aims to optimise management practices to improve maize productivity in the GRDC Northern Grains Region with respect to increasing climate variability.

Daniel Brieva has a Bachelor's degree (Licenciatura) in Physics, specialising in Geophysics from the University of Concepcion, Chile. His research has mainly focused on describing marine circulation through numerical models, particularly ROMS and FVCOM, along Chile's marine coast. His current PhD project aims to describe the circulation off the Southeast Queensland coast based on the analysis of data available and simulations done by ROMS_AGRIF. The Chilean government through the BECASCHILE program sponsors the PhD project.