How to use High Performance Computing and Cloud Clusters to enhance your research outcomes

Workshop

About your presenters

Dr Paul Coddington, Deputy Director
m: 0407 074 829 | e: paul.coddington@ersa.edu.au
Dr Paul Coddington is Deputy Director of eRSA, where he has worked on eResearch projects, infrastructure and user engagement since 2002. He has also been involved with several national eResearch projects. He has a PhD in computational physics and has over 25 years of experience in eResearch, including programming some of the earliest parallel computers and working on university research and development projects in high-performance and distributed computing, computational science and research data management.

Joey Gerlach, User Experience Coordinator
m: 0403 992 570 | e: joey.gerlach@ersa.edu.au | @joeyger_lach
Joey undertook post-graduate research in conservation genomics at the University of Adelaide from 2013 to 2015. She was constantly involved in teaching work in this time, including setting up the researcher educational groups collectively known as Stats Solutions. Joey has a passion for helping researchers reach their research goals more efficiently by knowing how to use the computing tools that are available. Her goal is to reduce the barriers to using eResearch resources by making them easier to learn and more user-friendly.

About eRSA

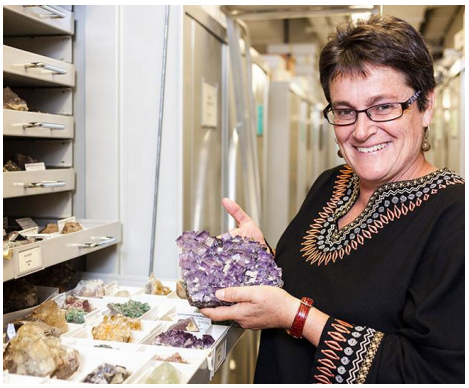
eRSA is South Australia's leading research data service provider. eRSA provides "not normal IT" services to researchers in South Australia. Founded in 2007 as provider of High Performance Computing to suits the needs of physics, astronomy, climate change and genomics researchers, we have broadened our capabilities to include cloud computing, big data storage, management and analysis, software development and consultancy.

eRSA is a not-for-profit collaborative joint venture between the University of Adelaide, Flinders University, and the University of South Australia.

By providing access to a suite of advanced ICT tools and services that are reliable, easy to use and secure, we enable all researchers to explore new and innovative research opportunities that would not otherwise be accessible.

Meet Professor Andrea Gerson

In such a complex and technical area of knowledge, Professor Andrea Gerson's research is proving invaluable for mining companies.



"I'm looking at two ends of the same process. One is extracting valuable metals and the other is reducing the environmental impact," Professor Gerson said.

"Much of my work is on the surface reactions of different types of minerals which can be broadly divided into two parts – one is looking at weathering and that's related to environmental remediation, and the other is focused on minerals processing and that's to do with extraction."

"Both of these processes are enormously valuable for mining companies as they are key to their operations."

"Quite often companies aren't necessarily sure how minerals and rock types will react with each other and how these interactions may be useful. They may not consider them to be immediately valuable, but they may be extremely important in environmental remediation."

Professor Gerson's work involves complex molecular modelling of rock surfaces, which wouldn't be possible without the powerful computer facilities provided by eResearch SA's supercomputers, Corvus and Tizard.

"Molecular modelling of minerals surfaces requires us to look at how materials such as water, oxygen or mineral processing chemicals interact with the surfaces to better understand the mechanisms of what is taking place," she said.

"The desired outcome is if you can understand the basic mechanisms you have a feel for what you can do to either improve them in terms of minerals processing or retard them, which is often the desired outcome for environmental applications."

“One of the issues with this modelling is that it’s computer intensive, so if you only have a limited CPU, you can only look at a very small model so it’s not necessarily very realistic.”

“The way that rock surfaces and minerals react is very specific, so the more exact the model the more accurate the outcome.”

“Having access to greater computer power provided through eResearch SA helps us put in place a more realistic model in terms of the number of atoms you can have in the system and that improves the predictive power of the calculations.” Professor Gerson said another benefit of working with eResearch SA was gaining access to molecular modeling software.

“Over the years, the interaction with eResearch SA has been extremely good and the service provided has been excellent. It has given us access to facilities that otherwise wouldn’t have been readily accessible,” she said.

Workbook Contents

About this workshop

This workshop is designed for new users wanting a complete introduction to HPC.

Participants will learn the basics of using a high performance computer including logging in, submitting jobs, and using the queuing system. They will also learn how to make best use of HPC for different types of tasks and applications.

The Tizard supercomputer can be used for complex data processing and analysis jobs that standard desktop computers would find it difficult or impossible to perform, and enables users to run many processing jobs with different parameters or input files more quickly.

eRSA's Cloud Cluster 'Emu' makes it easy for researchers or research groups to set up their own private clusters. Researchers can request an allocation of compute cores on the National Research Cloud, and then set up Cloud virtual machines so that they look just like a standard supercomputer cluster.

During this hands-on workshop you will:

- learn when and how to use the different components of the Tizard machine
- log in from a desktop machine
- learn how the queuing system works
- submit a job
- compile and run your own programs
- test your programs
- when to use clusters in the cloud vs the Tizard machine
- how to choose the right type of cluster for you

Workshop Agenda

9.30am – 10.40am	Introduction to eRSA and Introduction to High Performance Computing
10.40am – 11.00am	Break and refreshments
11.10am – 12.30pm	Using Tizard
12.30pm – 1:30pm	Lunch break
1.30pm – 3:00 pm	Hands-on session • Unix tutorial • Running jobs on Tizard
3.00pm – 4.00pm	Help with running your own applications on Tizard

HPC and Emu workshop - Quickstart Userguide

For more complete information on using the Tizard High Performance Computer, look at the eRSA user guides at http://support.ersa.edu.au/hpc/user-guide.html

In this Quickstart Userguide, the terms inside `<arrows>` should be replaced with user input.

Logging On to the HPC Machines

Windows

Use a program called Putty

- Here is the direct link to download the putty program "Putty.exe"
  http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html
- Open the PuTTY application.
- Enter **"tizard1.ersa.edu.au"** as the hostname.
- Ensure the "Connection type" is **"SSH"** and set the "Port" to **"22"**
- Click the "Open" button, and if you are asked if you trust the connection, enter "yes".
- Enter your eRSA username and password at the prompts.

Linux and Mac

Use ssh on the command line (in the Terminal app) with this command:

```
ssh <USERNAME>@tizard1.ersa.edu.au
```

Note: `<USERNAME>` is your eRSA HPC username
You will be prompted to enter your eRSA password.

FileZilla SFTP connection to Tizard

- Install FileZilla from https://filezilla-project.org/download.php?type=client
  (if you have Ubuntu OS use the Ubuntu Software Centre to install)
- Open FileZilla and select **"File" --> "Site Manager"**
- Create a "New Site" called tizard1.
  - Host : tizard1.ersa.edu.au
  - Port : 22
  - Protocol : SFTP
  - Logon Type : Normal
  - Enter your username and password and click "Connect"

Setting up a job submission

You are now logged on to the tizard submission node. You can run small jobs here directly, but for most jobs you will modify a PBS submission script to the Torque queueing system.

- See the submission script templates in a hidden folder called ".templates" located in your home directory.

```
ls -a
ls .templates
less .templates/tizard.sub
```

- Create a directory to work from in "/scratch" and create a symbolic link to this folder in the home directory

```
mkdir /scratch/$USER
```

- Copy the submission script to the directory

```
cp .templates/tizard.sub /scratch/$USER/
```

- Navigate to the directory (using the quick link, if you like)

```
cd /scratch/$USER/
```

Edit the submission script

We will be running an example job using the package "Stacks"

- Use the module commands to explore the availablity of the package, and the dependencies that must be loaded prior to loading the package.

```
module avail
module avail stacks
module whatis stacks
```

- Copy the package names listed under "Required Modules"
- Edit the submission script with `nano tizard.sub` (or use the file edit option in FileZilla)

Terms to replace when editing a submission script

Replace the following terms in the submission script (shown in **bold** *).*

Template Script

MyJobName	give your job a descriptive name
Your-email-Address	your email address is used to alert you of job completion
#PBS -l nodes=1:ppn=**Y**	enter the number of processors required
#PBS -l mem=**X**gb,vmem=**X**gb	enter the amount of memory required
#PBS -l walltime=**HH:MM:SS**	enter the maximum time the process will take.
module load **application_module**	load any required modules first (see the output from the `module whatis` command above). Then on a new line, load the program you will use.
MyProgram+Arguments	Enter the command to start the program, as you would enter it on the command line.

Running an Example Job Submission

We will be running an example job using an example data set that we can copy from a shared folder.

1. Copy the dataset and the submission script to your scratch folder. Explore the contents of the example folder.

```
cp -r /opt/shared/training/HPCworkshop ./workshop
cd workshop
ls -lh
ls -lh data
nano tizard_exp1.sub
```

2. Submit the job to the queueing system, then check the status of the job.

```
qsub tizard_exp1.sub
qstat
```

3. Check the log file and the results folder:

```
ls -l
less Experiment1.o<XXXXX>
ls results/
```

Emu - Cluster in the Cloud

Using Emu is almost identical to using Tizard. The same submission script can be modified to use in Emu. The full user guide is found on the eRSA support page http://support.ersa.edu.au/hpc/emu-guide.html

Logging on to Emu

This is the same as logging on to tizard, except:

Host : emu.ersa.edu.au

or

```
ssh <USERNAME>@emu.ersa.edu.au
```

Notice the home folder at log-in is the same as the tizard1 home folder.

Look at the Emu submission script template:

```
ls -la
ls .templates
less .templates/emu.sub
```

Running a job on the Emu Cluster

Navigate to the Emu '/scratch' directory to submit a job:

```
mkdir /scratch/$USER
cp .templates/emu.sub /scratch/$USER/
cd /scratch/$USER
```

Copy the example data folder to this location:

```
cp -r /opt/shared/training/HPCworkshop ./workshop
cd workshop
ls -lh
```

Edit the submission script for use with Emu

```
nano tizard_exp1.sub
```

Make the following changes:

```
#PBS -q tizard    -->     #PBS -q emu
```

```
#PBS -l mem=.....etc     -->   ### PBS -l mem=.....etc
```

Use control 'o' to save the modified file to the name **"emu_exp1.sub"**

Run the job:

```
qsub emu_exp1.sub
qstat
```

Check the log file and the results folder:

```
ls -l
less Experiment1.o<XXXXX>
ls results/
```

Transferring data

FileZilla and WinSCP

These sftp clients are an easy way of moving files between a remote computer (e.g. the Tizard or Emu /scratch folders) and your desktop computer. Just point and click to find the required folders, then drag and drop the files between the two locations.

SFTP via the Command Line

Secure file transfer using shell commands allows you to transfer data between two remote data storage locations (e.g. between your Emu /scratch folder and your eRSA /data long-term data storage).

Enter the 'sftp' command while logged on to 'emu.ersa.edu.au', and you will have access to the remote storage server.

`sftp username@sftp.ersa.edu.au` - you will be prompted for a password.

You are now accessing the remote data storage server, and you can navigate the files on the server as per usual with commands like `cd` and `ls`. The commands `get` and `put` will transfer data between the machines:

- `get <remote_server_file.txt> <localEMU_destination/>`
- `put <localEMU_file.zip> <remote_server_destination/>`

  to close the sftp connection, type `exit`.

Quick reference for commands

Command	Action
`module --help`	help page for the 'module' commands
`module avail`	list all available packages in the repository
`module avail <search term>`	list package names containing the search term
`module list`	list the packages that are already loaded on the VM
`module show <name>`	show info on the package, and lists the required modules to pre-load
`module whatis <name>`	other information about the package
`module load <name>`	load the package onto the VM
`module unload <name>`	remove the package 'cache' from the VM
`cd /scratch/$USER`	navigate to the folder from which jobs are run
`nano <file>.sub`	edit a submission script
`qsub <file>.sub`	send a submission script to the queueing system
`qstat`	check the status of submitted jobs
`qdel <jobID>`	delete a job from the queue (check jobid from `qstat`)

Can we help you more?
Check out our upcoming workshops!

Our training workshops can help advance your research further and faster than ever. Our interactive workshops teach generic skills and tools learnt will apply to many types of analysis in any discipline. C**heck out our workshop calendar: https://www.ersa.edu.au/service/workshop-training/.**

Upcoming eRSA Training Workshops – *Talk to us after the workshop to register!*	
April 2016 **Running a Virtual Machine in the Cloud** *14 April, University of Adelaide*	***Register Now***
R-Studio in the Cloud *15 April, University of Adelaide*	***Register Now***
Data Storage/Management for Your Research (University of Adelaide)	
July 2016 Running a Virtual Machine in the Cloud (UniSA, Flinders) Windows Remote Desktop in the Cloud (University of Adelaide) Advanced Cloud Computing (University of Adelaide) Intro to Big Data for Researchers (University of Adelaide) Data Storage/Management for Your Research (UniSA, Flinders) HPC Emu (University of Adelaide) Security Tips (University of Adelaide)	
September 2016 HPC Emu (UniSA, Flinders) Running a Virtual Machine in the Cloud – Remote Desktop (University of Adelaide) R-Studio in the Cloud (University of Adelaide) Data Storage/Management for Your Research (University of Adelaide)	
December 2016 Advanced Cloud Computing (University of Adelaide) HPC Emu (University of Adelaide) Intro to Big Data for Researchers (University of Adelaide) Running a Virtual Machine in the Cloud (University of Adelaide)	

Can't attend a workshop?

We can deliver our workshops onsite, customized to your needs. Email us at **comms@ersa.edu.au** and we'll touch base to discuss a training solution for you, your lab or office.