

# Running a Virtual Machine in the Cloud Workshop

## 25 February 2016

## About your presenter

**Joey Gerlach, User Experience Coordinator**

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.

t: 08 7228 5470 | m: 0403 992 570 | e: joey.gerlach@ersa.edu.au | w: [www.ersa.edu.au](http://www.ersa.edu.au)

facebook.com/eResearchSA  
twitter.com/eResearchSA

## 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 Dr Claudia Junge,**   
Cloud computing user and protector of Australia’s shark population

“The software and Nectar Cloud resources I have been able to use through eRSA have been incredibly useful, especially for the population analyses.”

Marine biologist Dr Claudia Junge has had a fascination with the ocean since she first went scuba diving at the age of 14. So after studying biology in Germany and undertaking a PhD at the University of Oslo, Norway, she decided to discover the deep seas of the South.

An ARC Research Associate in the School of Biological Sciences at the University of Adelaide, Dr Junge is using eRSA and the Nectar Cloud to power her research into Australia’s dusky shark and bronze whaler shark populations. Dr Junge is working with a multidisciplinary team of researchers as well as multiple government and industry partners to find out how many genetic stocks of dusky and bronze whaler sharks there are in Australia, in order to sustainably manage the species.

After extracting DNA from shark tissue samples, Dr Junge then uses next-generation sequencing (NGS), applying a method that involves cutting the genome into smaller fragments at specific recognition sites.

“Because of this, I end up with thousands of single nucleotide polymorphisms (SNPs) across the whole genome,” Dr Junge said.

“To run bioinformatic analyses on so many SNPs and samples you need a number of resources and it’s just not possible to have all of these on your desktop computer – this is where eRSA really comes in handy.”

“Once I have extracted the genetic information, such as levels of gene flow, we then work with modelling researchers, and the chemistry information our collaborators have found, to parameterise species specific spatially explicit population models.

“The software and Nectar Cloud resources I have been able to use through eRSA have been incredibly useful, especially for the population analyses, because you need specific programs, which they’re always willing to install.

“I have datasets that include over 10,000 SNPs for up to 300 different individuals and just one of these analyses can take 150 hours, not to mention that I then have to do this in replicates of 20 for 10 different settings – if I even attempted to do this on a desktop it would take forever.

“It’s great that researchers like me can use eRSA’s resources – I use the cluster in the cloud to access a high-performance computing cluster that uses virtual machines – it makes my life a lot easier and so much faster to get the results I need.”

The long-term result of Dr Junge’s research will be better fisheries management.

“As both species are fished in Australia – bronze whalers predominately in South Australia and dusky sharks predominately in Western Australia – and only produce very few offspring, compared to most commercially fished species, our studies are also important in ensuring that Australasian stocks are not being overfished.”

Contrary to previous studies, results from Dr Junge’s project have already shown that dusky shark populations around Australia are made up of the same genetic stock.

“We found that dusky sharks being fished in Indonesia belonged to the same genetic stock as our Australian samples,” Dr Junge said.

“This is important for fisheries management to keep in mind, particularly when making sustainability agreements interstate and internationally.

“Interestingly, bronze whaler sharks in Southern Australia are also very mobile and seem to be connected genetically.

“But from what we can tell, samples from Western Australia and also from around the Great Australian Bight are somewhat different, genetically, from the southern and eastern side of Australia as well as New Zealand.”

This research was supported under the Australian Research Council’s Linkage Projects funding scheme (project LP120100652).

## About this workshop

The aim of this workshop is to make using Nectar cloud computing a little less intimidating by walking through the common steps in running analyses on a VM.

This workshop is for researchers who:

* Haven't used cloud computing but would like to learn how to and how it can help their research.
* Have started using cloud computing, but are finding the learning curve steep.
* Currently use Nectar services and are looking for tips to make better use of the resources.

This tutorial will demonstrate some basics of setting up and running a virtual machine (VM) in the Nectar research cloud. We will install packages, download files, and run a demonstration analysis with some optimisation tips.

The workshop will teach generic skills, and the tools learnt will be applicable to many types of analysis in any discipline. The tutorial will involve a test analysis using an example program called "Structure" with the task manager application "GNU Parallel", such that the analysis is optimised for a VM with a large number of processing cores.  
   
Agenda

1. Introduction to the Nectar research cloud and the different ways it can benefit your research.
2. Launching a Nectar virtual machine instance.
   1. SSH keys
   2. Allocations and Security settings
   3. Launch an instance on the Nectar dashboard
   4. Connect to your VM
3. Exploring your VM
4. Installing packages with a package manager
5. Downloading and installing packages from the internet
6. Transferring files to and from the VM
7. Running the program "Structure" using "Parallel"
8. Disconnecting the process from the terminal shell.

## Today’s Tip

### Using Byobu on your Virtual Machine

Byobu is a “virtual console” that will save your Terminal/PuTTY session when you are disconnected from the VM. This keeps programs running and allows you to resume your computing from where you left off the last time you used it.

To load and begin using byobu, enter the following commands:

sudo apt-get install byobu

byobu-select-backend screen

byobu

Learning all the features can take a little time, but if you use the Linux command-line to access virtual computers frequently there is a great benefit in using the byobu virtual console.

* It keeps your terminal session running perpetually, you can reconnect at any time.
* This includes keeping active any running processes attached to the terminal session.
* The next time you log in, the terminal session will be preserved as you left it.
* You can create multiple tabs in the same session, all of which are saved and can be resumed later.
* You can split screens to see multiple tabs in one window.
* You can save a log file of all the activity in a tab.
* See <http://byobu.co/> for a video demo of byobu.

## See the following NeCTAR support page for more information: <https://support.nectar.org.au/support/solutions/articles/6000089713-tips-for-running-jobs-on-your-vm#screen>

# Workshop: Using a Nectar Cloud Virtual Machine

This guide will show you how to use your Nectar VM to perform jobs, including tips on making efficient use of cloud capabilities. This tutorial will demonstrate some basics of setting up and running a VM. We will install packages, download files, and run a demonstration analysis with some optimisation tips.

The aim of the workshop is to make using a Nectar cloud a less intimidating experience by walking through some of the common steps in running analyses on a VM.

The workshop will teach generic skills, and the tools learnt will be applicable to many types of analysis in any discipline. The tutorial will involve a test analysis using an example program called "Structure" with the task manager application "GNU Parallel", such that the analysis is optimised for a VM with a large number of processing cores.

## Launch a Virtual Machine (Instance)

This tutorial will provide an overview of launching and accessing a virtual machine (VM) in Nectar. These processes are covered in depth on the Nectar website Training Modules <http://training.nectar.org.au/> , and Support Pages <https://support.nectar.org.au/support/home>

1. Log on to the Nectar dashboard <https://dashboard.rc.nectar.org.au/project>
2. Select the correct project allocation in the top bar (if it starts with "pt- ", it is your default trial allocation)
3. Select "**Access & Security**" under the "Compute" subheading in the left side main menu.
   1. Click "Create Security Group"
   2. Name the security group 'SSH', with the description "port 22 for SSH". Click "Create Security Group”
   3. Click "**Manage Rules**" in the "Actions" drop-down menu. Click "**Add Rule**".
   4. Type "22" under "**Port**". Under "**CIDR**", you choose a range of IP addresses that can access your VM through port 22.
      * If you keep it as "0.0.0.0/0", you can access your VM from any computer, but others might find a way to access it also. Here are the IP ranges for SA University networks. Use one of these, and the VM can only be accessed form a University computer (including an external computer using a VPN to access the University network).
        + 129.127.0.0/16 - University of Adelaide
        + 129.96.0.0/16 - Flinders
        + 130.220.0.0/16 – UniSA
   5. Select "**Key Pairs**" and use "**Import Key Pair**" to upload the public key of a keypair you have created in PuttyGen (Windows users) or with

ssh-keygen -f ~/.ssh/keyname.  
See <https://support.nectar.org.au/support/solutions/articles/6000055376-launching-virtual-machines> for instructions on setting up keypairs.

1. Select "**Instances**" in the left side main menu. Click "**Launch Instance**".
2. Give your VM a name, choose a "**Flavor**" (size of the VM), and select an Image (pre-loaded operating system, e.g. Nectar Ubuntu 14.04).
3. Select the "**Access & Security**" tab. Select your Key Pair name, and select the "**SSH**" security group.
4. If you are going to use a volume storage attachment, select the "**Availability Zone**" and choose "**sa**" from the drop-down menu.
5. Click "**Launch**". It may take a few minutes for the instance to boot.

## Connect to your VM

See the support page:  
<https://support.nectar.org.au/support/solutions/articles/6000055446-accessing-instances>

1. Copy the IP Address of the instance.

* Windows - Set up a PuTTY connection with the IP address
* Mac/Linux - In the terminal app, enter ssh -i <keyname> ubuntu@<IPaddress>

## Getting acquainted with the VM

* Try some commands to look around the VM
  + lsblk ; df -hT ;ps ; top
* Install "**htop**" using Ubuntu's "Advanced Package Tool (apt)"

apt-cache search htop

sudo apt-get install htop

htop

* Connect to the VM with FileZilla to visualise the file structure and contents.  
   <https://support.nectar.org.au/support/solutions/articles/6000085114-transferring-data-to-your-vm#filezilla>

## Some handy commands to run the first time you access your VM

* sudo passwd ubuntu - Set a password for the user "ubuntu"
* sudo chown ubuntu /mnt - Make user "ubuntu" the owner of the mounted storage in /mnt

Run the following commands periodically to update the package manager and the installed packages. We won't run them today because they can be time-consuming.  
sudo apt-get update  
 sudo apt-get upgrade

## Running an example data analysis job

We will be running a job with the population genetic program "Structure". This program is a great example of an "embarrassingly parallel problem" that is ideal for running on a cloud VM with a large number of processors.

## Downloading and decompressing files

* To download "Structure", copy the URL for the latest release <http://pritchardlab.stanford.edu/structure_software/release_versions/>
* Then use the "wget" command to download the tar.gz file. e.g.  
   wget http://pritchardlab.stanford.edu/structure\_software/release\_versions/v2.3.4/release/structure\_linux\_console.tar.gz
* Decompress the archived folder with:  
   tar -zxvf structure\_linux\_console.tar.gz
* Download today's tutorial file:  
   wget https://github.com/joeygerlach/structure\_parallel/archive/master.zip
* Use the *unzip* command to decompress the tutorial file  
   sudo apt-get install zip unzip  
   unzip master.zip

## Moving the analysis to the mounted storage "/mnt"

Always perform analyses in a larger storage disc, not the primary (root) disc. In this tutorial we will use the secondary (ephemeral) disc, but it is recommended that users request volume storage to use for running analyses instead because the data will be more secure.

* First, we make the mounted storage disc writable:  
   sudo chown ubuntu /mnt
* Copy the tutorial files to the mounted disc:  
   cp -r ~/structure\_parallel-master/ /mnt/structure\_run1/  
   cd /mnt/structure\_run1
* Look at the files (you may wish to use FileZilla for this)  
   ls -lhF

## Running the program - using "Parallel" and "nohup"

* Look at the support page on running jobs on a VM:  
  <https://support.nectar.org.au/support/solutions/articles/6000089713-tips-for-running-jobs-on-your-vm>
* Install the task manager "GNU Parallel"  
   sudo apt-get install parallel
* Run Structure using "GNU parallel" by calling the shell script:  
   bash Parallel\_Structure.sh
* Look at the processes running:  
   htop
* Because we used the **nohup** command, we can close the window and the process will continue running. Try closing your terminal, then re-connecting to the VM. Enter htop.

## Transferring files

We have created a folder of results files. We will transfer them to your computer, or to a remote data storage server.

First, we can zip the results folder into a single file.  
 zip structure\_1.zip results/\* OR  
 tar -zcvf structure\_1.tar.gz results/

## Filezilla

Drag and drop files between your VM and your local computer.

## Secure File Transfer Protocol

Use the **sftp** command to access a remote computer, then use "get" to download or "put" to upload files.

sftp jgerlach@sftp.ersa.edu.au  
get test.txt /home/ubuntu/  
put structure\_1.tar.gz ./  
quit

**Congratulations, you have now:**

* launched a virtual machine in the NeCTAR cloud,
* installed applications,
* downloaded software and files,
* run an analysis program,
* used "parallel" and "nohup",
* and transferred data for backup.

**Enjoy adapting these skills to the analysis programs you use in your research!**

## Quick reference for shell commands

Connecting from the terminal app on your Mac:  
ssh -i <Nectar\_Key> ubuntu@XX.XX.XX.XX

### Commands for your Linux VM

|  |  |
| --- | --- |
| Command | Action |
| sudo passwd ubuntu | set a password for user 'ubuntu' |
| sudo chown ubuntu /mnt | make the ephemeral disk writable |
| lsblk | list the block storage |
| df -hT | display the disk usage |
| du -h <path/to/directory> | display directory and file sizes |
| top or htop | activity monitor for your VM |
| ps | list the running processes on your VM (with PID#) |
| kill <PID#> | terminate the process by PID number |
| control + 'c' | stops a process running in your terminal |
| sudo apt-get update | updates the list of packages available to install |
| sudo apt-get upgrade | upgrades the installed packages |
| apt-cache search <name> | search for a package to install |
| sudo apt-get install <name> | install a package |
| nohup <normal commands go here> 2>&1 & | keep a job running in the background |
| jobs | list the active jobs (with job numbers) |
| control + 'z' | pause a job running in the foreground |
| disown %n | detach a (paused) job from the terminal session (n=job number) |
| bg %n | move a (paused) job to the background (n=job number) |
| parallel < list\_of\_commands.txt | send a list of jobs through the task manager "Parallel" |
| nohup parallel < list\_of\_commands.txt 2>&1 | Using nohup and parallel together |
| wget <URL> | download a file from the internet |
| tar -zxvf file\_name.tar.gz | extract a gzip compressed tar archive |
| tar -zcvf archive.tar.gz folder/ | create a compressed archive of a folder |
| unzip file.zip | Uncompress a zip archive |
| zip archive.zip \*files | create a zip archive of one or more files or directories |

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

Our training workshops can help you advance your research further and faster than ever. Our interactive workshops teach generic skills to users, and the tools learnt will be applicable to many types of analysis in any discipline. **You can check out our workshop calendar here:** [**https://www.ersa.edu.au/service/workshop-training/**](https://www.ersa.edu.au/service/workshop-training/)**.**

|  |  |
| --- | --- |
| Upcoming eRSA Training Workshops – *Talk to us after the workshop to register!* | |
| **March 2016  How to use HPC and Cloud Clusters to enhance research outcomes**  *9 March, University of Adelaide* | ***Register Now*** |
| **April 2016 Running a Virtual Machine in the Cloud**  *14 April, University of Adelaide*  **R-Studio in the Cloud**  *15 April, University of Adelaide*  Data Storage/Management for Your Research (University of Adelaide) | ***Register Now   Register Now*** |
| **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**](mailto:comms@ersa.edu.au) **and we’ll touch base to discuss a training solution for you, your lab or office.**