Using a NeCTAR Cloud Virtual Machine

Joey Gerlach

25 February 2016

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.

1 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 adresses 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.

4. Select “**Instances**” in the left side main menu. Click “**Launch Instance**”.
5. 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).
6. Select the “**Access & Security**” tab. Select your Key Pair name, and select the “**SSH**” security group.
7. If you are going to use a volume storage attachment, select the “**Availability Zone**” and choose “**sa**” from the drop-down menu.
8. Click “**Launch**”. It may take a few minutes for the instance to boot.

2 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>`

2.1 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

2.2 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
```

3 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.

3.1 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_
  ```

- 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
  ```

3.2 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
  ```

3.3 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`.

4 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/`

4.1 Filezilla

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

4.2 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 anlysis program,
- used “parallel” and “nohup”,
- and transferred data for backup.

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

5 Quick reference for shell commands

Connecting from the terminal app on your Mac:

```
ssh -i <Nectar_Key> ubuntu@XX.XX.XX.XX
```

5.0.1 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