Course Docs at https://goo.gl/7D2YFn

Data transfer and RDS for HPC

Hayim Dar and Nathaniel Butterworth

sih.info@sydney.edu.au

Sydney Informatics Hub

A Core Research Facility

HPC Access

Example: ssh -Y <username>@hpc.sydney.edu.au

THE UNIVERSITY OF SYDNEY

Image: thecultureconcept.com

Introduction

- Sydney Informatics Hub (a Core Research Facility) alongside ICT

- Under the Research Portfolio, Core Research Facilities provide access to high-end research infrastructure, personnel and services: supporting quality research outcomes for USyd researchers

  http://sydney.edu.au/research/support/facilities.shtml

- Provides Artemis HPC and associated training and support

THE UNIVERSITY OF SYDNEY
Sydney Informatics Hub

Informatics and Research Computing Expertise

Provides research computing expertise, training, and support

- Data science, analyses and support (bioinformatics, modelling, simulation, visualisation, machine learning, statistics)

- Training and workshops
  - High Performance Computing (HPC)
  - Programming (R, Python, Matlab, Scripting, GPU)
  - Code management (Git)
  - Bioinformatics (RNA-Seq, Genomics)

- Research Computing Support
  - Artemis HPC
  - Argus Virtual Research Desktop
  - Bioinformatics software support (CLC Genomics Workbench, Ingenuity Pathways Analysis)

- Events and Competitions
  - HPC Publication Incentive – High quality papers that acknowledge SIH and/or HPC/VRD
  - Artemis HPC Symposium

Overview of today's course

- Research Data Store Introduction
- "Classic RDS" vs "RCOS"
- Data transfer between your computer, the research data store and Artemis
- Artemis data transfer queue (dtq)
- Getting accounts on HPC and RDS via RDMP

Research Data Store (RDS)

- The research data store is a central place to store digital data generated as part of your research
- https://sydneyuni.atlassian.net/wiki/spaces/RC/pages
- **No cost to USyd researchers!**
  - Access requires a UniKey and valid RDMP
- Request as much storage as you need, it is free (if it ever says there is a charge, it is wrong)
- Backed up
  - No need for an external hard drive backup!
- Secure
  - Only people you specify will have access. Not sending data to overseas/cloud storage providers with possibly non-regulatory compliant storage

Research Data Store (RDS)

- Two types of Research Data Store:
- Each project must choose one:

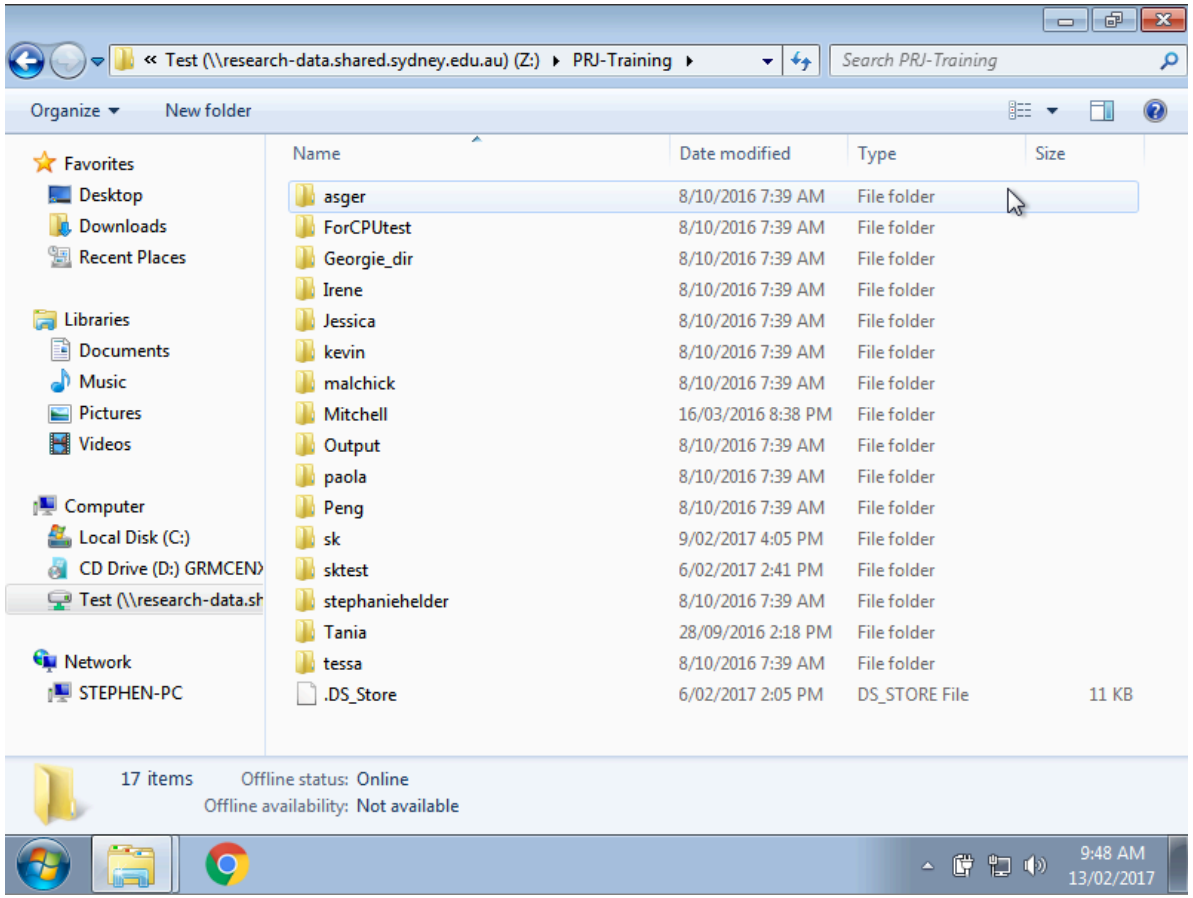
Research Data Store

Classic RDS

Research Computing Optimised Storage (RCOS)

Option 1: Classic RDS

- The best option if you don't use HPC
- Map a network drive and browse/transfer files in a graphical file explorer
- Open files in programs (eg Matlab) without transferring data to your local hard disk
- CIFS/Samba file access
  - Network accessible storage



Option 2: RCOS

- Use SFTP (secure File Transfer Protocol) to transfer data
- Command-line driven: Good for Linux and Artemis users
- RCOS is mounted on Artemis (NFS v4)
  - Transfer data to/from Artemis with cp/mv commands
- To open files on your computer, you must transfer them to your local machine first (not network accessible)

```
vlan-2689-10-17-109-95:~ stephenkolmann$ sftp ict_hpctrain1@hpc.sydney.edu.au
ict_hpctrain1@hpc.sydney.edu.au's password:
Connected to hpc.sydney.edu.au.
sftp> ls
montyhall.m                    montyhall.out                  runmonty.pbs
runmonty.pbs.e1054450          runmonty.pbs.e1268904          runmonty.pbs.o1054450
runmonty.pbs.o1054450_usage    runmonty.pbs.o1268904          runmonty.pbs.o1268904_usage
sftp> get montyhall.m
Fetching /home/ict_hpctrain1/montyhall.m to montyhall.m
/home/ict_hpctrain1/montyhall.m                               100%  527     0.5KB/s   00:00
sftp> lls
Applications      R                 computations      jnk                scaling
Desktop           RDSMount          dedalus_testing   jobs_running      sjk.kdbx
Documents         VirtualBox VMs    dell              junkdir            sourcecode
Downloads         acetaldehyde.xyz  django            mech_eng_abaqus    usage.png
Library           artemis.pub       get-pip.py        montyhall.m        veracrypt
Movies            bin               hello_world.c.txt pgadmin.log
Music             bt.tar.gz         hw48.pbs.txt      python_google
Pictures          cleantraining.sh  initdb            rocks_servers
Public            code              iso               runmonty.pbs
sftp> put bt.tar.gz
Uploading bt.tar.gz to /home/ict_hpctrain1/bt.tar.gz
bt.tar.gz                                                     100%   45KB  44.6KB/s   00:00
sftp>
vlan-2689-10-17-109-95:~ stephenkolmann$
```

Getting an account on HPC and RDS

1. Fill out Research Data Management Plan (RDMP) at https://rdmp.sydney.edu.au

   For more information on data management: http://sydney.edu.au/research_support/data/
2. Sent to Lead Chief Investigator (LCI) for approval
3. Sent to Research Data Support team for final approval
4. Sent to ICT and HPC administrators who create your accounts
5. You will receive an email when your accounts are created

- The whole process takes a few business days
- If you experience a delay, in most cases it is because the LCI has not approved it
  - You can check up on the status of your RDMP by emailing researchdatasupport@sydney.edu.au

Getting an account on HPC and RDS

- Key points re RDMP:
  - Abbreviated project name is used to name HPC and RDS directories, so make sure its meaningful/memorable to you (these are unique University-wide)
  - At least one LCI per project, must have Academic Profile Online
  - All collaborators + LCI can read and write to HPC and RDS Project directories
  - Must select 'Yes' at HPC question to have HPC access created for that project
  - You can have multiple projects
  - **Storage on RDS**: choose 'classic RDS' (CIFS) or RCOS (NFS)

Data Transfer

Overview of routes and tools for data transfer between different sources and destinations

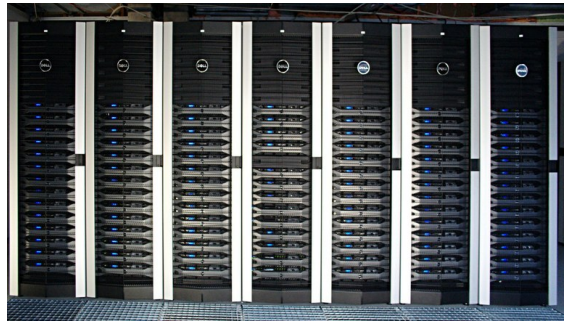
THE UNIVERSITY OF SYDNEY

Data transfer methods

- Depending on where your data is located and where you want to copy it to will dictate which transfer method to use
- **A few things to keep in mind:**
  - Artemis HPC and RDS are physically separate machines
  - Artemis HPC is NOT backed up, RDS is
  - All raw data and important output should be stored on RDS and only copied to HPC when it is required for computational analysis

RDS

RCOS (NFS) | Classic (CIFS)



http://www.clker.com/clipart-14276.html, https://www.hsn.com/products/dell-2tb-nal-hard-drive/7889994, http://www.123rf.com/stock-photo/internet_icon.html , otechlead.com/networking/dell-brings-new-hpc-for-research-manufacturing-and-genomics-

Data analysis workflow

1) Raw data obtained
2) Copy it from source to RDS
3) Copy it from RDS to HPC scratch directory for analysis
4) Run compute job, sending output to scratch directory
5) Check the job has completed successfully and the output is correct
6) Copy final output and scripts to RDS for long-term, safe storage
7) If space permits and you require certain output files for downstream computational analysis, copy them to your project directory
8) Confirm that you have all raw data and output files backed up on RDS
9) Delete raw data, scripts and output files from scratch to RDS

Data transfer between your computer and HPC

Local computer

Image: freelargeimages.com

scp, sftp, rsync
FileZilla

Artemis HPC

Image: itnews.com

Data transfer between your computer and 'classic' RDS (CIFS)

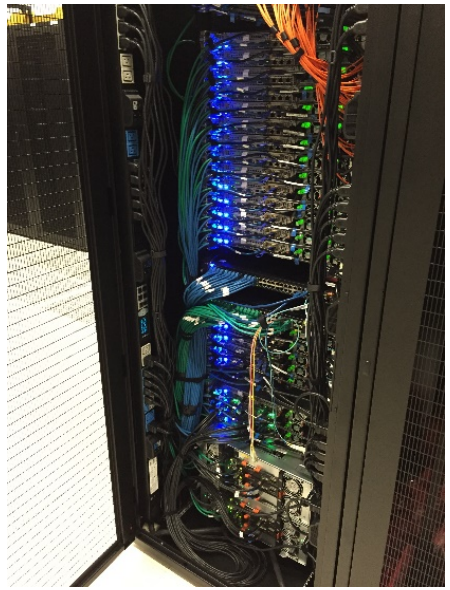
Research data store

Local computer

Image: freelargeimages.com

Mapped network drive

'classic' RDS



Artemis HPC

Image: itnews.com

Data transfer between HPC and 'classic' RDS (CIFS)

Research data store

'classic' RDS

Local computer

Image: freelargeimages.com

smbclient
FileZilla

Artemis HPC

Image: itnews.com

Data transfer between HPC and 'classic' RDS (CIFS)

Local computer

Image: freelargeimages.com

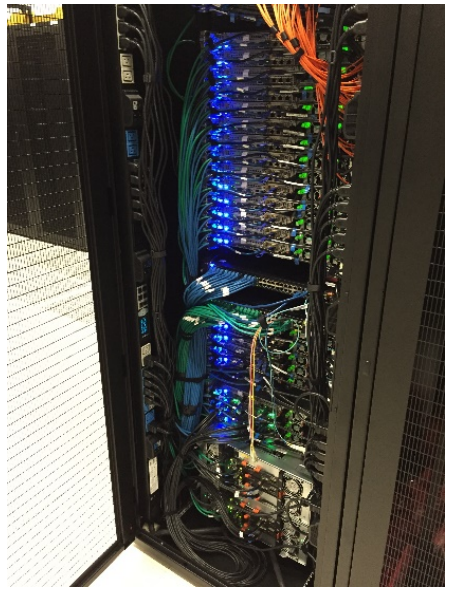
FileZilla

Research data store

'classic' RDS

FileZilla

smbclient
FileZilla



Artemis HPC

Image: itnews.com

FileZilla transfers data from Classic RDS to HPC via your local computer!

Data transfer between your computer and RCOS (NFS)

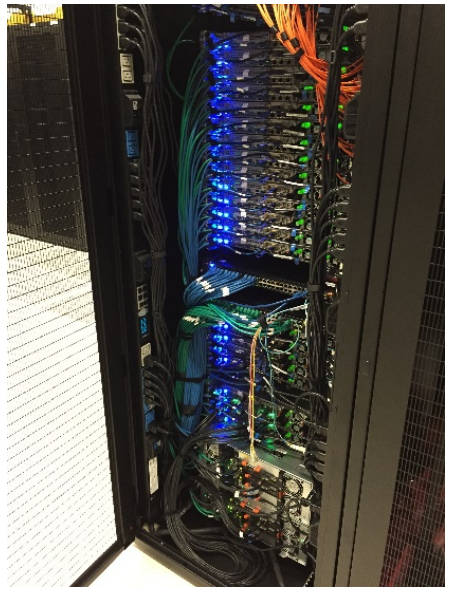
Research data store

Local computer

Image: freelargeimages.com

scp, sftp, rsync
FileZilla

RCOS



Artemis HPC

Image: itnews.com

Data transfer between HPC and RCOS (NFS)

Research data store

Local computer

Image: freelargeimages.com

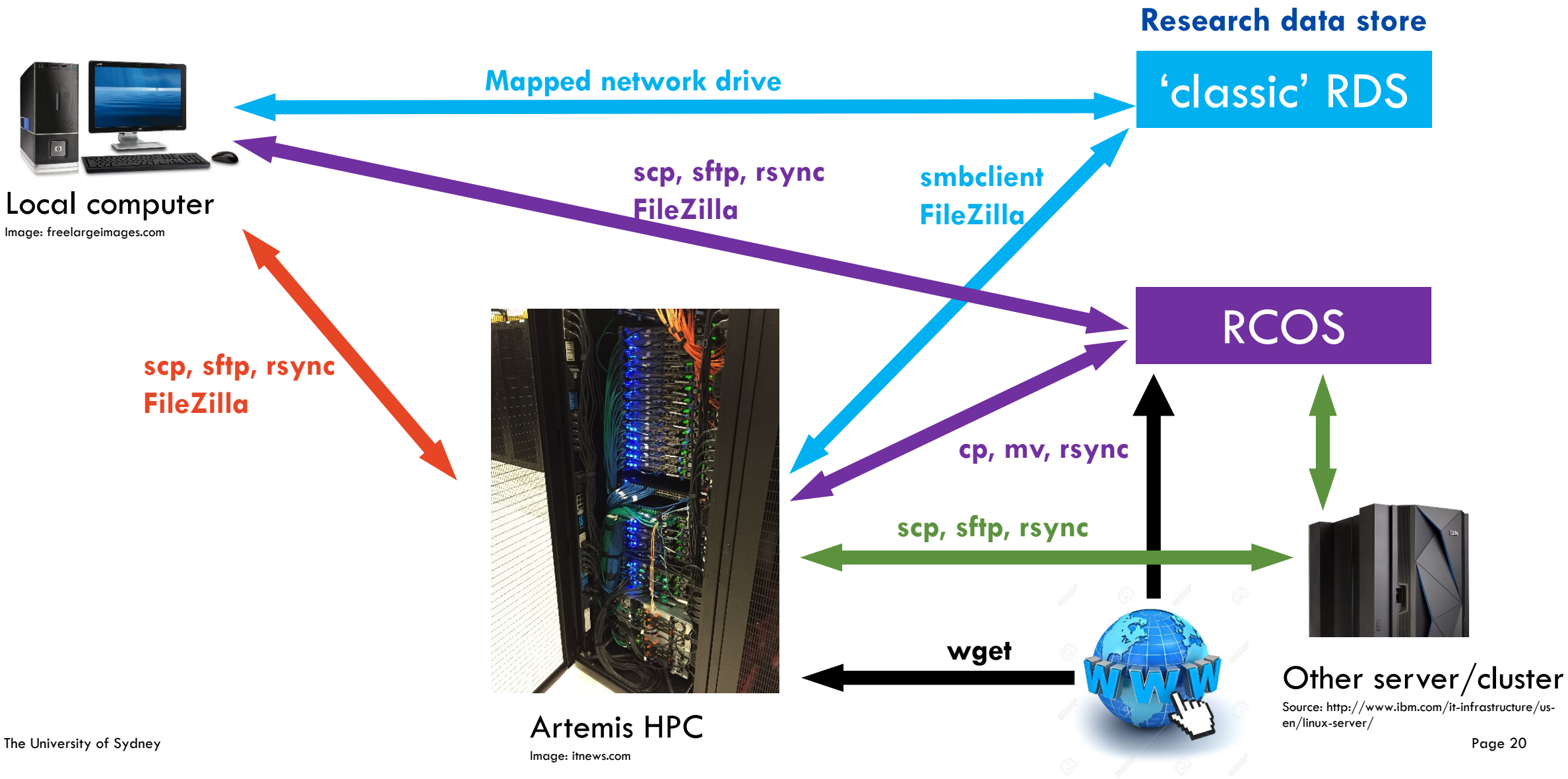
Artemis HPC

Image: itnews.com

RCOS

Artemis mount enables cp, mv, rsync

Data transfer methods summary



Data transfer: some practice

1. From local computer to HPC with scp, sftp, rsync or FileZilla
   - Any data
2. Web to RCOS with wget
   - Input data (Sample_data.tar.gz)
3. From RCOS to HPC with cp
   - Input Data (Sample_data.tar.gz)
4. From HPC to RCOS with rsync
   - Output files (Extracted data from Sample_data.tar.gz)
5. From local computer to RCOS with scp, sftp, rsync or FileZilla
   - Any data
6. From HPC to classic RDS with smbclient
   - See slides at the end of this document

1. Local to HPC using FileZilla

- Upload data to HPC with FileZilla

  - Mac or Cygwin/Ubuntu Windows users, if you are comfortable with scp or sftp you can use either of those tools instead

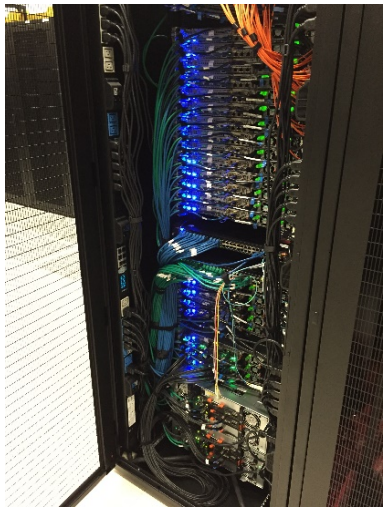
  - Use any data you like for this exercise. For example, you could download this file: ftp://ftp.ncbi.nlm.nih.gov/SampleData/Genomes/Analyzed_RNAseq/SRR1259333.bw

- FileZilla is a free file transfer client
  - Available from https://filezilla-project.org/
  - Windows, Linux or Mac

Local computer

Image: freelargeimages.com

scp, sftp, rsync FileZilla



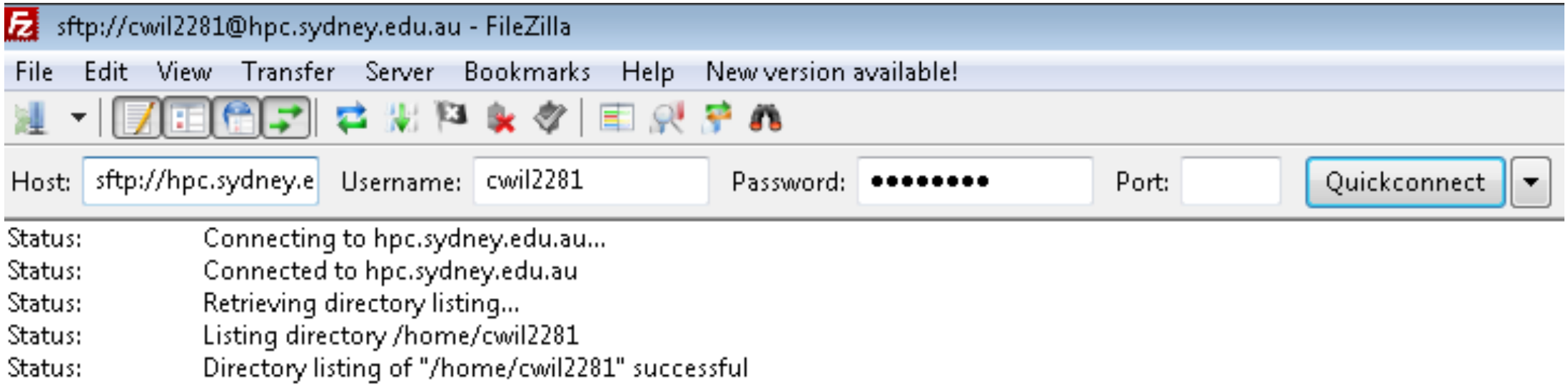
Artemis HPC

Image: itnews.com

- Open FileZilla by clicking the windows icon, then type FileZilla, then click the FileZilla icon

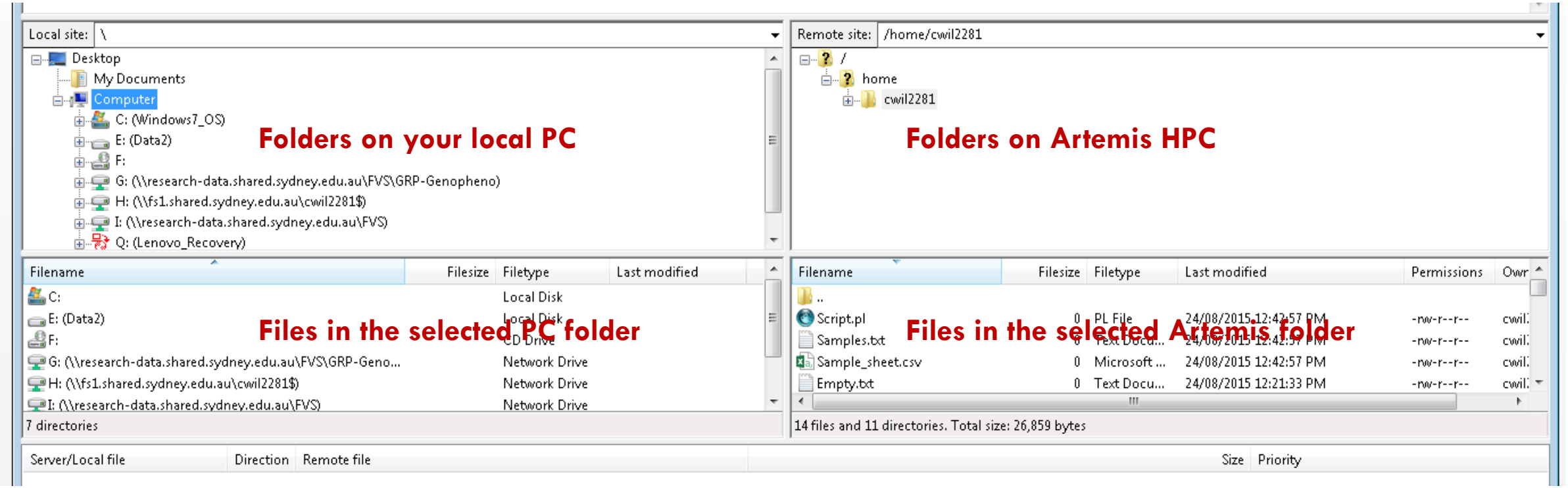
1. Local to HPC using FileZilla

- In the 'host' field, enter: `sftp://hpc.sydney.edu.au`
- In the 'Username' field, enter your unikey (or ict_hpctrain<N>)
- In the 'Password' field, enter your unikey password (or ict training password)
- You can leave port blank (default = 22)
- Click 'Quick connect'



1. Local to HPC using FileZilla

- The middle area of the FileZilla window is divided into four sections:



1. Local to HPC using FileZilla

- To copy from local to HPC, **first open the HPC directory that you want to copy the file to**

➢ Navigate to `/project/Training/`

➢ Make a directory for yourself (right click in the lower-right file list box and select "Create Directory")

You can navigate to `/project/Training` in one of two ways:

➢ Expand out the list of Artemis directories from the root by clicking

➢ Type the full directory path in the 'Remote site' text box

Single-click the question mark to expand list of directories

1. Local to HPC using FileZilla

- Next, locate the data to be uploaded using the left-hand panes of FileZilla

- Locate your data on your computer in the left hand side of FileZilla
- Upload the file to Artemis using your preferred method:
  - Double click
  - Right click → upload
  - Drag and drop
- You will see the copied file appear in your HPC directory (lower right pane)
- Optional: start an Artemis terminal session and confirm the file was transferred by running 'ls' on the directory containing the transferred file

1. Local to HPC on the command line

- From the Mac terminal, or from a Windows Cygwin or Ubuntu (W10) terminal, command line transfer is an alternative to FileZilla

Example scp command:

```
scp /path/to/file/SRR1259333.bw <unikey>@hpc.sydney.edu.au:/project/Training/<YourName>
```

Example rsync command:

```
rsync /path/to/file/SRR1259333.bw <unikey>@hpc.sydney.edu.au:/project/Training/<YourName>
```

Example sftp commands:

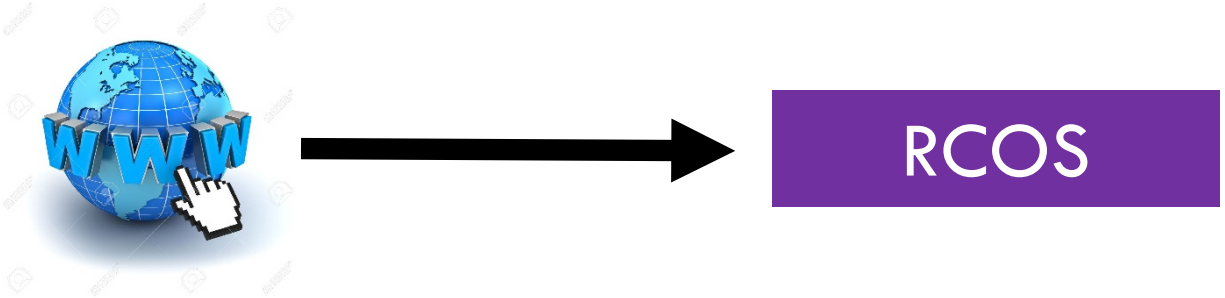
```
sftp <unikey>@hpc.sydney.edu.au
cd /project/Training/<YourName>
put /path/to/file/SRR1259333.bw
ctrl + d
```

2. From the web to RCOS

- Next we will use 'wget' to download data from the internet to RCOS

Important points regarding RCOS and the Artemis HPC:

- RCOS has been ***mounted*** on the Artemis HPC: this means that it is accessible from an Artemis session under the drive /rds
- The **same project name** as used in /project and /scratch space is used, with the addition of a prefix 'PRJ-'
- Eg, if your short project name is Training, your RCOS directory on HPC is `/rds/PRJ-Training`
- You can change into it using 'cd', move data to and from it with 'cp', 'mv', 'rsync' etc
- One thing you ***cannot*** do: read files on RCOS from a job running on the compute nodes
- This means that **input data for a compute job must first be copied to HPC** (/scratch or /project area)

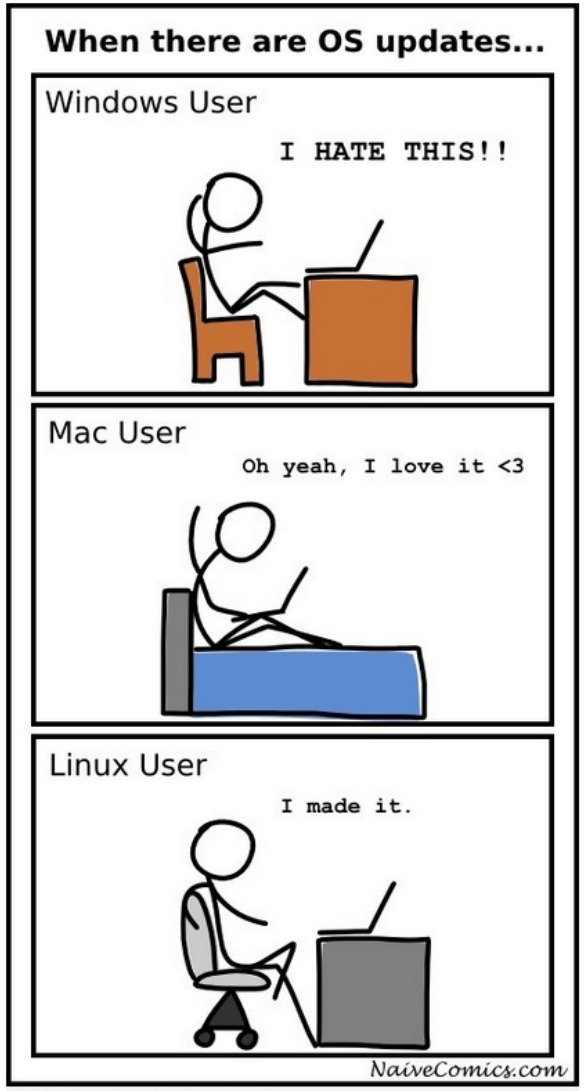


Accessing the Artemis HPC from your computer

- **Mac/Linux:** use the existing Terminal application

- **Windows:** install a terminal client – many available with various features, eg:

Client	License	Benefit
Xwin	Provided by USyd	Edit text in a graphical window
Cygwin	Free	Approximates a native Linux environment on Windows PC
Putty	Free	Easily change size and colour of text



Connecting to Artemis from: Mac

- Optional: Install XQuartz: https://www.xquartz.org/
- Go to Finder → Applications → Utilities
- Double-click Terminal
- Type the following into Terminal then hit enter:

```
ssh -Y ict_hpctrain<N>@hpc.sydney.edu.au
```

NOTE:
If you see text enclosed in arrows, this means replace the text with content relevant to you

NOTE:
The –Y option only works if you installed XQuartz

Connecting to Artemis from: Mac

- Optional: Install XQuartz: https://www.xquartz.org/
- Go to Finder → Applications → Utilities
- Double-click Terminal
- Type the following into Terminal then hit enter:

```
ssh -Y ict_hpctrain<N>@hpc.sydney.edu.au
```

NOTE:
If you see text enclosed in arrows, this means replace the text with content relevant to you

NOTE:
The –Y option only works if you installed XQuartz

- Type the password then press 'enter'

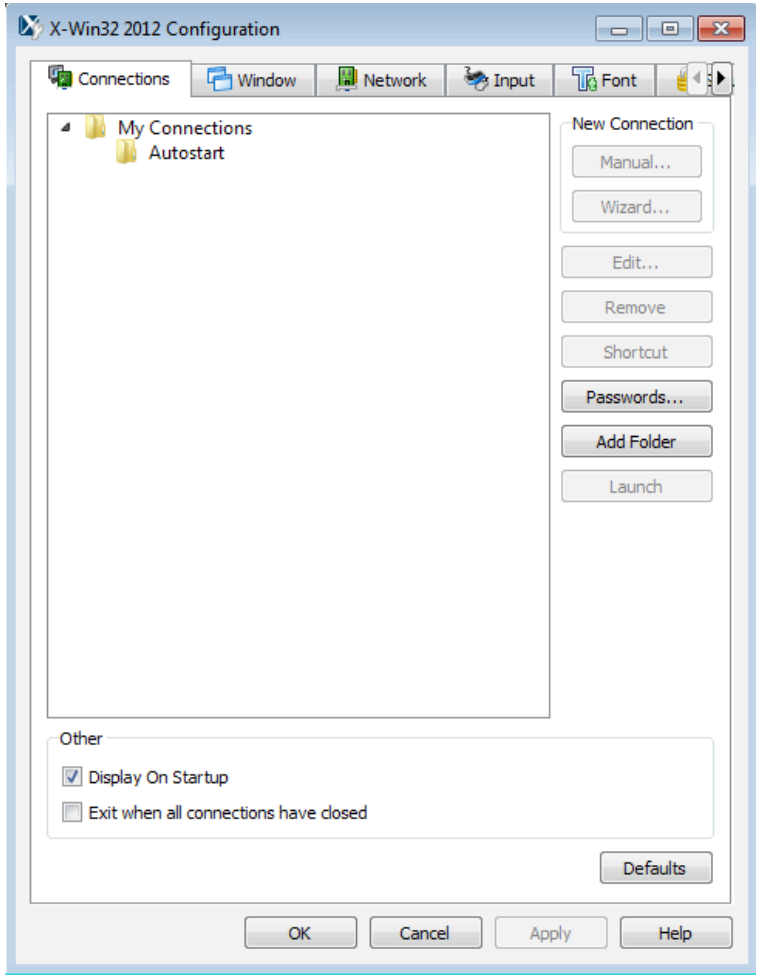
```
Stephens-MacBook-Pro:~ stephenlevo$ ssh -Y ict_hpctrain10@hpc.sydney.edu.au
The authenticity of host 'hpc.sydney.edu.au (10.250.96.204)' can't be established.
RSA key fingerprint is da:c3:1e:26:24:fc:f9:1c:74:99:99:fb:85:26:d7:50.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'hpc.sydney.edu.au,10.250.96.20 RSA) to the list of known hosts.
ict_hpctrain10@hpc.sydney.edu.au's password:
Last login: Fri Apr 24 10:39:55 2015 from vlan-2657-10-16-233-8.
-----------------------------------------------------------------
    Welcome to the University of Sydney HPC Linux Cluster - Artemis
-----------------------------------------------------------------

** NOTE: All jobs must request a project the user is a member of.
   Please specify -P <project> on the command line,
   or #PBS -P <project> in your PBS scripts.

** Please don't run jobs on the login node!!! Use the queuing syste

** Repeat! If you run jobs on the login nodes they will be subject to
   random termination if they cause load problems. Please don't do it.

[ict_hpctrain10@login2 ~]$
```

If connecting for the first time you will be asked if you trust the authenticity of host – type 'yes' then hit enter key

Logon success

Connecting to Artemis from: Windows with Xwin

- Download and install X-Win: http://softserv.usyd.edu.au/data/X-Win32/X-Win32-14.0/
- Click the windows icon in the lower-right hand corner of your screen
- Type X-Win into the search bar
- Click the X-Win32 or X-Config options
- A configuration window will appear



Where to download software for your own Windows machine?

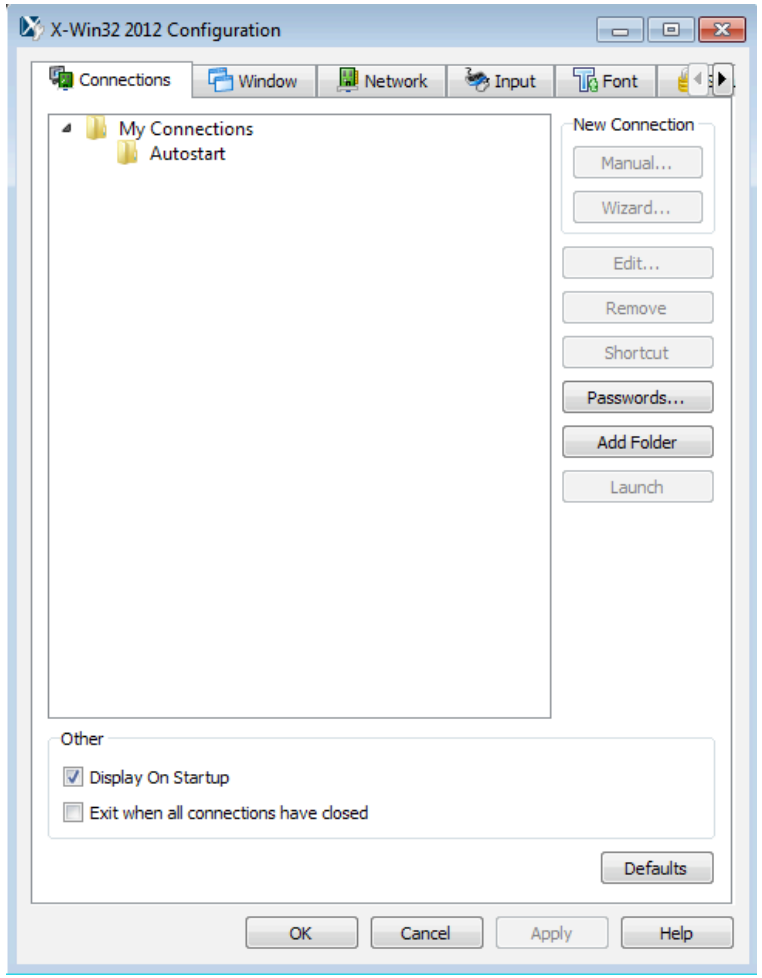
- Xwin: http://softserv.usyd.edu.au/data/X-Win32/X-Win32-14.0/
  -Requires unikey credentials
  -Follow the instructions to install license server details

- Putty: http://www.putty.org/
  -Download and run 'putty.exe'

- Cygwin: https://www.cygwin.com/
  -Download and run 'setup-x86_64.exe' (64 bit OS) or 'setup-x86.exe (32 bit OS)
  -During setup, select additional packages to install. Recommended: openssh, x-org server, xinit

Connecting to Artemis from: Windows with Xwin

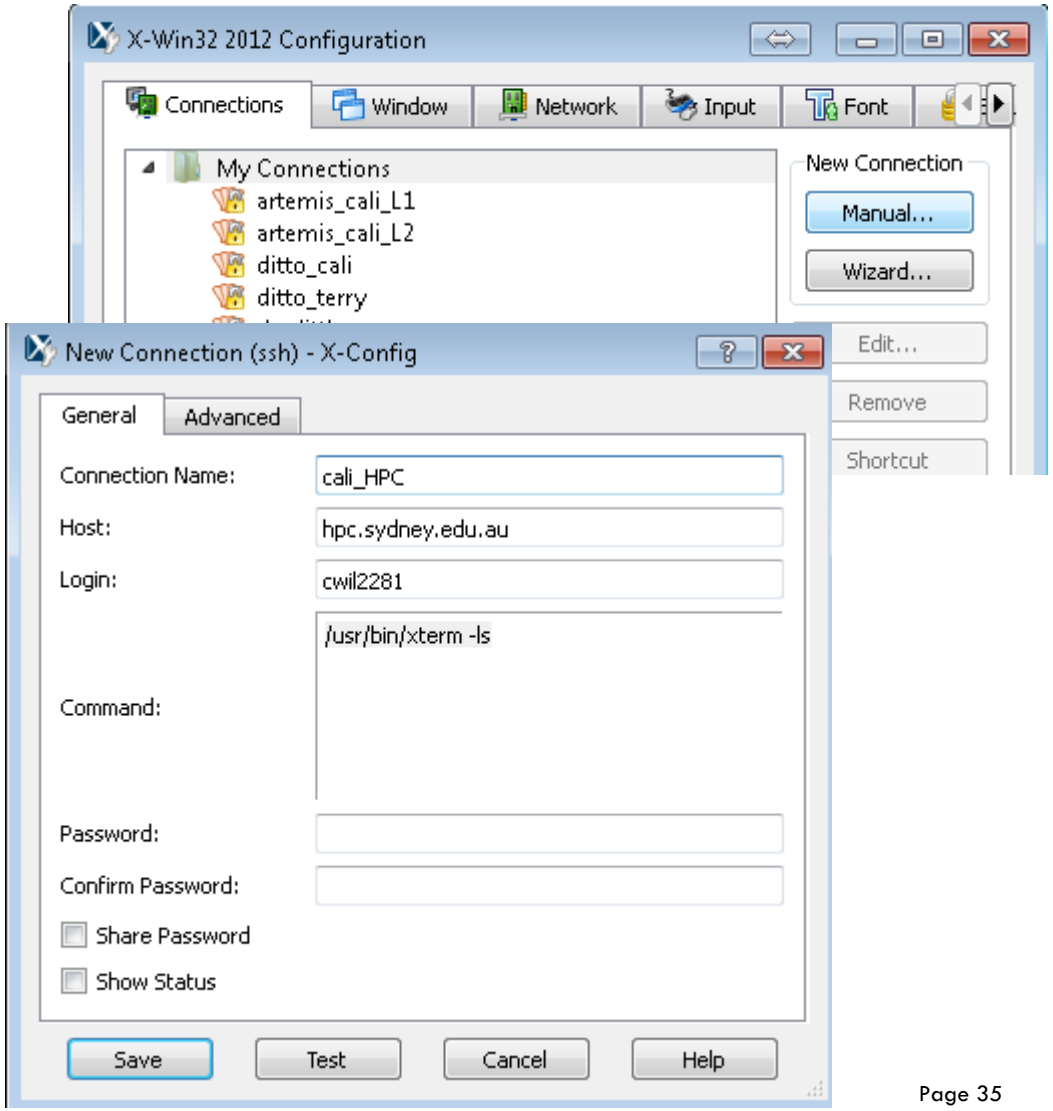
- Click the windows icon in the lower-right hand corner of your screen
- Type X-Win into the search bar
- Click the X-Win32 or X-Config options
- A configuration window will appear



Connecting to Artemis from: Windows with Xwin

- Along the right hand side under 'New Connection' select 'Manual'
- Select 'ssh' for connection method then click 'Next'
- In the 'Connection Name' field, enter any descriptive name eg **<name_HPC>**
- In the 'Host' field, enter: **hpc.sydney.edu.au**
- In the 'Login' field, enter your unikey
  - Training unikey: ict_hpctrain<N>
- In the 'Command' field, enter: **/usr/bin/xterm -ls**
- Click 'Save' then 'Launch'

Letter 'l' not numeral 1



Connecting to Artemis from: Windows with Putty

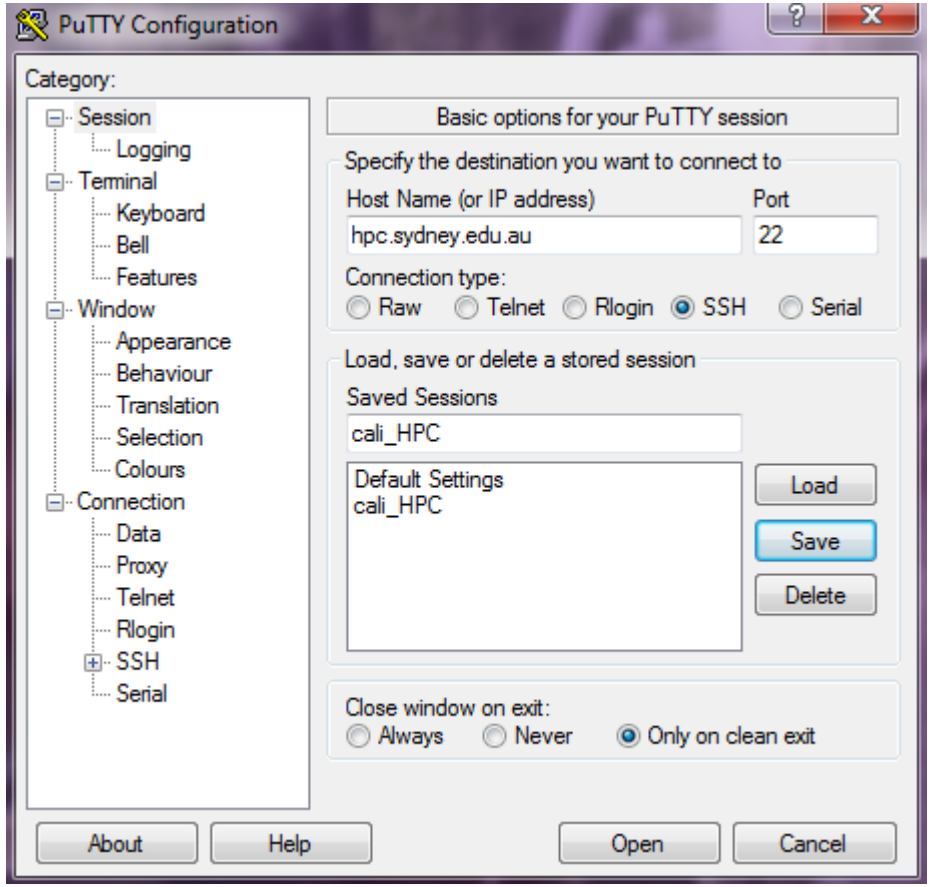
- Double click the Putty desktop icon

  A configuration window will appear

- In the 'Host Name' field, enter

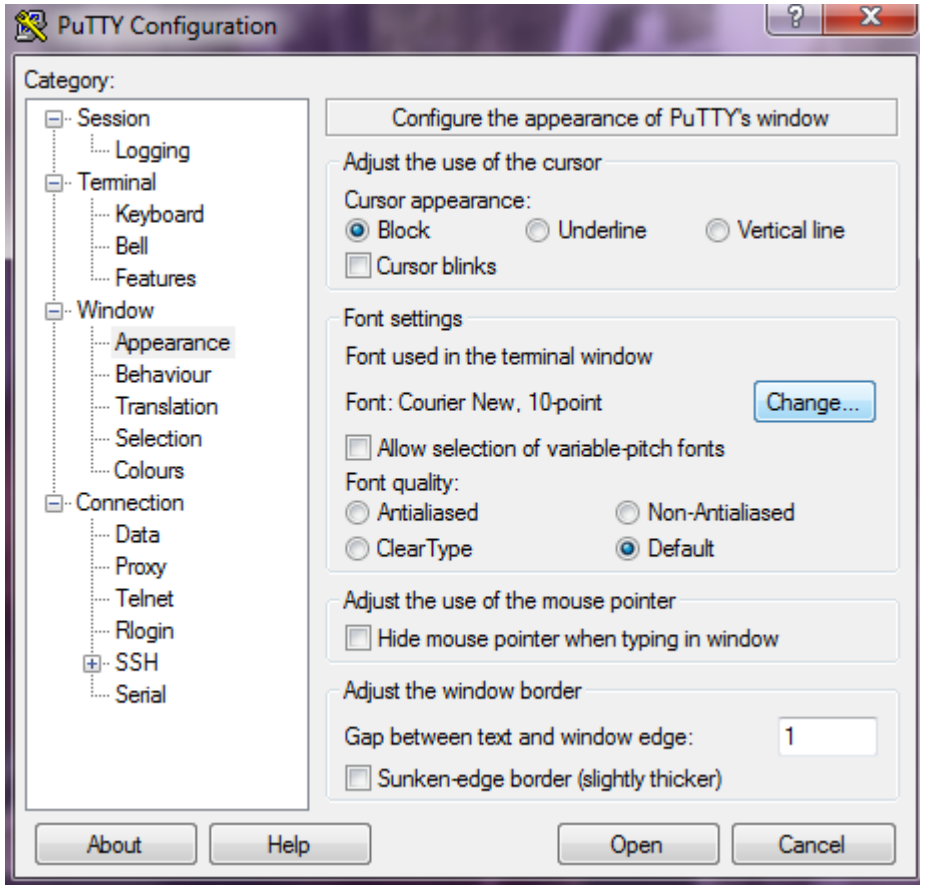
  `hpc.sydney.edu.au`

- Leave port as 22 (default)
- Type a name for your connection into 'Saved Sessions' and click 'Save'
- Click 'Open'
- At 'login as' enter your unikey (or ict_hpctrain<N>)
- Enter your unikey (or training) password



Adjusting the terminal font size/colour on Putty

1. Open your Putty configuration window again
2. Click on your saved session name
3. Click ‘Load’
4. Under ‘Window’, select ‘Appearance’
5. Under ‘Font Settings’, click ‘Change’
6. Change the font as desired
7. Click ‘OK’
8. Click ‘Session’
9. Click ‘Save’
10. Click ‘Open’ to launch your new-look terminal



2. From the web to RCOS

- In your HPC terminal, change into your RCOS directory*

  **Not all HPC accounts have RCOS. If you have a 'classic RDS' account, please copy the data to your project area instead. Instructions for using 'smbclient' (similar to ftp) to transfer data between 'classic RDS' and HPC are at the end of this presentation.*

  ```
  cd /rds/PRJ-Training
  ```

- Make a new directory and call it your name

  ```
  mkdir <YourName>_data
  ```

- Change into your new directory

  ```
  cd <YourName>_data
  ```

- In a web browser (any EXCEPT Internet Explorer), open http://hp580.angis.org.au/~cali/
- Right click the file 'Sample_data.tar.gz' and select 'copy link address' or 'copy link location'
- In your terminal, type 'wget' then a space, paste the copied link** then press 'enter'

  ***Control + V does not mean copy to Linux! To copy into a Linux terminal on Windows: laptop pad is usually right click, mouse is usually centre click (wheel). You might need to press Shift+Insert to paste. To copy into Mac terminal, use Command + V*

  ```
  wget http://hp580.angis.org.au/~cali/Sample_data.tar.gz
  ```

- Successful transfer should show a progress bar followed by a message: `"Sample_data.tar.gz" saved`

3. RCOS to HPC using cp

- We will transfer Sample_data.tar.gz from RCOS to Artemis and perform our "data processing" (which is simply extracting the archive today!)

➢ On Artemis, change directory to your Artemis /project directory:

```
mkdir /project/Training/<YourName>_data
cd /project/Training/<YourName>_data
```

➢ Copy Sample_data.tar.gz from /rds to your current working directory:

```
cp /rds/PRJ-Training/<YourName>_data/Sample_data.tar.gz .
```

➢ Extract the files from Sample_data.tar.gz and move them to your directory:

```
tar -zxf Sample_data.tar.gz
mv Sample_data/* .
rmdir Sample_data
```

Don't forget to include the full-stop!

Artemis HPC

Image: itnews.com

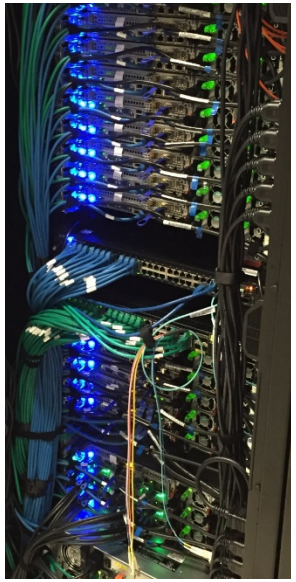
RCOS ⟷ Artemis HPC

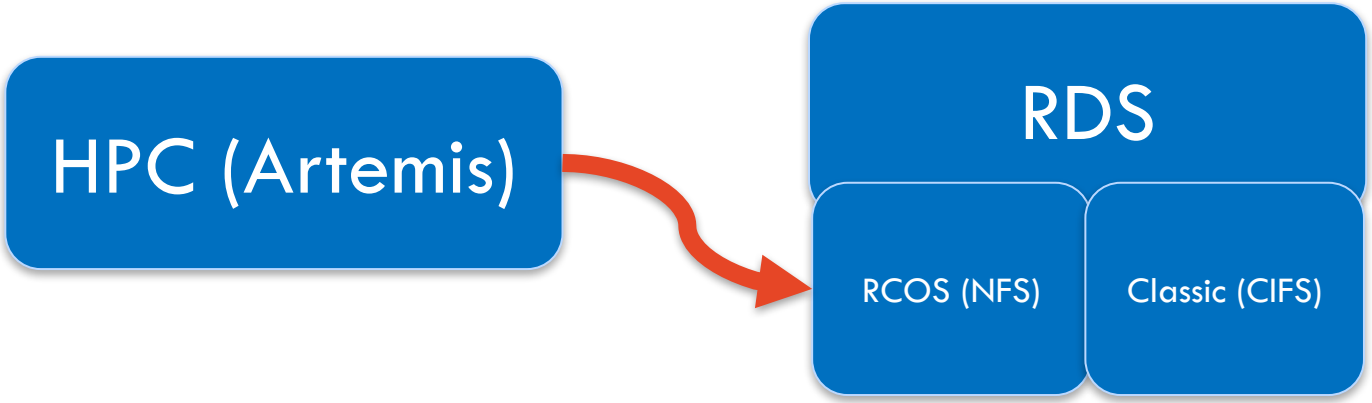
cp, mv, rsync

Moving data from HPC to RDS

Now that Artemis has processed our data, we must transfer our output elsewhere.

- Remember: Artemis is for processing data, not storing data
- Artemis storage is not backed up
- Move your important data to the Research Data Store
- Clean up your Artemis directories regularly



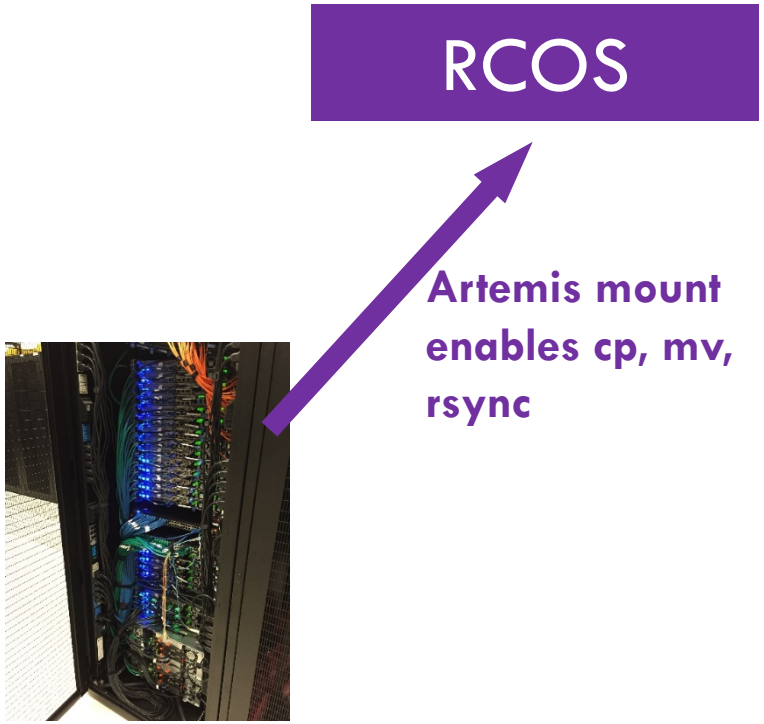


Data transfer: some practice

1. From local computer to HPC with scp, sftp, rsync or FileZilla
   - Any data ✓
2. Web to RCOS with wget ✓
   - Input data (Sample_data.tar.gz)
3. From RCOS to HPC with cp ✓
   - Input Data (Sample_data.tar.gz)
4. From HPC to RCOS with rsync
   - Output files (Extracted data from Sample_data.tar.gz)
5. From local computer to RCOS with scp, sftp, rsync or FileZilla
   - Any data
6. From HPC to classic RDS with smbclient
   - See slides at the end of this document

4. HPC to RCOS: rsync

- This is the reverse of step 3
- The pathname is: `/rds/PRJ-Training`
- You can back up data there with cp, mv or rsync commands
- rsync can preserve permissions and file creation/modification times and only copies files that are new or have changed



Artemis HPC

Image: itnews.com

➢ Backup your final output files from /project to RCOS using rsync:

```
rsync -rt /project/Training/<YourName>_data/* /rds/PRJ-Training/<YourName>_data
```

- r = resursive (copies directories too), t = preserves timestamps
- Run 'ls' on RCOS to confirm transfer

5a. local to RCOS: FileZilla

➢ Upload any local data to RCOS

FileZilla:

Local computer

Image: freelargeimages.com

FileZilla

RCOS

1. Open FileZilla and enter into hostname field:
   `sftp://rcos-int.sydney.edu.au`
   `sftp://rcos.sydney.edu.au`
2. Fill in your UniKey and password and click 'quick connect'
3. On the right hand panes (RCOS) find your project
   - Either type the path into 'Remote site' field, or use the mouse to traverse the folder icons
4. On the left hand side (local) find the unzipped files you copied from the HPC
5. Click one file, hold down control and then select the other 2
   - Copy to RCOS by either drag-and-drop or right click → upload (double-click won't work for multiple files)
6. You will see a status of the transfer down the bottom, and the files will appear in the folder on the right when the transfer is complete

5b. local to RCOS: sftp

- Upload any local data to RCOS from **Mac, Windows (with Cygwin) or Windows 10 (with the beta Ubuntu terminal)**

Local computer

Image: freelargeimages.com

sftp

RCOS

sftp:

1. Open a terminal on your **local** computer
2. Change into the directory containing the 3 unzipped files you downloaded from HPC
3. Connect to RCOS over sftp and upload the 3 files:

```
sftp <unikey>@rcos-int.sydney.edu.au
cd /rds/PRJ-Training
mput *
```

Internal hostname:
`rcos-int.sydney.edu.au`
External hostname:
`rcos.sydney.edu.au`

5c. local to RCOS: rsync

- Upload any local data to RCOS from **Mac, Windows Cygwin or Windows 10 Ubuntu**

Local computer

Image: freelargeimages.com

scp

RCOS

rsync:

1. Open a terminal on your **local** computer
2. Change into the directory containing the files you downloaded from HPC
3. Use rsync (or scp) to upload the 3 scripts to RCOS:

```
rsync * <unikey>@rcos-int.sydney.edu.au:/rds/PRJ-Training
```

Tip: to recursively copy a directory, add the 'r' flag to rsync, eg

```
rsync -r Directory <unikey>@rcos-int.sydney.edu.au:/rds/PRJ-Training
```

Data transfer: some practice

1. From local computer to HPC with scp, sftp or FileZilla ✓
   - Any data
2. Web to RCOS with wget
   - Input data (Sample_data.tar.gz) ✓
3. From RCOS to HPC with cp
   - Input Data (Sample_data.tar.gz) ✓
4. From HPC to RCOS with rsync
   - Output files (Extracted data from Sample_data.tar.gz) ✓
5. From local computer to RCOS with scp, rsync, sftp or FileZilla
   - Output files (Extracted data from Sample_data.tar.gz) ✓
6. From HPC to classic RDS with smbclient
   - See slides at end of this document

6. HPC to classic RDS: smbclient

- ‘smbclient’ is a command-line tool similar to sftp to transfer between a Linux system (Artemis) and a Windows system (RDS)
- FileZilla could also be used, but this is much slower
  - FileZilla moves data between HPC and RDS via your PC/laptop creating a network bottleneck
- See slides at the end for instructions

Research data store

‘classic’ RDS

smbclient
FileZilla

Artemis HPC

Image: itnews.com

Data transfer: some practice

1. From local computer to HPC with scp, sftp or FileZilla
   - Any data ✓
2. Web to RCOS with wget
   - Input data (Sample_data.tar.gz) ✓
3. From RCOS to HPC with cp
   - Input Data (Sample_data.tar.gz) ✓
4. From HPC to RCOS with rsync
   - Output files (Extracted data from Sample_data.tar.gz) ✓
5. From local computer to RCOS with scp, sftp rsync, or FileZilla
   - Output files (Extracted data from Sample_data.tar.gz) ✓
6. From HPC to classic RDS with smbclient
   - See slides at the end of this document ✓

Artemis Data Transfer Queue

Schedule data transfers using a data transfer PBS script

Create data transfer jobs using dt-script

Data transfer queue on Artemis

- There is an Artemis queue dedicated to running data transfer jobs
  - dtq
- This queue is for I/O intensive jobs (such as transferring data to and from Artemis)
- Commands such as `cp`, `tar`, `scp`, `rsync` can all be used in this queue
- This is the only queue that can read to or write from RCOS (/rds)
- This queue also has internet access, so you can run (for example) wget commands in a dtq script

Data transfer queue on Artemis

- An example data transfer script is below. Don't worry about making this now. It's just an example ☺

```
#!/bin/bash
#PBS -P Training
#PBS -l select=1:ncpus=1:mem=2gb
#PBS -l walltime=00:05:00
#PBS -q dtq

cd /project/Training/<YourName>_data
rsync /rds/PRJ-Training/<YourName>_data/* .
```

Using the data transfer queue on Artemis

- The end result: Data was copied from RCOS to Artemis. No big deal right?
  - We already know how to copy data from RCOS to Artemis
- BUT
- You can schedule a job to start straight after copying data to Artemis
- You don't have to worry about your computer going to "sleep"
- This is especially useful if you need to transfer large amounts of data to Artemis before running a compute job
  - Your job will automatically begin at 5am when the data transfer completes ☺

Automation using the data transfer queue on Artemis

- We can do a data transfer and compute job using two, linked PBS scripts:
  1. A script that downloads data from the internet, another cluster or server to Artemis using download.pbs (which is a dtq queue script)
  2. A script that runs a job using the downloaded data as input: index.pbs (which is a small-express queue script)

➢ Copy download.pbs and index.pbs to your own dtq_testing folder:

```
cd /project/Training/<YourName>_data
mkdir dtq_testing
cd dtq_testing
rsync /project/Training/data_transfer_scripts/* .
```

Automation using the data transfer queue on Artemis

- Script 1: download.pbs

download.pbs

```
#!/bin/bash
#PBS -P Training
#PBS -l select=1:ncpus=1:mem=2gb
#PBS -l walltime=00:05:00
#PBS -q dtq

cd $PBS_O_WORKDIR
wget http://hp580.angis.org.au/~cali/Sample_data.tar.gz
tar -zxf Sample_data.tar.gz
cp Sample_data/* .
rm Sample_data/*
rmdir Sample_data
```

Changes directory to the directory where you submit this script

This script grabs input data from the internet, extracts it, and reorganises it so it is ready for the analysis in the next script!

Automation using the data transfer queue on Artemis

- Script 2: index.pbs

index.pbs

```
#! /bin/bash
# Create indexes for reference sequence
#PBS -P Training
#PBS -N Index
#PBS -l select=1:ncpus=1:mem=4GB
#PBS -l walltime=00:10:00
#PBS -q small-express

# Load modules
module load bwa
module load samtools

io=$PBS_O_WORKDIR

# BWA index:
bwa index -a bwtsw ${io}/canfam3_chr5.fasta

# SAMtools index:
samtools faidx ${io}/canfam3_chr5.fasta
```

This script performs analysis on the data we set up in Script 1

Automation using the data transfer queue on Artemis

- Submit the two scripts to the PBS Scheduler

```
qsub download.pbs
```

1265834.pbsserver

```
qsub -W depend=afterok:1265834 index.pbs
```

Replace this number with the the job id assigned to your job

- -W depend=afterok is a qsub option that says "only run index.pbs if download.pbs terminated successfully"
- index.pbs won't run unless the files were successfully downloaded and extracted

- Check job status

```
qstat -xu <unikey>
```

Automation using the data transfer queue on Artemis

```
[abcd1234@login3 data_transfer_scripts]$ qstat -xu abcd1234

pbsserver:
                                                                  Req'd  Req'd   Elap
Job ID          Username Queue    Jobname    SessID NDS TSK Memory Time  S Time
--------------- -------- -------- ---------- ------ --- --- ------ ----- - -----
1426229.pbsserv abcd1234 dtq      download.p 110451   1   1    2gb 00:05 R 00:00
1426230.pbsserv abcd1234 small-ex Index         --   1   1    4gb 00:10 H   --
```

- "H" state means the job is "held". It won't start until it is manually released, or until it's job dependencies are met.
- We used -W depend=afterok:1426229, so index.pbs won't start until download.pbs successfully completed.

Data transfer queue helper script – dt-script

- There is a script available to help make using dtq simpler
  - dt-script
- We could have done the previous jobs with a single dt-script command!
- dt-script creates the dtq job for you. Specify where to copy data from and to, your project name, and (optionally) a PBS job to run after the data transfer completes.
- Syntax:

```
dt-script -P <Project> -f <from> -t <to> -j job.pbs
```

- `-P`, `-f` and `-t` are compulsory; `-j` is optional.
- There are other options available to dt-script. To see these, run the command:

```
dt-script --help
```

- Advanced users can make a copy of dt-script and tailor it to their needs

Data transfer queue helper script – dt-script

- Submit a copy/compute job using dt-script:

```
mkdir dt-script-testing
cd dt-script-testing
cp ../index.pbs .
dt-script -P Training -f "/rds/PRJ-Training/dt-data/*" -t `pwd`
-j index.pbs
```

Space, then full-stop

The above command uses two non-interchangeable types of quotes

Double quotes are required when using wildcards (eg *)

These quotes execute the command inside the quotes. In this case, pwd prints your current working directory

Data transfer queue helper script – dt-script

- Submit the dt-script job:

```
dt-script -P Training -f "/rds/PRJ-Training/dt-data/*" -t `pwd`
-j index.pbs
```

- This dt-script command submits two jobs: a data transfer job and a compute job
- The data transfer script was created automatically

```
qstat -u <unikey>
```

```
                                                          Req'd  Req'd   Elap
Job ID          Username Queue    Jobname    SessID NDS TSK Memory Time  S Time
--------------- -------- -------- ---------- ------ --- --- ------ ----- - -----
1277947.pbsserv skol2049 dtq      dt-script     --    1   1    4gb 24:00 Q   --
1277948.pbsserv skol2049 small-ex Index         --    1   1    4gb 00:10 H   --
```

The "H" means the job is in the "held" state. It will start once the dtq job successfully finishes.

Copy important data back to RCOS

- **Verify your output data is OK** before transferring back to RCOS or another computer
  - A machine cannot do this for you (in most cases). The onus is on **you** the researcher to ensure correct output before backing up your data and cleaning it off /project or /scratch

➢ To copy data back to RCOS, you can use dt-script:

```
mkdir /rds/PRJ-Training/<YourName>_data/dtq_output
dt-script -P Training \
-f `pwd` \
-t "/rds/PRJ-Training/<YourName>_data/dt-output/*"
```

No backslashes required if you type this all on one line

Some quick revision!

- Take 15 minutes to work through the revision activity
  - Write down your answers as you work through the tasks
  - We will then go through the answers together

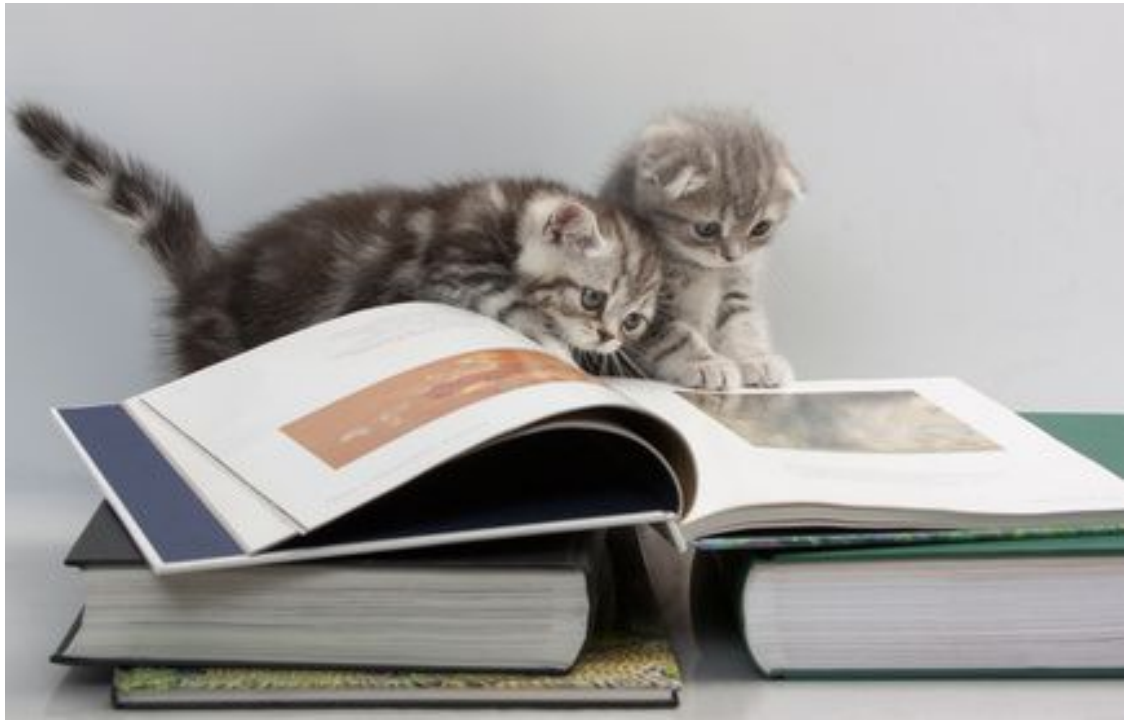


Image: animalbehaviorcollege.com

Revision activity

1. What types of Research Data Storage are available?

_____

2. What commands/programs can you use to transfer data from Artemis to RCOS?

_____

3. What is the name of the Artemis data transfer queue?

_____

4. What is the fair share weighting of dtq?

_____

5. What three flags must you always include in a data transfer script command?

_____

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

_____

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

_____

8. What is the maximum number of CPUs you can request in a dtq job?

_____

9. What is the maximum walltime for dtq?

_____

10. Where should all important input and output data be stored long-term?

_____

Revision activity

1. What types of Research Data Storage are available?

   `Classic RDS and RCOS`

2. What commands/programs can you use to transfer data from Artemis to RCOS?

_____

3. What is the name of the Artemis data transfer queue?

_____

4. What is the fair share weighting of dtq?

_____

5. What three flags must you always include in a data transfer script command?

_____

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

_____

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

_____

8. What is the maximum number of CPUs you can request in a dtq job?

_____

9. What is the maximum walltime for dtq?

_____

10. Where should all important input and output data be stored long-term?

_____

Revision activity

1. What types of Research Data Storage are available?

   `Classic RDS and RCOS`

2. What commands/programs can you use to transfer data from Artemis to RCOS?

   `cp, mv, rsync, dt-script`

3. What is the name of the Artemis data transfer queue?

_____

4. What is the fair share weighting of dtq?

_____

5. What three flags must you always include in a data transfer script command?

_____

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

_____

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

_____

8. What is the maximum number of CPUs you can request in a dtq job?

_____

9. What is the maximum walltime for dtq?

_____

10. Where should all important input and output data be stored long-term?

_____

Revision activity

1. What types of Research Data Storage are available?

   `Classic RDS and RCOS`

2. What commands/programs can you use to transfer data from Artemis to RCOS?

   `cp, mv, rsync, dt-script`

3. What is the name of the Artemis data transfer queue?

   `dtq`

4. What is the fair share weighting of dtq?

_____

5. What three flags must you always include in a data transfer script command?

_____

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

_____

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

_____

8. What is the maximum number of CPUs you can request in a dtq job?

_____

9. What is the maximum walltime for dtq?

_____

10. Where should all important input and output data be stored long-term?

_____

Revision activity

1. What types of Research Data Storage are available?

   Classic RDS and RCOS

2. What commands/programs can you use to transfer data from Artemis to RCOS?

   cp, mv, rsync, dt-script

3. What is the name of the Artemis data transfer queue?

   dtq

4. What is the fair share weighting of dtq?

   Zero!

5. What three flags must you always include in a data transfer script command?

   _____

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

   _____

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

   _____

8. What is the maximum number of CPUs you can request in a dtq job?

   _____

9. What is the maximum walltime for dtq?

   _____

10. Where should all important input and output data be stored long-term?

   _____

Revision activity

1. What types of Research Data Storage are available?

   Classic RDS and RCOS

2. What commands/programs can you use to transfer data from Artemis to RCOS?

   cp, mv, rsync, dt-script

3. What is the name of the Artemis data transfer queue?

   dtq

4. What is the fair share weighting of dtq?

   Zero!

5. What three flags must you always include in a data transfer script command?

   -P -f -t

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training"
```

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

   _____

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

   _____

8. What is the maximum number of CPUs you can request in a dtq job?

   _____

9. What is the maximum walltime for dtq?

   _____

10. Where should all important input and output data be stored long-term?

   _____

Revision activity

1. What types of Research Data Storage are available?

   Classic RDS and RCOS

2. What commands/programs can you use to transfer data from Artemis to RCOS?

   cp, mv, rsync, dt-script

3. What is the name of the Artemis data transfer queue?

   dtq

4. What is the fair share weighting of dtq?

   Zero!

5. What three flags must you always include in a data transfer script command?

   -P -f -t

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training"
```

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

   -W depend=afterok:<JobId>

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

   _____

8. What is the maximum number of CPUs you can request in a dtq job?

   _____

9. What is the maximum walltime for dtq?

   _____

10. Where should all important input and output data be stored long-term?

   _____

Revision activity

1. What types of Research Data Storage are available?

   Classic RDS and RCOS

2. What commands/programs can you use to transfer data from Artemis to RCOS?

   cp, mv, rsync, dt-script

3. What is the name of the Artemis data transfer queue?

   dtq

4. What is the fair share weighting of dtq?

   Zero!

5. What three flags must you always include in a data transfer script command?

   -P -f -t

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training"
```

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

   -W depend=afterok:<JobId>

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

   -j

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training" \
-j job.pbs
```

8. What is the maximum number of CPUs you can request in a dtq job?

   _____

9. What is the maximum walltime for dtq?

   _____

10. Where should all important input and output data be stored long-term?

   _____

Revision activity

1. What types of Research Data Storage are available?

   Classic RDS and RCOS

2. What commands/programs can you use to transfer data from Artemis to RCOS?

   cp, mv, rsync, dt-script

3. What is the name of the Artemis data transfer queue?

   dtq

4. What is the fair share weighting of dtq?

   Zero!

5. What three flags must you always include in a data transfer script command?

   -P -f -t

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training"
```

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

   -W depend=afterok:<JobId>

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

   -j

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training" \
-j job.pbs
```

8. What is the maximum number of CPUs you can request in a dtq job?

   Two

9. What is the maximum walltime for dtq?

   _____

10. Where should all important input and output data be stored long-term?

   _____

Revision activity

1. What types of Research Data Storage are available?

   Classic RDS and RCOS

2. What commands/programs can you use to transfer data from Artemis to RCOS?

   cp, mv, rsync, dt-script

3. What is the name of the Artemis data transfer queue?

   dtq

4. What is the fair share weighting of dtq?

   Zero!

5. What three flags must you always include in a data transfer script command?

   -P -f -t

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training"
```

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

   -W depend=afterok:<JobId>

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

   -j

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training" \
-j job.pbs
```

8. What is the maximum number of CPUs you can request in a dtq job?

   Two

9. What is the maximum walltime for dtq?

   10 days

10. Where should all important input and output data be stored long-term?

Revision activity

1. What types of Research Data Storage are available?

   Classic RDS and RCOS

2. What commands/programs can you use to transfer data from Artemis to RCOS?

   cp, mv, rsync, dt-script

3. What is the name of the Artemis data transfer queue?

   dtq

4. What is the fair share weighting of dtq?

   Zero!

5. What three flags must you always include in a data transfer script command?

   -P -f -t

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training"
```

6. What PBS directive tells the scheduler only to run a job after another job has successfully terminated?

   -W depend=afterok:<JobId>

7. What data transfer script flag lets you specify a PBS script to run after the data transfer job finishes?

   -j

```
e.g. dt-script -P Training \
-f "/scratch/Training/mydata" \
-t "/rds/PRJ-Training" \
-j job.pbs
```

8. What is the maximum number of CPUs you can request in a dtq job?

   Two

9. What is the maximum walltime for dtq?

   10 days

10. Where should all important input and output data be stored long-term?

    Research Data Store

HPC commands summary

PBS directives

`-P` = project

`-N` = job name

`-l select=1:ncpus=1:mem=1GB` = request nodes, processors and total RAM

`-l walltime=00:10:00` = walltime in hours:mins:secs

`-M` = email, use with –m

`-m` = mail when job: a = aborts, b = begins, e = ends

`-j oe` = create one job log containing both STDOUT and STDERR

`-q` = specify queue (can omit to submit to defaultQ)

Disk quota

`pquota` = show space allowed and used in all your accessible directories

PBS commands

`qsub job.pbs` = submit a job

`qstat -x jobID` = check status of specific job

`qstat -u unikey` = check status of all your submitted jobs

`qstat -T | grep unikey` = check estimated start time of all your submitted jobs

`qdel jobID` = delete job

`qsub -W depend=afterok:1234567 job.pbs` = submit a job and start it only if job 1234567 terminated successfully

`dt-script --help` = shows you how to use dt-script

smbclient and sftp commands

`put File`= transfer 1 file from HPC to RDS

`get File` = transfer 1 file from RDS to HPC

`mput *` = transfer multiple files from HPC to RDS using star wild card to match filenames

`mget *` = transfer multiple files from RDS to HPC using star wild card to match filenames

`lcd` = local cd

`!pwd` = local pwd

`!ls` = local ls

smbclient only

`prompt` = toggle on/off prompt

`recurse` = toggle on/off recurse

`mput Directory*` = run after the above 2 commands, will recursively copy Directory and all its files and subfolders (retaining directory structure) from HPC to RDS

`mget Directory*` = as above, from RDS to HPC

scp

Copy data to host: `scp data user@host:destinationDir`

Get data from host: `scp user@host:data destinationDir`

Thank you!

- That completes your introduction to using Artemis HPC DTQ
- Thanks to all for your attendance and participation
- Thanks also to Cali Willet for designing the courses!

Where to get help:

- Submit a HPC service request:
  - Go to the ICT Self-Service Portal at https://sydney.service-now.com/selfservice/ict_services.do
- Please feel free to contact us with any questions:

  sih.info@sydney.edu.au

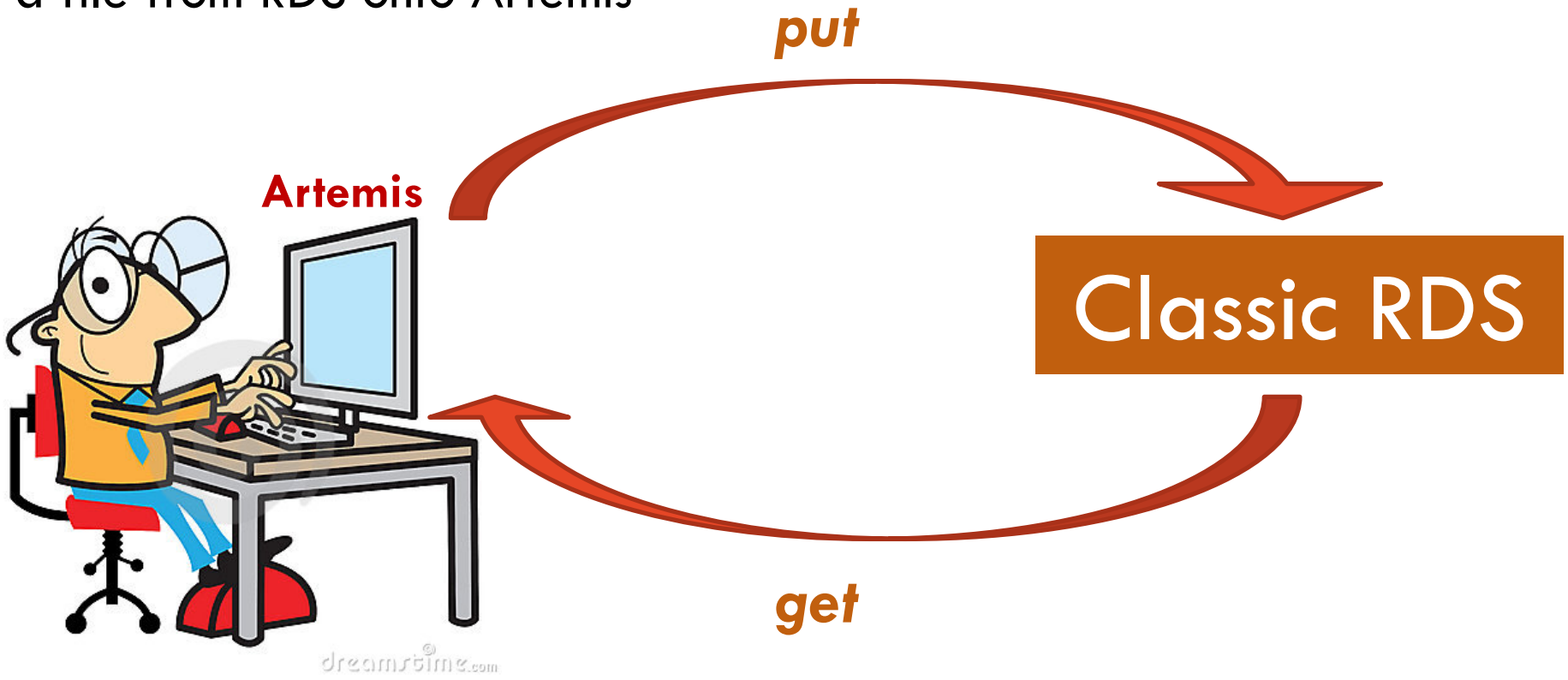
Image: iconshut.com

- We appreciate your feedback, please complete the survey:

  https://goo.gl/aMJSA7

7. Data transfer: HPC to Classic RDS - smbclient

- ‘`put`’ a file from Artemis onto RDS
- ‘`get`’ a file from RDS onto Artemis



Data transfer with smbclient: location matters!

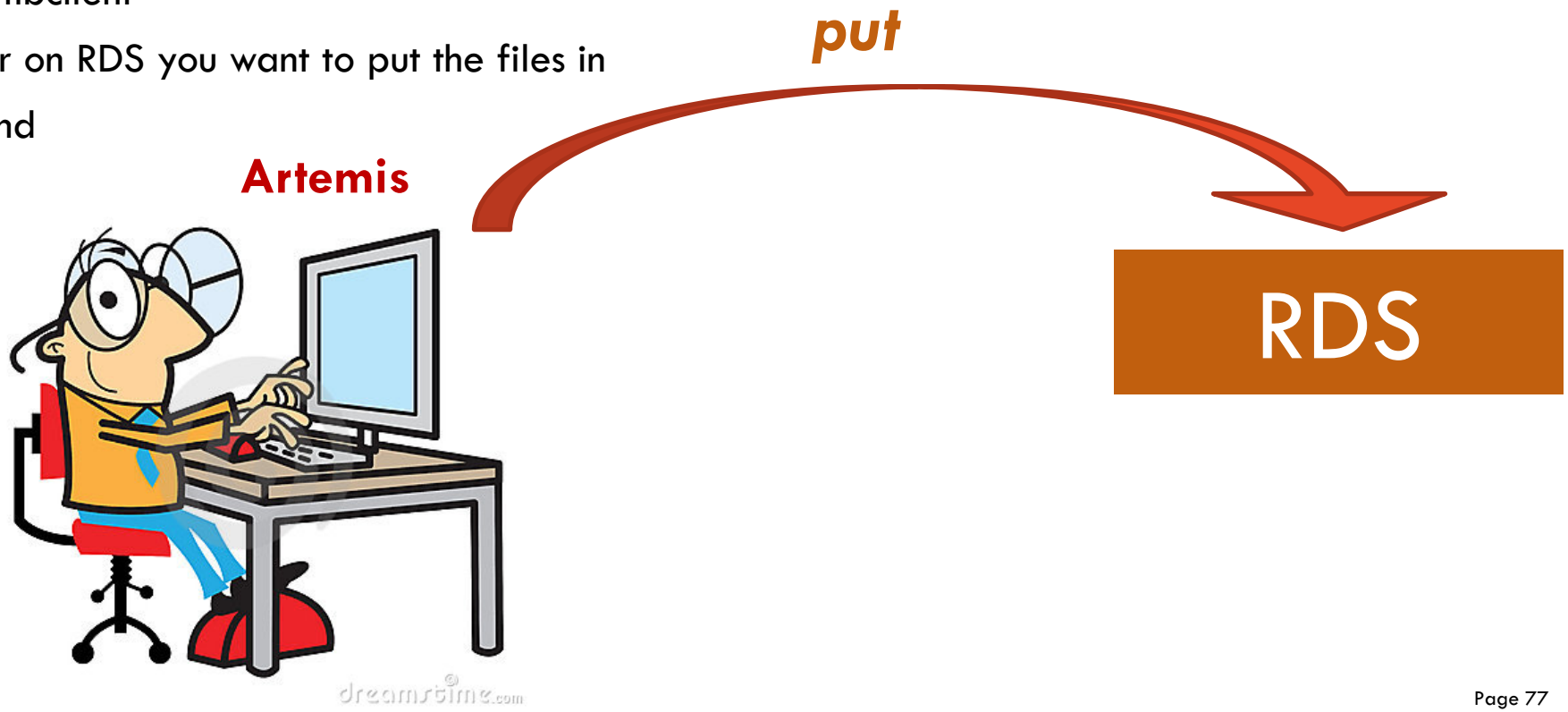
- '`put`' a file from Artemis onto RDS

Step 1: change into the directory on Artemis that contains the file/files you want to copy to RDS

Step 2: connect to RDS with smbclient

Step 3: change into the folder on RDS you want to put the files in

Step 4: issue the 'put' command



Data transfer with smbclient: location matters!

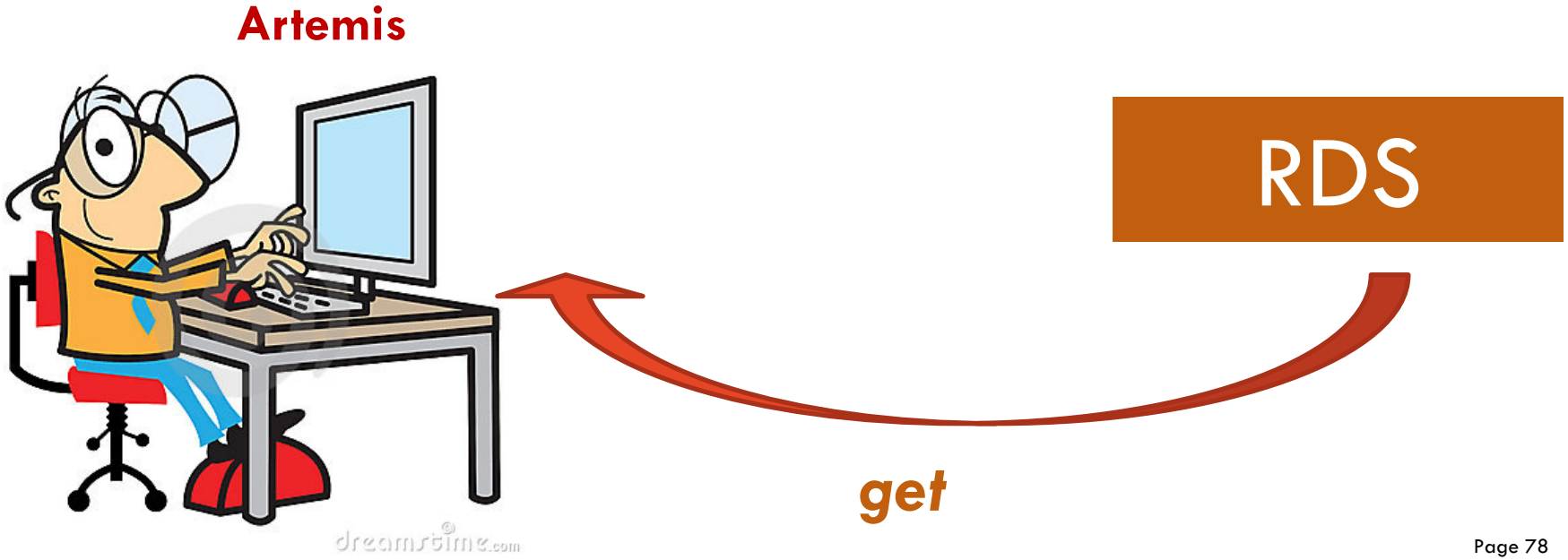
- ‘`get`’ a file from RDS onto Artemis

Step 1: change into the directory on Artemis where you want the files to end up

Step 2: connect to RDS with smbclient

Step 3: change into the folder on RDS that contains the file/files you want to get

Step 4: issue the ‘get’ command



Data transfer with smbclient: location matters!

- If you need to change or query a directory on Artemis (local) **while logged in to RDS** (remote) with smbclient, you need to use modified commands:

Operation	Classic RDS	Artemis
Change directory	`cd Directory`	`lcd Directory`
Present working directory	`pwd`	`!pwd`
List directory contents	`ls`	`!ls`

Data transfer: HPC to Classic RDS - smbclient

- ‘put’ canfam3_chr5.fasta from Artemis onto RDS

1. In your HPC terminal, change into your directory that contains the file
   - **cd /project/Training/<YourName>**
2. Enter the following command to connect to RDS via smbclient:

   smbclient //research-data.shared.sydney.edu.au/<FAC> -U <UniKey> -W SHARED

   *Make sure to replace ‘<FAC>’ with your 3 letter Faculty abbreviation, eg SMS, FAE, FVS

   ***Training login users**: use ‘Test’ in place of ‘FAC’

   *Make sure to replace ‘<UniKey>’ with your UniKey (or your HPC training login eg ict_hpctrain10)
3. Enter your UniKey (or training login) password when prompted

Data transfer: HPC to Classic RDS - smbclient

- ‘put’ canfam3_chr5.fasta from Artemis onto RDS

4. Run ‘ls’ to list the RDS project folders (prefix ‘PRJ’) you have permission to access

```
cwil2281@login2:~
[cwil2281@login2 ~]$ smbclient //research-data.shared.sydney.edu.au/FVS -U cwil2281 -W SHARED
Enter cwil2281's password:
Domain=[SHARED] OS=[Unix] Server=[DellFluidFS]
smb: \> ls
  .                                   D        0  Thu Jul  2 18:35:58 2015
  ..                                  D        0  Thu Jul  2 18:35:58 2015
  GRP-Genopheno                       D        0  Thu Jun 25 09:36:59 2015
  README.txt                          A     3514  Fri Jun 20 11:37:11 2014
  .DS_Store                          AH    10244  Wed Jul  8 17:03:29 2015
  PRJ-Cribbing                        D        0  Fri Jul  3 12:21:32 2015
  .clusterConfig                    DHS        0  Fri Jun 20 13:00:19 2014
  PRJ-DOG_DEAF                        D        0  Fri Aug 21 12:12:40 2015
  PRJ-BioServ                         D        0  Fri Aug 21 12:05:35 2015
Error in dskattr: NT_STATUS_NOT_SUPPORTED
smb: \>
```

Data transfer: HPC to Classic RDS - smbclient

- ‘put’ canfam3_chr5.fasta from Artemis onto RDS

5. Change into the RDS project folder you want to copy the files to

```
cd PRJ-Project
```

*Make sure to replace ‘project’ with your abbreviated project name

***Training login users,** your project is ‘Training’

```
[ict_hpctrain10@login2 RDS-ICT-HPCTRAINING-RW]$ smbclient //research-data.shared.sydney.edu.a
u/Test -W SHARED -U ict_hpctrain10
Enter ict_hpctrain10's password:
Domain=[SHARED] OS=[Unix] Server=[DellFluidFS]
smb: \> ls
  .                                   D        0  Thu Jan 28 09:20:32 2016
  ..                                  D        0  Thu Jan 28 09:20:32 2016
  PRJ-Training                        D        0  Wed Jan 20 18:15:30 2016
  README.txt                          A     3514  Fri Jun 20 11:37:11 2014
  .DS_Store                          AH    10244  Mon Nov 30 11:31:43 2015
  .clusterConfig                    DHS        0  Mon Nov 23 09:19:34 2015
  transactionManager.log.filepart      A 25720290570  Tue May 26 14:42:29 2015
Error in dskattr: NT_STATUS_NOT_SUPPORTED
smb: \> cd PRJ-Training\
smb: \PRJ-Training\>
```

Data transfer: HPC to Classic RDS - smbclient

- ‘put’ canfam3_chr5.fasta from Artemis onto RDS

6. **Extra step for training login users:**

   - Make a directory inside PRJ-Training, call it your name
   - Change into it
   - Use it for the rest of the data transfer activities

7. Use the ‘put’ command to copy canfam3_chr5.fasta to RDS

```
put canfam3_chr5.fasta
```

```
smb: \PRJ-Training\stephen\> put canfam3_chr5.fasta
putting file canfam3_chr5.fasta as \PRJ-Training\stephen\canfam3_chr5.fasta (51231.6 kb/s)
(average 51231.6 kb/s)
```

Multiple transfer: HPC to Classic RDS - smbclient

- ‘mput’ the two .pbs files from Artemis onto Classic RDS:

1. Verify your **local** directory (on Artemis) contains the .pbs data files

```
!pwd
!ls
```

```
smb: \PRJ-Training\stephen\> !pwd
/project/Training/stephen
smb: \PRJ-Training\stephen\> !ls
134_R1.fastq.gz  align.pbs          hello.pl   transfer
134_R2.fastq.gz  canfam3_chr5.fasta  index.pbs
```

Multiple transfer: HPC to Classic RDS - smbclient

- ‘mput’ the two .pbs files from Artemis onto Classic RDS:

2. Put both .pbs files onto Classic RDS using the star wild card:

```
mput *.pbs
```

3. You will be prompted ‘put file align.pbs?’ → enter ‘y’ for yes, ‘n’ for no

```
smb: \PRJ-Training\stephen\> mput *.pbs
Put file index.pbs? y
putting file index.pbs as \PRJ-Training\stephen\index.pbs (18.7 kb/s) (average 18.7 kb/s)
Put file align.pbs? y
putting file align.pbs as \PRJ-Training\stephen\align.pbs (29.3 kb/s) (average 24.3 kb/s)
```

Multiple transfer: HPC to Classic RDS - smbclient

- ‘mput’ the two .pbs files from Artemis onto Classic RDS:

4. Run the ‘`ls`’ command to see that your files were successfully copied to RDS:

```
smb: \PRJ-Training\stephen\> ls
  .                                   D        0  Tue Feb 14 13:02:03 2017
  ..                                  D        0  Tue Feb 14 13:01:48 2017
  align.pbs                           A      720  Tue Feb 14 12:59:57 2017
  index.pbs                           A      402  Tue Feb 14 12:59:56 2017
  canfam3_chr5.fasta                  A 40500003  Tue Feb 14 12:57:23 2017
Error in dskattr: NT_STATUS_NOT_SUPPORTED
```

5. Type ‘q’ then hit enter (or use ctrl + d) to **quit smbclient**

Data transfer: Classic RDS to HPC - smbclient

- ‘get’ bt.tar.gz from RDS onto Artemis

1. Enter the following command at the Artemis shell prompt (or use the tip):

```
smbclient //research-data.shared.sydney.edu.au/FAC
-U UniKey -W SHARED
```

2. Enter your UniKey password when prompted
3. Use ‘cd’ to change into the RDS project folder with the file:

```
cd PRJ-Training/ict
```

Tip: Use the ‘up’ arrow to retrieve the smbclient command you entered earlier, or run history | grep smbc then run !n ☺

Data transfer: Classic RDS to HPC - smbclient

- ‘get’ bt.tar.gz from Classic RDS onto Artemis

5. Use ‘get’ to copy bt.tar.gz to your ‘Transfer’ directory

```
get bt.tar.gz
```

You should see ‘getting file…’

```
smb: \> cd PRJ-Training
smb: \PRJ-Training\> get bt.tar.gz
getting file \PRJ-Training\bt.tar.gz of size 45646 as bt.tar.gz
(810.5 KiloBytes/sec) (average 810.5 KiloBytes/sec)
```

Multiple transfer: Classic RDS to HPC - smbclient

- ‘mget’ both .pbs files from Classic RDS to HPC

1. **Turn off prompt by typing ‘prompt’** then hit enter
2. Use ‘mget’ to copy both .pbs files to your Artemis directory

```
mget *.pbs
```

- Notice how both files are transferred without you needing to type ‘y’ for each

```
smb: \PRJ-Training\ict\> mget *.pbs
getting file \PRJ-Training\ict\align.pbs of size 720 as align.pbs (78.1 KiloBytes/sec)
 (average 707.5 KiloBytes/sec)
getting file \PRJ-Training\ict\index.pbs of size 402 as index.pbs (43.6 KiloBytes/sec)
 (average 625.6 KiloBytes/sec)
```

3. Confirm your ‘get’ and ‘mget’ commands have worked by running local ls:

```
!ls
```

‘**prompt**’ asks you before transferring every file in a multiple transfer- this is by default on.
To toggle it off (recommended for numerous and/or large files), simply enter the command ‘prompt’

Multiple transfer: recursive copy with smbclient

- If you want to copy a directory and it's contents, you need to perform a **recursive copy**
  - To do this you need to have 'recurse' on (off by default)
  - Ideal to also have 'prompt' off (on by default)

- This will also retain the directory structure of the directory you are copying
  - Very handy for copying whole projects or essential directory structures

- Make sure you are one directory level above the directory you want to copy

Multiple transfer: recursive copy with smbclient

- Recursively copy the ‘bt’ directory from Artemis HPC to Classic RDS

1. Exit smbclient by typing ‘q’

   ```
   q
   ```

2. Extract the bt.tar.gz file:

   ```
   tar -zxf bt.tar.gz
   ```

   You should now have a directory in your current directory on Artemis called ”bt”:

```
[ict_hpctrain1@login3 stephen]$ tar -zxf bt.tar.gz
[ict_hpctrain1@login3 stephen]$ ls
134_R1.fastq.gz  align.pbs  bt.tar.gz            hello.pl
134_R2.fastq.gz  bt         canfam3_chr5.fasta   index.pbs
```

Multiple transfer: recursive copy with smbclient

- Recursively copy the 'Output' directory from Classic RDS to Artemis HPC

Opposite to 'prompt', '**recurse**' is by default off. Toggle it on by entering 'recurse'.

3. Log into smbclient again:

```
smbclient //research-data.shared.sydney.edu.au/Test -U
ict_hpctrain1 -W SHARED
```

3. Change directories to your directory on Classic RDS

```
cd PRJ-Training/<YourName>
```

3. Toggle prompt off by typing '`prompt`'
4. Toggle on recursive copy by typing '`recurse`'

Multiple transfer: recursive copy with smbclient

- Recursively copy the 'Output' directory from Classic RDS to Artemis HPC

5. Recursively copy the 'bt' directory and all it's contents to RDS: **`mput bt`**

```
smb: \> cd PRJ-Training
smb: \PRJ-Training\> cd stephen
smb: \PRJ-Training\stephen\> prompt
smb: \PRJ-Training\stephen\> recurse
smb: \PRJ-Training\stephen\> put bt
bt does not exist
smb: \PRJ-Training\stephen\> mput bt
putting file bt/btC16.pbs as \PRJ-Training\stephen\bt\btC16.pbs (29.2 kb/s) (average 29.2 kb/s)
putting file bt/C16.o1016707 as \PRJ-Training\stephen\bt\C16.o1016707 (264.9 kb/s) (average 147.0 kb/s)
putting file bt/bt.C.16 as \PRJ-Training\stephen\bt\bt.C.16 (6897.6 kb/s) (average 3522.3 kb/s)
```

Classic RDS to Local Computer: Map a network Drive

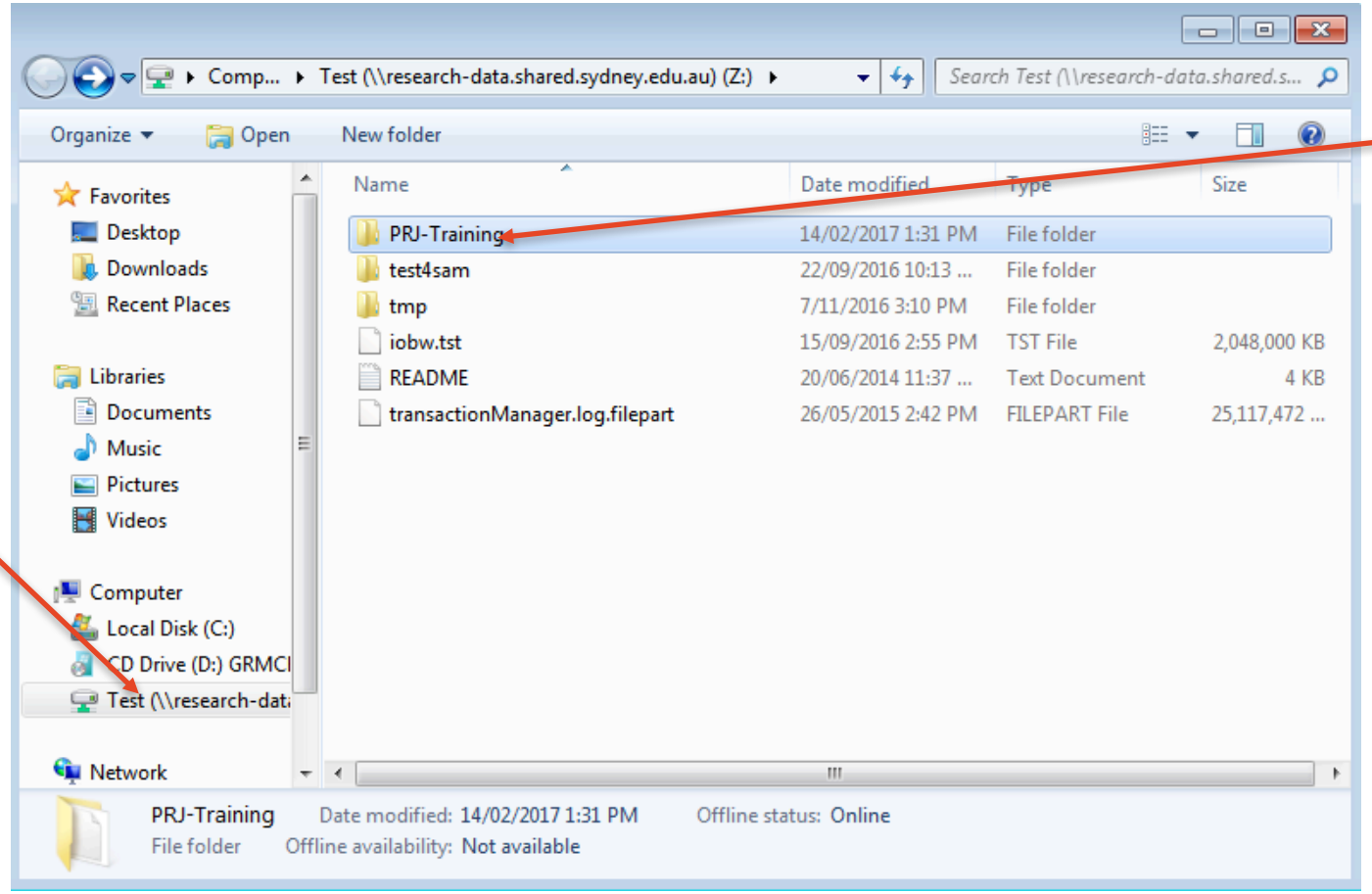
- To transfer data between your local computer and Classic RDS, you can map a network drive and drag-and-drop files
- Detailed instructions are available on the University Website: http://staff.ask.sydney.edu.au/app/answers/detail/a_id/174/~/how-do-i-map-my-network-drive-so-i-can-access-my-network-storage%3F

Classic RDS to Local Computer: Map a network Drive

- For Windows 7 and Classic RDS using the HPC training accounts:
- From the start menu click **Computer**
- When the Computer Window displays, click the **Map network drive button**.
- Select a letter from the dropdown list in the Drive field
- Enter the path of the shared drive you wish to connect to in the **Folder field**: `\\research-data.shared.sydney.edu.au\Test`
- Tick the **Connect using different credentials** box
- Enter **mcs\UniKey** and your **password**. Use your ict_hpctrain unikey and password.
- Click **OK**

Classic RDS to Local Computer: Map a network Drive

- If successful, you should see PRJ-Training in a file explorer window
- You can transfer data here just like you would on a normal windows computer



Your Classic RDS directory is here

Classic RDS is here