

The University of Queensland

Research Computing Centre

Transferring Files User Guide

Transferring Files User Guide

[Document Status](#)

[Background](#)

[Cluster Connection Tools](#)

[Options for Staying "Always" Connected](#)

[RDM Collection for Data Transfers](#)

[FUSE Filesystem Access](#)

[SMB Mounts](#)

[Individual Login Nodes and Cluster Aliases](#)

[Encryption](#)

[Host Key Fingerprints for HPC](#)

[Having Trouble?](#)

[Connections to Awoonga Timing Out](#)

[Connections Rejected](#)

[Connections Dropping Out](#)

[Some Best Practice Tips](#)

[PBS Job Scripts from Windows Computers ... beware of ^M](#)

[Small Files ... Lump Them](#)

[Text Files ... Compress Them](#)

[Messy Directories ... Get a Grip](#)

[Some Usage Scenarios](#)

[Windows + FileZilla + Tinaroo1](#)

[Windows + WinSCP + Tinaroo2](#)

[Linux + RSync + FlashLite2](#)

[Mac + Fugu + Awoonga](#)

[File Transfers within a Cluster](#)

Document Status

Last updated on July 8 2020 by David Green

Background

Most HPC users will need to upload input and download outputs at some point in time.

This user guide is an overview of how it can be achieved.

- The HPC login nodes are publicly accessible servers.
- The HPC login nodes only support encrypted file transfers (i.e. sftp and not ftp)
- The HPC compute nodes have several network interfaces with different names. The shortest name is not usually the fastest interface!

Cluster Connection Tools

The following table lists some of the tools available that can be used to connect to the HPCs so you can

- transfer files
- operate command line sessions
- display X11 graphics on your local computer

This list is not an endorsement of any particular tool.

Use at your own risk.

	Windows	Linux	Mac	Android

File Transfers	FileZilla CyberDuck WinSCP psftp	FileZilla CyberDuck scp at the command line gftp	FileZilla CyberDuck scp at the command line RBrowser Fugu (inactive)	Turbo Client Synchronise Ultimate
Command Line Sessions	PuTTY MobaXterm	ssh at the command line PuTTY	ssh at the command line Fugu (inactive)	Serverauditor JuiceSSH
X11 Servers	Xming MobaXterm	X11 is installed by default	X11 is now a separate project called XQuartz	Search for "X11 server" in Play Store

Just search the internet for them by name if any of these links are out-of-date.

You should be able to locate these tools for download using an internet search engine or your "app store". Once you have the software installed, you can start transferring files to/from the cluster.

Options for Staying "Always" Connected

It is possible to use RDM or to set up other less permanent connections to/from HPC for the purpose of transferring files.

RDM Collection for Data Transfers

HPC enabled RDM collections are able to automatically mount onto the HPC on demand. RDM collections can be accessed from workstations using the various mechanisms outlined in the RDM website. Many people are using this as a way to transfer data into the HPC.

If you transfer a text file using the RDM from a Windows computer and want to use it on the HPC, you may find that you need to "fix" the file using the `dos2unix` command.

FUSE Filesystem Access

Another option is to use a feature called FUSE filesystem access which is supported on the HPC. This allows you to connect into HPC using a special tool called `sshfs` and to access files like they were a "mapped drive".

It is straight forward to use on a **linux workstation (or virtual machine)** and resembles a `scp` or `rsync` invocation:

```
sshfs uqdgree5@tinaroo1.rcc.uq.edu.au:/home/uqdgree5 /mnt/HPC/home/uqdgree5
```

Once established, you can then map the drive using the Nautilus file browser (Places ... Network ...). I have visited with some of our HPC users who are quite effective in using this approach for their work.

There are also **options for Windows and Mac** (just search for "sshfs client for Mac Windows" using a search engine)

SMB Mounts

It is possible to use the `smbclient` commandline tool on the HPC to reach out and map an external storage service if it supports SMB/CIFS. On the HPC check the documentation using `man smbclient`.

Do NOT type your password on the command line invocation of the command.

This is because the password may appear in the process table for all to see! Instead enter your password only when prompted.

Individual Login Nodes and Cluster Aliases

Cluster	Cluster Login Alias	Login Nodes
Awoonga	awoonga.rcc.uq.edu.au	awoonga1.rcc.uq.edu.au
FlashLite	flashlite.rcc.uq.edu.au	flashlite1.rcc.uq.edu.au
		flashlite2.rcc.uq.edu.au
Tinaroo	tinaroo.rcc.uq.edu.au	tinaroo1.rcc.uq.edu.au
		tinaroo2.rcc.uq.edu.au

Wiener

wiener.hpc.net.uq.edu.au

wiener.hpc.net.uq.edu.au

Encryption

It is all about trust!

Your file transfer tool will need to establish an encrypted "tunnel" through which your username, password and data will pass between your computer and the server.

As part of establishing an encrypted connection, your file transfer tool will connect to the server and need to validate its identity.

If it has never connected to that server before (or the server has updated its identity) you will be prompted to confirm that you trust the server.

Only say "Yes" if the server identity that is offered to you by your software matches what it should be!

The host keys for HPC systems are as follows.

All the login nodes for a particular cluster should use the same key.

Host Key Fingerprints for HPC

The following table lists the md5sum "fingerprints" for each cluster for different key types.

Your connection tool should show you one or more of these when it challenges you about your connection.

Cluster	Key Type/Algorithm	Host Key Finger Print
Awoonga	ssh-rsa	99:c0:b3:70:d4:e5:47:0d:ef:89:45:5c:37:21:b4:96
	ssh-ed25519	5a:a2:c4:87:ae:28:77:6d:af:73:7b:9f:4f:fa:1d:6b
	ecdsa-sha2-nistp521	fc:b6:22:38:95:b8:8b:72:b2:f6:a7:6a:f2:03:fb:41
FlashLite	ssh-rsa	b8:5c:2f:d9:76:ce:d2:23:a6:00:21:db:73:8a:3a:f1
	ssh-ed25519	d6:76:cd:1a:79:c4:ed:c4:2d:3f:ff:59:61:fc:ea:ec
	ecdsa-sha2-nistp521	d5:d5:d7:1d:2b:6b:61:61:c1:67:f8:da:1c:bd:3d:96
Tinaroo	host	SHA256:1o+e0tTtxK51l+T3UIL3q7sVgbqSheVvcfdin20ELMA
	ssh-rsa	SHA256:yi4wnEfZ0v0seUWSvgMetBDaDGfovihej+Pjfn80gFg
	ssh-dsa	SHA256:0/p35uohoBWeEWniIGzbQrWAJCSLQN91qwQzA3ejifk
	ssh-ecdsa	SHA256:sr36dGBdatgkneopSSe1UJJ1b6Uo42jjgRXbGqFs9UY
	ssh-ed25519	SHA256:0c7633cZvPLUHVnJeBwDDau8XS+sDRw+bsjT5QNfjaA
Wiener	ssh-rsa	66:4b:0c:f0:ec:e2:43:c0:5b:d6:d0:0b:73:e3:ad:86
	ecdsa-sha2-nistp521	84:e4:c0:0b:fe:ef:38:89:39:9d:11:28:24:0e:12:58

You can verify these fingerprints using these commands on the login node:

```
ssh-keygen -lf /etc/ssh/ssh_host_key.pub
ssh-keygen -lf /etc/ssh/ssh_host_rsa_key.pub
ssh-keygen -lf /etc/ssh/ssh_host_dsa_key.pub
ssh-keygen -lf /etc/ssh/ssh_host_ecdsa_key.pub
ssh-keygen -lf /etc/ssh/ssh_host_ed25519_key.pub
```

Having Trouble?

Connections to Awoonga Timing Out

You may experience a connection timed out error.

```
Status:    Connecting to awoonga1.rcc.uq.edu.au...
Response:  fzSftp started
Command:   open "uqgree5@awoonga1.rcc.uq.edu.au" 22
Command:   Pass: *****
Error:     Connection timed out
```

```
Error:    Could not connect to server
Status:   Waiting to retry...
```

This can manifest itself in repeated attempts to login to the server, and eventually being blocked by a security feature that rejects frequent attempts to login to a specific HPC account. If that happens to you, apply the fix described here but wait 5-10 minutes before attempting to reconnect.

To work around this we recommend that you increase the Timeout setting in your file transfer tool to 120 seconds (2 minutes), or more.

Tool	How to Change Timeout
FileZilla	Edit ... Settings ... Connection ... Timeout (type in the value of 120)
WinSCP	SessionSettings ... Advanced ... Connection ... Timeouts (type or click up to 120)
CyberDuck	TBA

Connections Rejected

Users sometimes experience a problem with their file transfer connection dropping out shortly after connecting. This can happen if your **home directory is over quota**.

Connections Dropping Out

Another problem is when connections are made but drop out shortly after files begin to transfer. There was a case recently where the user was transferring large files to the HPC using an older laptop over wireless network.

The same sort of large file transfers were fine from a desktop computer.

Reducing the MTU value of the wireless network adapter to 1490 bytes fixed the problem.

Some Best Practice Tips

PBS Job Scripts from Windows Computers ... beware of ^M

Be careful if you are editing or saving your PBS script on a Windows computer and transferring it to HPC.

There can be control characters (^M) embedded in the file that are not visible but may cause the batch job to spontaneously fail when it starts to run.

You can avoid this by always transferring the text file using text transfer mode of WinSCP/FileZilla/CyberDuck.

If you want to be sure, you can also use the `dos2unix` command on the HPC to ensure the file is 100% Unix (and PBS) compatible.

Small Files ... Lump Them

You probably would never withdraw \$200 from an ATM using 10 distinct \$20 transactions, would you?

Transferring every file involves connection and filesystem overheads, in addition to the actual transfer of the contents of the file.

If you have a lot of small files, you will find that sending them individually can be a bit slow.

Instead, we recommend that you combine your small files into an archive file format.

Depending on your situation, you would create one or more TAR or ZIP format archive file(s) on the sending machine, before transferring the archive files to the receiving machine.

If you use tar, then it does not need to compress the file, so packing and unpacking is very quick.

On Linux

```
tar cf alltogether.tar DirectoryOfManyFiles/
zip -r alltogether.zip DirectoryOfManyFiles/
```

Text Files ... Compress Them

Why travel with a mostly suitcase when your stuff fits in an overnight bag?

ASCII text files can often be compressed quite a lot (to <20% file size).

Depending on your network connection, it may be advantageous for you to

- compress the text files on the sending machine using gzip or bzip2 or 7zip

- transfer the compressed files
- uncompress the files on the receiving machine.

Messy Directories ... Get a Grip

Your data can build up and before you know it you have a directory structure that is difficult to work with. You need to get organised.

For that situation, we have a couple of useful file management tools that will be able to aggregate and archive your gnarly data.

fpart dpart hsync parallel

More details to follow.

Some Usage Scenarios

Windows + FileZilla + Tinaroo1

Start FileZilla and enter the following parameters in the top of the application.

Parameter	Value	Notes
Host	sftp://tinaroo1.rcc.uq.edu.au	SFTP is better for large text files.
Username	Your UQ User Name	
Password:	Your UQ password	
Port:	22	Usually 22 for encryption

Press the Enter key ... Do not click the Quickconnect button when you connect to a host for the first time.

You can use the Quickconnect to activate a connection you have used previously.

Windows + WinSCP + Tinaroo2

Start WinSCP and enter the following parameters

Parameter	Value	Notes
File protocol	Either SFTP or SCP	SFTP is better for large text files
Host name	tinaroo2.rcc.uq.edu.au	tinaroo.rcc.uq.edu.au will also work as well
Port number	22	Usually 22 for encryption
User name	Your UQ User Name	
Password		Leave blank, do not save passwords in profiles

Click Save so you can re-use this profile default name for the profile is username@servername

Click Login to commence the connection to the server. You will be prompted for your password (do not save it).

The WinSCP application will have a split screen with your Windows computer on the left and the linux server (tinaroo2) on the right.

By default, you will land in your home directory on Tinaroo, however you should be able to save a different landing point for future connections (using bookmarks and/or preferences).

Linux + RSync + FlashLite2

Like the manual page says ... `rsync` is a fast and extraordinarily versatile file copying tool.

It has a convenient feature of transferring only the files that have changed or appeared on the sender side.

It has a wealth of options for controlling what, and how data, is transferred.

Some of the most commonly used options are

- `-a` the archive mode is a combination of other options for ownership, permissions, recursion etc.

- `--delete` will ensure that files that have been removed from the sender side will be removed on the receiver side
- `-v` verbose mode will list things out as they get considered and/or transferred
- `-n` the dry run mode will tell you what would happen but doesn't do anything about transfers

The following example will create a complete copy of the directory called `myDataDirectory` within my `/30days/uqgree5` folder on flashlite:

```
rsync -av myDataDirectory uqgree5@flashlite2.rcc.uq.edu.au:/30days/uqgree5/
```

The following example would make sense in a PBS Pro batch job.

It would copy all of my MATLAB mat files from the Inputs directory into local disk (`$TMPDIR`) where the job is running.

```
rsync -a ${PBS_O_WORKDIR}/Inputs/*.mat ${TMPDIR}/
```

Mac + Fugu + Awoonga

Sorry I don't use a Mac!

You might find this article from [Columbia University](#) useful.

File Transfers within a Cluster

The following section has been updated but may still contain torque syntax or sample torque output.

The fast interconnects between nodes in clusters can be used by appending `.ipoib` (FlashLite) and `.ib0` (Tinaroo) to the node's shortform hostname.

This arises because the cluster nodes have multiple interfaces and multiple nicknames.

```

davidg@tinaroo1:~> grep tn513a /etc/hosts
10.120.20.214    tn513a.local    tn513a
10.120.10.214   tn513a.bmc
10.120.30.214   tn513a.ib0

```

```

davidg@flashlite1:~> grep fl222 /etc/hosts
192.168.10.186  fl222.local    fl222
10.255.100.186 fl222.ipoib
192.168.30.186 fl222.ipmi fl222.bmc

```

The clusters automatically use the IB interfaces for accessing network storage (GPFS).

The `$PBS_NODEFILE` for a batch job will contain automatically contain the IB interface names.

Consequently, the message passing data within multi-node batch jobs will also use the IB interface.

```

##### Execution Started #####
JobId:83073
UserName:uqgree5
GroupName:gris-uq
ExecutionHost:tn307d
#####
uqgree5@tn307d:~> cat $PBS_NODEFILE
tn307d.ib0
tn307d.ib0
tn307d.ib0
tn307d.ib0

```

You may derive some benefit from copying files internally via that interface, although bear in mind that using an encrypted connection between nodes comes with overheads.

You will usually find it quicker/simpler to just copy data to/from the GPFS filestores (`/home/30days/90days`) using the `cp` command.