

VK7) Copying and moving your data to Viking.

Below we will summarise the typical way users can move/copy their data to Viking. As more methods or data sources become apparent we will add this information here.

Viking is a self contained machine, therefore you will notice your normal UoY home directories are not available. This is intentional for the following reasons.

- If the dedicated network link between campus and Viking goes down then this may cause slowness or jobs to fail. Instead, jobs should continue to run until the link is re-established.
- If a user tries to read/write from the university filestores using Viking, it is possible that they could bring the entire storage system down for the university.
- UoY home directories are not designed for high performance computing. We have a special high performance filestore available on Viking called scratch. You will see a link to your scratch area in your home area on Viking.

- [Windows](#)
- [Linux](#)
 - [scp](#)
 - [rsync](#)
 - [Moving files to/from Viking outside of the campus network](#)
 - [VPN](#)
 - [scp command via "jump host"](#)
- [Google Drive - Moving data from Viking directly to Google Drive](#)
 - [Prerequisites](#)
 - [Authorisation route](#)
 - [Client ID](#)
 - [Configuring rclone](#)
 - [Testing rclone](#)
 - [Issues](#)

Windows

If you are copying data from a Windows device it is recommended you use WinSCP. Details on how to set it up can be found [here](#)

[Copying files from another host or desktop to interactive Linux servers](#)

Whilst the content in the above link makes explicit reference to the "[research0.york.ac.uk](#)" server, the information is also applicable to Viking.

Also if you are trying to move lots of data, you will be better off copying it directly to your scratch area.

Linux

You can copy your data from any Linux device to Viking using the following commands

- `scp`
- `rsync`

Here are a couple of examples.

scp

Recommended for a small number of files

You wish to copy your data to Viking and place it in your scratch area on Viking.

```
#For an individual file
scp filename viking.york.ac.uk:~/scratch/

#For a folder with lots of files
scp -r dirname viking.york.ac.uk:~/scratch/
```

There are many options you can use with scp. To view these options either run

```
man scp
```

on the device you are using scp on or have a look at this [scp wiki page](#)

rsync

Recommend for a large number of files. Rsync will compare what is already in place, therefore if the network is interrupted you can run the command again and it will pick up where you left off. It will only copy files that do not exist on the other server or files that have been changed.

You wish to copy your data to Viking and place it in your scratch area on Viking.

Here is an example

```
rsync -avz dirname viking.york.ac.uk:~/scratch
```

There are many options you can use with rsync. To view these options either run

```
man rsync
```

on the device you are using rsync on or consult the [rsync webpage](#)

Moving files to/from Viking outside of the campus network

If your client device (e.g. personal laptop) is not connected to the campus network you can still access your files on Viking:

VPN

If you have an active VPN client connected to the University's VPN service you can access Viking as if you were using the campus network, with the instructions detailed above being applicable. More information on the University's VPN service can be found here:

<https://www.york.ac.uk/it-services/services/vpn/>

scp command via "jump host"

If you don't wish to make use of the University's VPN service to connect to Viking, and have access to the `scp` command (detailed above), then you can make use of the "jump host" option `-J` and the University's SSH gateway service, `ssh.york.ac.uk`:

scp with jumphost template

```
scp -J abc123@ssh.york.ac.uk abc123@viking.york.ac.uk:/path/to/files /path/to/destination
```

Replacing "abc123" with your University username and substituting the correct source and destination paths.

Below is an example in which the file "testfile" in the "abc123" user's Viking scratch directory is copied to their home directory on their local device:

scp files on viking via jumphost

```
scp -J abc123@ssh.york.ac.uk abc123@viking.york.ac.uk:/mnt/lustre/users/abc123/testfile /home/abc123/
```

More information on the SSH gateway service, including the implementation of two-factor authentication, can be found here:

<https://www.york.ac.uk/it-services/services/ssh/>

Google Drive - Moving data from Viking directly to Google Drive

We know a number of Viking users like to store data on Google Drive. It is possible to copy data directly from Viking to your google drive folder. Below we will provide instructions on how to set this up.

Prerequisites

Authorisation route

You will need to log into Google using a browser to authorise rclone. To generate the login token you will need, in order of complexity, one of:

1. rclone available on your local machine,
2. be using a graphical session on Viking, or
3. set up SSH port forwarding to the rclone process running on Viking.

For more information about the authorisation process review the rclone documentation here: https://rclone.org/remote_setup/

Client ID

As part of the process you will be asked to setup a client-id. **It is worth setting this up first before configuring rclone.** The instructions can be found here:

<https://rclone.org/drive/#making-your-own-client-id>

Once you have a Client ID you can begin configuring rclone.

Configuring rclone

Log in to Viking and load the rclone module:

```
module load tools/rclone
```

Verify the module has loaded with `module list` and then you can begin the configuration process:

```
rclone config
```

rclone will ask you a number of questions. We recommend the following responses as basic working defaults when setting it up:

1. `n` - new remote
2. Give it a name - don't use spaces, makes it awkward (I used "gdrive")
3. `17` - Google Drive (NB: do **not** select "Google Cloud Storage")
4. Enter the Application Client ID (see step above to create an ID)
5. Enter the Client Secret (see step above to create an ID)
6. `3` - "Access to files created by rclone only" (unless you want to access all files, in which case select "1")
7. ``Enter`` - ID of the root folder
8. ``Enter`` - Service account credentials
9. `n` - Edit advanced config
10. "Use auto config". If you are using a graphical session you can answer `y`, otherwise answer `n`. If you answered `n` then:
 - a. You will be prompted to authorise the rclone client. Run the command on your machine with a browser, then log in to the URL presented.
 - b. Authorise rclone to connect to Google Drive.
 - c. Copy and paste code back into terminal.
11. `n` - Do not configure this as a Team Drive (unless it is a Team Drive, in which case choose `y`)
12. `y` - This is okay
13. `q`

rclone should now be ready to use!

Testing rclone

To test your installation, create a test directory and run the following:

```
rclone copy --create-empty-src-dirs TEST_DIR gdrive:VIKING-DATA-TEST
```

If you login into Google Drive you should see the folder VIKING-DATA-TEST and any files that are in TEST_DIR.

Refer to the [rclone documentation](#) for full information about rclone usage.

Issues

If you see the following

```
2021/08/16 11:33:23 Fatal error: listing Team Drives failed: googleapi: Error 403: Insufficient Permission: Request had insufficient authentication scopes., insufficientPermissions
```

Please check you are not trying to sync to a Team Drive instead. If you are trying to sync to a Team Drive you will need to amend question 11 to `y`.