Data Transfer

On HPC systems, we can interact with the command line over SSH. We can also create and move file and directories. We can run a program and get some output. But the question is how can we

- Upload files, say our datasets, to the HPC machine?
- Download files, say the results of our analysis, from the HPC machine for further analysis locally?

There are two popular methods to do these:

- FTP = File Transfer Protocol; and
- SCP = Secure Copy

FTP file transfer

You need an FTP client, such as FileZilla. You tell the client which server to connect to e.g,: hpc.yourinstitution.edu and port: 22. You then authenticate with your username and password. Note that All FTP communication to the HPC machine will be 100% secure.

Secure Copy (`scp` command)

The secure Copy (`scp`) command is another way to transfer files to and from the HPC machine. Like FTP, it is secure, but you will need to invoke it from the command line. Therefore SCP is non-interactive, but the good news is- it can easily be scripted.

```
scp <file_name.ext> <user_name>@hpc.yourinstitution.edu:<dest_dir>
```

Get files from the Internet

To do this, we need to be at the command line (i.e. logged into the HPC machine by `ssh`) and type: `wget <URL_to_file>`. This will download the file

(e.g., a dataset) to your current working directory.

Dealing with archived files

In order to transfer data and save space, you will often find .zip files every where. This signifies that the file actually contains one or more files bundled for convenience and then compressed. It's called an archive. Common archive formats have extensions `tar`, `tgz`, `gz`, `bz2`. You can easily unzip these files as follows:

```
unzip <file_name>.zip
tar xvf <file_name>.tar
tar xzvf <file_name>.tgz
gunzip <file_name>.gz
bunzip2 <file_name>.bz2
```

Notice that the file ends with an extension in the left, you can use the corresponding command (`unzip` for `zip`, etc.) in the right!