Introduction to DBE Cluster Computing in Windows

Matt Shotwell
(shotwelm@musc.edu)

Some notes

The examples below use the username `shotwelm`. However, this should be replaced with your NetID, and you should use your password to log in to the cluster computer. The network address of the cluster computer is `ccrc.mdc.musc.edu`. You can see some information about the cluster by following the link http://ccrc.mdc.musc.edu. These examples are intended to accompany the live introductory session.

Install PuTTY

PuTTY is a bundle of software for connecting to other computers. You can get the software by following the link http://tartarus.org/~simon/putty-snapshots/x86/putty-installer.exe. This installer will install several programs, two of which are PuTTY and PSFTP. PuTTY is used to log on to the cluster computer interactively and issue commands. PSFTP is used to transfer files to and from the cluster computer.

Example PSFTP usage

The following is an example PSFTP session. This program lets us transfer files from our computer to the cluster computer and vice versa. This example demonstrates how the R script `test.R` is transfered from your PC to your home directory on the cluster computer. Text following the `psftp>` promt are typed by the user. You may type `help` to see a list of possible commands.

```
psftp>open ccrc.mdc.musc.edu
login as: shotwelm
shotwelm@ccrc.mdc.musc.edu's password: <password>
Remote working directory is /home/nfs/shotwelm
psftp>dir
Listing directory /home/nfs/shotwelm
drwxr-x--x   11 shotwelm shotwelm      512 Dec 16 09:33 .
drwxr-x--x    5 root     root            0 Jan 21 18:44 ..
psftp>put C:\test.R
local:C:\test.R => remote:/home/shotwelm/test.R
psftp>dir
drwxr-x--x   11 shotwelm shotwelm      512 Dec 16 09:33 .
drwxr-x--x    5 root     root            0 Jan 21 18:44 ..
-rw-r--r--    1 shotwelm shotwelm       12 Jan 21 18:49 test.R
psftp>exit
```

Example PuTTY usage

The following is an example PuTTY session. Once the program is started, the PuTTY configuration window is popped up. All that is necessary at this step is to enter the address of the cluster computer (`ccrc.mdc.musc.edu`) into the 'Host Name (or IP address)' box, and click the 'Open' button. The configuration window should close and a 'terminal' window will open. The following example starts where the 'terminal' window is opened. Text following the `$` symbol is typed by the user.

```
login as: shotwelm
shotwelm@ccrc.mdc.musc.edu's password: <password>
Last login: Thu Jan 21 18:57:17 2010 from shotwelm-labr.135can.musc.edu
[shotwelm@ccrc ~]$ ls
test.R
[shotwelm@ccrc ~]$ whoami
shotwelm
[shotwelm@ccrc ~]$ exit
```

Example cluster job script

The following example is the contents of a script file, with the extension `.sh`, indicating it is a *sh*ell script. A script file is used to tell the cluster computer what you want it to do. When the script file is finished, you submit it to the cluster computer for processing. This script tells the cluster computer to run the command `R CMD BATCH test.R`. The rest of the file is for configuration. You don't really need to worry about any of the other lines except the ones with `-N` and `-M`. These lines tell the cluster computer the name (`-N`) and your email address. If you specify your email address, you will be notified by email when your job is complete.(`-M`).

```
#!/bin/bash
#$ -N test
#$ -M shotwelm@musc.edu
#$ -m eas
#$ -S /bin/bash
#$ -e /dev/null
#$ -o /dev/null
#$ -V
#$ -cwd
R CMD BATCH test.R
```

Example job sumbit

Suppose that the script above is stored in a file called `test.sh` in your home directory on the cluster computer (you could write the script on your PC and transfer it to the cluster computer using PSFTP), along with the R script `test.R`. Then we can log onto the cluster computer using PuTTY and submit this job to the cluster computer using the `qsub` command. The following examlpe begins at the same point as the Example PuTTY usage above and demonstrates how the `test.sh` job is submitted to the cluster computer.

```
login as: shotwelm
shotwelm@ccrc.mdc.musc.edu's password: <password>
Last login: Thu Jan 21 18:57:17 2010 from shotwelm-labr.135can.musc.edu
[shotwelm@ccrc ~]$ ls
test.sh test.R
[shotwelm@ccrc ~]$ qsub test.sh
Your job 7930 ("test") has been submitted
[shotwelm@ccrc ~]$ exit
```

Example data retrieval

Once you have received an email that the job is complete (assuming you provided an email address), you may wish to retrieve the job results. In the case above, the command `R CMD BATCH test.R` will generate a file called `test.Rout` that will contain the results, in addition to any datafiles that

were written by the script. The PSFTP program may be used to retrieve these files from the cluster computer. The following script retrieves the `test.Rout` file and saves it on you PC desktop.

```
psftp>open ccrc.mdc.musc.edu
login as: shotwelm
shotwelm@ccrc.mdc.musc.edu's password: <password>
Remote working directory is /home/nfs/shotwelm
psftp>dir
Listing directory /home/nfs/shotwelm
drwxr-x--x   11 shotwelm shotwelm      512 Dec 16 09:33 .
drwxr-x--x    5 root     root          0 Jan 21 18:44 ..
-rw-r--r--    1 shotwelm shotwelm       12 Jan 21 18:49 test.R
-rw-r--r--    1 shotwelm shotwelm      874 Jan 21 22:53 test.Rout
-rw-r--r--    1 shotwelm shotwelm      139 Jan 21 21:35 test.sh
psftp>get test.Rout
remote:/home/shotwelm/test.Rout => local:test.Rout
psftp>exit
```

Basic Linux commands

- `ls` - list contents of a directory
- `cd <dir>` - change directories
- `rm <file>` - remove a file
- `cp <from> <to>` - copy a file
- `mv <from> <to>` - move a file
- `nano <file>` - edit a file
- `qsub <file>` - submit a job script to the cluster
- `exit` - logout