Bioinformatics Hub Phoenix cluster

Users

Bioinformatics Hub

Stephen Pederson
Hien To
Jimmy Breen (a1650598)
Alastair Ludington (a)
Lakshmi Chandrapaty (a1690027)

Robinson Research Institute

Jimmy Breen (a1650598)
Ben Mayne
Shalem Leemaqz

Setup

To login to Phoenix via ssh. You will need your university account id

ssh a1650598@phoenix.rc.adelaide.edu.au

It will ask you for a password. This is your university account password.

You dont really want to have keep on typing that everytime you login, so lets setup an ssh config file. If you've ever used ssh, you will have an .ssh directory in your home folder

```
cd ~/.ssh/
```

This contains all your ssh relevant stuff like ssh keypairs (will explain later) and authorized keys. To create an alias for the phoenix cluster, lets create the config file (if you dont have one already)

```
echo "Host phoenix
HostName phoenix.rc.adelaide.edu.au
User a1650598
ServerAliveInterval 60" >> .ssh/config
```

Now you can just login to Phoenix, by doing the following:

```
ssh phoenix
```

You still need to put in your password. But we can skip this step easily by using a key-pair. A key-pair is like a handshake between two networks (this being your local computer and the Phoennix cluster) that enables password-less access to the desired network. You can do this between any two networks that you can access via ssh.

The steps i'll run through now are contained in this link

Step One—Create the RSA Key Pair
The first step is to create the key pair on the client machine (there is a good chance that this will just be your computer):

```
ssh-keygen -t rsa
```

It will ask you for a pass-phrase (THIS IS VERY IMPORTANT TO REMEMBER...but then again, all passwords are fairly important i guess). We will use RSA encryption here, but there are others that you could potentially try (dsa for example). I would suggest using this for now

```
Enter passphrase (empty for no passphrase):
```

The entire process will probably look like this:

```
ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/demo/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/demo/.ssh/id_rsa.
Your public key has been saved in /home/demo/.ssh/id_rsa.pub.
The key fingerprint is:
4a:dd:0a:c6:35:4e:3f:ed:27:38:8c:74:44:4d:93:67 demo@a
The key's randomart image is:
+--[ RSA 2048]----+
|          .oo.   |
|         . o.E   |
|        +  . o   |
|       . = = .   |
|        = S = .  |
|       o + = +   |
|        . o + o .|
|            . o  |
|                 |
+-----------------+
```

What has been made are two things:
1. A encrypted key; and 2. A public key that needs to be transfered to the *other* server that

you want to talk to

For me, this would be:

```
ssh-copy-id a1650598@phoenix.rc.adelaide.edu.au

The authenticity of host '12.34.56.78 (12.34.56.78)' can't be established.
RSA key fingerprint is b1:2d:33:67:ce:35:4d:5f:f3:a8:cd:c0:c4:48:86:12.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '12.34.56.78' (RSA) to the list of known hosts.
user@12.34.56.78's password:
```

The keys should be in a "public key" file called:

```
~/.ssh/authorized_keys
```

Try ssh'ing in and seeing if everything went through

SSH keys are obviously very sensitive, so they wont work if the file permissions are not set to read-only. We dont want anyone editing them or deleting them.

Introduction

Ok we're in, now what do we do with the cluster. First, lets have a look at the file-system thats attached

```
[a1650598@phoenixl1 ~]$ df -h
Filesystem              Size  Used Avail Use% Mounted on
rootfs                   99G   36G   58G  39% /
10.33.139.1:/opt/boot/hpc6
                         99G   36G   58G  39% /
none                     24G  140K   24G   1% /dev
/dev/ram                 10G   29M   10G   1% /ram
/dev/sda1               7.8G  4.7G  2.7G  64% /localscratch
tmpfs                    24G   12K   24G   1% /dev/shm
phoenixnfs1.rc.adelaide.edu.au:/home
                         40T  7.7T   33T  20% /home
phoenixnfs1.rc.adelaide.edu.au:/opt/shared
                        1.0T  148G  876G  15% /opt/shared
phoenixnfs1.rc.adelaide.edu.au:/opt/shared/slurm/slurm.state
                        1.0T  148G  876G  15% /opt/shared/slurm/slurm.state
10.33.139.231@tcp:10.33.139.230@tcp:/data
                        349T  133T  199T  40% /data
```

The two areas which you will spend your time is:

- /home - This is where you put your custom programs, settings and store some data
- /localscratch: Scratch storage is the fastest storage, so ideally you want to run all your major analyses here

When you arrive in the machine, you should create a directory on localscratch. For example:

```
mkdir /localscratch/jbreen/
```

Modules & Programs

There are two types of programs on phoenix, ones that have been installed by us (i.e. me) locally and "modules"

Modules are convienient little commands that allow you to run a simple command to use important programs.

For example, If I wanted to use bowtie to run an alignment, but I dont have that program in my PATH, I can do the following:

```
[a1650598@phoenixl1 Programs]$ module avail

------------------------------------------------------------
/usr/share/Modules/modulefiles ------------------------------------------------
------------
dot          module-git   module-info modules      null         use.own


------------------------------------------------------------
/etc/modulefiles -----------------------------------------------------------
-----
OpenFOAM/2.1.1          cmake/2.8.12.1           gslib
mpfr/3.1.2             pythia/8.2.10
OpenFOAM/2.3.1          cuda/5.0                 gurobi/6.0.5
muscle/3.8.31          python/2.7.10
R/3.2.0               cuda/6.0                 hdf5/1.8.15
namd/2.10-cuda         qt/4.8.6
abaqus/6.14           cuda/6.5                 hdf5/1.8.15-patch1
ncbi-blast/2.2.24+      root/5.34.32
ansys/14.5            cuda/7.0                 hkl2000/1
ncbi-blast/2.2.27+      samtools/1.2
ansys/16.1            cufflinks/2.2.1           hoomd-blue/1.1.1
ncbi-blast/2.2.31+      scotch/6.0.0
ansys/16.2            epmr/15.04               hoomd-blue/1.2.0
netcdf/4.4.0-C         subread/1.4.6-p2
atlas/3.10.2          fastqc/0.11.3            hoomd-blue/hoomd-blue
netcdf/4.4.2-Fortran    terachem/v1.50k
bedtools/2.25.0        fastx_toolkit/0.0.14      intel/13.1.3.174
netcdf/4.4.2-Fortran-intel tophat/2.1.0
bioperl/1.6.924        flex/2.5.39              intelmpi/5.0.3.048
opencv/2.4.10          trim_galore/0.4.0
blast/2.2.24          gaussian/g09             java/1.8.0
openmm/6.2.0          trinity/2.0.6
boost/1.54.0          gcc/4.8.4                julia/0.3
openmm/6.2.0-cuda       uofa_utils/1
boost/1.55.0          gcc/5.1.0                lammps/2015.05.15
openmpi/gnu/1.8.4       wamit/v7101
bowtie/1.1.1          glibc/2.16.0             lammps/2015.08.10/gpu
openmpi/intel/1.8.1     wu-blast/2.0
bowtie/2.2.5          glog/0.3.3               lhapdf/5.9.1
openmpi-x86_64         xds/1
bwa/0.7.12           gmp/5.1.2                ls-dyna/8.0.0
paraview/4.1.0         zlib/1.2.8
ccp4/6.5             gnu-parallel/20150322      matlab/2014a
pgi/15.4
censor/4.2.29          go/1.5                  matlab/2015a
phenix/1.9-1692
cgal/4.3             gsl/1.9                  mpc/1.0.1
protobuf/2.6.1
```

I see the module for bowtie is listed there on the available modules. Now I want to load it:

```
[a1650598@phoenixl1 Programs]$ module load bowtie/2.2.5
```

Lets see if I can run the program:

```
[a1650598@phoenixl1 Programs]$ bowtie2
No index, query, or output file specified!
Bowtie 2 version 2.2.5 by Ben Langmead (langmea@cs.jhu.edu,
www.cs.jhu.edu/~langmea)
Usage:
  bowtie2 [options]* -x <bt2-idx> {-1 <m1> -2 <m2> | -U <r>} [-S <sam>]

  <bt2-idx>  Index filename prefix (minus trailing .X.bt2).
             NOTE: Bowtie 1 and Bowtie 2 indexes are not compatible.
```

So thats modules, but you might find that you dont have module command for a program that you need. This will happen quite often, so its best to locally install your program. Generally you will need the admins (Stephen, Hien or Jimmy) to do this step because it requires a admin password.

For programs that have already been installed, have a look in the directory /localscratch/Programs/

```
[a1650598@phoenixl1 Programs]$ ls -l /localscratch/Programs/
total 196968
drwxr-xr-x    8 a1650598      4096 Jun 11  2015 AdmixTools-20150610-git-
3065acc5
-rwxr-xr-x    1 a1650598 144533833 Jul 24  2015 AdmixTools.tar.gz
drwxr-xr-x    2 a1650598      4096 Jun  2  2015 Eigensoft_6_0_1
drwxr-xr-x    6 a1650598      4096 Feb 17 09:40 GBSX
drwxr-sr-x    7 a1650598      4096 Feb  4 21:27 HiC-Pro
drwxr-sr-x   10 a1650598      4096 Feb 13 10:46 HiCExplorer
-rwxr-xr-x    1 a1650598     35147 Nov 21  2013 LICENSE
drwxr-xr-x    2 a1650598      4096 Jun  5  2015 MafFilter
drwxr-sr-x    4 a1650598      4096 Sep  3  2015 Metaxa2_2.0.2
drwxr-sr-x    5 a1650598      4096 Aug 31  2015 PileOMeth
drwxr-sr-x   12 a1650598      4096 Jun  1  2015 PyVCF
drwxr-xr-x    6 a1650598      4096 Jun 11  2015 RAxML-git20150511-297bc55146
-rwxr-xr-x    1 a1650598       280 Sep  1  2015 README.afw-acad2-rsync
drwxr-sr-x  155 a1650598      4096 Feb 28 10:41 R_libs
drwxr-sr-x    4 a1650598      4096 Dec 19 04:09 SailfishBeta-
0.9.0_DebianSqueeze
drwxr-sr-x    4 a1650598      4096 Jan 13 06:31 SalmonBeta-
0.6.1_DebianSqueeze
drwxr-xr-x    8 a1650598      4096 Jun  2  2015 adapterremoval2-git-0afe24-
20150601
```

```
drwxr-xr-x    9 a1650598      12288 Aug 31   2015 angsd-20150831-git~8b89ba42
drwxr-xr-x    9 a1650598      12288 Aug 31   2015 angsd_0.902
-rwxr-xr-x    1 a1650598        144 Aug 31   2015 angsd_0.902_install.txt
drwxr-sr-x    9 a1650598       4096 Feb 14 13:06 autoconf-2.69
-rwxr-xr-x    1 a1650598    1927468 Apr 25   2012 autoconf-latest.tar.gz
drwxr-sr-x    4 a1650598       4096 Jun  6   2015 bamaddrg
drwxr-xr-x    6 a1650598       4096 Jun  2   2015 bamtools
drwxr-sr-x    7 a1650598       4096 Nov 17 10:45 bbmap
drwxr-xr-x    7 a1650598       4096 Jun 30   2015 bcftools
drwxr-xr-x    4 a1650598       4096 Jun  2   2015 bcftools-1.2
drwxr-xr-x    2 a1650598       4096 Sep 25 08:12 bedops_bin
-rwxr-xr-x    1 a1650598    5926099 Apr 22   2015 bedops_linux_x86_64-
v2.4.14.tar.bz2
drwxr-xr-x    8 a1650598       4096 Jun 11   2015 binutils-2.23.2
-rwxr-xr-x    1 a1650598   21440347 Mar 27   2013 binutils-2.23.2.tar.bz2
drwxr-xr-x    4 a1650598       4096 Jun  2   2015 bioawk
drwxr-xr-x    3 a1650598       4096 Jun  3   2015 bycatch
drwxr-xr-x   10 a1650598       4096 Sep  1   2015 dmd-2.067.0
drwxr-xr-x    3 a1650598       4096 Jul  1   2015 drift_test
drwxr-xr-x    6 a1650598       4096 Jul 29   2015 fastStructure-20150729-
gite47212f8
drwxr-sr-x    3 a1650598       4096 Sep  8   2015 gammytools
drwxr-xr-x    3 a1650598       4096 Sep  1   2015 gatk-3.3-0
drwxr-xr-x    4 a1650598       4096 Jun  2   2015 ghostscript-9.15
-rwxr-xr-x    1 a1650598     120234 Aug 17   2015 gmon.out
drwxr-xr-x    5 a1650598       4096 Jun  2   2015 gnuplot-4.6.6
drwxr-xr-x    4 a1650598       4096 Aug 25   2015 grg-utils
drwxr-xr-x    4 a1650598       4096 Sep  1   2015 heterozygosity_DoNotShare
drwxr-sr-x    3 a1650598       4096 Feb  4 21:03 hiCPro
drwxr-sr-x    6 a1650598       4096 Feb 13 10:30 hicup_v0.5.8
drwxr-sr-x    5 a1650598       4096 Apr 18   2015 hisat-0.1.6-beta
-rwxr-xr-x    1 a1650598   27352793 Apr 18   2015 hisat-0.1.6-beta-
Linux_x86_64.zip
drwxr-sr-x    5 a1650598       4096 Nov 20 07:32 hisat2-2.0.1-beta
drwxr-sr-x    7 a1650598       4096 Feb 27 13:04 homer
drwxr-xr-x    6 a1650598       4096 Jun 30   2015 htslib
drwxr-xr-x    6 a1650598       4096 Jun  2   2015 htslib-1.2.1
drwxr-sr-x    3 a1650598       4096 Oct 30 04:38 kallisto_linux-v0.42.4
drwxr-xr-x    5 a1650598       4096 Jun  2   2015 libevent-2.0.22
drwxr-xr-x   16 a1650598       4096 Aug 25   2015 libgd-gd-2.1.1
drwxr-sr-x    5 a1650598       4096 Sep  1   2015 malt
drwxr-sr-x    7 a1650598       4096 Sep 19   2015 methclone
drwxr-sr-x    5 a1650598       4096 Sep 19   2015 methpipe
drwxr-sr-x    9 a1650598       4096 Sep 28 17:14 methtuple
drwxr-xr-x    7 a1650598       4096 Aug  7   2015 mlRho
drwxr-xr-x    4 a1650598       4096 Jun  2   2015 msmc-20150413
drwxr-xr-x    3 a1650598       4096 Jun  2   2015 mugsy_x86-64-v1r2.2
drwxr-xr-x    7 a1650598       4096 Jun  2   2015 ncurses-5.9
drwxr-sr-x    7 a1650598       4096 Jan  5 15:50 ngi_visualizations
drwxr-xr-x    5 a1650598       4096 Aug 31   2015 ngsPopGen-20150831-
git~abeabb73
```

```
drwxr-xr-x    5 a1650598        4096 Jul 16   2015 nmap-6.49BETA4
drwxr-xr-x    3 a1650598        4096 Jun  3   2015 nuclearratescurve
drwxr-xr-x   11 a1650598        4096 Jun  9   2015 paleomix-git-2fc2560-20150601
drwxr-xr-x   10 a1650598        4096 Jun  2   2015 paleomix-git-2fc2560-
20150601-nope
drwxr-sr-x    3 a1650598        4096 Aug 24   2015 perl5
drwxr-xr-x    2 a1650598        4096 Jun  9   2015 picard-1.133
drwxr-sr-x    2 a1650598        4096 Jan 18 13:02 plink-1.90
drwxr-sr-x    2 a1650598        4096 Jul 12   2014 plinkseq-0.10
drwxr-sr-x    6 a1650598        4096 Feb 14 13:07 pullseq
drwxr-sr-x    5 a1650598        4096 Sep  9   2015 rnammer
drwxr-xr-x    8 a1650598        4096 Jun  3   2015 samtools
drwxr-xr-x    8 a1650598        4096 Jun  9   2015 samtools-0.1.19
drwxr-xr-x    4 a1650598        4096 Jun  2   2015 samtools-1.2
drwxr-xr-x    5 a1650598        4096 Jun  2   2015 samtools_1_1
drwxr-xr-x    3 a1650598        4096 Jun  2   2015 seqbility-20091110
drwxr-sr-x    4 a1650598        4096 Jan 15 09:46 simwreck
drwxr-sr-x    5 a1650598        4096 Jan 28 13:57 snpEff
drwxr-xr-x    6 a1650598        4096 Jul  1   2015 snpEff_4.1g
drwxr-sr-x    5 a1650598        4096 Jan 28 14:06 snpEff_4.2g
drwxr-xr-x    8 a1650598        4096 Nov 24 09:54 swDMR-1.0.6
drwxr-xr-x    4 a1650598        4096 Jun  2   2015 tmux-1.9a
drwxr-xr-x    4 a1650598        4096 Jul 29   2015 treemix-20150729-git361fac8d
drwxr-xr-x    2 a1650598        4096 Jun  2   2015 trim_galore_zip
drwxr-xr-x    4 a1650598        4096 Jun  2   2015 vcftools-0.1.12b
```

Within each should have executibles that you will need. For example:

```
[a1650598@phoenixl1 Programs]$ ll samtools-1.2/bin/
total 3528
-rwxr-xr-x 1 a1650598   71970 Jun  2  2015 ace2sam
-rwxr-xr-x 1 a1650598    5328 Jun  2  2015 blast2sam.pl
-rwxr-xr-x 1 a1650598    3364 Jun  2  2015 bowtie2sam.pl
-rwxr-xr-x 1 a1650598   18032 Jun  2  2015 export2sam.pl
-rwxr-xr-x 1 a1650598    6125 Jun  2  2015 interpolate_sam.pl
-rwxr-xr-x 1 a1650598   21100 Jun  2  2015 maq2sam-long
-rwxr-xr-x 1 a1650598   21020 Jun  2  2015 maq2sam-short
-rwxr-xr-x 1 a1650598   32675 Jun  2  2015 md5fa
-rwxr-xr-x 1 a1650598   21469 Jun  2  2015 md5sum-lite
-rwxr-xr-x 1 a1650598    6773 Jun  2  2015 novo2sam.pl
-rwxr-xr-x 1 a1650598   48036 Jun  2  2015 plot-bamstats
-rwxr-xr-x 1 a1650598    3208 Jun  2  2015 psl2sam.pl
-rwxr-xr-x 1 a1650598    8961 Jun  2  2015 sam2vcf.pl
-rwxr-xr-x 1 a1650598 3203836 Jun  2  2015 samtools
-rwxr-xr-x 1 a1650598   16478 Jun  2  2015 samtools.pl
-rwxr-xr-x 1 a1650598   11105 Jun  2  2015 seq_cache_populate.pl
-rwxr-xr-x 1 a1650598    3909 Jun  2  2015 soap2sam.pl
-rwxr-xr-x 1 a1650598    7888 Jun  2  2015 varfilter.py
-rwxr-xr-x 1 a1650598   51558 Jun  2  2015 wgsim
-rwxr-xr-x 1 a1650598    4157 Jun  2  2015 wgsim_eval.pl
-rwxr-xr-x 1 a1650598    3583 Jun  2  2015 zoom2sam.pl

[a1650598@phoenixl1 bin]$ ./samtools
./samtools: /lib64/libz.so.1: version `ZLIB_1.2.3.3' not found (required by
./samtools)
[a1650598@phoenixl1 bin]$ module load zlib
[a1650598@phoenixl1 bin]$ ./samtools

Program: samtools (Tools for alignments in the SAM format)
Version: 1.2 (using htslib 1.2.1)

Usage:   samtools <command> [options]

Commands:
  -- indexing
```

Now you can add some of these programs to your PATH variable that is set within bash. You will need to edit your ~/.bashrc file:

This is the start of my bashrc file:

```
[a1650598@afw Refs]$ cat ~/.bashrc

# .bashrc

# Source global definitions
if [ -f /etc/bashrc ]; then
        . /etc/bashrc
fi

# User specific aliases and functions
export PATH="/home/a1650598/.local/bin:\
/localscratch/Programs/bedops_bin:\
/localscratch/Programs/PileOMeth:\
/home/a1650598/acad_ngs_pipeline/bash_bin:\
/home/a1650598/perl5/bin:\
/home/a1650598/bin:\
/home/a1650598/local/HiC-Pro_2.7.3b/bin:\
/localscratch/Programs/bbmap:\
/localscratch/Programs/samtools-1.2/bin:\
/localscratch/Programs/plinkseq-0.10:\
/localscratch/Programs/plink-1.90:\
/localscratch/Programs/gammytools:\
/localscratch/Programs/SailfishBeta-0.9.0_DebianSqueeze/bin:\
/localscratch/Programs/hisat2-2.0.1-beta:\
/localscratch/Programs/SalmonBeta-0.6.1_DebianSqueeze/bin:\
/localscratch/Programs/HiCExplorer:\
/localscratch/Programs/HiCPlotter:\
/localscratch/Programs/segemehl:\
/localscratch/Programs/homer/.//bin":$PATH
```

Reference Sequences and/or Genomes

Obviously, the storage on the cluster gets full fairly quickly. So it makes sense to have a particular directory where you're reference sequences live. This way, multiple users can use the same sequences for their analysis, without having to download or transfer everything all the time

The directory of this will probably change

```
[a1650598@afw ~]$ cd /localscratch/Refs/
[a1650598@afw Refs]$ ll
total 7876
drwxrwxr-x.  3 a1650598 afw_users   12288 Jun  1  2015 blast_db
drwxrwxr-x. 10 a1650598 afw_users    4096 Mar 27  2014 Cereals
drwxrwsr-x.  2 a1650598 afw_users      47 May 28  2015 Cyanobacteria
-rw-rw-r--.  1 a1650598 afw_users   25922 May  8  2015 ERCC.fa.gz
drwxrwxr-x. 10 a1650598 afw_users    4096 Mar 24  2014 five_apes
drwxrwxr-x. 10 a1650598 afw_users    4096 Aug 25  2015 fungal_genomes
drwxrwsr-x. 11 a1650598 afw_users    4096 Jan 31 10:30 human
-rw-rw-r--.  1 a1118972 afw_users      47 Nov  2 10:00 info.txt
drwxrwsr-x.  4 a1650598 afw_users    4096 Feb 18 14:46 mouse
drwxrwsr-x.  2 a1650598 a1650598     4096 May 27  2015 PhiX
drwxrwxr-x. 45 a1650598 afw_users    4096 Nov 17 11:08 PlantGenomes
-rw-r-----.  1 a1650598 afw_users 7984536 Sep  3  2015
uchime_reference_dataset_11_03_2015.zip
```