**Hoffman2 questions**

**Read each question carefully and answer to the best of your ability. Questions may have more than one correct answer.**

1. Briefly (1 or 2 sentences) describe why somebody would want to use Hoffman2.

**Ans1.** If someone has to do a heavy duty computation or has huge amount of data, it might not be feasible to do them in personal computers. Hoffman2 uses the distributed computing approach to perform computation on different nodes or computers and return the result to the user. Therefore the user need not worry about the computation time or space.

1. How do you connect to Hoffman2 if you are a
   1. Unix, Linux, or Mac user?
   2. Windows user?

**Ans2.**

**Linux:** use the ‘ssh’ command in the terminal to log in: ssh *login-id@hoffman2.idre.ucla.edu*

**Windows:** we use ssh clients such as PuTTY, MobaXterm and Cygwin. Secondly, we can use the remote desktop option too via the **NoMachine** client.

1. Briefly describe the difference between a ‘login node’ and a ‘compute node’. When should you use each of these?

**Ans3.**

Login node is assigned when we log on to Hoffman2 where we can organize files and move around. On the other hand, a compute node is meant for heavy duty computations which is assigned when anyone submits a job to the scheduler. These computations should never be performed on the login node. The login nodes are meant for light-weight tasks such as submitting jobs and file operations (editing, copying). Any cpu-intensive or memory-intensive task, including compiling large programs, must be done either by batch job submission or in an interactive session on the compute nodes(Line taken from the website).

1. How much storage is available in your Hoffman2 home directory? Where can you access additional storage?

**Ans4.**

20GB is available for every user in the home directory. Temporary storage of 2 TB can be accessed at ‘/u/scratch’ and is valid for 7 days.

1. What will happen to a job that attempts to use more memory than what had been requested?

**Ans5.**

Jobs or interactive sessions using more processors or significantly more memory than was reserved with UGE may be terminated without prior notice by the System Administrator.

(Ref: https://www.hoffman2.idre.ucla.edu/computing/policy/)

1. You want to submit a job that requires 4 GB of memory and will complete in approximately 2 hours. Name several reasons why you should **not** request 16 GB and 24 hours of run time.

**Ans6.**

* It will delay the starting of the process because the scheduler would find it more difficult to allocate more memory and processing time for 1 particular user.
* It will also defeat the job scheduler’s back filling capability and waste cluster resources which can be used by other users.

**For questions 7-10, write down the appropriate commands (there can be more than one right answer!) See the IDRE website if you need help answering any of these questions.**

1. Request a compute node for interactive use; you believe the job requires 12 GB of memory and will take approximately 4 hours to complete.

**Ans7.**

qrsh -l h\_rt=4:00:00,h\_data=12G

1. Request a compute node for interactive use; you want 2 CPU cores, 4 GB of memory, and believe the job will take approximately 8 hours to complete.

**Ans8**

qrsh -l h\_rt=8:00:00,h\_data=4G -pe shared 2

1. Submit a job; you want to name the job ‘gene\_data\_lasso’, and you want to be emailed **only if the job is aborted**. You want 12 GB of memory, 8 hours run time, and the name of the job is ‘myjob.sh’.

**Ans9**

qsub -N gene\_data\_lasso -l h\_data=12G,h\_rt=8:00:00 -m a myjob.sh

1. Submit a job; this is the same as the job in question 9, with the following changes:
   1. You want to change the working directory to where you currently are in the file system.
   2. You also want to be emailed when the job starts and finishes.
   3. You only want to request 64 MB of memory.

**Ans10**

qsub -N gene\_data\_lasso -cwd -l h\_data = 64M,h\_rt = 8:00:00 -m bea myjob.sh