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.

   Hoffman2 is UCLA's supercomputing cluster. It is the largest and the most powerful cluster in the UC system. The Hoffman2 Cluster currently consists of 1,200+ 64-bit nodes and 13,340 cores, with an aggregate of over 50TB of memory. Hence, someone who has a huge amount of data and/or wants to perform heavy-duty computation would want to use the Hoffman2 cluster.

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

   - *Unix, Linux or Mac users can use the 'ssh' command in the terminal to log in as follows:*
     - *ssh login-id@hoffman2.idre.ucla.edu*
   - *Windows users can use an SSH client like MobaXterm, PuTTY, etc or they can use remote desktop such as 'NoMachine'*

3. Briefly describe the difference between a 'login node' and a 'compute node'. When should you use each of these?

   - *The Hoffman2 login nodes provide command line access to the Hoffman2 Cluster. 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.*

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

   Every user has a storage of 20 GB in the home directory for general use. One can access additional storage 2 TB for temporary use in '/u/stratch', which is kept for 7 days.

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

   Any job that attempts to use more memory than what had been requested will be terminated without prior notice by the System Administrator.

6. 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.

   One shouldn't request more time, memory or processors that what your job requires because that will delay its starting. It may also defeat the job scheduler's back-filling capability and waste cluster resources.

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.

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

   qrsh -l h_rt=4:00:00,h_data=12G,

8. 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.

   qrsh -l h_rt=8:00,h_data=4G, -pe shared 2

9. 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'.

   Qsub -N gene_data_lasso -l h_data=12G,h_rt=23:00:00 -m a myjob.sh

10. Submit a job; this is the same as the job in question 9, with the following changes:

    a. You want to change the working directory to where you currently are in the file system.
    b. You also want to be emailed when the job starts and finishes.
    c. You only want to request 64 MB of memory.

    Qsub -cwd -N gene_data_lasso -l h_data=64M,h_rt=23:00:00 -m bea myjob.sh