

Let's assume that you have given permission to use Pinbot server. That means you have access to directories '/home/USER', '/data1/USER', '/data2/USER', '/data3/USER' and '/data4/kanduriLab_repository'. 'USER' means your username on server, Example: '/home/santhilal', '/data1/santhilal' and so on.

- 1) Where (which directory) do you store/install your programs or save your scripts on Pinbot?
- 2) Assume you got some files from sequencing company/platform. Where (which directory) do you download the files? And what will you check before downloading the files to the server?
- 3) If the server is overloaded with the files and the administrator requests you to reduce the storage space. What will you do? Tip: Raw FASTQ files (uncompressed) and alignment files (BAM) occupy most of the storage A) Compress all uncompressed large files and folders in my user space B) Delete all the files and folders from my user space C) Delete unused/unwanted files and folders from my user space D) Delete all the reproducible files, such as alignment files E) Delete all raw FASTQ files from 'data1-3' F) Delete all raw FASTQ files from 'data4' F) Backup all old analysis files from 'data1-3' (finished projects) in external hard disk and delete them.
- 4) To perform analysis how do you choose between disks (data1, data2 and data3)?
- 5) What are the two key (logical) parameters required to run any program/tools?
- 6) What are the things to take care before submitting a job? What commands will you use?
- 7) How do you check total number of available CPUs on the server?
- 8) How to run a program/tool in background (so that you do not have to keep the terminal/bash open throughout the process)? If you want to see the progress of the tool running in background, how do you check?
- 9) Do all bioinformatics tools/programs support multi-threading (utilizing multiple CPUs/cores/threads for jobs)?
- 10) Give an example where a particular tool/task requires multiple CPUs/cores/threads for single job.
- 11) Assume you have 10 FASTQ raw files to align and your server has only 12 CPUs/cores. And all the CPUs are 100% free for your use. What is the **least efficient** way to do the alignment A) Submit 10 jobs parallel with one core each B) Submit one job at a time with 8 or more cores C) Submit 2 jobs parallel at a time with 6 cores each.
- 12) For the above options from **question 11**. Let's assume there are also other users submitting their jobs on the server. Which options you think is **more efficient** for you to use without overcrowding the CPUs?
- 13) How do you connect to the server (command)? Which directory are you in when you are logged into server? Once you are on server terminal, how do you come back to your local computer's home directory?