**Exploring the UAHPC with BASH: Commands Part 2**

**Manipulating Files in Unix**

Common tasks in bioinformatics are file manipulation to fit a particular format and parsing information. We will start to explore different file types common in bioinformatics applications later in the semester, but there are general methods to changing the structure and content of files that are applicable across most file types.

**Goals of this exercise**

1. Create new files using information from other files.
2. Write and append files.
3. Identify target information in a file.

**Parsing Data from Files**

There are a few file formats you will see across all areas of data management: tab-delimited and comma separated. Being able to pull desired information from these and presenting in a new file is a valuable skill.

For this exercise, you will manipulate existing files by pulling out target information and putting it into a new file in your directory.

You will be using the following file:

/scratch/mrmckain/Bioinformatics\_Class/Class\_Files/ McKain\_Collaborators.txt

1. Using the command cut,split the file McKain\_Collaborators.txt into three different files: McKain\_Collaborators\_Institutions.txt, McKain\_Collaborators\_Relationships.txt, and McKain\_Collaborators\_LastActive.txt.
   1. These files should be in your class directory.
   2. You will need to use the -f parameter for cut and the command > to redirect the standard out to a file.
   3. Provide a screenshot of the new files in your directory.
2. Write the command you used below.
3. Explain what the command is doing.
4. How would you change this command if you were using comma separated files?

**Appending an Existing File**

There are times when you will need to add to an existing file instead of making a completely new file. This is helpful when new data is added to a project.

You will be using the following file:

/scratch/mrmckain/Bioinformatics\_Class/Class\_Files/ McKain\_Addition.txt

1. Using a combination of cut and >>, add the information in the above file to each of the three files you created.
   1. Provide a screenshot of the contents of the files demonstrating you successfully added the new information.
2. Write the command you used below.
3. Explain what the command is doing.

**Getting Specific Information Out of Files**

Often, we need very specific information from files. There are many ways of getting this information, some easier than others.

In this exercise, you will have free reign to figure out how to answer the following questions.

1. Using the files you made above, create a file that combines all the information into a new file but only has those collaborators that are from The University of Alabama.
2. Paste your commands here.
3. Explain what these are doing.