

## DSCC 201/401 Homework Assignment #3

Due: **September 29, 2021 at 9 a.m. EDT**

Answers to these questions should be submitted via Blackboard. Please upload a file containing your answers and explanations to Blackboard (Homework #3: Linux Environment) as a Word document (docx), text file (txt), or PDF. All input and output should be plain text (cut and paste from a terminal window) consolidated into one file, with your name provided at the top. It is very important to include the command you used to give the answer to the question. Credit will only be given if the command to answer the question and the output have been provided.

For each of the following questions 1-8 provide the input command and the output generated.

1. Download the file `NIH_2020.tar.gz` from Blackboard. Upload the file to your scratch directory on BlueHive. What is the total amount of disk space used by the `NIH_2020.tar.gz` file that you copied to BlueHive? Show the command how you determined the size of the file.
2. Submit an interactive job to the debug partition on BlueHive to obtain a compute node with 1 core and 2 GB of RAM for 1 hour using a terminal session. What command did you type to obtain these resources? What is the hostname of the compute node you are using?
3. Read the manual pages for the `gunzip` and `tar` commands. Uncompress the `NIH_2020.tar.gz` file with the `gunzip` command and then subsequently unarchive the `NIH_2020.tar` file while logged into the compute node. Show the commands you used to uncompress and unarchive the file.
4. What is the combined total size of all files (amount of storage on disk) in megabytes located in the `NIH_2020` directory? Show the command you used to calculate this value.
5. How many files are located in the `set1` directory? Show the command you used to calculate this value.
6. Using the `find` command, what is the name of the single largest file located in the `set0` directory? What is the size of this file? Show the commands you used to determine the answer.
7. How many lines contain the string "Rochester" in all files in the `NIH_2020` directory and all of its subdirectories? Show the command you used to calculate this value.

8. All files in the `set0` and `set1` subdirectories of the `NIH_2020` directory are CSV (comma-separated values) files. Write a bash script to execute in the `NIH_2020` directory that will add the `.csv` extension to all file names in the `set0` and `set1` directories. For example, `nih200` should become `nih200.csv`. (**DSCC 401 Only:** Convert lowercase nih file names to uppercase NIH file names. For example, `nih200.csv` should become `NIH200.csv`, while `set0` and `set1` remain lowercase.)

**For questions 9, 10, and 11, create three separate files for each of the three Bash script shown below. Run the scripts on BlueHive, examine the outputs, and answer the following questions:**

```
#!/bin/bash
# script 1
pi=3.14159
two_pi=2*$pi
echo $two_pi
```

```
#!/bin/bash
# script 2
pi=3.14159
two_pi=$((2*pi))
echo $two_pi
```

```
#!/bin/bash
# script 3
pi=3
two_pi=$((2*pi))
echo $two_pi
```

9. Show the outputs of script 1, script 2, and script 3.
10. What is the purpose of the `#` in the Bash scripts?
11. What is the purpose of the double parentheses `(( ))` in the scripts?