
Tutorial 1

for

Big Data Programming in BASH
Version 1

Adam Ryan (14395076)

COMP47470

September 21, 2021

Contents

1	Questions	3
1.1	Purpose	3
1.2	Exercise 1	3
1.3	Exercise 2	3
2	Answers	5
2.1	Purpose	5
2.2	Exercise 1	5
2.3	Exercise 2	5

1 Questions

1.1 Purpose

This section charts the questions to the lab1 notebook in Big Data Programming which looks at using Bash and installing Docker.

1.2 Exercise 1

The following details the questions in section 1.

- What is the output of the commands:
 - `echo {0..9}`
 - `echo 1.{0..9}`
 - `echo {A..C}{0..2}`
- Use `sort` to sort the content of `/etc/passwd`
- Use `grep` to list only the line from `/proc/meminfo` that shows total memory (the name of the variable is `MemTotal`).
- Use `grep` to send all the lines with "to" in the `syslog` file (`/var/log/syslog_public`). Then redirect this output into the `syslog` in your home directory.
- List all of the files in a directory in reverse date order, so that the most recent is at the bottom (`ls`).

1.3 Exercise 2

The following details the questions in section 2.

- Download the dataset as `unirank.csv` using `wget` (capital o option).
- Explore the file.
- Find all the "colleges" in the list (e.g. `UniversityCollegeDublin`, think `grep`).
- List and group all the HEIs by state (`cut,sort`).
- Find the number of HEIs per state (`cut,sort,uniq`). (which state has the most institutes in the dataset?)

- Create a list (x, y) of couples containing the Tuitions and Fees (x) and the rank (y) for each institute. Get this file to your host machine and plot the values using Excel, gnuplot, etc. What do you notice?

2 Answers

2.1 Purpose

This section charts the solutions to the lab1 notebook in Big Data Programming which looks at using Bash and installing Docker.

2.2 Exercise 1

The following details the questions in section 1.

- What is the output of the commands:
 - `echo {0..9}` prints all of the characters 0 1 2 3 4 5 6 7 8 9
 - `echo 1.{0..9}` prints all of the characters 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9
 - `echo {A..C}{0..2}` prints all of the characters A.0 A.1 A.2 B.0 B.1 B.2 C.0 C.1 C.2
- Use sort to sort etc passwd was complete.
- `grep MemTotal /proc/meminfo`
- (File path for this exercise does not exist)
- `ls -ltr`

2.3 Exercise 2

The following details the questions in section 2.

- `wget -o unirank.csv http://csserver.ucd.ie/ thomas/unirank`
- `cat/sort/ etc`
- `grep College ./unirank`
- `cut -d"," -f1,3 unirank | sort -k2`
- `cut -d"," -f3 unirank | sort -k2 | uniq -c`
- `docker cp BigDataProgramming:/home/COMP47470/lab1/final_question.txt ./` and subsequently open in Excel.