

Course: Big Data

Lab 03

MapReduce

Fill answers of the questions below in the given tables.

Your screenshots must **contain commands** for required operations.

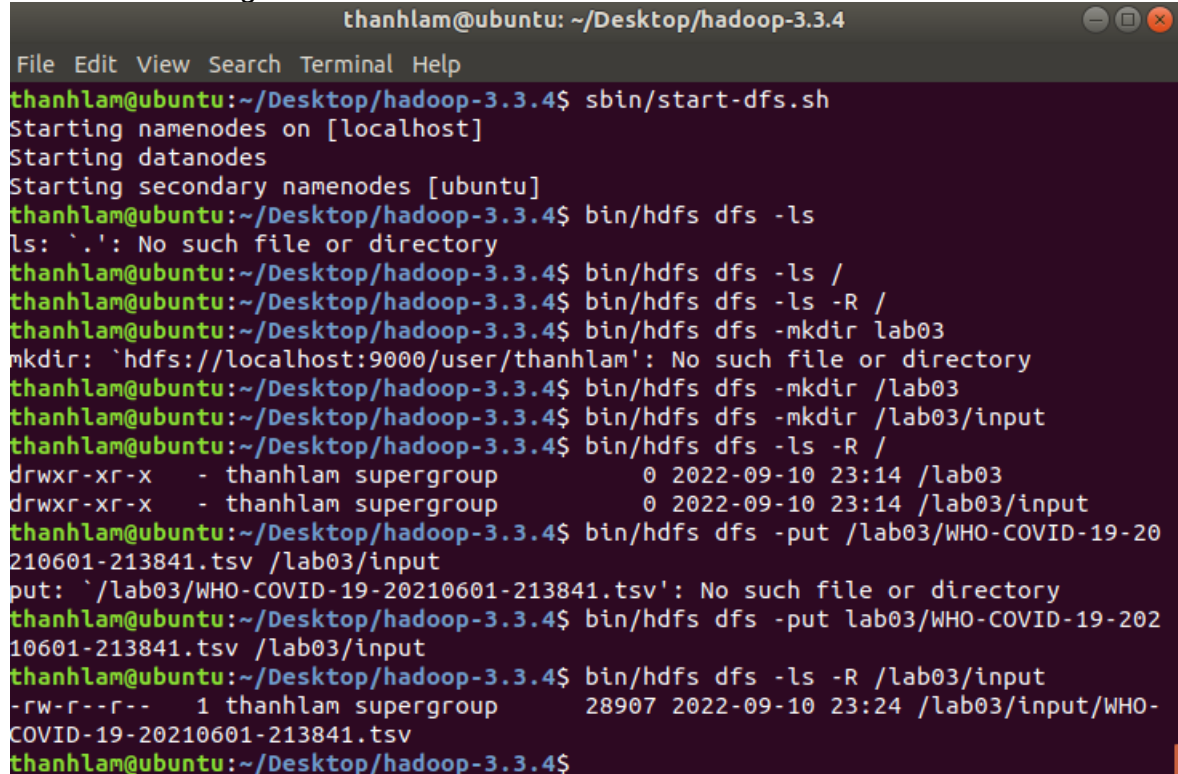
Question 1:

Given a tsv file [WHO-COVID-19-20210601-213841.tsv](#) which is corresponding to the [WHO Coronavirus \(COVID-19\) Dashboard](#).

Students are required to create a folder, named **lab03**, in HDFS and then copy the tsv to **lab03/input/**

Take a screenshot to show the content of **lab03/input/** in HDFS

Your screenshot goes here



```
thanhlam@ubuntu: ~/Desktop/hadoop-3.3.4
File Edit View Search Terminal Help
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -ls
ls: `.`: No such file or directory
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -ls /
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -ls -R /
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -mkdir lab03
mkdir: `hdfs://localhost:9000/user/thanhlam': No such file or directory
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -mkdir /lab03
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -mkdir /lab03/input
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -ls -R /
drwxr-xr-x  - thanhlam supergroup          0 2022-09-10 23:14 /lab03
drwxr-xr-x  - thanhlam supergroup          0 2022-09-10 23:14 /lab03/input
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -put /lab03/WHO-COVID-19-20210601-213841.tsv /lab03/input
put: `/lab03/WHO-COVID-19-20210601-213841.tsv': No such file or directory
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -put lab03/WHO-COVID-19-20210601-213841.tsv /lab03/input
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$ bin/hdfs dfs -ls -R /lab03/input
-rw-r--r--  1 thanhlam supergroup      28907 2022-09-10 23:24 /lab03/input/WHO-COVID-19-20210601-213841.tsv
thanhlam@ubuntu:~/Desktop/hadoop-3.3.4$
```

Question 2:

Create one and only one java file, named **ASEANCaseCount.java**, to run a MapReduce job that counts the number of cumulative total cases among ASEAN countries (*South-East Asia Region in the given data table*).

The output of the MapReduce job is located in **lab03/output-java/**.

Submit the source code file following the instructions in Submission Notice.

Question 3:

Create a pair of Python files, named **ASEANDeathCountMapper.py** and **ASEANDeathCountReducer.py**, to run a MapReduce job that counts the number of cumulative total deaths among ASEAN countries (*South-East Asia Region in the given data table*).

The output of the MapReduce job is located in **lab03/output-python/**.

Submit the source code files following the instructions in Submission Notice.

Submission Notice

- Export your answer file as pdf
- Rename the pdf following the format:
<student number>_HoTen.pdf
E.g. 123456_NguyenThanhAn.pdf
If you have not been assigned a student number yet, then use 123456 instead.
- Create a folder with the name as **<student number>_HoTen**, which contains
 - **<student number>_HoTen.pdf** → your answer
 - **java/** | → Java source code folder
| **ASEANCaseCount.java**
 - **python/** | → Python source code folder
| **ASEANDeathCountMapper.py**
| **ASEANDeathCountReducer.py**
- Compress the folder **<student number>_HoTen** in zip format and finally submit to the given form.
E.g. 123456_HoTen.zip
- Careless mistakes in filename, format, question order, etc. are not accepted (0 pts).