

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

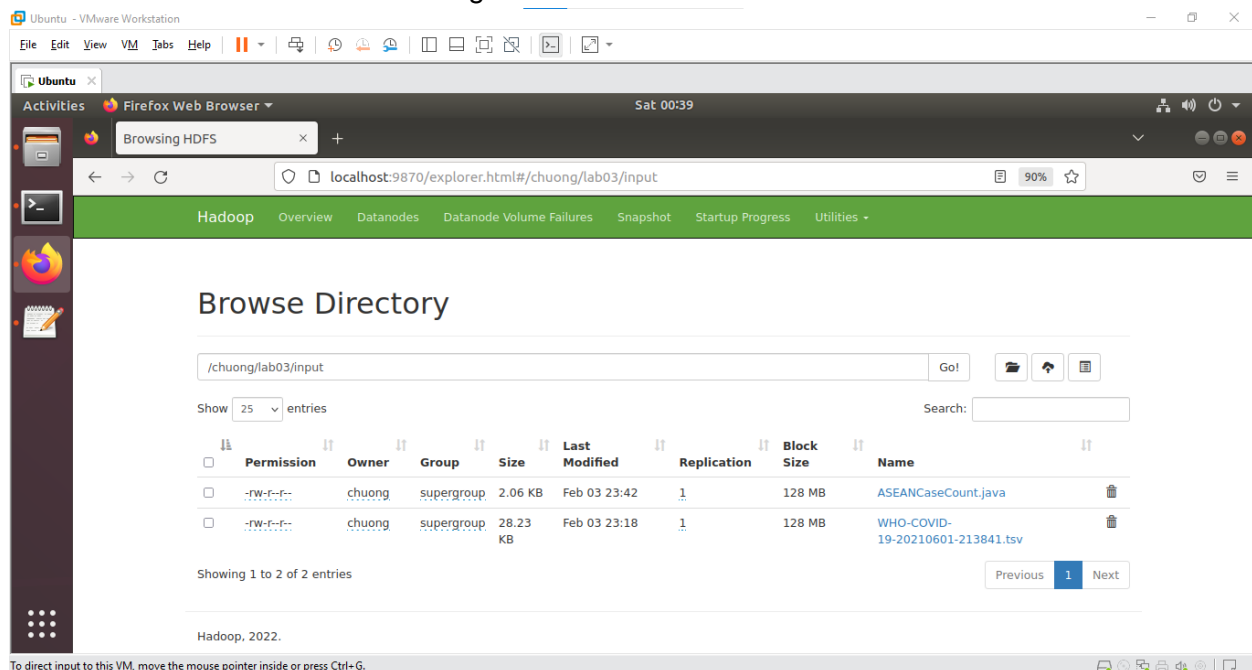
```
chuong@ubuntu: ~/hadoop-3.2.3$ hdfs dfs -put /home/chuong/WHO-COVID-19-20210601-213841.tsv /chuong/lab03/input
chuong@ubuntu:~/hadoop-3.2.3$ hdfs dfs -ls /chuong/lab03/input
Found 1 items
-rw-r--r-- 1 chuong supergroup 28907 2023-02-03 23:18 /chuong/lab03/input/WHO-COVID-19-20210601-213841.tsv
chuong@ubuntu:~/hadoop-3.2.3$ hdfs dfs -cat /chuong/lab03/input
cat: '/chuong/lab03/input': Is a directory
chuong@ubuntu:~/hadoop-3.2.3$ hdfs dfs -cat /chuong/lab03/input/*C
chuong@ubuntu:~/hadoop-3.2.3$ hdfs dfs -cat /chuong/lab03/input/WHO-COVID-19-20210601-213841.tsv
Name WHO Region Cases - cumulative total Cases - cumulative total per 100000 population Cases - newly reported in last 7 daysC
ases - newly reported in last 7 days per 100000 population Cases - newly reported in last 24 hours Deaths - cumulative total Deaths
- cumulative total per 100000 population Deaths - newly reported in last 7 days Deaths - newly reported in last 7 days per 100000 popu
lation Deaths - newly reported in last 24 hours Transmission Classification
Global - 170,363,852.000 2,182.386 3,383,026.000 43.337 358,691.000 3,546,870.000 45.436 76,025.000 0.974 7,370.
000
United States of America Americas 32,929,178.000 9,948.310 131,305.000 39.670 0.000 588,596.000 177.820 3,896.
000 1.180 0.000 Community transmission
India South-East Asia 28,175,044.000 2,041.660 1,226,170.000 88.850 127,510.000 331,895.000 24.050 24,664.000 1.7902
,795.000 Clusters of cases
Brazil Americas 16,515,120.000 7,769.650 431,862.000 203.170 43,520.000 461,931.000 217.320 12,863.000 6.0508
74.000 Community transmission
France Europe 5,566,214.000 8,558.220 59,437.000 91.390 8,541.000 108,558.000 166.910 756.000 1.160 15.000 Commun
ity transmission
Turkey Europe 5,242,911.000 6,216.470 56,424.000 66.900 6,933.000 47,405.000 56.210 1,137.000 1.350 134.00
Community transmission
Russian Federation Europe 5,071,917.000 3,475.480 62,006.000 42.490 8,475.000 121,501.000 83.260 2,700.000 1
.850 339.000 Clusters of cases
The United Kingdom Europe 4,484,060.000 6,605.280 21,518.000 31.700 3,111.000 127,781.000 188.230 60.000 0.0906
.000 Community transmission
Italy Europe 4,216,003.000 7,068.910 23,820.000 39.940 2,948.000 126,046.000 211.340 821.000 1.380 44.000 Cluste
of cases
```

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 (optional):

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).