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](https://drive.google.com/file/d/1TG6orBmU74s1_Z3NDsyntRb9-OAHIuy_/view?usp=sharing) which is corresponding to the [WHO Coronavirus (COVID-19) Dashboard](https://covid19.who.int/table).

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* |

## 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:

**lab03\_<student number>\_HoTen.pdf**

E.g. lab03\_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).