

Handson 3

Big Data Systems

Soheil Shirvani - 3720505



Contents

[Lunching Hadoop on Local 2](#_Toc86911783)

[Lunching Hadoop on Cluster 2](#_Toc86911784)

# Lunching Hadoop on Local

First we make the class CovidCaseNumberLocal similar to MaxTempertureLocal. We run our code on local and Test it on the VM Master computer. After each run we have to remove the output folder if we are using the same argument for running it.

After running and debugging on local and getting a result, we can run the code on cluster mode.

# Lunching Hadoop on Cluster

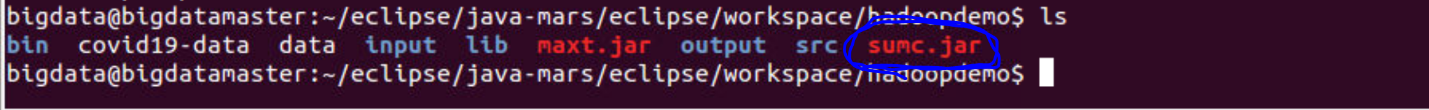
First we have to create a jar file including the classes (Mapper, Reducer, JobRunner):

Text

Description automatically generated

Then we have to move the Jar file to the project root directory:





Then we have to remove the output folder (if there is any from the previous runs):

Text

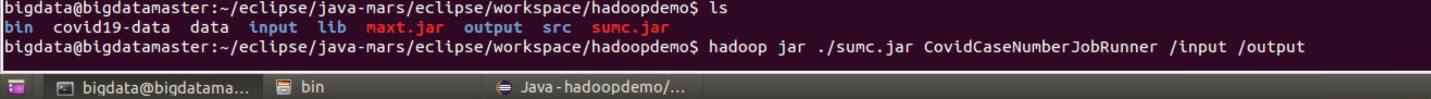
Description automatically generated

Then we have to remove the MaxTempertureInput from the folder and copy the new Covid-Data CSV file to the input folder in dfs directory:

Text

Description automatically generated

After that, we can run our Created Jar File in cluster mode:



And we can see it running on cluster:

A picture containing graphical user interface

Description automatically generated

Then after it finishes we can see:

Text

Description automatically generated

Text

Description automatically generated

And for the Output we can see the output folder and inside it:

Text

Description automatically generated

Here we can see the

\_Success folder. It means our code runs successfully and the results are saved in “part-r-00000” folder.

We can see inside the results folder like below:

Text

Description automatically generated

Here is just a sub part of the whole results. Only Countries with more than 1000000 new cases are shown in the output file. The results are provided in a txt file along with my submission.