

Hadoop Assignment

Ashvin Bachoo | Big Data Query | 17/11/2020

# INTRODUCTION:

This report consists of four questions.

1. Total number of deaths for Africa.
2. The country with the most number of cases as at now.
3. Which date had the highest number of new cases and in which country?
4. Top ten countries in terms of new cases in the last 2 months.

Steps:

1.Connect to Ambari.

2. files view/user/maria\_dev /upload the files.

3. python covid\_mauritius.py covid.csv.

4. Execute the code.

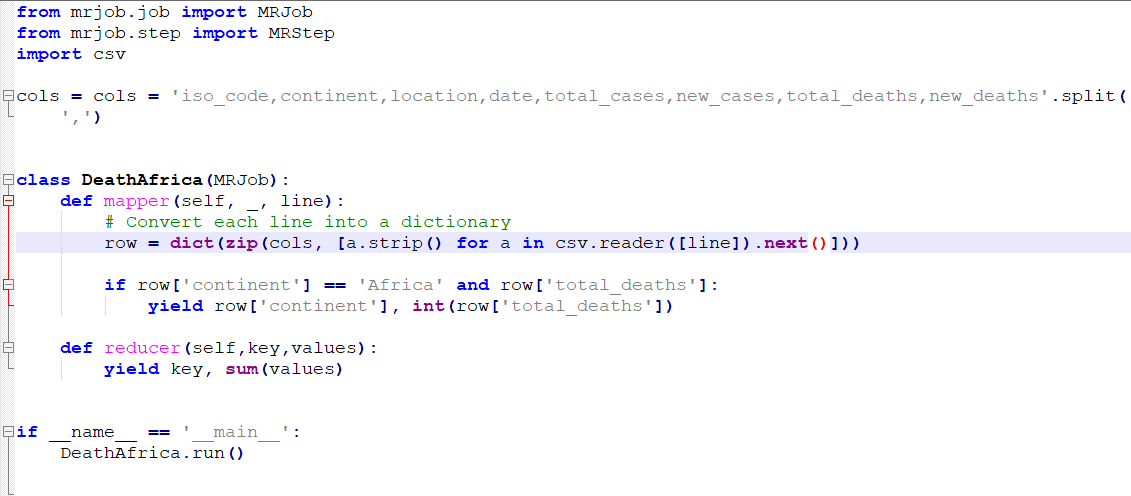
The admin part:

Ambari.

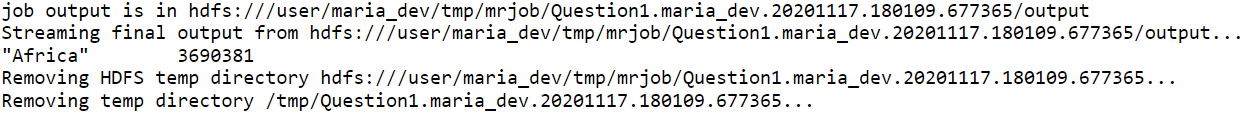
Log-in: root and password.

Install the necessary packages for the running of the program.

# Question 1 (Total number of deaths in Africa)

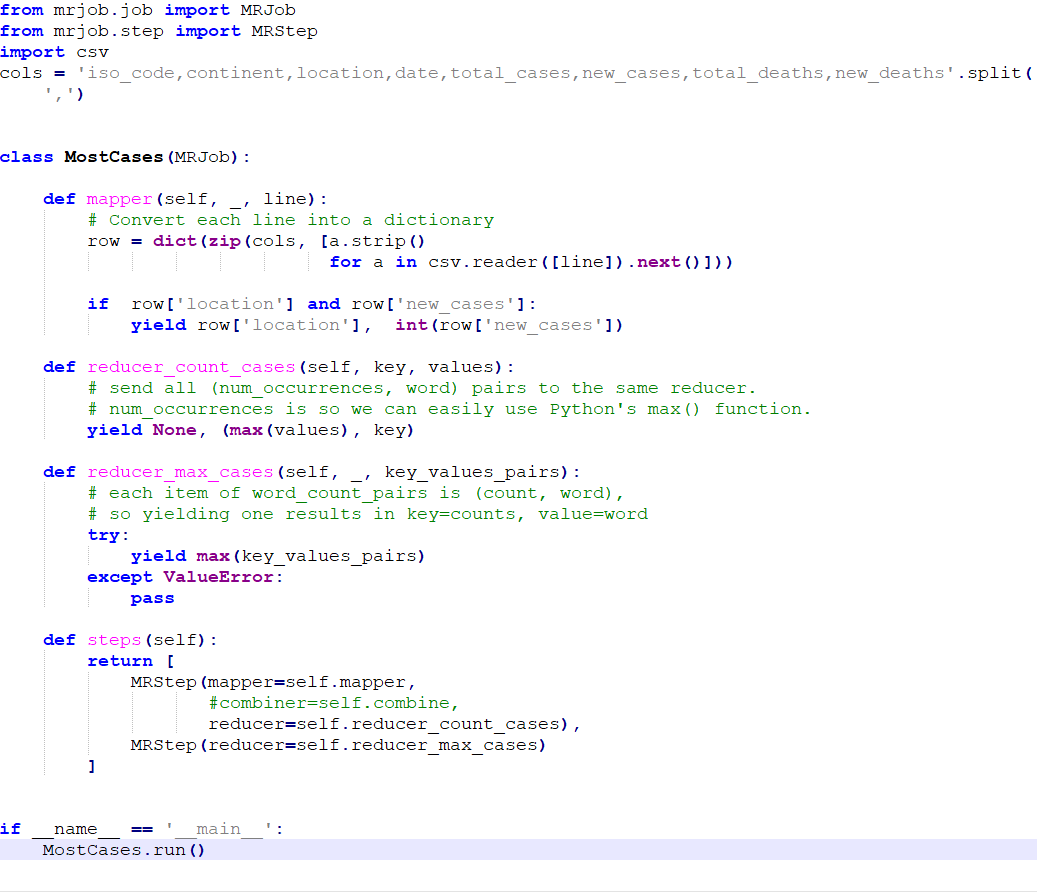


Result:

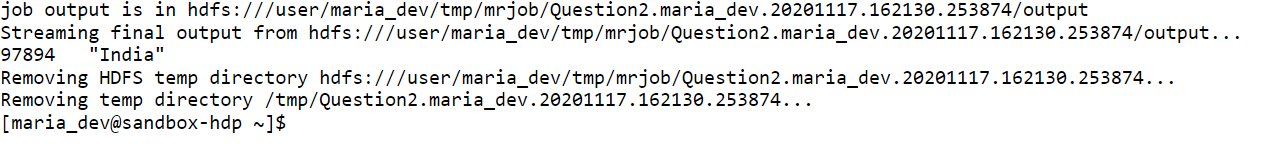


# QUESTION2: (Country with the greatest number of cases as at now)

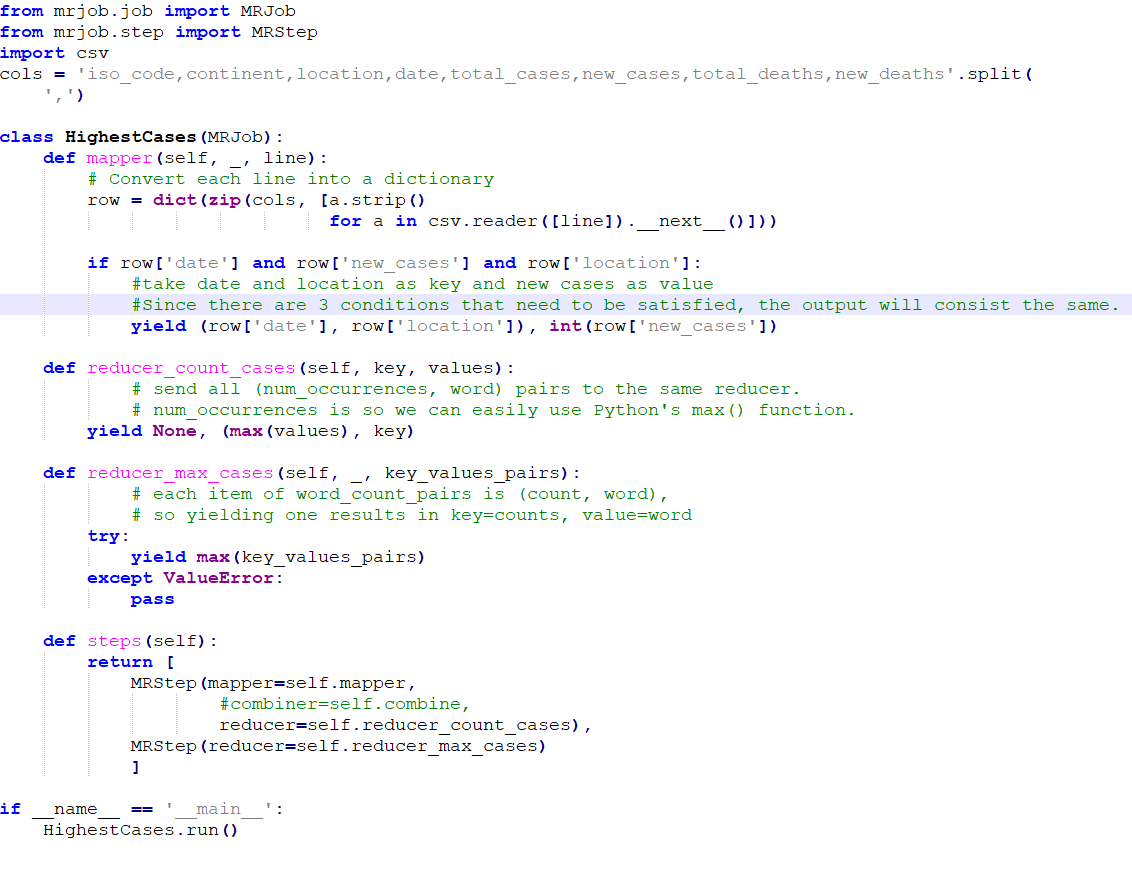
Code: This syntax filters the number of cases using the max () function and outputs the highest and latest number of cases.



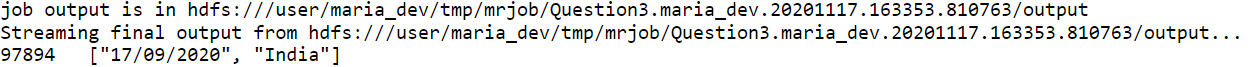
Result:



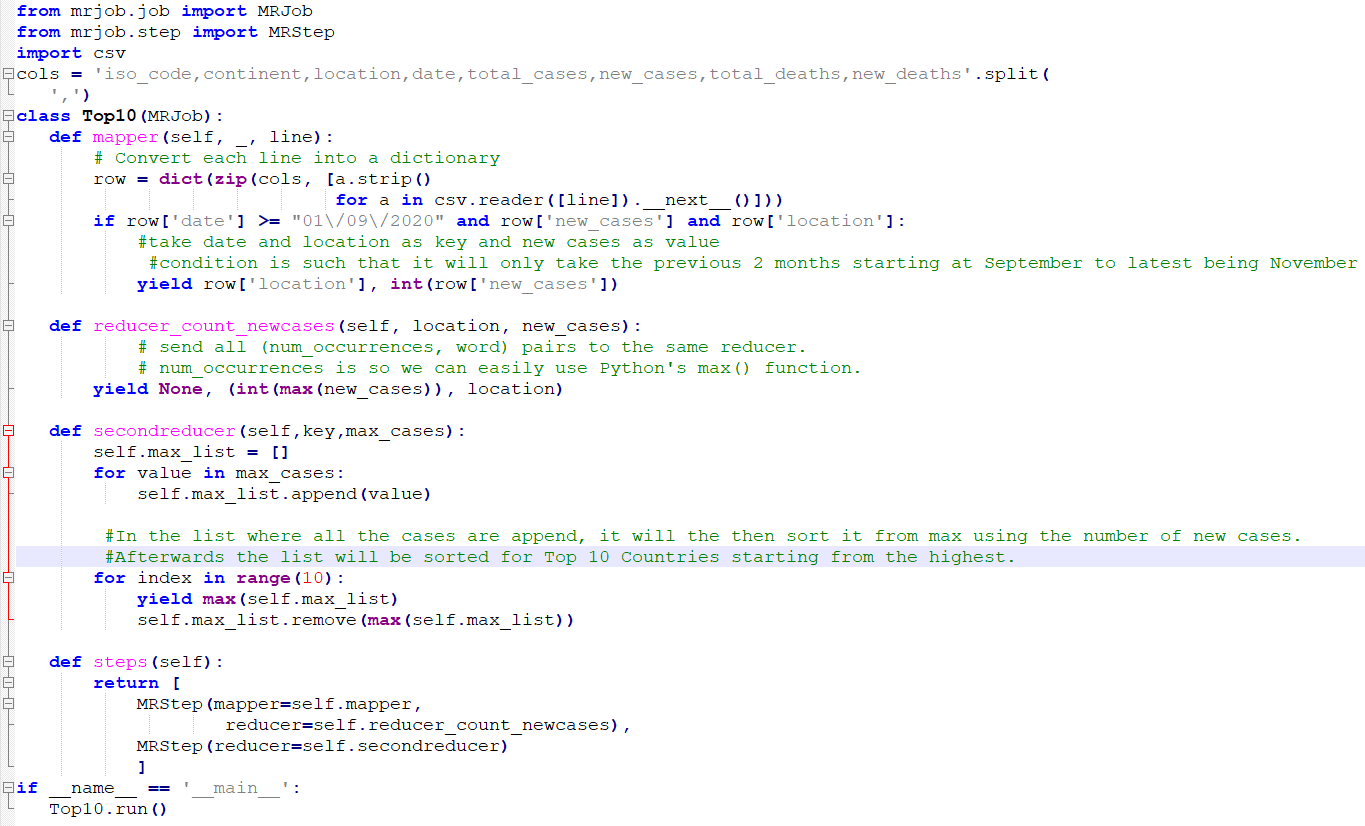
# QUESTION3: (Which date has the highest number of new cases and which country?)



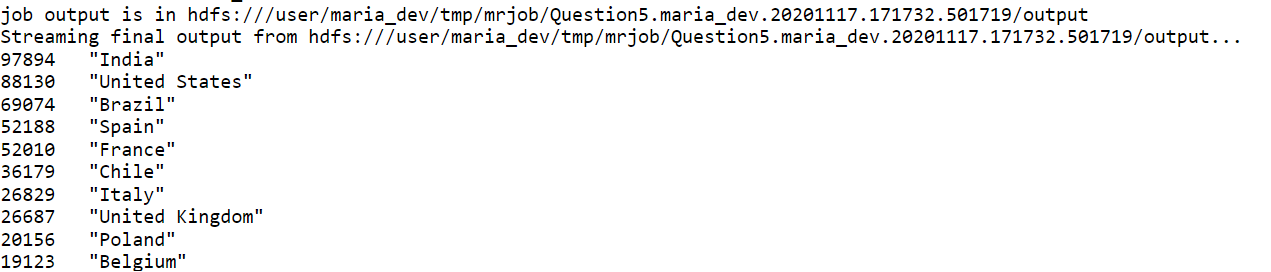
Result:



# QUESTION4: (Top 10 countries in terms of new cases in the last 2 months)



Results:



Explanation: This python file allows the selection of new cases and countries columns from the .csv file and then it introduces an ‘if’ condition that filters the result for the last two months only whilst also sorting the results for the top 10 countries starting from the highest.