Weather data analytics using MapReduce

ANNAPOORNIMA .S
225229101

Inputs and Outputs

- **Input file should be in : /weather/in00/**

  WAData.txt

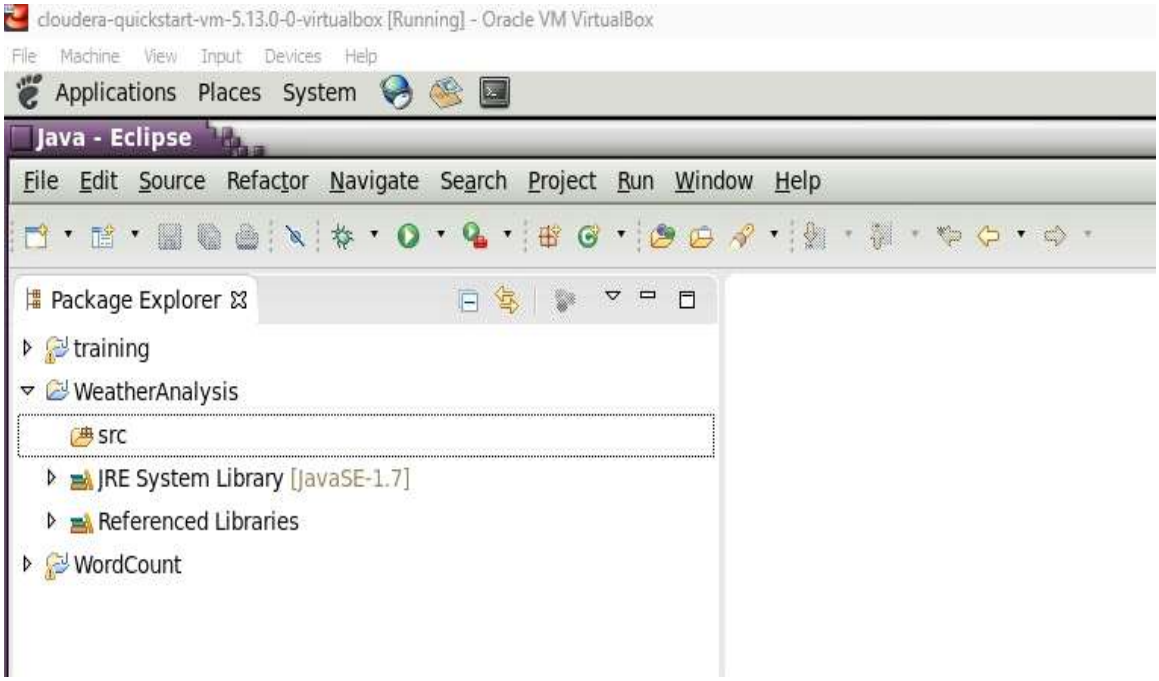
  Copy the content text from sample_weather.txt, which is attached in Googleclassroom.

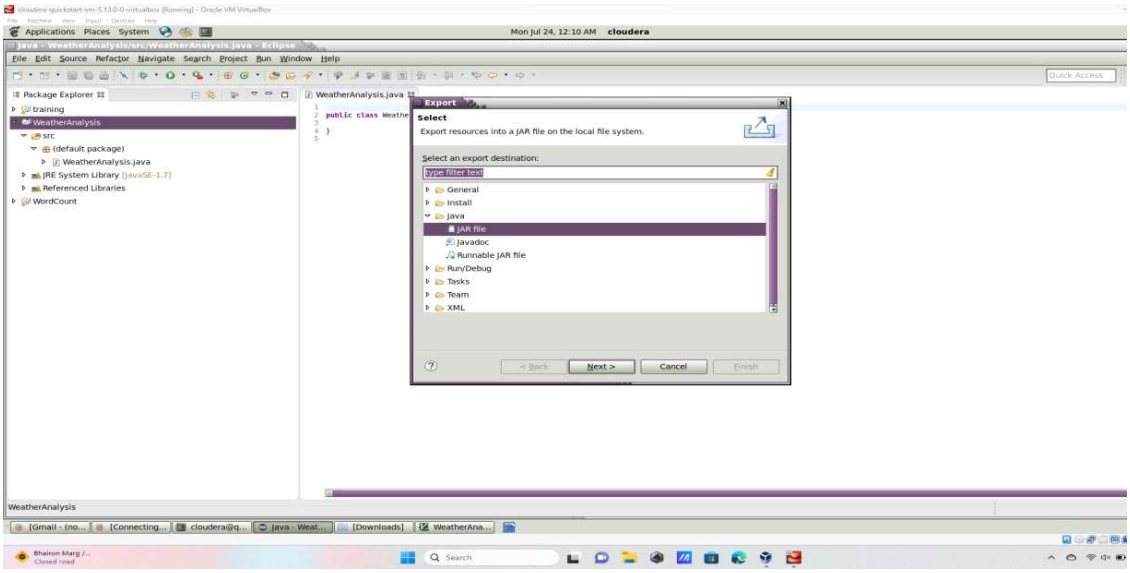
- **Output file should be in /weather/out00/**

Step 1

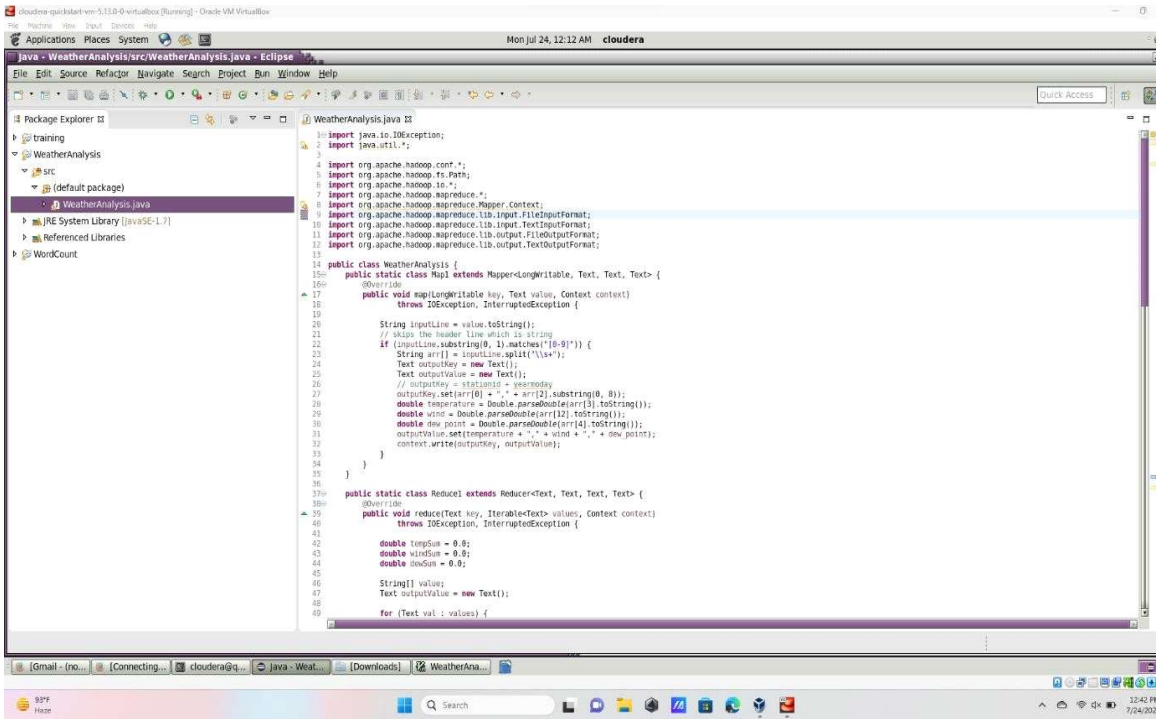
Compile WeatherAnalysis.javaand create a WeatherAnalysis.jar:

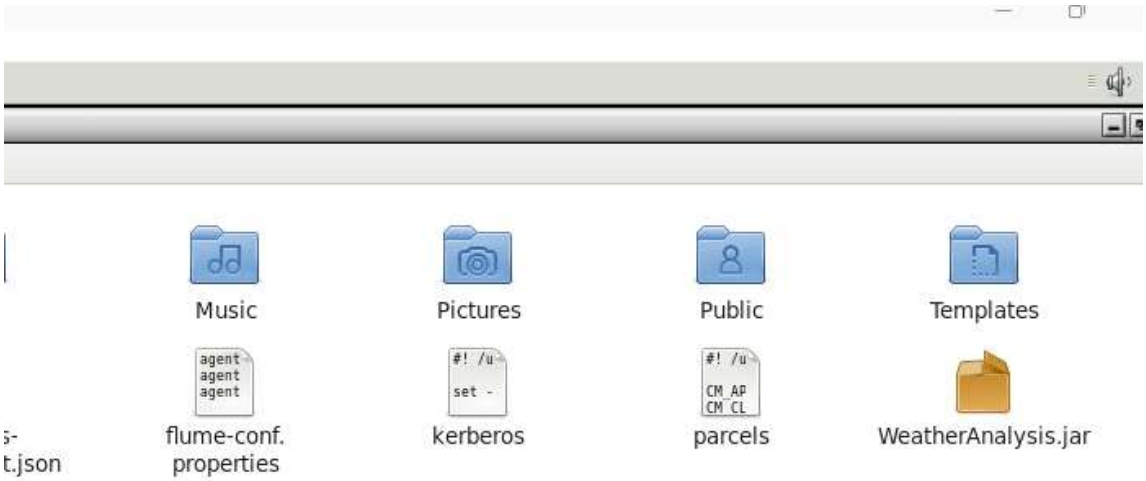
(i) Create WeatherAnalysis.java project.





(iii) Create Weathernalysis class file using Google classroom attached WeatherAnalysis.javafile.

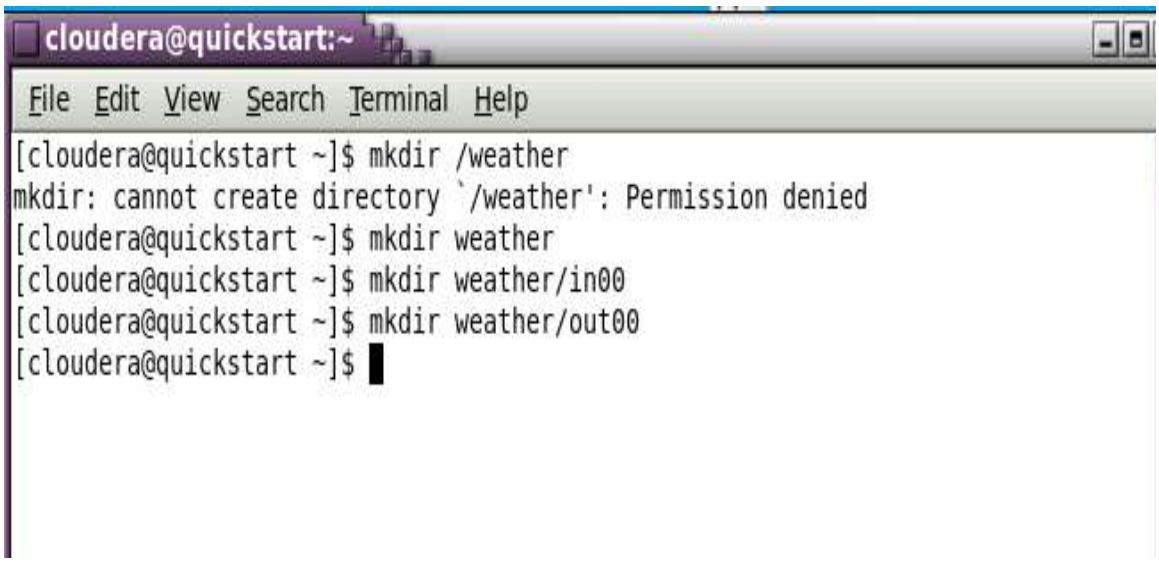




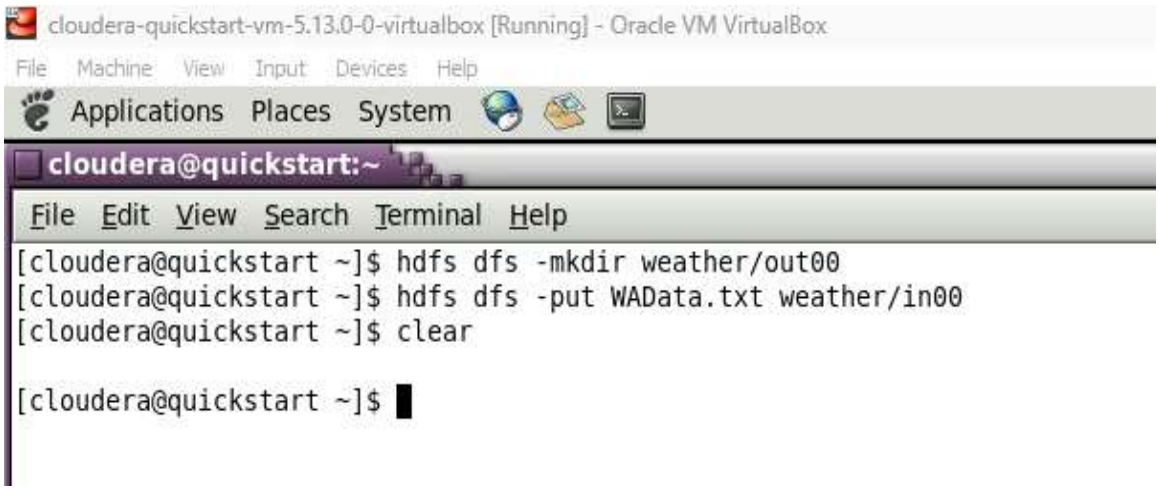
Step 2

Create following folders in HDFS:

- /weather/in00 - input directory in HDFS
- /weather/out00- output directory in HDFS



Create and copy WAData.txt-files into input folder:



[cloudera@quickstart ~]$ hdfs dfs -ls

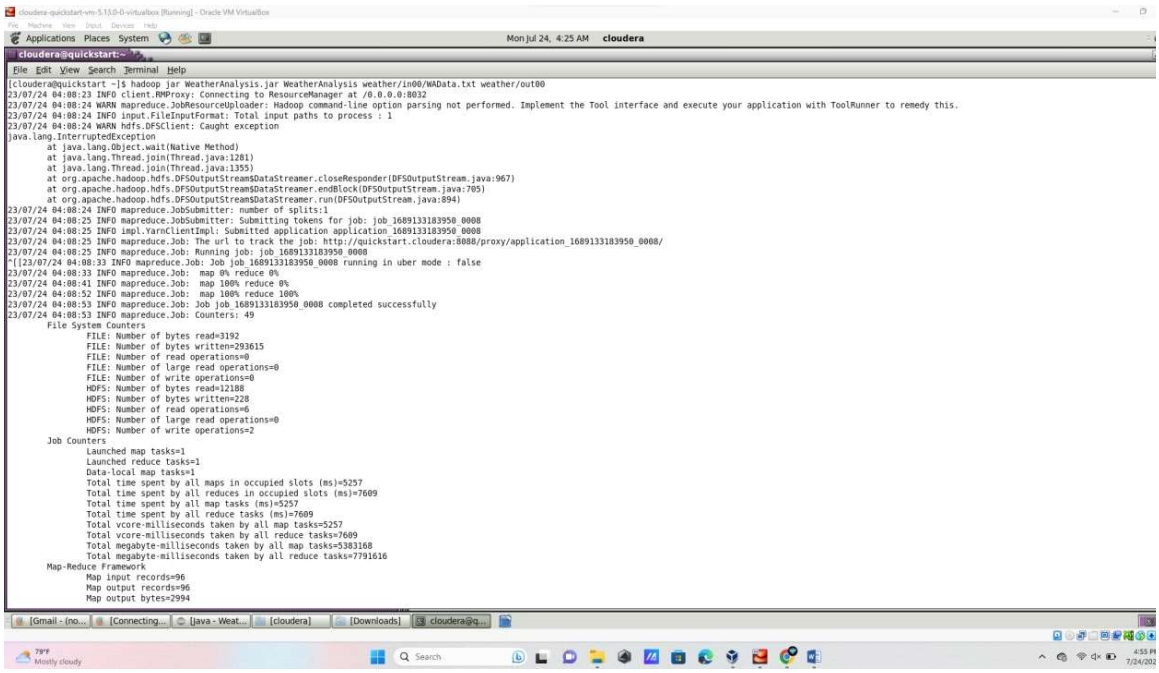
/weather/in00/Found 1 items
-rw-r--r-- 1 cloudera supergroup 12054 2021-08-26 15:48 /weather/in00/WAData.txt

Step 4

Run the MapReduce application :

[cloudera@quickstart ~]$ hadoop jar WeatherAnalysis.jar WeatherAnalysis /weather/in00/WAData.txt /weather/out00/

Show MapReduce Framework



Output:

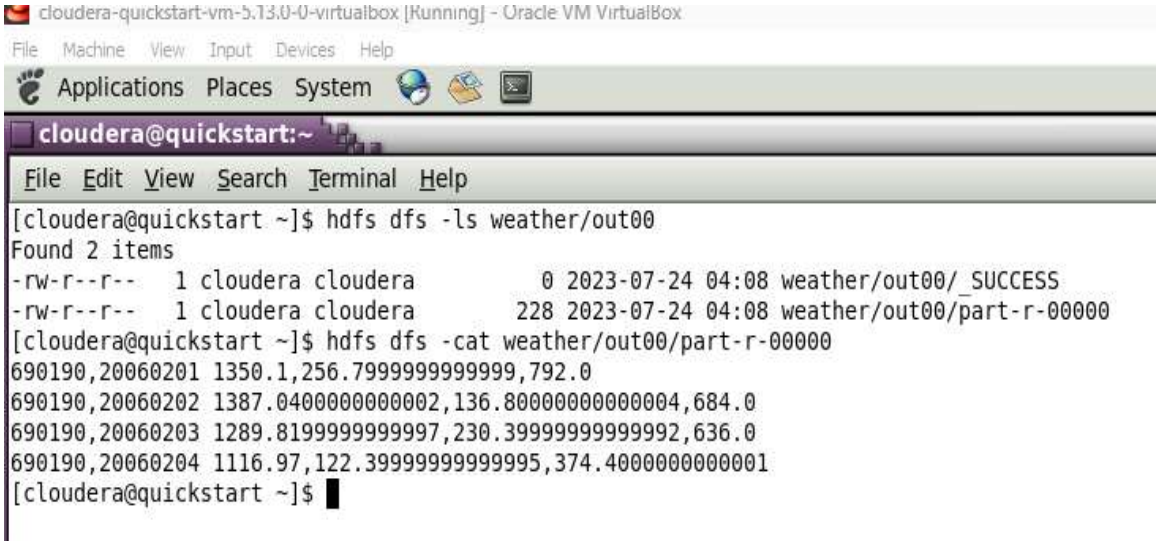
[cloudera@quickstart ~]$ hdfs dfs -ls

/weather/out00/Found 2 items
-rw-r--r-- 1 cloudera supergroup 0 2021-08-26 15:50 /weather/out00/_SUCCESS

-rw-r--r-- 1 cloudera supergroup 228 2021-08-26 15:50

/weather/out00/part-r-00000[cloudera@quickstart ~]$ hdfs dfs -cat

/weather/out00/part-r-00000


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls weather/out00
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2023-07-24 04:08 weather/out00/_SUCCESS
-rw-r--r-- 1 cloudera cloudera 228 2023-07-24 04:08 weather/out00/part-r-00000
[cloudera@quickstart ~]$ hdfs dfs -cat weather/out00/part-r-00000
690190,20060201 1350.1,256.7999999999999,792.0
690190,20060202 1387.0400000000002,136.80000000000004,684.0
690190,20060203 1289.8199999999997,230.39999999999992,636.0
690190,20060204 1116.97,122.39999999999995,374.4000000000001
[cloudera@quickstart ~]$