

Weather data analytics using MapReduce

ANNAPOORNIMA .S
225229101

Inputs and Outputs

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

WADData.txt

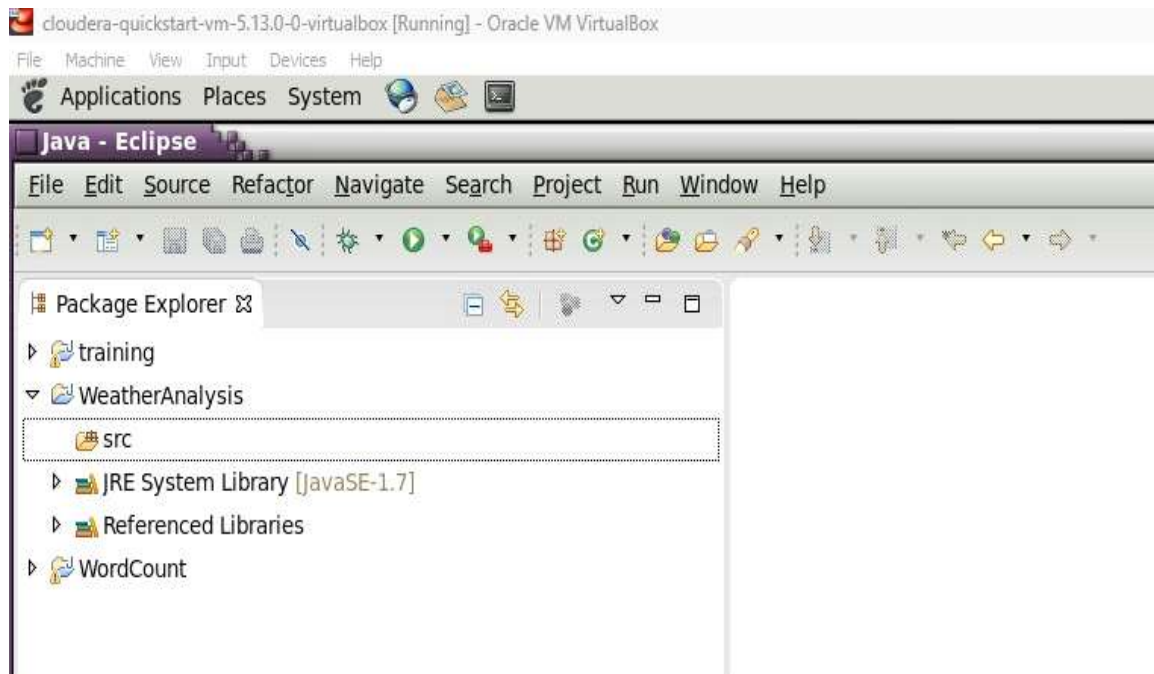
Copy the content text from sample_weather.txt, which is attached in Googleclassroom.

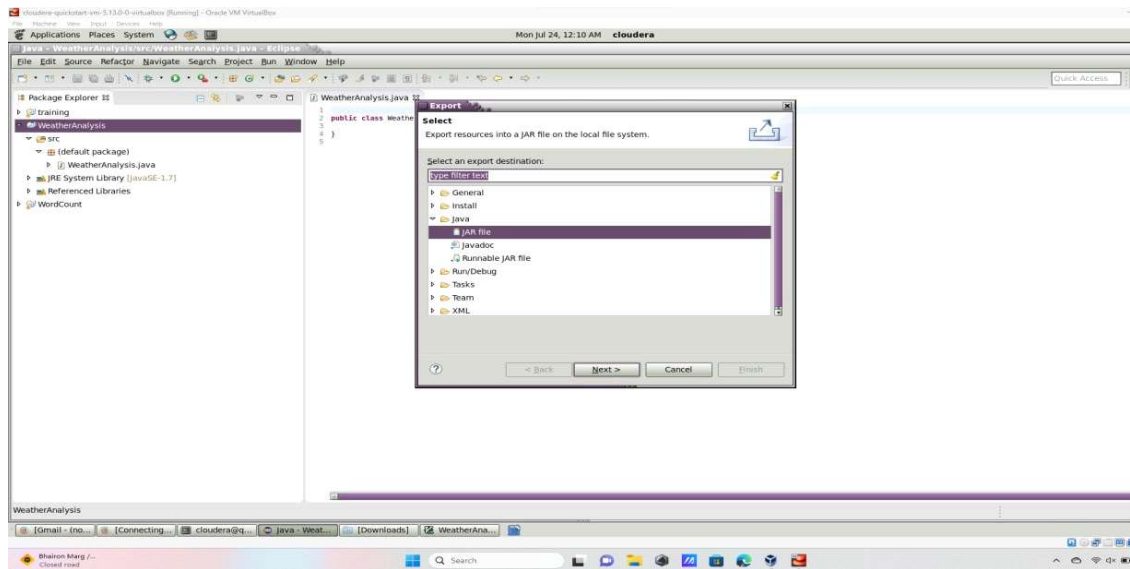
- Output file should be in /weather/out00/

Step 1

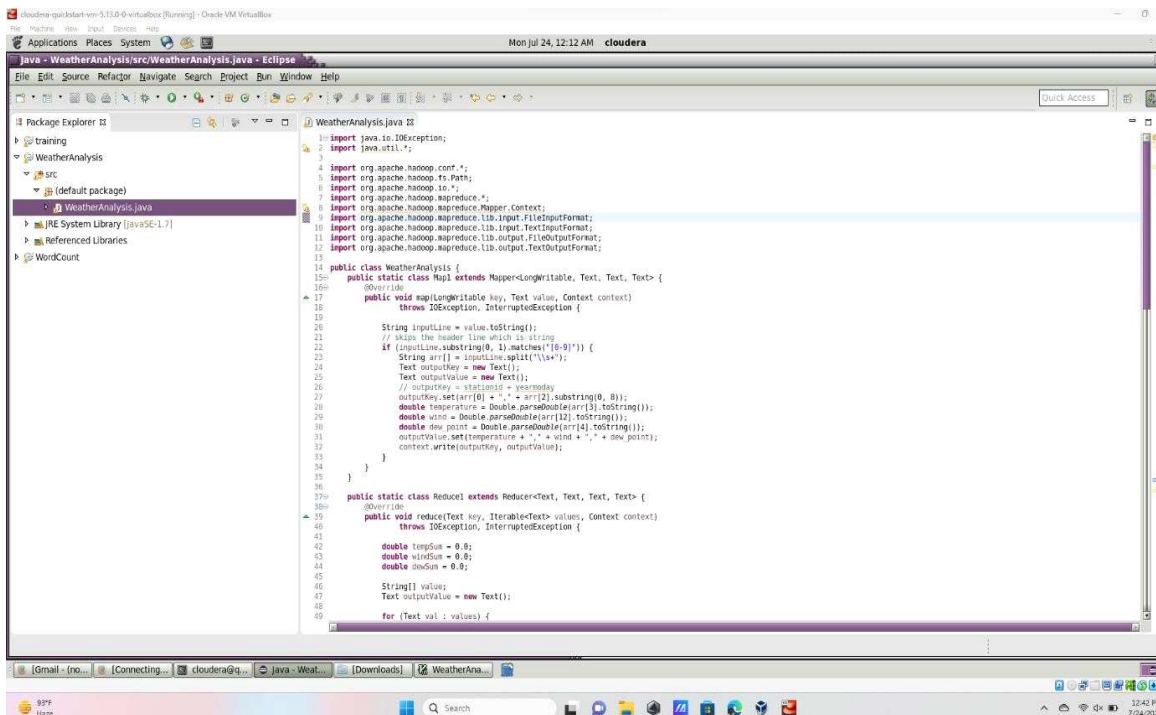
Compile WeatherAnalysis.java and create a WeatherAnalysis.jar:

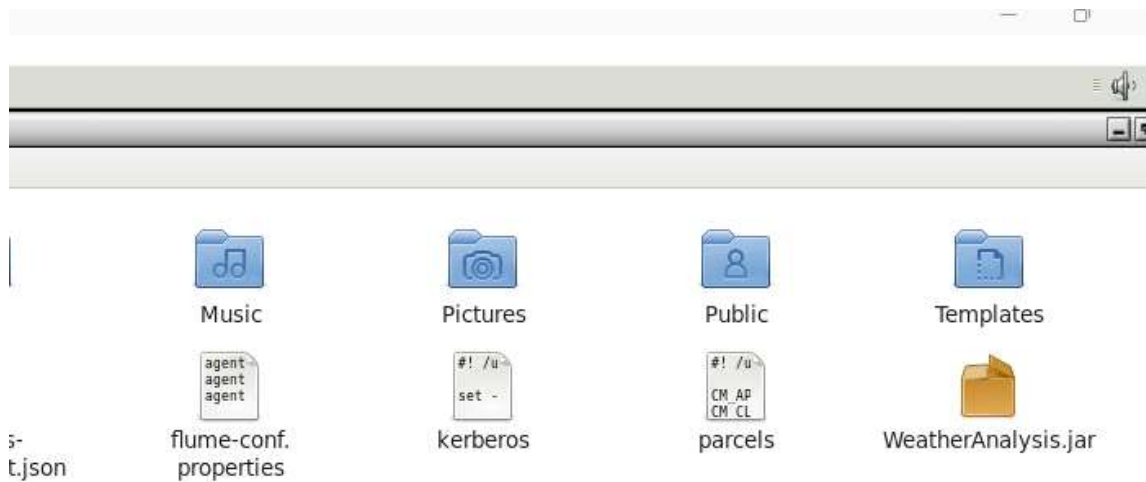
- (i) Create WeatherAnalysis.java project.





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





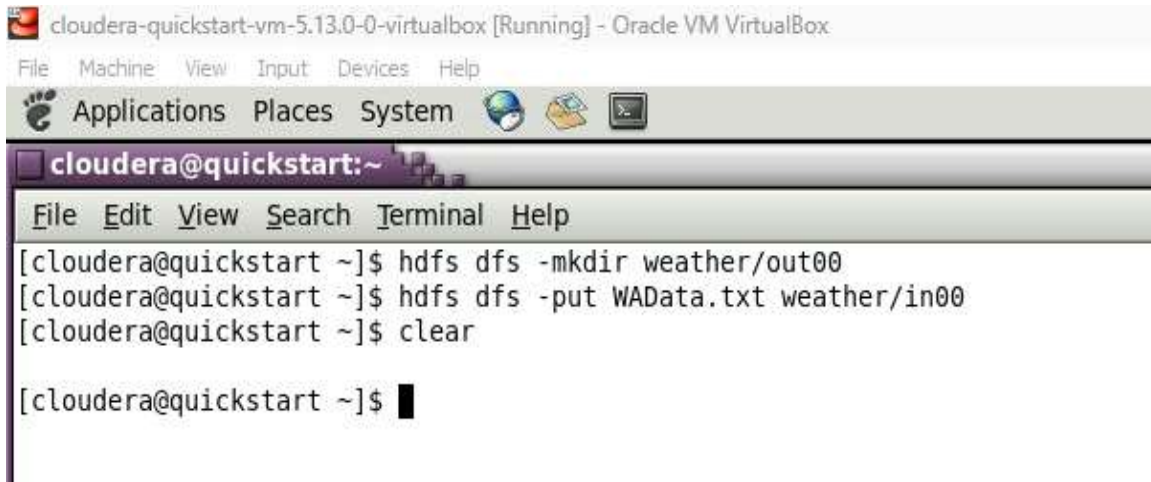
Step 2

Create following folders in HDFS:

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

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ mkdir /weather  
mkdir: cannot create directory `/weather': Permission denied  
[cloudera@quickstart ~]$ mkdir weather  
[cloudera@quickstart ~]$ mkdir weather/in00  
[cloudera@quickstart ~]$ mkdir weather/out00  
[cloudera@quickstart ~]$
```

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



```
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 -mkdir weather/out00
[cloudera@quickstart ~]$ hdfs dfs -put WADData.txt weather/in00
[cloudera@quickstart ~]$ clear
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$ hdfs dfs -ls
/weather/in00/Found 1 items
-rw-r--r-- 1 cloudera supergroup 12054 2021-08-26 15:48 /weather/in00/WADData.txt
```

Step 4

Run the MapReduce application :

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

Show MapReduce Framework

```

[cloudera@quickstart ~]$ hadoop jar WeatherAnalysis.jar WeatherAnalysis weather/in00/wData.txt weather/out00
23/07/24 04:08:23 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/07/24 04:08:24 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
23/07/24 04:08:24 INFO input.FileInputFormat: Total input paths to process : 1
23/07/24 04:08:24 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStreamDataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStreamDataStreamer.endBlock(DFSOutputStream.java:705)
    at org.apache.hadoop.hdfs.DFSOutputStreamDataStreamer.run(DFSOutputStream.java:894)
23/07/24 04:08:24 INFO mapreduce.JobSubmitter: number of splits:1
23/07/24 04:08:25 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1689133183950_0008
23/07/24 04:08:25 INFO impl.YarnClientImpl: Submitted application application_1689133183950_0008
23/07/24 04:08:25 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8080/proxy/application_1689133183950_0008/
23/07/24 04:08:25 INFO mapreduce.Job: Running job: job_1689133183950_0008
11/23/07/24 04:08:33 INFO mapreduce.Job: Job job_1689133183950_0008 running in uber mode : false
23/07/24 04:08:33 INFO mapreduce.Job: map 0% reduce 0%
23/07/24 04:08:41 INFO mapreduce.Job: map 100% reduce 0%
23/07/24 04:08:52 INFO mapreduce.Job: map 100% reduce 100%
23/07/24 04:08:53 INFO mapreduce.Job: Job job_1689133183950_0008 completed successfully
23/07/24 04:08:53 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=3192
        FILE: Number of bytes written=293615
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=12108
        HDFS: Number of bytes written=228
        HDFS: Number of read operations=6
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=5257
        Total time spent by all reduces in occupied slots (ms)=7609
        Total time spent by all map tasks (ms)=5257
        Total time spent by all reduce tasks (ms)=7609
        Total vcore-milliseconds taken by all map tasks=5257
        Total vcore-milliseconds taken by all reduce tasks=7609
        Total megabyte-milliseconds taken by all map tasks=5393168
        Total megabyte-milliseconds taken by all reduce tasks=7791616
    Map-Reduce Framework
        Map input records=96
        Map output records=96
        Map output bytes=2994
  
```

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 ~]$
```