

Exercise 03a: Weather data analytics using MapReduce

This exercise's MapReduce process is doing Weather analysis

Prerequisites

Ensure that Hadoop is installed, configured and is running. More details:

Single Node Setup for first-time users.

Cluster Setup for large, distributed clusters.

Inputs and Outputs

i. Input file should be in : **/weather/in00/**

WAData.txt

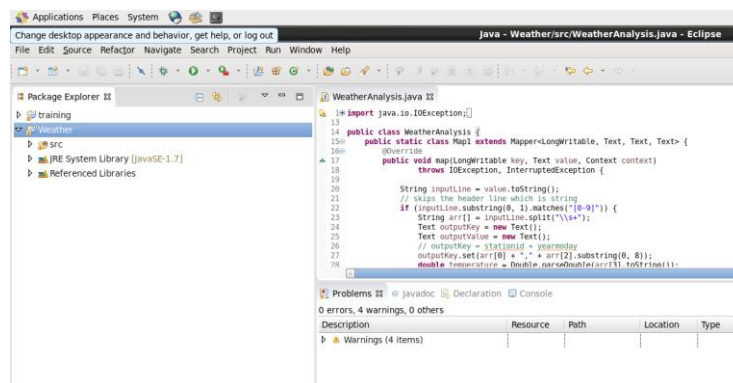
Copy the content text from sample_weather.txt, Which is attached in Google classroom.

ii. Output file should be in **/weather/out00/**

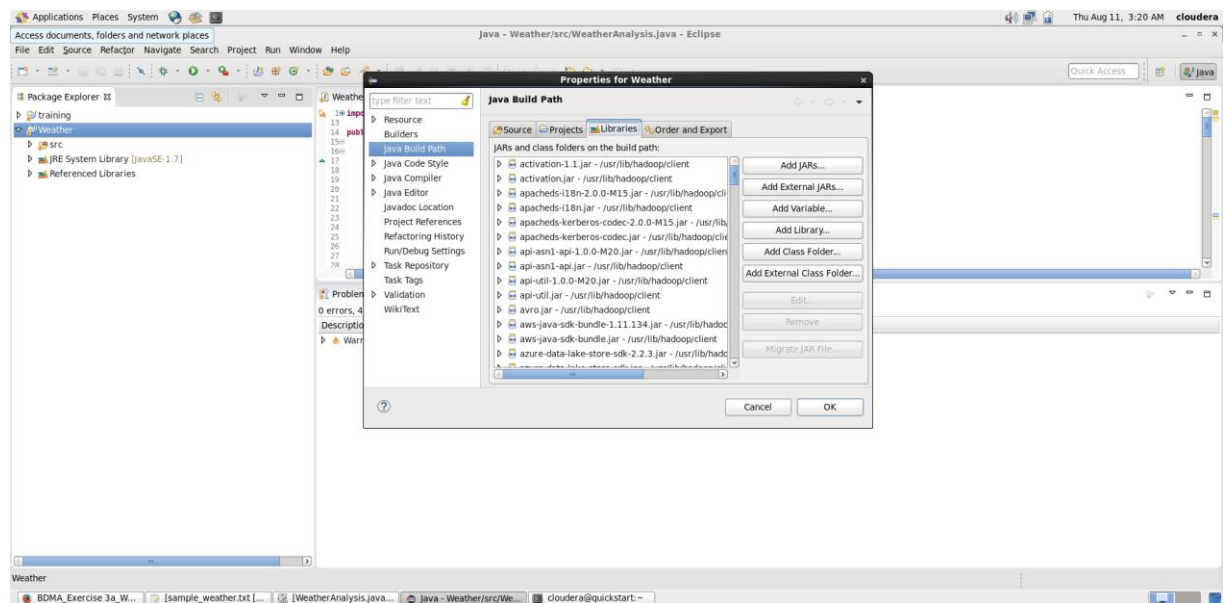
Step 1

Compile `WeatherAnalysis.java` and create a `WeatherAnalysis.jar`:

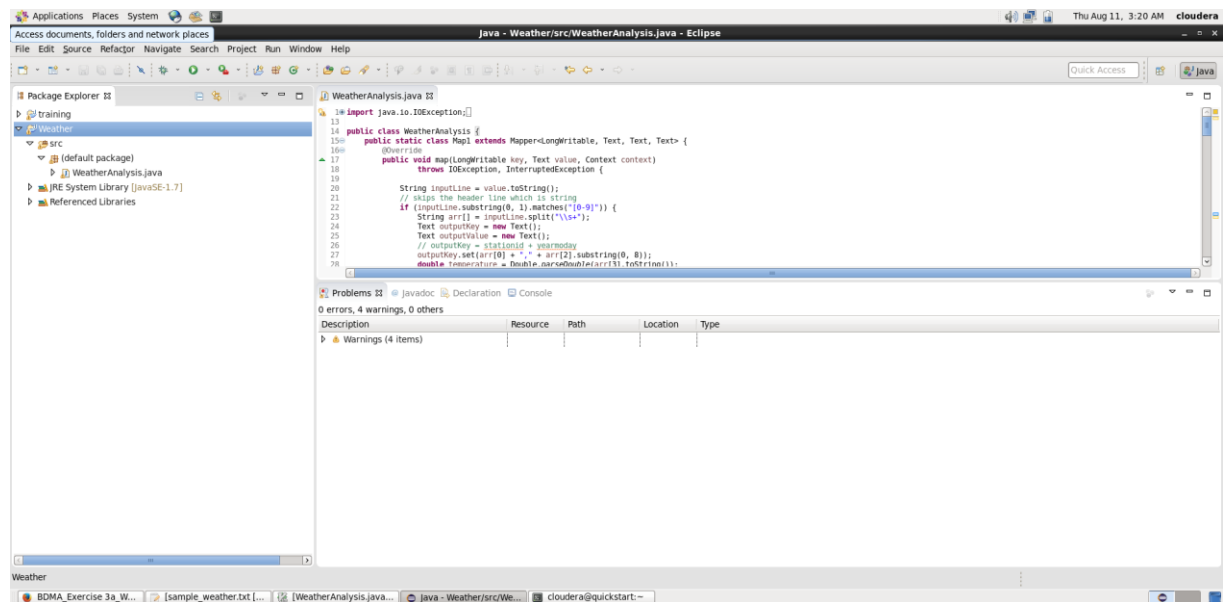
(i) Create `WeatherAnalysis.java` project.



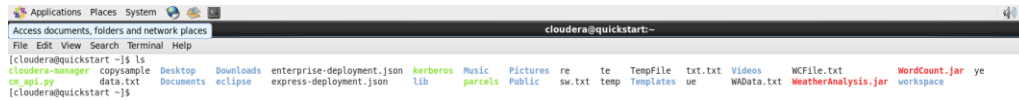
(ii) Import external .jar files



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



(iv) Create WeatherAnalysis.jar file



Step 2

Create following folders in HDFS:

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

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-08-11 02:40 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-07-28 08:53 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
drwxr-xr-x - cloudera supergroup 0 2022-08-11 03:14 /weather
[cloudera@quickstart ~]$ hdfs dfs -ls /weather
Found 2 items
drwxr-xr-x - cloudera supergroup 0 2022-08-11 03:22 /weather/in00
drwxr-xr-x - cloudera supergroup 0 2022-08-11 03:14 /weather/out00
[cloudera@quickstart ~]$ hdfs dfs -ls /weather/in00
[cloudera@quickstart ~]$ hdfs dfs -put WData.txt /weather/in00
[cloudera@quickstart ~]$ hdfs dfs -ls /weather/in00
Found 1 items
-rw-r--r-- 1 cloudera supergroup 12054 2022-08-11 03:23 /weather/in00/WAD
ata.txt
[cloudera@quickstart ~]$ █
```

Step 3

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

```
[cloudera@quickstart ~]$ hdfs dfs -put WData.txt /weather/in00
[cloudera@quickstart ~]$ hdfs dfs -ls /weather/in00
Found 1 items
-rw-r--r-- 1 cloudera supergroup 12054 2022-08-11 03:23 /weather/in00/WAD
ata.txt
[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/WData.txt
```

Step 4

Run the MapReduce application :

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

Show MapReduce Framework

```
Access documents, folders and network places cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop jar WeatherAnalysis.jar WeatherAnalysis /weather/in00/WAData.txt /weather/out01/  
22/08/11 03:26:36 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
22/08/11 03:26:38 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
22/08/11 03:26:39 INFO input.FileInputFormat: Total input paths to process : 1  
22/08/11 03:26:39 INFO mapreduce.JobSubmitter: number of splits:1  
22/08/11 03:26:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1660210785745_0001  
22/08/11 03:26:41 INFO impl.YarnClientImpl: Submitted application application_1660210785745_0001  
22/08/11 03:26:42 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8080/proxy/application_1660210785745_0001/  
22/08/11 03:26:42 INFO mapreduce.Job: Running job: job_1660210785745_0001  
22/08/11 03:27:07 INFO mapreduce.Job: Job job_1660210785745_0001 running in uber mode : false  
22/08/11 03:27:07 INFO mapreduce.Job: map 0% reduce 0%  
22/08/11 03:27:22 INFO mapreduce.Job: map 100% reduce 0%  
22/08/11 03:27:32 INFO mapreduce.Job: map 100% reduce 100%  
22/08/11 03:27:32 INFO mapreduce.Job: Job job_1660210785745_0001 completed successfully  
22/08/11 03:27:32 INFO mapreduce.Job: Counters: 49  
File System Counters  
FILE: Number of bytes read=3192  
FILE: Number of bytes written=293559  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=12174  
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)=12356  
Total time spent by all reduces in occupied slots (ms)=6090  
Total time spent by all map tasks (ms)=12356  
Total time spent by all reduce tasks (ms)=6090  
Total vcore-milliseconds taken by all map tasks=12356  
Total vcore-milliseconds taken by all reduce tasks=6090  
Total megabyte-milliseconds taken by all map tasks=12652544  
Total megabyte-milliseconds taken by all reduce tasks=6236160  
Map-Reduce Framework  
Map input records=96  
Map output records=96  
Map output bytes=2994  
Map output materialized bytes=3192  
Input split bytes=120  
Combine input records=0  
Combine output records=0  
Reduce input groups=4  
Reduce shuffle bytes=3192  
Reduce input records=96
```

Step 5

Output:

```
[cloudera@quickstart ~]$ hdfs dfs -ls /weather/out01/  
Found 2 items  
-rw-r--r-- 1 cloudera supergroup 0 2022-08-11 03:27 /weather/out01/_S  
UCCESS  
-rw-r--r-- 1 cloudera supergroup 228 2022-08-11 03:27 /weather/out01/pa  
rt-r-00000  
[cloudera@quickstart ~]$ hdfs dfs -cat /weather/out01/part-r-00000  
690190,20060201 1350.1,256.7999999999999,792.0  
690190,20060202 1387.0400000000000,136.8000000000000,684.0  
690190,20060203 1289.8199999999999,230.3999999999999,636.0  
690190,20060204 1116.97,122.3999999999999,374.4000000000001  
[cloudera@quickstart ~]$
```

```
BDMA_Exercise 3a_W... [sample_weather.txt [... [WeatherAnalysis.java... [Jav
```

```
[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
```