**Practical 4**

**Aim: Whether Analysis Application using Java and MapReduce**

**What is MapReduce ?**

A MapReduce is a data processing tool which is used to process the data parallely in a

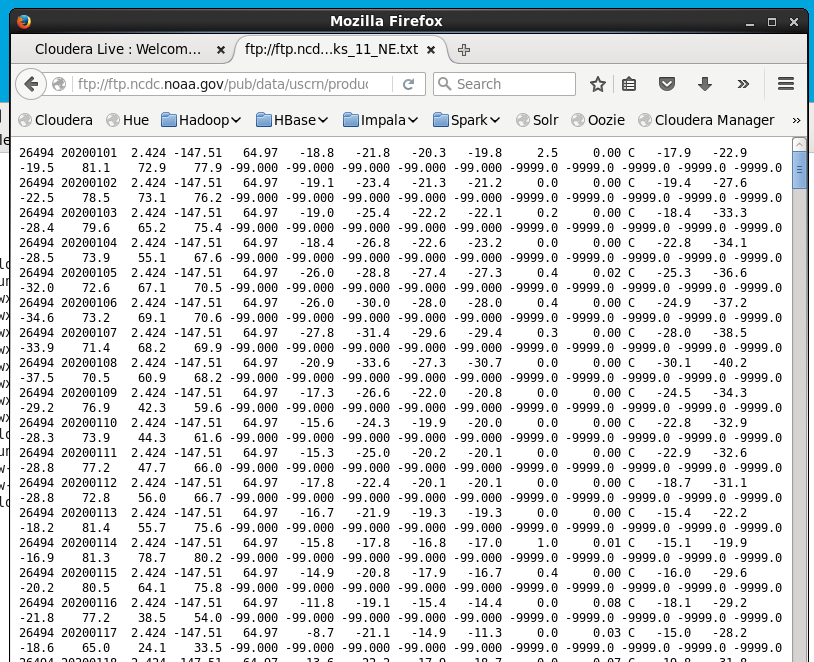
distributed form. It was developed in 2004, on the basis of paper titled as "MapReduce:

Simplified Data Processing on Large Clusters," published by Google.

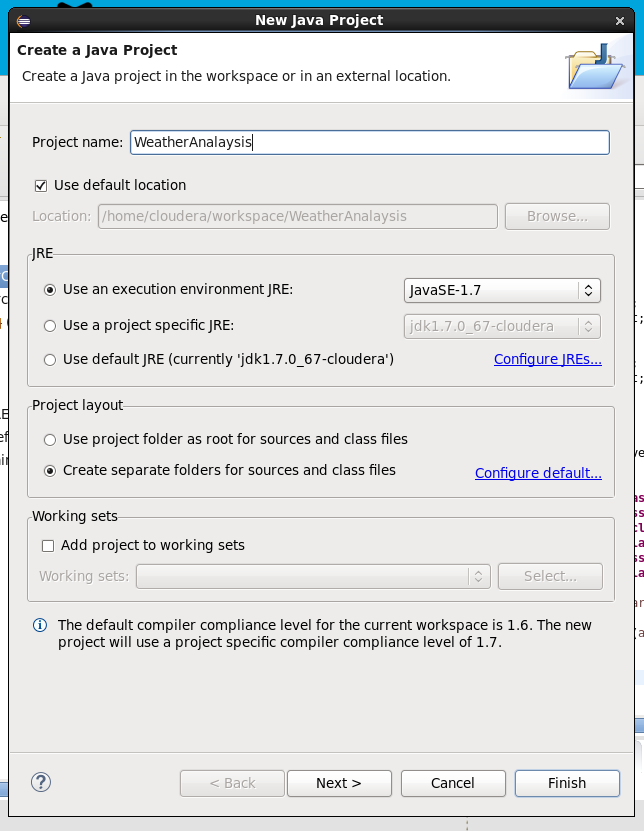
**Steps:**

1. Open browser >

Go to ftp://ftp.ncdc.noaa.gov/pub/data/uscrn/products/daily01/ and download the dataset file and save it

****

1. File -> New -> Java Project -> Name: WeatherAnalaysis

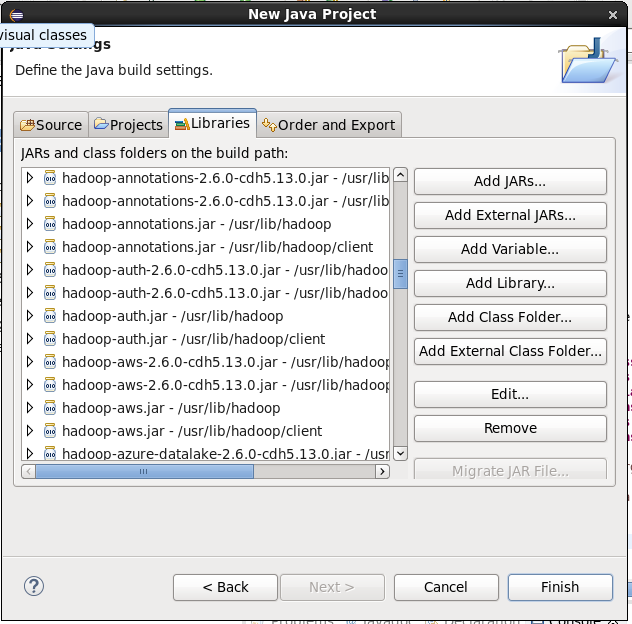
****

1. Add Hadoop Libraries to the Project

JAR Files -> usr/lib/hadoop/

JAR Files -> usr/lib/hadoop/client/

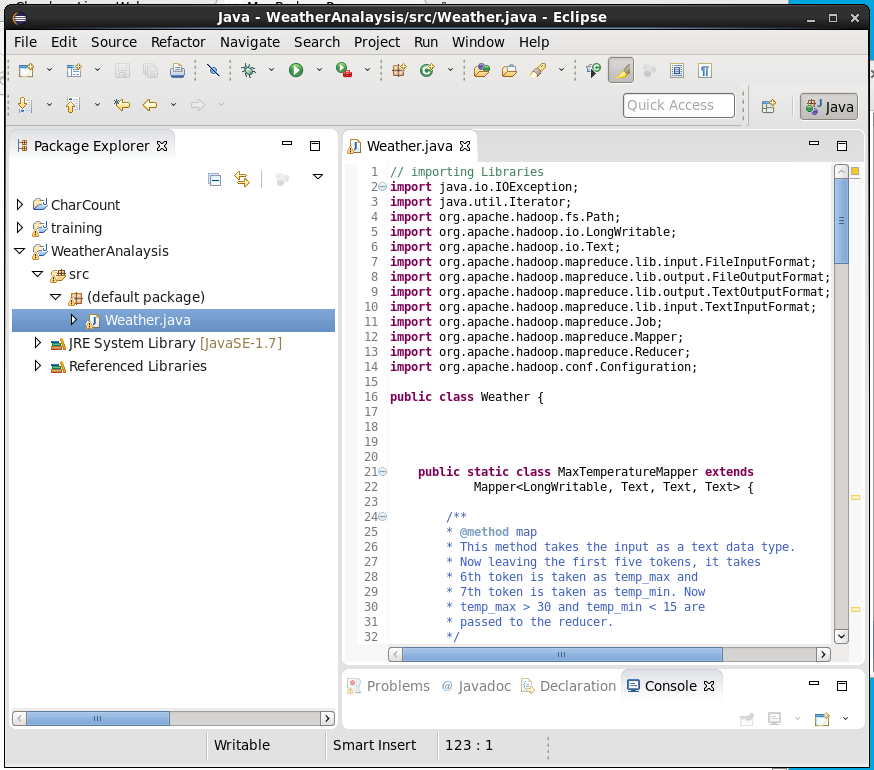
Click Finish

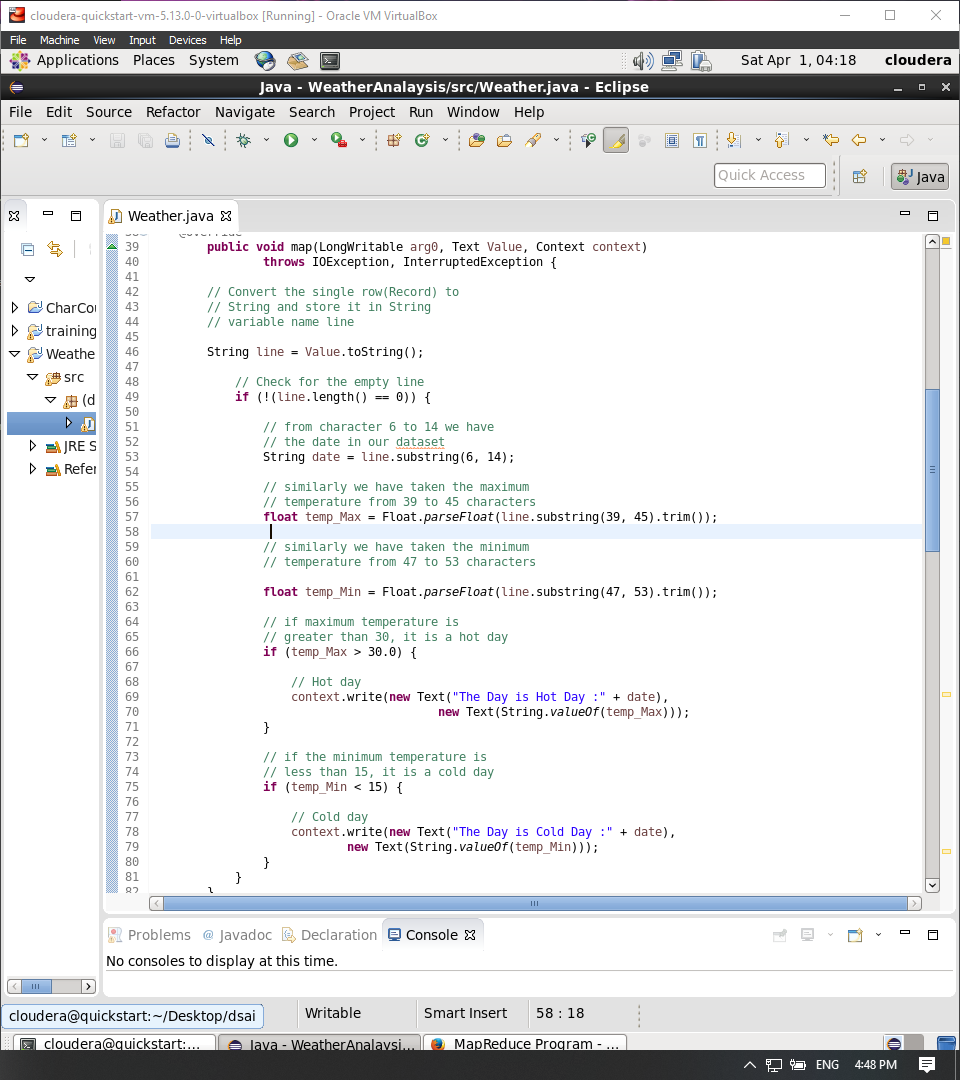
****

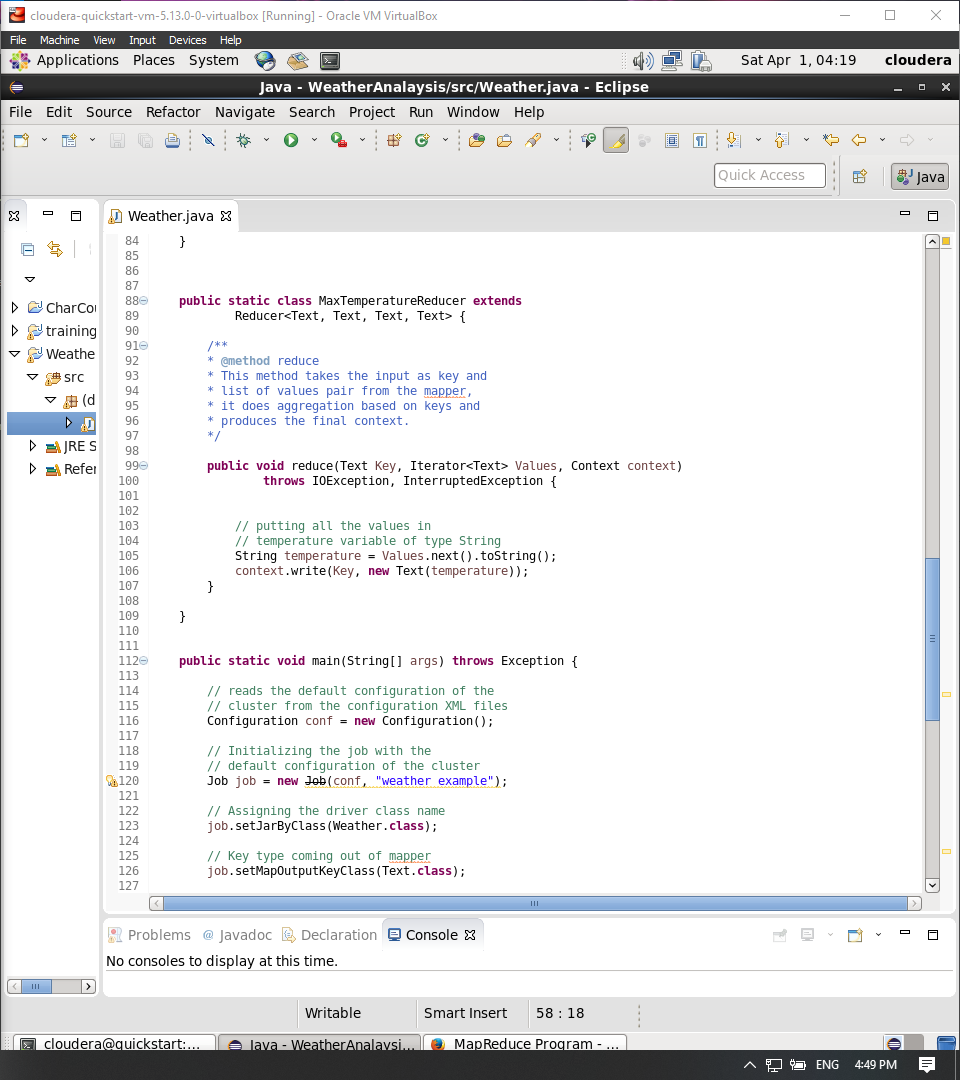
1. Create Classes: Right Click on Project -> New -> Class

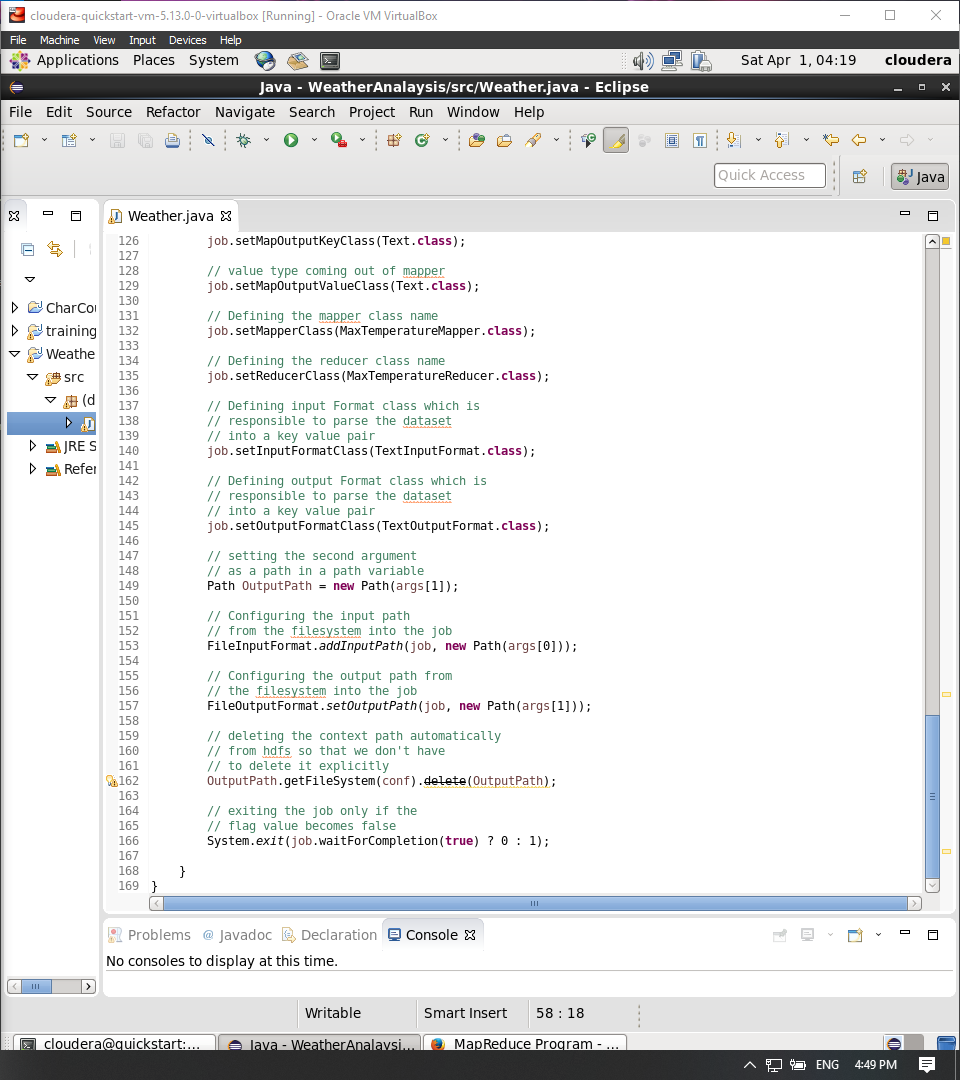
Weather.java

And write the code for it

****

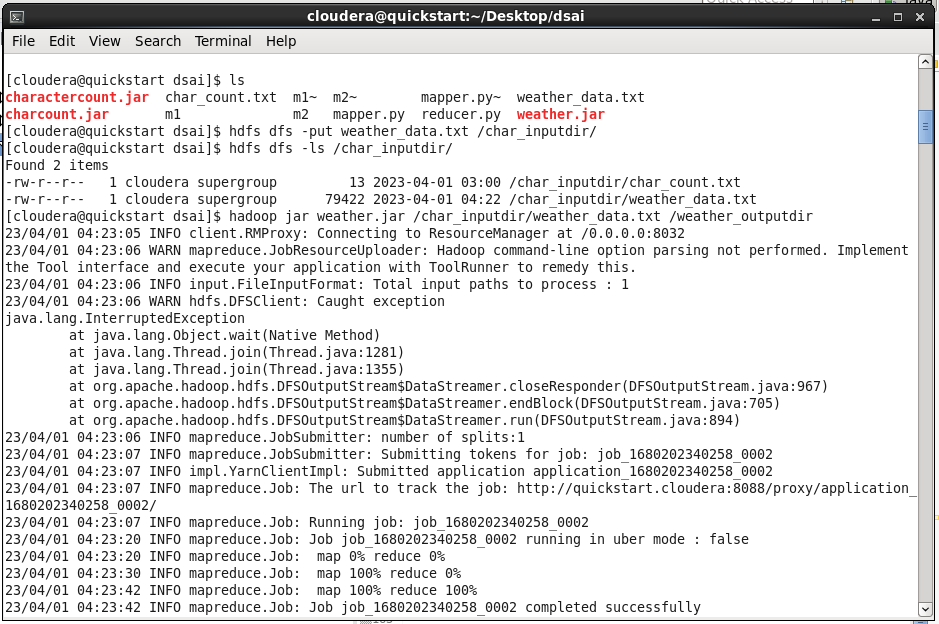
****

****

****

1. Write following commands in terminal

* ls
* hdfs dfs -put weather\_data.txt /char\_inputdir/
* hdfs dfs -ls /char\_inputdir/
* hadoop jar weather.jar /char\_inputdir/weather\_data.txt /weather\_outputdir



* hdfs dfs -ls /
* hdfs dfs -ls /weather\_outputdir
* hdfs dfs -cat /weather\_outputdir/part-r-00000

