EXPERIMENT -6

MAPREDUCEPROGRAM2

OBJECTIVE:

Write a Map Reduce programthat mines weather data. Hint: Weather sensors collecting data every hourat many locations across the globe gather a large volume of log data, which is a good candidate foranalysis with Map Reduce, since it is semi structured and record- oriented.

RESOURCES:

VMWare, Webbrowser, 4GBRAM,HardDisk80GB.

PROGRAMLOGIC:

WordCount is a simple program which counts the number of occurrences of each word in a given text input data set. WordCount fits very well with the MapReduce programmingmodelmaking it agreat exampleto understandtheHadoopMap/Reduceprogrammingstyle. Our implementation consists of three main parts:

1. Mapper
2. Reducer
3. Mainprogram

Step-1.WriteaMapper

AMapperoverridesthe—mapfunctionfromtheClass"org.apache.hadoop.mapreduce.Mapper" which provides <key, value> pairs as the input. AMapper implementation may output <key,value>pairsusingtheprovidedContext.

Input valueofthe WordCountMaptaskwillbe a lineoftext fromthe input datafileandthekey wouldbethe linenumber <line_number, line_of_text>. Maptaskoutputs<word,one>for each word in the line of text.

Pseudo-code

```
voidMap(key,value){
```

```
foreachmax_tempxinvalue:
output.collect(x, 1);

}
voidMap(key,value){
    foreachmin_tempxinvalue:

output.collect(x,1);
}
```

Step-2 Write aReducer

AReducercollectstheintermediate<key,value>outputfrommultiple maptasksandassemblea singleresult.Here,the WordCountprogramwillsumuptheoccurrenceofeachwordtopairs as

<word, occurrence>.

Pseudo-code

```
voidReduce(max_temp,<listofvalue>){ for
each x in <list of value>:

sum+=x;
final_output.collect(max_temp,sum);

}
voidReduce(min_temp,<listofvalue>){ for
each x in <list of value>:

sum+=x;
final_output.collect(min_temp,sum);

}
```

3. WriteDriver

TheDriverprogramconfiguresandruntheMapReduce job.Weusethe mainprogramto perform basic configurations such as:

JobName : name ofthisJobExecutable (Jar)

Class:themainexecutableclass.Forhere,WordCount.

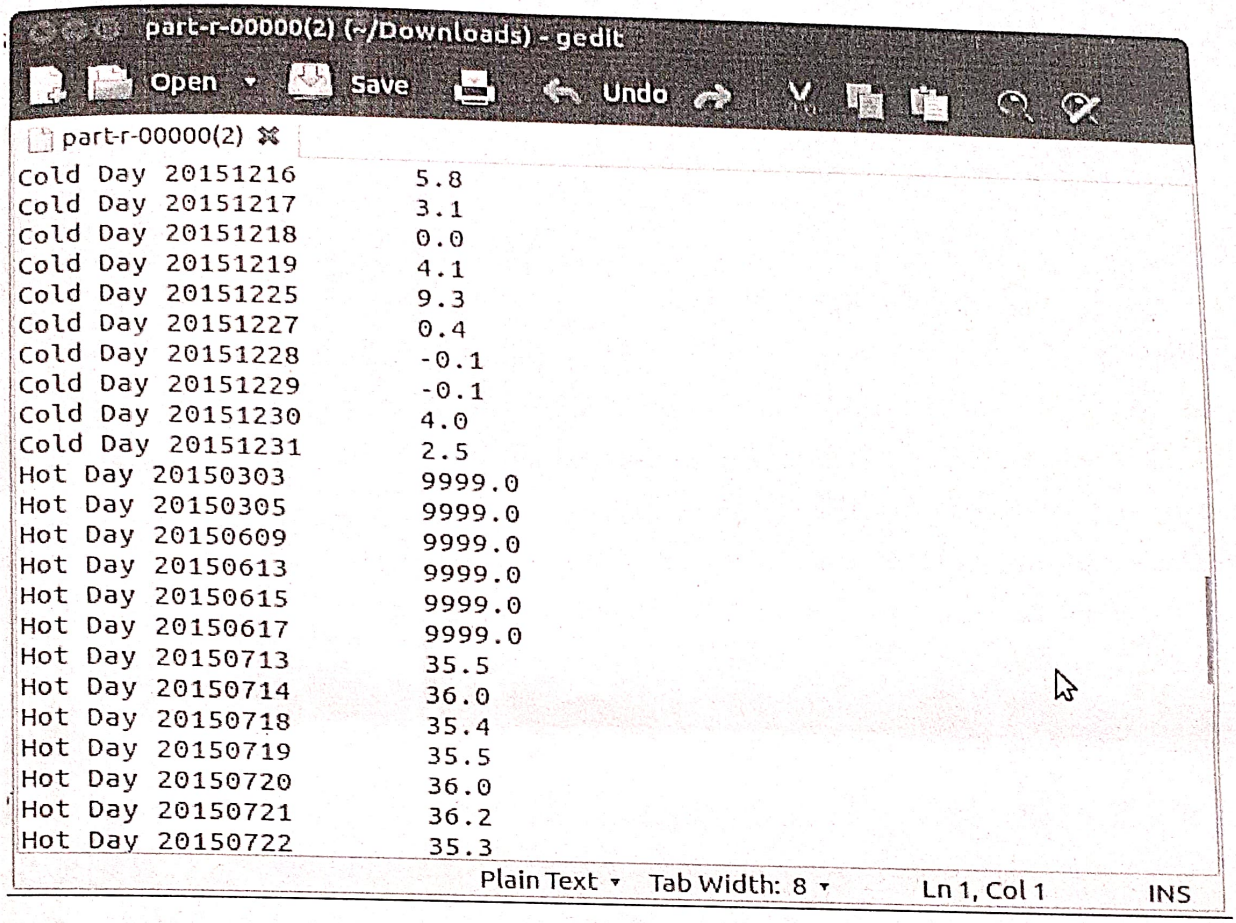
MapperClass:classwhichoverridesthe"map"function.Forhere,Map.
Reducer: classwhichoverride the "reduce" function. For here ,Reduce.
Output Key: type of output key. For here, Text.Output

Value:typeofoutputvalue.Forhere,IntWritable.
File Input Path

FileOutput Path

INPUT/OUTPUT:

SetofWeatherDataovertheyears



PRE-LABVIVAQUESTIONS:

1) ExplainthefunctionofMapReducer partitioner?
2) WhatisthedifferencebetweenanInputSplitandHDFS Block?
3) WhatisSequencefileinputformat?

LAB ASSIGNMENT:

1. UsingMapReducejobtoIdentifylanguagebymergingmultilanguagedictionaryfiles into a single dictionary file.
2. JoinmultipledatasetsusingaMapReduceJob.

POST-LABVIVA QUESTIONS: