EXPERIMENT -5

MAPREDUCEPROGRAM1

OBJECTIVE:

RunabasicwordcountMapReduceprogramtounderstandMapReduce Paradigm.

RESOURCES:

VMWarestack,4GBRAM,Webbrowser,HardDisk80 GB.

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 programming model making it a great example to understand the Hadoop Map/Reduce programming style. Our implementation consists of three main parts:

1. Mapper
2. Reducer
3. Driver

Step-1. Writea Mapper

A Mapper overrides the —mapǁ function from the Class "org.apache.hadoop.mapreduce.Mapper" which provides <key, value> pairs as the input. A Mapper implementation may output <key,value>pairsusingtheprovidedContext.

Input value of the WordCount Map task will be a line of text from the input data file and the key would bethe line number<line_number, line_of_text>.Maptaskoutputs<word,one> foreach word in the line of text.

Pseudo-code

voidMap(key,value){

```
        foreachwordxinvalue:

            output.collect(x,1);


}
```

Step-2. WriteaReducer

AReducercollectstheintermediate<key,value>outputfrommultiplemaptasksandassembleasingle result.

Here, the WordCount program will sum up the occurrence of each word to pairs as <word,occurrence>.

Pseudo-code

```
voidReduce(keyword,<listofvalue>){for
    each x in <list of value>:
      sum+=x;
  final_output.collect(keyword, sum);
}
```

INPUT/OUTPUT:

```
lendi@ubuntu: ~/Desktop
16/08/17 01:17:45 INFO impl.YarnClientImpl: Submitted application application_14
71410736896_0001
16/08/17 01:17:45 INFO mapreduce.Job: The url to track the job: http://ubuntu.ub
untu-domain:8088/proxy/application_1471410736896_0001/
16/08/17 01:17:45 INFO mapreduce.Job: Running job: job_1471410736896_0001
16/08/17 01:17:52 INFO mapreduce.Job: Job job_1471410736896_0001 running in uber
 mode : false
16/08/17 01:17:52 INFO mapreduce.Job:  map 0% reduce 0%
16/08/17 01:17:59 INFO mapreduce.Job:  map 100% reduce 0%
16/08/17 01:18:06 INFO mapreduce.Job:  map 100% reduce 100%
16/08/17 01:18:06 INFO mapreduce.Job: Job job_1471410736896_0001 completed succe
ssfully
16/08/17 01:18:06 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=3772644
                FILE: Number of bytes written=7775215
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1764718
                HDFS: Number of bytes written=510970
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
```

PRE-LABVIVAQUESTIONS:

1. Justifyhowhadooptechnologysatisfiesthebusinessinsightsnow-a–days?