LARGE SCALE DATA PROCESSING (CSE 3025)
WIN SEMESTER 2017-18

NAME: B.SHUBANKAR

REGNO: 15BCE1123

PROFESSOR: DR. Maheswari N

LAB-3

AIM: To identify word that occurs maximum no of time and total number of words available.

PROGRAM:

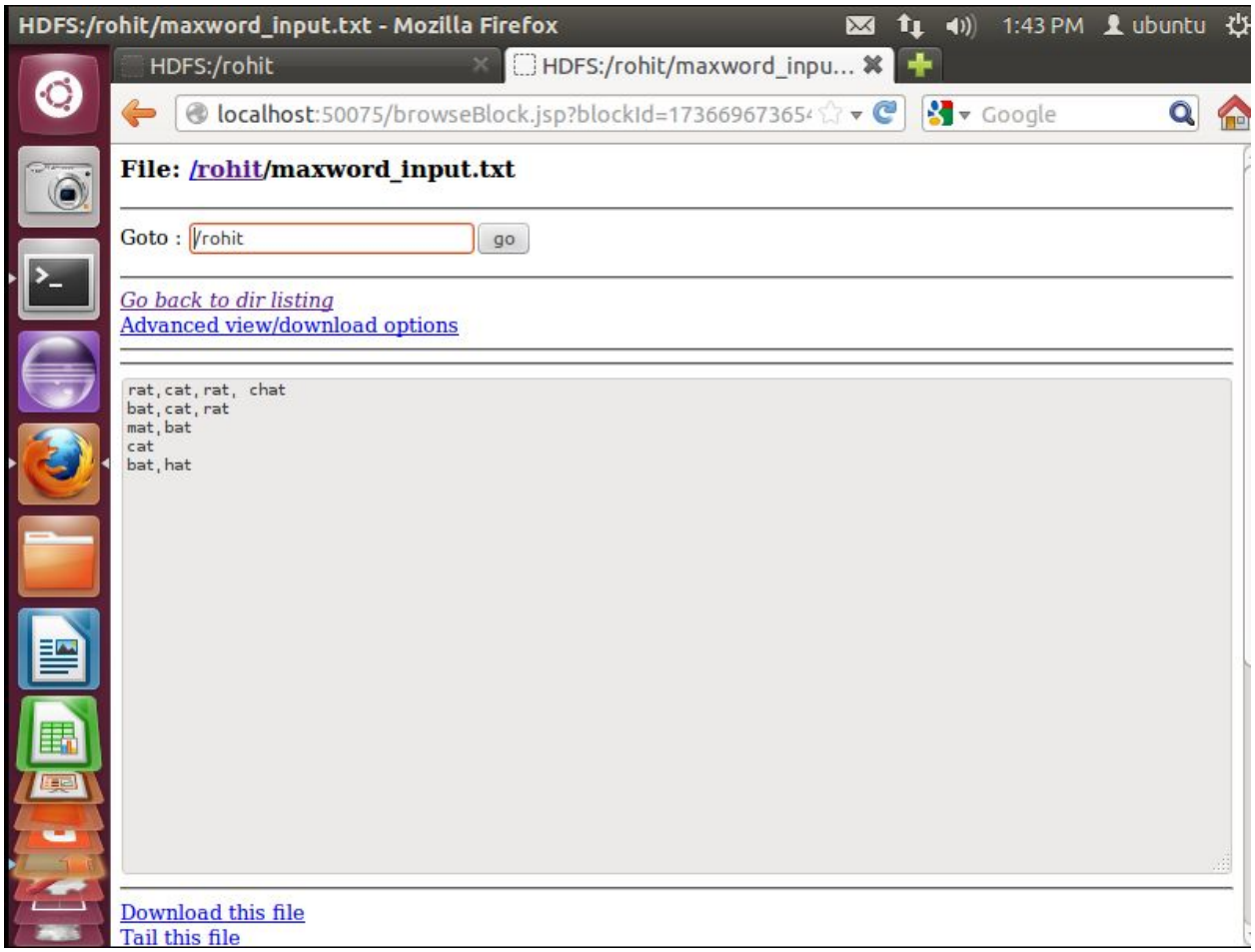
```
import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
public class WordCountmp1 {
public static class Map extends Mapper<LongWritable, Text, Text,
```

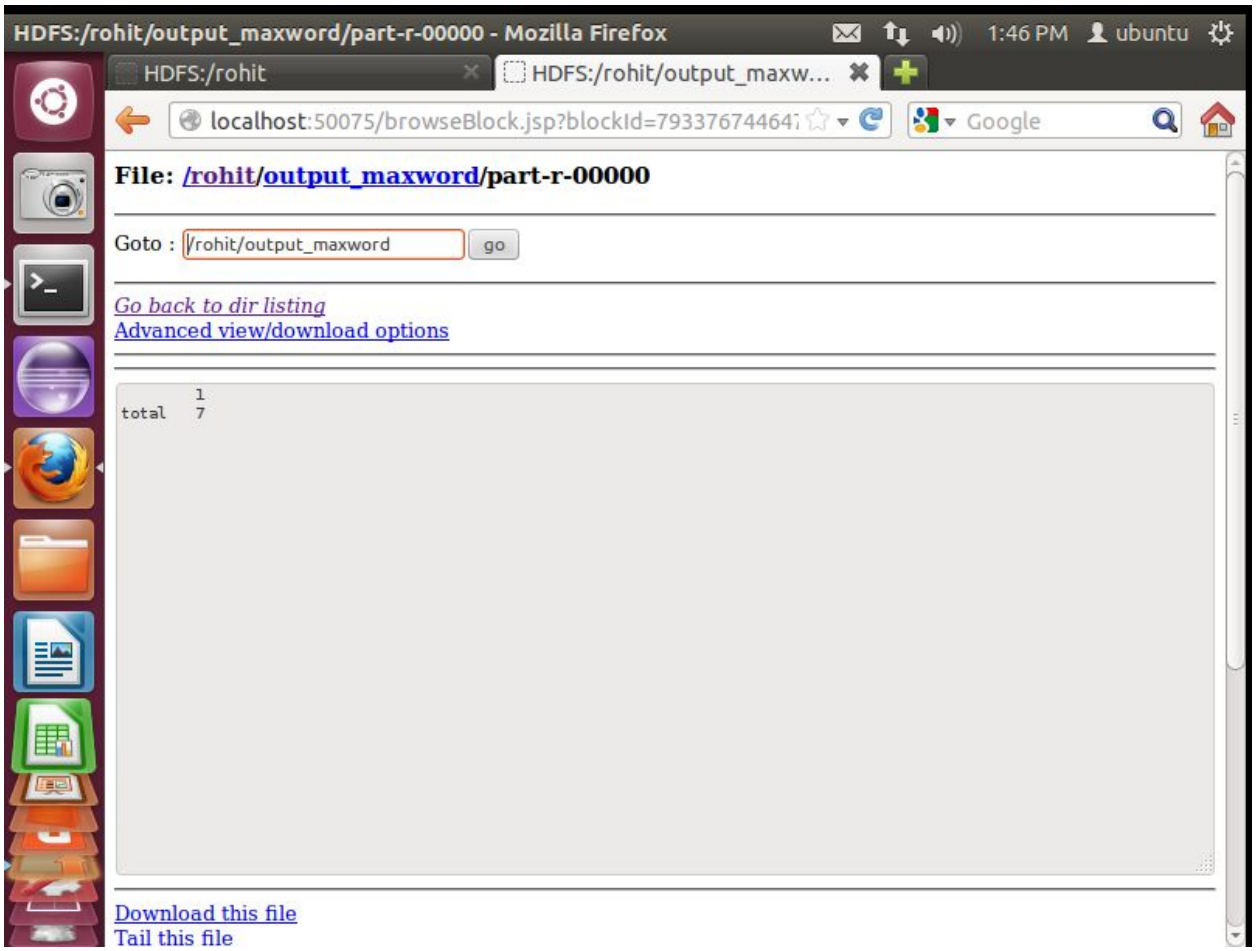
```
IntWritable> {
private final static IntWritable one = new IntWritable(1);
private Text word = new Text();
private Text w= new Text("Total no of words");
int tot=0;
public void map(LongWritable key, Text value, Context context) throws
IOException, InterruptedException {
String line = value.toString();
StringTokenizer tokenizer = new StringTokenizer(line);
while (tokenizer.hasMoreTokens()) {
word.set(tokenizer.nextToken());
tot++;
//context.write(word, one);
}

}
public void cleanup(Context context)throws IOException, InterruptedException {
        context.write(w, new IntWritable(tot));
}
}
public static class Reduce extends Reducer<Text, IntWritable, Text,
IntWritable> {
private Text mword = new Text();
private final static IntWritable count = new IntWritable(1);
```

```
int c=0;
public void reduce(Text key, Iterable<IntWritable> values, Context
context)throws IOException, InterruptedException {
int sum = 0;
for (IntWritable val : values) {
sum += val.get();
if(sum>c)
{
        c=sum;
        mword.set(key);
        count.set(c);
}
}
context.write(key, new IntWritable(sum));
}
public void cleanup(Context context)throws IOException, InterruptedException {
        context.write(mword,count);
}
}

public static void main(String[] args) throws Exception {
Configuration conf = new Configuration();
Job job = new Job(conf, "wordcountmp1");
job.setJarByClass(WordCountmp1.class);
```

```
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
job.setMapperClass(Map.class);
job.setReducerClass(Reduce.class);
job.setInputFormatClass(TextInputFormat.class);
job.setOutputFormatClass(TextOutputFormat.class);
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
job.waitForCompletion(true);
}
}
```

OUTPUT :


HDFS:/rohit/maxword_input.txt - Mozilla Firefox
1:43 PM
ubuntu
HDFS:/rohit
HDFS:/rohit/maxword_inpu...
localhost:50075/browseBlock.jsp?blockId=17366967365
Google
File: /rohit/maxword_input.txt
Goto :
/rohit
go
Go back to dir listing
Advanced view/download options
rat,cat,rat, chat
bat,cat,rat
mat,bat
cat
bat,hat
Download this file
Tail this file


HDFS:/rohit/output_maxword/part-r-00000 - Mozilla Firefox
1:46 PM
ubuntu
HDFS:/rohit
HDFS:/rohit/output_maxw...
localhost:50075/browseBlock.jsp?blockId=793376744647
Google
File: /rohit/output_maxword/part-r-00000
Goto : /rohit/output_maxword
go
Go back to dir listing
Advanced view/download options
1
total 7
Download this file
Tail this file