VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

ITA-6008 Big Data Analytics

Assignment - 2

By :

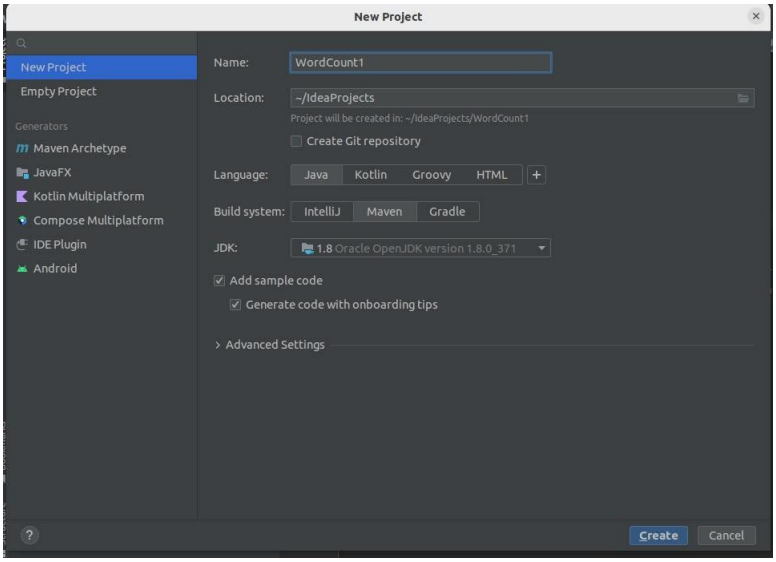
Rajat Singh 22MCA0139

Submitted to :

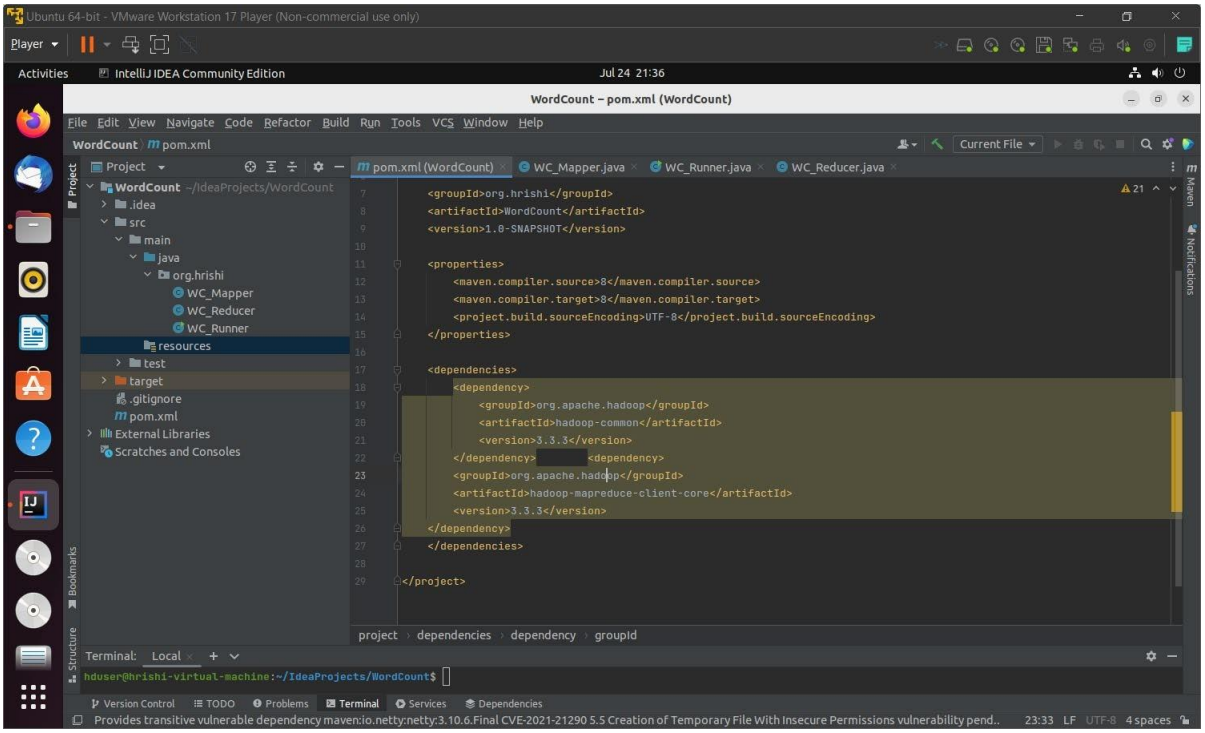
Prof. POUNAMBAL M

Write and execute any map reduce program in your Hadoop environment and submit the same with the snapshot.

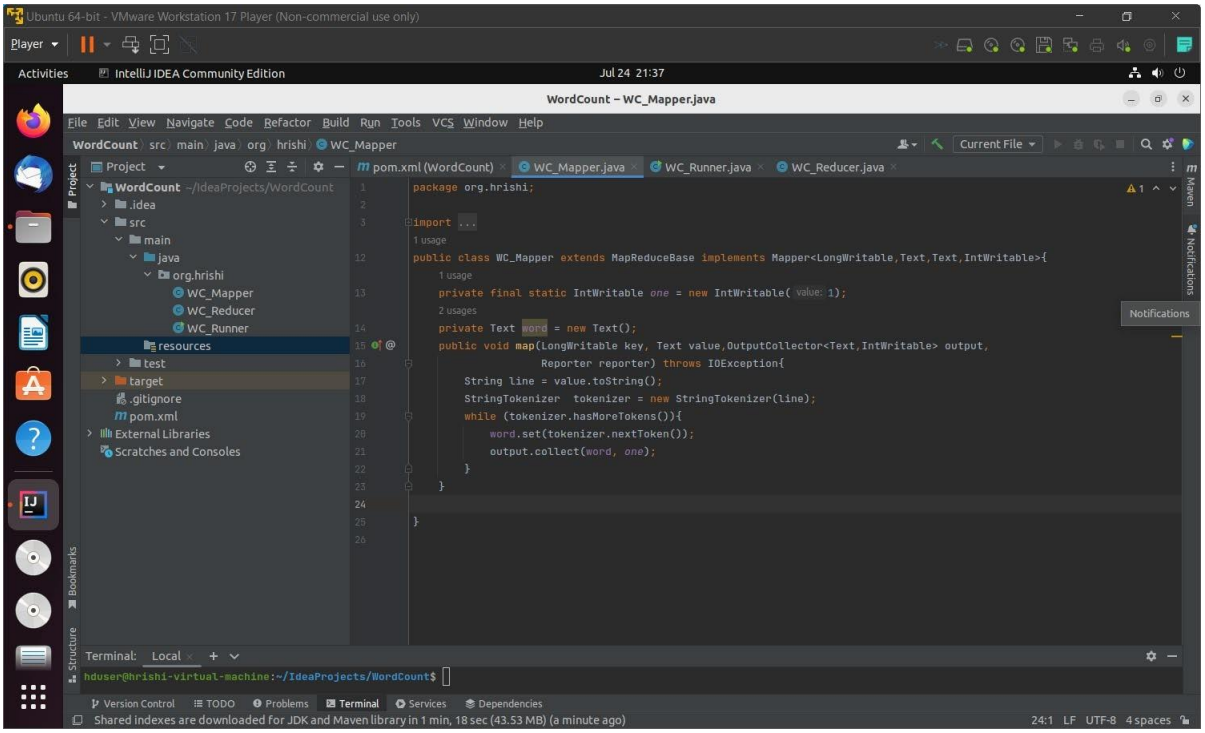
1. Create java new project:



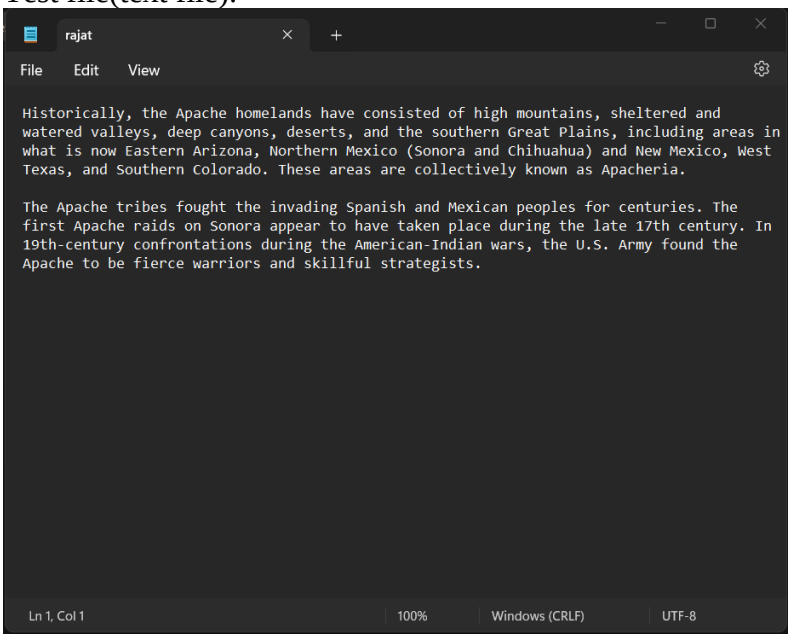
2. Add Dependencies:



3. Write the code for word count:



Test file(text file):



Code:

WC_Mapper.java

```
import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;
public class WC_Mapper extends MapReduceBase implements
Mapper<LongWritable,Text,Text,IntWritable>{
private final static IntWritable one = new IntWritable(1);
private Text word = new Text();
public void map(LongWritable key, Text value,OutputCollector<Text,IntWritable> output,
Reporter reporter) throws IOException{
String line = value.toString();
StringTokenizer tokenizer = new StringTokenizer(line);
while (tokenizer.hasMoreTokens()){
word.set(tokenizer.nextToken());
output.collect(word, one);
}
```

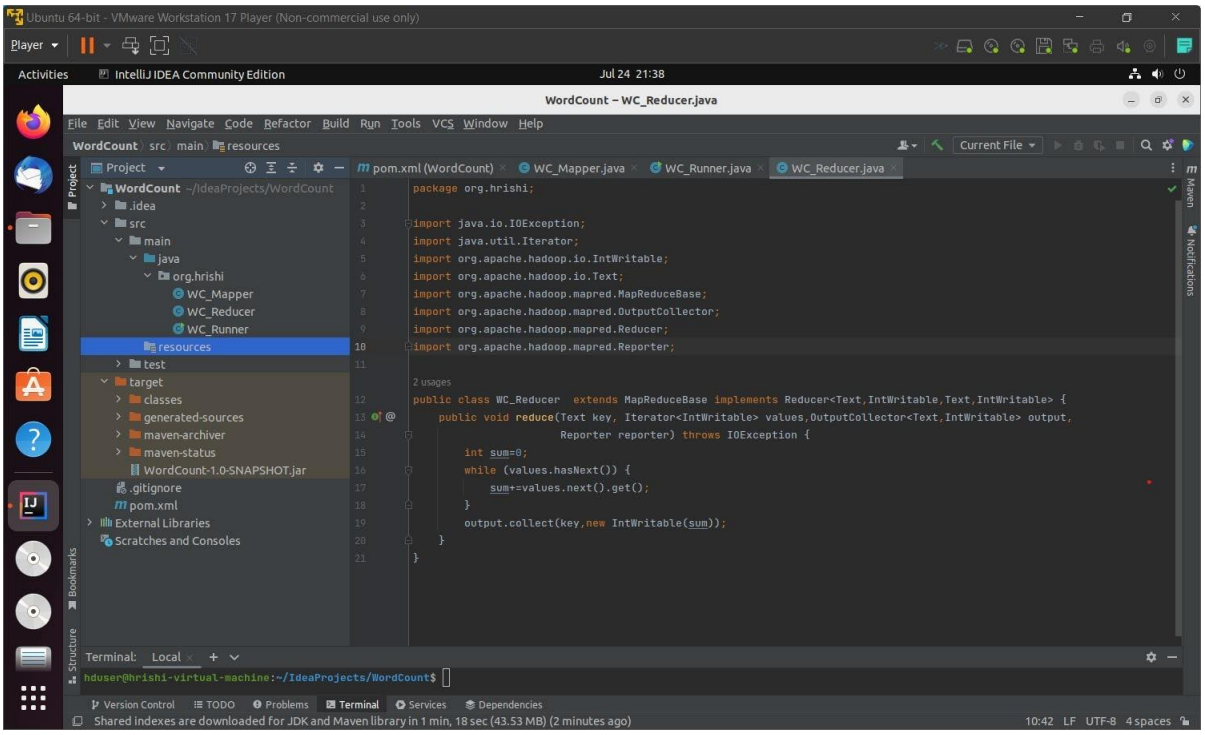
```
}
```

WC_Reducer.java

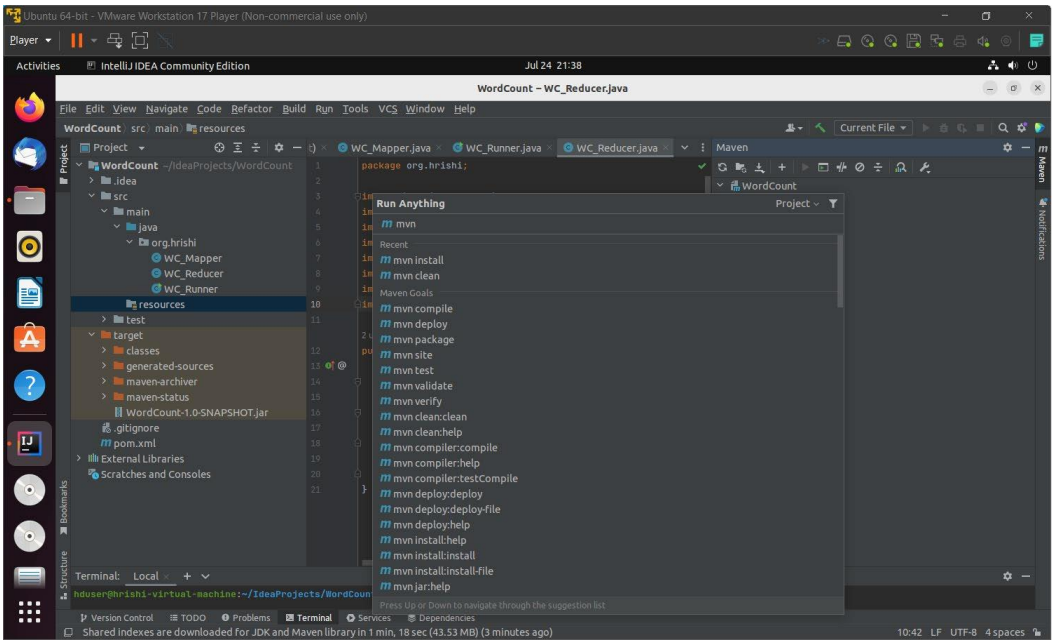
```
import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;

public class WC_Reducer extends MapReduceBase implements
Reducer<Text,IntWritable,Text,IntWritable> {
    public void reduce(Text key, Iterator<IntWritable>
values,OutputCollector<Text,IntWritable> output,
                       Reporter reporter) throws IOException {
        int sum=0;
        while (values.hasNext()) {
            sum+=values.next().get();
        }
        output.collect(key,new IntWritable(sum));
    }
}
```

WC_Runner.java

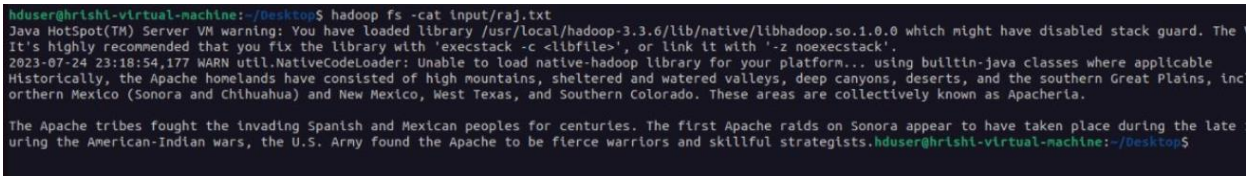
```
import java.io.IOException;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.mapred.TextInputFormat;
import org.apache.hadoop.mapred.TextOutputFormat;
public class WC_Runner {
    public static void main(String[] args) throws IOException{
        JobConf conf = new JobConf(WC_Runner.class);
        conf.setJobName("WordCount");
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        conf.setMapperClass(WC_Mapper.class);
        conf.setCombinerClass(WC_Reducer.class);
        conf.setReducerClass(WC_Reducer.class);
        conf.setInputFormat(TextInputFormat.class);
        conf.setOutputFormat(TextOutputFormat.class);
        FileInputFormat.setInputPaths(conf,new Path(args[0]));
        FileOutputFormat.setOutputPath(conf,new Path(args[1]));
        JobClient.runJob(conf);
    }
}
```



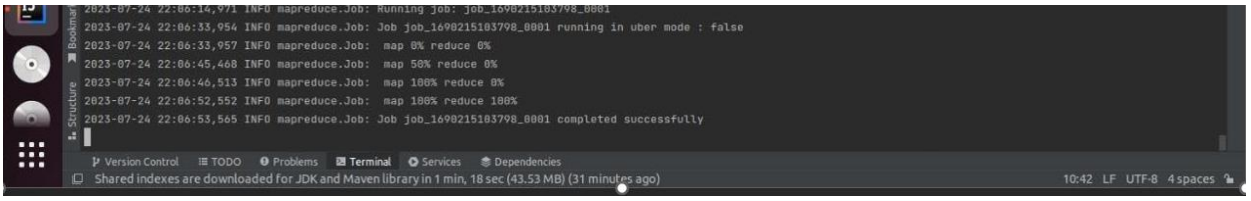
4. Create a jar file using maven install:



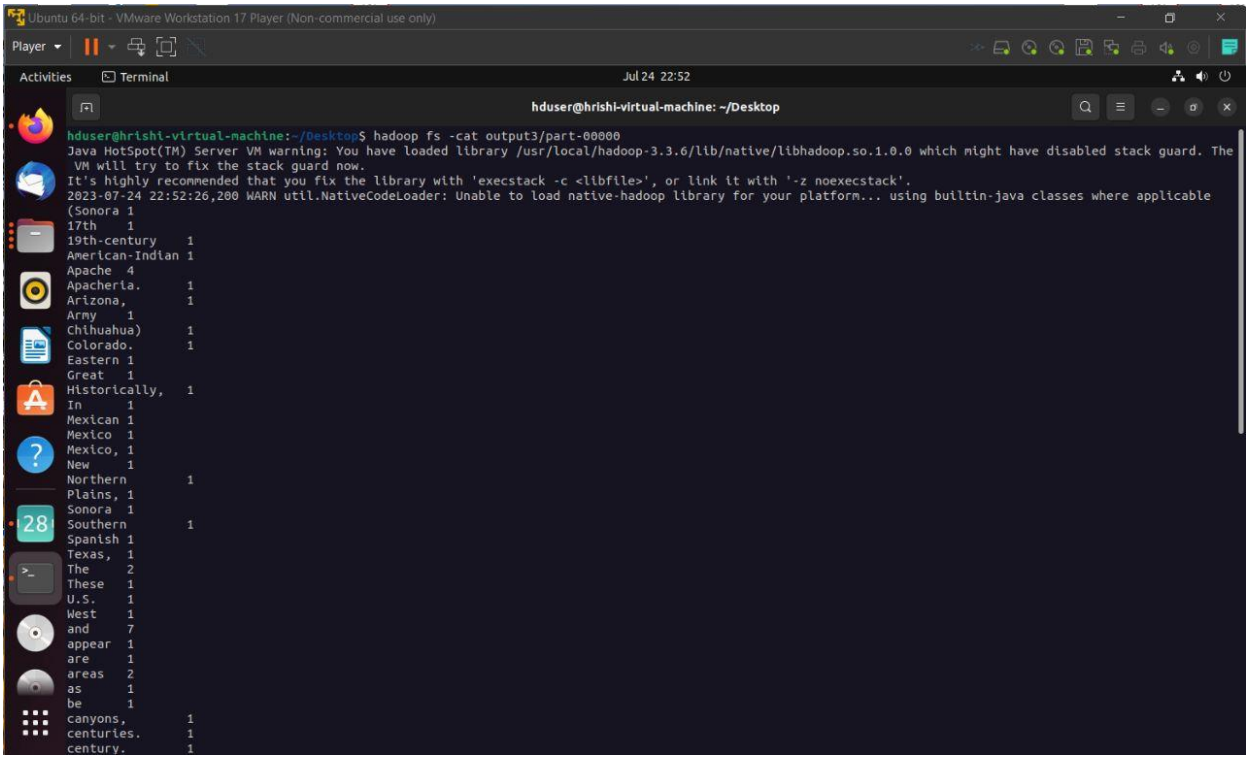
5. Create a file in Hadoop fs:

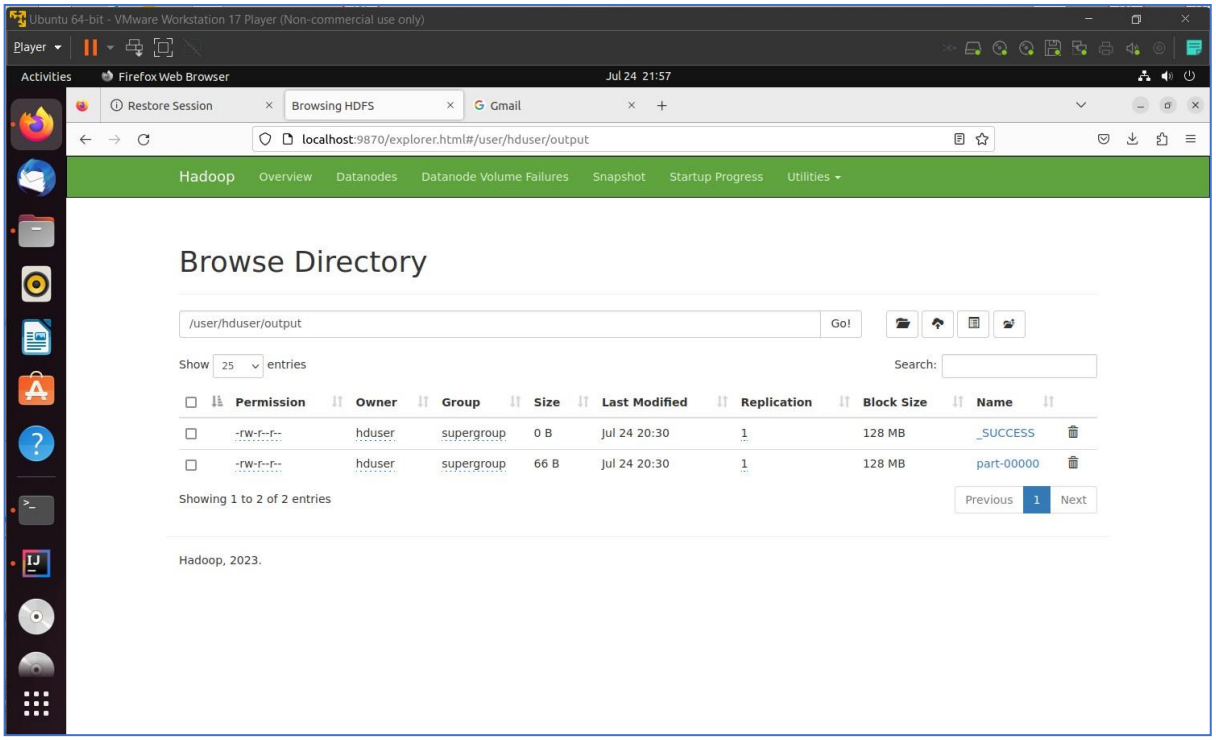


6. Run the following command to run the code and get the output:



Output:


Ubuntu 64-bit - VMware Workstation 17 Player (Non-commercial use only)
Player
Activities
Terminal
Jul 24 22:52
hduser@hrishi-virtual-machine: ~/Desktop
hduser@hrishi-virtual-machine:~/Desktop$ hadoop fs -cat output3/part-00000
Java HotSpot(TM) Server VM warning: You have loaded library /usr/local/hadoop-3.3.6/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
2023-07-24 22:52:26,200 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
(Sonora 1
17th 1
19th-century 1
American-Indian 1
Apache 4
Apacheria. 1
Arizona, 1
Army 1
Chihuahua) 1
Colorado. 1
Eastern 1
Great 1
Hlstorically, 1
In 1
Mexican 1
Mexico 1
Mexico, 1
New 1
Northern 1
Plains, 1
Sonora 1
Southern 1
Spanish 1
Texas, 1
The 2
These 1
U.S. 1
West 1
and 7
appear 1
are 1
areas 2
as 1
be 1
canyons, 1
centuries. 1
century. 1


Ubuntu 64-bit - VMware Workstation 17 Player (Non-commercial use only)
Player
Activities
Firefox Web Browser
Jul 24 21:57
Restore Session
Browsing HDFS
Gmail
localhost:9870/explorer.html#/user/hduser/output
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/user/hduser/output
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
-rw-r--r--
hduser
supergroup
0 B
Jul 24 20:30
1
128 MB
_SUCCESS
-rw-r--r--
hduser
supergroup
66 B
Jul 24 20:30
1
128 MB
part-00000
Showing 1 to 2 of 2 entries
Previous
1
Next
Hadoop, 2023.