

ITA-6008 Big Data Analytics Assignment - 2

By :

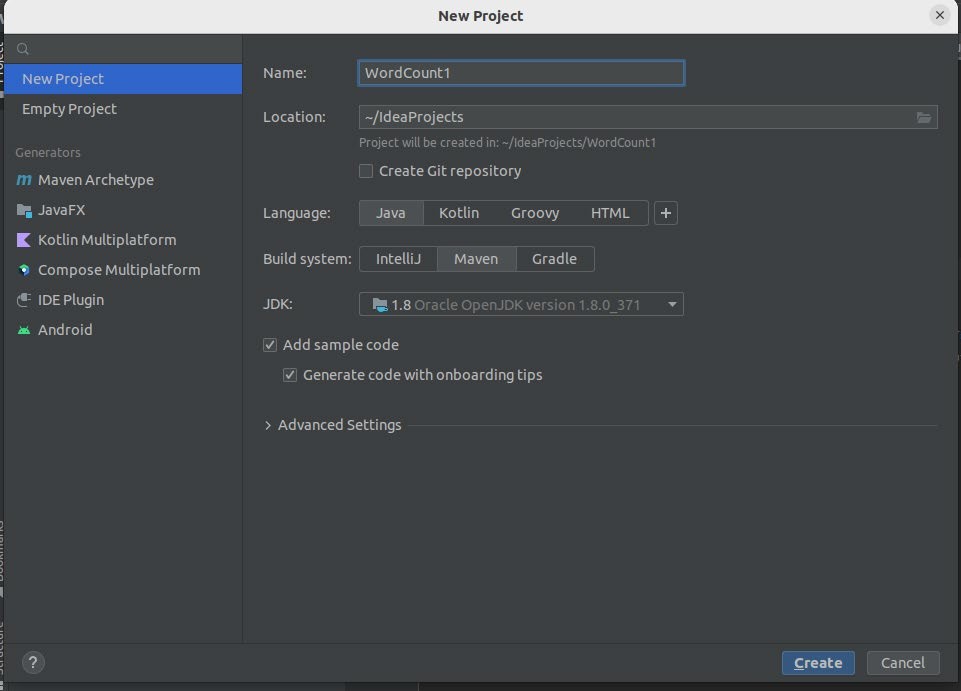
Rajat Singh 22MCA0139

Submitted to :

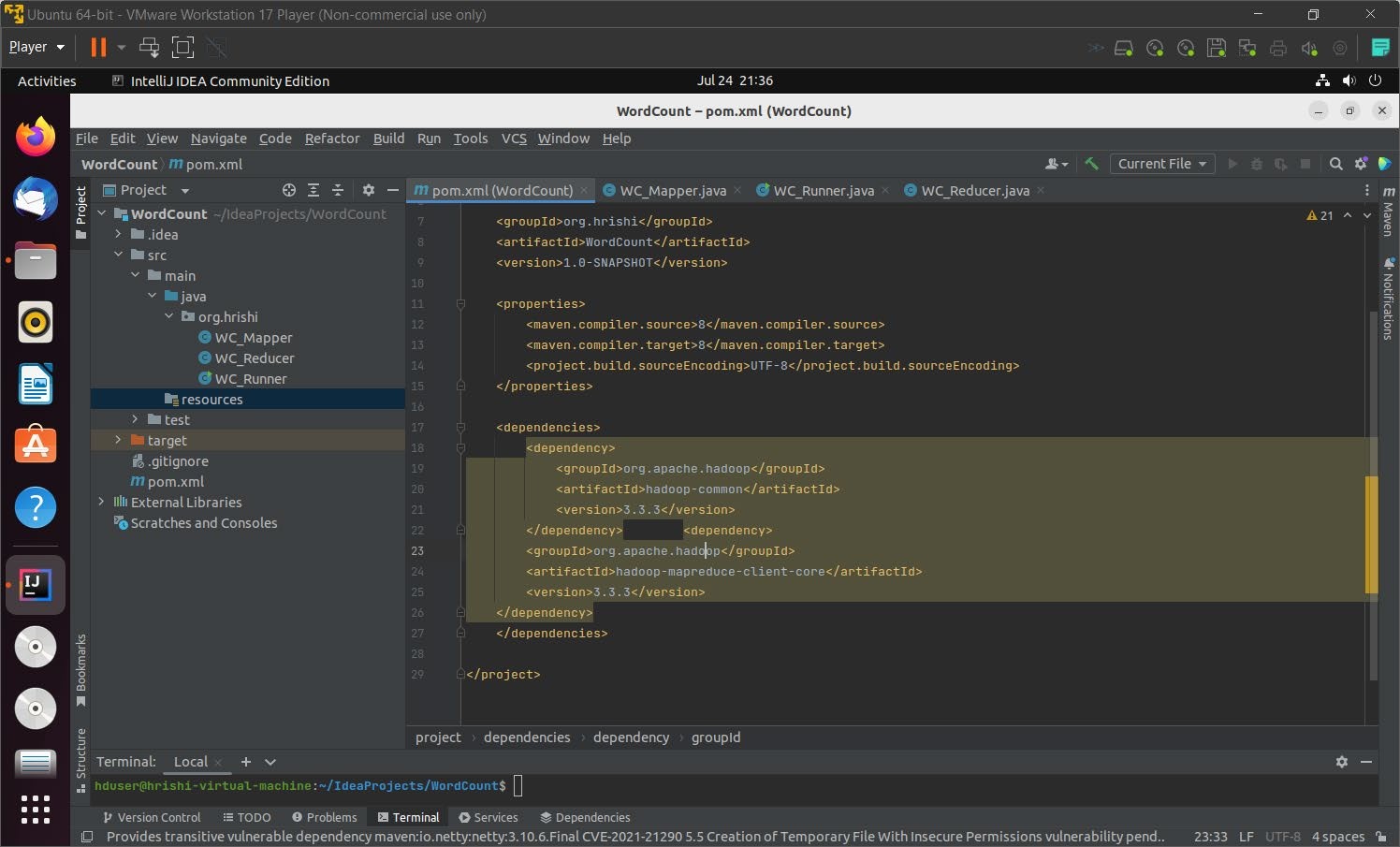
**Prof. POUNAMBAL M**

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

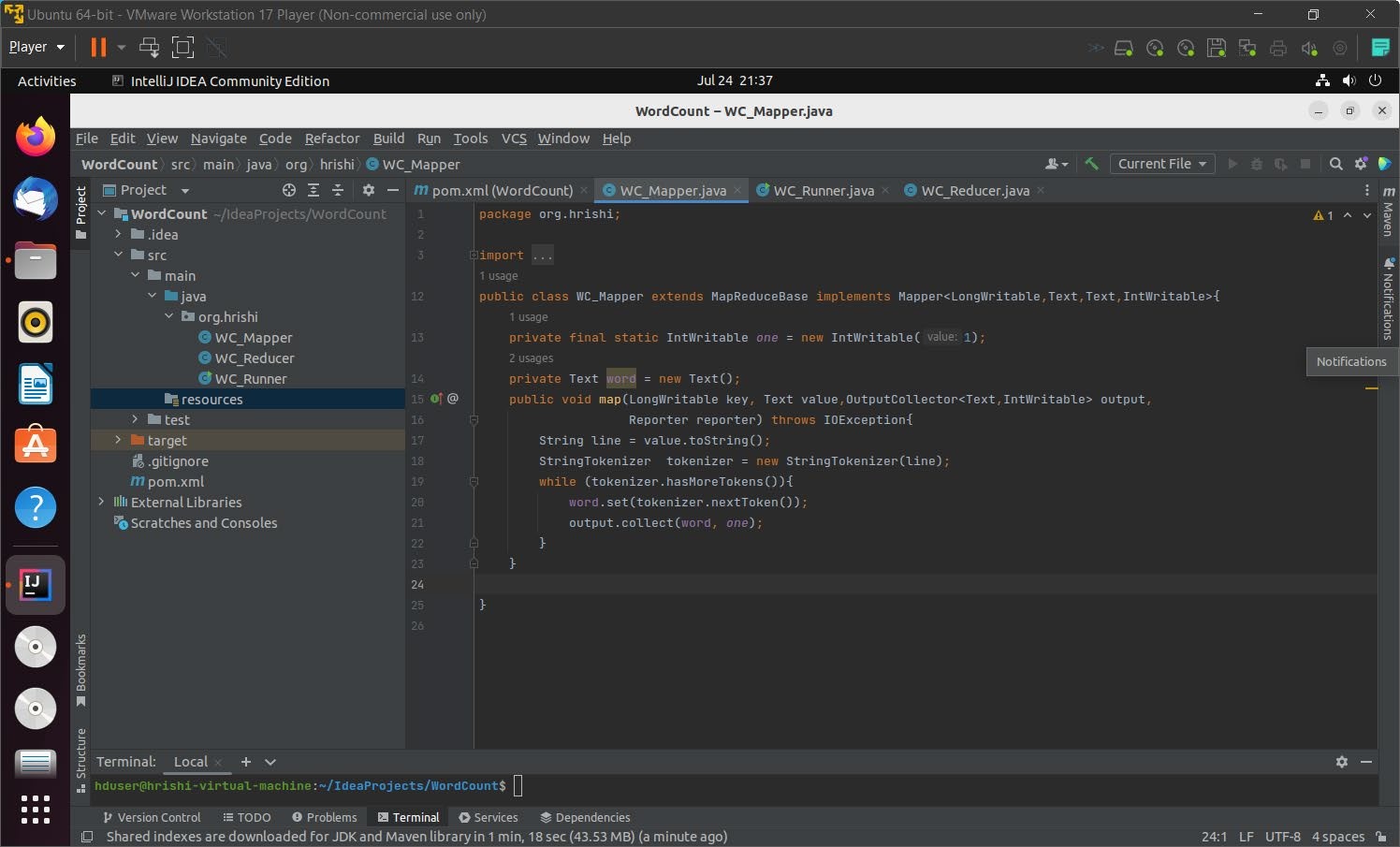
1. Create java new project:

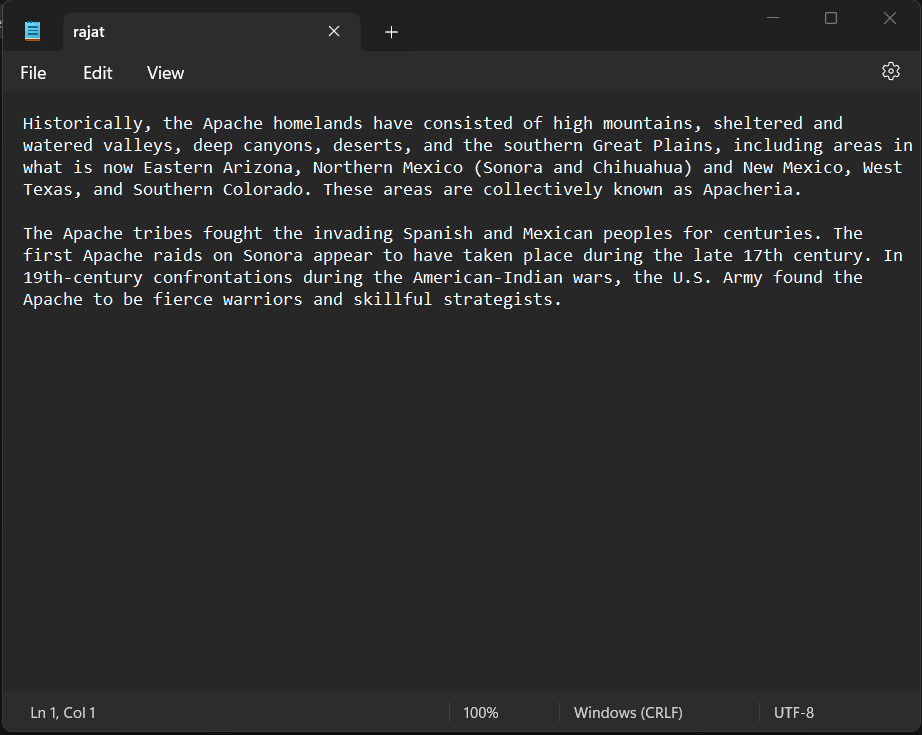


1. Add Dependencies:



1. 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;

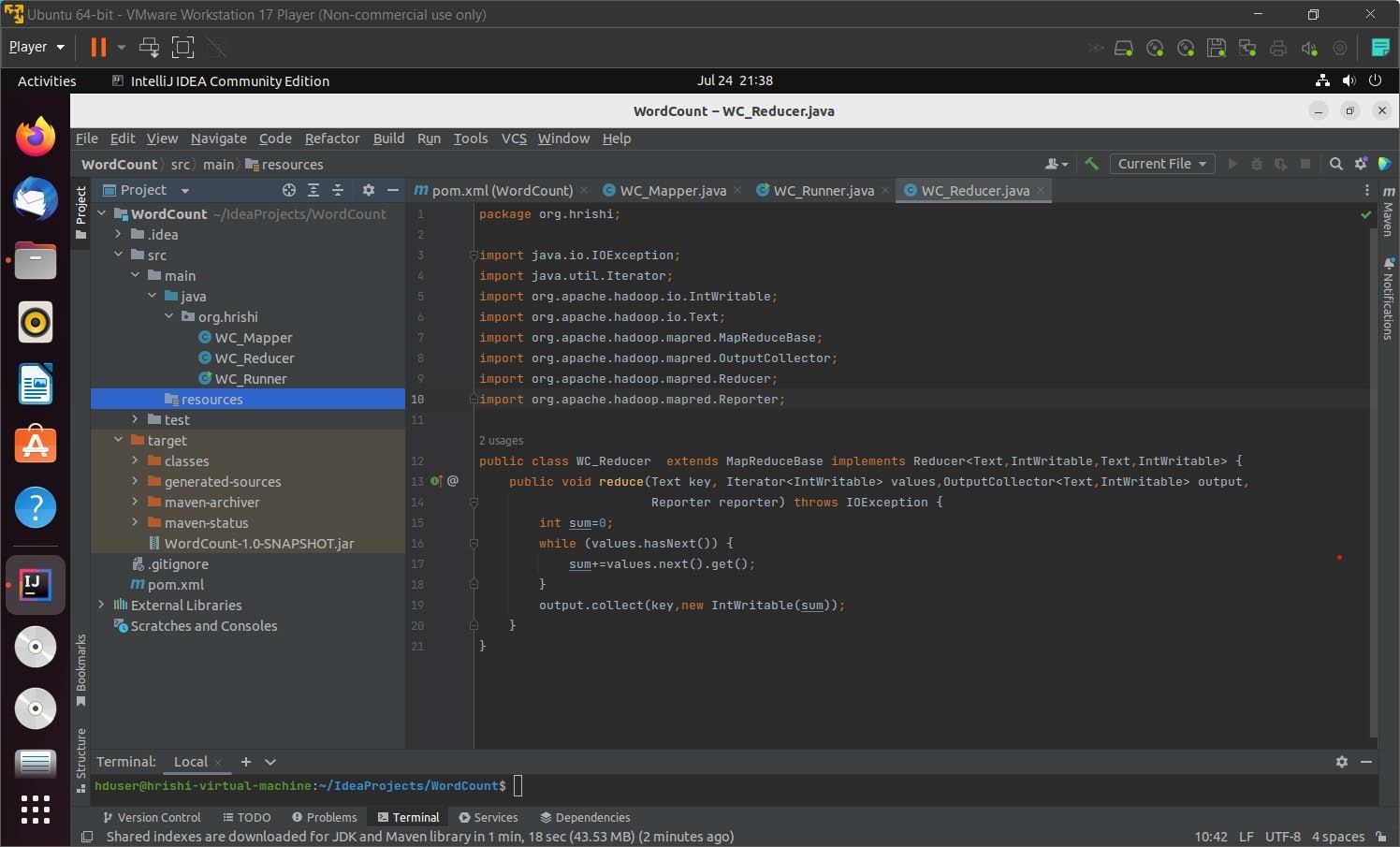
import org.apache.hadoop.mapred.FileInputFormat; import org.apache.hadoop.mapred.FileOutputFormat; import org.apache.hadoop.mapred.JobClient; import org.apache.hadoop.mapred.JobConf;

import org.apache.hadoop.mapred.TextInputFormat; import org.apache.hadoop.mapred.TextOutputFormat; public class WC\_Runner {

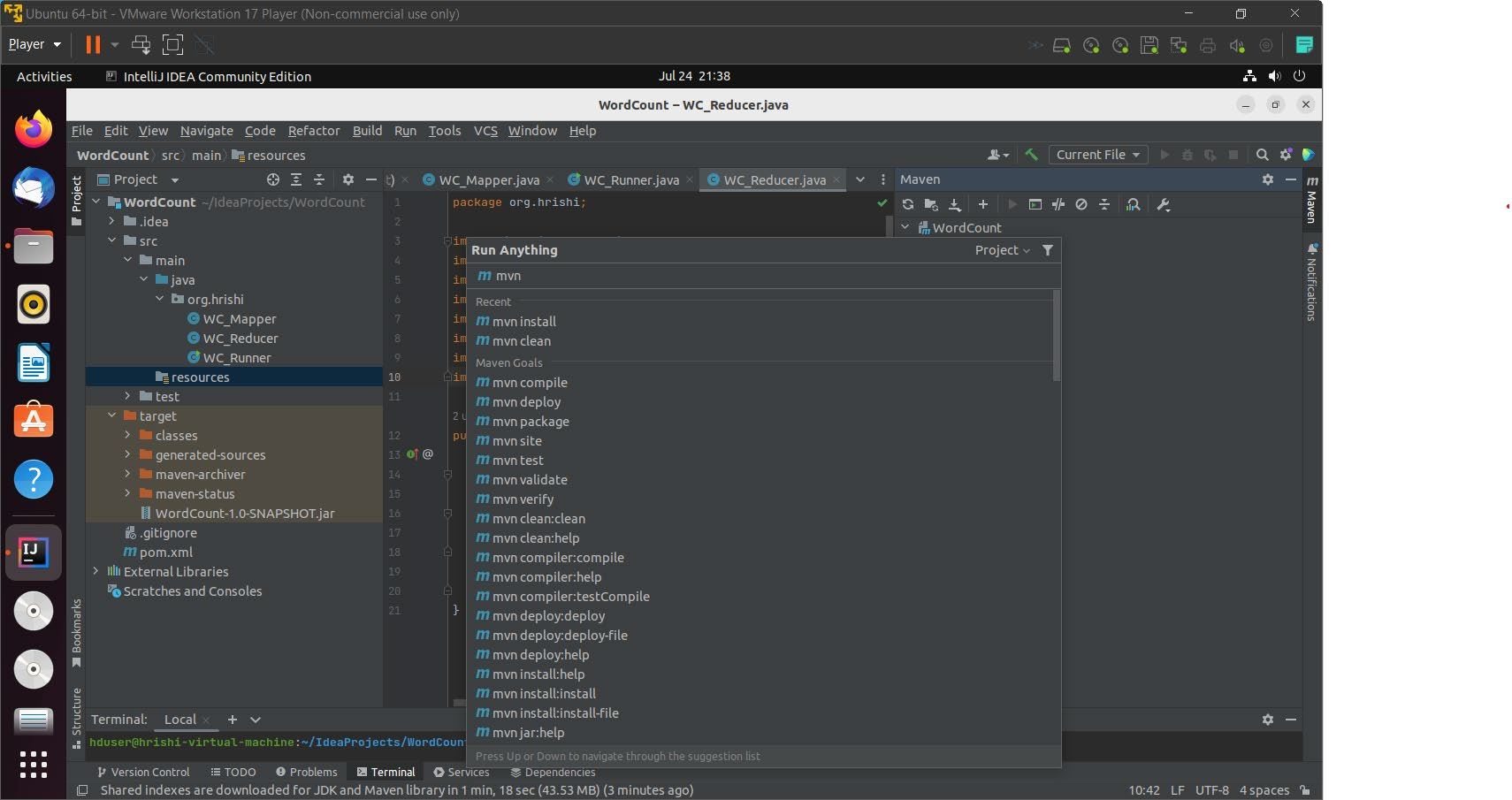
public static void main(String[] args) throws IOException{ JobConf conf = new JobConf(WC\_Runner.class); conf.setJobName("WordCount"); conf.setOutputKeyClass(Text.class); conf.setOutputValueClass(IntWritable.class); conf.setMapperClass(WC\_Mapper.class); conf.setCombinerClass(WC\_Reducer.class); conf.setReducerClass(WC\_Reducer.class); 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);

}

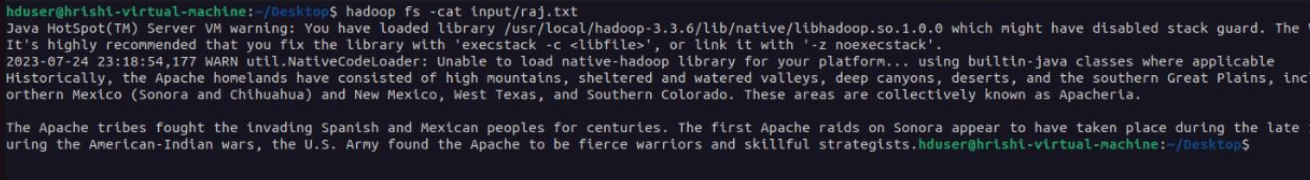
}



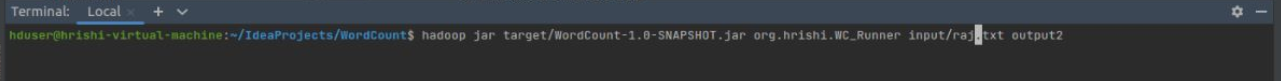
1. Create a jar file using maven install:



1. Create a file in Hadoop fs:



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



A screenshot of a computer

Description automatically generated

# Output:

A screenshot of a computer

Description automatically generated

