



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

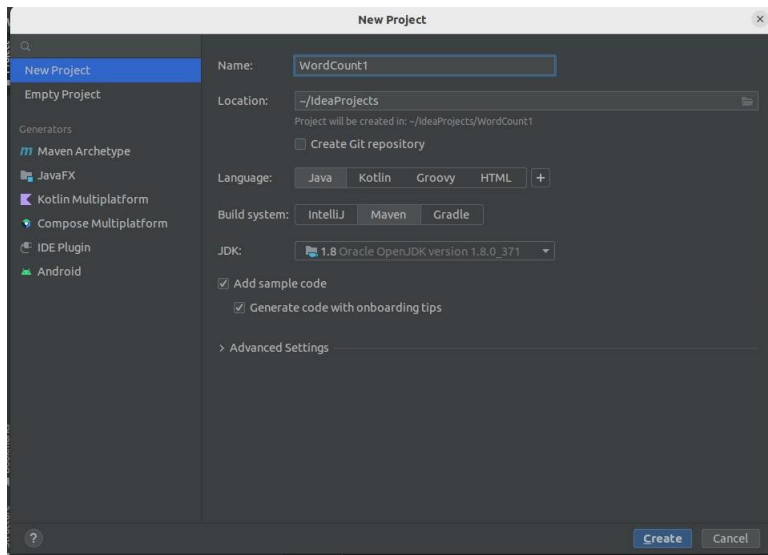
22MCA0139

Submitted to :

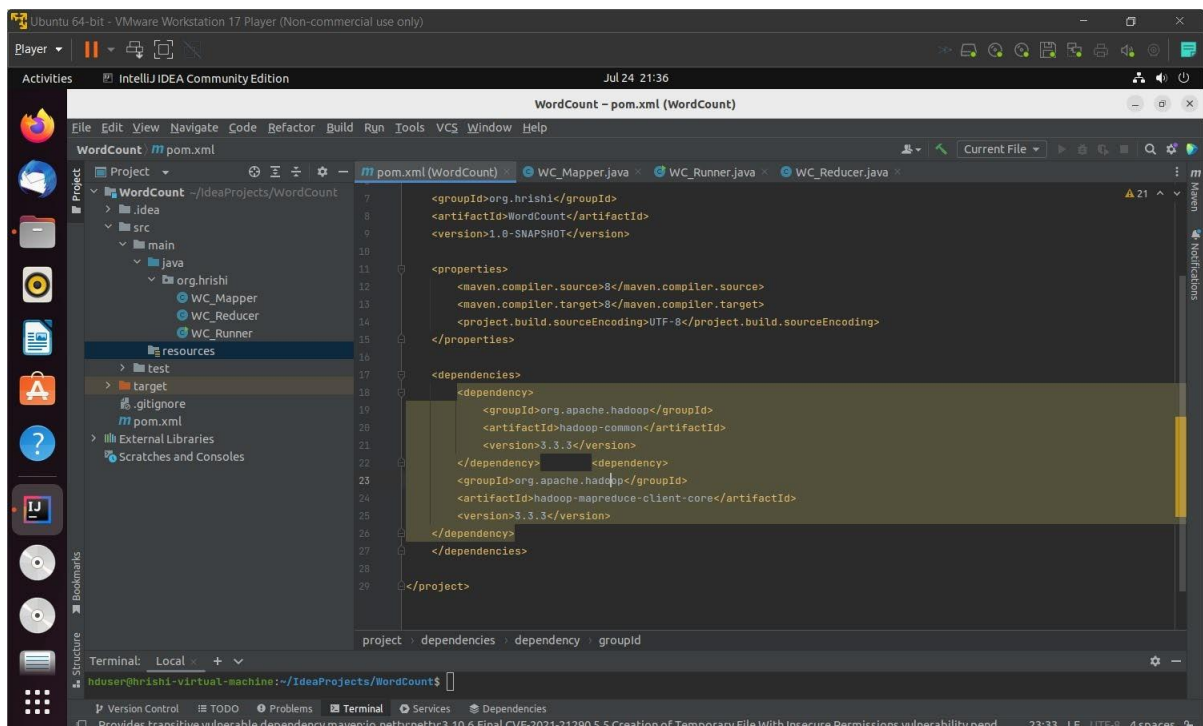
Prof. POUNAMBAL M

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

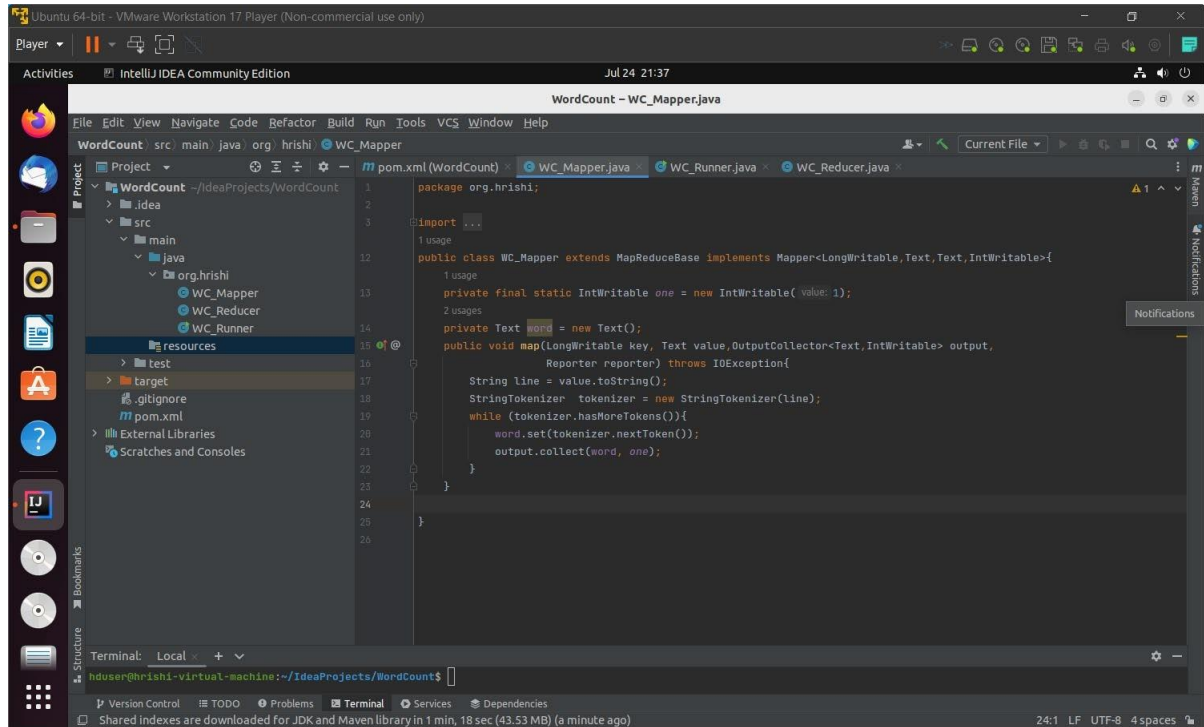
1. Create java new project:



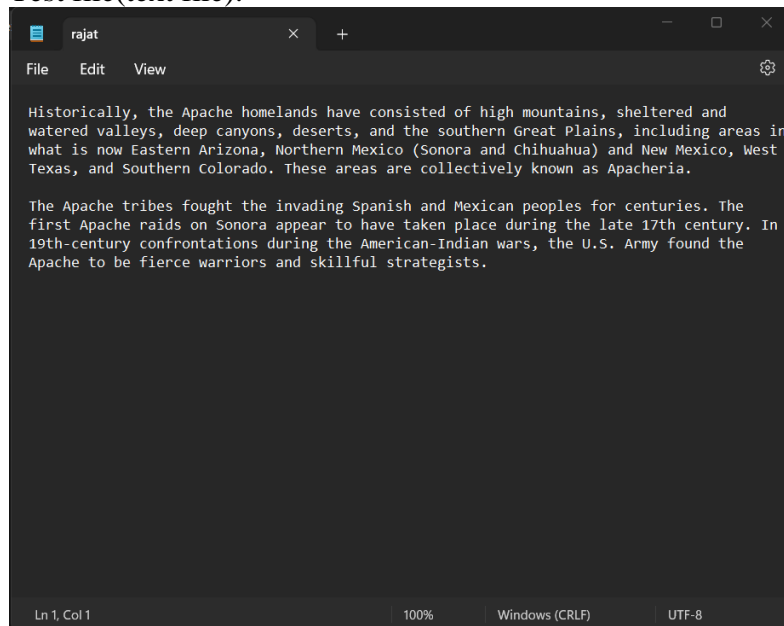
2. Add Dependencies:



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);
}
}
```

```
}
```

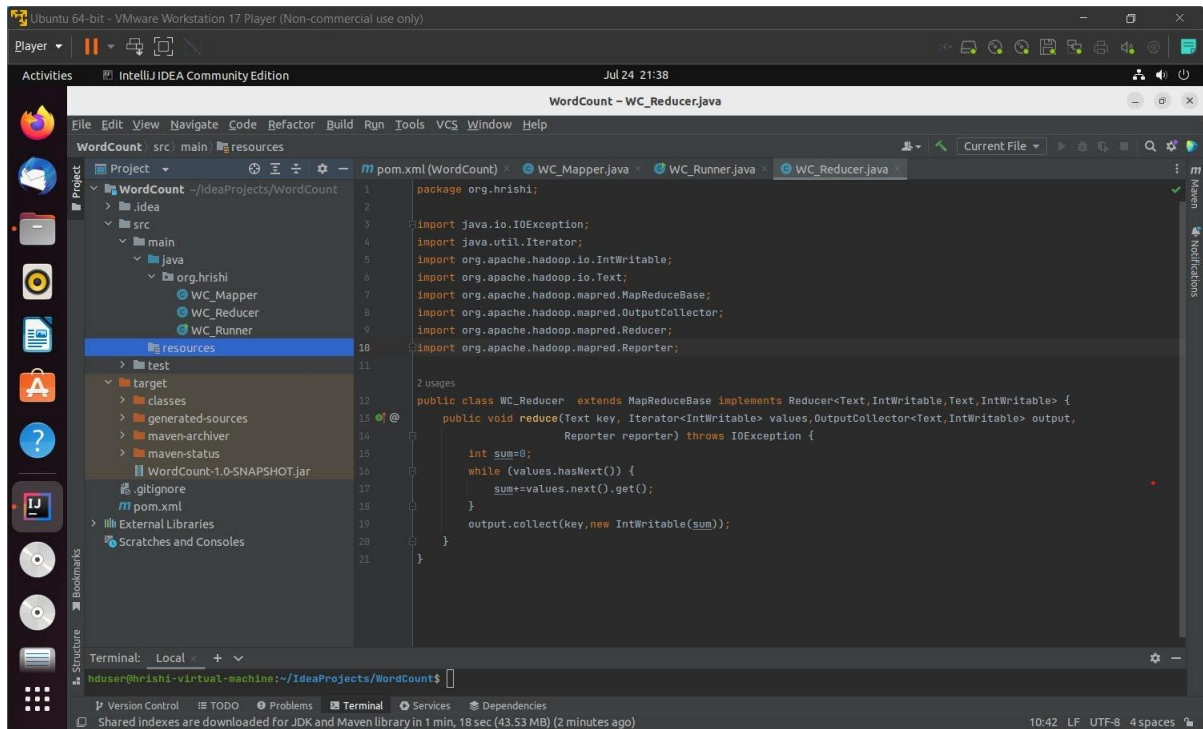
WC_Reducer.java

```
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;

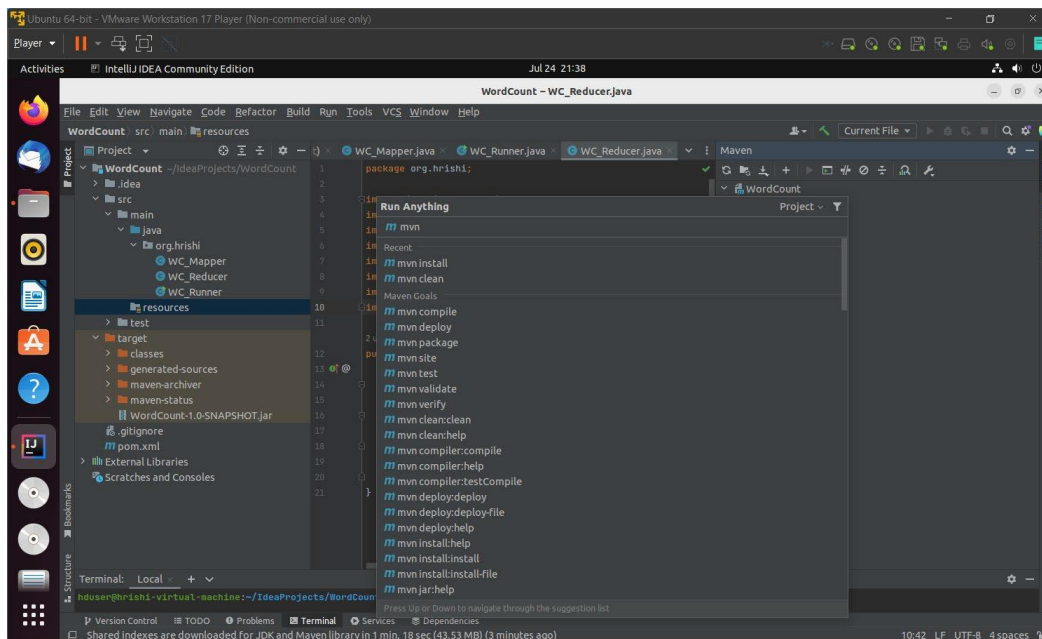
public class WC_Reducer extends MapReduceBase implements
Reducer<Text,IntWritable,Text,IntWritable> {
    public void reduce(Text key, Iterator<IntWritable>
values,OutputCollector<Text,IntWritable> output,
                        Reporter reporter) throws IOException {
        int sum=0;
        while (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:

```
hduser@hrishi-virtual-machine:~/Desktop$ hadoop fs -cat input/raj.txt
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
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
2023-07-24 23:18:54,177 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Historically, the Apache homelands have consisted of high mountains, sheltered and watered valleys, deep canyons, deserts, and the southern Great Plains, inc
orthern Mexico (Sonora and Chihuahua) and New Mexico, West Texas, and Southern Colorado. These areas are collectively known as Apacheria.

The Apache tribes fought the invading Spanish and Mexican peoples for centuries. The first Apache raids on Sonora appear to have taken place during the late
uring the American-Indian wars, the U.S. Army found the Apache to be fierce warriors and skillful strategists.hduser@hrishi-virtual-machine:~/Desktop$
```

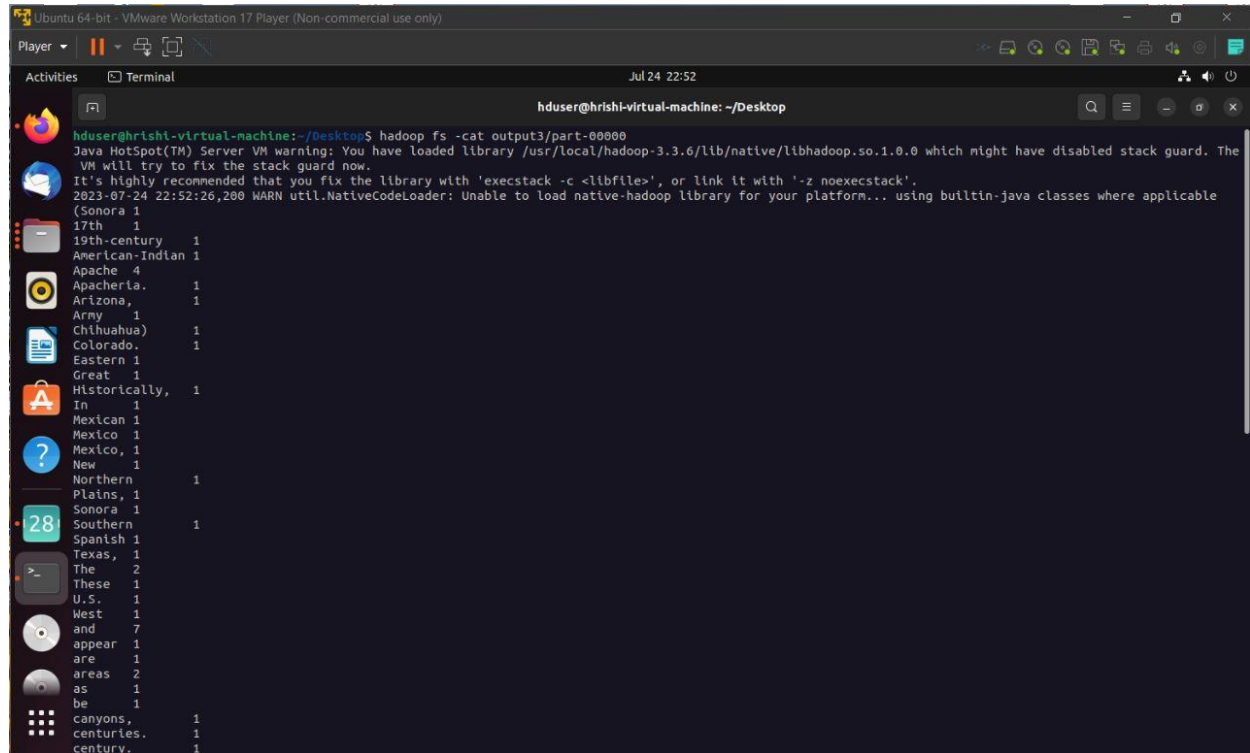
6. Run the following command to run the code and get the output:

```
Terminal: Local + -
hduser@hrishi-virtual-machine:~/IdeaProjects/WordCount$ hadoop jar target/WordCount-1.0-SNAPSHOT.jar org.hrishi.WC.Runner input/raj.txt output2
```

```
2023-07-24 22:06:14,971 INFO mapreduce.Job: Running job: job_1690215103798_0001
2023-07-24 22:06:33,954 INFO mapreduce.Job: Job job_1690215103798_0001 running in uber mode : false
2023-07-24 22:06:45,467 INFO mapreduce.Job: map 0% reduce 0%
2023-07-24 22:06:46,513 INFO mapreduce.Job: map 58% reduce 0%
2023-07-24 22:06:52,552 INFO mapreduce.Job: map 100% reduce 0%
2023-07-24 22:06:53,565 INFO mapreduce.Job: map 100% reduce 100%
2023-07-24 22:06:53,565 INFO mapreduce.Job: Job job_1690215103798_0001 completed successfully

P Version Control  IDE TODO  Problems  Terminal  Services  Dependencies
Shared indexes are downloaded for JDK and Maven library in 1 min, 18 sec (43.53 MB) (31 minutes ago) 10:42 LF UTF-8 4spaces
```

Output:



The screenshot shows a terminal window titled "hduser@hrishi-virtual-machine: ~/Desktop". The terminal output displays the command `hadoop fs -cat output3/part-00000` and its results. The output is a list of words and their frequencies, such as "17th 1", "19th-century 1", "American-Indian 1", "Apache 4", "Apacheria. 1", "Arizona, 1", "Army 1", "Chihuahua) 1", "Colorado. 1", "Eastern 1", "Great 1", "Historically, 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", and "century. 1".

```
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
Historically, 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
```