

Experiment 9.

Student Name: Mohit Kumar Gupta
Branch: MCA
Semester: 6
Subject Name: DABli Lab

UID: 19MCA8265
Section/Group: C/2
Date of Performance: 02/05/2021
Subject Code: 867

1. Aim/Overview of the practical:

Implementing Frequent Item set algorithms using Map-Reduce

2. Task to be done:

Frequent Item Set Algorithms

3. Algorithm/Flowchart:

Eclipse IDE — JAR file — Hadoop terminal — Jar file Execution.

4. Program:

```
import java.io.BufferedReader;  
  
import java.util.*;  
  
import java.io.InputStreamReader;
```



```
import java.io.IOException;  
import java.util.ArrayList;  
import java.util.Collections;  
  
import java.util.StringTokenizer;
```

```
import org.apache.hadoop.conf.Configuration; import
org.apache.hadoop.fs.FSDataOutputStream; import
org.apache.hadoop.fs.FileSystem;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Job;

import org.apache.hadoop.mapreduce.Mapper;

import org.apache.hadoop.mapreduce.Reducer;

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat; import
org.apache.hadoop.mapreduce.lib.output.FileOutputFormat; import
org.apache.hadoop.util.GenericOptionsParser;
```

```
public class Apriori {

    public static class TokenizerMapper extends Mapper<Object, Text, Text,
IntWritable>

    {

        private final static IntWritable one = new IntWritable();

        private Text word = new Text(); private
String t = new String(); private String
subsets = new String();
```

```
//Get conf from main function
```

```
// static Configuration test;
```

```
/* public void setup(Context context) throws IOException
{
    //Add Feature for read file

    Path pt = new Path("hdfs:/ethonwu/Anotherfile.txt"); //Location on HDFS
    FileSystem fs = FileSystem.get(new Configuration());
    BufferedReader br = new BufferedReader(new InputStreamReader(fs.open(pt)));
    String line;
    line = br.readLine();
    while(line!=null)
    {

    }
}
```

```
//end
```

```
}*/
```

```
public void map(Object key, Text value, Context context) throws IOException,  
InterruptedException
```

```
{
```

```
    // Path pt = new Path("hdfs:/ethonwu/A.txt");
```

```
    //FileSystem fs = FileSystem.get(test);
```

```
    //FSDDataOutputStream fsout = fs.create(pt);
```

```
    Map<String, List<String>> tmpList = new HashMap<String, List<String>>();
```

```
    StringTokenizer itr = new StringTokenizer(value.toString());
```

```
    while (itr.hasMoreTokens())
```

```
    {
```

```
        t = itr.nextToken().toString();
```

```
        int list_l = t.length();
```

```
        StringTokenizer st = new StringTokenizer(t, ",");
```

```
        int[] list = new int[list_l];
```

```
        int k = 0;
```

```
while (st.hasMoreTokens())
{
    // Word count here

    // word.set(st.nextToken().toString());

    //context.write(word, one);

    list[k] = Integer.valueOf(st.nextToken());
    k++;
}

for(int i=1;i<k;i++) {
    List<String> people = new ArrayList<String>();
}

/*

for(int i=1;i<=k;i++) {
    word.set(Integer.toString(i));
    context.write(word, one);
}*/
```

```
// String[][] Subests_Array = new String[k][k];

// int[] Array_index = new int[k];

//for(int index = 0 ; index<k ; index++) {
    //Array_index[index] = 0;
//}

int n = k;

subsets = new String();

char tmp = '\0';

// for(int l = 0 ; l<k ; l++) {
for (int i = 0; i < (1<<n); i++)
{
    subsets = new String();

    subsets = "";

    int flag = 0;

    tmp = '\0';

    for (int j = 0; j < n; j++)
    {

        tmp = '\0';

        if ((i & (1 << j)) > 0)
        {
```

```
if(subsets=="")
{
    subsets = subsets + Integer.toString(list[j]);
}

else
{
    subsets = subsets + "," + Integer.toString(list[j]);
    flag++;
}

}

}

if (!"".equals(subsets)) {
    // if(l==flag) {
        word.set(subsets);
        // one.set(flag+1);
    }
```

```
// fsout.writeUTF(subsets);
```

```
//fsout.flush();
```

```
//fsout.sync();
```

```
//fsout.close();
```

```
}
```

```
}
```

```
public static class IntSumCombiner extends
```

```
Reducer<Text,IntWritable,Text,IntWritable>
```

```
{
```

```
    private IntWritable pre_result = new IntWritable();
```

```
    public void reduce(Text key, Iterable<IntWritable> values,Context  
context)throws IOException, InterruptedException
```

```
{
```

```
int sum = 0;

for (IntWritable val : values)
{
    context.write(key,new IntWritable(val.get()));
}

// if (sum >= 3) {
// pre_result.set(sum);
//context.write(key, pre_result);
//}

}

}
```

```
public static class IntSumReducer extends Reducer<Text,IntWritable,Text,IntWritable>
{
    private IntWritable result = new IntWritable();
```

```
private Text halfresult = new Text();  
  
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 >= 2)
```

```
    {
```

```
        result.set(sum);  
  
        context.write(key, result);  
    }  
  
}
```

```
    }  
}
```

```
public static void main(String[] args) throws Exception {  
    Configuration conf = new Configuration();  
  
    // TokenizerMapper.test = conf;  
  
    String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs();  
    if (otherArgs.length != 2) {  
        System.err.println("Usage: wordcount <in> <out>");  
        System.exit(2);  
    }  
  
    Job job = new Job(conf, "word count");  
  
    //New for in MR job  
  
    /*  
  
    Path inPath = new Path(args[0]);  
    Path outputPath = null;  
    for(int i= 0;i<=2 ; i++)  
    {  
        outputPath = new Path(args[1]+i);
```

— Job job = new Job(conf, "word count");

```
job.setJarByClass(Apriori.class);  
job.setMapperClass(TokenizerMapper.class);  
job.setReducerClass(IntSumReducer.class);  
job.setOutputKeyClass(Text.class);  
job.setOutputValueClass(IntWritable.class);  
FileInputFormat.addInputPath(job,inPath);  
FileOutputFormat.setOutputPath(job,outPath);  
job.waitForCompletion(true);  
  
inPath = outPath;
```

```
}
```

```
*/
```

```
//end
```

```
Path outputPath = new Path(args[1]);  
outputPath.getFileSystem(conf).delete(outputPath, true);  
job.setJarByClass(Apriori.class);
```

```
job.setMapperClass(TokenizerMapper.class);  
job.setCombinerClass(IntSumCombiner.class);  
job.setReducerClass(IntSumReducer.class);
```

```
job.setOutputKeyClass(Text.class);  
job.setOutputValueClass(IntWritable.class);
```

```
//job.setNumReduceTasks(3); FileInputFormat.addInputPath(job,  
new Path(otherArgs[0]));  
FileOutputFormat.setOutputPath(job, new Path(otherArgs[1]));
```

```
System.exit(job.waitForCompletion(true) ? 0 : 1);
```

```
/*
```

```
while(job.waitForCompletion(true)) { }
```

```
Path pt = new Path(args[1]); //Location on HDFS
FileSystem fs = FileSystem.get(new Configuration());

BufferedReader br = new BufferedReader(new InputStreamReader(fs.open(pt)));

String line;

line = br.readLine();

int[] list = new int[line.length()]; ArrayList<Integer>
key_list = new ArrayList<>();

ArrayList<Integer> value_list = new ArrayList<>();

int k=0;
while(line!=null)
{
    StringTokenizer st = new StringTokenizer(line, ",");
    while (st.hasMoreTokens())
    {
        list[k] = Integer.valueOf(st.nextToken());
```




DEPARTMENT OF ACADEMIC AFFAIRS

Discover. Learn. Empower.



```
        k++;  
    }  
    key_list.add(list[0]);  
    value_list.add(list[1]);  
  
    }  
Collections.sort(key_list);  
*/  
  
}  
}
```



5. Result/Output/Writing Summary:

```
aarti@aarti-VirtualBox:~/Desktop$ su - hadoop
Password:
hadoop@aarti-VirtualBox:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [0.0.0.0]
Starting datanodes
Starting secondary namenodes [aarti-VirtualBox]
Starting resource manager
Starting node managers
hadoop@aarti-VirtualBox:~$ hdfs dfs -ls /
Found 18 items
drwxr-xr-x - hadoop supergroup      0 2021-04-05 10:13 /aarti
drwxr-xr-x - hadoop supergroup      0 2021-03-31 19:32 /abc
drwxr-xr-x - hadoop supergroup      0 2021-03-31 19:49 /hello
drwxr-xr-x - hadoop supergroup      0 2021-04-12 11:16 /insert
drwxr-xr-x - hadoop supergroup      0 2021-04-05 18:18 /m1
drwxr-xr-x - hadoop supergroup      0 2021-04-07 12:26 /mac
drwxr-xr-x - hadoop supergroup      0 2021-04-05 17:55 /multiply
drwxr-xr-x - hadoop supergroup      0 2021-04-21 11:39 /opdatastream
drwxr-xr-x - hadoop supergroup      0 2021-04-07 12:28 /outputjoins
drwxr-xr-x - hadoop supergroup      0 2021-04-12 11:18 /outputsearch
drwxr-xr-x - hadoop supergroup      0 2021-04-21 11:35 /r1
drwxr-xr-x - hadoop supergroup      0 2021-04-05 10:24 /result
drwxr-xr-x - hadoop supergroup      0 2021-03-31 19:04 /result2
drwxr-xr-x - hadoop supergroup      0 2021-03-31 19:01 /rr1
drwxr-xr-x - hadoop supergroup      0 2021-04-12 10:58 /search
drwxr-xr-x - hadoop supergroup      0 2021-03-31 18:48 /stdnt
drwx----- hadoop supergroup      0 2021-03-31 18:50 /tmp
-rw-r--r-- 1 hadoop supergroup    4182 2021-03-31 19:49 /xyz
hadoop@aarti-VirtualBox:~$ hdfs dfs -put /home/aarti/Downloads/apriori.input /mac
hadoop@aarti-VirtualBox:~$ hdfs dfs -put /home/aarti/Downloads/FrequentSet.jar /mac
hadoop@aarti-VirtualBox:~$ hadoop jar /home/aarti/Downloads/FrequentSet.jar /mac/apriori.input /outputf
Exception in thread "main" java.lang.Error: Unresolved compilation problem:
    The type org.apache.commons.cli.Options cannot be resolved. It is indirectly referenced from required .class files

    at Apriori.main(Apriori.java:209)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
```



DEPARTMENT OF ACADEMIC AFFAIRS

Discover. Learn. Empower.

**NAAC
GRADE A+**
ACCREDITED UNIVERSITY

<

>

Home

Downloads

Q

≡

▼

≡

—

□

×

Recent

★ Starred

Home

Desktop

Documents

Downloads

Music

Pictures

Videos

Trash

+ Other Locations

eclipse-installer

ReduceJoin.jar

hadoop-3.2.2.tar.gz

Frequent.jar

cust.txt

FrequentSet.jar

tnxn.txt

apriori.Input

Reduce join.txt

apriori algo(1)

mapred.docx

apriori algo

one step matrix multiply.txt

Driver.jar

hadoop-mapreduce-client-co...

Searching.jar

hadoop-common-3.2.2.jar

search5.txt

eclipse-inst-jre-linux64.tar.gz

Searching(1).txt

sort.docx



DEPARTMENT OF ACADEMIC AFFAIRS

Discover. Learn. Empower.



```
apriori.input
~/Downloads

1 100 1,2,5
2 200 2,4
3 300 2,3
4 400 1,2,4
5 500 1,3
6 600 2,3
7 700 1,3
8 800 1,2,3,5
9 900 1,2,3
```

Plain Text ▾ Tab Width: 8 ▾ Ln 1, Col 1 ▾ INS



DEPARTMENT OF ACADEMIC AFFAIRS

Discover. Learn. Empower.

**NAAC
GRADE A+**
ACCREDITED UNIVERSITY

eclipse-workspace - FrequentSet/src/Apriori.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

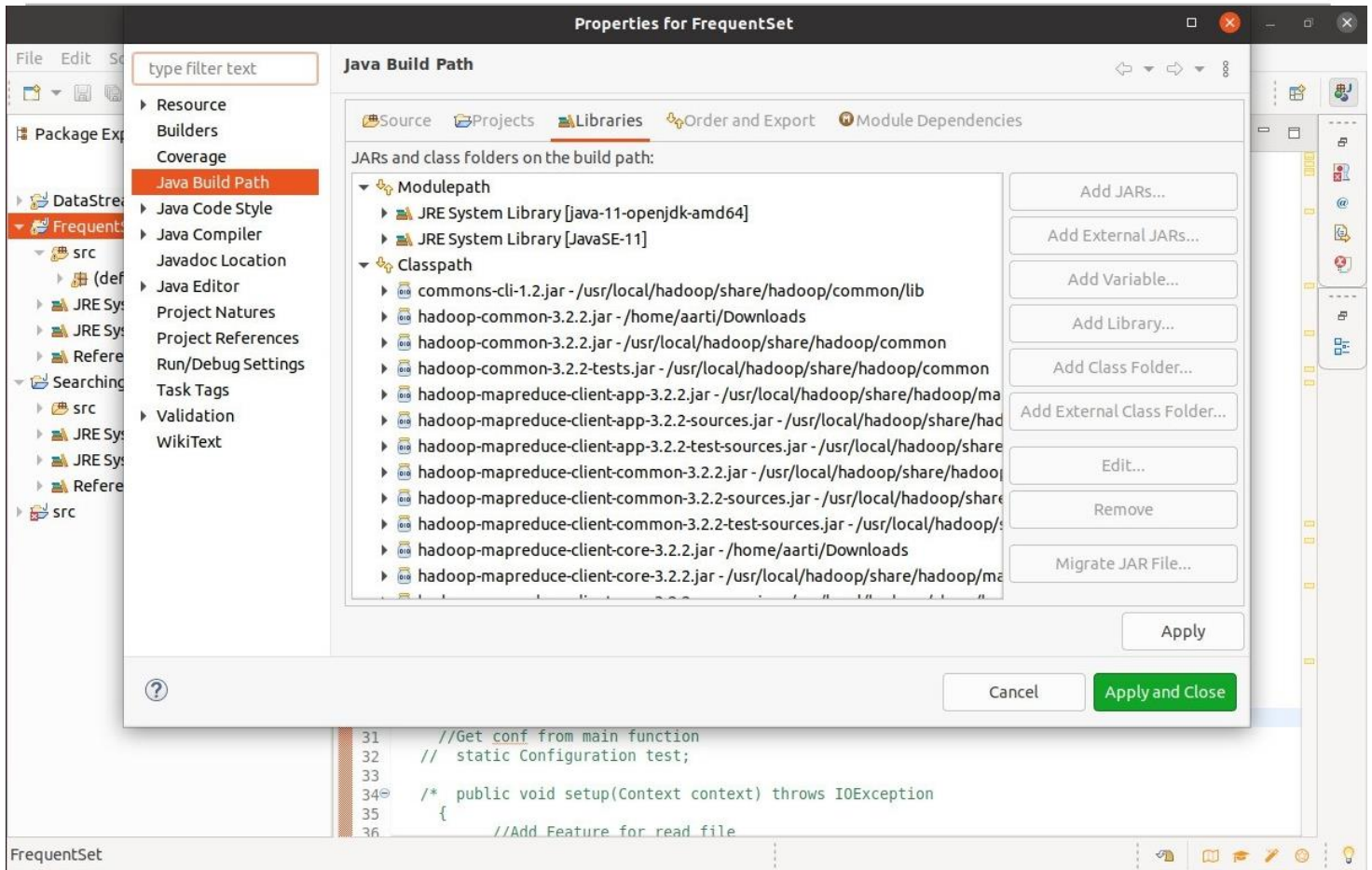
Package Explorer

- DataStream
- FrequentSet
 - src
 - (default package)
- JRE System Library [java-11-openjdk]
- JRE System Library [JavaSE-11]
- Referenced Libraries
- Searching
 - src
 - JRE System Library [java-11-openjdk]
 - JRE System Library [JavaSE-11]
 - Referenced Libraries
- src

Apriori.java

```
1 import java.io.BufferedReader;
2 import java.util.*;
3 import java.io.InputStreamReader;
4 import java.io.IOException;
5 import java.util.ArrayList;
6 import java.util.Collections;
7 import java.util.StringTokenizer;
8 import org.apache.hadoop.conf.Configuration;
9 import org.apache.hadoop.fs.FSDataOutputStream;
10 import org.apache.hadoop.fs.FileSystem;
11 import org.apache.hadoop.fs.Path;
12 import org.apache.hadoop.io.IntWritable;
13 import org.apache.hadoop.io.Text;
14 import org.apache.hadoop.mapreduce.Job;
15 import org.apache.hadoop.mapreduce.Mapper;
16 import org.apache.hadoop.mapreduce.Reducer;
17 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
18 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
19 import org.apache.hadoop.util.GenericOptionsParser;
20
21
22 public class Apriori {
23     public static class TokenizerMapper extends Mapper<Object, Text, Text, IntWritable>
24     {
25         private final static IntWritable one = new IntWritable();
26         private Text word = new Text();
27         private String t = new String();
28         private String subsets = new String();
29
30         //Get conf from main function
31         // static Configuration test;
32
33         /* public void setup(Context context) throws IOException
34         {
35             //Add Feature for read file
36
```

Writable Smart Insert 30 : 6 : 1058



Properties for FrequentSet

type filter text

Package Explorer: Package Explorer, DataStream, FrequentSet, src, (def), JRE Sy, JRE Sy, JRE Sy, Refer, Searching, src, JRE Sy, JRE Sy, Refer, src

Left Panel: Resource, Builders, Coverage, **Java Build Path**, Java Code Style, Java Compiler, Javadoc Location, Java Editor, Project Natures, Project References, Run/Debug Settings, Task Tags, Validation, WikiText

Java Build Path

Source Projects Libraries Order and Export Module Dependencies

JARs and class folders on the build path:

- Modulepath
 - JRE System Library [java-11-openjdk-amd64]
 - JRE System Library [JavaSE-11]
- Classpath
 - commons-cli-1.2.jar - /usr/local/hadoop/share/hadoop/common/lib
 - hadoop-common-3.2.2.jar - /home/aarti/Downloads
 - hadoop-common-3.2.2.jar - /usr/local/hadoop/share/hadoop/common
 - hadoop-common-3.2.2-tests.jar - /usr/local/hadoop/share/hadoop/common
 - hadoop-mapreduce-client-app-3.2.2.jar - /usr/local/hadoop/share/hadoop/m
 - hadoop-mapreduce-client-app-3.2.2-sources.jar - /usr/local/hadoop/share/had
 - hadoop-mapreduce-client-app-3.2.2-test-sources.jar - /usr/local/hadoop/share
 - hadoop-mapreduce-client-common-3.2.2.jar - /usr/local/hadoop/share/hadoop
 - hadoop-mapreduce-client-common-3.2.2-sources.jar - /usr/local/hadoop/share
 - hadoop-mapreduce-client-common-3.2.2-test-sources.jar - /usr/local/hadoop/s
 - hadoop-mapreduce-client-core-3.2.2.jar - /home/aarti/Downloads
 - hadoop-mapreduce-client-core-3.2.2.jar - /usr/local/hadoop/share/hadoop/ma

Buttons: Add JARs..., Add External JARs..., Add Variable..., Add Library..., Add Class Folder..., Add External Class Folder..., Edit..., Remove, Migrate JAR File..., Apply, Cancel, Apply and Close

Editor: FrequentSet

```

31 //Get conf from main function
32 // static Configuration test;
33
34 /* public void setup(Context context) throws IOException
35 {
36 //Add Feature for read file
  
```



DEPARTMENT OF ACADEMIC AFFAIRS

Discover. Learn. Empower.



```
at java.base/java.lang.reflect.Method.invoke(Method.java:566)
at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
hadoop@aarti-VirtualBox:~$ hdfs dfs -put /home/aarti/Downloads/apriori.input /r1
hadoop@aarti-VirtualBox:~$ hdfs dfs -put /home/aarti/Downloads/Frequent.jar /r1
hadoop@aarti-VirtualBox:~$ hadoop jar /home/aarti/Downloads/Frequent.jar /r1/apriori.input /output2
2021-04-28 12:10:00,853 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-04-28 12:10:03,372 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1619590008435_0001
2021-04-28 12:10:04,722 INFO input.FileInputFormat: Total input files to process : 1
2021-04-28 12:10:05,257 INFO mapreduce.JobSubmitter: number of splits:1
2021-04-28 12:10:06,702 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1619590008435_0001
2021-04-28 12:10:06,704 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-04-28 12:10:07,392 INFO conf.Configuration: resource-types.xml not found
2021-04-28 12:10:07,392 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-04-28 12:10:08,483 INFO impl.YarnClientImpl: Submitted application application_1619590008435_0001
2021-04-28 12:10:08,712 INFO mapreduce.Job: The url to track the job: http://aarti-VirtualBox:8088/proxy/application_1619590008435_0001/
2021-04-28 12:10:08,713 INFO mapreduce.Job: Running job: job_1619590008435_0001
2021-04-28 12:10:50,306 INFO mapreduce.Job: Job job_1619590008435_0001 running in uber mode : false
2021-04-28 12:10:50,320 INFO mapreduce.Job: map 0% reduce 0%
2021-04-28 12:11:13,215 INFO mapreduce.Job: map 100% reduce 0%
2021-04-28 12:11:37,718 INFO mapreduce.Job: map 100% reduce 100%
2021-04-28 12:11:39,843 INFO mapreduce.Job: Job job_1619590008435_0001 completed successfully
2021-04-28 12:11:40,041 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=578
  FILE: Number of bytes written=470635
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=186
  HDFS: Number of bytes written=72
  HDFS: Number of read operations=8
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
```



DEPARTMENT OF ACADEMIC AFFAIRS

Discover. Learn. Empower.



```
Merged Map outputs=1
GC time elapsed (ms)=239
CPU time spent (ms)=2100
Physical memory (bytes) snapshot=378957824
Virtual memory (bytes) snapshot=5343297536
Total committed heap usage (bytes)=230821888
Peak Map Physical memory (bytes)=240390144
Peak Map Virtual memory (bytes)=2666020864
Peak Reduce Physical memory (bytes)=138567680
Peak Reduce Virtual memory (bytes)=2677276672

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=85
File Output Format Counters
Bytes Written=72

hadoop@aarti-VirtualBox:~$ hdfs dfs -ls /output2
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-04-28 12:11 /output2/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 72 2021-04-28 12:11 /output2/part-r-000000
hadoop@aarti-VirtualBox:~$ hdfs dfs -cat /output2/part-r-000000
1 6
1,2 4
1,2,3 2
1,2,5 2
1,3 4
1,5 2
2 7
2,3 4
2,4 2
2,5 2
3 6
4 2
5 2
hadoop@aarti-VirtualBox:~$
```

Learning outcomes (What I have learnt):

1. Creating jar file
2. Creating project file for apriori project
3. executing project in hadoop terminal
4. starting all nodes in cluster
5. implementing jar file to project



DEPARTMENT OF ACADEMIC AFFAIRS

Discover. Learn. Empower.



Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.	Worksheet		10
2.	Pre Lab		5
3.	Post Lab		5