ASSIGNMENT: CASE STUDY - 1

Problem Statement:

a. What are the movie titles that the user has rated?
b. How many times a movie has been rated by the user?
c. In question 2 above, what is the average rating given for a movie?

SOLUTION:

MOVIES MAPPER CODE

```
package acadgild:

import java.io.IOException;
import java.io.IOException;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class CaseStudyIUseCasesMoviesMapper extends Mapper<LongWritable,
Text, Text, Text> {

      public void map(LongWritable key, Text value, Context context)throws
IOException, InterruptedException {
      try {
            if (key.get() == 0 && value.toString().contains("movieId")){
            return;
            }
            else {
                  String record = value.toString();
                  String[] parts = record.split(",");
                  context.write(new Text(parts[0]), new Text("movies\t" +
parts[1]));
               }
      }
      catch (Exception e) {
                  e.printStackTrace();
         }
    }
}
```

RATINGS MAPPER CODE

```
package acadgild:

import java.io.IOException;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class CaseStudyIUseCasesRatingsMapper extends Mapper<LongWritable,
```

```
Text, Text, Text> {

      public void map(LongWritable key, Text value, Context context)throws
IOException, InterruptedException {
      try {
            if (key.get() == 0 && value.toString().contains("userId")){
                  return;
            }
            else {
                  String record = value.toString();
                  String[] parts = record.split(",");
                  context.write(new Text(parts[1]), new Text("ratings\t" +
parts[2]));
            }
      }
      catch (Exception e) {
                  e.printStackTrace();
      }
     }
}
```

REDUCER CODE

```
package acadgild;

import java.io.IOException;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class CaseStudyIUseCasesReducer extends Reducer<Text, Text, Text,
Text> {

      public void reduce(Text key, Iterable<Text> values, Context
context)throws IOException, InterruptedException {
                  String titles = "";
                  double total = 0.0;
                  int count = 0;
            System.out.println("Text Key =>"+key.toString());
            for (Text t : values)
            {
                  String parts[] = t.toString().split("\t");
                  System.out.println("Text values =>"+t.toString());

               if (parts[0].equals("ratings")) {
                  count++;
                  String rating = parts[1].trim();
                  System.out.println("Rating is =>"+rating);
                  total += Double.parseDouble(rating);
               }
               else if (parts[0].equals("movies")) {
                  titles = parts[1];
               }
            }
```

```
            double average = total / count;

//for calculating average

      String str = String.format("Number of times rated = %d and
averagerated time = %f", count, average);

//String str = String.format("%d", count);

      context.write(new Text(titles), new Text(str));


      }
}
```

DRIVER CODE

```
package acadgild:

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.MultipleInputs;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class CaseStudyIUseCasesDriver {
@SuppressWarnings("deprecation")

      public static void main(String[] args) throws Exception
      {
            if (args.length != 3) {
                  System.err.println("Usage: CaseStudyIUseCase2Driver
<input path1> <input path2><output path>");
                  System.exit(-1);
            }

//Job Related Configurations
            Configuration conf = new Configuration();
            Job job = new Job(conf, "CaseStudyIUseCase2Driver");

            job.setJarByClass(CaseStudyIUseCasesDriver.class);
            job.setNumReduceTasks(2);

//Since there are multiple input, there is a slightly different way of
specifying inputpath, input format and mapper

            MultipleInputs.addInputPath(job, new
Path(args[0]),TextInputFormat.class,
            CaseStudyIUseCasesMoviesMapper.class);
```

```
            MultipleInputs.addInputPath(job, new
Path(args[1]),TextInputFormat.class,
            CaseStudyIUseCasesRatingsMapper.class);

//Set the reducer

            job.setReducerClass(CaseStudyIUseCasesReducer.class);

//set the out path

            Path outputPath = new Path(args[2]);
            FileOutputFormat.setOutputPath(job, outputPath);
            outputPath.getFileSystem(conf).delete(outputPath, true);

//set up the output key and value classes

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

//execute the job

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

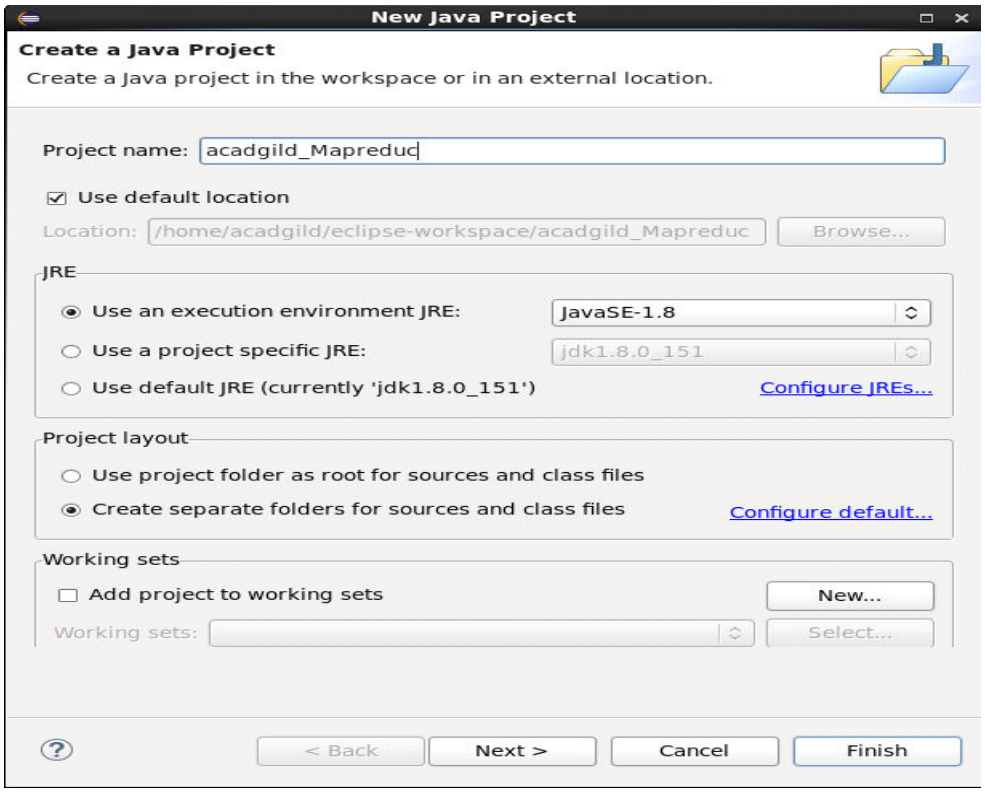
EXPLANATION: SO HERE WE NEED TO CREATE A MAP REDUCE PROGRAM TO CALCULATE "WHICH ARE THE MOVIE TITLES RATED BY USER", "HOW MANY TIMES THE MOVIE HAS BEEN RATED BY THE USER" AND "AVERAGE RATING GIVEN FOR THE MOVIE".

SO TO DO ABOVE TASKS FIRST DOWNLOADED THE INPUT FILE i.e OUR DATASET **"movies.csv" AND "ratings.csv."** FILE AND THEN USING "put" COMMAND COPIED THE FILE INTO HDFS DIRECTORY:-

PATH : "/user/acadgild/movies.csv"
PATH : "/user/acadgild/ratings.csv"

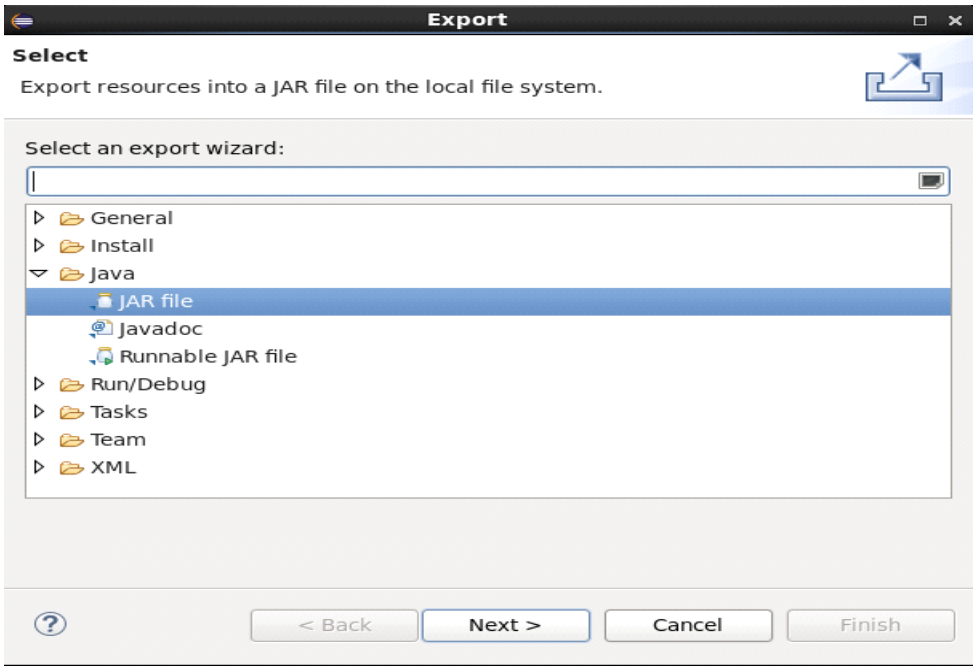
CODE: hadoop fs -put movies.csv ratings.csv /user/acadgild

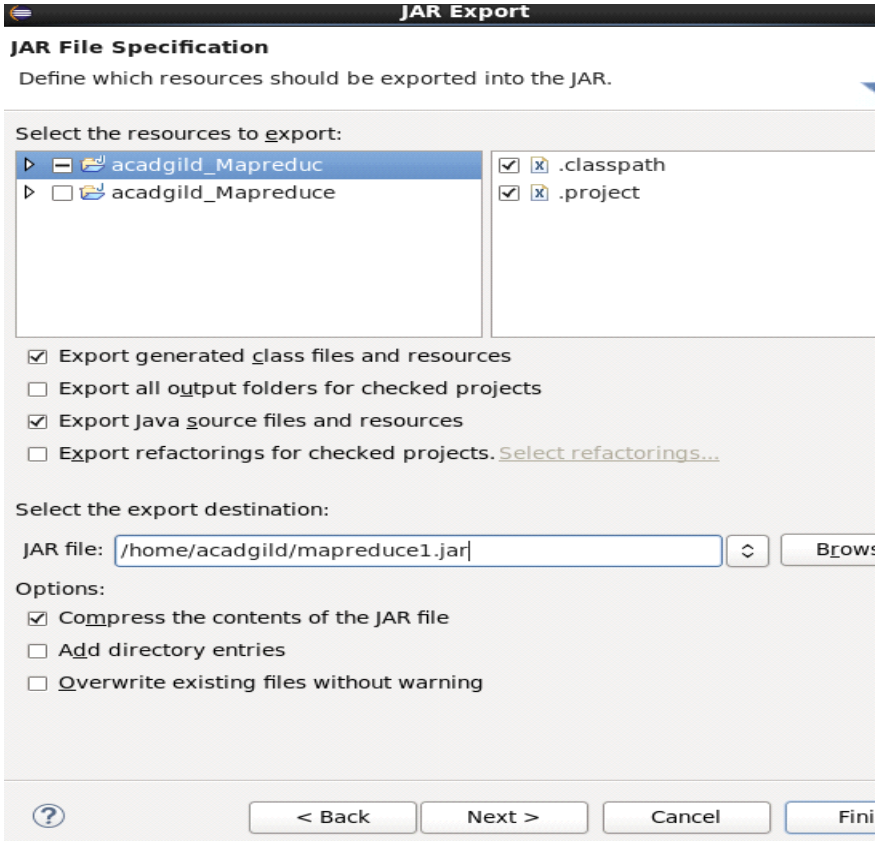
THEN CREATED A JAVA PROJECT IN ECLIPSE FOR THAT CLICK **"New --> Java Project --> ENTER NAME OF THE PROJECT --> Finish"** AS SHOWN IN THE SCREENSHOT.

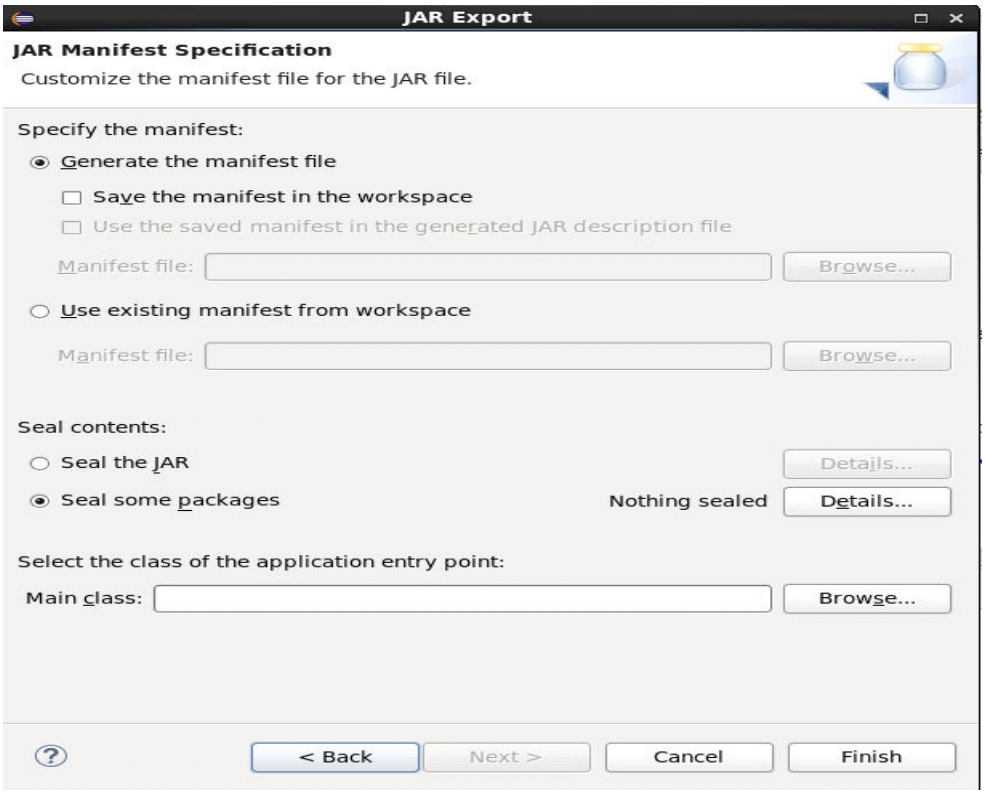


NOW U CAN SEE A JAVA PROJECT IS CREATED WHICH CAN BE SEEN ON LEFT HAND SIDE. NOW WE NEED TO CREATE A JAVA CLASS FOR OUR "MAPPER, DRIVER AND REDUCER CODE". SO WE RIGHT CLICK ON OUR CREATED JAVA PROJECT AND GO TO **"New --> CLASS --> ENTER NAME OF THE CLASS --> Finish"**. WE DO SAME TILL ALL THE REQUIRED CLASSES ARE MADE.

NOW ENTER THE CODE. AFTER ENTERING THE CODE U WILL SEE MANY ERRORS. ITS BECAUSE WE NEED TO ADD LIBRARIES i.e JAR FILES FOR OUR MAP-REDUCE ENVIRONMNET TO WORK. SO FOR THAT RIGHT CLICK ON THE CREATED JAVA PROJECT AND THEN CLICK ON **"Build Path --> Configure Build Path --. GO TO Libraries --> Add External JARS"** THEN BROWSE THE LOCATION AND THEN ADD IT.

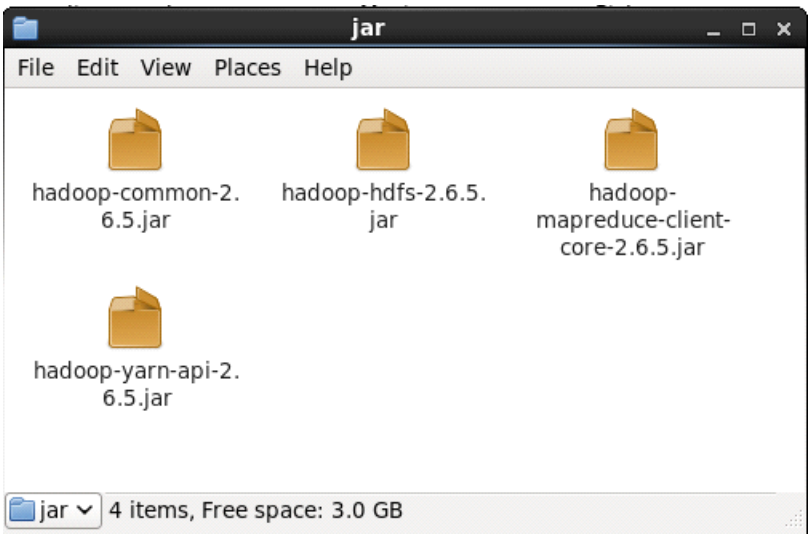

Export
Select
Export resources into a JAR file on the local file system.
Select an export wizard:
General
Install
Java
JAR file
Javadoc
Runnable JAR file
Run/Debug
Tasks
Team
XML
< Back
Next >
Cancel
Finish


JAR Export
JAR File Specification
Define which resources should be exported into the JAR.
Select the resources to export:
acadgild_Mapreduc
acadgild_Mapreduce
.classpath
.project
Export generated class files and resources
Export all output folders for checked projects
Export Java source files and resources
Export refactorings for checked projects. Select refactorings...
Select the export destination:
JAR file: /home/acadgild/mapreduce1.jar
Brows
Options:
Compress the contents of the JAR file
Add directory entries
Overwrite existing files without warning
< Back
Next >
Cancel
Fini



JAR FILES EXPORTED FROM PATH "/home/acadgild/jar".

THE JAR FILES REQUIRED ARE :



CAN BE DOWNLOADED FROM LINK "https://mvnrepository.com/artifact/org.apache.hadoop"

AFTER ADDING THE NECESSARY JAR FILES THE ERRORS ARE ELIMINATED.
NOW WE NEED TO EXPORT THESE JAVA CLASS AS A "JAR" FILE INTO OUR LINUX MACHINE SO THAT WE CAN RUN THE MAP REDUCE PROGRAM DIRECTLY THROUGH COMMANDLINE.

SO FOR CREATING A JAR FILE. SELECT BOTH THE CLASSES THEN RIGHT CLICK. THEN CLICK ON **"Export --> JAR FILE --> ENTER JAR FILE NAME --> NEXT --> THEN GO TO OPTION "MAIN CLASS" AND BROWSE THE CLASS NAME i.e "driver" --> FINISH.** WITH THIS OUR JAR FILE WITH NAME **"AdMapReduceNew.jar"** IS CREATED. USING THE BELOW CODE WE CAN CHECK THE JAR FILE CREATED. FIRST OPEN THE TERMINAL THEN ENTER:-

CODE: ls

SOLUTION REPORT:

```
[acadgild@localhost ~]$ ls
AdMapReduce.jar     commandline  Documents  eclipse
hadoop-common-2.6.5.jar  jar    new        new file~  project
sports.txt  Videos
AdMapReduceNew.jar  Desktop      Downloads  eclipse-workspace  install
Music  new file  Pictures   Public   Templates
```

EXPLANATION: NOW ITS TIME TO RUN OUR JAR FILE. SO WE USE THE BELOW CODE. THE OUTPUT WILL BE SAVED IN HDFS DIRECTORY.

PATH:"/user/acadgild/newOut"

CODE: ~]$ hadoop jar AdMapReduceNew.jar /user/acadgild/movies.csv /user/acadgild/ratings.csv /user/acadgild/newOut

SOLUTION REPORT:

```
[acadgild@localhost ~]$ hadoop jar AdMapReduceNew.jar
/user/acadgild/movies.csv /user/acadgild/ratings.csv
/user/acadgild/newOut
19/03/29 18:36:03 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
19/03/29 18:36:05 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
19/03/29 18:36:06 WARN mapreduce.JobResourceUploader: Hadoop command-line
option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
19/03/29 18:36:06 INFO input.FileInputFormat: Total input paths to process
: 1
19/03/29 18:36:06 INFO input.FileInputFormat: Total input paths to process
: 1
19/03/29 18:36:06 INFO mapreduce.JobSubmitter: number of splits:7
19/03/29 18:36:07 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1553853639219_0004
19/03/29 18:36:07 INFO impl.YarnClientImpl: Submitted application
application_1553853639219_0004
19/03/29 18:36:07 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1553853639219_0004/
19/03/29 18:36:07 INFO mapreduce.Job: Running job: job_1553853639219_0004
19/03/29 18:36:17 INFO mapreduce.Job: Job job_1553853639219_0004 running
in uber mode : false
19/03/29 18:36:17 INFO mapreduce.Job:  map 0% reduce 0%
19/03/29 18:36:51 INFO mapreduce.Job:  map 2% reduce 0%
```

```
19/03/29 18:36:53 INFO mapreduce.Job:  map 6% reduce 0%
19/03/29 18:36:56 INFO mapreduce.Job:  map 9% reduce 0%
19/03/29 18:36:57 INFO mapreduce.Job:  map 10% reduce 0%
19/03/29 18:36:58 INFO mapreduce.Job:  map 11% reduce 0%
19/03/29 18:36:59 INFO mapreduce.Job:  map 15% reduce 0%
19/03/29 18:37:00 INFO mapreduce.Job:  map 16% reduce 0%
19/03/29 18:37:01 INFO mapreduce.Job:  map 18% reduce 0%
19/03/29 18:37:02 INFO mapreduce.Job:  map 20% reduce 0%
19/03/29 18:37:03 INFO mapreduce.Job:  map 22% reduce 0%
19/03/29 18:37:04 INFO mapreduce.Job:  map 24% reduce 0%
19/03/29 18:37:05 INFO mapreduce.Job:  map 25% reduce 0%
19/03/29 18:37:06 INFO mapreduce.Job:  map 26% reduce 0%
19/03/29 18:37:07 INFO mapreduce.Job:  map 28% reduce 0%
19/03/29 18:37:08 INFO mapreduce.Job:  map 29% reduce 0%
19/03/29 18:37:09 INFO mapreduce.Job:  map 30% reduce 0%
19/03/29 18:37:10 INFO mapreduce.Job:  map 32% reduce 0%
19/03/29 18:37:11 INFO mapreduce.Job:  map 33% reduce 0%
19/03/29 18:37:13 INFO mapreduce.Job:  map 35% reduce 0%
19/03/29 18:37:14 INFO mapreduce.Job:  map 36% reduce 0%
19/03/29 18:37:17 INFO mapreduce.Job:  map 37% reduce 0%
19/03/29 18:37:18 INFO mapreduce.Job:  map 38% reduce 0%
19/03/29 18:37:21 INFO mapreduce.Job:  map 44% reduce 0%
19/03/29 18:37:24 INFO mapreduce.Job:  map 45% reduce 0%
19/03/29 18:37:41 INFO mapreduce.Job:  map 46% reduce 0%
19/03/29 18:37:44 INFO mapreduce.Job:  map 47% reduce 0%
19/03/29 18:37:45 INFO mapreduce.Job:  map 49% reduce 0%
19/03/29 18:37:46 INFO mapreduce.Job:  map 51% reduce 0%
19/03/29 18:37:47 INFO mapreduce.Job:  map 54% reduce 0%
19/03/29 18:37:48 INFO mapreduce.Job:  map 55% reduce 0%
19/03/29 18:37:49 INFO mapreduce.Job:  map 56% reduce 0%
19/03/29 18:37:50 INFO mapreduce.Job:  map 57% reduce 0%
19/03/29 18:37:51 INFO mapreduce.Job:  map 59% reduce 0%
19/03/29 18:37:53 INFO mapreduce.Job:  map 60% reduce 0%
19/03/29 18:37:56 INFO mapreduce.Job:  map 75% reduce 0%
19/03/29 18:37:57 INFO mapreduce.Job:  map 76% reduce 0%
19/03/29 18:38:18 INFO mapreduce.Job:  map 77% reduce 0%
19/03/29 18:38:22 INFO mapreduce.Job:  map 78% reduce 0%
19/03/29 18:38:23 INFO mapreduce.Job:  map 79% reduce 0%
19/03/29 18:38:25 INFO mapreduce.Job:  map 80% reduce 0%
19/03/29 18:38:26 INFO mapreduce.Job:  map 81% reduce 0%
19/03/29 18:38:28 INFO mapreduce.Job:  map 83% reduce 0%
19/03/29 18:38:32 INFO mapreduce.Job:  map 84% reduce 0%
19/03/29 18:38:33 INFO mapreduce.Job:  map 86% reduce 0%
19/03/29 18:38:36 INFO mapreduce.Job:  map 88% reduce 0%
19/03/29 18:38:39 INFO mapreduce.Job:  map 91% reduce 0%
19/03/29 18:38:42 INFO mapreduce.Job:  map 94% reduce 0%
19/03/29 18:38:45 INFO mapreduce.Job:  map 96% reduce 0%
19/03/29 18:38:48 INFO mapreduce.Job:  map 97% reduce 0%
19/03/29 18:38:50 INFO mapreduce.Job:  map 99% reduce 0%
19/03/29 18:38:53 INFO mapreduce.Job:  map 100% reduce 0%
19/03/29 18:39:11 INFO mapreduce.Job:  map 100% reduce 43%
19/03/29 18:39:18 INFO mapreduce.Job:  map 100% reduce 46%
19/03/29 18:39:21 INFO mapreduce.Job:  map 100% reduce 67%
```

```
19/03/29 18:39:27 INFO mapreduce.Job:  map 100% reduce 68%
19/03/29 18:39:36 INFO mapreduce.Job:  map 100% reduce 69%
19/03/29 18:39:45 INFO mapreduce.Job:  map 100% reduce 70%
19/03/29 18:39:54 INFO mapreduce.Job:  map 100% reduce 71%
19/03/29 18:40:04 INFO mapreduce.Job:  map 100% reduce 72%
19/03/29 18:40:13 INFO mapreduce.Job:  map 100% reduce 73%
19/03/29 18:40:22 INFO mapreduce.Job:  map 100% reduce 74%
19/03/29 18:40:28 INFO mapreduce.Job:  map 100% reduce 75%
19/03/29 18:40:37 INFO mapreduce.Job:  map 100% reduce 76%
19/03/29 18:40:46 INFO mapreduce.Job:  map 100% reduce 77%
19/03/29 18:40:55 INFO mapreduce.Job:  map 100% reduce 78%
19/03/29 18:41:04 INFO mapreduce.Job:  map 100% reduce 79%
19/03/29 18:41:13 INFO mapreduce.Job:  map 100% reduce 80%
19/03/29 18:41:22 INFO mapreduce.Job:  map 100% reduce 81%
19/03/29 18:41:31 INFO mapreduce.Job:  map 100% reduce 82%
19/03/29 18:41:38 INFO mapreduce.Job:  map 100% reduce 83%
19/03/29 18:41:47 INFO mapreduce.Job:  map 100% reduce 84%
19/03/29 18:41:56 INFO mapreduce.Job:  map 100% reduce 85%
19/03/29 18:42:05 INFO mapreduce.Job:  map 100% reduce 86%
19/03/29 18:42:14 INFO mapreduce.Job:  map 100% reduce 87%
19/03/29 18:42:23 INFO mapreduce.Job:  map 100% reduce 88%
19/03/29 18:42:29 INFO mapreduce.Job:  map 100% reduce 89%
19/03/29 18:42:38 INFO mapreduce.Job:  map 100% reduce 90%
19/03/29 18:42:47 INFO mapreduce.Job:  map 100% reduce 91%
19/03/29 18:42:56 INFO mapreduce.Job:  map 100% reduce 92%
19/03/29 18:43:05 INFO mapreduce.Job:  map 100% reduce 93%
19/03/29 18:43:14 INFO mapreduce.Job:  map 100% reduce 94%
19/03/29 18:43:20 INFO mapreduce.Job:  map 100% reduce 95%
19/03/29 18:43:30 INFO mapreduce.Job:  map 100% reduce 96%
19/03/29 18:43:39 INFO mapreduce.Job:  map 100% reduce 97%
19/03/29 18:43:48 INFO mapreduce.Job:  map 100% reduce 98%
19/03/29 18:43:58 INFO mapreduce.Job:  map 100% reduce 99%
19/03/29 18:44:07 INFO mapreduce.Job:  map 100% reduce 100%
19/03/29 18:44:10 INFO mapreduce.Job: Job job_1553853639219_0004 completed
successfully
19/03/29 18:44:10 INFO mapreduce.Job: Counters: 50
      File System Counters
            FILE: Number of bytes read=961719024
            FILE: Number of bytes written=1457594135
            FILE: Number of read operations=0
            FILE: Number of large read operations=0
            FILE: Number of write operations=0
            HDFS: Number of bytes read=711856049
            HDFS: Number of bytes written=3869465
            HDFS: Number of read operations=27
            HDFS: Number of large read operations=0
            HDFS: Number of write operations=4
      Job Counters
            Killed map tasks=2
            Launched map tasks=9
            Launched reduce tasks=2
            Data-local map tasks=9
            Total time spent by all maps in occupied slots (ms)=898683
```

```
        Total time spent by all reduces in occupied slots (ms)=629178
        Total time spent by all map tasks (ms)=898683
        Total time spent by all reduce tasks (ms)=629178
        Total vcore-milliseconds taken by all map tasks=898683
        Total vcore-milliseconds taken by all reduce tasks=629178
        Total megabyte-milliseconds taken by all map tasks=920251392
        Total megabyte-milliseconds taken by all reduce tasks=644278272
    Map-Reduce Framework
        Map input records=26070134
        Map output records=26070132
        Map output bytes=442789828
        Map output materialized bytes=494930183
        Input split bytes=1832
        Combine input records=0
        Combine output records=0
        Reduce input groups=45843
        Reduce shuffle bytes=494930183
        Reduce input records=26070132
        Reduce output records=45843
        Spilled Records=76775523
        Shuffled Maps =14
        Failed Shuffles=0
        Merged Map outputs=14
        GC time elapsed (ms)=10522
        CPU time spent (ms)=369300
        Physical memory (bytes) snapshot=1710927872
        Virtual memory (bytes) snapshot=18509070336
        Total committed heap usage (bytes)=1242103808
    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=0
    File Output Format Counters
        Bytes Written=3869465
```

```
[acadgild@localhost ~]$ hadoop jar AdMapReduceNew.jar /user/acadgild/movies.csv /user/acadgild/ratings.csv /user/acadgild/newOut
19/03/29 18:36:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/03/29 18:36:05 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
19/03/29 18:36:06 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
19/03/29 18:36:06 INFO input.FileInputFormat: Total input paths to process : 1
19/03/29 18:36:06 INFO input.FileInputFormat: Total input paths to process : 1
19/03/29 18:36:06 INFO mapreduce.JobSubmitter: number of splits:7
19/03/29 18:36:07 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1553853639219_0004
19/03/29 18:36:07 INFO impl.YarnClientImpl: Submitted application application_1553853639219_0004
19/03/29 18:36:07 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1553853639219_0004/
19/03/29 18:36:07 INFO mapreduce.Job: Running job: job_1553853639219_0004
19/03/29 18:36:17 INFO mapreduce.Job: Job job_1553853639219_0004 running in uber mode : false
19/03/29 18:36:17 INFO mapreduce.Job:  map 0% reduce 0%
19/03/29 18:36:51 INFO mapreduce.Job:  map 2% reduce 0%
19/03/29 18:36:53 INFO mapreduce.Job:  map 6% reduce 0%
19/03/29 18:36:56 INFO mapreduce.Job:  map 9% reduce 0%
19/03/29 18:36:57 INFO mapreduce.Job:  map 10% reduce 0%
19/03/29 18:36:58 INFO mapreduce.Job:  map 11% reduce 0%
19/03/29 18:36:59 INFO mapreduce.Job:  map 15% reduce 0%
19/03/29 18:37:00 INFO mapreduce.Job:  map 16% reduce 0%
19/03/29 18:37:01 INFO mapreduce.Job:  map 18% reduce 0%
19/03/29 18:37:02 INFO mapreduce.Job:  map 20% reduce 0%
19/03/29 18:37:03 INFO mapreduce.Job:  map 22% reduce 0%
19/03/29 18:37:04 INFO mapreduce.Job:  map 24% reduce 0%
19/03/29 18:37:05 INFO mapreduce.Job:  map 25% reduce 0%
19/03/29 18:37:06 INFO mapreduce.Job:  map 26% reduce 0%
19/03/29 18:37:07 INFO mapreduce.Job:  map 28% reduce 0%
19/03/29 18:37:08 INFO mapreduce.Job:  map 29% reduce 0%
19/03/29 18:37:09 INFO mapreduce.Job:  map 30% reduce 0%
19/03/29 18:37:10 INFO mapreduce.Job:  map 32% reduce 0%
19/03/29 18:37:11 INFO mapreduce.Job:  map 33% reduce 0%
19/03/29 18:37:13 INFO mapreduce.Job:  map 35% reduce 0%
19/03/29 18:37:14 INFO mapreduce.Job:  map 36% reduce 0%
19/03/29 18:37:17 INFO mapreduce.Job:  map 37% reduce 0%
19/03/29 18:37:18 INFO mapreduce.Job:  map 38% reduce 0%
19/03/29 18:37:21 INFO mapreduce.Job:  map 44% reduce 0%
19/03/29 18:37:24 INFO mapreduce.Job:  map 45% reduce 0%
19/03/29 18:37:41 INFO mapreduce.Job:  map 46% reduce 0%
19/03/29 18:37:44 INFO mapreduce.Job:  map 47% reduce 0%
19/03/29 18:37:45 INFO mapreduce.Job:  map 49% reduce 0%
19/03/29 18:37:46 INFO mapreduce.Job:  map 51% reduce 0%
19/03/29 18:37:47 INFO mapreduce.Job:  map 54% reduce 0%
19/03/29 18:37:48 INFO mapreduce.Job:  map 55% reduce 0%
19/03/29 18:37:49 INFO mapreduce.Job:  map 56% reduce 0%
19/03/29 18:37:50 INFO mapreduce.Job:  map 57% reduce 0%
19/03/29 18:37:51 INFO mapreduce.Job:  map 59% reduce 0%
19/03/29 18:37:53 INFO mapreduce.Job:  map 60% reduce 0%
19/03/29 18:37:56 INFO mapreduce.Job:  map 75% reduce 0%
19/03/29 18:37:57 INFO mapreduce.Job:  map 76% reduce 0%
19/03/29 18:38:18 INFO mapreduce.Job:  map 77% reduce 0%
19/03/29 18:38:22 INFO mapreduce.Job:  map 78% reduce 0%
19/03/29 18:38:23 INFO mapreduce.Job:  map 79% reduce 0%
19/03/29 18:38:25 INFO mapreduce.Job:  map 80% reduce 0%
19/03/29 18:38:26 INFO mapreduce.Job:  map 81% reduce 0%
19/03/29 18:38:28 INFO mapreduce.Job:  map 83% reduce 0%
19/03/29 18:38:32 INFO mapreduce.Job:  map 84% reduce 0%
19/03/29 18:38:33 INFO mapreduce.Job:  map 86% reduce 0%
19/03/29 18:38:36 INFO mapreduce.Job:  map 88% reduce 0%
19/03/29 18:38:39 INFO mapreduce.Job:  map 91% reduce 0%
19/03/29 18:38:42 INFO mapreduce.Job:  map 94% reduce 0%
19/03/29 18:38:45 INFO mapreduce.Job:  map 96% reduce 0%
19/03/29 18:38:48 INFO mapreduce.Job:  map 97% reduce 0%
19/03/29 18:38:50 INFO mapreduce.Job:  map 99% reduce 0%
19/03/29 18:38:53 INFO mapreduce.Job:  map 100% reduce 0%
19/03/29 18:39:11 INFO mapreduce.Job:  map 100% reduce 43%
19/03/29 18:39:18 INFO mapreduce.Job:  map 100% reduce 46%
19/03/29 18:39:21 INFO mapreduce.Job:  map 100% reduce 67%
19/03/29 18:39:27 INFO mapreduce.Job:  map 100% reduce 68%
19/03/29 18:39:36 INFO mapreduce.Job:  map 100% reduce 69%
19/03/29 18:39:45 INFO mapreduce.Job:  map 100% reduce 70%
19/03/29 18:39:54 INFO mapreduce.Job:  map 100% reduce 71%
19/03/29 18:40:04 INFO mapreduce.Job:  map 100% reduce 72%
19/03/29 18:40:13 INFO mapreduce.Job:  map 100% reduce 73%
19/03/29 18:40:22 INFO mapreduce.Job:  map 100% reduce 74%
19/03/29 18:40:28 INFO mapreduce.Job:  map 100% reduce 75%
19/03/29 18:40:37 INFO mapreduce.Job:  map 100% reduce 76%
```

```
19/03/29 18:40:46 INFO mapreduce.Job:  map 100% reduce 77%
19/03/29 18:40:55 INFO mapreduce.Job:  map 100% reduce 78%
19/03/29 18:41:04 INFO mapreduce.Job:  map 100% reduce 79%
19/03/29 18:41:13 INFO mapreduce.Job:  map 100% reduce 80%
19/03/29 18:41:22 INFO mapreduce.Job:  map 100% reduce 81%
19/03/29 18:41:31 INFO mapreduce.Job:  map 100% reduce 82%
19/03/29 18:41:38 INFO mapreduce.Job:  map 100% reduce 83%
19/03/29 18:41:47 INFO mapreduce.Job:  map 100% reduce 84%
19/03/29 18:41:56 INFO mapreduce.Job:  map 100% reduce 85%
19/03/29 18:42:05 INFO mapreduce.Job:  map 100% reduce 86%
19/03/29 18:42:14 INFO mapreduce.Job:  map 100% reduce 87%
19/03/29 18:42:23 INFO mapreduce.Job:  map 100% reduce 88%
19/03/29 18:42:29 INFO mapreduce.Job:  map 100% reduce 89%
19/03/29 18:42:38 INFO mapreduce.Job:  map 100% reduce 90%
19/03/29 18:42:47 INFO mapreduce.Job:  map 100% reduce 91%
19/03/29 18:42:56 INFO mapreduce.Job:  map 100% reduce 92%
19/03/29 18:43:05 INFO mapreduce.Job:  map 100% reduce 93%
19/03/29 18:43:14 INFO mapreduce.Job:  map 100% reduce 94%
19/03/29 18:43:20 INFO mapreduce.Job:  map 100% reduce 95%
19/03/29 18:43:30 INFO mapreduce.Job:  map 100% reduce 96%
19/03/29 18:43:39 INFO mapreduce.Job:  map 100% reduce 97%
19/03/29 18:43:48 INFO mapreduce.Job:  map 100% reduce 98%
19/03/29 18:43:58 INFO mapreduce.Job:  map 100% reduce 99%
19/03/29 18:44:07 INFO mapreduce.Job:  map 100% reduce 100%
19/03/29 18:44:10 INFO mapreduce.Job: Job job_1553853639219_0004 completed successfully
19/03/29 18:44:10 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=961719024
                FILE: Number of bytes written=1457594135
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=711856049
                HDFS: Number of bytes written=3869465
                HDFS: Number of read operations=27
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
        Job Counters
                Killed map tasks=2
                Total time spent by all reduces in occupied slots (ms)=629178
                Total time spent by all map tasks (ms)=898683
                Total time spent by all reduce tasks (ms)=629178
                Total vcore-milliseconds taken by all map tasks=898683
                Total vcore-milliseconds taken by all reduce tasks=629178
                Total megabyte-milliseconds taken by all map tasks=920251392
                Total megabyte-milliseconds taken by all reduce tasks=644278272
        Map-Reduce Framework
                Map input records=26070134
                Map output records=26070132
                Map output bytes=442789828
                Map output materialized bytes=494930183
                Input split bytes=1832
                Combine input records=0
                Combine output records=0
                Reduce input groups=45843
                Reduce shuffle bytes=494930183
                Reduce input records=26070132
                Reduce output records=45843
                Spilled Records=76775523
                Shuffled Maps =14
                Failed Shuffles=0
                Merged Map outputs=14
                GC time elapsed (ms)=10522
                CPU time spent (ms)=369300
                Physical memory (bytes) snapshot=1710927872
                Virtual memory (bytes) snapshot=18509070336
                Total committed heap usage (bytes)=1242103808
        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=0
        File Output Format Counters
                Bytes Written=3869465
```

EXPLANATION: WE CAN CHECK THE OUTPUT USING THE BELOW COMMAND(REFER THE SCREENSHOT)

CODE:

```
hadoop fs -ls /user/acadgild/newOut
hadoop fs -cat /user/acadgild/newOut/part-r-00000 | head
hadoop fs -cat /user/acadgild/newOut/part-r-00001 | head
```

SOLUTION REPORT:

```
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/newOut
19/03/29 18:48:28 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 3 items
-rw-r--r--   1 acadgild supergroup          0 2019-03-29 18:44
/user/acadgild/newOut/_SUCCESS
-rw-r--r--   1 acadgild supergroup    1945036 2019-03-29 18:44
/user/acadgild/newOut/part-r-00000
-rw-r--r--   1 acadgild supergroup    1924429 2019-03-29 18:43
/user/acadgild/newOut/part-r-00001

[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/newOut/part-r-00000
| head
19/03/29 18:49:04 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Toy Story (1995) Number of times rated = 66008 and averagerated time =
3.888157
GoldenEye (1995) Number of times rated = 32534 and averagerated time =
3.431841
City Hall (1996) Number of times rated = 4436 and averagerated time =
3.232304
Curdled (1996)  Number of times rated = 217 and averagerated time =
3.099078
First Daughter (1999) Number of times rated = 3 and averagerated time =
3.333333
"Flaw Number of times rated = 14 and averagerated time = 3.714286
Jason Becker: Not Dead Yet (2012)     Number of times rated = 9 and
averagerated time = 3.444444
Chicago Massacre: Richard Speck (2007)Number of times rated = 2 and
averagerated time = 2.500000
Keep the Lights On (2012)  Number of times rated = 25 and averagerated
time = 3.100000
True Blue (2001)      Number of times rated = 3 and averagerated time =
3.000000
cat: Unable to write to output stream.

[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/newOut/part-r-00001
| head
19/03/29 18:49:27 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
"Comic     Number of times rated = 1 and averagerated time = 4.000000
Up in Smoke (1957)    Number of times rated = 3 and averagerated time =
```

3.666667
Battle of Los Angeles (2011) Number of times rated = 44 and averagerated time = 2.522727
Beauty Is Embarrassing (2012) Number of times rated = 15 and averagerated time = 3.600000
Girl Model (2011) Number of times rated = 32 and averagerated time = 3.281250
Crossfire Hurricane (2012) Number of times rated = 18 and averagerated time = 3.388889
Middle of Nowhere (2012) Number of times rated = 11 and averagerated time = 3.454545
Enola Gay and the Atomic Bombing of Japan (1995) Number of times rated = 1 and averagerated time = 3.500000
Red Hook Summer (2012) Number of times rated = 11 and averagerated time = 2.045455
Stella Maris (1918) Number of times rated = 2 and averagerated time = 3.750000
cat: Unable to write to output stream.

```
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/newOut
19/03/29 18:48:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 3 items
-rw-r--r--   1 acadgild supergroup          0 2019-03-29 18:44 /user/acadgild/newOut/_SUCCESS
-rw-r--r--   1 acadgild supergroup    1945036 2019-03-29 18:44 /user/acadgild/newOut/part-r-00000
-rw-r--r--   1 acadgild supergroup    1924429 2019-03-29 18:43 /user/acadgild/newOut/part-r-00001
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/newOut/part-r-00000 | head
19/03/29 18:49:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Toy Story (1995)        Number of times rated = 66008 and averagerated time = 3.888157
GoldenEye (1995)        Number of times rated = 32534 and averagerated time = 3.431841
City Hall (1996)        Number of times rated = 4436 and averagerated time = 3.232304
Curdled (1996)  Number of times rated = 217 and averagerated time = 3.099078
First Daughter (1999)   Number of times rated = 3 and averagerated time = 3.333333
"Flaw   Number of times rated = 14 and averagerated time = 3.714286
Jason Becker: Not Dead Yet (2012)       Number of times rated = 9 and averagerated time = 3.444444
Chicago Massacre: Richard Speck (2007)  Number of times rated = 2 and averagerated time = 2.500000
Keep the Lights On (2012)       Number of times rated = 25 and averagerated time = 3.100000
True Blue (2001)        Number of times rated = 3 and averagerated time = 3.000000
cat: Unable to write to output stream.
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/newOut/part-r-00001 | head
19/03/29 18:49:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
"Comic  Number of times rated = 1 and averagerated time = 4.000000
Up in Smoke (1957)      Number of times rated = 3 and averagerated time = 3.666667
Battle of Los Angeles (2011)    Number of times rated = 44 and averagerated time = 2.522727
Beauty Is Embarrassing (2012)   Number of times rated = 15 and averagerated time = 3.600000
Girl Model (2011)       Number of times rated = 32 and averagerated time = 3.281250
Crossfire Hurricane (2012)      Number of times rated = 18 and averagerated time = 3.388889
Middle of Nowhere (2012)        Number of times rated = 11 and averagerated time = 3.454545
Enola Gay and the Atomic Bombing of Japan (1995)        Number of times rated = 1 and averagerated time = 3.500000
Red Hook Summer (2012)  Number of times rated = 11 and averagerated time = 2.045455
Stella Maris (1918)     Number of times rated = 2 and averagerated time = 3.750000
cat: Unable to write to output stream.
```