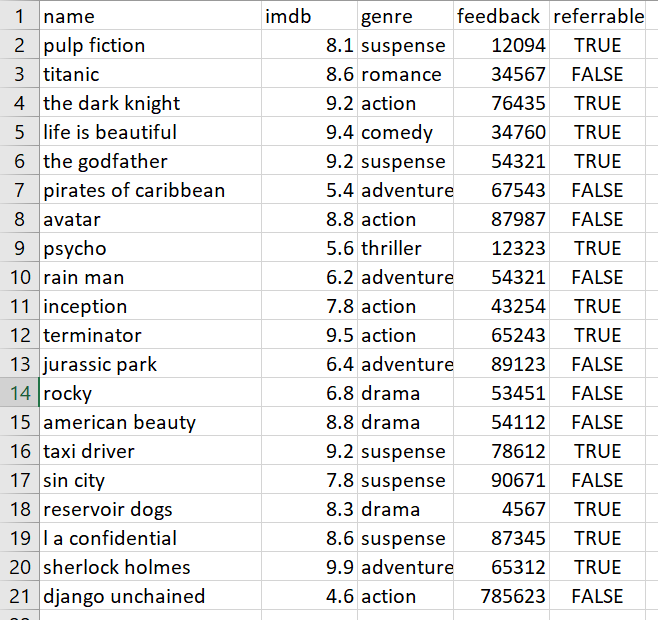
DATASET



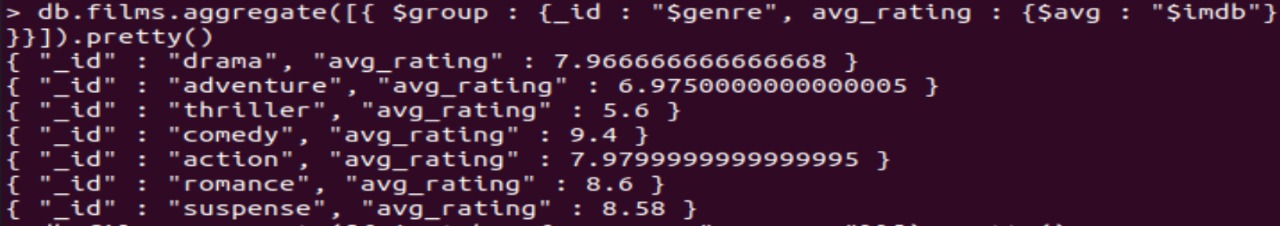
EXECUTION-DETAILS

MONGODB PART

> show dbs;  
Institute  0.000GB  
admin      0.000GB  
bd         0.000GB  
bdb        0.000GB  
bookdb     0.000GB  
config     0.000GB  
local      0.000GB  
movies     0.000GB  
myDb       0.000GB  
test       0.000GB  
> use movies  
switched to db movies  
> db.films.find().pretty()  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb4d"),  
"name" : "pulp fiction",  
"imdb" : 8.1,  
"genre" : "suspense",  
"feedback" : 12094,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb4e"),  
"name" : "titanic",  
"imdb" : 8.6,  
"genre" : "romance",  
"feedback" : 34567,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb4f"),  
"name" : "the dark knight",  
"imdb" : 9.2,  
"genre" : "action",  
"feedback" : 76435,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb50"),  
"name" : "life is beautiful",  
"imdb" : 9.4,  
"genre" : "comedy",  
"feedback" : 34760,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb51"),  
"name" : "the godfather",  
"imdb" : 9.2,  
"genre" : "suspense",  
"feedback" : 54321,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb52"),  
"name" : "avatar",  
"imdb" : 8.8,  
"genre" : "action",  
"feedback" : 87987,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb53"),  
"name" : "psycho",  
"imdb" : 5.6,  
"genre" : "thriller",  
"feedback" : 12323,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb54"),  
"name" : "rain man",  
"imdb" : 6.2,  
"genre" : "adventure",  
"feedback" : 54321,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb55"),  
"name" : "inception",  
"imdb" : 7.8,  
"genre" : "action",  
"feedback" : 43254,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb56"),  
"name" : "terminator",  
"imdb" : 9.5,  
"genre" : "action",  
"feedback" : 65243,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb57"),  
"name" : "jurassic park",  
"imdb" : 6.4,  
"genre" : "adventure",  
"feedback" : 89123,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb58"),  
"name" : "pirates of caribbean",  
"imdb" : 5.4,  
"genre" : "adventure",  
"feedback" : 67543,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb59"),  
"name" : "rocky",  
"imdb" : 6.8,  
"genre" : "drama",  
"feedback" : 53451,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb5a"),  
"name" : "reservoir dogs",  
"imdb" : 8.3,  
"genre" : "drama",  
"feedback" : 4567,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb5b"),  
"name" : "l a confidential",  
"imdb" : 8.6,  
"genre" : "suspense",  
"feedback" : 87345,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb5c"),  
"name" : "sherlock holmes",  
"imdb" : 9.9,  
"genre" : "adventure",  
"feedback" : 65312,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb5d"),  
"name" : "django unchained",  
"imdb" : 4.6,  
"genre" : "action",  
"feedback" : 785623,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb5e"),  
"name" : "sin city",  
"imdb" : 7.8,  
"genre" : "suspense",  
"feedback" : 90671,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb5f"),  
"name" : "american beauty",  
"imdb" : 8.8,  
"genre" : "drama",  
"feedback" : 54112,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb60"),  
"name" : "taxi driver",  
"imdb" : 9.2,  
"genre" : "suspense",  
"feedback" : 78612,  
"referrable" : "TRUE"  
}

1. Demonstrate usage of $match, $group and aggregate pipelines.

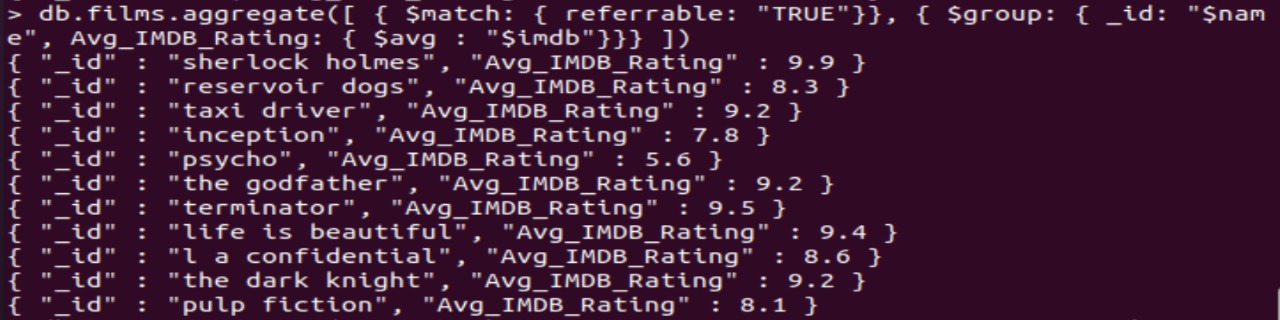
> db.films.aggregate([{ $group : {\_id : "$genre", avg\_rating : {$avg : "$imdb"}}}]).pretty()  
{ "\_id" : "drama", "avg\_rating" : 7.966666666666668 }  
{ "\_id" : "adventure", "avg\_rating" : 6.9750000000000005 }  
{ "\_id" : "thriller", "avg\_rating" : 5.6 }  
{ "\_id" : "comedy", "avg\_rating" : 9.4 }  
{ "\_id" : "action", "avg\_rating" : 7.9799999999999995 }  
{ "\_id" : "romance", "avg\_rating" : 8.6 }  
{ "\_id" : "suspense", "avg\_rating" : 8.58 }



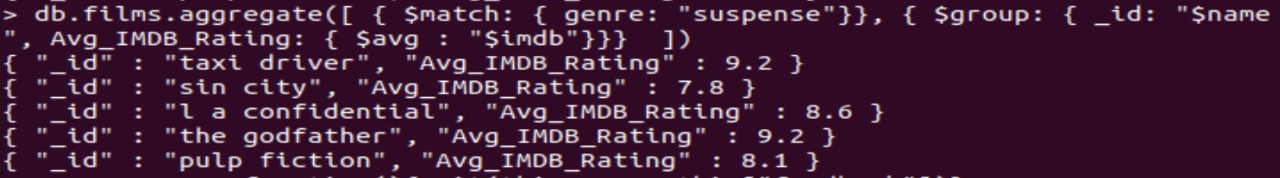
> db.films.aggregate([{ $match : { genre : "suspense"}}]).pretty()  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb4d"),  
"name" : "pulp fiction",  
"imdb" : 8.1,  
"genre" : "suspense",  
"feedback" : 12094,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb51"),  
"name" : "the godfather",  
"imdb" : 9.2,  
"genre" : "suspense",  
"feedback" : 54321,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb5b"),  
"name" : "l a confidential",  
"imdb" : 8.6,  
"genre" : "suspense",  
"feedback" : 87345,  
"referrable" : "TRUE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb5e"),  
"name" : "sin city",  
"imdb" : 7.8,  
"genre" : "suspense",  
"feedback" : 90671,  
"referrable" : "FALSE"  
}  
{  
"\_id" : ObjectId("60eb38a47fd1ce1433ddbb60"),  
"name" : "taxi driver",  
"imdb" : 9.2,  
"genre" : "suspense",  
"feedback" : 78612,  
"referrable" : "TRUE"  
}



> db.films.aggregate([ { $match: { referrable: "TRUE"}}, { $group: { \_id: "$name", Avg\_IMDB\_Rating: { $avg : "$imdb"}}} ])  
{ "\_id" : "sherlock holmes", "Avg\_IMDB\_Rating" : 9.9 }  
{ "\_id" : "reservoir dogs", "Avg\_IMDB\_Rating" : 8.3 }  
{ "\_id" : "taxi driver", "Avg\_IMDB\_Rating" : 9.2 }  
{ "\_id" : "inception", "Avg\_IMDB\_Rating" : 7.8 }  
{ "\_id" : "psycho", "Avg\_IMDB\_Rating" : 5.6 }  
{ "\_id" : "the godfather", "Avg\_IMDB\_Rating" : 9.2 }  
{ "\_id" : "terminator", "Avg\_IMDB\_Rating" : 9.5 }  
{ "\_id" : "life is beautiful", "Avg\_IMDB\_Rating" : 9.4 }  
{ "\_id" : "l a confidential", "Avg\_IMDB\_Rating" : 8.6 }  
{ "\_id" : "the dark knight", "Avg\_IMDB\_Rating" : 9.2 }  
{ "\_id" : "pulp fiction", "Avg\_IMDB\_Rating" : 8.1 }

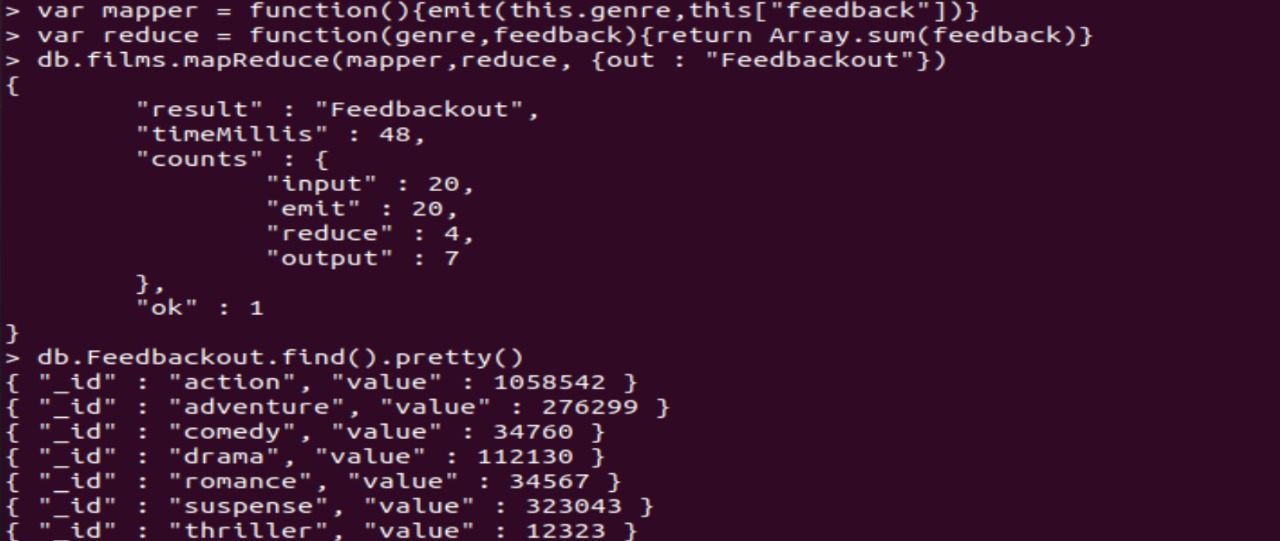


> db.films.aggregate([ { $match: { genre: "suspense"}}, { $group: { \_id: "$name", Avg\_IMDB\_Rating: { $avg : "$imdb"}}}  ])  
{ "\_id" : "taxi driver", "Avg\_IMDB\_Rating" : 9.2 }  
{ "\_id" : "sin city", "Avg\_IMDB\_Rating" : 7.8 }  
{ "\_id" : "l a confidential", "Avg\_IMDB\_Rating" : 8.6 }  
{ "\_id" : "the godfather", "Avg\_IMDB\_Rating" : 9.2 }  
{ "\_id" : "pulp fiction", "Avg\_IMDB\_Rating" : 8.1 }



2. Demonstarte map-reduce aggregate function

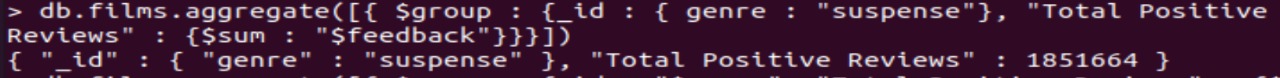
> var mapper = function(){emit(this.genre,this["feedback"])}  
> var reduce = function(genre,feedback){return Array.sum(feedback)}  
> db.films.mapReduce(mapper,reduce, {out : "Feedbackout"})  
{  
"result" : "Feedbackout",  
"timeMillis" : 48,  
"counts" : {  
"input" : 20,  
"emit" : 20,  
"reduce" : 4,  
"output" : 7  
},  
"ok" : 1  
}  
> db.Feedbackout.find().pretty()  
{ "\_id" : "action", "value" : 1058542 }  
{ "\_id" : "adventure", "value" : 276299 }  
{ "\_id" : "comedy", "value" : 34760 }  
{ "\_id" : "drama", "value" : 112130 }  
{ "\_id" : "romance", "value" : 34567 }  
{ "\_id" : "suspense", "value" : 323043 }  
{ "\_id" : "thriller", "value" : 12323 }



3. Count the number of movies which belong to suspense category and find the total number of positive reviews in that category

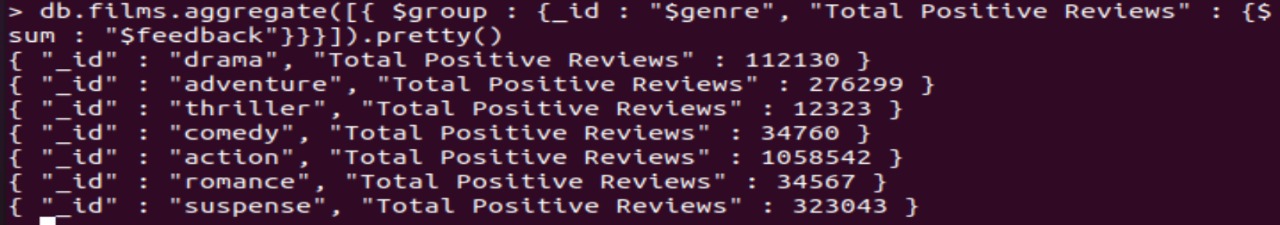
(I took suspense instead thriller as a genre because in my dataset only one movie had thriller category)

> db.films.aggregate([{ $group : {\_id : { genre : "suspense"}, "Total Positive Reviews" : {$sum : "$feedback"}}}])  
{ "\_id" : { "genre" : "suspense" }, "Total Positive Reviews" : 1851664 }



4. Group all the records by genre and find the total number of positive reviews in that genre

> db.films.aggregate([{ $group : {\_id : "$genre", "Total Positive Reviews" : {$sum : "$feedback"}}}]).pretty()  
{ "\_id" : "drama", "Total Positive Reviews" : 112130 }  
{ "\_id" : "adventure", "Total Positive Reviews" : 276299 }  
{ "\_id" : "thriller", "Total Positive Reviews" : 12323 }  
{ "\_id" : "comedy", "Total Positive Reviews" : 34760 }  
{ "\_id" : "action", "Total Positive Reviews" : 1058542 }  
{ "\_id" : "romance", "Total Positive Reviews" : 34567 }  
{ "\_id" : "suspense", "Total Positive Reviews" : 323043 }  
>

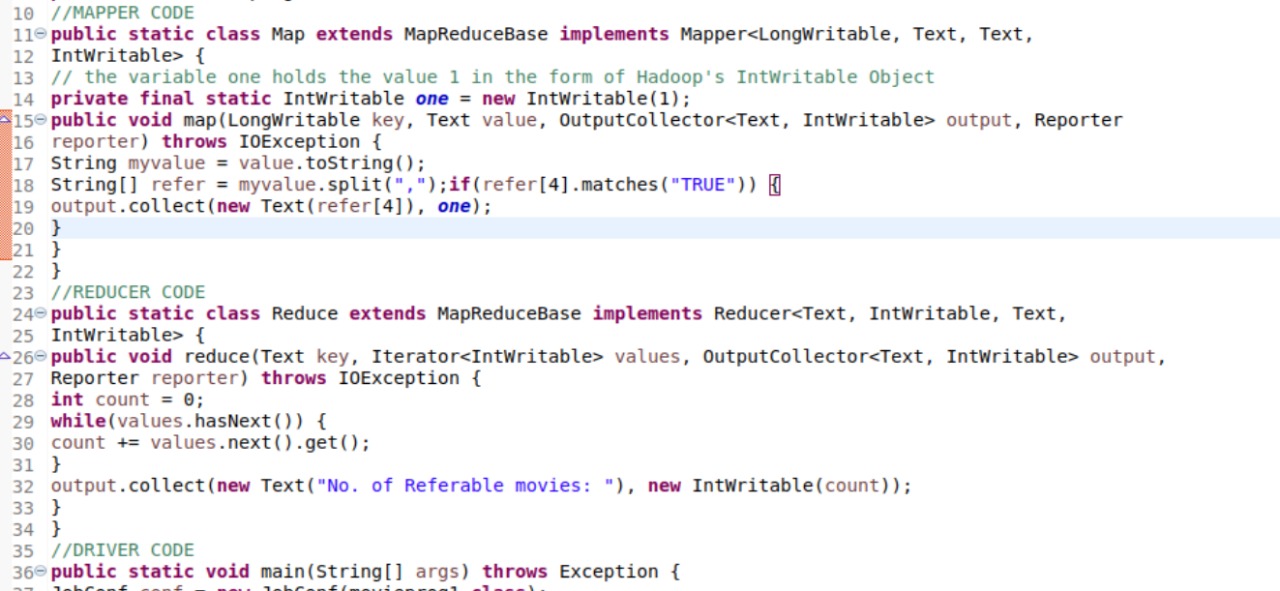


HADOOP PART

//1st program

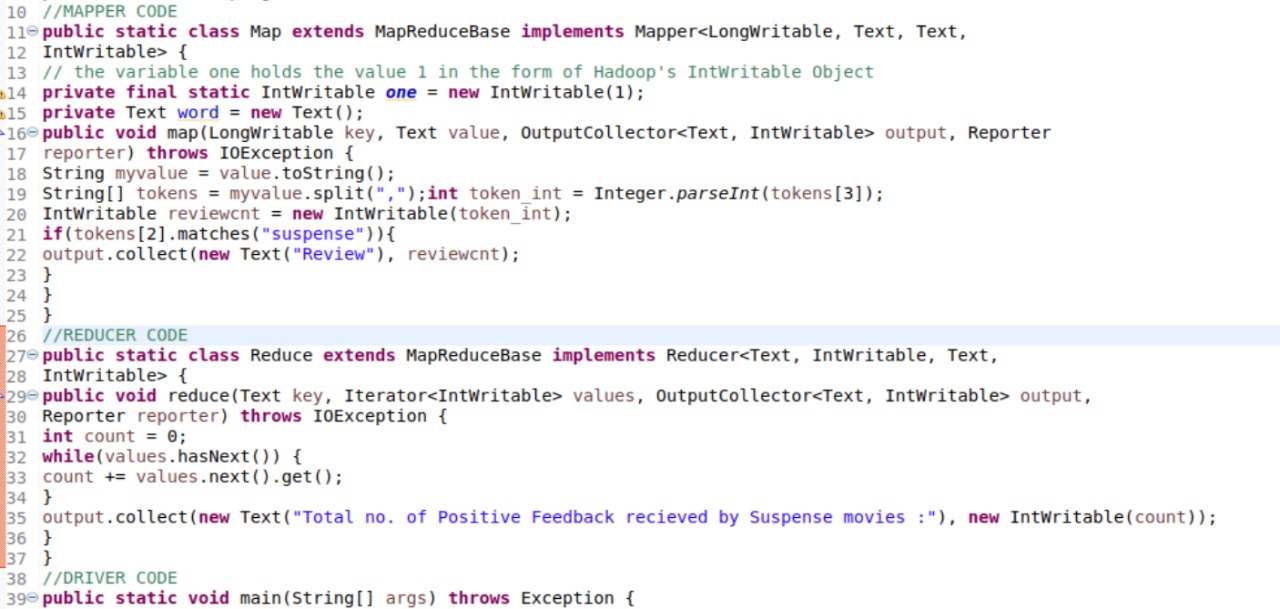
package bdla1;  
import java.io.IOException;  
import java.util.\*;  
import org.apache.hadoop.fs.Path;  
import org.apache.hadoop.conf.\*;  
import org.apache.hadoop.io.\*;  
import org.apache.hadoop.mapred.\*;  
import org.apache.hadoop.util.\*;  
public class movieprog1 {  
//MAPPER CODE  
public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text,  
IntWritable> {  
// the variable one holds the value 1 in the form of Hadoop's IntWritable Object  
private final static IntWritable one = new IntWritable(1);  
public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter  
reporter) throws IOException {  
String myvalue = value.toString();  
String[] refer = myvalue.split(",");if(refer[4].matches("TRUE")) {  
output.collect(new Text(refer[4]), one);  
}  
}  
}  
//REDUCER CODE  
public static class Reduce 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 count = 0;  
while(values.hasNext()) {  
count += values.next().get();  
}  
output.collect(new Text("No. of Referable movies: "), new IntWritable(count));  
}  
}  
//DRIVER CODE  
public static void main(String[] args) throws Exception {  
JobConf conf = new JobConf(movieprog1.class);  
conf.setJobName("Refrable Movie Counting");  
conf.setOutputKeyClass(Text.class);  
conf.setOutputValueClass(IntWritable.class);  
conf.setMapperClass(Map.class);  
conf.setCombinerClass(Reduce.class);  
conf.setReducerClass(Reduce.class);  
conf.setInputFormat(TextInputFormat.class);  
conf.setOutputFormat(TextOutputFormat.class); // hadoop jar jarname classpath //inputfolder  
FileInputFormat.setInputPaths(conf, new Path(args[0]));  
FileOutputFormat.setOutputPath(conf, new Path(args[1]));  
JobClient.runJob(conf);  
}

}

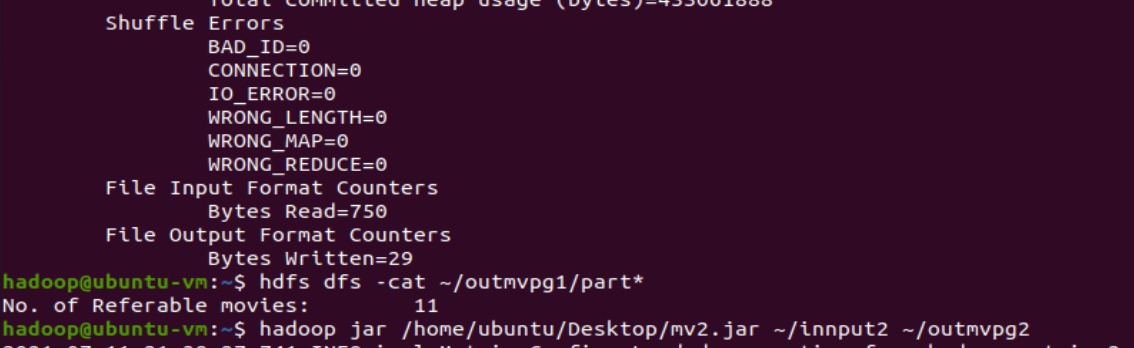


//2nd Program

package bdla1;  
import java.io.IOException;  
import java.util.\*;  
import org.apache.hadoop.fs.Path;  
import org.apache.hadoop.conf.\*;  
import org.apache.hadoop.io.\*;  
import org.apache.hadoop.mapred.\*;  
import org.apache.hadoop.util.\*;  
public class movieprog2 {  
//MAPPER CODE  
public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text,  
IntWritable> {  
// the variable one holds the value 1 in the form of Hadoop's IntWritable Object  
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 myvalue = value.toString();  
String[] tokens = myvalue.split(",");int token\_int = Integer.parseInt(tokens[3]);  
IntWritable reviewcnt = new IntWritable(token\_int);  
if(tokens[2].matches("suspense")){  
output.collect(new Text("Review"), reviewcnt);  
}  
}  
}  
//REDUCER CODE  
public static class Reduce 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 count = 0;  
while(values.hasNext()) {  
count += values.next().get();  
}  
output.collect(new Text("Total no. of Positive Feedback recieved by Suspense movies :"), new IntWritable(count));  
}  
}  
//DRIVER CODE  
public static void main(String[] args) throws Exception {  
JobConf conf = new JobConf(movieprog2.class);  
conf.setJobName("Positive Reviews Counting");  
conf.setOutputKeyClass(Text.class);  
conf.setOutputValueClass(IntWritable.class);  
conf.setMapperClass(Map.class);  
conf.setCombinerClass(Reduce.class);  
conf.setReducerClass(Reduce.class);  
conf.setInputFormat(TextInputFormat.class);  
conf.setOutputFormat(TextOutputFormat.class); // hadoop jar jarname classpath inputfolder  
//outputfolder  
FileInputFormat.setInputPaths(conf, new Path(args[0]));  
FileOutputFormat.setOutputPath(conf, new Path(args[1]));JobClient.runJob(conf);  
}  
}



ubuntu@ubuntu-vm:~$ sudo su - hadoop  
[sudo] password for ubuntu:  
hadoop@ubuntu-vm:~$ cd $HADOOP\_HOME/sbin && ./start-all.sh  
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.  
WARNING: This is not a recommended production deployment configuration.  
WARNING: Use CTRL-C to abort.  
Starting namenodes on [localhost]  
Starting datanodes  
Starting secondary namenodes [ubuntu-vm]  
Starting resourcemanager  
Starting nodemanagers  
hadoop@ubuntu-vm:/usr/local/hadoop/sbin$ jps  
8064 ResourceManager  
8228 NodeManager  
7670 DataNode  
7878 SecondaryNameNode  
8556 Jps  
7485 NameNode  
hadoop@ubuntu-vm:/usr/local/hadoop/sbin$ hdfs dfs -mkdir -p ~/innput  
hadoop@ubuntu-vm:/usr/local/hadoop/sbin$ cd ~  
hadoop@ubuntu-vm:~$ hdfs dfs -mkdir -p ~/innput  
hadoop@ubuntu-vm:~$  hdfs dfs -put /home/ubuntu/Desktop/bdla.csv ~/innput2  
hadoop@ubuntu-vm:~$ hadoop jar /home/ubuntu/Desktop/mv1.jar ~/innput2 ~/outmvpg1  
2021-07-11 21:37:27,773 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties  
2021-07-11 21:37:27,841 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).  
2021-07-11 21:37:27,841 INFO impl.MetricsSystemImpl: JobTracker metrics system started  
2021-07-11 21:37:27,851 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!  
2021-07-11 21:37:27,999 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
2021-07-11 21:37:28,068 INFO mapred.FileInputFormat: Total input files to process : 1  
2021-07-11 21:37:28,116 INFO mapreduce.JobSubmitter: number of splits:1  
2021-07-11 21:37:28,210 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_local2018836498\_0001  
2021-07-11 21:37:28,210 INFO mapreduce.JobSubmitter: Executing with tokens: []  
2021-07-11 21:37:28,301 INFO mapreduce.Job: The url to track the job: <http://localhost:8080/>  
2021-07-11 21:37:28,303 INFO mapreduce.Job: Running job: job\_local2018836498\_0001  
2021-07-11 21:37:28,303 INFO mapred.LocalJobRunner: OutputCommitter set in config null  
2021-07-11 21:37:28,305 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter  
2021-07-11 21:37:28,311 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2  
2021-07-11 21:37:28,312 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false  
2021-07-11 21:37:28,357 INFO mapred.LocalJobRunner: Waiting for map tasks  
2021-07-11 21:37:28,359 INFO mapred.LocalJobRunner: Starting task: attempt\_local2018836498\_0001\_m\_000000\_0  
2021-07-11 21:37:28,398 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2  
2021-07-11 21:37:28,398 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false  
2021-07-11 21:37:28,426 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]  
2021-07-11 21:37:28,436 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/home/hadoop/innput2:0+750  
2021-07-11 21:37:28,483 INFO mapred.MapTask: numReduceTasks: 1  
2021-07-11 21:37:28,531 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)  
2021-07-11 21:37:28,531 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100  
2021-07-11 21:37:28,531 INFO mapred.MapTask: soft limit at 83886080  
2021-07-11 21:37:28,531 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600  
2021-07-11 21:37:28,531 INFO mapred.MapTask: kvstart = 26214396; length = 6553600  
2021-07-11 21:37:28,536 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer  
2021-07-11 21:37:28,682 INFO mapred.LineRecordReader: Found UTF-8 BOM and skipped it  
2021-07-11 21:37:28,685 INFO mapred.LocalJobRunner:  
2021-07-11 21:37:28,685 INFO mapred.MapTask: Starting flush of map output  
2021-07-11 21:37:28,685 INFO mapred.MapTask: Spilling map output  
2021-07-11 21:37:28,685 INFO mapred.MapTask: bufstart = 0; bufend = 99; bufvoid = 104857600  
2021-07-11 21:37:28,685 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214356(104857424); length = 41/6553600  
2021-07-11 21:37:28,692 INFO mapred.MapTask: Finished spill 0  
2021-07-11 21:37:28,706 INFO mapred.Task: Task:attempt\_local2018836498\_0001\_m\_000000\_0 is done. And is in the process of committing  
2021-07-11 21:37:28,713 INFO mapred.LocalJobRunner: hdfs://localhost:9000/home/hadoop/innput2:0+750  
2021-07-11 21:37:28,713 INFO mapred.Task: Task 'attempt\_local2018836498\_0001\_m\_000000\_0' done.  
2021-07-11 21:37:28,725 INFO mapred.Task: Final Counters for attempt\_local2018836498\_0001\_m\_000000\_0: Counters: 24  
File System Counters  
FILE: Number of bytes read=13370  
FILE: Number of bytes written=558400  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=750  
HDFS: Number of bytes written=0  
HDFS: Number of read operations=5  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=1  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework  
Map input records=20  
Map output records=11  
Map output bytes=99  
Map output materialized bytes=38  
Input split bytes=93  
Combine input records=11  
Combine output records=1  
Spilled Records=1  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=15  
Total committed heap usage (bytes)=216530944  
File Input Format Counters  
Bytes Read=750  
2021-07-11 21:37:28,730 INFO mapred.LocalJobRunner: Finishing task: attempt\_local2018836498\_0001\_m\_000000\_0  
2021-07-11 21:37:28,731 INFO mapred.LocalJobRunner: map task executor complete.  
2021-07-11 21:37:28,737 INFO mapred.LocalJobRunner: Waiting for reduce tasks  
2021-07-11 21:37:28,737 INFO mapred.LocalJobRunner: Starting task: attempt\_local2018836498\_0001\_r\_000000\_0  
2021-07-11 21:37:28,744 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2  
2021-07-11 21:37:28,744 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false  
2021-07-11 21:37:28,748 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]  
2021-07-11 21:37:28,752 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@39353fc5  
2021-07-11 21:37:28,753 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!  
2021-07-11 21:37:28,774 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=810339520, maxSingleShuffleLimit=202584880, mergeThreshold=534824096, ioSortFactor=10, memToMemMergeOutputsThreshold=10  
2021-07-11 21:37:28,797 INFO reduce.EventFetcher: attempt\_local2018836498\_0001\_r\_000000\_0 Thread started: EventFetcher for fetching Map Completion Events  
2021-07-11 21:37:28,829 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt\_local2018836498\_0001\_m\_000000\_0 decomp: 34 len: 38 to MEMORY  
2021-07-11 21:37:28,831 INFO reduce.InMemoryMapOutput: Read 34 bytes from map-output for attempt\_local2018836498\_0001\_m\_000000\_0  
2021-07-11 21:37:28,832 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 34, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->34  
2021-07-11 21:37:28,833 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning  
2021-07-11 21:37:28,834 INFO mapred.LocalJobRunner: 1 / 1 copied.  
2021-07-11 21:37:28,834 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs  
2021-07-11 21:37:28,840 INFO mapred.Merger: Merging 1 sorted segments  
2021-07-11 21:37:28,841 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 6 bytes  
2021-07-11 21:37:28,842 INFO reduce.MergeManagerImpl: Merged 1 segments, 34 bytes to disk to satisfy reduce memory limit  
2021-07-11 21:37:28,842 INFO reduce.MergeManagerImpl: Merging 1 files, 38 bytes from disk  
2021-07-11 21:37:28,843 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce  
2021-07-11 21:37:28,843 INFO mapred.Merger: Merging 1 sorted segments  
2021-07-11 21:37:28,844 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 6 bytes  
2021-07-11 21:37:28,845 INFO mapred.LocalJobRunner: 1 / 1 copied.  
2021-07-11 21:37:28,948 INFO mapred.Task: Task:attempt\_local2018836498\_0001\_r\_000000\_0 is done. And is in the process of committing  
2021-07-11 21:37:28,951 INFO mapred.LocalJobRunner: 1 / 1 copied.  
2021-07-11 21:37:28,951 INFO mapred.Task: Task attempt\_local2018836498\_0001\_r\_000000\_0 is allowed to commit now  
2021-07-11 21:37:28,978 INFO output.FileOutputCommitter: Saved output of task 'attempt\_local2018836498\_0001\_r\_000000\_0' to hdfs://localhost:9000/home/hadoop/outmvpg1  
2021-07-11 21:37:28,979 INFO mapred.LocalJobRunner: reduce > reduce  
2021-07-11 21:37:28,979 INFO mapred.Task: Task 'attempt\_local2018836498\_0001\_r\_000000\_0' done.  
2021-07-11 21:37:28,980 INFO mapred.Task: Final Counters for attempt\_local2018836498\_0001\_r\_000000\_0: Counters: 30  
File System Counters  
FILE: Number of bytes read=13478  
FILE: Number of bytes written=558438  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=750  
HDFS: Number of bytes written=29  
HDFS: Number of read operations=10  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=3  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=38  
Reduce input records=1  
Reduce output records=1  
Spilled Records=1  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=0  
Total committed heap usage (bytes)=216530944  
Shuffle Errors  
BAD\_ID=0  
CONNECTION=0  
IO\_ERROR=0  
WRONG\_LENGTH=0  
WRONG\_MAP=0  
WRONG\_REDUCE=0  
File Output Format Counters  
Bytes Written=29  
2021-07-11 21:37:28,980 INFO mapred.LocalJobRunner: Finishing task: attempt\_local2018836498\_0001\_r\_000000\_0  
2021-07-11 21:37:28,980 INFO mapred.LocalJobRunner: reduce task executor complete.  
2021-07-11 21:37:29,318 INFO mapreduce.Job: Job job\_local2018836498\_0001 running in uber mode : false  
2021-07-11 21:37:29,320 INFO mapreduce.Job:  map 100% reduce 100%  
2021-07-11 21:37:29,321 INFO mapreduce.Job: Job job\_local2018836498\_0001 completed successfully  
2021-07-11 21:37:29,329 INFO mapreduce.Job: Counters: 36  
File System Counters  
FILE: Number of bytes read=26848  
FILE: Number of bytes written=1116838  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=1500  
HDFS: Number of bytes written=29  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=4  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework  
Map input records=20  
Map output records=11  
Map output bytes=99  
Map output materialized bytes=38  
Input split bytes=93  
Combine input records=11  
Combine output records=1  
Reduce input groups=1  
Reduce shuffle bytes=38  
Reduce input records=1  
Reduce output records=1  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=15  
Total committed heap usage (bytes)=433061888  
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=750  
File Output Format Counters  
Bytes Written=29  
hadoop@ubuntu-vm:~$ hdfs dfs -cat ~/outmvpg1/part\*  
No. of Referable movies: 11

  
hadoop@ubuntu-vm:~$ hadoop jar /home/ubuntu/Desktop/mv2.jar ~/innput2 ~/outmvpg2  
2021-07-11 21:38:27,741 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties  
2021-07-11 21:38:27,813 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).  
2021-07-11 21:38:27,813 INFO impl.MetricsSystemImpl: JobTracker metrics system started  
2021-07-11 21:38:27,822 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!  
2021-07-11 21:38:27,978 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
2021-07-11 21:38:28,050 INFO mapred.FileInputFormat: Total input files to process : 1  
2021-07-11 21:38:28,088 INFO mapreduce.JobSubmitter: number of splits:1  
2021-07-11 21:38:28,185 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_local54023548\_0001  
2021-07-11 21:38:28,185 INFO mapreduce.JobSubmitter: Executing with tokens: []  
2021-07-11 21:38:28,276 INFO mapreduce.Job: The url to track the job: <http://localhost:8080/>  
2021-07-11 21:38:28,277 INFO mapreduce.Job: Running job: job\_local54023548\_0001  
2021-07-11 21:38:28,277 INFO mapred.LocalJobRunner: OutputCommitter set in config null  
2021-07-11 21:38:28,278 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter  
2021-07-11 21:38:28,290 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2  
2021-07-11 21:38:28,290 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false  
2021-07-11 21:38:28,336 INFO mapred.LocalJobRunner: Waiting for map tasks  
2021-07-11 21:38:28,341 INFO mapred.LocalJobRunner: Starting task: attempt\_local54023548\_0001\_m\_000000\_0  
2021-07-11 21:38:28,366 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2  
2021-07-11 21:38:28,366 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false  
2021-07-11 21:38:28,397 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]  
2021-07-11 21:38:28,406 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/home/hadoop/innput2:0+750  
2021-07-11 21:38:28,446 INFO mapred.MapTask: numReduceTasks: 1  
2021-07-11 21:38:28,482 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)  
2021-07-11 21:38:28,482 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100  
2021-07-11 21:38:28,483 INFO mapred.MapTask: soft limit at 83886080  
2021-07-11 21:38:28,483 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600  
2021-07-11 21:38:28,483 INFO mapred.MapTask: kvstart = 26214396; length = 6553600  
2021-07-11 21:38:28,488 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer  
2021-07-11 21:38:28,594 INFO mapred.LineRecordReader: Found UTF-8 BOM and skipped it  
2021-07-11 21:38:28,597 INFO mapred.LocalJobRunner:  
2021-07-11 21:38:28,597 INFO mapred.MapTask: Starting flush of map output  
2021-07-11 21:38:28,597 INFO mapred.MapTask: Spilling map output  
2021-07-11 21:38:28,597 INFO mapred.MapTask: bufstart = 0; bufend = 55; bufvoid = 104857600  
2021-07-11 21:38:28,597 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214380(104857520); length = 17/6553600  
2021-07-11 21:38:28,603 INFO mapred.MapTask: Finished spill 0  
2021-07-11 21:38:28,619 INFO mapred.Task: Task:attempt\_local54023548\_0001\_m\_000000\_0 is done. And is in the process of committing  
2021-07-11 21:38:28,625 INFO mapred.LocalJobRunner: hdfs://localhost:9000/home/hadoop/innput2:0+750  
2021-07-11 21:38:28,625 INFO mapred.Task: Task 'attempt\_local54023548\_0001\_m\_000000\_0' done.  
2021-07-11 21:38:28,632 INFO mapred.Task: Final Counters for attempt\_local54023548\_0001\_m\_000000\_0: Counters: 24  
File System Counters  
FILE: Number of bytes read=16614  
FILE: Number of bytes written=556443  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=750  
HDFS: Number of bytes written=0  
HDFS: Number of read operations=5  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=1  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework  
Map input records=20  
Map output records=5  
Map output bytes=55  
Map output materialized bytes=73  
Input split bytes=93  
Combine input records=5  
Combine output records=1  
Spilled Records=1  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=10  
Total committed heap usage (bytes)=216006656  
File Input Format Counters  
Bytes Read=750  
2021-07-11 21:38:28,632 INFO mapred.LocalJobRunner: Finishing task: attempt\_local54023548\_0001\_m\_000000\_0  
2021-07-11 21:38:28,632 INFO mapred.LocalJobRunner: map task executor complete.  
2021-07-11 21:38:28,636 INFO mapred.LocalJobRunner: Waiting for reduce tasks  
2021-07-11 21:38:28,636 INFO mapred.LocalJobRunner: Starting task: attempt\_local54023548\_0001\_r\_000000\_0  
2021-07-11 21:38:28,642 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2  
2021-07-11 21:38:28,642 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false  
2021-07-11 21:38:28,642 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]  
2021-07-11 21:38:28,647 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@2bc7819b  
2021-07-11 21:38:28,648 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!  
2021-07-11 21:38:28,669 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=810339520, maxSingleShuffleLimit=202584880, mergeThreshold=534824096, ioSortFactor=10, memToMemMergeOutputsThreshold=10  
2021-07-11 21:38:28,672 INFO reduce.EventFetcher: attempt\_local54023548\_0001\_r\_000000\_0 Thread started: EventFetcher for fetching Map Completion Events  
2021-07-11 21:38:28,699 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt\_local54023548\_0001\_m\_000000\_0 decomp: 69 len: 73 to MEMORY  
2021-07-11 21:38:28,709 INFO reduce.InMemoryMapOutput: Read 69 bytes from map-output for attempt\_local54023548\_0001\_m\_000000\_0  
2021-07-11 21:38:28,710 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 69, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->69  
2021-07-11 21:38:28,711 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning  
2021-07-11 21:38:28,712 INFO mapred.LocalJobRunner: 1 / 1 copied.  
2021-07-11 21:38:28,712 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs  
2021-07-11 21:38:28,717 INFO mapred.Merger: Merging 1 sorted segments  
2021-07-11 21:38:28,717 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 6 bytes  
2021-07-11 21:38:28,718 INFO reduce.MergeManagerImpl: Merged 1 segments, 69 bytes to disk to satisfy reduce memory limit  
2021-07-11 21:38:28,718 INFO reduce.MergeManagerImpl: Merging 1 files, 73 bytes from disk  
2021-07-11 21:38:28,719 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce  
2021-07-11 21:38:28,719 INFO mapred.Merger: Merging 1 sorted segments  
2021-07-11 21:38:28,720 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 6 bytes  
2021-07-11 21:38:28,721 INFO mapred.LocalJobRunner: 1 / 1 copied.  
2021-07-11 21:38:28,806 INFO mapred.Task: Task:attempt\_local54023548\_0001\_r\_000000\_0 is done. And is in the process of committing  
2021-07-11 21:38:28,809 INFO mapred.LocalJobRunner: 1 / 1 copied.  
2021-07-11 21:38:28,809 INFO mapred.Task: Task attempt\_local54023548\_0001\_r\_000000\_0 is allowed to commit now  
2021-07-11 21:38:28,826 INFO output.FileOutputCommitter: Saved output of task 'attempt\_local54023548\_0001\_r\_000000\_0' to hdfs://localhost:9000/home/hadoop/outmvpg2  
2021-07-11 21:38:28,827 INFO mapred.LocalJobRunner: reduce > reduce  
2021-07-11 21:38:28,827 INFO mapred.Task: Task 'attempt\_local54023548\_0001\_r\_000000\_0' done.  
2021-07-11 21:38:28,827 INFO mapred.Task: Final Counters for attempt\_local54023548\_0001\_r\_000000\_0: Counters: 30  
File System Counters  
FILE: Number of bytes read=16792  
FILE: Number of bytes written=556516  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=750  
HDFS: Number of bytes written=68  
HDFS: Number of read operations=10  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=3  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=73  
Reduce input records=1  
Reduce output records=1  
Spilled Records=1  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=0  
Total committed heap usage (bytes)=216006656  
Shuffle Errors  
BAD\_ID=0  
CONNECTION=0  
IO\_ERROR=0  
WRONG\_LENGTH=0  
WRONG\_MAP=0  
WRONG\_REDUCE=0  
File Output Format Counters  
Bytes Written=68  
2021-07-11 21:38:28,827 INFO mapred.LocalJobRunner: Finishing task: attempt\_local54023548\_0001\_r\_000000\_0  
2021-07-11 21:38:28,827 INFO mapred.LocalJobRunner: reduce task executor complete.  
2021-07-11 21:38:29,281 INFO mapreduce.Job: Job job\_local54023548\_0001 running in uber mode : false  
2021-07-11 21:38:29,281 INFO mapreduce.Job:  map 100% reduce 100%  
2021-07-11 21:38:29,282 INFO mapreduce.Job: Job job\_local54023548\_0001 completed successfully  
2021-07-11 21:38:29,286 INFO mapreduce.Job: Counters: 36  
File System Counters  
FILE: Number of bytes read=33406  
FILE: Number of bytes written=1112959  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=1500  
HDFS: Number of bytes written=68  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=4  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework  
Map input records=20  
Map output records=5  
Map output bytes=55  
Map output materialized bytes=73  
Input split bytes=93  
Combine input records=5  
Combine output records=1  
Reduce input groups=1  
Reduce shuffle bytes=73  
Reduce input records=1  
Reduce output records=1  
Spilled Records=2  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=10  
Total committed heap usage (bytes)=432013312  
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=750  
File Output Format Counters  
Bytes Written=68  
hadoop@ubuntu-vm:~$ hdfs dfs -cat ~/outmvpg2/part\*  
Total no. of Positive Feedback recieved by Suspense movies : 323043  
hadoop@ubuntu-vm:~$

