Program :

package movie;

import java.io.IOException;

import org.apache.hadoop.conf.Configuration;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.\*;

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;

import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

import org.apache.hadoop.mapreduce.\*;

public class Movie {

public static class MovieMapper extends Mapper<LongWritable, Text, Text, DoubleWritable> {

@Override

public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {

String[] parts = value.toString().split(",");

if (parts.length == 2) {

String movieName = parts[0].trim();

double rating = Double.parseDouble(parts[1].trim());

context.write(new Text(movieName), new DoubleWritable(rating));

}

}

}

public static class MovieReducer extends Reducer<Text, DoubleWritable, Text, DoubleWritable> {

@Override

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

double sum = 0;

int count = 0;

for (DoubleWritable value : values) {

sum += value.get();

count++;

}double averageRating = sum / count;

context.write(key, new DoubleWritable(averageRating));

}

}

public static void main(String[] args) throws Exception {

Configuration conf = new Configuration();

Job job = Job.getInstance(conf, "Movie Recommendation");

job.setJarByClass(Movie.class);

job.setMapperClass(MovieMapper.class);

job.setReducerClass(MovieReducer.class);job.setOutputKeyClass(Text.class);

job.setOutputValueClass(DoubleWritable.class);

FileInputFormat.addInputPath(job, new Path(args[0]));

FileOutputFormat.setOutputPath(job, new Path(args[1]));

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

}

}

File :

diary of a young Gril , 9.0

the knight nation , 8.7

businessman ,9.0

mere humsafar, 10

Pyare Afzal ,9.9

yeh rista kya kehelata hai 9.9

Output :

[cloudera@quickstart ~]$ hadoop fs -put Desktop/picture.txt picture.txt

[cloudera@quickstart ~]$ hadoop jar moviename.jar movie.Movie picture.txt pictureoutput

24/05/08 20:14:20 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

24/05/08 20:14:21 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

24/05/08 20:14:22 INFO input.FileInputFormat: Total input paths to process : 1

24/05/08 20:14:22 INFO mapreduce.JobSubmitter: number of splits:1

24/05/08 20:14:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1715223867191\_0001

24/05/08 20:14:23 INFO impl.YarnClientImpl: Submitted application application\_1715223867191\_0001

24/05/08 20:14:23 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application\_1715223867191\_0001/

24/05/08 20:14:23 INFO mapreduce.Job: Running job: job\_1715223867191\_0001

24/05/08 20:14:41 INFO mapreduce.Job: Job job\_1715223867191\_0001 running in uber mode : false

24/05/08 20:14:41 INFO mapreduce.Job: map 0% reduce 0%

24/05/08 20:14:52 INFO mapreduce.Job: map 100% reduce 0%

24/05/08 20:15:04 INFO mapreduce.Job: map 100% reduce 100%

24/05/08 20:15:05 INFO mapreduce.Job: Job job\_1715223867191\_0001 completed successfully

24/05/08 20:15:05 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=134

FILE: Number of bytes written=220629

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=257

HDFS: Number of bytes written=99

HDFS: Number of read operations=6

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

Job Counters

Launched map tasks=1

Launched reduce tasks=1

Data-local map tasks=1

Total time spent by all maps in occupied slots (ms)=8478

Total time spent by all reduces in occupied slots (ms)=9614

Total time spent by all map tasks (ms)=8478

Total time spent by all reduce tasks (ms)=9614

Total vcore-seconds taken by all map tasks=8478

Total vcore-seconds taken by all reduce tasks=9614

Total megabyte-seconds taken by all map tasks=8681472

Total megabyte-seconds taken by all reduce tasks=9844736

Map-Reduce Framework

Map input records=6

Map output records=5

Map output bytes=118

Map output materialized bytes=134

Input split bytes=122

Combine input records=0

Combine output records=0

Reduce input groups=5

Reduce shuffle bytes=134

Reduce input records=5

Reduce output records=5

Spilled Records=10

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=197

CPU time spent (ms)=1750

Physical memory (bytes) snapshot=346222592

Virtual memory (bytes) snapshot=3007389696

Total committed heap usage (bytes)=226365440

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=135

File Output Format Counters

Bytes Written=99

[cloudera@quickstart ~]$ hadoop fs -cat pictureoutput/part-r-00000

Pyare Afzal 9.9

businessman 9.0

diary of a young Gril 9.0

mere humsafar 10.0

the knight nation 8.7

[cloudera@quickstart ~]$