ACADGILD

APACHE PIG: CASE STUDY1 – MOVIE RATING

Assignment 7.2

ASHISH KUMAR
ASHISHASHUU0602@GMAIL.COM

Apache PIG – Case Study 1- Movie Rating

To DO

Problem Statement:

What are the movie titles that the user has rated?

How many times a movie has been rated by the user?

In question 2 above, what is the average rating given for a movie?

Input Files:

1. movies.csv
2. ratings.csv

Jar file used to execute: **Session7.jar** (created using provided .java files)

Syntax of Command to be used to execute Jar File

hadoop jar <path of jar file> <Input 1 path> <Input 2 path> <Output Path>

Command used to execute JAR file

hadoop jar /home/acadgild/Desktop/Assignment7/Session7.jar
/hadoopdata/hdfs/casestudy/movies.csv
/hadoopdata/hdfs/casestudy/ratings.csv /hadoopdata/hdfs/out1

```
[acadgild@192 ~]$ hadoop jar /home/acadgild/Desktop/Assignment7/Session7.jar /hadoopdata/hdfs/casestudy/movies.csv /hadoopdata/hdfs/casestudy/ratings.csv /hadoopdata/hdfs/out1
18/08/07 05:23:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/07 05:23:18 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/07 05:23:19 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/08/07 05:23:19 INFO input.FileInputFormat: Total input paths to process : 1
18/08/07 05:23:19 INFO input.FileInputFormat: Total input paths to process : 1
18/08/07 05:23:19 INFO mapreduce.JobSubmitter: number of splits:7
18/08/07 05:23:19 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533597014224_0011
18/08/07 05:23:20 INFO impl.YarnClientImpl: Submitted application application_1533597014224_0011
18/08/07 05:23:20 INFO mapreduce.Job: The url to track the job: http://192.168.0.4:8088/proxy/application_1533597014224_0011/
18/08/07 05:23:20 INFO mapreduce.Job: Running job: job_1533597014224_0011
18/08/07 05:23:26 INFO mapreduce.Job: Job job_1533597014224_0011 running in uber mode : false
18/08/07 05:23:26 INFO mapreduce.Job:  map 0% reduce 0%
18/08/07 05:23:57 INFO mapreduce.Job:  map 9% reduce 0%
18/08/07 05:24:00 INFO mapreduce.Job:  map 16% reduce 0%
18/08/07 05:24:03 INFO mapreduce.Job:  map 18% reduce 0%
18/08/07 05:24:06 INFO mapreduce.Job:  map 23% reduce 0%
18/08/07 05:24:09 INFO mapreduce.Job:  map 24% reduce 0%
18/08/07 05:24:10 INFO mapreduce.Job:  map 26% reduce 0%
18/08/07 05:24:13 INFO mapreduce.Job:  map 28% reduce 0%
18/08/07 05:24:16 INFO mapreduce.Job:  map 29% reduce 0%
18/08/07 05:24:19 INFO mapreduce.Job:  map 30% reduce 0%
18/08/07 05:24:22 INFO mapreduce.Job:  map 32% reduce 0%
18/08/07 05:24:26 INFO mapreduce.Job:  map 34% reduce 0%
18/08/07 05:24:29 INFO mapreduce.Job:  map 36% reduce 0%
18/08/07 05:24:32 INFO mapreduce.Job:  map 38% reduce 0%
18/08/07 05:24:35 INFO mapreduce.Job:  map 40% reduce 0%
18/08/07 05:24:38 INFO mapreduce.Job:  map 42% reduce 0%
18/08/07 05:24:42 INFO mapreduce.Job:  map 43% reduce 0%
18/08/07 05:24:45 INFO mapreduce.Job:  map 44% reduce 0%
18/08/07 05:24:48 INFO mapreduce.Job:  map 46% reduce 0%
18/08/07 05:24:51 INFO mapreduce.Job:  map 48% reduce 0%
18/08/07 05:24:54 INFO mapreduce.Job:  map 51% reduce 0%
18/08/07 05:24:58 INFO mapreduce.Job:  map 53% reduce 0%
18/08/07 05:25:04 INFO mapreduce.Job:  map 54% reduce 0%
```

```
18/08/07 05:25:51 INFO mapreduce.Job:  map 80% reduce 0%
18/08/07 05:25:53 INFO mapreduce.Job:  map 81% reduce 0%
18/08/07 05:25:54 INFO mapreduce.Job:  map 82% reduce 0%
18/08/07 05:25:57 INFO mapreduce.Job:  map 83% reduce 0%
18/08/07 05:26:00 INFO mapreduce.Job:  map 84% reduce 0%
18/08/07 05:26:02 INFO mapreduce.Job:  map 85% reduce 0%
18/08/07 05:26:03 INFO mapreduce.Job:  map 86% reduce 0%
18/08/07 05:26:11 INFO mapreduce.Job:  map 87% reduce 0%
18/08/07 05:26:13 INFO mapreduce.Job:  map 88% reduce 0%
18/08/07 05:26:18 INFO mapreduce.Job:  map 89% reduce 0%
18/08/07 05:26:22 INFO mapreduce.Job:  map 90% reduce 0%
18/08/07 05:26:28 INFO mapreduce.Job:  map 91% reduce 0%
18/08/07 05:26:36 INFO mapreduce.Job:  map 93% reduce 0%
18/08/07 05:26:41 INFO mapreduce.Job:  map 95% reduce 0%
18/08/07 05:26:44 INFO mapreduce.Job:  map 96% reduce 0%
18/08/07 05:26:48 INFO mapreduce.Job:  map 97% reduce 0%
18/08/07 05:26:51 INFO mapreduce.Job:  map 98% reduce 0%
18/08/07 05:26:54 INFO mapreduce.Job:  map 99% reduce 0%
18/08/07 05:27:00 INFO mapreduce.Job:  map 100% reduce 0%
18/08/07 05:27:29 INFO mapreduce.Job:  map 100% reduce 10%
18/08/07 05:27:38 INFO mapreduce.Job:  map 100% reduce 14%
18/08/07 05:27:47 INFO mapreduce.Job:  map 100% reduce 19%
18/08/07 05:27:57 INFO mapreduce.Job:  map 100% reduce 24%
18/08/07 05:28:03 INFO mapreduce.Job:  map 100% reduce 29%
18/08/07 05:28:12 INFO mapreduce.Job:  map 100% reduce 61%
18/08/07 05:28:15 INFO mapreduce.Job:  map 100% reduce 67%
18/08/07 05:28:27 INFO mapreduce.Job:  map 100% reduce 68%
18/08/07 05:28:39 INFO mapreduce.Job:  map 100% reduce 69%
18/08/07 05:28:48 INFO mapreduce.Job:  map 100% reduce 70%
18/08/07 05:29:00 INFO mapreduce.Job:  map 100% reduce 71%
18/08/07 05:29:12 INFO mapreduce.Job:  map 100% reduce 72%
18/08/07 05:29:21 INFO mapreduce.Job:  map 100% reduce 73%
18/08/07 05:29:33 INFO mapreduce.Job:  map 100% reduce 74%
18/08/07 05:29:45 INFO mapreduce.Job:  map 100% reduce 75%
18/08/07 05:29:57 INFO mapreduce.Job:  map 100% reduce 76%
18/08/07 05:30:09 INFO mapreduce.Job:  map 100% reduce 77%
18/08/07 05:30:21 INFO mapreduce.Job:  map 100% reduce 78%
18/08/07 05:30:34 INFO mapreduce.Job:  map 100% reduce 79%
18/08/07 05:30:46 INFO mapreduce.Job:  map 100% reduce 80%
18/08/07 05:30:58 INFO mapreduce.Job:  map 100% reduce 81%
```

```
18/08/07 05:31:10 INFO mapreduce.Job:  map 100% reduce 82%
18/08/07 05:31:22 INFO mapreduce.Job:  map 100% reduce 83%
18/08/07 05:31:34 INFO mapreduce.Job:  map 100% reduce 84%
18/08/07 05:31:47 INFO mapreduce.Job:  map 100% reduce 85%
18/08/07 05:31:59 INFO mapreduce.Job:  map 100% reduce 86%
18/08/07 05:32:11 INFO mapreduce.Job:  map 100% reduce 87%
18/08/07 05:32:20 INFO mapreduce.Job:  map 100% reduce 88%
18/08/07 05:32:32 INFO mapreduce.Job:  map 100% reduce 89%
18/08/07 05:32:44 INFO mapreduce.Job:  map 100% reduce 90%
18/08/07 05:32:56 INFO mapreduce.Job:  map 100% reduce 91%
18/08/07 05:33:08 INFO mapreduce.Job:  map 100% reduce 92%
18/08/07 05:33:20 INFO mapreduce.Job:  map 100% reduce 93%
18/08/07 05:33:32 INFO mapreduce.Job:  map 100% reduce 94%
18/08/07 05:33:44 INFO mapreduce.Job:  map 100% reduce 95%
18/08/07 05:33:56 INFO mapreduce.Job:  map 100% reduce 96%
18/08/07 05:34:08 INFO mapreduce.Job:  map 100% reduce 97%
18/08/07 05:34:20 INFO mapreduce.Job:  map 100% reduce 98%
18/08/07 05:34:29 INFO mapreduce.Job:  map 100% reduce 99%
18/08/07 05:34:41 INFO mapreduce.Job:  map 100% reduce 100%
18/08/07 05:34:48 INFO mapreduce.Job: Job job_1533597014224_0011 completed successfully
18/08/07 05:34:49 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=961694264
                FILE: Number of bytes written=1457488374
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=711856469
                HDFS: Number of bytes written=1669001
                HDFS: Number of read operations=24
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Killed map tasks=1
                Launched map tasks=8
                Launched reduce tasks=1
                Data-local map tasks=8
                Total time spent by all maps in occupied slots (ms)=1170703
                Total time spent by all reduces in occupied slots (ms)=540290
                Total time spent by all map tasks (ms)=1170703
                Total time spent by all reduce tasks (ms)=540290
                Total vcore-milliseconds taken by all map tasks=1170703
                Total vcore-milliseconds taken by all reduce tasks=540290
```

```
                Total megabyte-milliseconds taken by all map tasks=1198799872
                Total megabyte-milliseconds taken by all reduce tasks=553256960
        Map-Reduce Framework
                Map input records=26070134
                Map output records=26070132
                Map output bytes=442789828
                Map output materialized bytes=494930141
                Input split bytes=2252
                Combine input records=0
                Combine output records=0
                Reduce input groups=45843
                Reduce shuffle bytes=494930141
                Reduce input records=26070132
                Reduce output records=45843
                Spilled Records=76775523
                Shuffled Maps =7
                Failed Shuffles=0
                Merged Map outputs=7
                GC time elapsed (ms)=7679
                CPU time spent (ms)=569900
                Physical memory (bytes) snapshot=2043879424
                Virtual memory (bytes) snapshot=16725954560
                Total committed heap usage (bytes)=1552941056
        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=1669001
```

OUTPUT

```
[acadgild@192 ~]$ hadoop fs -ls /hadoopdata/hdfs/
18/08/07 05:35:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 3 items
drwxr-xr-x   - acadgild supergroup          0 2018-08-07 05:20 /hadoopdata/hdfs/casestudy
drwxr-xr-x   - acadgild supergroup          0 2018-08-07 05:12 /hadoopdata/hdfs/out
drwxr-xr-x   - acadgild supergroup          0 2018-08-07 05:34 /hadoopdata/hdfs/out1
[acadgild@192 ~]$ hadoop fs -ls /hadoopdata/hdfs/out1
18/08/07 05:36:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-08-07 05:34 /hadoopdata/hdfs/out1/_SUCCESS
-rw-r--r--   1 acadgild supergroup    1669001 2018-08-07 05:34 /hadoopdata/hdfs/out1/part-r-00000
[acadgild@192 ~]$
```

```
[acadgild@192 ~]$ hadoop fs -ls /hadoopdata/hdfs/out1
18/08/07 06:03:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-08-07 05:34 /hadoopdata/hdfs/out1/_SUCCESS
-rw-r--r--   1 acadgild supergroup    1669001 2018-08-07 05:34 /hadoopdata/hdfs/out1/part-r-00000
[acadgild@192 ~]$ hadoop fs -cat /hadoopdata/hdfs/out1/part-r-00000 | head
18/08/07 06:03:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Toy Story (1995)        66008   3.888157
GoldenEye (1995)        32534   3.431841
City Hall (1996)        4436    3.232304
Curdled (1996)  217     3.099078
"Comic  1       4.000000
Up in Smoke (1957)      3       3.666667
First Daughter (1999)   3       3.333333
"Flaw   14      3.714286
Battle of Los Angeles (2011)    44      2.522727
Jason Becker: Not Dead Yet (2012)       9       3.444444
cat: Unable to write to output stream.
[acadgild@192 ~]$
```