

# Case Study-I

## Note:

1. In **Task1** of this case study, movies that were never rated was also found out.
2. This code is tested on **local as well as on HDFS (MR mode)**. Map-Reduce Output follows local output.
3. The output count has been kept to **10** on purpose, to demonstrate that the logic is correct.

## Input Files:

### Local Mode

```
[acadgild@localhost case_study]$ ll
total 695188
-rw-rw-r--. 1 acadgild acadgild 2283388 Jul 17 05:28 movies.csv
-rw-rw-r--. 1 acadgild acadgild 842 Jul 19 10:49 movies_small.csv
-rw-rw-r--. 1 acadgild acadgild 2604 Jul 17 10:23 pig_1531803218149.log
-rw-rw-r--. 1 acadgild acadgild 709550294 Jul 17 05:28 ratings.csv
-rw-rw-r--. 1 acadgild acadgild 410 Jul 19 10:49 ratings_small.csv
-rw-rw-r--. 1 acadgild acadgild 3341 Jul 31 18:33 Task1.pig
-rw-rw-r--. 1 acadgild acadgild 2990 Jul 31 18:34 Task2.pig
-rw-rw-r--. 1 acadgild acadgild 4146 Jul 31 18:41 Task3.pig
[acadgild@localhost case_study]$ pwd
/home/acadgild/Desktop/TestHadoop/pig/case_study
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost case_study]$
```

### MapReduce Mode

```
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/pig/case_study/
18/07/31 20:54:21 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 2283388 2018-07-31 19:38 /hadoopdata/pig/case_study/movies.csv
-rw-r--r-- 1 acadgild supergroup 709550294 2018-07-31 19:38 /hadoopdata/pig/case_study/ratings.csv
[acadgild@localhost ~]$
```

## Task 1(Local Mode):

What are the movie titles that the user has rated? & Not rated???

**Ans:**

**pig -brief -x local Task1.pig**

**-brief** - Ignores Info messages & other unwanted messages dump over screen )

**ScreenShot:**

**Output (Rated):**

```
job_1533041762263_0014 -> job_1533041762263_0015
```

```
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Success!
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
Total input paths to process : 1
Total output paths to process : 1
(Jumanji (1995))
(Grumpier Old Men (1995))
(Waiting to Exhale (1995))
(Father of the Bride Part II (1995))
(Heat (1995))
(Sabrina (1995))
(Tom and Huck (1995))
(Sudden Death (1995))
(GoldenEye (1995))
(Toy Story (1995))
```

## Output (Not Rated):

```
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1672383806_0004    ->    job_local909731179_0005,
job_local909731179_0005 ->    job_local986208195_0006,
job_local986208195_0006

Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Success!
fs.default.name is deprecated. Instead, use fs.defaultFS
SchemaTupleBackend has already been initialized
Total input paths to process : 1
Total input paths to process : 1
("Trespasser")
("Blue Blood (2006)")
("Operator 13 (1934)")
("White Banners (1938)")
("Music in the Air (1934)")
("Parenti serpenti (1992)")
("Man on a Tightrope (1953)")
("Bling: A Planet Rock (2007)")
("Jane Austen in Manhattan (1980)")
("Turtles Are Surprisingly Fast Swimmers (Turtles Swim Faster Than Expected) (Kame wa igai to hayaku oyogu) (2005)")
Pig script completed in 5 minutes, 23 seconds and 532 milliseconds (323532 ms)
You have new mail in /var/spool/mail/acadgild
acadgild@acadgild:~$ cat /dev/null
```

## Task 1(Map-Reduce Mode):

What are the movie titles that the user has rated? & Not rated???

**Ans:**

**pig -brief** Task1.pig

(**pig** - runs pig script in MR mode. This is the default mode.

**-brief** - Ignores Info messages & other unwanted messages dump over screen )

**Note:** Program files are properly documented for a detailed description of each instruction used within the program along with sample inputs.

```
[acagild@localhost case_study]$ pig -brief Task1.pig
18/07/31 19:47:31 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/31 19:47:31 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/31 19:47:31 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Loading error messages to: /home/acagild/Desktop/TestHadoop/pig/case_study/pig_1533046651431.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acagild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acagild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Default bootstrap file /home/acagild/.pigbootstrap not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: hdfs://localhost:8020
Pig Script ID for the session: PIG-Task1.pig-b500930e-b615-4c31-90b2-8c78adc1934b
ATS is disabled since yarn.timeline-service.enabled set to false
fs.default.name is deprecated. Instead, use fs.defaultFS
fs.default.name is deprecated. Instead, use fs.defaultFS
Pig features used in the script: HASH_JOIN, GROUP_BY, LIMIT
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schema tuple] was not set... will not generate code.
TRUE, ENABLED=>{AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter}}
Columns pruned for A: $2
Columns pruned for A1: $0, $2
Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
File concatenation threshold: 100 optimistic? false
Rewrite: POPackage->POForEach to POPackage(JoinPackager)
MR plan size before optimization: 3
MR plan size after optimization: 3
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to ResourceManager at localhost/127.0.0.1:8032
Pig script settings are added to the job
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
```

```

job_1533041762263_0015      ->      job_1533041762263_0015
job_1533041762263_0015

Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Success!
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
Total input paths to process : 1
Total input paths to process : 1
(Jumanji (1995))
(Grumpier Old Men (1995))
(Waiting to Exhale (1995))
(Father of the Bride Part II (1995))
(Heat (1995))
(Sabrina (1995))
(Tom and Huck (1995))
(Sudden Death (1995))
(GoldenEye (1995))
(Toy Story (1995))

```

```

Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Success!
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
Total input paths to process : 1
Total input paths to process : 1
(*Trespasser)
(Blue Blood (2006))
(Operator 13 (1934))
(White Banners (1938))
(Music in the Air (1934))
(Parents: serpenti (1992))
(Man on a Tightrope (1953))
(Bling: A Planet Rock (2007))
(Jane Austen in Manhattan (1980))
(Turtles Are Surprisingly Fast Swimmers (Turtles Swim Faster Than Expected) (Kame wa igai to hayaku oyogu) (2005))
Pig script completed in 14 minutes, 27 seconds and 342 milliseconds (867342 ms)
You have new mail in /var/spool/mail/acadgild

```

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

**pig -brief -x local Task2.pig**

**-brief** - Ignores Info messages & other unwanted messages dump over screen )

**Note:** Program files are properly documented for a detailed description of each instruction used within the program along with sample inputs.

## ScreenShot:

```
[acadgild@localhost case_study]$ pig -brief -x local Task2.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/07/31 19:20:31 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/31 19:20:31 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/case_study/pig_1533045031570.log
user.name is deprecated. Instead, use mapreduce.job.user.name
Default bootstrap file /home/acadgild/.pigbootstrap not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: file:///
Pig Script ID for the session: PIG-Task2.pig-29f3a2bd-7ed8-44fb-a377-971c3b9c5133
ATS is disabled since yarn.timeline-service.enabled set to false
fs.default.name is deprecated. Instead, use fs.defaultFS
fs.default.name is deprecated. Instead, use fs.defaultFS
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Pig features used in the script: HASH_JOIN, GROUP_BY, LIMIT
fs.default.name is deprecated. Instead, use fs.defaultFS
{RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
Columns pruned for A: $2
Selected heap (Tenured) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
File concatenation threshold: 100 optimistic? false
Choosing to move algebraic foreach to combiner
Rewrite: POForEach->POForEach to POPackage(JoinPackager)
MR plan size before optimization: 3
MR plan size after optimization: 3
fs.default.name is deprecated. Instead, use fs.defaultFS
session.id is deprecated. Instead, use dfs.metrics.session-id
Initializing JVM Metrics with processName=JobTracker, sessionId=
Pig script settings are added to the job
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
```

## Output:

```
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

```
Job DAG:
job_local1209966198_0001 -> job_local386285371_0002,
job_local386285371_0002 -> job_local2119680997_0003,
job_local2119680997_0003
```

```
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Success!
```

```
fs.default.name is deprecated. Instead, use fs.defaultFS
SchemaTupleBackend has already been initialized
Total input paths to process : 1
Total input paths to process : 1
```

```
(Jumanji (1995),26060)
(Grumpier Old Men (1995),15497)
(Waiting to Exhale (1995),2981)
(Father of the Bride Part II (1995),15258)
(Heat (1995),27895)
(Sabrina (1995),15157)
(Tom and Huck (1995),1521)
(Sudden Death (1995),4423)
(GoldenEye (1995),32537)
(Toy Story (1995),66008)
```

```
Pig script completed in 3 minutes, 20 seconds and 249 milliseconds (200249 ms)
You have new mail in /var/spool/mail/acadgild
```

## Task 2(Map-Reduce Mode):

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

**Ans:**

**pig -brief Task2.pig**

(pig - runs pig script in MR mode. This is the default mode.)



**Note:** Program files are properly documented for a detailed description of each instruction used within the program along with sample inputs.

```

acadgild@localhost: case_study$ pig -brief Task2.pig
18/07/31 20:21:10 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/31 20:21:10 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/31 20:21:10 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/case_study/pig_1533048670637.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Default bootstrap file /home/acadgild/.pigbootstrap not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: hdfs://localhost:8020
Pig Script ID for the session: PIG-Task2.pig-a9009206-cc0f-41cl-a5e8-e5baeb2f1eb0
ATS is disabled since yarn.timeline-service.enabled set to false
fs.default.name is deprecated. Instead, use fs.defaultFS
fs.default.name is deprecated. Instead, use fs.defaultFS
Pig features used in the script: HASH, JOIN, GROUP BY, LIMIT
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
{RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
Columns pruned for A: $2
Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
File concatenation threshold: 100 optimistic? false
Choosing to move algebraic foreach to combiner
Rewrite: POPackage->POForEach to POPackage(JoinPackager)
MR plan size before optimization: 3
MR plan size after optimization: 3
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to ResourceManager at localhost/127.0.0.1:8032
Pig script settings are added to the job
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent

```

```
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Success!
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
Total input paths to process : 1
Total input paths to process : 1
(Jumanji (1995),26060)
(Grumpier Old Men (1995),15497)
(Waiting to Exhale (1995),2981)
(Father of the Bride Part II (1995),15258)
(Heat (1995),27895)
(Sabrina (1995),15157)
(Tom and Huck (1995),1521)
(Sudden Death (1995),44237)
(GoldenEye (1995),32534)
(Toy Story (1995),65088)
```

Pig script completed in 7 minutes, 37 seconds and 248 milliseconds (457248 ms)  
You have new mail in /var/spool/mail/acadgild

## Task 3:

What is the average rating given for a movie?

**Ans:**

**pig -brief -x local Task3.pig**

(**pig -x local** - runs pig command in local mode<since it is very large file, running locally>

**-brief** - Ignores Info messages & other unwanted messages dump over screen )

**Note:** Program files are properly documented for a detailed description of each instruction used within the program along with sample inputs.

**ScreenShot:**

```
[acadgild@localhost case_study]$ pig -brief -x local Task3.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/07/31 19:25:30 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/31 19:25:30 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/case_study/pig_1533045330889.log
user.name is deprecated. Instead, use mapreduce.job.user.name
Default bootstrap file /home/acadgild/.pigbootstrap not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: file:///
Pig Script ID for the session: PIG-Task3.pig-4d128c67-3d9c-4edf-b984-c9b75e21883d
ATS is disabled since yarn.timeline-service.enabled set to false
fs.default.name is deprecated. Instead, use fs.defaultFS
fs.default.name is deprecated. Instead, use fs.defaultFS
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Pig features used in the script: HASH JOIN, GROUP BY, LIMIT
fs.default.name is deprecated. Instead, use fs.defaultFS
{RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
Columns pruned for A: $2
Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
File concatenation threshold: 100 optimistic? false
Choosing to move algebraic foreach to combiner
Rewrite: POPackage->POForEach to POPackage(JoinPacker)
MR plan size before optimization: 3
MR plan size after optimization: 3
fs.default.name is deprecated. Instead, use fs.defaultFS
session.id is deprecated. Instead, use dfs.metrics.session-id
Initializing JVM Metrics with processName=JobTracker, sessionId=
Pig script settings are added to the job
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
```



## Output:

```
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1150606481_0001      ->      job_local518922245_0002,
job_local518922245_0002 ->      job_local968437510_0003,
job_local968437510_0003

Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
Success!
fs.default.name is deprecated. Instead, use fs.defaultFS
SchemaTupleBackend has already been initialized
Total input paths to process : 1
Total input paths to process : 1
(Jumanji (1995),3.2369531849577897)
(Grumpier Old Men (1995),3.1755501064722202)
(Waiting to Exhale (1995),2.8757128480375713)
(Father of the Bride Part II (1995),3.079564818455892)
(Heat (1995),3.841763756945689)
(Sabrina (1995),3.372105297882167)
(Tom and Huck (1995),3.1291913214990137)
(Sudden Death (1995),3.008365362875876)
(GoldenEye (1995),3.431840536054589)
(Toy Story (1995),3.8881574960610834)
Pig script completed in 3 minutes, 10 seconds and 328 milliseconds (190328 ms)
You have new mail in /var/spool/mail/acadgild
facadnild@localhost: case studies █
```

## Task 3(Map-Reduce Mode):

What is the average rating given for a movie?

**Ans:**

**pig -brief Task3.pig**

(**pig** - runs pig script in MR mode. This is the default mode.

**-brief** - Ignores Info messages & other unwanted messages dump over screen )

**Note:** Program files are properly documented for a detailed description of each instruction used within the program along with sample inputs.

```
[acagild@localhost case_study]$ pig -brief Task3.pig
18/07/31 20:37:10 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/31 20:37:10 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/31 20:37:10 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
Apache Pig version 0.16.0 (r1746538) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acagild/Desktop/TestHadoop/pig/case_study/pig_1533049630345.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acagild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acagild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Default bootstrap file /home/acagild/.pigbootstrap not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: hdfs://localhost:8020
Pig Script ID for the session: PIG-Task3.pig-53e31870-1e99-4be6-bbbc-1eea9bd56edd
ATS is disabled since yarn.timeline-service.enabled set to false
fs.default.name is deprecated. Instead, use fs.defaultFS
fs.default.name is deprecated. Instead, use fs.defaultFS
Pig features used in the script: HASH_JOIN, GROUP_BY, LIMIT
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
RULES_ENABLED={AdaptJoin, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInsertter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInsertter}}
Columns pruned for A: $2
Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
File concatenation threshold: 100 optimistic? false
Choosing to move algebraic foreach to combiner
Rewrite: POPackage->POForEach->POPackage(JoinPackager)
MR plan size before optimization: 3
MR plan size after optimization: 3
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to ResourceManager at localhost/127.0.1:8032
Pig script settings are added to the job
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
```

**Output:**

```
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCESSFUL. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCESSFUL. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCESSFUL. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCESSFUL. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCESSFUL. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCESSFUL. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCESSFUL. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCESSFUL. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCESSFUL. Redirecting to job history server
Success!
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
Total input paths to process : 1
Total input paths to process : 1
(Jumanji (1995),3.2369531849577897)
(Grumpier Old Men (1995),3.175501064722202)
(Waiting to Exhale (1995),2.8757128480375713)
(Father of the Bride Part II (1995),3.079564018455892)
(Heat (1995),3.841763756945689)
(Sabrina (1995),3.372105297882167)
(Tom and Huck (1995),3.12919131214990137)
(Sudden Death (1995),3.008365362875876)
(GoldenEye (1995),3.431846536054589)
(Toy Story (1995),3.88815749660610834)
Pig script completed in 8 minutes, 34 seconds and 366 milliseconds (514366 ms)
You have new mail in /var/spool/mail/acadgild
load_devel localhost_casa_studylib
```

\*\*\*\*\*

End

\*\*\*\*\*