

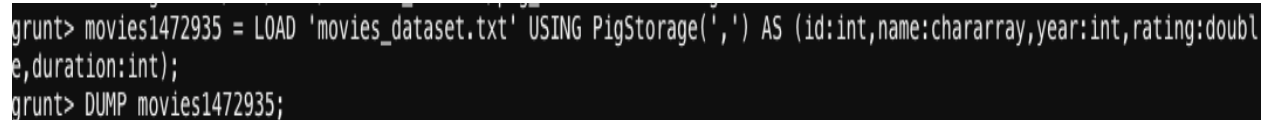
Step 1:

Loading of dataset into movies relation:

Command:

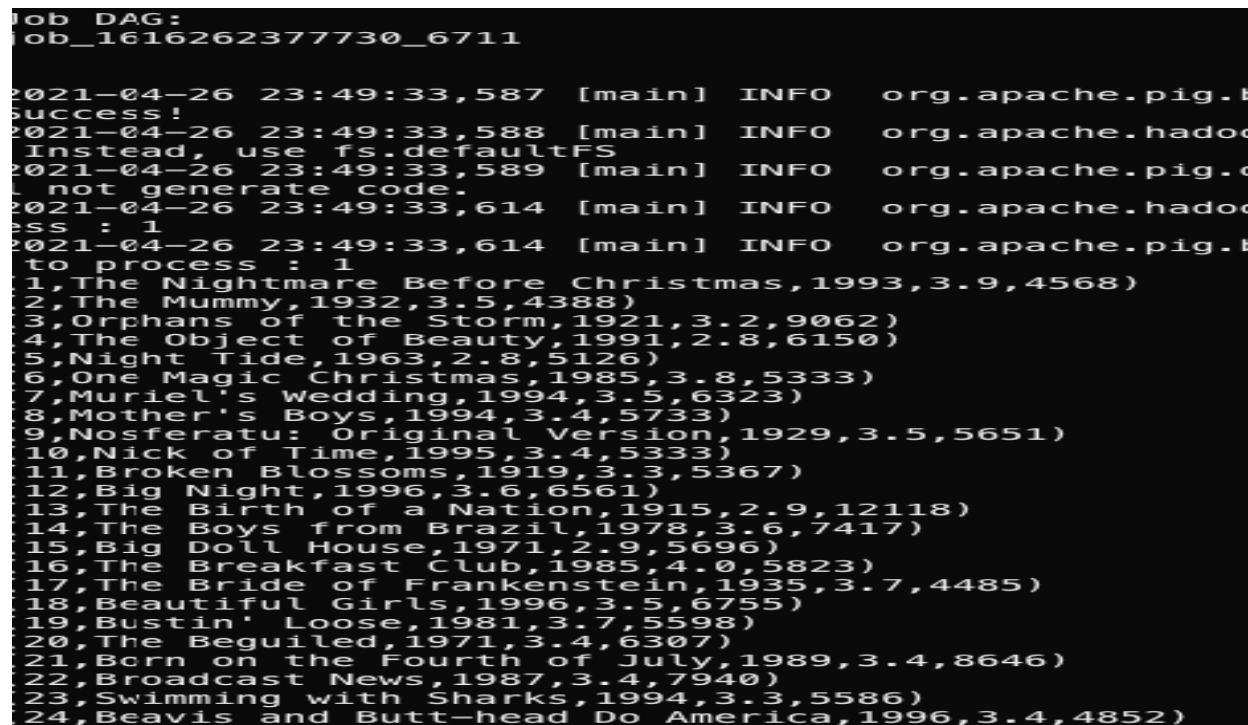
```
movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS
(id:int,name:chararray,year:int,rating:double,duration:int);
```

Screenshot:



```
grunt> movies1472935 = LOAD 'movies_dataset.txt' USING PigStorage(',') AS (id:int,name:chararray,year:int,rating:double,duration:int);
grunt> DUMP movies1472935;
```

Results:



```
Job DAG:
Job_1616262377730_6711

2021-04-26 23:49:33,587 [main] INFO org.apache.pig.b
success!
2021-04-26 23:49:33,588 [main] INFO org.apache.hadoop
Instead, use fs.defaultFS
2021-04-26 23:49:33,589 [main] INFO org.apache.pig.c
not generate code.
2021-04-26 23:49:33,614 [main] INFO org.apache.hadoop
ess : 1
2021-04-26 23:49:33,614 [main] INFO org.apache.pig.b
to process : 1
1,The Nightmare Before Christmas,1993,3.9,4568)
2,The Mummy,1932,3.5,4388)
3,Orphans of the Storm,1921,3.2,9062)
4,The Object of Beauty,1991,2.8,6150)
5,Night Tide,1963,2.8,5126)
6,One Magic Christmas,1985,3.8,5333)
7,Muriel's Wedding,1994,3.5,6323)
8,Mother's Boys,1994,3.4,5733)
9,Nosferatu: Original Version,1929,3.5,5651)
10,Nick of Time,1995,3.4,5333)
11,Broken Blossoms,1919,3.3,5367)
12,Big Night,1996,3.6,6561)
13,The Birth of a Nation,1915,2.9,12118)
14,The Boys from Brazil,1978,3.6,7417)
15,Big Doll House,1971,2.9,5696)
16,The Breakfast Club,1985,4.0,5823)
17,The Bride of Frankenstein,1935,3.7,4485)
18,Beautiful Girls,1996,3.5,6755)
19,Bustin' Loose,1981,3.7,5598)
20,The Beguiled,1971,3.4,6307)
21,Born on the Fourth of July,1989,3.4,8646)
22,Broadcast News,1987,3.4,7940)
23,Swimming with Sharks,1994,3.3,5586)
24,Beavis and Butt-head Do America,1996,3.4,4852)
```

A. Find the number of movies released between 1950 and 1960.

Command:

```
movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS
(id:int,name:chararray,year:int,rating:double,duration:int);
filtered_movies = FILTER movies BY year > 1949 AND year < 1961;
grouped_movies = GROUP filtered_movies ALL;
counted_movies = FOREACH grouped_movies GENERATE group, COUNT(filtered_movies);
DUMP counted_movies;
```

Screenshot:

```

grunt> movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS (id:int,name:
chararray,year:int,rating:double,duration:int);
grunt> filtered_movies = FILTER movies BY year > 1949 AND year < 1961;
grunt> grouped_movies = GROUP filtered_movies ALL;
grunt> counted_movies = FOREACH grouped_movies GENERATE group, COUNT(filtered_mo
vies);
grunt> DUMP counted_movies;

```

Results:

```

Input(s):
Successfully read 49590 records (2942953 bytes) from: "hdfs://nameservice1/user/
edureka_1472935/movies_dataset.txt"


Output(s):
Successfully stored 1 records (13 bytes) in: "hdfs://nameservice1/tmp/temp-20356
41038/tmp1680437410"

Counters:
Total records written : 1
Total bytes written : 13
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6743

2021-04-27 02:12:52,260 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2021-04-27 02:12:52,260 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-27 02:12:52,261 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2021-04-27 02:12:52,263 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2021-04-27 02:12:52,263 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(all,547)
grunt>

```



B. Find the number of movies having rating more than 4.

Command:

```

movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS
(id:int,name:chararray,year:int,rating:double,duration:int);
filtered_movies = FILTER movies BY rating > 4.0;
grouped_movies = GROUP filtered_movies all;
counted_movies = FOREACH grouped_movies GENERATE group, COUNT(filtered_movies);
DUMP counted_movies;

```

Screenshot:

```
grunt> movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS (id:int,name:
chararray,year:int,rating:double,duration:int);
grunt> filtered_movies = FILTER movies BY rating > 4.0;
grunt> grouped_movies = GROUP filtered_movies all;
grunt> counted_movies = FOREACH grouped_movies GENERATE group, COUNT(filtered_mo
vies);
grunt> DUMP counted_movies;
```

Result:

```
Input(s):
Successfully read 49590 records (2942953 bytes) from: "hdfs://nameservice1/user/
edureka_1472935/movies_dataset.txt"

Output(s):
Successfully stored 1 records (13 bytes) in: "hdfs://nameservice1/tmp/temp-20356
41038/tmp-535700543"

Counters:
Total records written : 1
Total bytes written : 13
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6742

2021-04-27 02:10:39,274 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2021-04-27 02:10:39,274 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-27 02:10:39,275 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2021-04-27 02:10:39,277 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2021-04-27 02:10:39,277 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(all,897)
grunt> █
```

C. Find the movies whose rating are between 3 and 4.**Command:**

```
movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS
(id:int,name:chararray,year:int,rating:double,duration:int);
filtered_movies = FILTER movies BY rating < 4.01 AND rating > 2.99;
DUMP filtered_movies;
```

Screenshot:

```
grunt> movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS (id:int,name:
chararray,year:int,rating:double,duration:int);
grunt> filtered_movies = FILTER movies BY rating < 4.01 AND rating > 2.99;
grunt> DUMP filtered_movies;
```

Result:

```
(1,The Nightmare Before Christmas,1993,3.9,4568)
(2,The Mummy,1932,3.5,4388)
(3,Orphans of the Storm,1921,3.2,9062)
(6,One Magic Christmas,1985,3.8,5333)
(7,Muriel's Wedding,1994,3.5,6323)
(8,Mother's Boys,1994,3.4,5733)
(9,Nosferatu: Original Version,1929,3.5,5651)
(10,Nick of Time,1995,3.4,5333)
(11,Broken Blossoms,1919,3.3,5367)
(12,Big Night,1996,3.6,6561)
(14,The Boys from Brazil,1978,3.6,7417)
(16,The Breakfast Club,1985,4.0,5823)
(17,The Bride of Frankenstein,1935,3.7,4485)
(18,Beautiful Girls,1996,3.5,6755)
(19,Bustin' Loose,1981,3.7,5598)
(20,The Beguiled,1971,3.4,6307)
(21,Born on the Fourth of July,1989,3.4,8646)
(22,Broadcast News,1987,3.4,7940)
(23,Swimming with Sharks,1994,3.3,5586)
(24,Beavis and Butt-head Do America,1996,3.4,4852)
(25,Brighton Beach Memoirs,1986,3.4,6564)
(26,The Best of Times,1986,3.4,6247)
(27,Brassed Off,1996,3.5,6040)
(28,Last Tango in Paris,1972,3.1,7732)
(29,Leprechaun 2,1994,3.2,5125)
(30,Incident at Oglala: The Leonard Peltier Story,1992,3.7,5487)
(31,Kalifornia,1993,3.4,7095)
(32,The Lady Vanishes,1938,3.7,5762)
(33,Jingle All the Way,1996,3.6,5371)
(34,Killing Zoe,1993,3.4,5773)
```

D. Find the number of movies with duration more than 2 hours (7200 second).

Command:

```
movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS
(id:int,name:chararray,year:int,rating:double,duration:int);
filtered_movies = FILTER movies BY duration > 7200;
grouped_movies = GROUP filtered_movies ALL;
counted_movies = FOREACH grouped_movies GENERATE group, COUNT(filtered_movies);
DUMP counted_movies;
```

Screenshot:

```

grunt> movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS (id:int,name:
chararray,year:int,rating:double,duration:int);
grunt> filtered_movies = FILTER movies BY duration > 7200;
grunt> grouped_movies = GROUP filtered_movies ALL;
grunt> counted_movies = FOREACH grouped_movies GENERATE group, COUNT(filtered_mo
vies);
grunt> DUMP counted_movies;

```

Result:

```

Input(s):
Successfully read 49590 records (2942953 bytes) from: "hdfs://nameservice1/user/
edureka_1472935/movies_dataset.txt"

Output(s):
Successfully stored 1 records (13 bytes) in: "hdfs://nameservice1/tmp/temp-20356
41038/tmp-807611883"

Counters:
Total records written : 1
Total bytes written : 13
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6746

2021-04-27 02:16:48,190 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2021-04-27 02:16:48,190 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-27 02:16:48,191 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2021-04-27 02:16:48,193 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2021-04-27 02:16:48,193 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(all,641)
grunt> 

```

E. Find the list of years and number of movies released each year.

Command:

```

movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS
(id:int,name:chararray,year:int,rating:double,duration:int);
grouped_movies = GROUP movies BY year;
counted_movies = FOREACH grouped_movies GENERATE group, COUNT(movies.id);
DUMP counted_movies;


```


Screenshot:

```
grunt> movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS (id:int,name:
chararray,year:int,rating:double,duration:int);
grunt> grouped_movies = GROUP movies BY year;
grunt> counted_movies = FOREACH grouped_movies GENERATE group, COUNT(movies.id);
grunt> DUMP counted_movies;
```

Result:

(1913,3)	(1950,10)	
(1914,20)	(1951,33)	
(1915,1)	(1952,15)	(1989,421)
(1916,1)	(1953,17)	(1990,470)
(1918,1)	(1954,17)	(1991,364)
(1919,3)	(1955,14)	(1992,479)
(1920,6)	(1956,60)	(1993,564)
(1921,2)	(1957,98)	(1994,517)
(1922,2)	(1958,73)	(1995,592)
(1923,4)	(1959,87)	(1996,688)
(1924,5)	(1960,123)	(1997,788)
(1925,5)	(1961,119)	(1998,843)
(1926,2)	(1962,124)	(1999,1181)
(1927,4)	(1963,88)	(2000,902)
(1928,2)	(1964,126)	(2001,1173)
(1929,5)	(1965,104)	(2002,1117)
(1930,5)	(1966,103)	(2003,1399)
(1931,3)	(1967,279)	(2004,1381)
(1932,4)	(1968,173)	(2005,1937)
(1933,7)	(1969,124)	(2006,2416)
(1934,8)	(1970,141)	(2007,2892)
(1935,11)	(1971,131)	(2008,3358)
(1936,7)	(1972,166)	(2009,4451)
(1937,4)	(1973,168)	(2010,5107)
(1938,5)	(1974,178)	(2011,5511)
(1939,6)	(1975,134)	(2012,4339)
(1940,9)	(1976,118)	(2013,981)
(1941,7)	(1977,136)	(2014,1)
(1942,3)	(1978,231)	
(1943,7)	(1979,140)	
(1944,10)	(1980,107)	
(1945,9)	(1981,112)	
(1946,6)	(1982,153)	
(1947,9)	(1983,270)	
(1948,13)	(1984,303)	
(1949,9)	(1985,334)	
	(1986,287)	
	(1987,280)	
	(1988,334)	

grunt> 

F. Find the total number of movies in the dataset.

Command:

```
movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS
(id:int,name:chararray,year:int,rating:double,duration:int);
grouped_movies = GROUP movies ALL;
sum_movies = FOREACH grouped_movies GENERATE group, COUNT(movies.id);
DUMP sum_movies;
```

Screenshot:

```
grunt> movies = LOAD 'movies_dataset.txt' USING PigStorage(',') AS (id:int,name:
chararray,year:int,rating:double,duration:int);
grunt> grouped_movies = GROUP movies ALL;
grunt> sum_movies = FOREACH grouped_movies GENERATE group, COUNT(movies.id);
grunt> DUMP sum_movies;
```

Result:

```
Input(s):
Successfully read 49590 records (2942953 bytes) from: "hdfs://nameservice1/user/
edureka_1472935/movies_dataset.txt"

Output(s):
Successfully stored 1 records (15 bytes) in: "hdfs://nameservice1/tmp/temp-20356
41038/tmp-1286971903"

Counters:
Total records written : 1
Total bytes written : 15
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1616262377730_6758

2021-04-27 02:35:09,747 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2021-04-27 02:35:09,747 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-04-27 02:35:09,747 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2021-04-27 02:35:09,749 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2021-04-27 02:35:09,749 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(all,49590)
grunt> 
```