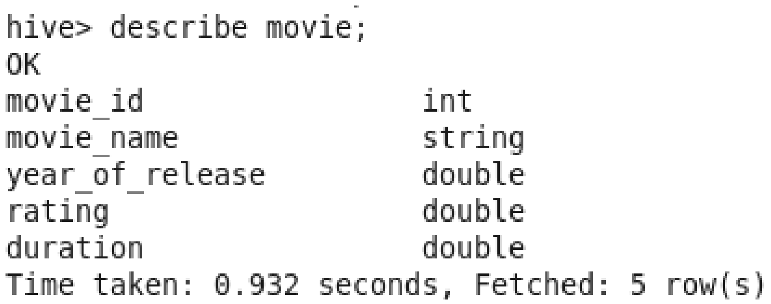
Problem Statement

1. Create table movie (movie_id int, movie_name string, year_of_release double, rating double, duration double) row format delimited by fields terminated by ‘,’;

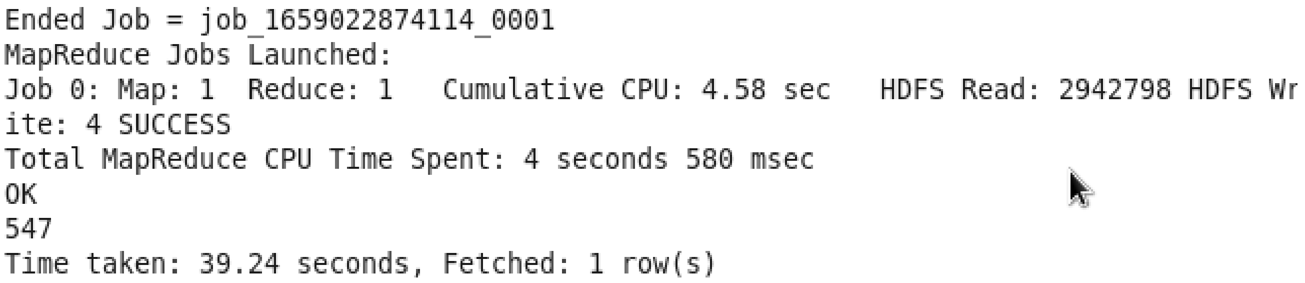


```
hive> describe movie;
OK
movie_id                int
movie_name              string
year_of_release         double
rating                  double
duration                double
Time taken: 0.932 seconds, Fetched: 5 row(s)
```

2. Load data local inpath ‘Desktop/Movie Dataset.txt’ into table movie;

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

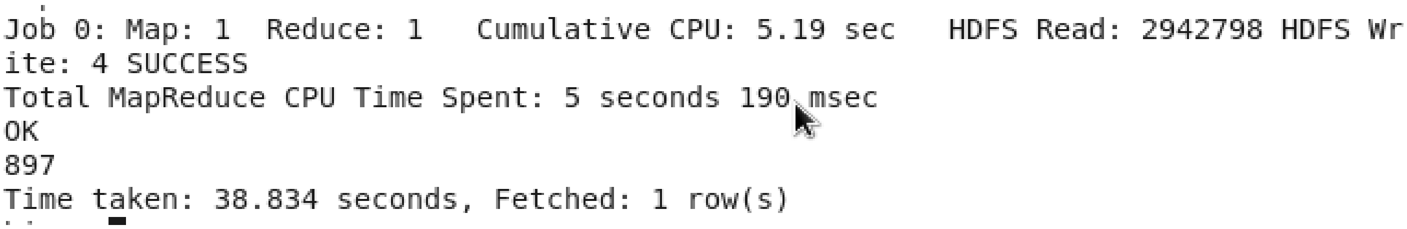
3. Select count(movie_name) from movie where year_of_release between 1950 and 1960;



```
Ended Job = job_1659022874114_0001
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 4.58 sec   HDFS Read: 2942798 HDFS Wr
ite: 4 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 580 msec
OK
547
Time taken: 39.24 seconds, Fetched: 1 row(s)
```

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

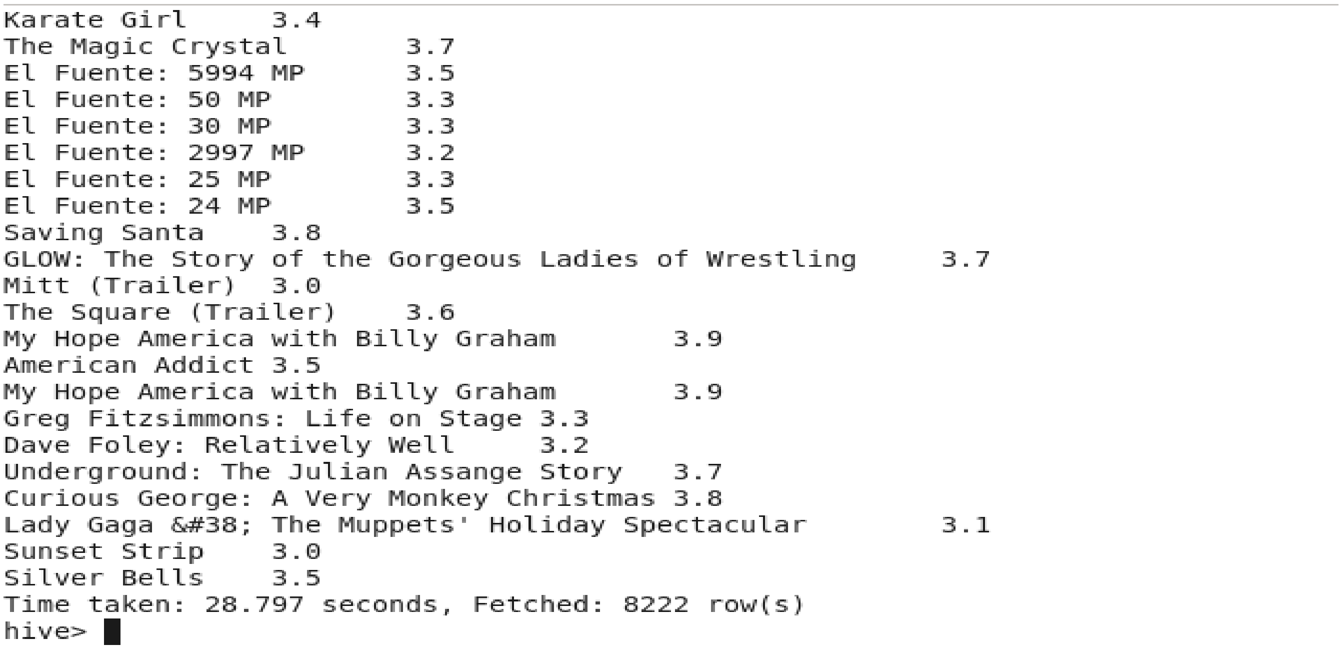
4. Select count(movie_name) from movie where rating > 4;



```
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 5.19 sec   HDFS Read: 2942798 HDFS Wr
ite: 4 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 190 msec
OK
897
Time taken: 38.834 seconds, Fetched: 1 row(s)
```

C. Find the movies whose ratings are between 3 and 4.

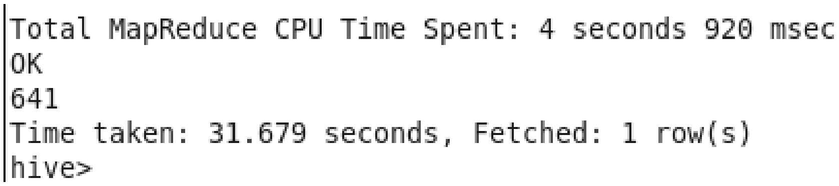
5. select movie_name, rating from movie where rating between 3 and 4;



```
Karate Girl     3.4
The Magic Crystal       3.7
El Fuente: 5994 MP      3.5
El Fuente: 50 MP        3.3
El Fuente: 30 MP        3.3
El Fuente: 2997 MP      3.2
El Fuente: 25 MP        3.3
El Fuente: 24 MP        3.5
Saving Santa    3.8
GLOW: The Story of the Gorgeous Ladies of Wrestling     3.7
Mitt (Trailer)  3.0
The Square (Trailer)    3.6
My Hope America with Billy Graham       3.9
American Addict 3.5
My Hope America with Billy Graham       3.9
Greg Fitzsimmons: Life on Stage 3.3
Dave Foley: Relatively Well     3.2
Underground: The Julian Assange Story   3.7
Curious George: A Very Monkey Christmas 3.8
Lady Gaga &#38; The Muppets' Holiday Spectacular        3.1
Sunset Strip    3.0
Silver Bells    3.5
Time taken: 28.797 seconds, Fetched: 8222 row(s)
hive>
```

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

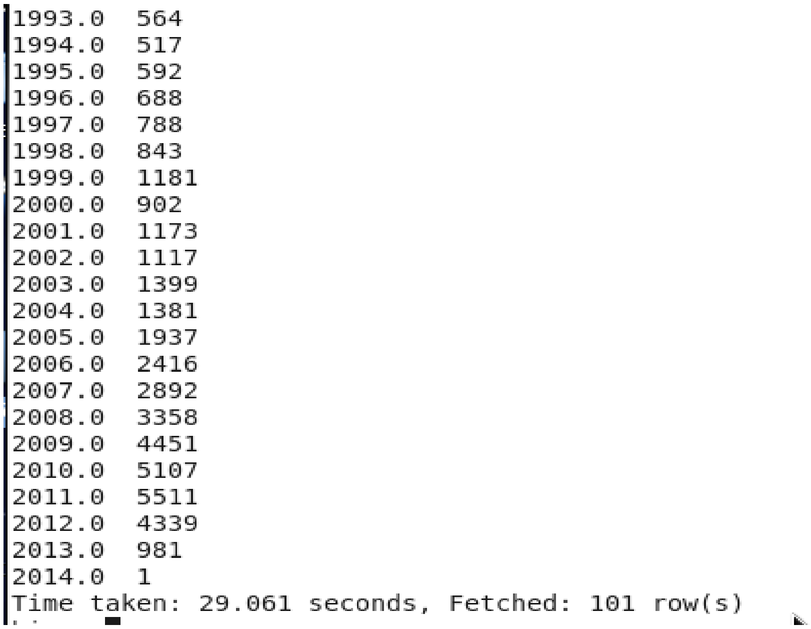
6. Select count(movie_name) from movie where duration > 7200;



```
Total MapReduce CPU Time Spent: 4 seconds 920 msec
OK
641
Time taken: 31.679 seconds, Fetched: 1 row(s)
hive>
```

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

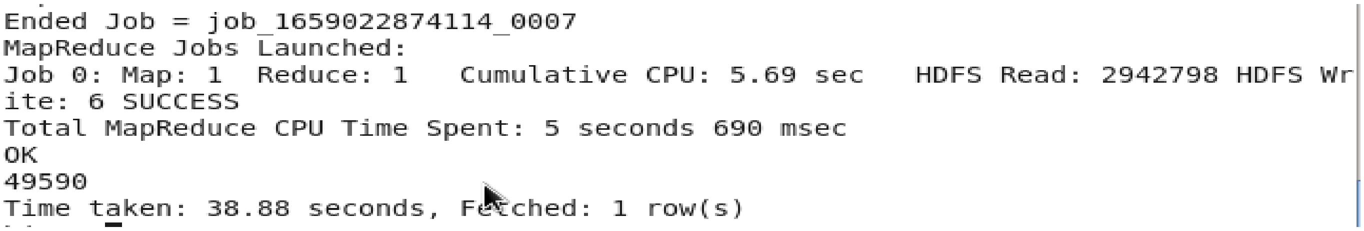
7. Select year_of_release, count(movie_name) from movie group by year_of_release;



```
1993.0  564
1994.0  517
1995.0  592
1996.0  688
1997.0  788
1998.0  843
1999.0  1181
2000.0  902
2001.0  1173
2002.0  1117
2003.0  1399
2004.0  1381
2005.0  1937
2006.0  2416
2007.0  2892
2008.0  3358
2009.0  4451
2010.0  5107
2011.0  5511
2012.0  4339
2013.0  981
2014.0  1
Time taken: 29.061 seconds, Fetched: 101 row(s)
```

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

8. Select count(movie_name) from movie;



```
Ended Job = job_1659022874114_0007
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 5.69 sec   HDFS Read: 2942798 HDFS Wr
ite: 6 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 690 msec
OK
49590
Time taken: 38.88 seconds, Fetched: 1 row(s)
```