Submitted by: Sri Harsha

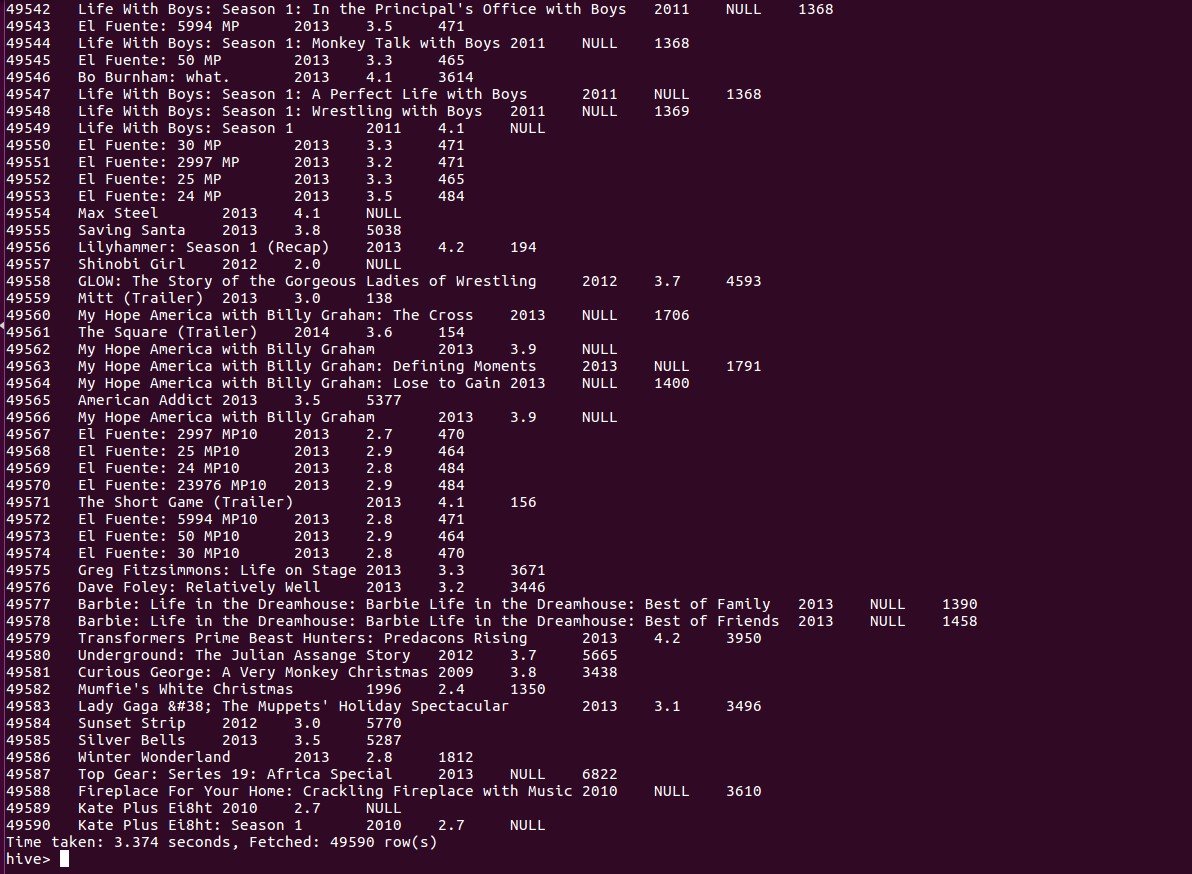
Solution for Hive:

Create table movie( movie\_id int, name string, year int, rating double, duration int ) row format delimited fields terminated by ',';

LOAD DATA LOCAL INPATH '/home/sri/Downloads/movie-datasets.txt' into TABLE movie;

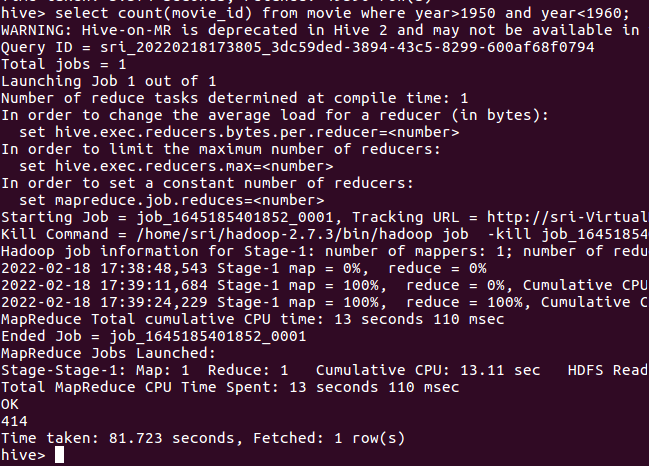
Select \* from movie

Displayed complete table after loading data:



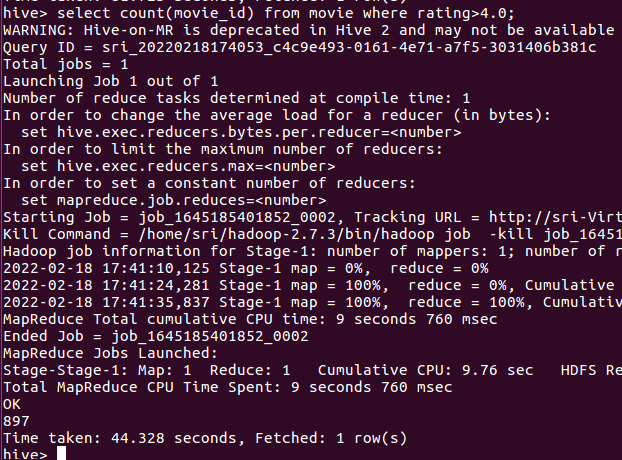
A.

select \* from movie where year>1950 and year<1960;



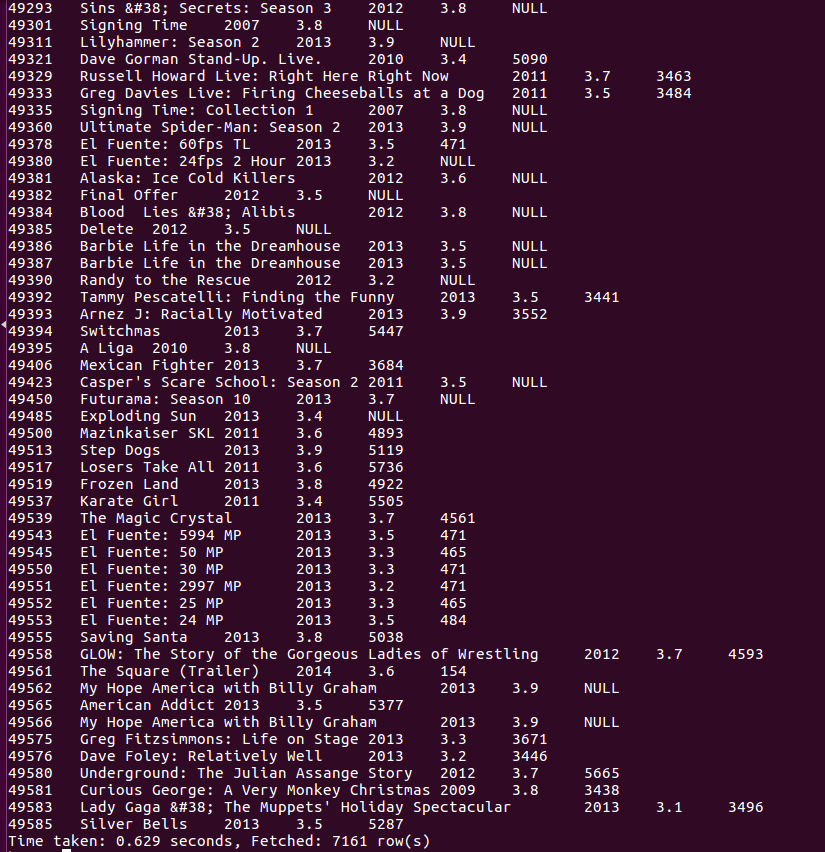
B.

select count(movie\_id) from movie where rating>4.0;



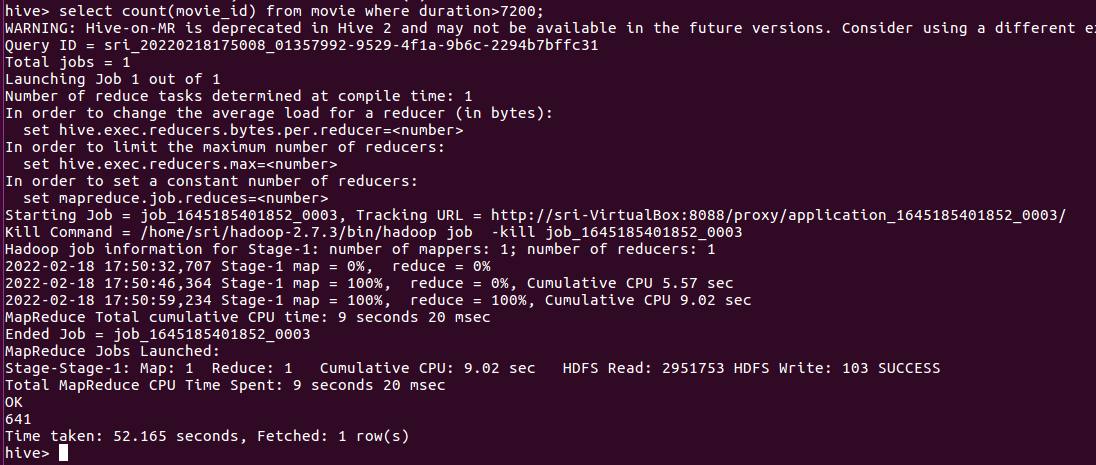
C.

select \* from movie where rating between 3 and 4;



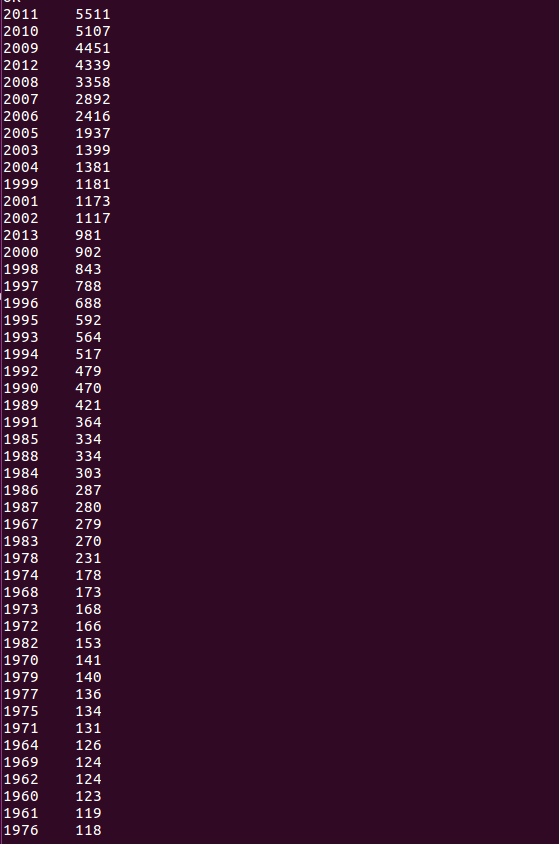
D.

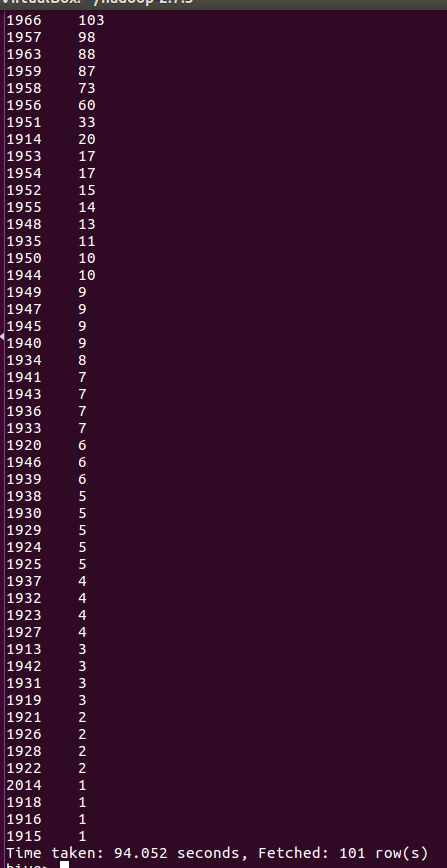
select count(movie\_id) from movie where duration>7200;



E.

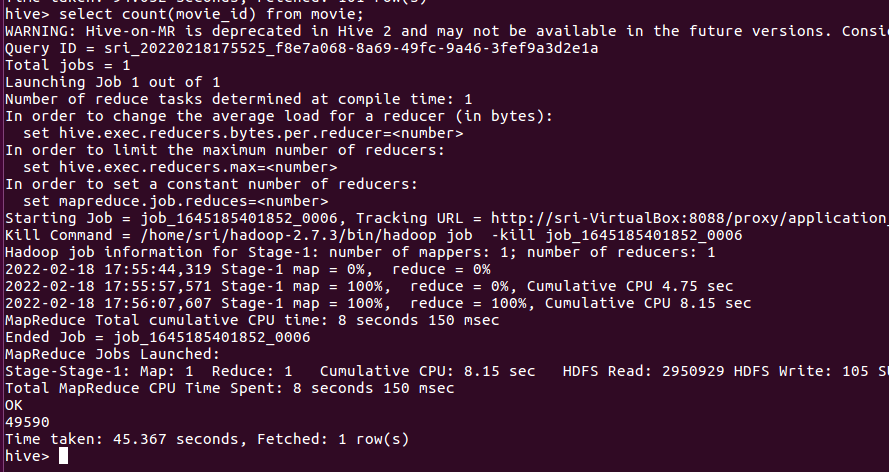
select year, count(movie\_id) as c from airport group by year order by c desc;





F.

select count(movie\_id) from movie;



Pig Solutions:

bin/hdfs dfs -put /home/sri/Downloads/movie-datasets.txt /practice\_pig

movie = load '/practice\_pig/movie-datasets.txt' USING PigStorage (',') as ( movie\_id:int, name:chararray, year:int, rating:double, duration:int );

A.

A1 = FILTER movie By year > 1950 and year < 1960;

A2 = COUNT(a1);

dump A2;

B.

B1 = FILTER movie By rating > 4;

B2 = COUNT(B1);

dump B2;

C.

C1 = FILTER movie By year > 3 and year < 4;

dump C1;

D.

D1 = FILTER movie By duration > 7200;

D2 = COUNT(B1);

dump D2;

E1 = foreach movie generate year;

E2 = group E1 by year;

E3 = foreach E2 generate group, COUNT(E1) as Movies;

dump E3;

F.

F1=COUNT(movie);

Dump F1;