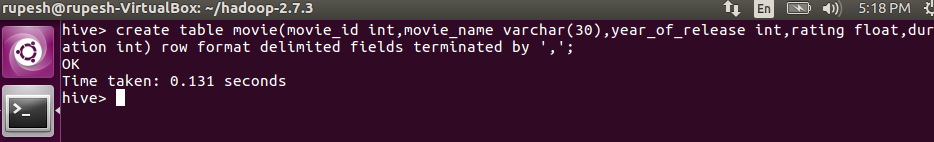
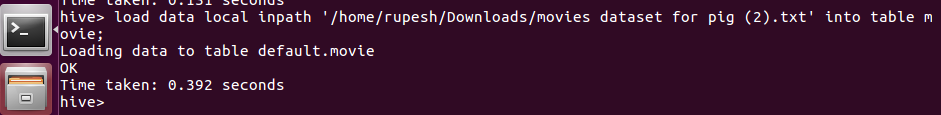
Hive

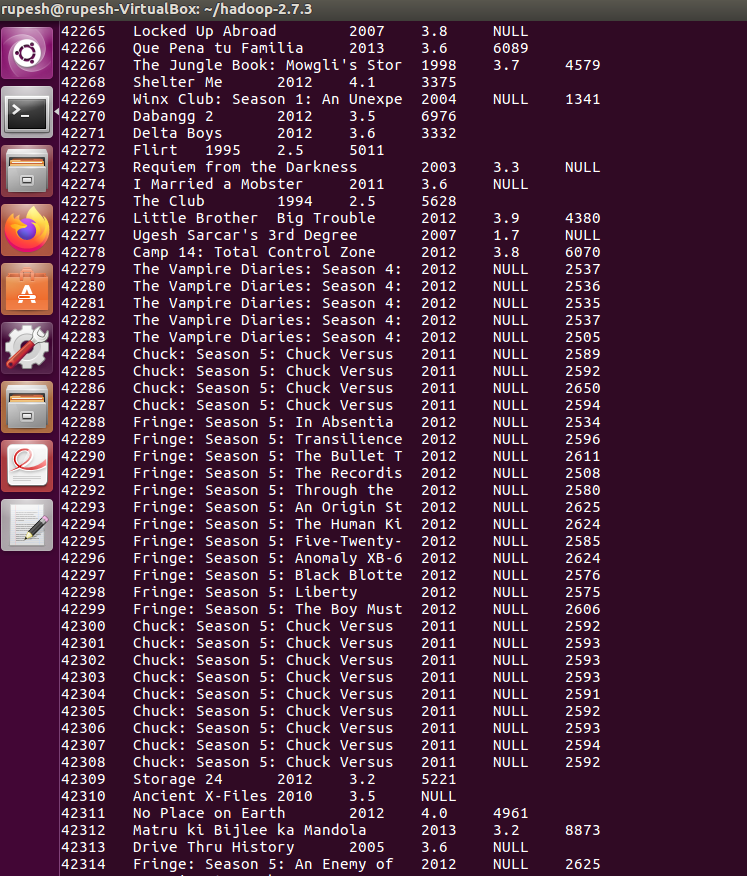
create table movie(movie\_id int,movie\_name varchar(30),year\_of\_release int,rating float,duration int) row format delimited fields terminated by ',';



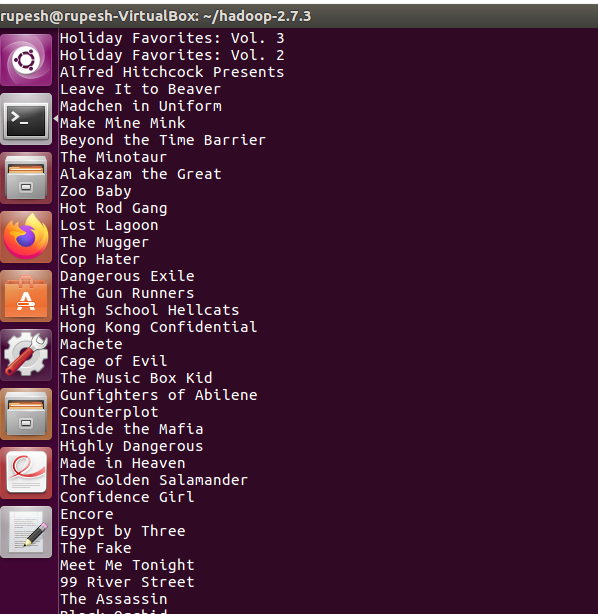
load data local inpath '/home/rupesh/Downloads/movies dataset for pig (2).txt' into table movie;



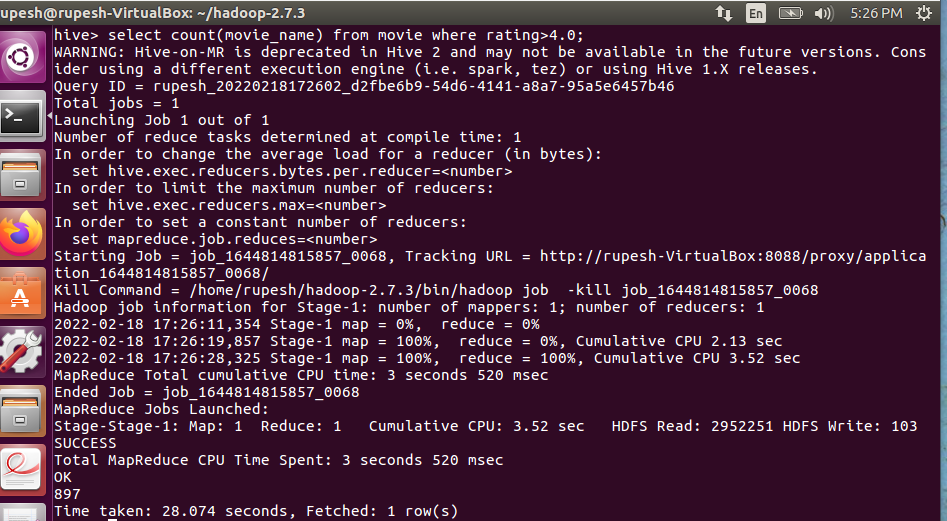
select \* from movie;



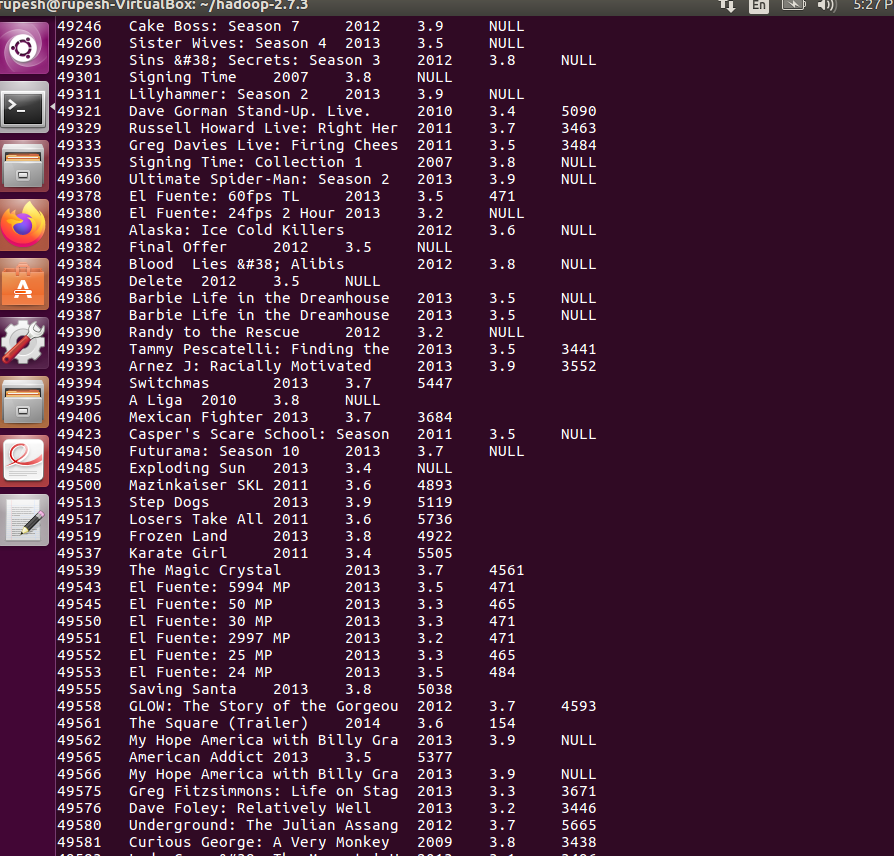
1. select \* from movie where year\_of\_release >=1950 and year\_of\_release<=1960;



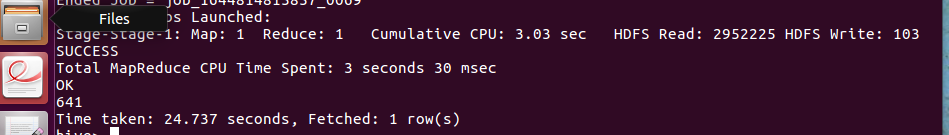
1. select count(movie\_name) from movie where rating>4.0;



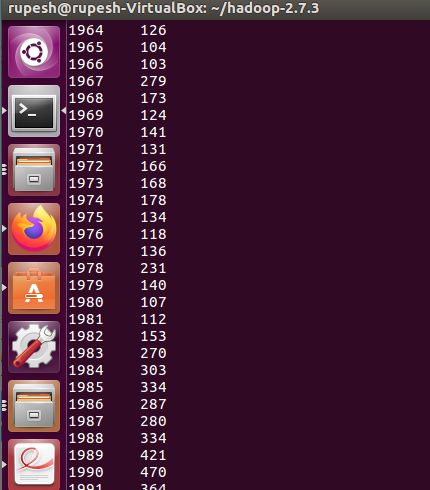
c) select \* from movie where rating>3.0 and rating<4.0;



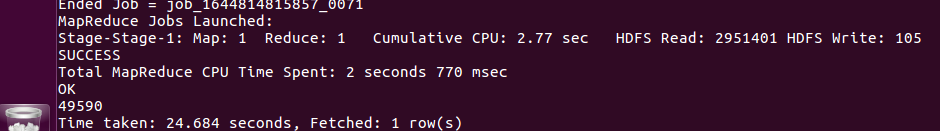
d) select count(movie\_id) from movie where duration>7200;



E) select year\_of\_release,count(movie\_id) from movie group by year\_of\_release;



F) select count(movie\_id) from movie;

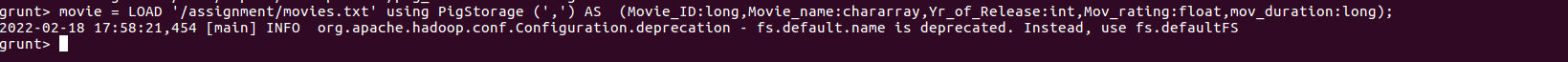


PIG

bin/hdfs dfs -put /home/rupesh/Downloads/movies.txt /assignment;



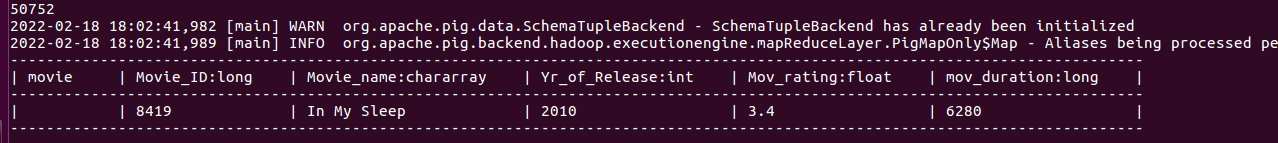
movie = LOAD '/assignment/movies.txt' using PigStorage (',') AS (Movie\_ID:long,Movie\_name:chararray,Yr\_of\_Release:int,Mov\_rating:float,mov\_duration:long);



describe movie;



illustrate movie;



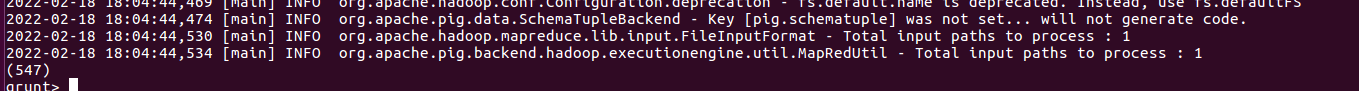
B = FILTER movie by Yr\_of\_Release >= 1950;

C = FILTER B by Yr\_of\_Release <= 1960;

Mov\_count1 = GROUP C All;

count = FOREACH Mov\_count1 GENERATE COUNT(C);

DUMP count;

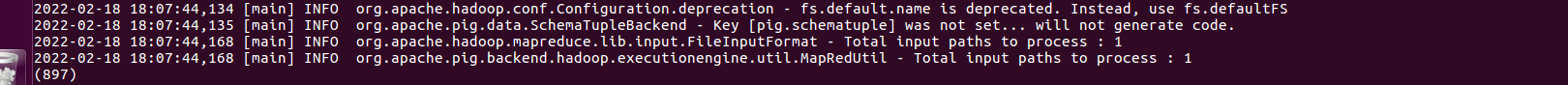


B = FILTER movie by Mov\_rating > 4.0;

Mov\_count2 = GROUP B All;

Count2 = FOREACH Mov\_count2 GENERATE COUNT(B);

DUMP count2;

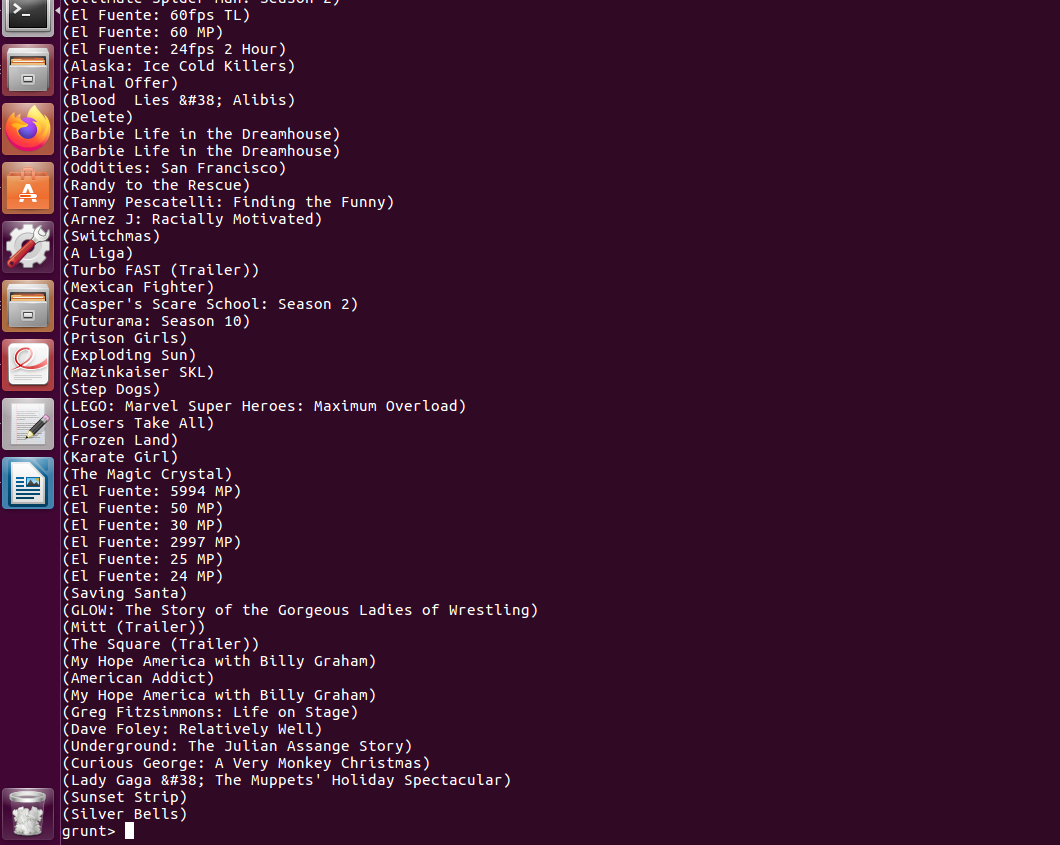


B = FILTER movie by Mov\_rating >= 3.0;

C = FILTER B by Mov\_rating <= 4.0;

Mov\_list = FOREACH C GENERATE Movie\_name;

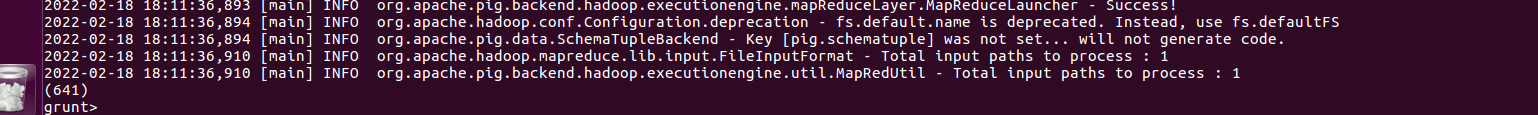
DUMP Mov\_list;

  
B = FILTER movie by mov\_duration > 7200;

Mov\_count3 = GROUP B All;

count3 = FOREACH Mov\_count3 GENERATE COUNT(B);

DUMP count3;

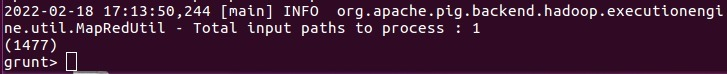


B= FOREACH movie GENERATE Yr\_of\_Release;

Yr\_list = GROUP B by Yr\_of\_Release ;

List = FOREACH Yr\_list GENERATE group, count(B) AS Num\_Movies;

DUMP List;



Mov\_count4 = GROUP A All;

count4 = FOREACH Mov\_count4 GENERATE COUNT(A);

DUMP count4;

