1. **Load the movies data into movie relation**

movies = load '/home/hduser/smallproject/movies\_data.csv' using PigStorage(‘,’) as(movieid:int,moviename:chararray,year:int,rating:float,duration:int);

1. **List all the movies that have rating greater than 4**

movie\_rating\_grt\_4 = filter movies by rating > 4;

1. **List the movies that were released between 1950 and 1960**

movies\_btwn\_1950\_1960 = filter movies by year > 1950 and year < 1960;

1. **List the movies that start with the Alphabet A**

movies\_starting\_with\_A = filter movies by moviename matches 'A.\*';

1. ***List the* movies *that have duration greater that 2 hours***

movies\_duration\_grt\_2\_hours = filter movies by duration > 7200;

1. ***List the movies that have rating between 3 and 4***

movies\_rating\_btwn\_3\_4 = filter movies by rating > 3 and rating < 4;

1. ***List the movie names its duration in minutes***

movie\_name\_duration = foreach movies generate moviename,(duration/60) as duration\_min;

1. ***List the years and the number of movies released each year.***

grpd = group movies by year;

movie\_count = foreach grpd generate group as year,COUNT(movies.movieid) as moviecount;

1. ***List all the movies analysed in dataset***

fewdetails = foreach movies generate movieid,moviename;

1. ***List all the movies in the ascending order of year.***

movies\_year\_wise\_asc\_order = order movies by year asc;

1. ***List all the movies in the descending order of year.***

movies\_year\_wise\_desc\_order = order movies by year desc;

1. ***list the distinct records***

no\_duplicates = distinct movies;

1. ***Just get first 10 movies***

top\_10\_movies = limit movies 10;

1. ***Get 10% of the whole dataset***

sample\_10\_percent = sample movies 0.1;

Verify if you have got 10% of records

grpd = group movies ALL;

movies\_count = foreach grpd generate COUNT(movies.movieid) as moviecount;

result : 49590;

grpd = group sample\_10\_percent ALL;

movies\_count = foreach grpd generate COUNT(sample\_10\_percent) as

moviecount;

result:4964;