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

grunt> movies = LOAD '/home/hduser/pig/myscripts/movies\_data.csv' USING PigStorage(',') as (id,name,year,rating,duration);

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

grunt> movies\_greater\_than\_four = FILTER movies BY (float)rating>4.0;

grunt> DUMP movies\_greater\_than\_four;

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

grunt> movies\_between\_50\_60 = FILTER movies by year>1950 and year<1960;

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

grunt> movies\_starting\_with\_A = FILTER movies by name matches 'A.\*';

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

grunt> movies\_duration\_2\_hrs = FILTER movies by duration > 7200;

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

grunt> movies\_rating\_3\_4 = FILTER movies BY rating>3.0 and rating<4.0;

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

grunt> movie\_duration = FOREACH movies GENERATE name, (double)(duration/60);

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

grunt> grouped\_by\_year = group movies by year;

grunt> count\_by\_year = FOREACH grouped\_by\_year GENERATE group, COUNT(movies);

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

grunt> group\_all = GROUP count\_by\_year ALL;

grunt> sum\_all = FOREACH group\_all GENERATE SUM(count\_by\_year.$1);

grunt> DUMP sum\_all;

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

grunt> desc\_movies\_by\_year = ORDER movies BY year ASC;

grunt> DUMP desc\_movies\_by\_year;

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

grunt> asc\_movies\_by\_year = ORDER movies by year DESC;

grunt> DUMP asc\_movies\_by\_year;

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

grunt> no\_dups = DISTINCT movies\_with\_dups;

grunt> DUMP no\_dups;

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

grunt> top\_10\_movies = LIMIT movies 10;

grunt> DUMP top\_10\_movies;

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

grunt> sample\_10\_percent = sample movies 0.1;

grunt> dump sample\_10\_percent;

Verify if you have got 10% of records

grunt> sample\_group\_all = GROUP sample\_10\_percent ALL;

grunt> sample\_count = FOREACH sample\_group\_all GENERATE COUNT(sample\_10\_percent.$0);

grunt> dump sample\_count;