To-do

Problem Statement:

What are the movie titles that the user has rated?

How many times a movie has been rated by the user?

In question 2 above, what is the average rating given for a movie?

Assignment as described above was to work with movie dataset and generate pig script and run it to generate required output.

PIG Script

```
a = load '/home/acadgild/input/movies.csv' using PigStorage(',') as (movieid:int, title:chararray, genres:chararray);
```

a1 = filter a by not title matches 'title*';

b = load '/home/acadgild/input/ratings.csv' using PigStorage(',') as (userId:int,movieId:int,rating:float,timestamp:bytearray);

c = group b by movieId;

d = foreach c generate group as movieId, COUNT(b.userId) as RatingUserPerMovieCount, SUM(b.rating)/COUNT(b.userId) as AverageRatingPerMovie;

store d into '/home/acadgild/output/MovieRatingAssignment.txt' using PigStorage(',');

e = load '/home/acadgild/output/MovieRatingAssignment.txt/part-r-00000' using PigStorage(',') as (MovieID:int, RatingUserPerMovieCount:int, AverageRatingPerMovie:float);

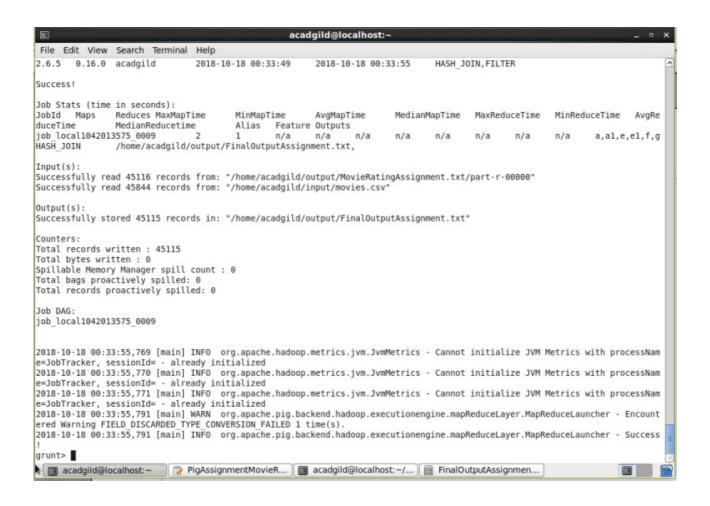
e1 = filter e by not (MovieID == 0);

f = join a1 by \$0, e1 by \$0;

g = foreach f generate \$1, \$4, \$5;

store g into '/home/acadgild/output/FinalOutputAssignment.txt' using PigStorage(',');

After running pig script successfully



Post running script final output created displaying head of the file

