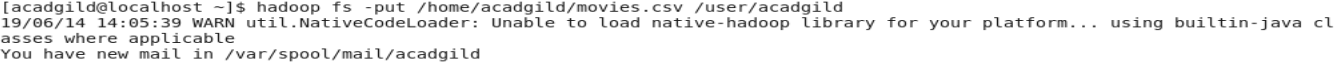
**ASSIGNMENT OF SESSION 7(PIG)**

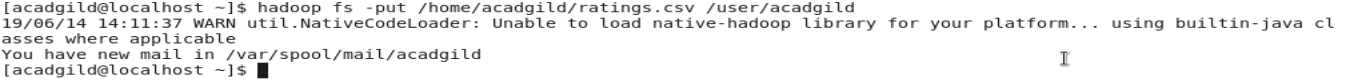
**Command to upload movies.csv into hdfs:-**

**hadoop fs –put /home/acadgild/movies.csv /user/acadgild**



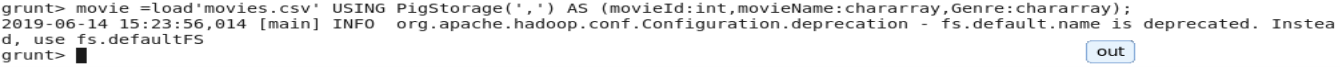
**Command to upload ratings.csv into hdfs:-**

**hadoop fs –put /home/acadgild/ratings.csv /user/acadgild**



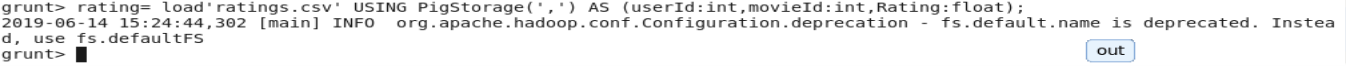
**Load movies.csv into movie:-**

**movie= load’movies.csv’ USING PigStorage(‘,’) AS (movieId:int,movieName:chararray:Genre:chararray);**



**Load ratings.csv into rating:-**

**rating= load’ratings.csv’ USING PigStorage(‘,’) AS (userId:int,movieId:int,Rating:float);**



**Join movie and rating using the foreign key column movieId.**

**movie\_rating= join movie by movieId,rating by movieId;**



**Grouping movieName in the movie\_rating**

**b = group movie\_rating by movieName;**



**For each movie generate the count and average rating using aggregate function:-**

**c= foreach b generate group as movieName,COUNT(movie\_rating)as cnt,AVG(movie\_rating.Rating) as average;**



**For each movie ,print its Title,total times it is rated and average rating**

**d = foreach c generate movieName,cnt,average;**



**Print the output on the screen:**

**It will show the movie Title, no of times it is rated and average rating**

**dump d;**

