create text file named "movies.txt" in local file system and add the following content:

1,The Nightmare Before Christmas,1993,3.9,4568

2,The Mummy,1932,3.5,4388

3,Orphans of the Storm,1921,3.2,9062

4,The Object of Beauty,1991,2.8,6150

5,Night Tide,1963,2.8,5126

6,One Magic Christmas,1985,3.8,5333

7,Muriel's Wedding,1994,3.5,6323

8,Mother's Boys,1994,3.4,5733

9,Nosferatu: Original Version,1929,3.5,5651

10,Nick of Time,1995,3.4,5333

Load the file to Hadoop

hadoop fs -put movies.txt /user/hduser/

Load contents into a variable

Movies = LOAD '/user/hduser//movies.txt' USING PigStorage(',') as (id,name,year,rating,duration);

Display contents

DUMP Movies;

Check format

Describe Movies;

Filter movies with rating

movies_greater_than_zero = FILTER Movies BY rating!=0;

extract values and store in variables

foreachexample= foreach movies_greater_than_zero generate year,rating,name;

dump foreachexample

pv_group_by_movie = GROUP Movies by name

pv_with_count = FOREACH pv_group_by_viewee {

GENERATE name, COUNT_STAR(rating) as pv_count;

Store in HDFS

STORE movies_greater_than_zero INTO '/user/hduser/movies_greater_than_three_point_five' USING PigStorage (',');

Display the processed file

cat /user/hduser/movies_greater_than_three_point_five