**EXPERIMENT-7**

**AIM:** Movie data Set Analysis using Pig, Database Operations in PIG Using CLOUDER

* **Copy the files form local system Download folder to root directory:**

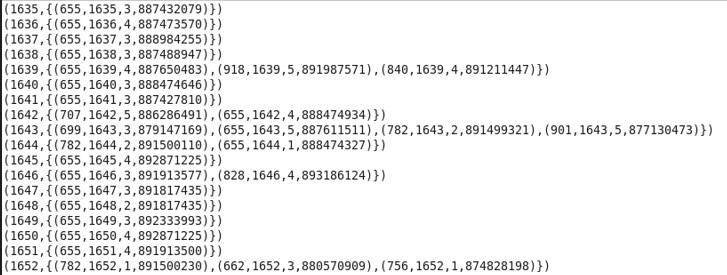
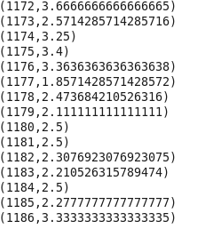
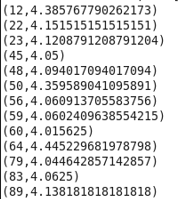




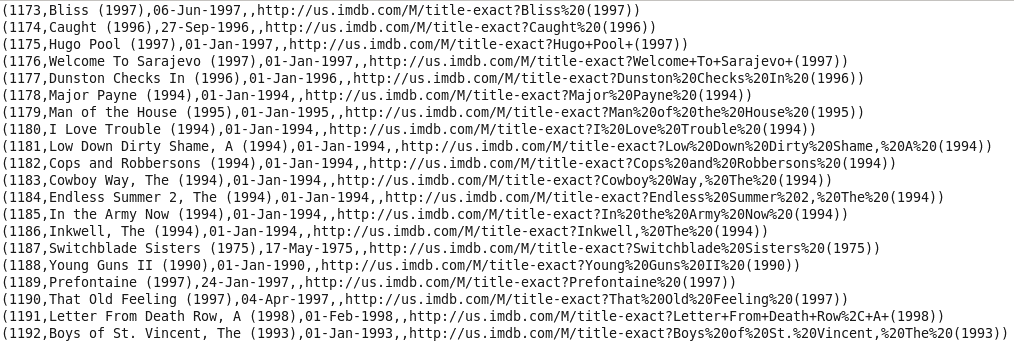
* **Load the movie data into pig**

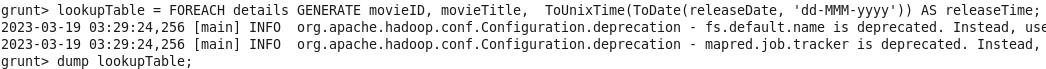
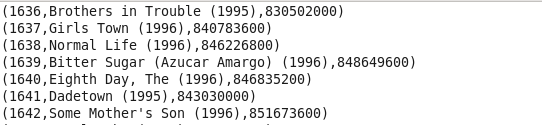
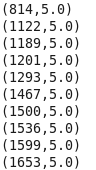
ratings = LOAD '/u.data' AS (userID:int, movieID:int, rating:int, ratingTime:int);

dump ratings;

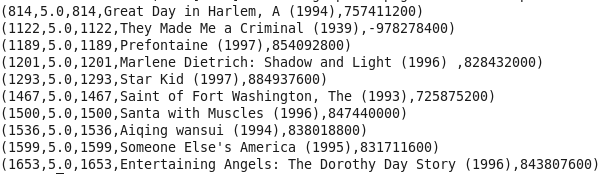
1. **Find the movie with avg rating >4.0 from u.data dataset.**
2. **Group the data according to movieID:**
3. **Find average ratings for the grouped data:**
4. **Filter the required results:**
5. **Find the oldest 5-star movies from u.data and u.item datasets.**
6. Load the u.item in pig:

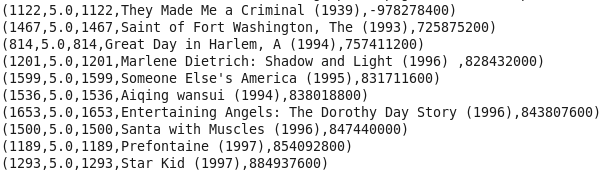
details = LOAD '/u.item' USING PigStorage('|') AS (movieID:int, movieTitle:chararray, releaseDate:chararray, videoRelease:chararray, imdbLink:chararray);

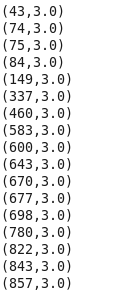
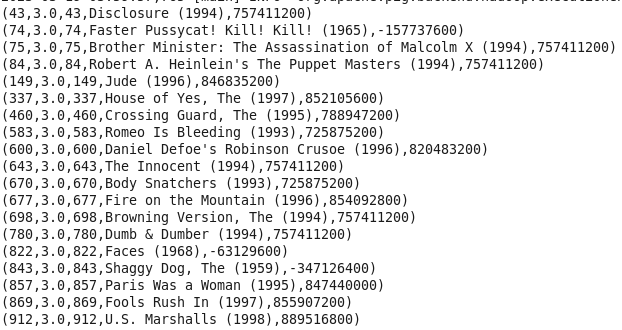
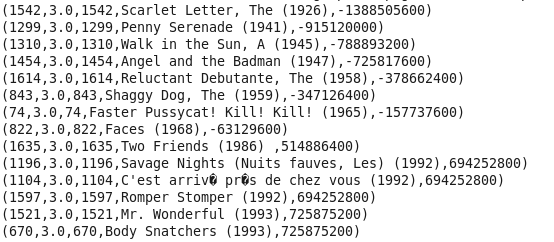
dump details;

1. Create a scalable timestamp column to compare the time:
2. Filter the movies with average rating as 5:

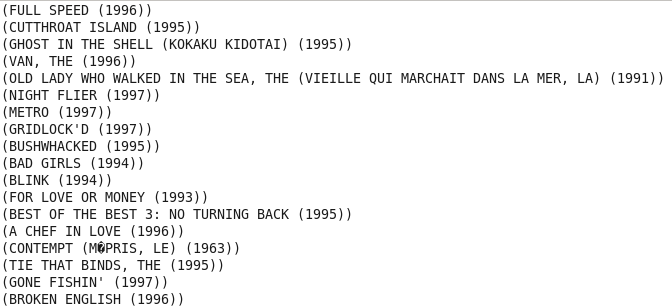


1. Join this table with the lookup table:
2. Order the results by time (year):



1. Find the oldest 3-star movies from u.data and u.item datasets.
2. Join this table with the lookup table found in part (b):
3. Order the results by time (year):
4. Display name of all movies in uppercase.

 If while loading the data we don’t use USING PigStorage(‘,) then while executing the following commands entire columns would be uppercase as it is reqd for a delimiter!



Submitted By:

Name: Soumya Ranjan Maharana

Regd. No.: 20010414

Branch: CSE (AI & ML)

Group: 4 (H)