**DETAILED DESCRIPTIONS OF DATA FILES**

**==============================================**

**u.data -- The full u data set, 100000 ratings by 943 users on 1682 items.**

**Each user has rated at least 20 movies. Users and items are**

**numbered consecutively from 1. The data is randomly**

**ordered. This is a tab separated list of**

**user id | item id | rating | timestamp.**

**The time stamps are unix seconds since 1/1/1970 UTC**

**u.info -- The number of users, items, and ratings in the u data set.**

**u.item -- Information about the items (movies); this is a tab separated**

**list of**

**movie id | movie title | release date | video release date |**

**IMDb URL | unknown | Action | Adventure | Animation |**

**Children's | Comedy | Crime | Documentary | Drama | Fantasy |**

**Film-Noir | Horror | Musical | Mystery | Romance | Sci-Fi |**

**Thriller | War | Western |**

**The last 19 fields are the genres, a 1 indicates the movie**

**is of that genre, a 0 indicates it is not; movies can be in**

**several genres at once.**

**The movie ids are the ones used in the u.data data set.**

**u.genre -- A list of the genres.**

**u.user -- Demographic information about the users; this is a tab**

**separated list of**

**user id | age | gender | occupation | zip code**

**The user ids are the ones used in the u.data data set.**

**u.occupation -- A list of the occupations.**

Problems:

##### [**1. Create an external table for u.data, item.data, u.genre, u.user, u.occupation, u.info so that it can be used efficiently for queries. Load data into external tables using load command.**](https://www.dezyre.com/hadoop-course/hive#collapse3-1)

2. Find the 5 star movies

3. Find all movies with an average rating less than 2.0

4. Sort movies by the total number of rating

5. Sort only **'Comedy'** movies order by rating (5,4,3,2,1) in DESC order