

# CODING PROBLEM

The problem mentioned below revolves around movies dataset. The dataset contains 4 files which are follows,

| File Name   | Description / Schema  |
|-------------|---|
| movies.dat  | MovieID – Title – Genres  |
| ratings.dat | UserID – MovieID – Rating – Timestamp                           |
| users.dat   | UserID – Gender – Age – Occupation – ZipCode                    |
| README      | Additonal information / explanation about the above three files |

The dataset can be downloaded from the link : <http://grouplens.org/datasets/movielens/1m/>

You are required to write a code in Hadoop Map Reduce to get results for the following questions:

1. Top ten most viewed movies with their movies Name (Ascending or Descending order)
2. Top twenty rated movies (Condition: The movie should be rated/viewed by at least 40 users)
3. We wish to know how have the genres ranked by Average Rating, for each profession and age group. The age groups to be considered are: 18-35, 36-50 and 50+.