Task 1

Please follows the instructions as following:

1. Create tables: movies, ratings, tags
2. Load the relevant data files to those tables
3. Prepare a script file to load movie_stats table.
   Following is the column detail with sample data for movie_stats table.
   movie_name,avg_rating,hash_tag,hash_tag_cnt
   Avengers,4.2,thriller,19
   Avatar,4.9,fiction,90
   Amelie,4.5,comedy,150

   Note: Choose appropriate column data type as per your knowledge.

Task 2

Prepare a SQL to update avg_rating for a given movie. Ask your peer to review it.

Task 3

Prepare a SQL to delete a record for a given movie. Ask your peer to review it.

==========

Task 4

We will receive one data file for each base table (movie, rating and tag). We need to enhance our ETL process to load these data files.

Following are the checks that you need to follow during load process to maintain data integrity.

- Movie ID is unique field. That means you have to check for any unique key violation
- Rating should be for a valid movie. That means movieId in rating data file should be present in Movie table
- There should not be any duplicate record in any table. In case of duplicate record, we will consider the latest one as the candidate to load and others to discard.

Task 5

We have some clients using MySQL database to generate reports. Hence we need to push our movie_stats table to that MySQL database.