INFO20003 Semester 1, 2021

Assignment 1: ER Modelling

Due: 3rd April 2021 10:00am

Submission: Via LMS https://canvas.lms.unimelb.edu.au/

Netflix

Netflix is an American content platform and production company headquartered in Los Gatos, California with approximately 200 million users.

Netflix not only hosts the entertainment content but also has entered in content-production industry by producing blockbusters such as House of Cards, Stranger Things, and The Crown. Netflix is a data-driven company relying heavily on data to make informed decisions about content creation and delivery.

Your team has been asked to design a MySQL database to store information about the viewing history of users, and the details of the content they have watched or plan to watch. With this data, Netflix can make better decisions and ultimately make users happier with their service. The following specifications have been provided to your team to assist in your design.

Subscription

Netflix offers a variety of plans to meet a user's needs. Users must first subscribe to a plan to create a Netflix user account, and gain access to the service. One user can purchase a subscription that can be shared with friends and family, each of whom will receive their own user account. This will result in multiple user accounts associated with the same subscription. Separate user accounts help Netflix to effectively divide the interest of the users: each will receive individualized suggestions for shows and have their own list of favourite shows (distinct from the other users sharing the same subscription). Each user has a nickname, and a profile avatar photo.

Subscriptions can be to three different plan types currently with different monthly rates for each of these: "basic", "standard" or "premium", but more plans might be added in future. Some plans can stream in HD, and others cannot. The plan a user chooses will also determine the number of devices that they can watch Netflix on at the same time. E.g., for "basic", "standard" and "premium" plan subscriptions, 1, 2 and 4 users can watch simultaneously, respectively (but again, these values might change in future). For each subscription, the number of users currently watching Netflix should be tracked, so that if more users try to simultaneously watch than are allowed an error can be shown.

For each subscription, Netflix keeps track of the purchaser's contact details such as their first name, last name, address, up to two phone numbers and a unique email address tied to the subscription. The date of purchase for a subscription should also be stored. You must store the country where the user who purchased the subscription is residing and will be using Netflix, as the Netflix library differs by country. Since it is a subscription, payment details also need to be stored for the recurring deduction. Netflix wants to record the account name (e.g., 'Farah Khan'), 16-digit credit card number (e.g., '0198 2345 3435 8822'), and the expiry month and year (e.g., 01, 2024). These details can be updated at any time, but the history of the changes does not need to be kept. The date at which the next payment is due and the date of payment of all past payments associated with that subscription and their amount need to be stored.

When a subscription is cancelled, we need to record the date that the subscription ended. User accounts associated with this subscription are not removed, and the subscription details are not removed from the database. When a subscription is cancelled the credit card details are deleted permanently, but the remaining information for the subscription stays. In this system, storing a user accounts' viewing history from past subscriptions is not necessary (see the 'user profiles' section for more on user viewing histories).

If a user later decides that they again want access to Netflix after their subscription was cancelled (using the same email address they used for the past subscription), a new subscription is created with a new purchase date and purchaser details, and the users from the old subscription are transferred to the new subscription. If a subscription is upgraded or downgraded, the existing subscription is effectively 'cancelled', and a new subscription is created in a similar manner.

Shows

Netflix hosts two types of Shows: Films and Series. Along with the title of the show, Netflix may also store few keywords such as *witty*, *humorous*, *inspiring*, *teen drama, Netflix original,* etc for each show. There is also a more standard maturity rating given to each show: Suitable for General Audiences (G), Parental Guidance suggested (PG), Recommended for mature Audiences (M), Not suitable for ages 14 and under (MA 15+), and Restricted to ages 18 and over (R 18+). These ratings and keywords help users make informed choices about the content to watch. The maturity rating level can also be set up for a user profile. Profiles with a set maturity level will only show titles that do not exceed the selected maturity rating level.

For films, the date of release, and name of the film studio need to be stored. For series, the creators name (assume only one), date of first episode's release and date of final episode's release (if the series is finished) need to be accessible. Series have one or more seasons, each of which have one or more episodes. A season may contain a name (which may be empty, in which case it is displayed

to users as "season X"). An episode must have the date of release stored, and a title (which may similarly be empty).

This database will not store the actual media objects, but rather a link to the media objects. Each film or episode is linked to one media object. For each media object, store the URL link, duration in minutes, and size in GB.

Netflix may decommission older shows. In such cases, the date that the show is available until is stored and displayed to the users.

User Profile Activity

Netflix maintains a list of films/episodes that have been watched by each user profile, the 'stopped time' for each of the viewings (i.e., how far through the show the user stopped watching, for instance '16 minutes 30 seconds in, the user stopped watching') and the date and time of viewing start. This allows the service to suggest finishing a film or episode they started and didn't finish as well as offer to resume playback from when they stopped. If an episode/film has been watched in its entirety, the stopped time corresponds to the episode/film's full duration. Note that viewing history is recorded on a per-episode basis for series.

Netflix has implemented a Thumbs Up/Down (i.e., Like/Dislike) Rating System for rating shows (note this applies to the show as a whole, not per episode). A user profile can also like or dislike a show without watching it. If a user changes a rating from thumbs up to thumbs down and vice versa, only the last rating is saved in the database. With this additional personal input, Netflix is able to improve personalization, making a user's homepage on Netflix even more relevant.

Top 10 shows

Netflix manually curates a list of top 10 shows and updates it every fortnight depending on the popularity of the show. A show can make it to top 10 more than once over time. We must be able to determine how many times a show has been in the top 10. A show can stay in the top 10 at the same or different rank. For example, Vikings can be number 1 from January 1st – January 14th, 2021 but the rank drops to number 3 from January 15th-January 31st, 2021. For each such rank, we must keep a record of start date when a show reached that position and end date when it was removed from the position (and potentially from the list altogether). Every time a show reaches one of the top 10 positions, this is marked as one occurrence on the list, thus the provided Vikings example would count as two appearances on the top 10 list.

Business Requirements

Your database design needs to be able to meet the business' needs to answer questions as:

- How many users have purchased the "Basic" subscription?

- How many users had active Netflix subscriptions in 2020?

- Which film had the most viewings in 2021 so far?

- Which series has the highest number of thumbs ups?

- Which series have been watched more than once in their entirety by "John Oliver"?

- How many users started watching "Forest Gump" but did not complete it?

- How many times was the series "Outlander" in the top-10 list?

- How many films are labelled with keyword "horror"?

- Has "Jenny Taylor" cancelled and renewed her subscription more than three times?

Instructions:

Assignment 1 is worth 10% of your final mark. The assignment will be graded out of 100 marks as described in the table below:

ER Physical Model with assumptions	80%
Conceptual Model in Chen's Notation	20%
.mwb Physical Model File submitted	Assignment Hurdle

- You are to analyse this business case and design a **Conceptual ER Model** in **Chen's notation** (can be hand drawn) as taught in class and a **Physical ER Model** for a MySQL Relational Database in **Crow's foot notation** (modelled with MySQL Workbench).

- You may list any assumptions you have made about the model on a separate page. There is a 200-word limit for assumptions. Assumptions must not be used to simplify the assignment, but only to justify your decision about any ambiguity in the study.

- As part of your submission you must submit your final MySQL Workbench .mwb file of your Physical ER Model.

    NOTE: Items 1 (Conceptual Model & Physical ER model) and 2 (Assumptions) must be submitted as **a single PDF**. The title of the PDF document must be **your student id** (e.g. 987654.pdf). Similarly, the title of your .mwb file should be your studentid.mwb (e.g. 987654.mwb).

Assignment Submission:

You are to submit the assignment under the Assignments tab on Canvas LMS. The submission will require you to **_submit two files_**:

1. ONE PDF document named as your student id (e.g. 987654.pdf) containing:
    i. a legible (hand drawn, or otherwise modelled) picture of a conceptual model in Chen's notation
    ii. a screenshot/export of your Physical ER Model done in MySQL workbench. *Ensure that all attributes are readable, and tables are fully expanded.*
    iii. any assumptions you made (limit this to a maximum of 200 words)
2. A copy of your final .mwb MySQL Workbench file of the Physical ER model.
   Note: this model will only be used by staff in circumstances where the screenshot/export of your model in the PDF is unreadable (i.e. if we have to open this file, a penalty of 10% will occur), so remember to include the export of your physical model in the PDF!

Note: If you fail to submit clear and legible models your assignment will be penalised

Unless you have an approved extension (see below), you will be penalised -10% of the total number of marks in the assignment per day that your submission is late. For instance, if you received a 78% raw score, but submitted 2 days late, you'd receive a 58% for the assignment.

Requesting a Submission Deadline Extension

If you need an extension due to a valid (medical) reason, you will need to provide evidence to support your request by 5pm, 2nd of April. Medical certificates need to be at least two days in length.

To request an extension:

* Email Farah Zaib Khan (farah.khan@unimelb.edu.au) from your university email address, supplying your student ID, the extension request and supporting evidence.
* If your submission deadline extension is granted you will receive an email reply granting the new submission date. Do not lose this email!

Reminder: INFO20003 Hurdle Requirements

To pass INFO20003, you must pass two hurdles:

* **Hurdle 1:** Obtain at least 50% (15/30) or higher for the three assignments (each worth 10%)
* **Hurdle 2:** Obtain at least 50% (35/70) or higher for the combination of quizzes and end of semester exam

Therefore, it is our recommendation to students that you attempt every assignment and every question in the exam.

GOOD LUCK!