Relational Database for Streaming Service

Problem Statement

Streaming services have taken over as the chief medium for consuming media content. Many forms of digital content, including motion pictures, are now being streamed through the internet in overwhelming amounts. There are several major studios today that prefer to release their movies directly on these streaming platforms instead of putting them in theatres.

Most successful streaming services take a data-driven approach to doing business. They collect vast amounts of data from the users and use it to serve targeted advertisements, deliver personalized recommendations, and understand user behavior. Because of the copious amounts of data collected, it is imperative that they have a robust DBMS at the backend to store the data in an organized fashion. Such a database would be necessary to manage the movie catalogue, subscriptions, and transactions made by the users, as well as to store user data that can be analyzed to understand viewing patterns, optimize content delivery, and deliver targeted advertisements.

Here, we propose a DBMS model to manage some of the data requirements posed by a movie streaming platform.

Data Requirements

1) Movies: The catalogue of movies currently available on the platform.
   - Movie ID: A unique ID for each movie
   - Title
   - Genre ('Action', 'Adventure', 'Comedy', 'Drama', 'Fantasy', 'Horror', 'Mystery', 'Romance', 'Sci-Fi', or 'Thriller')
   - Metadata: runtime (number of minutes), release year (min. 1888), Censor's rating (e.g. PG-13)
   - Available audio languages, available subtitle languages
   - Actors' information:
     i) Name
     ii) Date of Birth
     iii) Nationality
     iv) Actor ID: A unique ID for each actor
   - Critics' score (IMDB)
   - Total time (hours) streamed (sum of all watch session durations for the movie)

2) Accounts: Each household has one account; an account may have multiple profiles for different members of that household.
   - Account ID: A unique ID for each account
   - Account holder's details: Name, Date of Birth, Email, Phone Number
   - Password
   - Devices: Devices linked to the account (the number of devices is limited by the tier)
   - Payment history: Every payment the account has made

- Subscription information:
  i) Current tier (or no active subscription)
  ii) Date of last subscription
  iii) Subscription expiry date (one month after last subscription)
  iv) Auto renew (on/off): Decides whether the subscription automatically renews the following month.
- Profiles: List of Profiles under the account
- Total watch time: sum of total watch times of all profiles under the account.

3) User Profiles: Each individual member of a household can create a profile (under the household's account) which will be personalized for that individual.
   - Profile ID: A unique ID for each profile
   - Account under which the Profile was created
   - Profile Name
   - Date of Birth
   - Gender
   - Age group: Child, Teen, or Adult (based on Date of Birth)
   - Watch sessions: all watch sessions made by the user.
   - Total watch time (sum of all watch session durations for the profile)

4) Subscription Tiers: Tiers may be added, removed, or modified according to the needs of the business. All subscriptions are monthly.
   - Tier ID: A unique ID for each tier
   - Tier Name (Must also be unique)
   - Tier Price and currency
   - Tier Features:
     i) Number of devices allowed (per account)
     ii) Maximum Video Streaming Quality (e.g. 1080p, 4K, 4K+HDR, etc.)
     iii) Maximum Audio Streaming Quality (e.g. Stereo, Dolby Digital, Dolby Atmos, etc.)
     iv) Ad Frequency (e.g. No ads, Occasional, Frequent, etc.)
   - Accounts subscribed to that tier

5) Payments: Every payment made by an account, manually or automatically (auto-renewal), must be recorded. For a given account, the timestamp of each payment will be unique.
   - Payment ID: A unique ID for each transaction
   - Account that made the payment
   - Payment timestamp
   - Subscription Tier for which the payment was made
   - Payment Type: First-Time Subscription, Reactivated Subscription, Auto-renewal, Tier Upgrade, Tier downgrade
   - Payment method used (Credit Card, E-check, PayPal, Apple Pay, Venmo, or Google Pay)
   - Base amount = Subscription Tier Price
   - Tax rate
   - Discount code and rate (if applicable)
   - Total amount = Base amount - Discount Rate * Base Amount + Tax Rate * Base Amount
   - Currency Used

6) Watch Sessions: Each watch session made by a user profile on each device must be recorded.
   - Unique Session ID
   - Profile that initiates the session
   - Movie watched during the session
   - Session Start Date and Time
   - Session End Date and Time
   - Session Duration = Session End Time – Session Start Time
   - Device on which the movie was watched
   - IP address the movie was watched from

7) Devices
   - Unique Device ID
   - Device Type: TV, PC, Mobile or Console
   - Device Model (e.g. SamsungOLED-18, PlayStation5, etc.)
   - Device OS
   - User-given device name (e.g. “Bedroom TV”)
   - Device linked date: The date on which the device was linked to the account.
   - Screen Resolution
   - HDR Support?
   - Watch sessions made on the device
   - The account which the device is linked to (only one account can be linked to a given device)

<u>Additional Constraints</u>

- An account can have multiple linked devices (minimum one and maximum decided by the subscription tier), but a given device can be linked to one account only.
- There can be multiple user profiles under an account (at least one by default), but each profile belongs to exactly one account.
- Many actors can star in many movies. There must be at least one actor in each movie, and each actor must be in at least one movie.
- An account can be subscribed to only one subscription tier at a time, but many accounts can subscribe to the same tier. An account may not yet be subscribed to any tier, and a tier may not have any subscriptions yet.
- An account may have made multiple payments, but each payment is linked to exactly one account. An account may not have yet made any payments.
- Each payment is linked to a specific subscription tier, and multiple payments can be made for the same tier. A subscription tier may not have any received any payments towards it.
- Only one movie can be watched during a watch session, and each watch session is conducted on a single device and by a single user profile.
- A movie may never have been watched, and a user profile or device may never have had a watch session.

Business Goals

The database should collect data with the following goals:

1) Generate a report identifying the most popular movies (top 5, ranked by total watch session duration) watched by each demographic group (young boys, young girls, teen boys, teen girls, adult men, and adult women). This data can be used for targeted advertising by displaying ads between these movies for viewers of that age and gender group (as required by the advertiser) during the next quarter.

2) Identify the most popular movies (top 5, ranked by total watch session duration) watched in each region/area based on Watch Session IP address. This data can be used for targeted advertising by displaying ads between these movies for viewers in the respective regions (as required by the advertiser) during the next quarter.

3) Determine which genre receives the most watch time (total watch session duration) among each demographic group (young boys, young girls, teen boys, teen girls, adult men, and adult women).

4) Determine which actor/actress is popular with each demographic group (young boys, young girls, teen boys, teen girls, adult men, and adult women).

5) Did movies with high critic scores get more viewings? (Analyze the relationship between critic scores and total streaming hours for each movie.)

6) Determine the average total watch time (over the entire quarter) for a user of each age group.

7) Which type of device (TV, PC, Mobile, or Console) did the most traffic (highest total watch session duration) come from?

8) Which subscription tier was the most preferred (has the most number of accounts currently subscribed to it)?

9) How many accounts purchased first-time subscriptions, how many reactivated their subscriptions, how many auto-renewed, how many upgraded, and how many downgraded? How many accounts turned off auto-renew?

10) Which mode of payment is the most preferred (most payments made)?

11) For each account, generate a report of all IP addresses across all watch sessions used by each device authorized by the account. Analysis of this report will help to identify instances of password sharing between households.