421 Group Project Deliverable 1

Group 46

Petar Basta	260735072
Deneille Guiseppi	260722102
Luke Ma	260745824

1. Requirement Analysis

Introduction

Purpose

This application will serve as the backend for a video streaming service, of the likes of Hulu or Netflix. It describes users, the accounts they use, as well as their payments and the types of media that they can watch. Video streaming is a massive industry, projected to be worth over 130 billion by 2025. Such a database lets us explore the structure of the service and identify any shortcomings or weaknesses in the model. This model can thus be utilized as an efficient tool to slash costs and to streamline performance.

Scope and Requirements

We've limited the depth of our model to simply identify users and their actions, such as ratings, media type, and payments. We included the region in order to tailor the selection offered to a given viewer, as well as tags like actors, director or genre, which are common filters used when searching for media. This ignores any special cases of video streaming such as live shows, interactive choose-your-own-adventures, podcasts or any other hyper specific content. We also do not include any metadata collected across bigger swathes of the population, such as user preferences when compared with others, generally popular movies or viewing time by users (retention rate).

Database Description

Entities and their attributes

The following entities will be stored in tables of a relational database:

Account: An account is a digital record that is used by a person(s) to access the streaming service. An account is associated with a single unique attribute called ID.

Payment: A payment refers to the amount paid to the streaming service provider to access/create an account. It has multiple attributes, including data, amount and is identified by a primary key, ID.

User: A user is a weak entity that stores a user's data. The user(s) of an account is defined by the user's name and the foreign key of its account, ID.

Geographic Region: A geographic region is the country/region where the IP address of the user is located. The streaming service is located in multiple countries throughout the world, each of which has a unique primary key (name) for reference.

Media: A media is a Movie or TV Show. The media attributes are the title of the media and a unique primary key, ID.

Tag: This is a marker for users to look for as they search for media. This might denote the actors starring in the film, the director, the genre, or any other identifiers that the user might associate to the video in their minds.

Rating: This is an optional action that users can utilize as they watch videos, which helps us create a preference profile and tailor their viewing experience in the future

Movie: A movie is a subcategory of the media entity. Its attribute includes a release year attribute which makes note of the movie release date, as well as the other attributes inherited from the more general media entity.

TV Show: A TV Show is another sub-category of the media entity. It inherits all the attributes from media and has its own attribute, CompletionStatus which tracks the user's completion of a TV Show.

Season: A season is a weak entity that stores/keeps track of the tv show according to its seasons. It is defined by a unique number which is the season number a user has reached

in the tv show as well as a foreign ID key of the TV Show. A season also has a release year attribute.

Episode: An episode is another weak entity that keeps track of the season of a TV Show by the episode number. The episode is identified by its number, the foreign season number and the foreign key, ID from the media entity. It also includes a title attribute.

Relationships

pays: A *payment* is paid by an *account*. In this relationship, each payment is paid by exactly one account, and an account can make multiple payments.

admins: An *account* has an administrator who is a *user*. This relationship has a participation constraint where each account has exactly one user administrator, and a user can be administrator of an account.

uses: An *user* uses an *account*. In this relationship, each user uses exactly one account whereas an account can be used by multiple users.

is in: A *user* is in a *geographic region*. This relationship has a participation constraint from user to geographic region. This means that each user is in exactly one geographic region, and a geographic region can have multiple users in it. An account can have users in different geographic regions and has no constraints.

queues: A *user* queues up a *media* programme. This is a many-to-many relationship since a user can queue up multiple media, and a media programme can be queued up by multiple users. A user can then give the media a rating from 1 to 5.

available in: A *media programme* is available in a *geographic region*. This relationship has a many to many constraint such that a geographic region can have multiple media programmes available in it, and where a media programme must be available in at least one geographic region.

of type: A *media* programme is of a type *tag*. This relationship has a participation constraint such that a tag can describe multiple media programmes, and each media programme must have at least one descriptor tag.

2. Relations

Entities:

Payment(pid, date, amount, accid) (accid ref Account)

Account(accid, admin) (name ref User)

Region(regName)

Tag(tid, name)

Media(mid, title, releaseY, completionStat)

Weak Entities:

User(name, accid, regName) (accid ref Account, regName ref Region)

Episode(epNum, seasonNum, mid, epTitle) (seasonNum ref Season, mid ref Media)

Season(seasonNum, mid) (mid ref Media)

Rating(mid, name, accid, rid, value) (mid ref Media, name ref User, accid ref User[ref Account])

Relationships:

available_in(mid, regName) (mid ref Media, regName ref Region)

queues(name, accid, mid) (name ref User, accid ref User[ref account], mid ref Media, name/accid/mid ref Rating)

of_type(tid, mid) (tid ref tag, mid ref Media)

NOTE: Media is modeled as one large table where Movies will have NULL completionStatus and TV Shows will have NULL releaseYear. Modeling TV Show and Movie as their own table would cause problems with the relationship between Region-Media and Tag-Media since they would reference the “mid” of the Media table which wouldn’t exist, or would have to have a relationship with both TV Show and Movie.

NOTE: Since an Account can have several Users who all use the account, of which one must be the admin, there is a kind of circular reference between User and Account.

3. Discussion:

There are no remaining ways to combine any of the entities we have. For example, if we were to combine User and Region, just having region as an attribute of a User, then it would disturb the relationship between Region and Media since a Media must have a list of Regions where it is available. Another example would be Users and Account, which we could not combine since each User of an Account would no longer be able to have their own queue of Media. On the other hand, we had originally considered cast and genre as separate entities, but it made more sense to combine them under Tag, since that is how they are valued by the end user. We could have also implemented a viewed or a favorites data field, but they could also be included under Tag, since there is not much difference in the long run.