PRACTICAL 5

COS 221

Lwando Msindo 21507067 Matthew Gravette 23545977 Onthatile Moleboloa 22659812 Retshepile Nkwana 23580578 Bob Dlamini 21339120 Bridget Nkosi 22664638

Table of Contents

Task 1: Research	2
Task 2 – EER Diagram	4
Task 3 – Relational Mapping	7
Task 4 – Relational Schema	18
Task 5 – Web-Based Application	23
Task 6 – Data	25
Task 7 – Ontimization	28

Task 1: Research

Overview and explanation

The entertainment industry is the set of entities that generate a piece of work which involves the creation, production, promotion, or distribution for an audience - including theatre productions, motion pictures, radio broadcasts, television broadcasts and more. One of the many sub-categories of the entertainment industry is the Film and Television Industry. It establishes the process of generating films and television series. The medium of the Film and Television industry has transformed numerous times over several decades, with the most recent medium being streaming and online platforms.

Genres

Action depicts the Hero's Journey, involving high-stakes situations. Comedy adds humour to life's discontentments, aiming to make audiences laugh through exaggerated language, action, or narratives. Drama presents stories with conflicts to evoke powerful emotions, often reflecting societal issues. Documentary filmmaking depicts reality authentically to educate or record historical moments. Fantasy transcends natural laws, depicting imaginary worlds with magic and mythological elements. Horror aims to instil fear or dread, using fear as a narrative tool. Science Fiction blends science with imagination, expanding on scientific observations to create its world. Romance explores intimate relationships and their conflicts, focusing on how characters overcome challenges together.

Content rating and categorising

The content produced by the Film and Television industry may be distributed into three categories: Unrestricted, Age-restrictive, and Restrictive. They are rated based on topics covered in a piece of work that an audience may need to be informed about. Unrestricted has 2 rating groups - A (all ages) and PG (Parental Guidance). Age-restrictive has 5 rating groups - 7-9PG, 10-12PG, 13PG, 16 and 18. Restrictive has 2 rating groups - X18 and XX. Furthermore, ratings may be accompanied by content warnings. These include PPS (Photo Pattern Sensitivity), P (Prejudice), S (Sexually Related Activity), V (Violence), SV (Sexual Violence), CT (Criminal Techniques), D (Substance Abuse), IAT (Imitative Acts and Techniques, H (Horror), L (Language), N (Nudity). These categories follow the rating system used by the Film and Publication Board (Fpb). Alternative rating systems by different entities or countries exist.

Additional features/information provided

Recent research indicates a significant shift in film genre preferences: while action movies are gaining popularity, adventure films are declining. Comedy and drama are becoming strong competitors to action, and interest in documentaries is rising. Additionally, there's growing demand for diverse, original content, particularly short films. Key factors for choosing a streaming service include the efficiency of its recommendation algorithm and the quality of its media offerings. Effective recommendation algorithms should integrate with users' social media profiles and adapt to their moods, ensuring relevant suggestions. To enhance user satisfaction, streaming services must offer an intuitive interface, including media reviews and user engagement features. They should provide a diverse range of content, including new and exclusive offerings, to keep users interested. Affordability is crucial, with services offering price discounts, free trials, and minimal, relevant ads. Understanding audience demographics, such as age, gender, and location, helps cater to specific user preferences. Successful streaming services are also accessible across various devices and regions, allowing users to enjoy content anytime, anywhere.

References

Law Cornell. "Entertainment Industry Entity." Definition: entertainment industry, 2019, https://www.law.cornell.edu/definitions/uscode.php?width=840&height=800&iframe=true&def_id=50-USC-798752819-

341428507&term_occur=999&term_src=title:50:chapter:45:subchapter:II:section:3332%20https://www.premiumbeat.com/blog/guide-to-basic-film-genres/. Accessed 07 May 2024.

Indeed Editorial Team. "15 Types of Film Styles and Genres (And How They Differ)." Indeed, 31 July 2023, https://www.indeed.com/career-advice/career-development/types-of-film-styles.

"Ratings | FPB." Film & Publications Board, https://www.fpb.org.za/ratings/. Accessed 7 May 2024.

https://www.the-numbers.com/market/genres

The Numbers (2019). The Numbers - Leading Genres 1995-2019. [online] The-numbers.com. Available at: https://www.the-numbers.com/market/genres.

https://www.statista.com/forecasts/1371221/preferred-tv-content-by-genre-in-south-africa

Statista. (n.d.). Preferred TV content by genre in South Africa 2023. [online] Available at: https://www.statista.com/forecasts/1371221/preferred-tv-content-by-genre-in-south-africa.

https://explodingtopics.com/blog/streaming-trends

Byers, K. (2020). 11 Huge Streaming Trends In 2021. [online] Exploding Topics. Available at: https://explodingtopics.com/blog/streaming-trends.

PWC (2021). A period of videostreaming peace: Consumers have found their video consumption groove: PwC. [online] PwC. Available at:

https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/consumer-video-streaming-behavior.html.

www.small-screen.co.uk. (2021). 8 Reasons To Still Choose Netflix Over Other Streaming Services. [online] Available at: https://www.small-screen.co.uk/reasons-to-still-choose-netflix-over-other-streaming-services/#google_vignette [Accessed 8 May 2024].

Beaty, A. (2020). *How Are Streaming Services Attracting and Maintaining Subscribers?* [online] Cord Cutters News. Available at: https://cordcuttersnews.com/how-are-streaming-services-attracting-and-maintaining-subscribers/ [Accessed 8 May 2024].

FlixPatrol. (n.d.). *Preferences on Netflix by genre in Q1 2022 • FlixPatrol*. [online] Available at: https://flixpatrol.com/preferences/netflix/genre/[Accessed 8 May 2024]. FlixPatrol. (n.d.). *Preferences on HBO by genre in Q2 2024 • FlixPatrol*. [online] Available at: https://flixpatrol.com/preferences/hbo/genre/ [Accessed 8 May 2024].

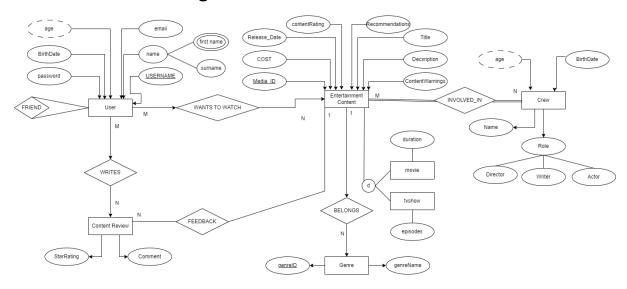
FlixPatrol. (n.d.). *Preferences on Amazon by genre in Q2 2024 • FlixPatrol*. [online] Available at: https://flixpatrol.com/preferences/amazon/genre/ [Accessed 8 May 2024].

FlixPatrol. (n.d.). *Preferences on Disney+ by genre in Q2 2024 • FlixPatrol*. [online] Available at: https://flixpatrol.com/preferences/disney/genre/ [Accessed 8 May 2024].

The Streaming Audience Analysis Metrics That Matter | Similarweb

Task 2 – EER Diagram

Version 1 - Initial Design



Key Features:

Entities:

- User → Store key information about the users of the website like their username and password
- Entertainment_content →v Information about the media stored including the mediaID according to the imdb ID and the title.
- Crew → Name and role of each person responsible for working on movie. Roles include actor, writer and director.
- Genre → As we assumed that there would a multivariate variable in entertainment_content for the genre we pre-emptively created this entity to store the genres of a certain media

Relationships:

- M:N → WRITES, INVOLVED_IN, WANTS_TO_WATCH, FRIEND, RECCOMENDED
- 1:N → FEEDBACK, BELONGS
- 1:1 → None

Assumptions:

- Recommendations are based off media and not user preferences
- Reviews will work on a 1 5-star rating

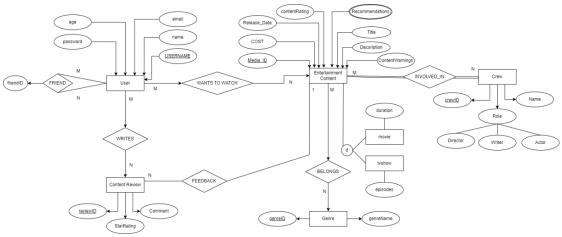
Complex Attributes:

• Name is the only complex attribute with it being made up of first names and a surname, the idea is that a user can have multiple first names and one surname.

Derived:

• Age in the users and crew based of the birthdate

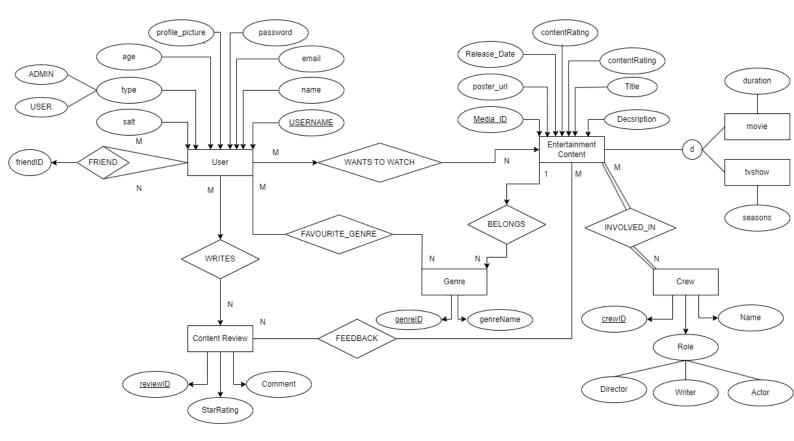
Version 2 – Revised Version



Changes:

- Added a friendID attribute to the friends relationship.
- Changed the relationship between genre and entertainment_content (BELONGS) to a many to many
- Removed the recommendations relationship in favour of a multivalued attribute
- Removed birthdate and derived age from crew since the information does not add to the website or user experiance
- Removed birthdate from user and only have age. Storing birthdate is only useful if the website had a celebrate birthday and the website does not have this feature
- The names were simplified to a simple attribute as the website does not make use of names.

Version 3 – Final Version



Changes:

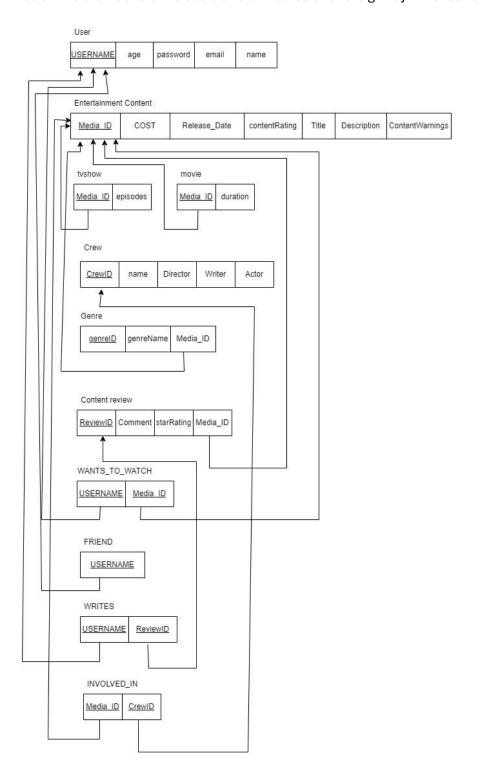
- Added attributes to user
 - o type to demark if a user is an admin or not
 - o profile-picture to store the users image, mostly likely from a predefined list
 - o salt to increase security for pass.
- Replaced the idea of recommendations in entertainment_content with a new many to many relationship (favourite_genre) between the user and genre. So now recommendation is not based on media but instead on the user's favourite genre.
- Dropped content_warning and cost due to data restrictions and overall simplification of the database.

Task 3 – Relational Mapping

First Draft

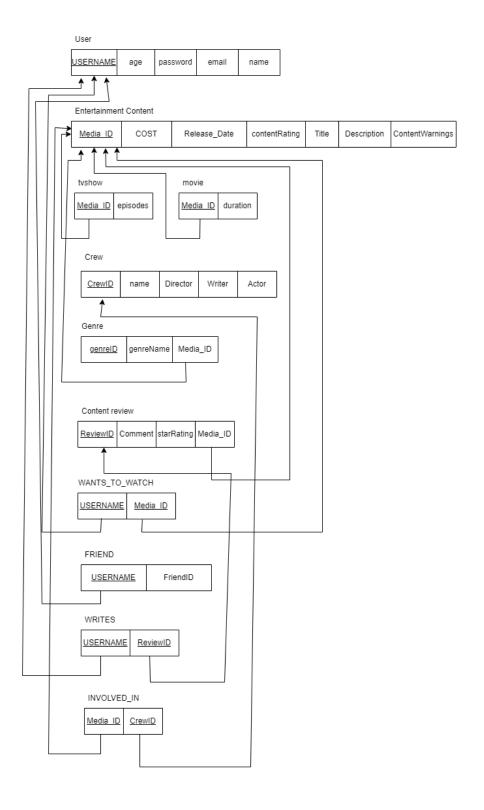
Our first draft was made after the third and before the final ER Diagram had been finalised, and so we

did not focus on step-by-step breakdowns until we had finalised the ER in case of changes made. This draft covers the basic and attributes and foreign key links found within our database.



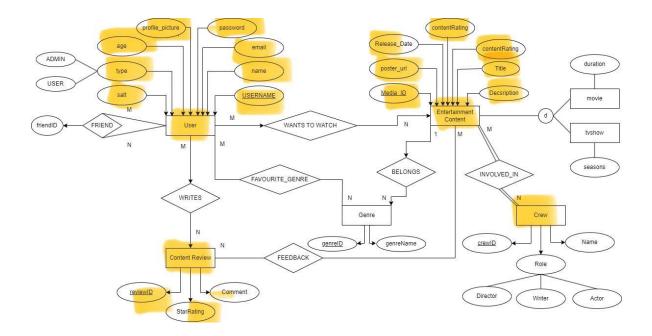
We decided to simplify the crew and so removed any excess information

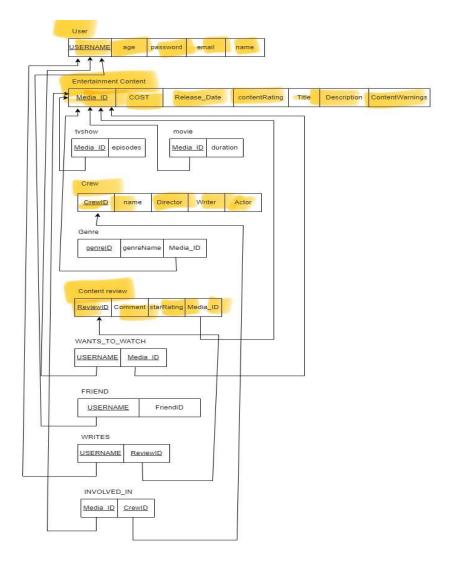
from the mapping. We decided that the only entertainment content that is important would be Country for statistical purposes whilst the nearby locations would be determined through the Winery city and the search function. We decided to remove UserID and make the email the username. We ensured that the Specializations linked correctly to the User generalization. Missing attributes were added.



STEP 1: REGULAR ENTITY TYPES

Mapping of regular (strong) entity types, Include only the simple component attributes
of composite attributes.





STEP 2: WEAK ENTITY TYPES

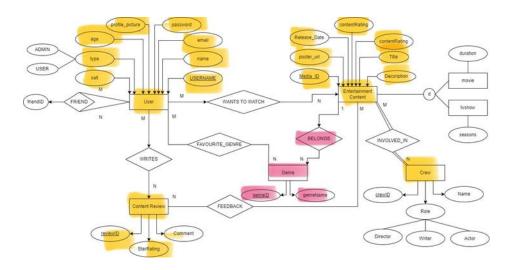
There are no weak entities

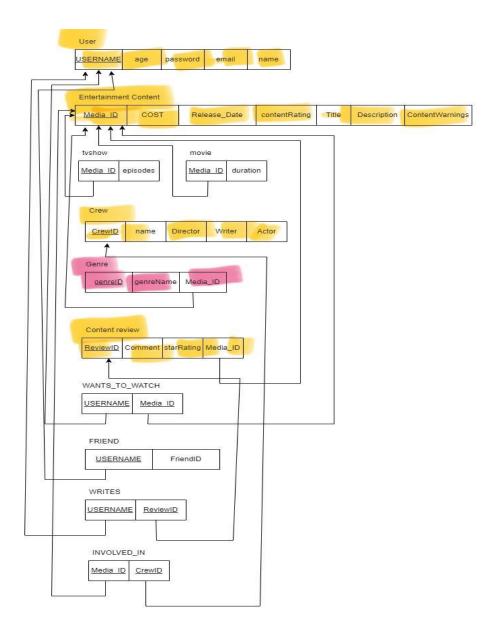
STEP 3: 1:1 RELATIONSHIPS

• There are no one to one relationships

STEP 4: 1:N RELATIONSHIP

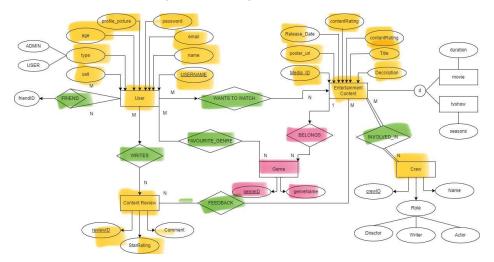
• All 1:N relationships were mapped. The Genre relationship was therefore added as a table due to its attributes. As identified, there exists only one 1:N relationships.

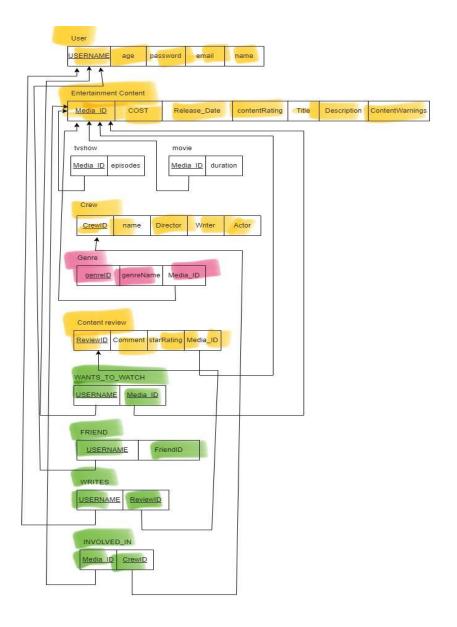




STEP 5: M:N RELATIONSHIPS

- There exists only six M:N relationship, existing between the following:
 - User and Friend
 - o Entertainment Content and User
 - Entertainment Content and Crew
 - o Entertainment Content and Content Review
 - o Content Review and User
 - o Genre and User
- This exists through connecting the primary keys.





STEP 6: MULTIVALUED ATTRIBUTES

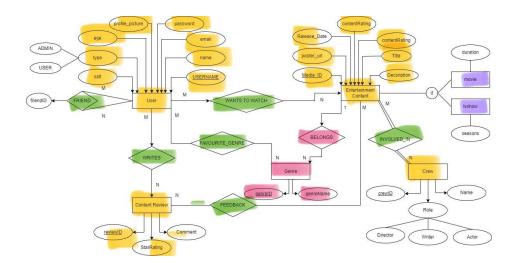
No multivalued attributes

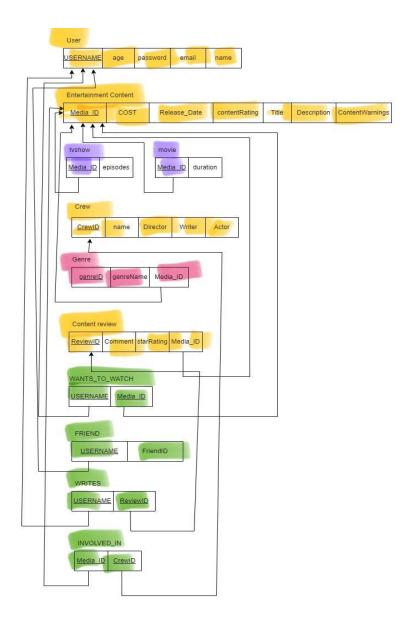
STEP 7: N-ARY RELATIONSHIP

• No n-ary relationships

STEP 8: MAPPING SPECIALISATION AND GENERALISATION

• The disjointness of the Entertainment Content as Tv Show and Movies was mapped in Relational schema. The appropriate keys were used to connect them to any necessary entities/tables.(Some arrowing was missed but was corrected in the final map coloring)



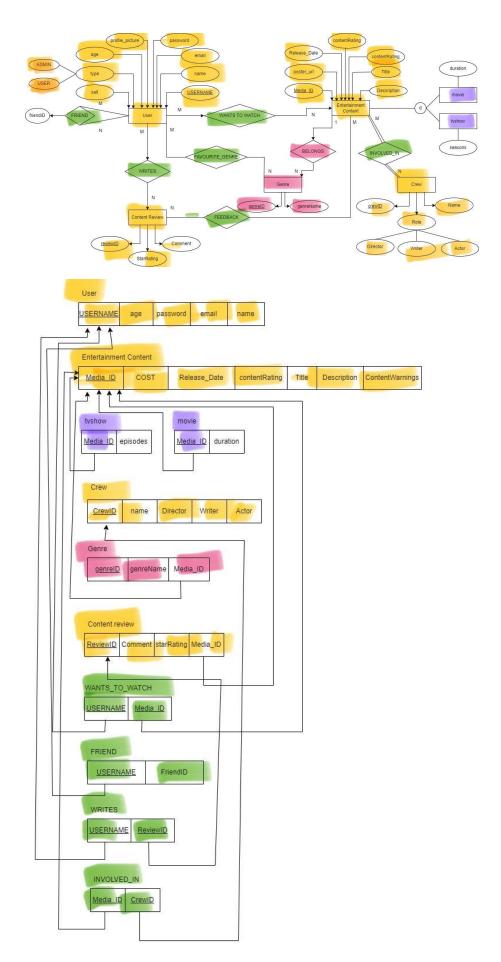


STEP 9: UNIONS

• No unions

FINAL (E)ER AND MAPPING WITH COLOURING

• This is the final colouring with all the relevant steps colour-coded. This can be properly viewed above with the step description.



Assumptions:

Unique Identifiers:

USERNAME is unique to each user.

Media_ID uniquely identifies each piece of entertainment content.

CrewID uniquely identifies each crew member.

genreID uniquely identifies each genre.

ReviewID uniquely identifies each review.

Relationships:

A user can have multiple friends.

A user can write multiple reviews.

Each review is linked to a specific piece of content.

A piece of content can belong to multiple genres.

A piece of content can have multiple reviews.

Crew members can be involved in multiple pieces of content.

Attributes:

COST is likely the cost to watch or produce the content.

Release_Date indicates when the content was released.

contentRating could indicate the age rating (e.g., PG, R).

Title, Description, and ContentWarnings are descriptive fields for the content.

episodes and duration specify TV show and movie details respectively.

Possible Issues or Considerations:

Normalization:

Ensure that the database is normalized to reduce redundancy.

For example, if ContentWarnings is repeated often, consider creating a separate table for it.

Constraints and Validations:

Ensure appropriate constraints are in place for attributes (e.g., valid email format for email). Scalability:

Consider how the system will scale with a large number of users, reviews, and content.

Privacy and Security:

Protect sensitive information like password.

Ensure user data is handled in compliance with data protection regulations.

This analysis provides a foundational understanding of the EER diagram and the relationships between the entities involved. Further details or refinements might be needed based on specific application requirements or additional context.

Task 4 – Relational Schema

```
genre
```

```
CREATE TABLE genre(
    genreID int NOT NULL AUTO_INCREMENT,
    genreName varchar(255) NOT NULL ,
    PRIMARY KEY (genreID),
    UNIQUE (genreName)
);
```

Keys: genreID(PRIMARY)

Constraints: All fields are NOT NULL

Checks: None

entertainment_content

```
CREATE TABLE entertainment_content(
    media_ID int NOT NULL AUTO_INCREMENT,
    title varchar(255) NOT NULL,
    release_Date date NOT NULL,
    description text NOT NULL,
    content_rating enum("PG","PG 9","PG 13","16","18") default "PG",
    PRIMARY KEY (media_ID)
);

Keys: media_ID (PRIMARY)
```

Constraints: All fields are NOT NULL

Checks: None

movie

```
CREATE TABLE movie(
    media_ID int NOT NULL,
    duration int NOT NULL,
    FOREIGN KEY(media_ID) REFERENCES entertainment_content(media_ID) ON DELETE CASCADE,
    CHECK (duration > 0)
);
```

Keys: media_ID (FOREIGN)

Constraints: All fields are NOT NULL

Checks: duration must be positive

tv_show

```
CREATE TABLE tvShow(
    media_ID int NOT NULL,
    seasons int NOT NULL,
    FOREIGN KEY(media_ID) REFERENCES entertainment_content(media_ID) ON DELETE CASCADE,
    CHECK (seasons > 0)
);
```

Keys: media_ID (FOREIGN)

Constraints: All fields are NOT NULL

Checks: seasons must be positive

belongs

```
CREATE TABLE BELONGS(
    genreID int NOT NULL,
    media_ID int NOT NULL,
    FOREIGN KEY(media_ID) REFERENCES entertainment_content(media_ID) ON DELETE CASCADE,
    FOREIGN KEY(genreID) REFERENCES genre(genreID)
);

Keys: media_ID (FOREIGN), genreID (FOREIGN)

Constraints: All fields are NOT NULL
```

favourite_genre

Checks: None

```
CREATE TABLE favourite_genre(
    username varchar(50) NOT NULL,
    genreID int NOT NULL,
    FOREIGN KEY(username) REFERENCES user(username) ON DELETE CASCADE,
    FOREIGN KEY(genreID) REFERENCES genre(genreID)
);
Keys: username (FOREIGN), genreID (FOREIGN)
Constraints: All fields are NOT NULL
```

crew

Checks: None

```
CREATE TABLE crew(
    crewID int NOT NULL AUTO_INCREMENT,
    name varchar(255),
    role enum("actor","director","writer"),
    PRIMARY KEY (crewID)
);
Keys: crewID (PRIMARY)
```

Checks: None

Constraints: crewID are NOT NULL

involved_in

```
CREATE TABLE INVOLVED_IN(
          crewID int NOT NULL,
          media_ID int NOT NULL,
          FOREIGN KEY(media_ID) REFERENCES entertainment_content(media_ID) ON DELETE CASCADE,
          FOREIGN KEY(crewID) REFERENCES crew(crewID) ON DELETE CASCADE
);
```

Keys: crewID (FOREIGN), media_ID(FOREIGN)

Constraints: All field are NOT NULL

Checks: None

friend

```
CREATE TABLE friend (
    username VARCHAR(50),
    friendID INT,
    PRIMARY KEY (username, friendID),
    FOREIGN KEY (username) REFERENCES users(username)
);
```

Keys: username (PRIMARY)

Constraints: username is NOT NULL

Checks: None

writes

```
CREATE TABLE writes (
    reviewID INT,
    username VARCHAR(50),
    PRIMARY KEY (reviewID, username),
    FOREIGN KEY (reviewID) REFERENCES content_review(reviewID),
    FOREIGN KEY (username) REFERENCES users(username)
);
```

Keys: reviewID(FOREIGN), username(FOREIGN)

Partial key: username, reviewID

Checks: None

content_review

```
CREATE TABLE content_review (
    reviewID INT PRIMARY KEY AUTO_INCREMENT,
    comments TEXT,
    starRating INT,
    mediaID INT,
    created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
);
```

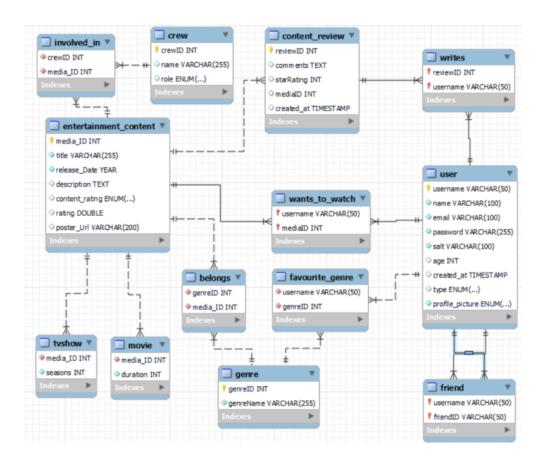
Keys: reviewID(PRIMARY)

wants_to_watch

```
CREATE TABLE wants_to_watch (
    username VARCHAR(50),
    mediaID INT,
    PRIMARY KEY (username, mediaID),
    FOREIGN KEY (username) REFERENCES users(username),
    FOREIGN KEY (mediaID) REFERENCES media(mediaID)
);
```

Keys: mediaID(FOREIGN), username(FOREIGN)

Partial key: username, mediaID



Task 5 – Web-Based Application

Functional Requirements

1. Authorization

- Login
- Signup
- Logout

Conclusion: The authorization features, including login, signup, and logout, are fundamental for secure and personalized user account management. These functions enable users to access and manage their profiles, participate in social features, and securely interact with the platform. This ensures a user-friendly and interactive streaming service, aligning with the project's goal.

2. View Movie Information

- Title, description, genre, rating, content warnings, release date
- Series (Episode title, episode description, episode length)
- Cast, director, user reviews

Conclusion: Providing information about movies and series is crucial for enhancing the user experience. Detailed content pages allow users to make informed decisions by accessing titles, descriptions, genres, ratings, and more. This requirement addresses the need for showcasing existing content with additional data.

3. Create/Edit/Delete Movies/Series

- Add movies/series
- Edit movies/series
- Delete movies/series
- Add/Create/Edit Actors, directors, genres, and production studios Conclusion: The ability to create, edit, and delete movies and series, along with managing related entities like actors and genres, ensures the platform remains current and relevant. These functionalities support dynamic content management, allowing continuous updates to the database. This dynamic approach is crucial for maintaining a curated collection that meets user demands, as outlined in the project scenario.

4. Discovery

- Search
- Filter
- Sort
- Recommendation

Conclusion: Implementing robust search, filter, sort, and recommendation features significantly enhances the user experience by making it easier to discover content that matches users' interests. These discovery tools are essential for navigating the content library.

- 5. Reviews
- Create review
- Add stars
- Create comment

Conclusion: The review system, including the ability to create reviews, add star ratings, and comment, fosters a sense of community and interactivity among users. These social features encourage user engagement and provide valuable insights into the content.

- 6. WatchList
- Adding a movie to the list

Conclusion: The watchlist feature allows users to create and manage personalized lists of movies and series they plan to watch. This functionality enhances user engagement by providing a convenient way to organize content. It also facilitates content sharing among friends, promoting a more connected user experience. Overall Alignment with Project Specifications Conclusion: The functional requirements ensure that the product meets milestones of the project specifications. By focusing on authorization, detailed content viewing, dynamic content management, discovery tools, interactive reviews, and personalized watchlists, the product is designed to create a user-friendly streaming platform. These features enhance the user experience and ensure that the Hoop streaming service meets the expectations outlined in the project scenario.

Task 6 - Data

The data was gotten from the following sites:

• https://developer.imdb.com/non-commercial-datasets/

- https://grouplens.org/datasets/movielens/latest/
- https://theposterdb.com

The data was entered into the database semi-manually. The main source of the data was from grouplen which was about 5000 media. The imdb database was then joined onto this database so that the only data remaining was relevant to the chosen media content. This is the reason that media_ID is the same as the imdbID. After this was done the list of media was further reduced to 130 fbased on how recent it was. This was done since the posters had to be entered manually and +/- 100 entities was a reasonable amount of data. All the while genre and crew are updating so only relevant entries remain, done because of the cascade attributes of relevant tables.

Entities like user, reviews and favourite genres will be entered manually, by the user or by admin. It was not got from any data source.

belong

genreID	media_ID
1	346631
3	362120
3	366548
4	373883

content_review

reviewID	comments	starRating	medialD	created_at
9	An absolute masterpiece! The storytelling and cine	5	69049	2024-05-30 08:16:05
10	A unique and refreshing take on the genre. Highly	4	196291	2024-05-30 08:16:05
11	Interesting concept, but the execution could have	3	219432	2024-05-30 08:16:05
12	Not really my cup of tea. The plot felt a bit disj	2	262158	2024-05-30 08:16:05
13	Fantastic visuals and a gripping story. Loved ever	5	416236	2024-05-30 08:16:05
14	Exciting and well-paced. The characters were well	4	417325	2024-05-30 08:16:05

crew

crewID	name	role
40	Stanley Kubrick	director
56	Paul Newman	actor
80	Orson Welles	director
84	Gong Li	actor

entertainment_content

r	media_ID	title	release_Date	description	content_rating	rating	poster_Url
	381061	Casino Royale	2006	Le Chiffre, a banker to the world's terrorists, is	PG	3.25	https://image.tmdb.org/t/p/original/
	385880	Monster House	2006	Monsters under the bed are scary enough, but what	PG	3.25	https://image.tmdb.org/t/p/original/
	395495	Catch and Release	2006	Gray Wheeler just lost everything. But it could be	PG 13	1.51	https://image.tmdb.org/t/p/original/

favourite_genre

username	genreID
Charlie_worldwide	10
Charlie_worldwide	3
Charlie_worldwide	7

friend

username	friendID
alex_wong	jane_smith
jane_smith	emily_jones
thatGirlTrudy45	alex_wong

genre

genreID	genreName
10	action
12	animation
3	comedy
9	crime

involved_in

crewID	media_ID
80	69049
123	465538
230	462499
243	1037705

movie

media_ID	duration
346631	93
362120	83
366548	108
373883	109

tv_show

media_ID	seasons
903747	5
948103	1
972534	6



user

wants_to_watch

username	medialD
Charlie_worldwide	332902
Charlie_worldwide	337926
Charlie_worldwide	341957

writes

reviewID	username
9	sophia_star
10	benji_m
11	mia_a
12	alex_thomas

Task 7 – Optimization

Query:

```
SELECT title, GROUP_CONCAT(genreName) AS genre, release_Date,
description, content_rating, rating FROM entertainment_content
INNER JOIN belongs ON entertainment_content.media_ID =
belongs.media_ID
LEFT JOIN genre ON belongs.genreID = genre.genreID
WHERE genre.genreID IN (
    SELECT genreID FROM favourite_genre
    INNER JOIN user on favourite_genre.username = user.username
    WHERE user.username = 'alex_wong'
    )
GROUP BY entertainment_content.media_ID
HAVING COUNT(belongs.media_ID) > 1
ORDER BY COUNT(belongs.media_ID) DESC;
```

Purpose

This query gets a list of all films that have 2 or more of the users favourite genres in it. This query is the backbone of the recommendation aspect of the website. Without it, the ability to give the user relevant recommendations would be hampered.

Performance

- Mysql → 9 row(s) returned 0.016 sec / 0.000 sec
- DataGrips → 9 rows retrieved starting from 1 in 156 ms

Interpretation

- Large amount of JOINS
- Redundant check on USER table
- Nested Query slows down query

Optimise

- All JOINS move to FROM parameter with join conditions in WHERE parameter
- Remove the redundant user check with a check directly to the favourite_genre table
- Remove the Nested Query

Optimised Query

```
SELECT title, GROUP_CONCAT(genreName) AS genre, release_Date,
description, content_rating, rating
FROM entertainment_content AS e, belongs AS b, genre AS g,
favourite_genre AS f
WHERE e.media_ID = b.media_ID AND b.genreID = g.genreID AND
f.genreID = g.genreID AND f.username = 'alex_wong'
GROUP BY e.media_ID
HAVING COUNT(b.media_ID) > 1
ORDER BY COUNT(b.media_ID) DESC;
```

Performance

Mysql → 9 row(s) returned 0.000 sec / 0.000 sec

Datagrips → 9 rows retrieved starting from 1 in 68 ms