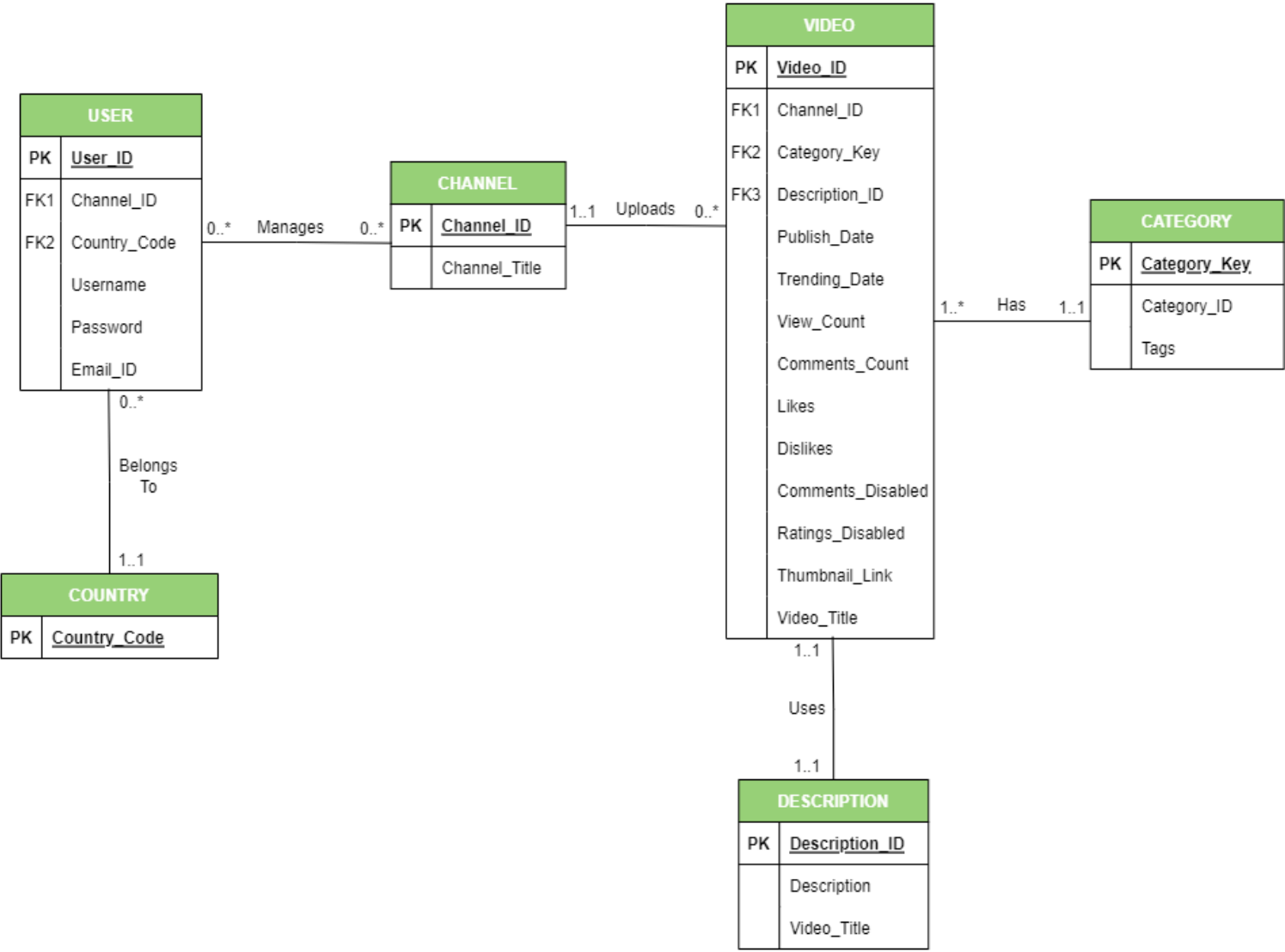


STAGE 2: Database Design

UML Diagram :



Entity Assumptions:

1. User
 - User has the primary key user_ID which is used to identify each user
 - A user has a username, password, and email_ID for logging in
 - Each user can also input a channel_ID for their channel if they have one
 - Each user also has a country_code which corresponds to the country they reside in
2. Channel
 - Channel has primary key channel_ID which is the identifier for their channel
 - Each channel has a channelTitle
3. Video
 - Video has primary key video_ID which identifies each video
 - Has a channel_ID corresponding to the channel that uploaded the video
 - Has a category_key that is given to give the category across each video a unique identifier
 - Has a description_ID that is given to the description across each video a unique identifier
 - A video also has a publish_date, trending_date, view_count, comments_count, likes, dislikes, comments_disabled, ratings_disabled, thumbnail_link, and a video_Title
4. Category
 - Category has a primary key category_key which gives all videos a unique identifier for its category
 - Category also has a category_ID and tags
5. Description
 - Description has a primary key description_ID used to give each videos description a unique identifier
 - Each description_ID has a corresponding description and video_title that we will run sentiment analysis on to give more information about
6. Country
 - Country has a country_code given to each country

Relationships:

1. Manages

‘Manages’ is the relation between user and channel. A user doesn’t have to manage a channel, and a user can also manage multiple channels. A channel can be managed by multiple people because multiple people may own one Youtube channel. At the same time, not every Youtuber may use our website. This is a many-many relationship.

2. Uploads

‘Uploads’ is the relation between channel and video. A channel can upload 0 or more videos, but a video can only belong to 1 channel. This is a 1-many relationship.

3. Has

‘Has’ is the relation between video and category. Each video only has 1 category, but each category can belong to multiple videos. This is a 1-many relationship.

4. Uses

‘Uses’ is the relation between video and description. Each video uses only 1 description. Among trending videos, the description will have a lot of information, so each description only has one video. This is a 1-1 relation.

5. BelongsTo

‘BelongsTo’ is the relation between user and country. Each user belongs to only 1 country since they can only be in one country at a time. But a country can have 0 or more users. This is a 1-many relation.

Normalization:

Boyce-Codd Normal Form (BCNF) was chosen because it eliminates redundancy resulting in a smaller database and perhaps faster query performance. Moreover, It prevents insertion, deletion, and update errors. Data inconsistency can result from anomalies. Lastly, it maintains consistency and accuracy in the database to ensure data integrity. In addition to this, it is a stricter form of 3NF, hence, if our database satisfies BCNF, it automatically satisfies 3NF.

To show that the following Entity Relationship schema adheres to BCNF, we need to demonstrate that “ A relation is in BCNF if, and only if, for every non-trivial functional dependency $X \rightarrow Y$, X is a superkey. A superkey is a set of attributes that can uniquely identify a tuple in a relation.

1. User: UserID, username, password, email, Channel Id, Country_code are all attributes.

Applying BCNF decomposition:

FD_User_1: UserID \rightarrow email, username, password, Country_code

For the first FD, since UserID is the primary key for each user, it is a superkey, and this FD follows BCNF.

FD_User_2: email, username, password \rightarrow Channel Id

The combination of email, username and password is considered a superkey for this FD because it uniquely defines other properties of the channel Id

Thus, FD_User_1 and FD_User_2 satisfies BCNF criteria ----- a)

2. Country: Country_code is the only attribute. Because Country_code is the single attribute, there are no functional dependencies with combination attributes.

Thus, FD_Country is None which satisfies BCNF criteria ----- b)

3. Channel: Channel Id, Channel Title are the existing attributes.

FD_Channel: Channel Id \rightarrow Channel Title

Because any non-trivial FD in a two-attribute relation indicates that one of the attributes is a superkey, two-attribute relations always meet the BCNF criterion.

Thus, FD_Channel satisfies BCNF criteria ----- c)

4. Category: Category key, Category Id, Tags are the existing attributes.

FD_Category: Category key \rightarrow Category Id, Tags

Category key is a superkey that determines both Category Id and Tags.

Thus, FD_Category satisfies BCNF criteria ----- d)

5. Description: Description_Id, Description, Video_title

FD_Description: Description_Id \rightarrow Description, Video_title

Description_Id is a superkey that determines both Description and Video_title

Thus, FD_Description satisfies BCNF criteria ----- e)

6.Video: Video_Id, Channel Id, Category key, Description Id, Publish date, Trending Date, view_Count, comment_count, likes, dislikes, Comments_Disabled, Ratings- Disabled, thumbnail - link, Video Title are the attributes:

Applying BCNF decomposition,

FD_Video_1: Video_Id \rightarrow Channel Id, Description Id, Category key, Video Title

FD_Video_2: Channel Id, Description Id \rightarrow likes, dislikes, Comments_Disabled, Ratings- Disabled, thumbnail - link

Each unique video (identified by its Video_Id) has a specific channel to which it belongs (Channel Id), a specific description linked with it (Description Id), a specific category to which it belongs (Category key), and a specific title (Video Title). There is an exact amount of likes, dislikes, a status of whether comments are disabled, a status of whether ratings are disabled, and a specific thumbnail link for each unique combination of a channel (identified by Channel Id) and a description (identified by Description Id).

Thus, FD_Video_1 and FD_Video_2 satisfies BCNF criteria ----- f)

From a), b), c), d), e) and f), we can see that “whenever there is a nontrivial FD, its left side is a superkey”

Hence, the ER diagram adheres to the BCNF form.

Relational Schema

- User(
 User_ID: INT [PK],
 Username: VARCHAR(255),
 Password: VARCHAR(255),
 Email_ID: VARCHAR(255),
 Channel_ID: VARCHAR(255) [FK to Channel.ChannelId],
 Country_Code: VARCHAR(255) [FK to Country.Country_code]
)
- Country(
 Country_Code: VARCHAR(255) [PK]
)
- Channel(
 Channel_ID: VARCHAR(255) [PK],
 Channel_Title: VARCHAR(255)
)
- Category(
 Category_Key: INT [PK],
 Category_ID: INT,
 Tags: VARCHAR(255)
)
- Description(
 Description_ID: INT [PK],
 Description: VARCHAR(255),
 Video_Title: VARCHAR(255)
)
- Video(
 Video_ID: VARCHAR(11) [PK],
 Channel_ID: VARCHAR(255) [FK to Channel.ChannelId],
 Category_Key: INT [FK to Category.CategoryKey],
 Description_ID: INT [FK to Description.DescriptionId],
 Publish_Date: DATETIME,
 Trending_Date: DATETIME,
 View_Count: INT,

Comment_Count: INT,
Likes: INT,
Dislikes: INT,
Comments_Disabled: BOOLEAN,
Ratings_Disabled: BOOLEAN,
Thumbnail_Link: VARCHAR(255),
Video_Title: VARCHAR(255)
)

