ASSIGNMENT 2 REPORT

1. ABSTRACT

In this assignment, we are picking the entertainment domain particularly movies and building a movie database gathering real world data. Social media platform twitter is then analyzed in relation to the database built. The domain has three entities, the movie which represents the company entity, production house which represents the producer entity and the social media users on twitter discussing the movies who represent the consumer entity. Conceptual model and E-R Diagrams are defined specific to the domain and a physical database is built based on the models and queries are run to get information from the database related to the domain.

2. DATA SOURCES AND DATA FIELDS:

The data for the three entities company, producer and consumer are gathered from the following sources:

(i) Company – movies:

Movies data for films released in 2002 are gathered from the social media site IMDB using the TMDB web API. An api_key is used to gain access and get information from the TMDB API. Information is obtained from multiple pages (about 800 records of movies) and stored in a data frame and exported to a CSV file.

URL page accessed:
https://api.themoviedb.org/3/discover/movie?api_key=21aefd57526da8bc838461554ca6cc49&primary_release_year=2002&page=1
(100 consecutive pages are fetched)

Data fields obtained from the web API include:
- Title (Name of the movie)
- Release_date (The Release date of the movie)
- Overview (Overview of the movie)
- Adult (Adult nature of the movie)
- Vote_count (Vote_count for the movie)
- Vote_average (Vote_average of the movie)
- Popularity (Popularity of the movie)

(ii) Producer – production house

Production house details of films released in 2002 are gathered from a Kaggle dataset. The dataset has about 5000 records of movies out of which the movies released in 2002 are selected as it is the data we are working on. The data is a csv file which is imported using the pandas framework and manipulated. The columns that are of interest are fetched from the dataset.

Kaggle dataset:
https://www.kaggle.com/tmdb/tmdb-movie-metadata

Data fields obtained from the Kaggle dataset include:
- Title (Name of the movie)
- Production_companies (Production companies of the movie)
- Production_countries (Countries involved in the production of the movie)
- Revenue (Revenue earned by the movie)

(iii) Consumer – movie users(on twitter)

The viewers of the movies form the consumers. Information of the set of people who have tweeted about the movies is collected. Twitter API is used to fetch the user details.
From each user the past 100 tweets and the related information is gathered.

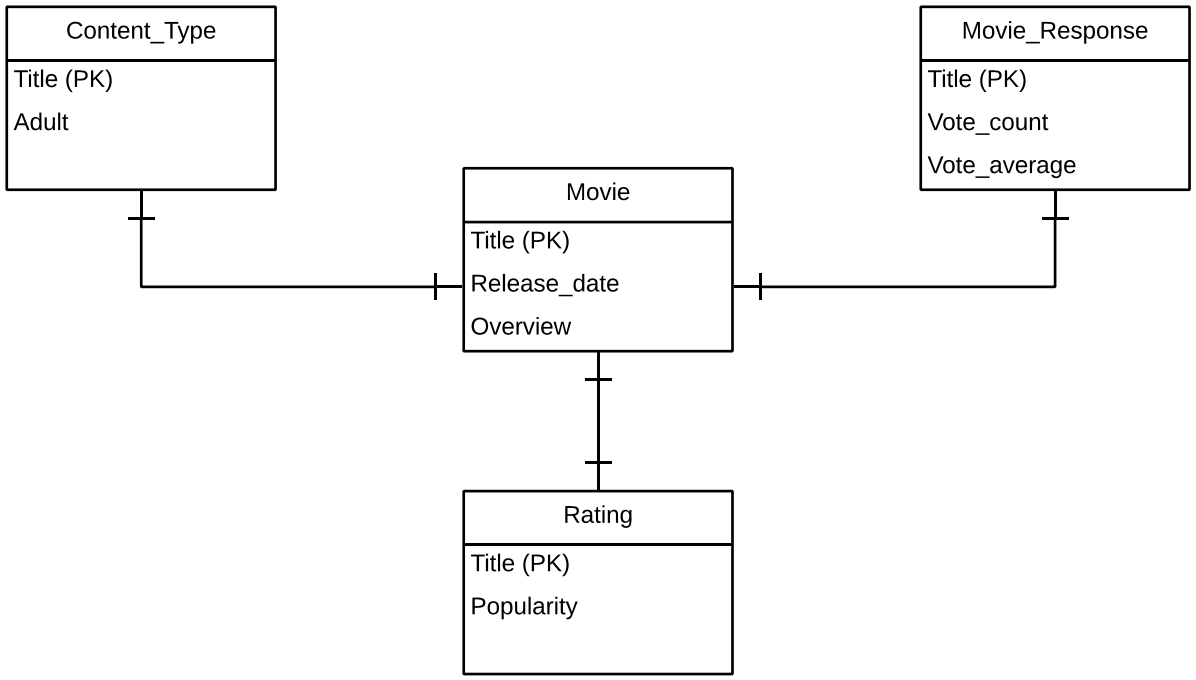
Data fields obtained from the Twitter API:
- Name (Screen name of the twitter user)
- Title (Movie the user has tweeted about)
- User-id (Twitter User ID)
- Followers (Followers count of the user)
- Post (Tweet posted by the user)
- Hashtags (Hashtags used in the tweet)
- Posted_time (Time the tweet was posted)
- Favorites (Favorites count)

3. CONCEPTUAL SCHEMA EXPLANATION:

The movie database has three entities. They are –
Company – Movie
Producer – Production house
Consumer – Movie viewer

The conceptual schemas for the three entities are as follows:

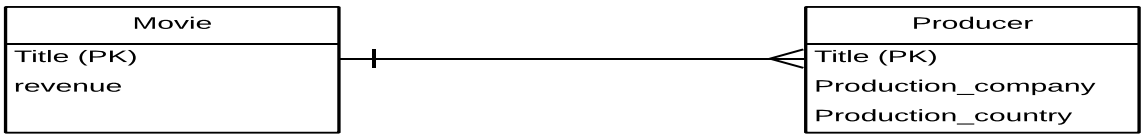
Company:



The conceptual schema shows the entities and the relationship between the entities. The various entities are as
follows:
Movies, Release_date, Overview, Adult, Vote_count, Vote_average, Popularity.
- Movie entity (Primary Key: Title) has attributes such as movie title, release_date and overview.
- Rating (Primary Key: Title) entity has attributes such as movie title, Popularity.
- Movie_response (Primary Key: Title) has attributes such as movie title, vote_count and vote_average.
- Content_type (Primary Key: Title) has attributes such as movie title, adult.

Relationships between the entities are as follows:
- Movies and Rating has a one to one relationship as one entry in movies as one movie could have one matching entry in rating table.
- Movie_response and rating has a one to one relationship as one entry in movie_response could have one
matching entry in rating table.
- Movies and content_type has a one to one relationship as one entry in ref_movie_genres could have many matching entry in movie_genre table.

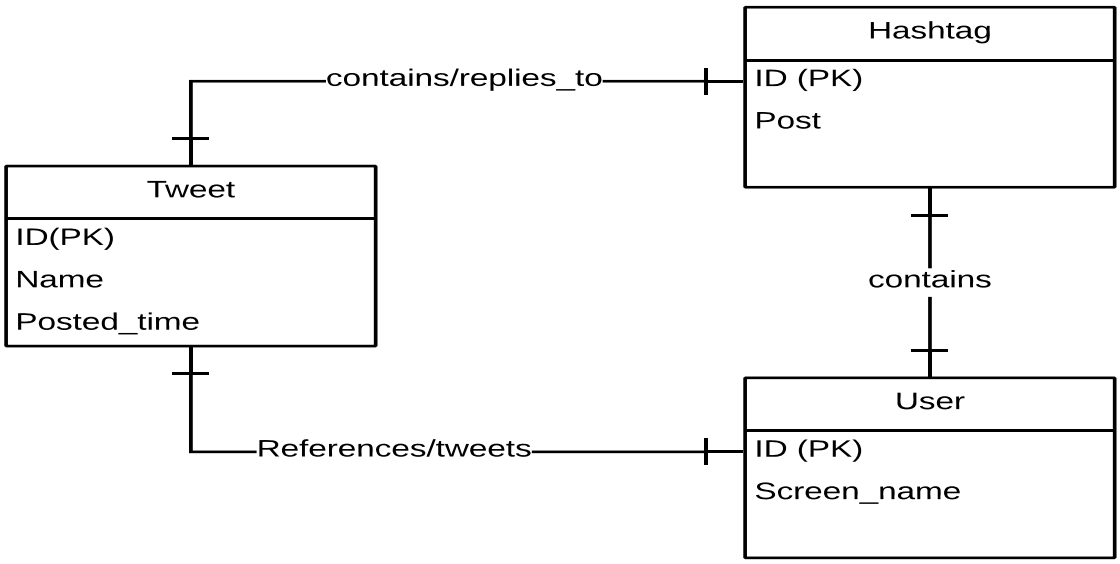
Producer:



The conceptual schema shows the entities and the relationship between the entities. The various entities are as
follows:
Movies, Producer.
- Movie entity (Primary Key: Title) has attributes such as title, revenue.
- Producer (Primary Key: Title) entity has attributes such as title, Production_companies, Production_countries.

Relationships between the entities are as follows:
- Movies and Producer has a one to many relationship as one entry in movies as one movie could have multiple matching entries in Producer table.

Consumer:



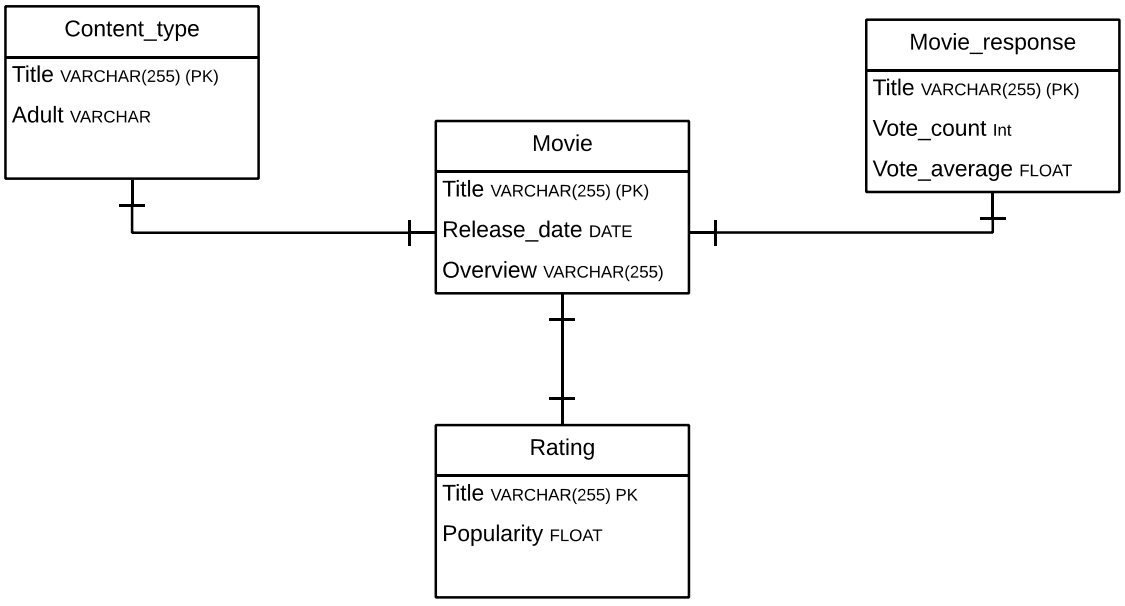
The conceptual schema shows the entities and the relationship between the entities. The various entities are as
follows:
Tweet, Hashtag, User.
- Tweet entity (Primary Key: ID) has attributes such as ID, Name, Posted_time.
- Hashtag (Primary Key: ID) entity has attributes such as ID, Post.
- User (Primary Key: ID) has attributes such as movie ID, Screen_name.

Relationships between the entities are as follows:
- Tweet and Hashtag has a one to one relationship as one entry in tweet could have one matching entry in Hashtag table.
- Tweet and User has a one to one relationship as one entry in tweet can have one matching entry in user table.
- Hashtag and User has a one to one relationship as one entry in hashtag could have one matching entry in user table.

4. ER DIAGRAMS:

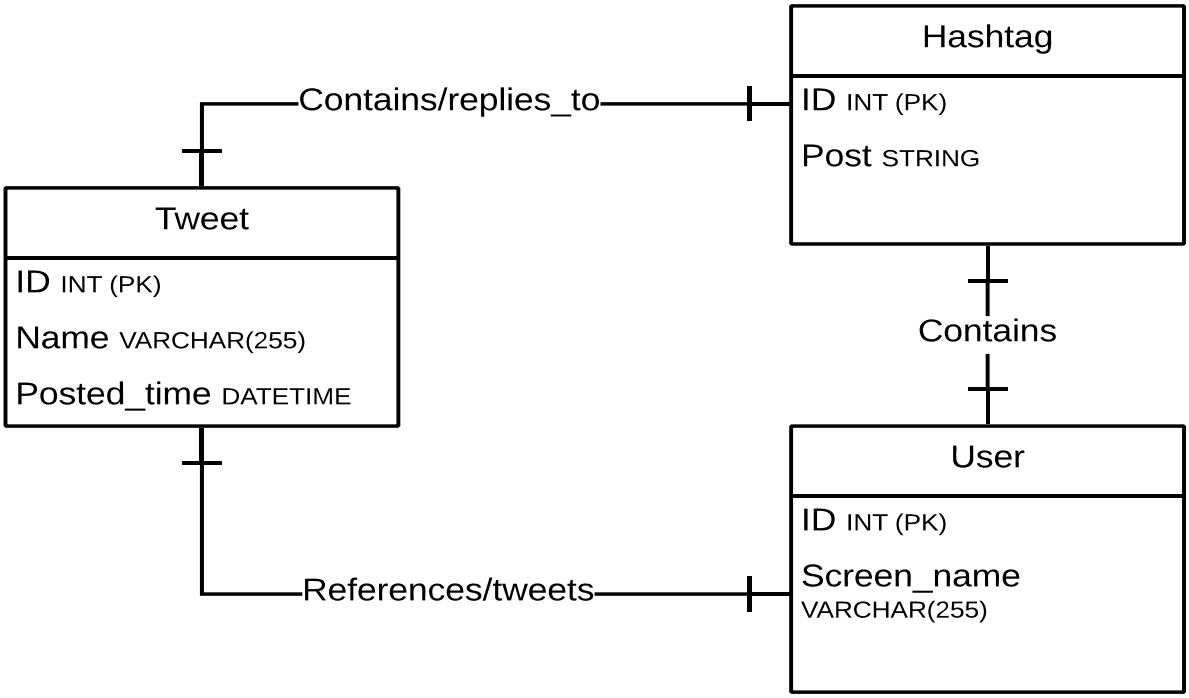
The ER Diagrams for the three entities are as follows:

Company:



Producer:

Consumer:



5. QUESTIONS ABOUT YOUR CONCEPTUAL MODEL:

1. What are the ranges, data types and format of all of the attributes in your entities?

Attributes in company:

- Title (Name of the movie) – VARCHAR(255)
  Eg. Harry Potter and the chamber of secrets
  (Range is less than or equal to 255 variable characters)

- Release_date (The Release date of the movie) – DATE.
  Eg. 2002-11-13
  (Range is 2002-01-01 to 2002-12-31)
  (Format – yyyy-mm-dd)

- Overview (Overview of the movie) - VARCHAR(255).
  Eg. Ignoring threats to his life, Harry returns to Hogwarts to investigate – aided by Ron and Hermione – a mysterious series of attacks.
  (Range is less than or equal to 255 variable characters)

- Adult (Adult nature of the movie) – VARCHAR
  Eg. False
  (Range is it can take a value of either True or False)

- Vote_count (Vote_count for the movie) – INT
  Eg. 11040
  (Range is -32768 to 32767)

- Vote_average (Vote_average of the movie) – FLOAT
  Eg. 7.7
  (Range includes 8-byte IEEE floating point number)

- Popularity (Popularity of the movie) – FLOAT
  Eg. 39.232
  (Range includes 8-byte IEEE floating point number)

Attributes in Producer:

- Title (Name of the movie) – VARCHAR(255)
  Eg. Harry Potter and the chamber of secrets
  (Range is less than or equal to 255 variable characters)

- Production_companies (Production companies of the movie) – VARCHAR(255)
  Eg. "1492 Pictures", "Heyday films", "Warner Bros."
  (Range is less than or equal to 255 variable characters)

- Production_countries (Countries involved in the production of the movie) – VARCHAR(255)
  Eg. "United States of America"
  (Range is less than or equal to 255 variable characters)

- Revenue (Revenue earned by the movie) – INTEGER
  Eg. 876688482
  (Range from -2**63 to (2**63-1))

Attributes in Consumer:

- Name (Screen name of the twitter user) - VARCHAR(255)
  Eg. Complex
  (Range is less than or equal to 255 variable characters)

- Title (Movie the user has tweeted about) – VARCHAR(255)
  Eg. Harry Potter and the chamber of secrets
  (Range is less than or equal to 255 variable characters)

- User-id (Twitter User ID) – INTEGER
  Eg. 215298957
  (Range from -2**63 to (2**63-1))

- Followers (Followers count of the user) – INTEGER
  Eg. 91674
  (Range from -2**63 to (2**63-1))

- Post (Tweet posted by the user) – STRING
  Eg. “Welcome to your interview, which Dark Souls game would you like to see me play and why?”
  (Range contains characters, numbers and special characters)

- Hashtags (Hashtags used in the tweet) – STRING
  Eg. FACT
  (Range contains characters, numbers and special characters)

- Posted_time (Time the tweet was posted) – STRING
  Eg. Mon Feb 25 17:19:00 +0000 2019
  (Range contains characters, numbers and special characters)

- Favorites (Favorites count) – INTEGER
  Eg. 1169
  (Range from -2**63 to (2**63-1))

2. When should you use an entity versus attribute? (Example: address of a person could be modeled as either)

- **Entity** is an **entity** is a single person, place, or thing about which data can be stored. The database stores all the values of **entities** and relationships. **Attribute** is the characteristic property of an existing **entity**.
  Entity is used when we can define the concept and identify multiple attributes. Attribute is used when it is a simple measure or classification.
  Eg. If we consider an address to be an entity or attribute, it can be an attribute for a person entity, however, it becomes an entity if we are talking of a real estate table.

3. When should you use an entity or relationship, and placement of attributes? (Example: a manager could be modeled as either)

- Entities are the things we need to store data about. Relationships describe how entities relate to each other. Manager could be a entity if it is a company database or an attribute if it is an employee database. Manager manages employees becomes a relationship.

4. How did you choose your keys? Which are unique?

- In company and producer tables the primary key is title of the movie. Since the movie titles cannot be the same, it was chosen as the primary key.
  The consumer entity has two tables. In the table which contains title of the movie and name of the twitter screen name, twitter screen name is the primary key as it is unique for every person.
  In the other table an ID which autoincrements is used as the primary key as others could have duplicate values.

5. Did you model hierarchies using the "ISA" design element? Why or why not?

- Yes. "ISA" design is used when if in a two relational database tables, one table has a foreign key that references the primary key of the other table exists. In our case, the title field in the Movie_tweet table references the title in company and producer and hence a "ISA" relationship is being used.

6. Were there design alternatives? What are their tradeoffs: entity vs. attribute, entity vs. relationship, binary vs. ternary relationships?

- There weren't many design alternatives as data was being modeled in the company-producer-consumer fashion. As it is relatively small database the entities, attributes ad relationships were fairly clear.

7. Where are you going to find real-world data to populate your model?

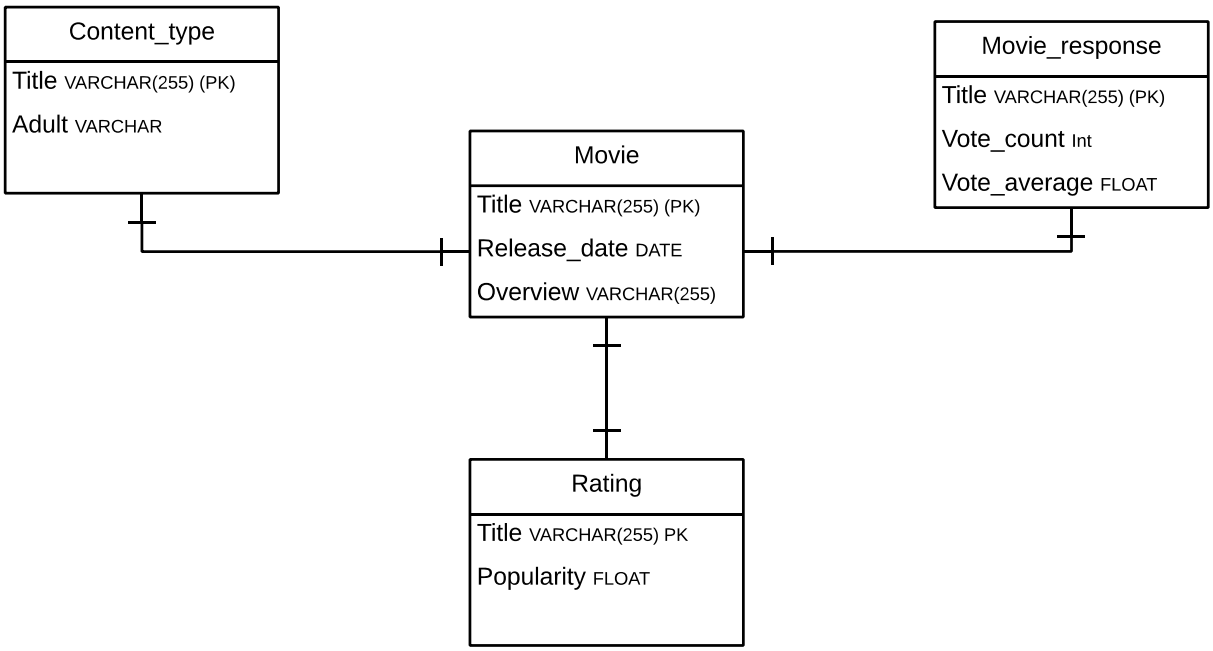
- Real world data pertaining to movies (from IMDB and TMDB) is used to populate the model.

Information can be gathered from API's or Web Scraping can be performed on relevant websites and datasets readily available can be used to populate the model.

PART II

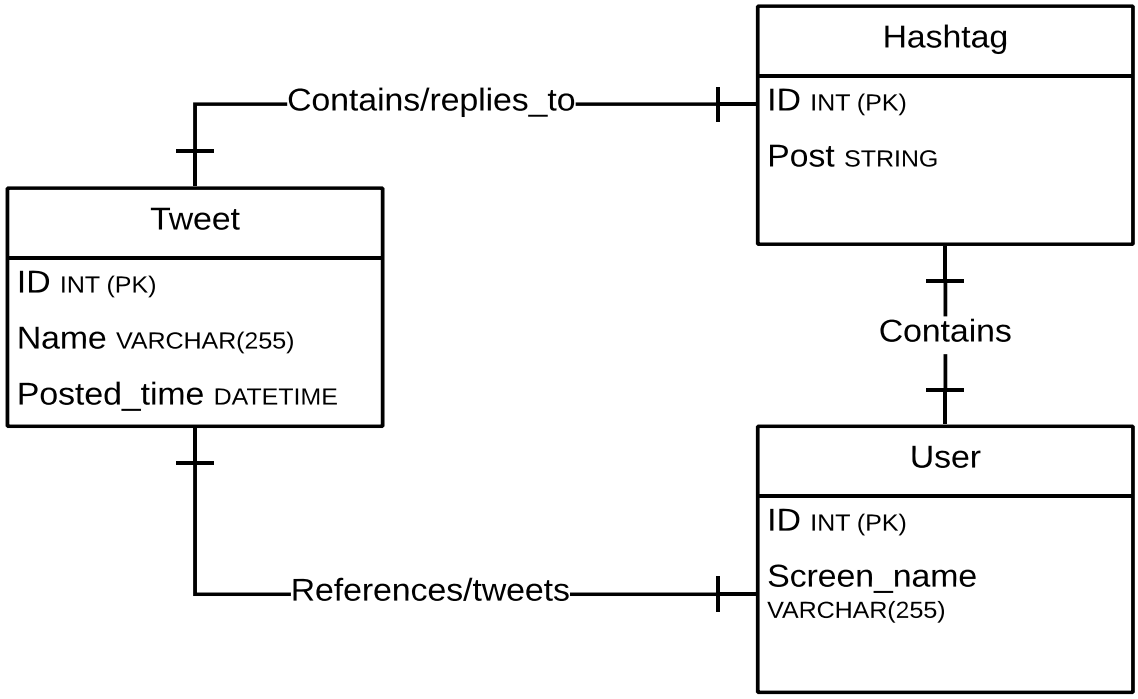
6. SQL AND DIAGRAM FOR THE PHYSICAL MODEL

1. Company:



2. **Producer:**

3. **Consumer:**



DATABASE SCHEMAS:

```
CREATE TABLE company(
title VARCHAR(255) PRIMARY KEY NOT NULL,
Release_date DATE,
Overview VARCHAR(255),
Adult VARCHAR,
Vote_count INT,
Vote_average FLOAT,
Popularity FLOAT
);

CREATE TABLE producer(
title VARCHAR(255) PRIMARY KEY NOT NULL,
production_companies VARCHAR(255),
production_countries VARCHAR(255),
revenue INTEGER
);

CREATE TABLE consumer(
Name VARCHAR(255) PRIMARY KEY NOT NULL,
User_id INTEGER,
Followers INTEGER,
Post STRING,
Hashtags STRING,
Posted_time STRING,
Favourites INTEGER
);

CREATE TABLE tweets(
title VARCHAR(255) PRIMARY KEY NOT NULL,
name VARCHAR(255),
FOREIGN KEY(name) REFERENCES consumer(Name)
);
```

SAMPLE DATABASE TABLES:

Company – Movies

title	Release_date	Overview	Adult	Vote_count	Vote_average	Popularity
Harry Potter and the Chamber of Secrets	2002-11-13	Ignoring threats to his life, Harry returns to Hogwarts to i...	False	11040	7.7	39.232
Hart's War	2002-02-15	Fourth-generation Army Col. William McNamara is imprison...	False	383	6.2	9.285
He Loves Me... He Loves Me...	2002-03-27	Angelique, a young student, is in love with a married doctor....	False	120	6.7	6.234
Heart of America	2002-09-05	Starring Jurgen Prochnow ('The Replacement Killers' 'Air Force...	False	11	4.8	1.487
Heartlands	2002-08-20	Gentle Colin 'Col' Lawes happily lead a quiet life, running a ne...	False	11	6.3	1.79
Heaven	2002-02-06	A woman takes the law into her own hands after police ignore...	False	114	6.6	7.649

Producer – Production house

title	production_companies	production_countries	revenue
Harry Potter and the Chamber of Secrets	[{"name": "1492 Pictures", "id": 436}, {"name": "Heyday films", "id": 437}, {"name": "Warner Bros.", "id": 6194}]	[{"iso_3166_1": "DE", "name": "Germany"}, {"iso_3166_1": "GB", "name": "United Kingdom"}, {"iso_3166_1": "US", "name": "United...	876688482
The Lord of the Rings: The Two Towers	[{"name": "WingNut Films", "id": 11}, {"name": "New Line Cinema", "id": 12}, {"name": "The Saul Zaentz Company", "id": 5237}]	[{"iso_3166_1": "NZ", "name": "New Zealand"}, {"iso_3166_1": "US", "name": "United States of America"}]	926287400
Men in Black II	[{"name": "Columbia Pictures", "id": 5}, {"name": "Amblin Entertainment", "id": 56}, {"name": "Columbia Pictures Corporatio...	[{"iso_3166_1": "US", "name": "United States of America"}]	441818803
Catch Me If You Can	[{"name": "Kemp Company", "id": 367}, {"name": "Splendid Pictures", "id": 368}, {"name": "Parkes/MacDonald Productions",...	[{"iso_3166_1": "US", "name": "United States of America"}]	352114312
Spider-Man	[{"name": "Columbia Pictures", "id": 5}, {"name": "Marvel Enterprises", "id": 19551}]	[{"iso_3166_1": "US", "name": "United States of America"}]	821708551
Ice Age	[{"name": "Twentieth Century Fox Film Corporation", "id": 306}, {"name": "Blue Sky Studios", "id": 9383}, {"name": "Twentieth Ce...	[{"iso_3166_1": "US", "name": "United States of America"}]	383257136

Consumer – Movie viewers

Name	User_id	Followers	Post	Hashtags	Posted_time	Favourite
Kevin	215298957	91674	@iTempp I could never pass the candy levels...	[]	Tue Mar 05 17:11:27 +0000 2019	69
Kevin	215298957	91674	@cookingsim could I please get a code for your game? I'm a really good chef	[]	Tue Mar 05 17:08:11 +0000 2019	9
Kevin	215298957	91674	What's your favorite ps1 game?	[]	Tue Mar 05 17:06:55 +0000 2019	1169
Kevin	215298957	91674	@iBallisticSquid @ashdubh Hahaha thank you for spreading the word of the cult	[]	Mon Mar 04 13:33:19 +0000 2019	33
Kevin	215298957	91674	@ashdubh Cheers dude, really appreciate it :)	[]	Mon Mar 04 13:14:50 +0000 2019	383
Kevin	215298957	91674	@_ejayb WHAT	[]	Mon Feb 25 22:54:37 +0000 2019	131
Kevin	215298957	91674	@cherryrsr don't make me do this	[]	Mon Feb 25 22:12:24 +0000 2019	385

title	Name
Harry Potter and the Chamber of Secrets	CallMeKevin1811
Harry Potter and the Chamber of Secrets	R3xoss
Harry Potter and the Chamber of Secrets	FortniteGame
Harry Potter and the Chamber of Secrets	Complex
Harry Potter and the Chamber of Secrets	Keggerade
The Lord of the Rings: The Two Towers	BertShowBert
The Lord of the Rings: The Two Towers	sadrianom

7. QUERIES YOU MUST ANSWER ABOUT YOUR PHYSICAL MODEL (IN SQL):

a. What user posted this (e.g. tweet, facebook post, IG post, etc.)?

- SELECT Name from consumer where Post="Welcome to your interview, which Dark Souls game would you like to see me play and why?";

=> Kevin

b. When did the user post this (e.g. tweet, facebook post, IG post, etc.)??
- SELECT Posted_time from consumer where Post="Welcome to your interview, which Dark Souls game would you like to see me play and why?"

=> Mon Feb 25 22:03:29 +0000 2019

c. What posts has this user posted in the past 24 hours?

- SELECT Post FROM consumer WHERE (Name="Kevin" AND Posted_time LIKE "%Mar%05%");

=>@iTempp I could never pass the candy levels :(
@cookingsim could I please get a code for your game? I'm a really good chef
What's your favorite ps1 game?

d. How many post has this user posted in the past 24 hours?

- SELECT count(Post) FROM consumer WHERE (Name="Kevin" AND Posted_time LIKE "%Mar%05%");

=> 3

e. What keywords/ hashtags are popular?

- SELECT Hashtags,COUNT(Hashtags) AS cnt FROM consumer GROUP BY Hashtags ORDER BY cnt DESC

=>	GameofThrones	17
	FortniteCreative	13
	LeavingNeverland	7
	ComplexNews	6
	ad	4
	CaptainMarvel	3
	ForTheThrone	3
	GameOfThrones	3
	ATLUTD	2
	GoT	2
	HotOnes	2
	JonSnow	2
	OneLastDance	2
	PancakeDay	2
	SharetheLove	2
	TheBurgerShow	2
	classof2017	2
	Beto	1
	Beto2020	1
	CheesedChallenge	1
	DaenerysTargaryen	1
	ESLKatowiceRoyale	1
	Emmys	1
	FACT	1
	FOOLishlove	1

FatherOfAsahd	1
FortniteWorldCup	1
GameofThronesLive	1
GenZ	1
GovGroup	1
JobsUnlisted	1
JonasBrothers	1
KevinUnderwood	1
Lyft	1
MPHS	1
MakeAmericaWorse	1
MarcPeeples	1
Mediterraneandiet	1
PeteDavidson	1
PeterDinklage	1
PokemonSwordShield	1
SMHdoEurope	1
SMHdoSoutheastAsia	1
SoFarGone	1
SonicTheHedgehog	1
SuperBowlLIII	1
The80sIn5Words	1
TheFinalSeason	1
UT21	1
VansChallenge	1
WomensHistoryMonth	1
almostthere	1
bba2019	1
cutmerealdeepjustnow	1
greatwolflodge	1
icameto	1
literallyanything	1
mphosa	1
prom2k17	1
soccer	1
socloseyetsofaraway	1
whydoesthisalwayshappentome	1

f. What posts are popular?

- SELECT Post,COUNT(Post) AS cnt FROM consumer GROUP BY Post ORDER BY cnt DESC;

=> "Anyone Can Wear the Mask"

This quote relates to twitch as well, anyone can be a streamer. https://	1
"My story has always been deeper than rap."	

@RickRoss' memoir, 'Hurricanes,' is set to hit shelves 1
"Shall we begin?" -Daenerys Targaryen

HELL YEA BABY! This season is off to an incredible start! #Ga 1
"You better know how to swim..."

Offset has some advice for 6ix9ine → https://t.co/kcRnYW3myy http 1
"You was supposed to be my Beyonce now you this n****s fiance?" 🎶

Is Meek Mill addressing Nicki Min 1

8. USE CASES:

Use case 1: Query for finding name and revenue of the movie which made the highest revenue

- SELECT title, MAX(revenue) FROM Producer;
- ⇨ The Lord of the Rings: The Two Towers 926287400

Use case 2: Query for overview and popularity of Harry Potter and the Chamber of Secrets

- SELECT Overview,Popularity from company where title="Harry Potter and the Chamber of Secrets";
- ⇨ Ignoring threats to his life, Harry returns to Hogwarts to investigate – aided by Ron and Hermione – a mysterious series of attacks. 39.232

Use case 3: Query for finding release_date, overview, vote_count, vote_average, popularity of the movie Highway

- SELECT Release_date, Overview, Vote_count, Vote_average, Popularity from company where title="Highway";
- ⇨ 2002-01-01 Jack is caught with the wife of his employer, a Vegas thug. The thug sends goons after Jack, who convinces his best friend, Pilot, to flee with him. Pilot insists that they head for Seattle, but doesn't tell Jack why. The goons learn from Pilot's drug source where the youths are headed, and they follow, hell bent on breaking Jack's feet. On the road, Jack and Pilot give a ride to Cassie, a distressed young woman. She and Jack hit it off. They pick up an aging stoner headed to Seattle for Kurt Cobain's memorial, and they help a circus sideshow family. Why is Pilot so set on Seattle, will the goons catch Jack, and is there any way the friends' competing needs can be resolved? 64 5.7 3.642

Use case 4: Query for finding revenue of the movie Hero

- SELECT title, revenue from Producer where title="Hero";
- ⇨ Hero 177394432

Use case 5: Query for finding the movies and their revenue for movies whose revenue exceeds 500000000

- SELECT title, revenue from Producer where revenue>500000000;

⇨ Harry Potter and the Chamber of Secrets 876688482

The Lord of the Rings: The Two Towers 926287400

Spider-Man 821708551

Star Wars: Episode II - Attack of the Clones 649398328

Cinderella 543514353

Use case 6: Query for finding twitter user with maximum followers

- SELECT Name,MAX(Followers) from consumer;

⇨ Fortnite 7179588

Use case 7: Query to get the tweets and the posted time of the tweets of the user with maximum followers

- SELECT Post,Posted_time from Consumer where Followers=(SELECT MAX(Followers) from Consumer);
  (Sample Output)

⇨ Become a pinball wizard with bumpers and flippers in #FortniteCreative! □ https://t.co/6vHBOGjaO4 Tue Mar 05 21:00:07 +0000 2019

X marks the spot! Downtime begins tomorrow, March 6th at 4 AM ET (0900 UTC). Tue Mar 05 20:01:00 +0000 2019

Catch some air and find some coin in Dolphindom's Driftin' Dolphins Skate Park! #FortniteCreative… https://t.co/A1cuEP4HLn Tue Mar 05 19:00:25 +0000 2019

Love is war. Hearts will break in Save the World's Love Storm Event! https://t.co/uA7Qd5Z7Gn Tue Mar 05 17:00:54 +0000 2019

Coming soon! Embrace exploration and uncover Buried Treasure X marks the spot! https://t.co/xJsT48PKwG Tue Mar 05 02:00:03 +0000 2019

Use case 8: Query to find the movie with highest revenue among the movies tweeted by a twitter user

- SELECT title, revenue from Producer where title=(SELECT title from Movie_tweet where Name="Complex");

⇨ Harry Potter and the Chamber of Secrets 876688482

Use case 9: Query to get the movie overview, release date and name of the twitter user for movies that the users have tweeted about

- SELECT company.Overview, company.Release_date, Movie_tweet.Name FROM company INNER JOIN Movie_tweet on Movie_tweet.title = company.title;

⇨ Ignoring threats to his life, Harry returns to Hogwarts to investigate – aided by Ron and Hermione – a mysterious series of attacks. 2002-11-13 CallMeKevin1811

Ignoring threats to his life, Harry returns to Hogwarts to investigate – aided by Ron and Hermione – a mysterious series of attacks. 2002-11-13 R3xoss

Ignoring threats to his life, Harry returns to Hogwarts to investigate – aided by Ron and Hermione – a mysterious series of attacks. 2002-11-13 FortniteGame

Ignoring threats to his life, Harry returns to Hogwarts to investigate – aided by Ron and Hermione – a mysterious series of attacks. 2002-11-13 Complex

Ignoring threats to his life, Harry returns to Hogwarts to investigate – aided by Ron and Hermione – a mysterious series of attacks. 2002-11-13 Keggerade

Frodo and Sam are trekking to Mordor to destroy the One Ring of Power while Gimli, Legolas and Aragorn search for the orc-captured Merry and Pippin. All along, nefarious wizard Saruman awaits the Fellowship members at the Orthanc Tower in Isengard. 2002-12-18 BertShowBert

Frodo and Sam are trekking to Mordor to destroy the One Ring of Power while Gimli, Legolas and Aragorn search for the orc-captured Merry and Pippin. All along, nefarious wizard Saruman awaits the Fellowship members at the Orthanc Tower in Isengard. 2002-12-18 sadrianom

Frodo and Sam are trekking to Mordor to destroy the One Ring of Power while Gimli, Legolas and Aragorn search for the orc-captured Merry and Pippin. All along, nefarious wizard Saruman awaits the Fellowship members at the Orthanc Tower in Isengard. 2002-12-18 Weinbach

Frodo and Sam are trekking to Mordor to destroy the One Ring of Power while Gimli, Legolas and Aragorn search for the orc-captured Merry and Pippin. All along, nefarious

wizard Saruman awaits the Fellowship members at the Orthanc Tower in Isengard. 2002-12-18 AlphaOmegaSin

Frodo and Sam are trekking to Mordor to destroy the One Ring of Power while Gimli, Legolas and Aragorn search for the orc-captured Merry and Pippin. All along, nefarious wizard Saruman awaits the Fellowship members at the Orthanc Tower in Isengard. 2002-12-18 _Rico21

Use case 10: Query for the twitter user who has tweeted a particular post

- SELECT Name from Consumer where Post= "We've re-enabled text chat for Xbox players.";

⇨ Fortnite

Use case 11: Query for title, overview, release date of the movie the twitter user complex has tweeted about

- SELECT title, Overview, Release_date from company where title=(SELECT title from Movie_tweet where Name="Complex"));

⇨ Harry Potter and the Chamber of Secrets Ignoring threats to his life, Harry returns to Hogwarts to investigate – aided by Ron and Hermione – a mysterious series of attacks. 2002-11-13