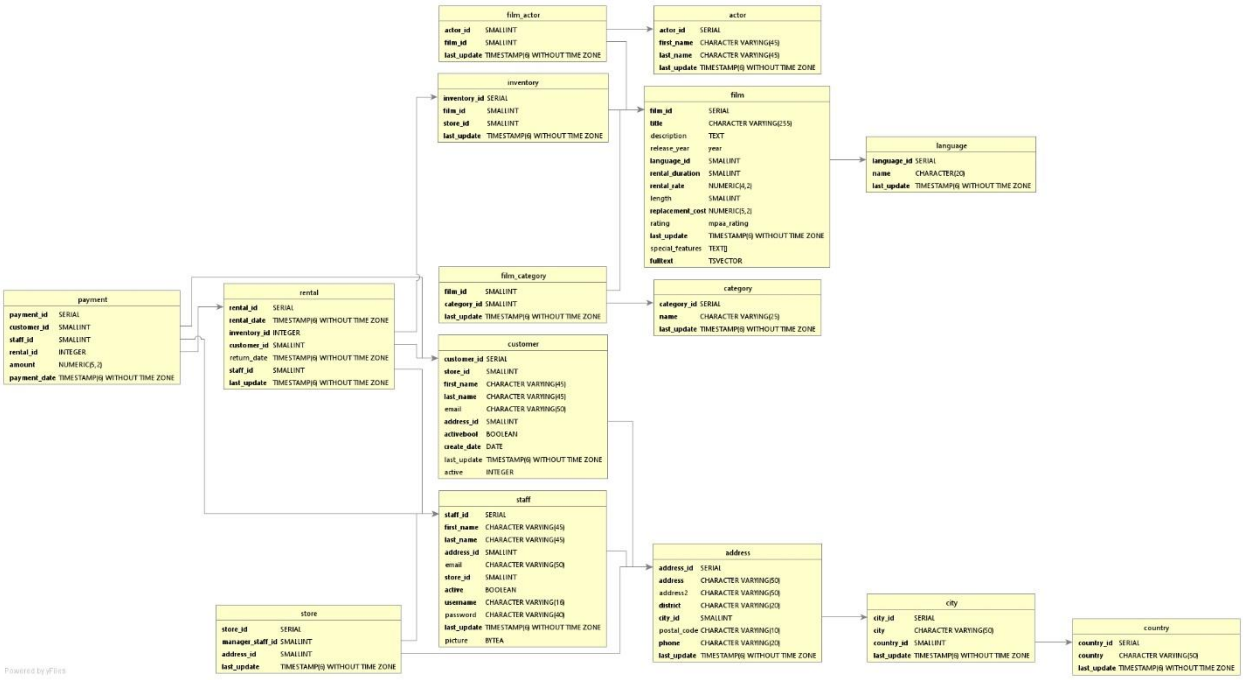
William Su

Task 3.2 – Data Storage and Structure



- Take a moment to examine your ERD. Does the Rockbuster database have a snowflake schema or a star schema? Write a brief explanation for your answer.
    - A snowflake schema because of how there are be sub-dimension tables. For example, in the ‘address’ table, there is a ‘city’ sub-dimension table and then a ‘country’ sub-dimension table.
- List all the fact tables and all the dimension tables in the schema. For each table, list every column and its data type, and write a brief description of the column. To get an idea of what this should look like, check out these example fact and dimension tables.
- If a column name doesn't tell you enough to write a description, you can also view the tables in pgAdmin 4. The SQL syntax for selecting a table is `SELECT * FROM table_name`. So `SELECT * FROM film` would return the film table, for example.

Fact Table

Rental

Columns	Data Type	Description
Rental_id	SERIAL	The unique ID assigned to each rental
Rental_date	TIMESTAMP (6) WITHOUT TIME ZONE	The time and date for when the rental occurred
Inventory_id	INTEGER	ID number of the product being rented
Customer_id	SMALLINT	ID number of the customer
Return_date	TIMESTAMP (6) WITHOUT TIME ZONE	The time and date for when the rental is returned.
Staff_id	SMALLINT	ID number of the staff who processed the rental/return
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	The time of the last update of the data entry

Dimension Table

Payment

Columns	Data Type	Description
Payment_id	SERIAL	The unique ID assigned to the transaction
Customer_id	SMALLINT	ID number of the customer
Staff_id	SMALLINT	ID number of the staff who processed the transaction
Rental_id	INTEGER	The unique ID assigned to each rental
Amount	NUMERIC (5,2)	The amount paid by the customer
Payment_date	TIMESTAMP (6) WITHOUT TIME ZONE	The time of the transaction

Film Actor

Columns	Data Type	Description
Actor_id	SMALLINT	ID of the actor
Film_id	SMALLINT	ID of the film
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of the last updated status

Actor

Columns	Data Type	Description
Actor_id	SERIAL	Unique ID number for each actor
First_name	CHARACTER VARYING (45)	First name of the Actor limited to 45 characters
Last_name	CHARACTER VARYING (45)	Last name of the Actor limited to 45 characters
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of the last update

Inventory

Columns	Data Type	Description
Inventory_id	SERIAL	Unique ID of each copy of the film
Film_id	SMALLINT	ID number of the film
Store_id	SMALLINT	ID number of the store that has the film
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of last update

Film

Columns	Data Type	Description
Film_id	SERIAL	Unique ID of the film
Title	CHARACTER VARYING (255)	Title of the film limited to 255 characters
Description	TEXT	Description of the film
Release_year	YEAR	Year film was released
Language_id	SMALLINT	ID of film language
Rental_duration	SMALLINT	How long the film was rented for
Rental_rate	NUMERIC (4,2)	How much the film was rented for per day
Length	SMALLINT	Total length of film
Replacement_cost	NUMERIC (5,2)	Cost to replace film
Rating	Mpaa_rating	MPAA Rating of film
Last_update	TIMESTAMP (6) WITHOUT TIMEZONE	Time of last update
Special features	TEXT	Special features available in the film
fulltext	TSVECTOR	

Language

Columns	Data Type	Description
Language_id	SERIAL	Unique ID of the language
Name	CHARACTERS (20)	Name of the language
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of last update

Film Category

Columns	Data Type	Description
Film_id	SMALLINT	ID number of the film
Category_id	SMALLINT	ID number of the category
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of last update

Category

Columns	Data Type	Description
Category_id	SERIAL	Unique ID of the film category
Name	CHARACTER VARYING (25)	Name of the film category
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of last update

Customer

Columns	Data Type	Description
Customer_ID	SERIAL	Unique ID of the customer
Store_id	SMALLINT	ID number of the store
First_name	CHARACTERS VARYING (45)	First name of the customer
Last_name	CHARACTERS VARYING (45)	Last name of the customer
Email	CHARACTER VARYING (50)	E-mail address of the customer
Address_id	SMALLINT	ID number of the customer's address
activebool	BOOLEAN	Customer is active or inactive
Create_date	DATE	The date where the customer was added to the database
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of last update
Active	INTEGER	Customer is active or inactive

Store

Columns	Data Type	Description
Store_id	SERIAL	Unique ID of the store
Manager_staff_id	SMALLINT	ID of the manager of the store
Address_id	SMALLINT	ID of the store's address
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of last update

Staff

Columns	Data Type	Description
Staff_id	SERIAL	Unique ID of the staff member
First_name	CHARACTER VARYING (45)	First name of the staff member
Last_name	CHARACTER VARYING (45)	Last name of the staff member
Address_id	SMALLINT	ID of their address
Email	CHARACTER VARYING (50)	E-Mail address of the staff member
Store_id	SMALLINT	ID of the store they work at
Active	BOOLEAN	Staff is active or inactive
Username	CHARACTER VARYING (45)	Username of the Staff member
Password	CHARACTER VARYING (45)	Password of the staff member
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of last update
picture	BYTEA	Picture of staff member

Address

Columns	Data Type	Description
Address_ID	SERIAL	Unique ID of the address
Address	CHARACTER VARYING (50)	First line of address
Address2	CHARACTER VARYING (50)	Second line of address
District	CHARACTER VARYING (20)	District of address
City_id	SMALLINT	ID number of City
Postal_code	CHARACTER VARYING (10)	Postal Code
phone	CHARACTER VARYING (20)	Phone number
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of last update

City

Columns	Data Type	Description
City_id	SERIAL	Unique ID number of city
City	CHARACTER VARYING (50)	City name
Country_id	SMALLINT	ID number of country
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	Time of last update

Country

Columns	Data Type	Description
Country_id	SERIAL	Unique ID number of country
country	CHARACTER VARYING (50)	Country name
Last_update	TIMESTAMP (6) WITHOUT TIME ZONE	The time of the last update of the data entry

Now that your data dictionary and ERD are ready to use, your manager has given you a list of business questions to answer. Use your data dictionary to figure out which tables you'd need to answer the questions below:

- Which actors brought Rockbuster the most revenue?
    - Film Actor, Film, and Rental.
- What language are the majority of movies in the collection?
    - Film and Language