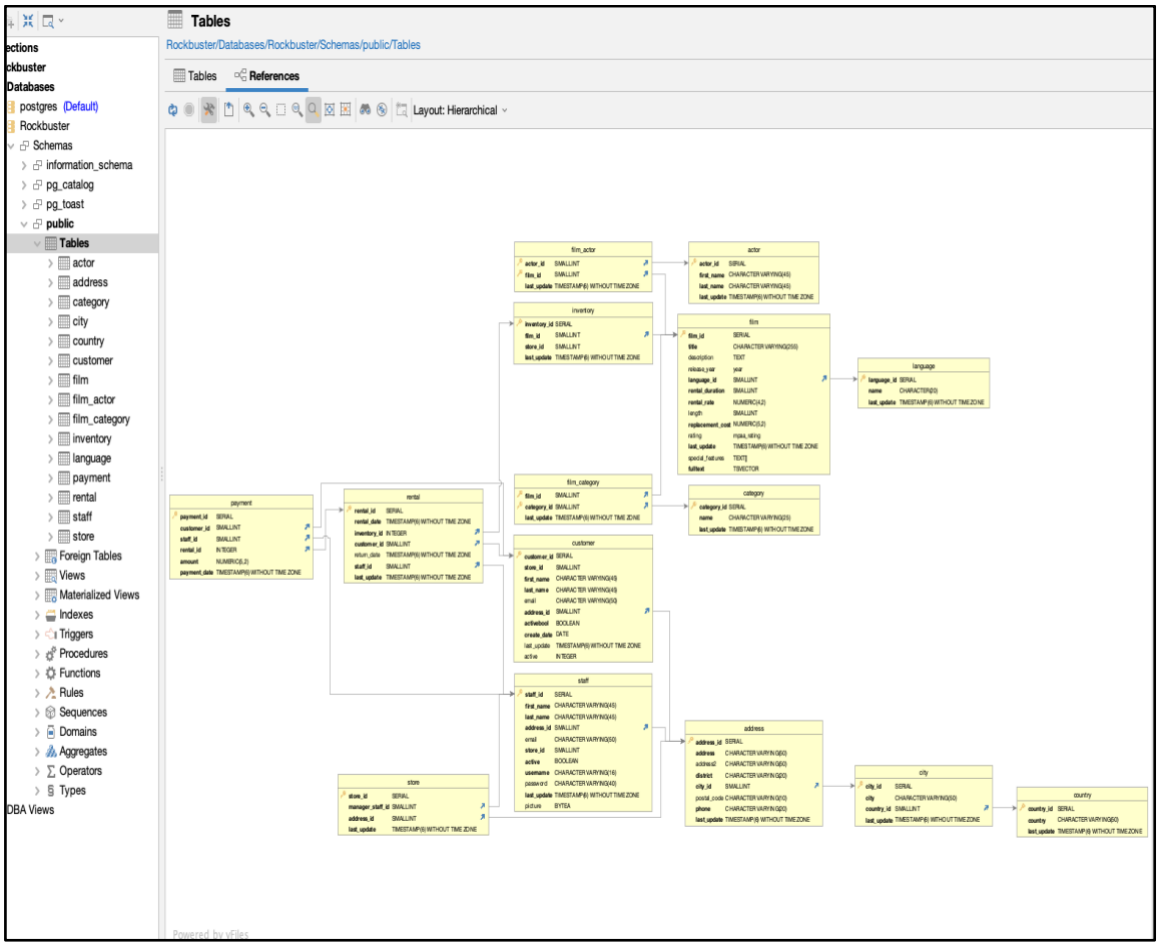
3.2 Data Storage and Structure

Step 2. Extract the ERD

- Download and install DbVisualizer (if you haven't already done so).
- Extract the ERD from the Rockbuster database and save it as an image (PNG or JPEG) using the instructions in the Exercise.
- Copy-paste the ERD into your answers document.



Step 3. Create the first draft of a data dictionary

- 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.

  The ERD has a Snowflake Schema. It is represented by a centralized fact tables which are connected to multiple dimensions and sub-dimensions tables

- 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.

Fact Table

Rental

Column	Data Type	Description
rental_id	Serial	Number assigned to rental
rental_date	TIMESTAMP(6) WITHOUT TIMEZONE	Date of rental
inventory_id	INTEGER	Number assigned to item in the table
customer_id	SMALLINT	Number assigned to customer
return_date	TIMESTAMP(6) WITHOUT TIMEZONE	Date rental was returned
staff_id	SMALLINT	Number assigned to employee
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

Dimension Tables

Payment

Column	Data Type	Description
payment_id	SERIAL	Number assigned to payment
customer_id	SMALLINT	Number assigned to customer
staff_id	SMALLINT	Number assigned to employee
rental_id	INTEGER	Number assigned to rental
amount	NUMERIC (5,2)	Amount paid
payment_date	TIMESTAMP(6) WITHOUT TIMEZONE	Date of payment

store

Column	Data Type	Description
store_id	SERIAL	Number assigned to store
manager_staff_id	SMALLINT	Number assigned to store manager
address_id	SMALLINT	Number assigned to store address
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

film_actor

Column	Data Type	Description
actor_id	SMALLINT	Number assigned to actor
film_id	SMALLINT	Number assigned to film
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

inventory

Column	Data Type	Description
inventory_id	SERIAL	Number assigned to item
film_id	SMALLINT	Number assigned to film
store_id	SMALLINT	Number assigned to store
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

film_category

Column	Data Type	Description
film_id	SMALLINT	Number assigned to film
film_category	SMALLINT	Number assigned to genre/category
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

customer

Column	Data Type	Description
customer_id	SERIAL	Number assigned to customer
store_id	SMALLINT	Number assigned to store
first_name	CHARACTERVARYING(50)	First name of customer
last_name	CHARACTERVARYING(45)	Last name of customer
email	CHARACTERVARYING(50)	Email address of customer
address_id	SMALLINT	Number assigned to customer's address
activebool	BOOLEAN	Customer's active status
create_date	DATE	Date entry was created
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated
active	INTGER	Customer's active status

Staff

Column	Data Type	Description
staff_id	SERIAL	Number assigned to employee
first_name	CHARACTERVARYING(45)	First name of employee
last_name	CHARACTERVARYING(45)	Last name of employee
address_id	SMALLINT	Number assigned to employee's address
email	CHARACTERVARYING(50)	Email address of employee
store_id	SMALLINT	Number assigned to store
active	BOOLEAN	Employee active status
username	CHARACTERVARYING(16)	Username of employee
password	CHARACTERVARYING(40)	Password of employee
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated
picture	BYTEA	Picture of employee

actor

Column	Data Type	Description
actor_id	SERIAL	Number assigned to actor
first_name	CHARACTERVARYING(45)	First name of actor
last_name	CHARACTERVARYING(45)	Last name of actor
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

film

Column	Data Type	Description
film_id	SERIAL	Number assigned to film
title	CHARACTERVARYING(45)	Title of film
description	TEXT	Description of film
release_year	year	Release year of film
language_id	SMALLINT	Number assigned to film language
rental_duration	SMALLINT	Length of film rental
rental_rate	NUMERIC(4,2)	Price of film
length	SMALLINT	Length of film
replacement_cost	NUMERIC(5,2)	Cost of replace of film
rating	mpaa_rating	Film rating
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated
special_features	TEXT []	Special features included with film
fulltext	TSVECTOR	Keywords associated with film

category

Column	Data Type	Description
category_id	SERIAL	Number assigned to film category
name	CHARACTERVARYING(20)	Name of language
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

Address

Column	Data Type	Description
address_id	SERIAL	Number assigned to address
address	CHARACTERVARYING(50)	Street address
address2	CHARACTERVARYING(50)	Supplementary street address
district	CHARACTERVARYING(20)	District
city_id	SMALLINT	Number assigned to City
postal_code	CHARACTERVARYING(10)	Postal code
phone	CHARACTERVARYING(20)	Phone number
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

language

Column	Data Type	Description
language_id	SERIAL	Number assigned to language
name	CHARACTERVARYING(20)	Name of language
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

city

Column	Data Type	Description
city_id	SERIAL	Number assigned to city
city	CHARACTERVARYING(20)	Name of city
country_id	SMALLINT	Number assigned to country
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

country

Column	Data Type	Description
country_id	SERIAL	Number assigned to country
country	CHARACTERVARYING(50)	Name of country
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

Step 4. Use your data dictionary to figure out which tables you'd need to answer the questions below:

- Which actors brought Rockbuster the most revenue?

  I would need to query the “actor table” (for the name of the actors), “film_actor table” (for films associated with each actor) and the “film table” (for the rental rates for each film)

- What language are the majority of movies in the collection?

  I would need to query the “language table” (for the name of each language), then the “film table” (for the language associated with each film)