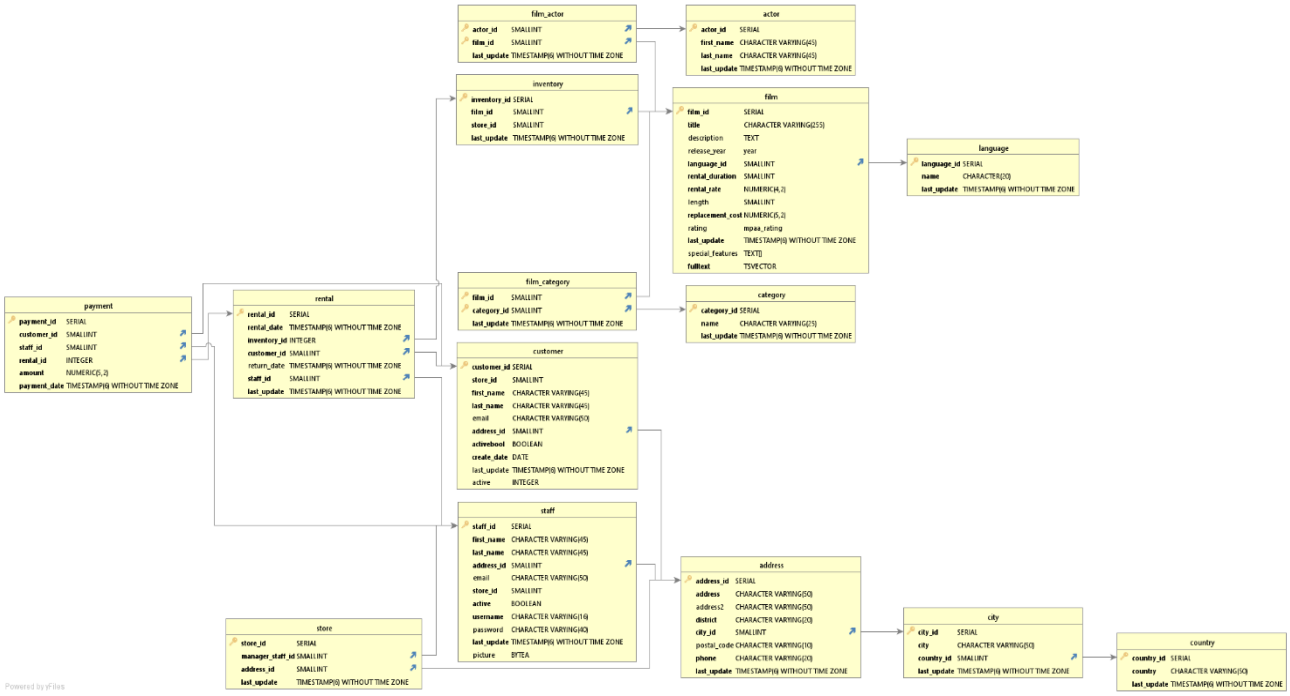
3.2 Data Storage & Structure

1. **Step 1 Create your Answers document:**
   Create a new text document and call it “Answers 3.2.” You'll save a copy of your ERD, data dictionary, and written answers in this document.
2. **Step 2 Extract the ERD:**
   - Download and install DbVisualizer.
   - 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.



3. **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?
     Snowflake schema because the arrangement is multidimensional; One fact table surrounded by dimension table which are in turn surrounded by 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.
     Green is fact table and blue are dimension tables.

rental - fact table		
Columns	**Data Type**	**Description**
rental_id	SERIAL	Identification number assigned to each rental order
rental_date	TIMESTAMP(6) WIHOUT TIME ZONE	Date of rental
inventory_id	INTEGER	Inventory identification number
customer_id	SMALLINT	Customer identification number
return_date	TIMESTAMP(6) WIHOUT TIME ZONE	Date of return
staff_id	SMALLINT	Staff identification number processing rental order
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

payment - dimension table		
Columns	**Data Type**	**Description**
payment_id	SERIAL	Identification number assigned to each payment
customer_id	SMALLINT	Customer identification number
staff_id	SMALLINT	Staff identification number processing payment
rental_id	INTEGER	Identification number assigned to each rental order
amount	NUMERIC (5,2)	Rental price
payment_date	TIMESTAMP(6) WIHOUT TIME ZONE	Date and time when payment made

film_category - dimension table		
Columns	**Data Type**	**Description**
film_id	SERIAL	Identification number assigned to each film
category_id	SMALLINT	Film category (comedy, action, etc.)
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

category - dimension table		
Columns	**Data Type**	**Description**
category_id	SERIAL	Identification number assigned to each film catrgory
name	CHARACTER VARYING(25)	Film category (comedy, action, etc.)
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

fim_actor - dimension table		
Columns	**Data Type**	**Description**
actor_id	SMALLINT	Identification number assigned to each film actor
film_id	SMALLINT	Identification number assigned to each film
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Staff identification number processing payment

inventory - dimension table		
Columns	**Data Type**	**Description**
inventory_id	SERIAL	Inventory identification number
film_id	SMALLINT	Identification number assigned to each film
store_id	SMALLINT	Identification number assigned to each store
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

actor - dimension table		
Columns	**Data Type**	**Description**
actor_id	SERIAL	Identification number assigned to each film actor
first_name	CHARACTER VARYING(45)	First name of the film actor
last_name	CHARACTER VARYING(45)	Last name of the film actor
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

film - dimension table		
Columns	**Data Type**	**Description**
film_id	SERIAL	Identification number assigned to each film
title	CHARACTER VARYING(25)	Film title
description	TEXT	Short description of the film story (short summary)
release_year	year	Release year of the film
languange_id	SMALLINT	Identification number assigned to each language
rental_duration	SMALLINT	Duration of the film rental
rental_rate	NUMERIC(4,2)	Total number of a film ever being rented
length	SMALLINT	Length of film (in minutes)
replacement_cost	NUMERIC(5,2)	Cost to re-order film in case defect or missing
rating	mpaa_rating	Film rating
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry
special_feature	TEXT	Special feature included in the film (sign language is included, etc.)
fulltext	TSVECTOR	

language - dimension table		
Columns	**Data Type**	**Description**
languange_id	SERIAL	Identification number assigned to each language
name	CHARACTER(20)	Languane name
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

customer - dimension table		
Columns	**Data Type**	**Description**
customer_id	SERIAL	Customer identification number
store_id	SMALLINT	Identification number assigned to each store
first_name	CHARACTER VARYING(45)	First name of the customer
last_name	CHARACTER VARYING(45)	Last name of the customer
email	CHARACTER VARYING(50)	Customer email address
address_id	SMALLINT	Customer address
activebool	BOOLEAN	Whether or not this customer is active member
create_date	DATE	Creation date of membership
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry
active	INETEGER	Membership status (silver, gold, etc.)

staff - dimension table		
Columns	**Data Type**	**Description**
staff_id	SERIAL	Staff identification number processing rental order
first_name	CHARACTER VARYING(45)	First name of the staff
last_name	CHARACTER VARYING(45)	Last name of the staff
address_id	SMALLINT	Staff address
email	CHARACTER VARYING(50)	Staff email address
store_id	SMALLINT	Identification number assigned to each store
active	BOOLEAN	Whether or not this staff is active employee
username	CHARACTER VARYING(16)	Staff username
password	CHARACTER VARYING(40)	Staff account's password
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry
picture	BYTEA	Staff password

store - dimension table		
Columns	**Data Type**	**Description**
store_id	SERIAL	Identification number assigned to each store
manager_staff_id	SMALLINT	Identification number assigned to each store manager
address_id	SMALLINT	Identification number assigned to store location
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

address - dimension table		
Columns	**Data Type**	**Description**
address_id	SERIAL	Identification number assigned to store location
address	CHARACTER VARYING(50)	Store address
address2	CHARACTER VARYING(50)	Supplementary store address information
district	CHARACTER VARYING(20)	Store district
city_id	SMALLINT	City name where store is located
postal_code	CHARACTER VARYING(10)	Postal code of city where store is located
phone	CHARACTER VARYING(20)	Store phone
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

city - dimension table		
Columns	**Data Type**	**Description**
city_id	CHARACTER VARYING(50)	City identification numbe where store is located
country_id	SMALLINT	Country identification number where store is located
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

country - dimension table		
Columns	**Data Type**	**Description**
country_id	SERIAL	Country identification number where store is located
country	CHARACTER VARYING(50)	Country name where store is located
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

- 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

4. **Step 4 Find information**

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?
  The most revenue of a film is given by rental rate located in the film table. To get connection to film table associated with the actor, we need 2 tables; namely film_actor table and actor table.
- What language are the majority of movies in the collection?
  To query all film associated with languages, we need 2 tables: film table and language table.