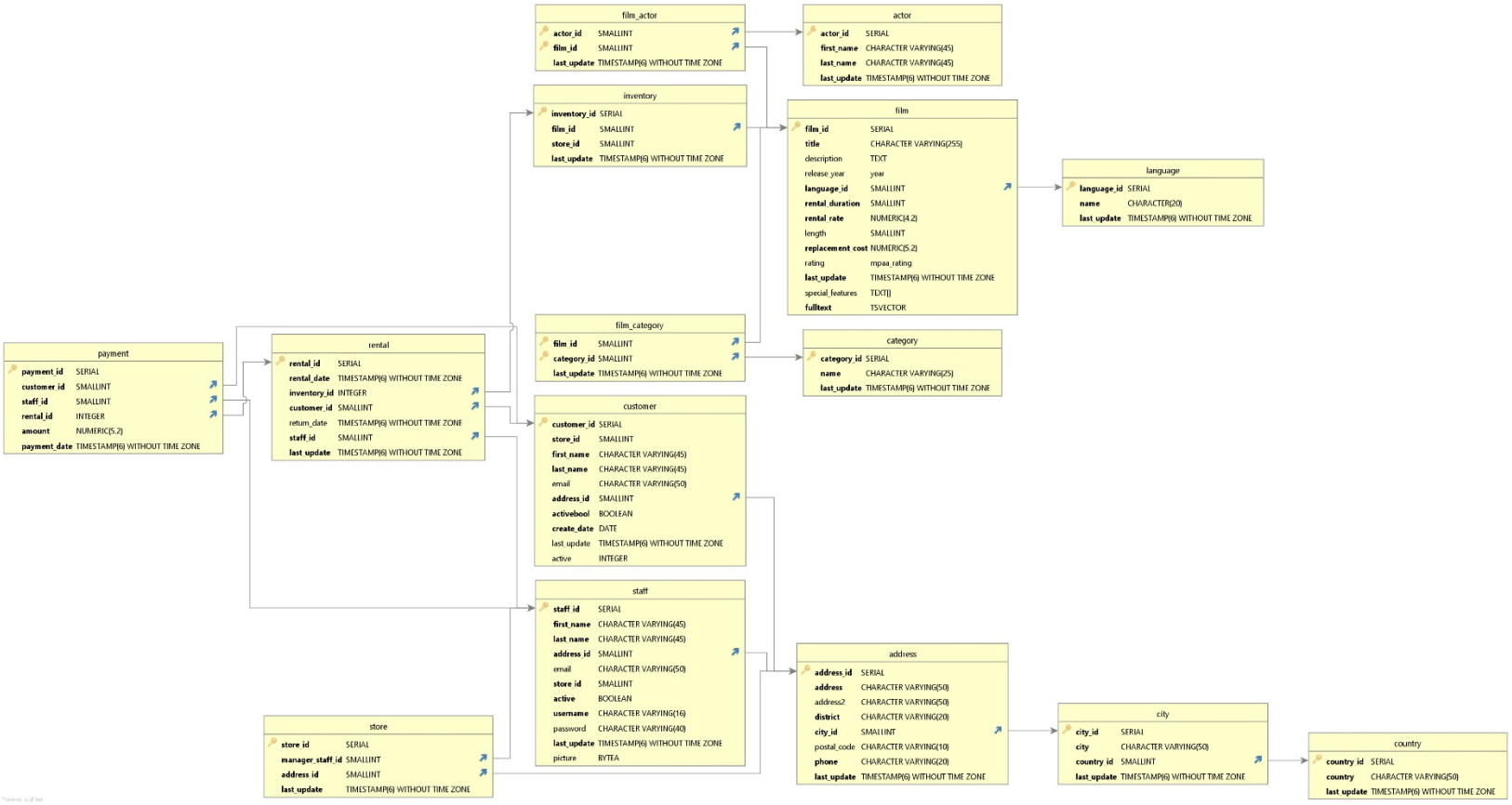
3.2 Data Storage & Structure Task

Step 1. Create your Answers document:

Step 2. Extract the ERD:



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.
    - This is a snowflake schema because there are sub dimension tables connected to the fact 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. 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		
Column	**Data Type**	**Description**
rental_id	SERIAL	Unique ID assigned to each rental order
rental_date	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of rental
inventory_id	INTEGER	Inventory number
customer_id	SMALLINT	Customer identification number
return_date	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of rental return
staff_id	SMALLINT	Staff identification number
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: Payment		
Column	**Data Type**	**Description**
payment_id	SERIAL	Unique ID assigned to each payment
customer_id	SMALLINT	Customer identification number
staff_id	SMALLINT	Staff identification number
rental_id	INTEGER	Unique ID assigned to each rental order
amount	NUMERIC(5,2)	Amount paid during transaction
payment_date	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of payment

Dimension Table: Store		
Column	**Data Type**	**Description**
store_id	SERIAL	Unique ID assigned to each store
manager_staff_id	SMALLINT	Manager identification number
address_id	SMALLINT	Address of Store
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: film_actor		
Column	**Data Type**	**Description**
actor_id	SMALLINT	Unique ID assigned to each actor
film_id	SMALLINT	Film identification number
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: inventory		
Column	**Data Type**	**Description**
inventory_id	SERIAL	Unique ID assigned to each inventory item
film_id	SMALLINT	Film identification number
store_id	SMALLINT	Unique ID assigned to each store
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: film_category		
Column	**Data Type**	**Description**
film_id	SMALLINT	Film identification number
category_id	SMALLINT	Category identification number
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: customer		
Column	**Data Type**	**Description**
customer_id	SERIAL	Unique ID assigned to each customer
store_id	SMALLINT	Unique ID assigned to each store
first_name	CHARACTER VARYING(45)	Customer first name
last_name	CHARACTER VARYING(45)	Customer last name
email	CHARACTER VARYING(50)	Customer email address
address_id	SMALLINT	Customer address
activebool	BOOLEAN	Status of customer: active or inactive
create_date	DATE	Date customer record created
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update
active	INTEGER	Status of customer

Dimension Table: staff		
Column	**Data Type**	**Description**
staff_id	SERIAL	Unique ID assigned to each staff member
first_name	CHARACTER VARYING(45)	Staff first name
last_name	CHARACTER VARYING(45)	Staff last name
address_id	SMALLINT	Staff address
email	CHARACTER VARYING(50)	Staff email address
store_id	SMALLINT	Unique ID assigned to each store
active	BOOLEAN	Status ofstaff: active or inactive
username	CHARACTER VARYING(16)	Staff username
password	CHARACTER VARYING(40)	Staff password
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update
picture	INTEGER	Picture of staff

Dimension Table: actor		
Column	**Data Type**	**Description**
actor_id	SERIAL	Unique ID assigned to each actor
first_name	CHARACTER VARYING(45)	Actor first name
last_name	CHARACTER VARYING(45)	Actor last name
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: film		
Column	**Data Type**	**Description**
film_id	SERIAL	Unique ID assigned to each film
title	CHARACTER VARYING(255)	Title of film
description	TEXT	Description of film
release_year	year	Film release year
language_id	SMALLINT	Film language
rental_duration	SMALLINT	Duration of rental
rental_rate	NUMERIC(4,2)	Price of rental
length	SMALLINT	Duration of film
replacement_cost	NUMERIC(5,2)	Amount needed to replace the rental
rating	mpaa_rating	Film rating
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update
special_features	TEXT[]	Description of films special features, if any
fulltext	TSVECTOR	Text

Dimension Table: category		
Column	**Data Type**	**Description**
category_id	SERIAL	Unique ID assigned to each category
name	CHARACTER VARYING(25)	Category name
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: language		
Column	**Data Type**	**Description**
language_id	SERIAL	Unique ID assigned to each language
name	CHARACTER VARYING(20)	Language name
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: address		
Column	**Data Type**	**Description**
address_id	SERIAL	Unique ID assigned to each store address
address	CHARACTER VARYING(50)	Store address first line
address2	CHARACTER VARYING(50)	Store address second line
district	CHARACTER VARYING(20)	Address district
city_id	SMALLINT	Unique ID assigned to each city
postal_code	CHARACTER VARYING(10)	Store address postal code
phone	CHARACTER VARYING(20)	Store phone number
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: city		
Column	**Data Type**	**Description**
city_id	SERIAL	Unique ID assigned to each city
city	CHARACTER VARYING(45)	City name
county_id	SMALLINT	Unique ID assigned to each county
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

Dimension Table: county		
Column	**Data Type**	**Description**
county_id	SERIAL	Unique ID assigned to each county
county	CHARACTER VARYING(50)	County name
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Time and date of last update

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?
    - Film, film_actor, and actor table
- What language are the majority of movies in the collection?
    - Film, and language table

Step 5. Save and upload your file: