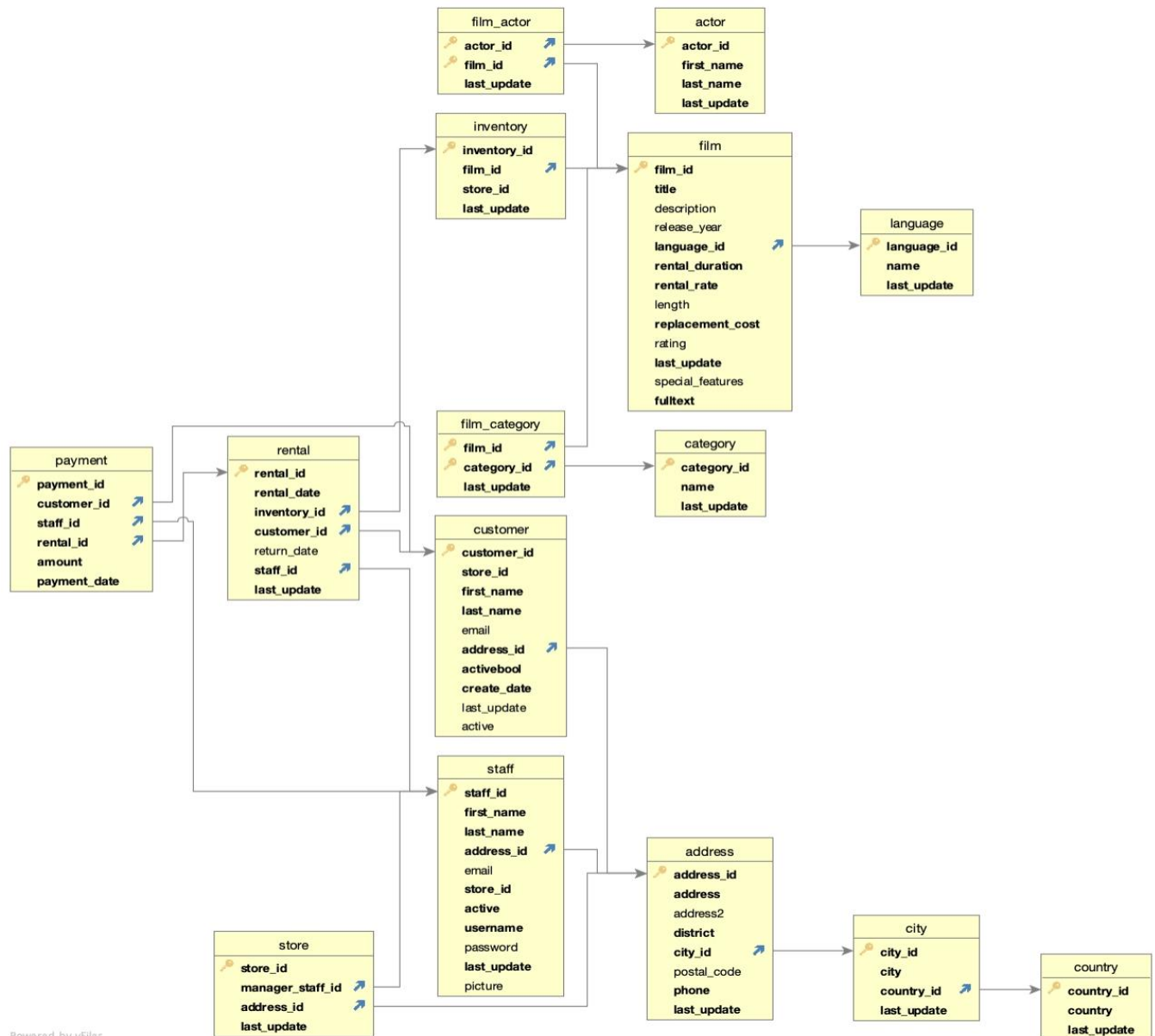# Answers 3.2 Data Storage & Structure

Step 2. Extract the ERD:

- Download and install [DbVisualizer](#) or [Lucidchart](#) (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.

It is a snowflake schema, as there is no central part in which all segments are connected. The fact table have multiple branches and are connected to multiple tables, creating subdimensions, typical to snowflake.

- 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.

Rental – FACT TABLE

Column	Data Type	Description
rental_id	SERIAL	Identification number assigned to each rental order
rental_date	TIMESTAMP(6) WITHOUT TIMEZONE	Date of rental
inventory_id	INTEGER	Inventory identification number
customer_id	SMALLINT	Customer identification number
rentum_date	TIMESTAMP(6) WITHOUT TIMEZONE	Date of returned
staff_id	SMALLINT	Staff identification number processing rental order
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Last update of the data entry

Payment Dimension table

Column	Data Type	Description
payment_id	SERIAL	Serial number assigned to rental
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 Dimension table

Column	Data Type	Description
store_id	SERIAL	Serial number assigned to rental
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 Dimension table

Column	Data Type	Description
actor_id	SMALLINT	Serial number assigned to rental
film_id	SMALLINT	Number assigned to film
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Last update of the data entry

Inventory Dimension table

Column	Data Type	Description
inventory_id	SERIAL	Serial number assigned to rental
film_id	SMALLINT	Number assigned to film
store_id	SMALLINT	Number assigned to store
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Last update of the data entry

Film category Dimension table

Column	Data Type	Description
film_id	SMALLINT	Identification number assigned to each film
film_category	SMALLINT	Film category (comedy, action, etc.)
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	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

Columns	Data Type	Description
last_update	TIMESTAMP(6) WIHOUT TIME ZONE	Last update of the data entry

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

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

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	

Category Dimension table

Columns	Data Type	Description
category_id	SERIAL	Identification number assigned to each film category
name	CHARACTER VARYING(25)	Film category (comedy, action, etc.)
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

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

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

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 - film_actor table and actor table.

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

  film and Language