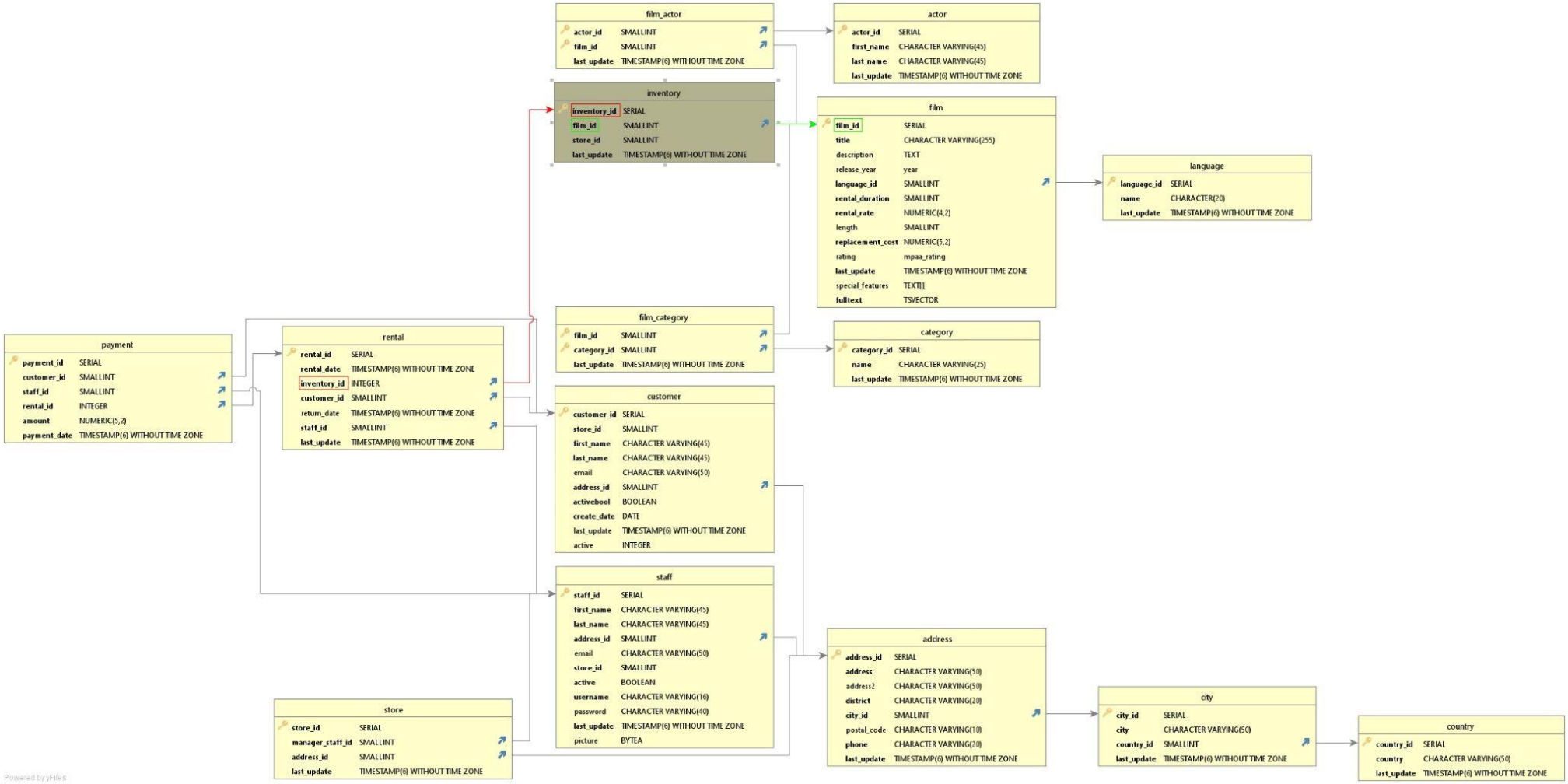
3.2: Data Storage & Structure

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.**
    - The Rockbuster database has a snowflake schema as the fact table (inventory) is in the center surrounded by dimension tables but, unlike the star schema, they're surrounded by sub-dimension 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 Tables: rental**

Columns	Data type	Description
rental_id	SERIAL	rental id number
rental_date	TIMESTAMP(6) WITHOUT TIME ZONE	payment date (yyyy-mm-dd) without time zone
inventory_id	INTEGER	inventory id number
customer_id	SMALLINT	customer id number
return_date	TIMESTAMP(6) WITHOUT TIME ZONE	return date (yyyy-mm-dd) without time zone
staff_if	SMALLINT	staff id number
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: payment**

Columns	Data type	Description
payment_id	SERIAL	payment id number
customer_id	SMALLINT	customer id number
staff_if	SMALLINT	staff id number
rental_id	INTEGER	rental id number
amount	NUMERIC (5,2)	the amount paid in currency
payment_date	TIMESTAMP(6) WITHOUT TIME ZONE	payment date (yyyy-mm-dd) without time zone

- **Dimension Tables: customer**

Columns	Data type	Description
customer_id	SERIAL	customer id number
store_id	SMALLINT	store id number
first_name	CHARACTER VARYING (45)	first name
last_name	CHARACTER VARYING (45)	last name

email	CHARACTER VARYING (50)	email address
address_id	SMALLINT	address id number
activebool	BOOLEAN	active bool, can be true or false
create_date	DATE	create date (yyyy-mm-dd) without time zone
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone
active	INTEGER	active status

- **Dimension Tables: staff**

Columns	**Data type**	**Description**
staff_id	SERIAL	staff id number
first_name	CHARACTER VARYING (45)	first name
last_name	CHARACTER VARYING (45)	last name
address_id	SMALLINT	address id number
email	CHARACTER VARYING (50)	email address
store_id	SMALLINT	store id number
active	BOOLEAN	active status, can be true or false

username	CHARACTER VARYING (16)	username
password	CHARACTER VARYING (40)	password
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone
picture	BYTEA	staff picture

- **Dimension Tables: store**

Columns	Data type	Description
store_id	SERIAL	store id number
manager_staff_id	SMALLINT	manager staff id number
address_id	SMALLINT	address id number
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: address**

Columns	Data type	Description
address_id	SERIAL	address id number
address	CHARACTER VARYING (50)	address
address 2	CHARACTER VARYING (50)	second address
district	CHARACTER VARYING (20)	district
city_id	SMALLINT	city id number
postal_code	CHARACTER VARYING (10)	postal code
phone	CHARACTER VARYING (20)	phone number
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: city**

Columns	Data type	Description
city_id	SERIAL	city id number
city	CHARACTER VARYING (50)	city

country_id	SMALLINT	country id number
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: country**

Columns	Data type	Description
country_id	SERIAL	country id number
country	CHARACTER VARYING (50)	country
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: film_category**

Columns	Data type	Description
film_id	SMALLINT	film id number
category_id	SMALLINT	category id number

last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: category**

Columns	Data type	Description
category_id	SERIAL	category id number
name	CHARACTER VARYING (25)	name
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: inventory**

Columns	Data type	Description
inventory_id	SERIAL	inventory id number
film_id	SMALLINT	film id number
store_id	SMALLINT	story id number

last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: film**

Columns	Data type	Description
film_id	SERIAL	film id number
title	CHARACTER VARYING (225)	film title
description	TEXT	film description
release_year	YEAR	release year of the movie
language_id	SMALLINT	language id number
rental_duration	SMALLINT	rental duration
rental_rate	NUMERIC(4,2)	the rate amount paid in currency
lenght	SMALLINT	length of the rental
replacement_cost	NUMERIC(5,2)	replacement cost in currency
rating	mpaa_rating	mpaa rating of the movie
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

special_features	TEXT []	special features included in the rental
fulltext	TSVECTOR	text

- **Dimension Tables: film_actor**

Columns	Data type	Description
actor_id	SMALLINT	actor id number
film_id	SMALLINT	name of the language
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: actor**

Columns	Data type	Description
actor_id	SERIAL	actor id number
first_name	CHARACTER VARYING (45)	actor first name

last_name	CHARACTER VARYING (45)	actor last name
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

- **Dimension Tables: language**

Columns	Data type	Description
language_id	SERIAL	language id number
name	CHARACTER (20)	name of the language
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	last update date (yyyy-mm-dd) without time zone

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?
  - We would need actor, film_actor, film, payment, rental.

- What language are the majority of movies in the collection?
    - We would need language, film, inventory.