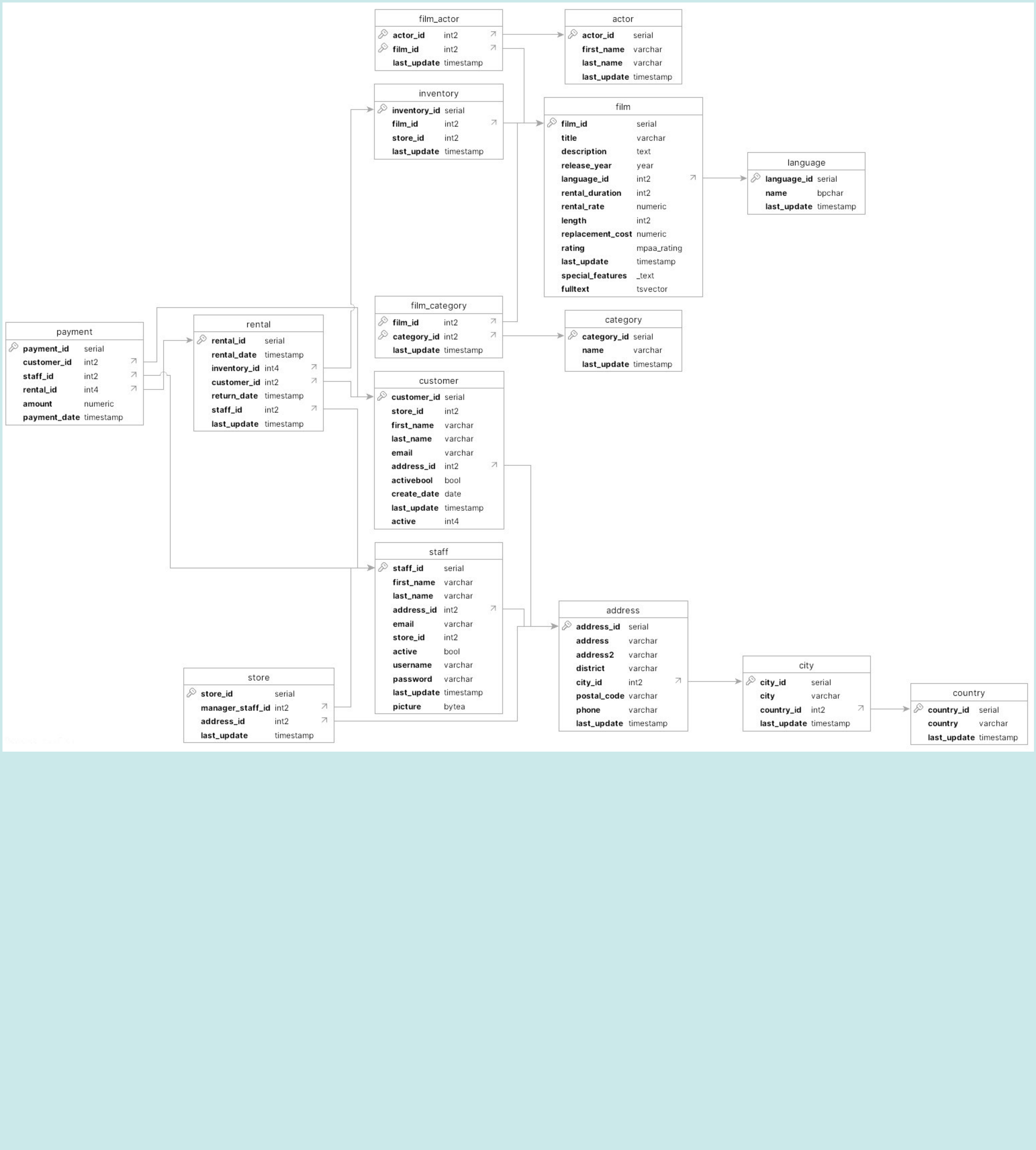
DATA STORAGE & STRUCTURE.

ENTITY RELATIONSHIP DIAGRAM (ERD)

Rockbuster Stealth

STEP 2. Entity Relational Diagram

Tool used: DbVisualizer



STEP 3. 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. By examining the **payment** table (fact table), I can see that it connects to the **staff** table, which is a dimension table. The **staff** table is further linked to sub-dimension tables, including **address**, **city**, and **country**.*
*Since this structure involves fact tables, dimension tables, and additional layers of sub-dimension tables, it is best categorized as a **snowflake schema**.*

Fact Tables
- *Payment*
- *Rental*

Dimension tables
- *Staff*
- *Customer*
- *Inventory*
- *Film actor*
- *Film category*
- *Store*

Sub-Dimension tables
- *Actor*
- *Film*
- *Language*
- *Category*
- *Address*
- *City*
- *Country*

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

payment Table

Columns	Data Type	Description
payment_id	integer	Unique identifier for the payment
customer_id int2	int2	Unique identifier for the customer
staff_id	int2	Unique identifier for the staff member
rental id	Int4	Unique identifier for the rental
amount	numeric	Payment amount
payment_date	timestamp	Payment date

rental Table

Columns	Data Type	Description
rental_id	integer	Unique identifier for the rental
rental_date	timestamp	Date of rental
inventory_id	Int4	Unique identifier for the inventory item
customer_id	int2	Unique identifier for the customer
return_date	timestamp	Date the rental was returned
staff_id	int2	Unique identifier for the staff member
last_update	timestamp	Timestamp of the last update to the record

Dimensional tables

actor Table

Columns	Data Type	Description
actor_id	Serial	Unique identifier for the actor
first_name	Varchar	First name of the individual
last_name	Varchar	Last name of the individual
last_update	timestamp	Timestamp of the last update to the record

address Table

Columns	Data Type	Description
address_id	Serial	Unique identifier for the address
address	varchar	Street address
address2	varchar	Additional address information
district	varchar	District or region of the address
city_id	int2	Unique identifier for the city
postal_code	varchar	Postal or ZIP code
phone	varchar	Phone number
last_update	timestamp	Timestamp of the last update to the record

Dimensional tables

category Table

Columns	Data Type	Description
category_id	Serial	Unique identifier for the category
name	Varchar	Name of the category
last_update	timestamp	Timestamp of the last update to the record

city Table

Columns	Data Type	Description
city_id	Serial	Unique identifier for the city
city	varchar	Name of the city
country_id	int2	Unique identifier for the country
last_update	timestamp	Timestamp of the last update to the record

Dimensional tables

country Table

Columns	Data Type	Description
country_id	Serial	Unique identifier for the country
country	varchar	Name of the country
last_update	timestamp	Timestamp of the last update to the record

customer Table

Columns	Data Type	Description
customer_id	Serial	Unique identifier for the customer
store_id	int2	ID of the store
first_name	varchar	First name of the individual
last_name	varchar	Last name of the individual
email	varchar	Email address
address_id	int2	ID of the related address
activebool	Bool	Indicates if the customer is active
create_date	date	Date the customer account was created
last_update	timestamp	Timestamp of the last update to the record
active	Int4	Indicates if the customer is active

Dimensional tables

film Table

Columns	Data Type	Description
film_id	Serial	Unique identifier for the film
title	varchar	Title of the film
description	text	Description of the film
release_year	Year	Year the film was released
language_id	int2	Unique identifier for the language
rental_duration	int2	Duration (in days) the film can be rented
rental_rate	numeric	Cost to rent the film
length	int2	Length of the film in minutes
replacement_cost	numeric	Cost to replace the film
rating	mpaa_rating	Film rating (e.g., PG, R)
last_update	timestamp	Timestamp of the last update to the record
special_features	_text	List of special features included
fulltext	tsvector	Full text search vector (PostgreSQL)

Dimensional tables

film_actor Table

Columns	Data Type	Description
actor_id	int2	Unique identifier for the actor
film_id	int2	Unique identifier for the film
last_update	timestamp	Timestamp of the last update to the record

film_category Table

Columns	Data Type	Description
film_id	int2	Unique identifier for the film
category_id	int2	Unique identifier for the category
last_update	timestamp	Timestamp of the last update to the record

inventory Table

Columns	Data Type	Description
inventory_id	Serial	Unique identifier for the inventory item
film_id	int2	Unique identifier for the film
store_id	int2	ID of the store
last_update	timestamp	Timestamp of the last update to the record

language Table

Columns	Data Type	Description
language_id	Serial	Unique identifier for the language
name	char(20)bpchar	Language name
last_update	timestamp	Timestamp of the last update to the record

Dimensional tables

staff Table

Columns	Data Type	Description
staff_id	Serial	Unique identifier for the staff member
first_name	varchar	First name of the individual
last_name	varchar	Last name of the individual
address_id	int2	ID of the related address
email	varchar	Email address
store_id	int2	ID of the store
active	Bool	Indicates if the staff member is active
username	varchar	Login username of the staff member
password	varchar	Login password of the staff member
last_update	timestamp	Timestamp of the last update to the record
picture	Bytea	Photo of staff member

store Table

Columns	Data Type	Description
store_id	Serial	Unique identifier for the store
manager_staff_id	int2	ID of the store manager
address_id	int2	ID of the related address
last_update	timestamp	Timestamp of the last update to the record

STEP 3. 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?

Relevant Tables:

payment (contains revenue amounts)
rental (links payments to inventory)
inventory (links to films)
film_actor (links films to actors)
actor (contains actor names)

Query Plan:

Join the tables starting from payment → rental → inventory → film → film_actor → actor
Aggregate revenue by actor to find which actor generated the most revenue.

What language are the majority of movies in the collection?

Relevant Tables:

film (contains language_id)
language (contains language names)

Query Plan:

Join film and language on language_id
Count the number of films per language
Identify the language with the highest count