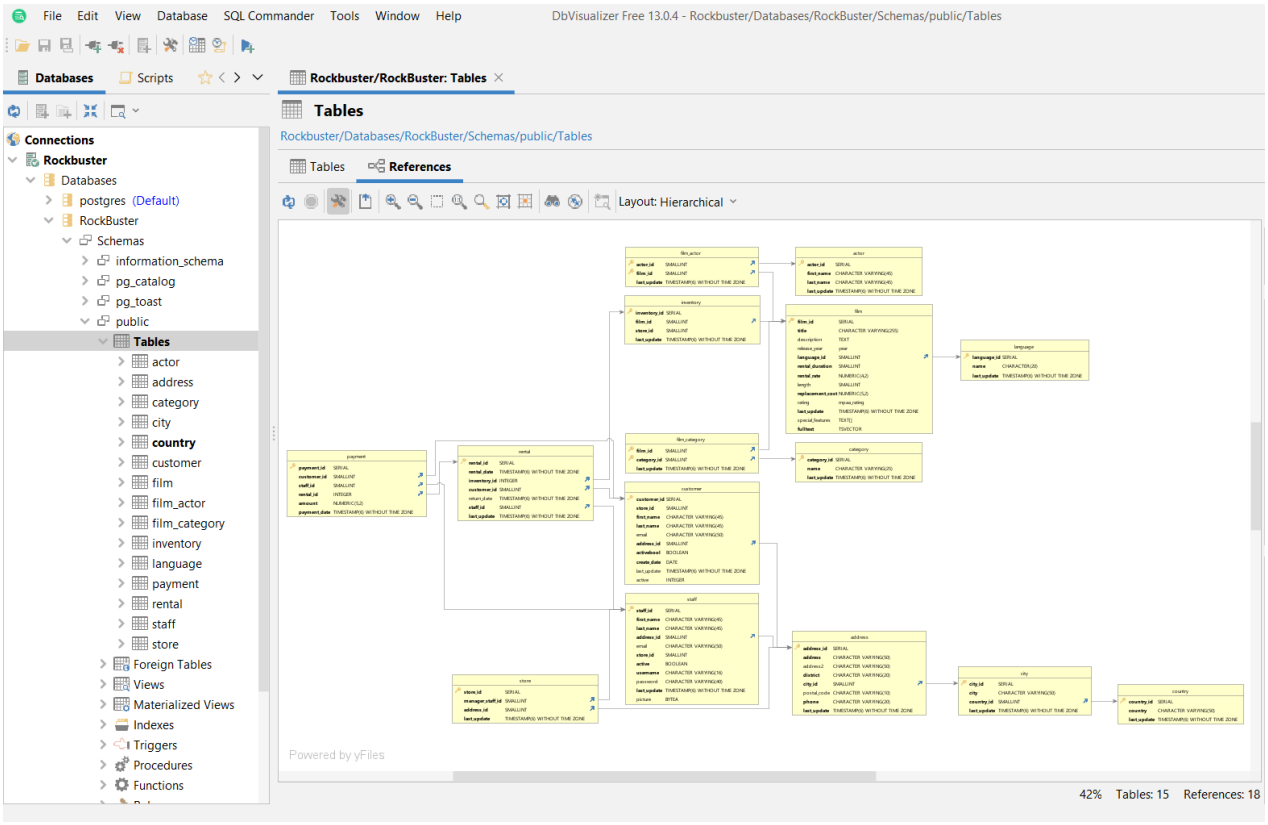
3.2 Data Storage and Structure

Step 2: Extract the ERD



Step 3: Create the first draft 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 ERD of the Rockbuster database has a snowflake schema. A snowflake schema represents by a centralized fact tables which connects to multiple dimensions, the point dimensions are further divided into more points

Fact table

Rental

Column	Data Type	Description
rental_id	Serial	Serial number assigned to rental
rental_date	TIMESTAMP(6) WITHOUT TIMEZONE	Date of rental
inventory_id	INTEGER	Number assigned to item in the table
customer_id	SMALLINT	Number assigned to customer
return_date	TIMESTAMP(6) WITHOUT TIMEZONE	Date rental was returned
staff_id	SMALLINT	Number assigned to employee
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

Dimension tables

Store

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

Payment

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

Film actor

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	Date entry was last updated

Inventory

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	Date entry was last updated

Film_category

Column	Data Type	Description
film_id	SMALLINT	Serial number assigned to rental
film_category	SMALLINT	Number assigned to genre/category
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

Customer

Column	Data Type	Description
customer_id	SERIAL	Serial number assigned to customer
store_id	SMALLINT	Number assigned to store
first_name	CHARACTERVARYING(50)	First name of customer
last_name	CHARACTERVARYING(45)	Last name of customer
email	CHARACTERVARYING(50)	Email address of customer
address_id	SMALLINT	Number assigned to customer's address
activebool	BOOLEAN	Customer's active status
create_date	DATE	Date entry was created
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated
active	INTGER	Customer's active status

Staff

Column	Data Type	Description
staff_id	SERIAL	Serial number assigned to rental
first_name	CHARACTERVARYING(45)	First name of employee
last_name	CHARACTERVARYING(45)	Last name of employee
address_id	SMALLINT	Number assigned to employee's address
email	CHARACTERVARYING(50)	Email address of employee
store_id	SMALLINT	Number assigned to store
active	BOOLEAN	Employee active status
username	CHARACTERVARYING(16)	Username of employee
password	CHARACTERVARYING(40)	Password of employee
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated
picture	BYTEA	Picture of employee

Actor

Column	Data Type	Description
actor_id	SERIAL	Serial number assigned to rental
first_name	CHARACTERVARYING(45)	First name of actor
last_name	CHARACTERVARYING(45)	Last name of actor
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

Film

Column	Data Type	Description
film_id	SERIAL	Serial number assigned to rental
title	CHARACTERVARYING(45)	Title of film
description	TEXT	Description of film
release_year	year	Release year of film
language_id	SMALLINT	Number assigned to film language
rental_duration	SMALLINT	Length of film rental
rental_rate	NUMERIC(4,2)	Price of film
length	SMALLINT	Length of film
replacement_cost	NUMERIC(5,2)	Cost of replace of film
rating	mpaa_rating	Film rating
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated
special_features	TEXT []	Special features included with film
fulltext	TSVECTOR	Keywords associated with film

Category

Column	Data Type	Description
Category_id	SERIAL	Serial number assigned to language
name	CHARACTERVARYING(20)	Name of language
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

Address

Column	Data Type	Description
address_id	SERIAL	Serial number assigned to rental
address	CHARACTERVARYING(50)	Street address
address2	CHARACTERVARYING(50)	Supplementary street address
district	CHARACTERVARYING(20)	District
city_id	SMALLINT	Number assigned to City
postal_code	CHARACTERVARYING(10)	Postal code
phone	CHARACTERVARYING(20)	Phone number
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

Language

Column	Data Type	Description
language_id	SERIAL	Serial number assigned to rental
name	CHARACTERVARYING(20)	Name of language
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

City

Column	Data Type	Description
city_id	SERIAL	Serial number assigned to rental
city	CHARACTERVARYING(20)	Name of city
country_id	SMALLINT	Number assigned to country
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

Country

Column	Data Type	Description
country_id	SERIAL	Serial number assigned to rental
country	CHARACTERVARYING(50)	Name of country
last_update	TIMESTAMP(6) WITHOUT TIMEZONE	Date entry was last updated

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?

  I will extract the information by joining three tables: actor, film actor and film to get the actors with the highest revenue.

- What language are the majority of movies in the collection

  I will combine the two tables: language and film table to know the language with majority of movie.