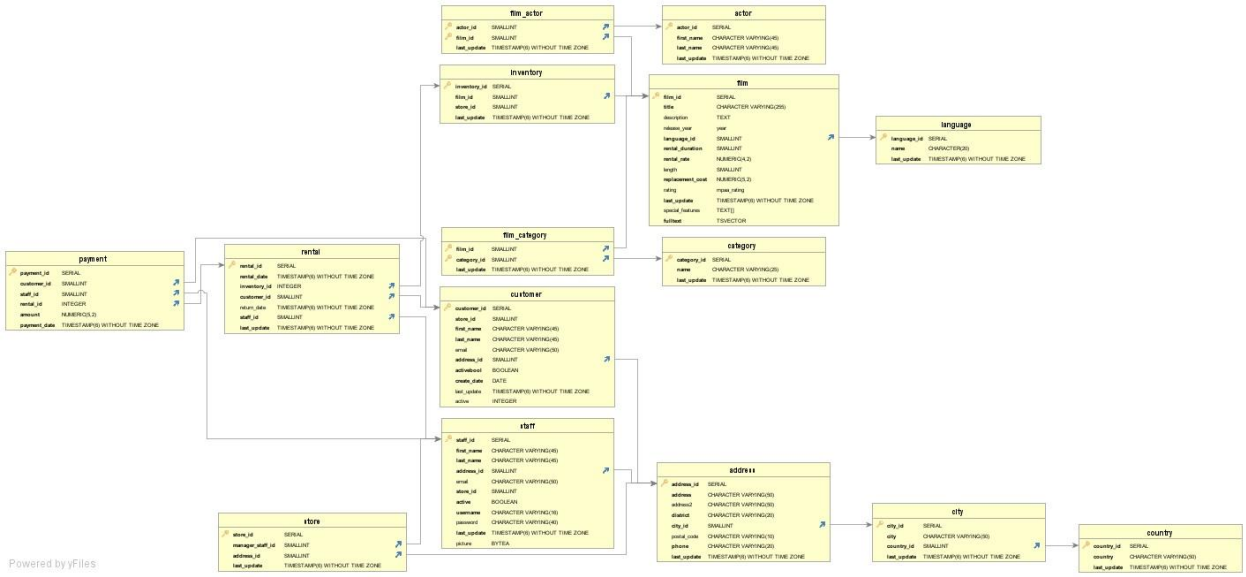
3.2: Data Storage & Structure

Step 1. Create your Answers document:

Create a new text document and call it “Answers 3.2.” You'll save a copy of your ERD, data dictionary, and written answers in this document.

Step 2. Extract the ERD:

- Download and install DbVisualizer (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.
    - This ERD is a snowflake schema. The main reason this classification works is because each table the results is a subcategory relating back to a primary one; in this case, payment. Every iteration after payment relates to payment either directly (rental, customer, address, etc) or indirectly as a descriptor for the later categories related to payment (such as payment>rental>inventory>film which has descriptors like film actor and actor).

- 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.

Fact tables:

Payment

columns	data type	description
payment_id	serial	Primary key for payment records. Identity increment for payment
customer_id	smallint	Identity increment for customers
staff_id	smallint	Identity increment for staff
rental_	integer	Identity increment for film rental
amount	numeric (5,2)	Monetary amount of payment
payment_date	timestamp (6) w/o timezone	Date/time of payment

FILM_CATEGORY

columns	data type	description
film_id	serial	Primary key for film records. Identity increment for film
category_id	smallint	Candidate key for film records. Identity increment for film category
last_update	serial	Date/time film last updated

Store

columns	data type	description
store_id	serial	Primary key for store records. Identity increment for store
manager_staff_id	smallint	Identity increment for manager & staff at store
address_id	smalint	Identity increment for store's address
last_update	timestamp(6) w/o timezone	Date/time store last updated

FILM_ACTOR

columns	data type	description
actor_id	smallint	Primary key for actor records. Identity increment for actor

film_id	smallint	Candidate key for actor records. Identity increment for film category
last_update	timestamp (6) w/o timezone	WITHOUT TIME ZONE Date/time actor last updated

Dimension Tables:

Rental

columns	data type	description
Rental_id	serial	Primary key for rental records. Identity increment for rental
Rental_date	timestamp(6) w/o timezone	Date/time of rental
Inventory_id	integer	Identity increment for inventory
Customer_id	smallint	Identity increment for inventory
Return_date	timestamp(6) w/o timezone	Date/time of return
Staff_id	smallint	Identity increment for staff
Last_update	timestamp(6) w/o timezone	Date/time rental last updated

Customer

columns	data type	description
Customer_id	serial	Primary key for customer records. Identity increment for customer
Store_id	smallint	Identity increment for store
First_name	character varying(45)	Customer's first name
Last_name	character varying(45)	Customer's last name
Email	character varying(45)	Customer's email
Address_id	smallint	Identity increment for customer's address
Activebool	boolean	True or False confirming whether customer active or not in system
Create_date	date	Date/time customer created in system
Last_update	timestamp(6) w/o timezone	Date/time customer's profile last updated
active	integer	0 or 1 integer for whether customer is active or not

Address

columns	data type	description
address_id	serial	Primary key for address records. Identity increment for address
address	character varying(50)	First line of address
address2	character varying(50)	Second line of address
district	character varying(50)	Address district
city_id	smallint	Address city
postal_code	character varying(50)	Address postal code
phone	character varying(20)	Phone number of staff
last_update	timestamp(6) w/o timezone	Date/time address profile last updated

STAFF

columns	data type	description
staff_id	serial	Primary key for staff records. Identity increment for staff
First_name	character varying(45)	Staff's first name
Last_name	character varying(45)	Staff's last name
Address_id	smallint	Identity increment for Staff's address
Email	character varying(50)	Staff's email
store_id	smallint	Identity increment for store where staff is employed
active	boolean	True or False confirming whether staff active or not in system
username	character varying(40)	Staff's username
password	character varying(40)	Staff's password
last_update	timestamp(6) w/o timezone	Date/time Staff's profile last updated
picture	bytea	Staff's picture

CITY

columns	data type	description
city_id	serial	Primary key for city records. Identity increment for city
city	character varying(50)	Name of city
country_id	smallint	Identity increment for country
last_update	timestamp(6) w/o timezone	Date/time city last updated

COUNTRY

columns	data type	description

country_id	serial	Primary key for country records. Identity increment for country
country	character varying(50)	Name of country
last_update	timestamp(6) w/o timezone	Date/time country last updated

Inventory

columns	data type	description
inventory_id	serial	Primary key for inventory records. Identity increment for inventory
film_id	smallint	Identity increment for film
store_id	smallint	Identity increment for store
last_update	timestamp(6) w/o timezone	Date/time inventory last updated

Actor

columns	data type	description
actor_id	serial	Primary key for actor records. Identity increment for actor
first_name	character varying(45)	Actor's first name
Last_name	character varying(45)	Actor's last name
last_update	timestamp(6) w/o timezone	Date/time actor last updated

Film

columns	data type	description
film_id	serial	Primary key for film records. Identity increment for film
title	CHARACTER VARYING(255)	Title of film
description	text	Description of film
release_year	year	Release year of film
language_id	smallint	Language of film
rental_duration	smallint	Rental duration of film
rental_rate	numeric(4,2)	Rental rate of film
length	smallint	Length of film
replacement_cost	numeric(5,2)	Replacement cost of film
rating	mpaa_rating	Rating of film
last_update	timestamp(6) w/o timezone	Date/time film's profile last updated
special_features	text[]	Description of what special features are included (Trailers, Behind the scenes, etc…)
fulltext	tsvector	Summary of text describing film

Actor

columns	data type	description
language_id	serial	Primary key for language records. Identity increment for language
name	character(20)	Name of language
last_update	timestamp(6) w/o timezone	Date/time language last updated

Category

columns	data type	description
category_id	serial	Primary key for category records. Identity increment for category
name	character(25)	Name of language
last_update	timestamp(6) w/o timezone	Date/time language last updated

- 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.

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

Step 5. Save and upload your file:

Save your “Answers 3.2” document as a PDF and upload it here for your tutor to review. Make sure to keep a copy for yourself! You’ll be finalizing your data dictionary at the end of this Achievement.