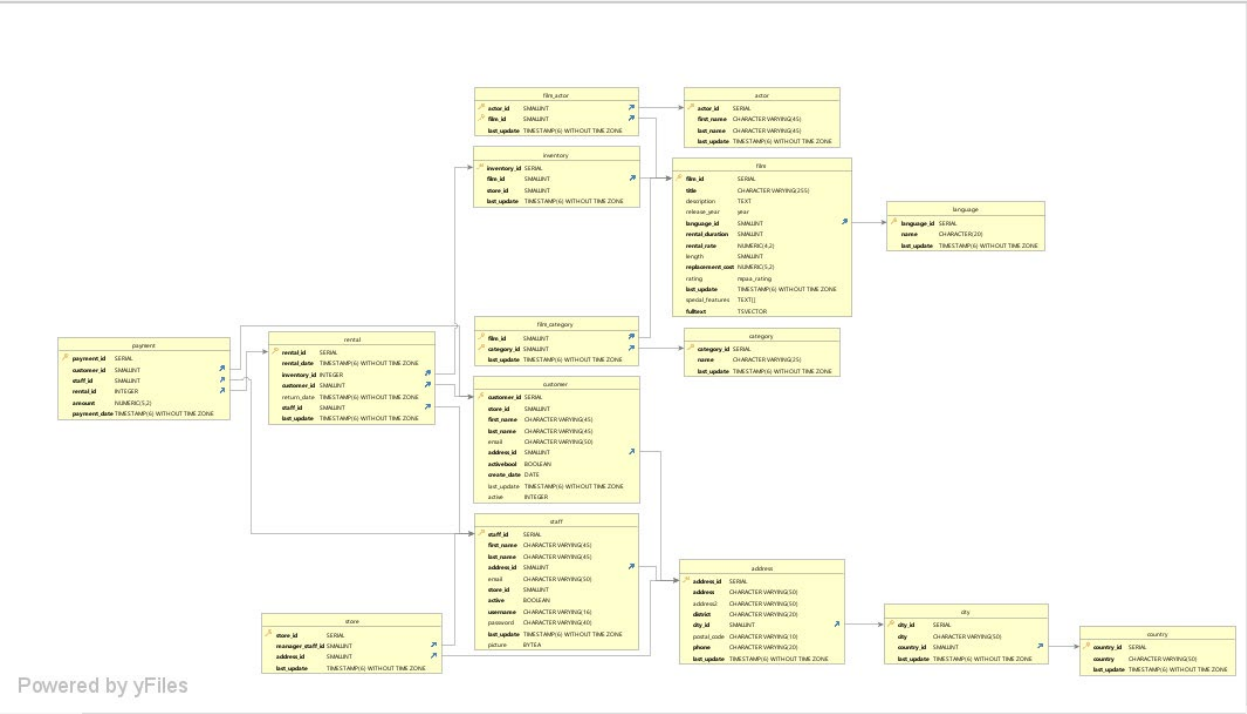
Data Immersion

May 20, 2022

Exercise 3.2

Step 2



Step 3

I would consider this a Snowflake Schema as the various tables are interconnected, and not just connected to the center of the Star.

Fact Table

rental		
Column	**Data Type**	**Description**
rental_id	SERIAL	Rental identification number
rental_date	TIMESTAMP(6) WITHOUT TIME ZONE	Date of Rental
inventory_id	INTEGER	Inventory identification number
customer_id	SMALLINT	Customer identification number
return_date	TIMESTAMP(6) WITHOUT TIME ZONE	Date the rental was returned
staff_id	SMALLINT	Staff identification number
last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Dimension Tables

Payment		
Payment_id	SERIAL	Payment identification number
Customer_id	SMALLINT	Customer identification number
Staff_id	SMALLINT	Staff identification number
Rental_id	INTEGER	Rental identification number
Amount	NUMERIC(5,2)	Amount paid (including cents)
Payment_date	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Store		
Store_id	SERIAL	Store identification number
Manager_staff_id	SMALLINT	Staff manager identification number
Address_id	SMALLINT	Store address identification number
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Film_Actor		
Actor_id	SERIAL	Actor identification number
Film_id	SMALLINT	Film identification number
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

inventory		
Actor_id	SERIAL	Actor identification number
Film_id	SMALLINT	Film identification number
Store_id	SMALLINT	Store identification number
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Film_category		
Film_id	SMALLINT	Film identification number
Category_id	SMALLINT	Film category identification number
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Customer		
Customer_id	SERIAL	Customer identification number
Store_id	SMALLINT	Store identification number
First_name	CHARACTER VARYING(45)	First name of the customer
Last_name	CHARACTER VARYING(45)	Last name of the customer
Email	CHARACTER VARYING(50)	Email address of the customer
Address_id	SMALLINT	Address of the customer
Activebool	BOOLEAN	Is the customer still active?
Create_date	DATE	Date the customer account was created
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated
active	INTEGER	Is the customer still active?

Staff		
staff_id	SERIAL	Staff identification number
First_name	CHARACTER VARYING(45)	First name of the Staff Member
Last_name	CHARACTER VARYING(45)	Last name of the Staff Member
Address_id	SMALLINT	Staff address identification number
Email	CHARACTER VARYING(50)	Staff Member email address
Store_id	SMALLINT	Store identification number
Active	BOOLEAN	Is the Staff Member active?
Username	CHARACTER VARYING(16)	Staff Member Username
Password	CHARACTER VARYING(40)	Staff Member Password
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated
picture	BYTEA	Employee's Picture

Actor		
Actor_id	SERIAL	Actor identification number
First_name	CHARACTER VARYING(45)	First name of the actor
Last_name	CHARACTER VARYING(45)	Last name of the actor
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Film		
Film_id	SERIAL	Film identification number
Title	CHARACTER VARYING(255)	Film Title
Description	TEXT	Description of the film
Release_year	YEAR	Year the film was released
Language_id	SMALLINT	Languages the film is in
Rental_duration	SMALLINT	Length of the rental duration
Rental_rate	NUMERIC(4,2)	Price of the rental
Length	SMALLINT	Movie length
Replacement_cost	NUMERIC(5,2)	Cost of replacing the movie
Rating	MPAA_RATING	Movie rating
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated
Special_features	TEXT	Any special features included
fulltext	TSVECTOR	Specific words associated with the film

Category		
Category_id	SERIAL	Movie category identification number
Name	CHARACTER VARYING(25)	Name of the movie category
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Address		
Address_id	SERIAL	Address identification number
Address	CHARACTER VARYING(50)	Address
Address2	CHARACTER VARYING(50)	Additional address space
District	CHARACTER VARYING(20)	District of the address
City_id	SMALLINT	City identification number
Postal_code	CHARACTER VARYING(10)	Postal code of the address
Phone	CHARACTER VARYING(20)	Phone number associated with the address
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Language		
Language_id	SERIAL	Language identification number
Name	CHARACTER(20)	Name of the language
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

City		
City_id	SERIAL	City identification number
City	CHARACTER VARYING(50)	City
Country_id	SMALLINT	Country identification number
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Country		
Country_id	SERIAL	Country identification number
Country	CHARACTER VARYING(50)	Country
Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date this entry was last updated

Step 4

To figure out which actor has grossed the most revenue through their rentals, we would need the following tables:

- Actor
- Film_Actor
- Film

To find out which language most of the movies are in, we would need the following tables:

- Language
- Film