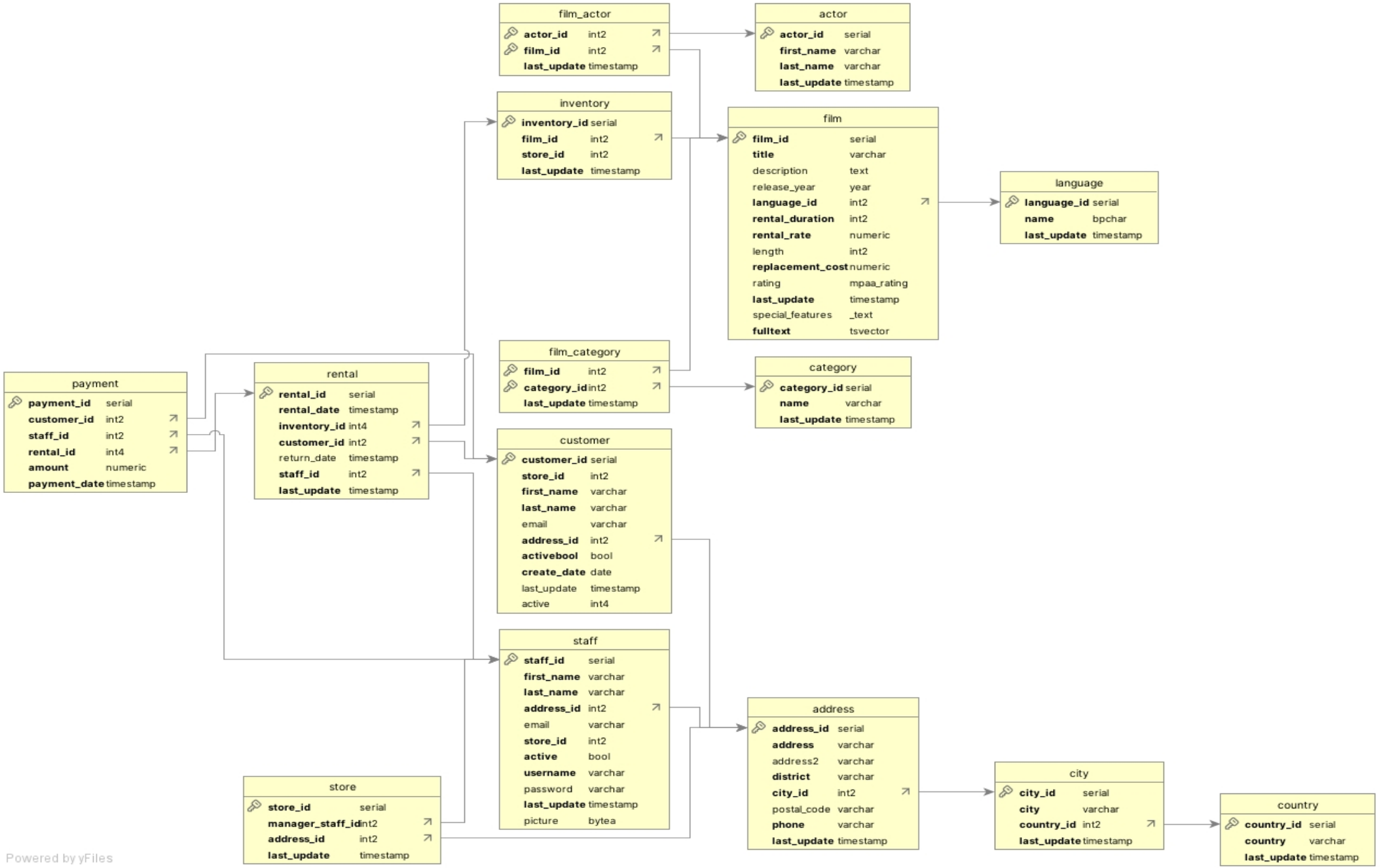
Databases & SQL for Analysts

3.2: Data Storage & Structure

Step 2. Extract the ERD


film_actor
actor_id int2
film_id int2
last_update timestamp
actor
actor_id serial
first_name varchar
last_name varchar
last_update timestamp
inventory
inventory_id serial
film_id int2
store_id int2
last_update timestamp
film
film_id serial
title varchar
description text
release_year year
language_id int2
rental_duration int2
rental_rate numeric
length int2
replacement_cost numeric
rating mpaa_rating
last_update timestamp
special_features _text
fulltext tsvector
language
language_id serial
name bpchar
last_update timestamp
film_category
film_id int2
category_id int2
last_update timestamp
category
category_id serial
name varchar
last_update timestamp
payment
payment_id serial
customer_id int2
staff_id int2
rental_id int4
amount numeric
payment_date timestamp
rental
rental_id serial
rental_date timestamp
inventory_id int4
customer_id int2
return_date timestamp
staff_id int2
last_update timestamp
customer
customer_id serial
store_id int2
first_name varchar
last_name varchar
email varchar
address_id int2
activebool bool
create_date date
last_update timestamp
active int4
staff
staff_id serial
first_name varchar
last_name varchar
address_id int2
email varchar
store_id int2
active bool
username varchar
password varchar
last_update timestamp
picture bytea
address
address_id serial
address varchar
address2 varchar
district varchar
city_id int2
postal_code varchar
phone varchar
last_update timestamp
city
city_id serial
city varchar
country_id int2
last_update timestamp
country
country_id serial
country varchar
last_update timestamp
store
store_id serial
manager_staff_id int2
address_id int2
last_update timestamp
Powered by yFiles

Step 3. Create the first draft of a data dictionary:

Does the Rockbuster database have a snowflake schema or a star schema? Write a brief explanation for your answer.

Snowflake schema:

The fact table are linked to dimensions, these dimensions are themselves normalized into sub-tables us we can see in the ERD

Fact table

table_name	column_name	data_type	description
payment	payment_id	integer	payment_id
payment	customer_id	smallint	customer_id
payment	staff_id	smallint	staff_id
payment	rental_id	integer	rental_id
payment	amount	numeric	Amount of payment
payment	payment_date	timestamp without time zone	payment_date
rental	rental_id	integer	rental_id
rental	rental_date	timestamp without time zone	rental_date
rental	inventory_id	integer	inventory_id
rental	customer_id	smallint	customer_id
rental	return_date	timestamp without time zone	return_date

rental	staff_id	smallint	staff_id
rental	last_update	timestamp without time zone	last_update

Dimension table

table_name	column_name	data_type	description
actor	actor_id	integer	actor_id
actor	first_name	varchar(45)	first_name of actor
actor	last_name	varchar(45)	last_name of actor
actor	last_update	timestamp without time zone	last_update
address	address_id	integer	address_id
address	address	varchar(50)	
address	address2	varchar(50)	
address	district	varchar(20)	district
address	city_id	smallint	city_id
address	postal_code	varchar(10)	postal_code
address	phone	varchar(20)	Phone number
address	last_update	timestamp without time zone	last_update
category	category_id	integer	category_id
category	name	varchar(25)	name
category	last_update	timestamp without time zone	last_update
city	city_id	integer	city_id
city	city	varchar(50)	city
city	country_id	smallint	country_id

city	last_update	timestamp without time zone	last_update
country	country_id	integer	country_id
country	country	varchar(50)	country
country	last_update	timestamp without time zone	last_update
customer	customer_id	integer	customer_id
customer	store_id	smallint	store_id
customer	first_name	varchar(45)	first_name of custmer
customer	last_name	varchar(45)	last_name of custmer
customer	email	varchar(50)	Email adress
customer	address_id	smallint	address_id
customer	activebool	boolean	activebool
customer	create_date	date	create_date
customer	last_update	timestamp without time zone	last_update
customer	active	integer	active
film	film_id	integer	film_id
film	title	varchar(255)	title
film	description	text	Description of film
film	release_year	integer	Date of release
film	language_id	smallint	language_id
film	rental_duration	smallint	rental_duration
film	rental_rate	numeric	rental_rate
film	length	smallint	length
film	replacement_cost	numeric	replacement_cost
film	rating	USER-DEFINED	rating
film	last_update	timestamp without time zone	last_update
film	special_features	ARRAY	special_features
film	fulltext	tsvector	fulltext

film_actor	actor_id	smallint	actor_id
film_actor	film_id	smallint	film_id
film_actor	last_update	timestamp without time zone	last_update
film_category	film_id	smallint	film_id
film_category	category_id	smallint	category_id
film_category	last_update	timestamp without time zone	last_update
inventory	inventory_id	integer	inventory_id
inventory	film_id	smallint	film_id
inventory	store_id	smallint	store_id
inventory	last_update	timestamp without time zone	last_update
language	language_id	integer	language_id
language	name	char(20)	name
language	last_update	timestamp without time zone	last_update
staff	staff_id	integer	staff_id
staff	first_name	varchar(45)	first_name
staff	last_name	varchar(45)	last_name
staff	address_id	smallint	address_id
staff	email	varchar(50)	email
staff	store_id	smallint	store_id
staff	active	boolean	active
staff	username	varchar(16)	username
staff	password	varchar(40)	password
staff	last_update	timestamp without time zone	last_update
staff	picture	bytea	picture
store	store_id	integer	store_id
store	manager_staff_id	smallint	manager_staff_id

store	address_id	smallint	address_id
store	last_update	timestamp without time zone	last_update

Step 4. Find information:

- **Which actors brought Rockbuster the most revenue?**

  Payment >> rental >> inventory >> film_actor > actor

- **What language are the majority of movies in the collection?**

  Film>>Language