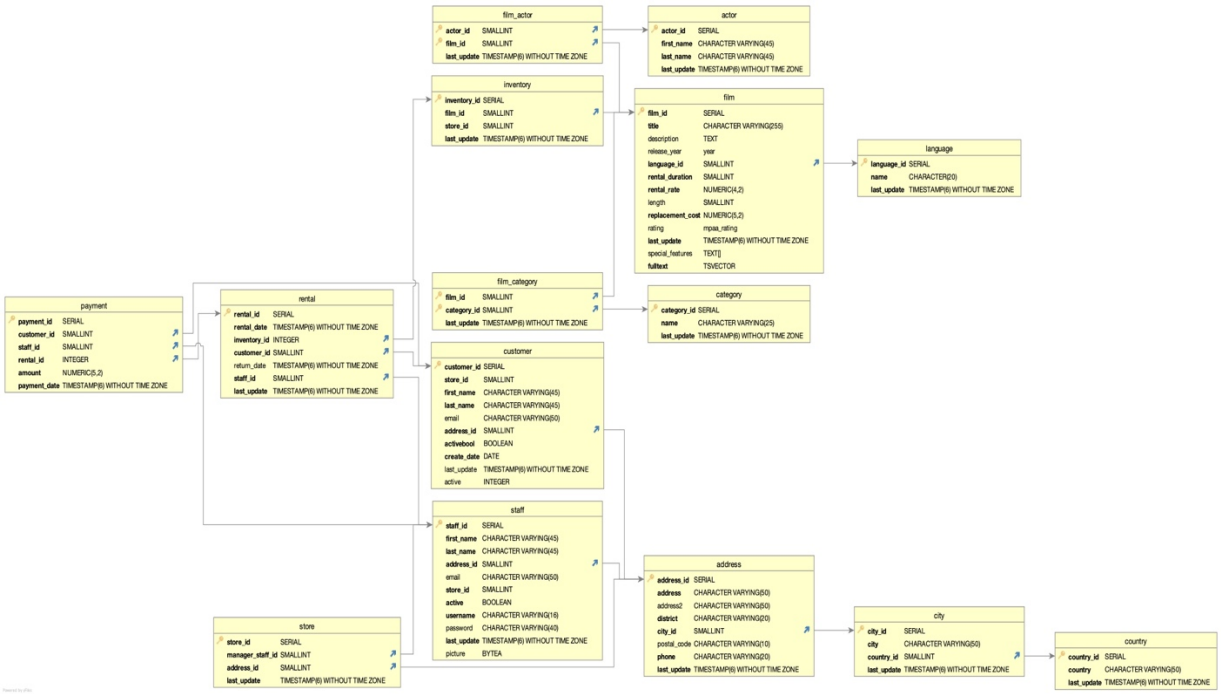
Juan Ignacio Galvalisi

Exercise 3.2: Data Storage & Structure

Step 2. Extract the ERD



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.

It has a snowflake schema for the following reason. The " ad-hoc " tables serve as a deeper explanation of the fact tables. In this sense, there are more interconnected dimension and sub-dimension tables than in a star schema. It is more probably here that you need to use many joins to fletch the data.

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 Table

rental

	Column	Data Type	Description
🔑	rental_id	SERIAL	Number assigned to rental
	rental_date	TIMESTAMP(6) WITHOUT TIME ZONE	Date of rental

	inventory_id	INTEGER	Number assigned to inventory
	customer_id	SMALLINT	Number assigned to customer
	return_date	TIMESTAMP(6) WITHOUT TIME ZONE	Date rental was returned
	staff_id	SMALLINT	Number assigned to staff
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

Dimension Tables

payment

	Column	Data Type	Description
🔑	payment_id	SERIAL	Number assigned to payment
	customer_id	SMALLINT	Number assigned to customer
	staff_id	SMALLINT	Number assigned to staff
	rental_id	INTEGER	Number assigned to rental
	amount	NUMERIC(5,2)	Amount paid
	payment_date	TIMESTAMP(6) WITHOUT TIME ZONE	Date of payment

inventory

	Column	Data Type	Description
🔑	inventory_id	SERIAL	Number assigned to inventory
	film_id	SMALLINT	Number assigned to film
	store_id	SMALLINT	Number assigned to store
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

customer

	Column	Data Type	Description
🔑	customer_id	SERIAL	Number assigned to customer
	store_id	SMALLINT	Number assigned to store
	first_name	CHARACTER VARYING(45)	First name of customer
	last_name	CHARACTER VARYING(45)	Last name of customer
	email	CHARACTER VARYING(50)	Email of customer
	address_id	SMALLINT	Number assigned to address
	activebool	BOOLEAN	Customer's active status
	create_date	DATE	Date entry was created
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated
	active	INTEGER	Is customer active or inactive?

staff

	Column	Data Type	Description
🔑	staff_id	SERIAL	Number assigned to employee
	first_name	CHARACTER VARYING(45)	First name of employee
	last_name	CHARACTER VARYING(45)	Last name of employee
	address_id	SMALLINT	Number assigned to address
	email	CHARACTER VARYING(50)	Email of employee
	store_id	SMALLINT	Number assigned to store
	active	BOOLEAN	Is employee active or inactive?

	username	CHARACTER VARYING(16)	Username of employee
	password	CHARACTER VARYING(40)	Password of employee
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated
	picture	BYTEA	Picture of employee

film

	Column	Data Type	Description
🔑	film_id	SERIAL	Number assigned to film
	title	CHARACTER VARYING(255)	Title of film
	description	TEXT	Film description
	release_year	Year	Release year of film
	language_id	SMALLINT	Number assigned to language
	rental_duration	SMALLINT	Length of film rental
	rental_rate	NUMERIC(4,2)	Price of film rental
	length	SMALLINT	Length of film
	replacement_cost	NUMERIC(5,2)	Costo to replace film
	rating	mpaa_rating	Film rating
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated
	special_features	TEXT[]	Special features included with film
	fulltext	TSVECTOR	Keywords related with film

language

	Column	Data Type	Description
🔑	language_id	SERIAL	Number assigned to language
	name	CHARACTER(20)	Language name
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

film_category

	Column	Data Type	Description
🔑	film_id	SMALLINT	Number assigned to film
🔑	category_id	SMALLINT	Number assigned to category
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

category

	Column	Data Type	Description
🔑	category_id	SERIAL	Number assigned to category
	name	CHARACTER VARYING(25)	Name of genre
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

actor

	Column	Data Type	Description
🔑	actor_id	SERIAL	Number assigned to actor
	first_name	CHARACTER VARYING(45)	Name of actor
	last_name	CHARACTER VARYING(45)	Last name of actor

	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

film_actor

	Column	Data Type	Description
🔑	actor_id	SMALLINT	Number assigned to actor
🔑	film_id	SMALLINT	Number assigned to film
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

store

	Column	Data Type	Description
🔑	store_id	SERIAL	Number assigned to store
	manager_staff_id	SMALLINT	Number assigned to store manager
	address_id	SMALLINT	Number assigned to address
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

address

	Column	Data Type	Description
🔑	address_id	SERIAL	Number assigned to address
	address	CHARACTER VARYING(50)	Street address
	address2	CHARACTER VARYING(50)	Additional street address
	district	CHARACTER VARYING(20)	Name of district
	city_id	SMALLINT	Number assigned to city
	postal_code	CHARACTER VARYING(10)	Postal code number
	phone	CHARACTER VARYING(20)	Phone number
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

city

	Column	Data Type	Description
🔑	city_id	SERIAL	Number assigned to city
	city	CHARACTER VARYING(50)	Name of city
	country_id	SMALLINT	Number assigned to country
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

country

	Column	Data Type	Description
🔑	country_id	SERIAL	Number assigned to country
	country	CHARACTER VARYING(50)	Name of country
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Data entry was last updated

Step 4. Find information

Which actors brought Rockbuster the most revenue?

We should consider three different tables:

1. The “actor” table has the name of the actors.
2. The “film_actor” table tells us which films are related to this particular actor.
3. The “film” table can show us the rental rate of this film.

We would use the keys *actor_id* and *film_id* with the aim of connecting these tables to each other.

What language are the majority of movies in the collection?

We should look at two different tables. First, the “film” table has the name of the film. Second, the “language” table tells us the language name. We should use the key *language_id* in order to connect these tables together.