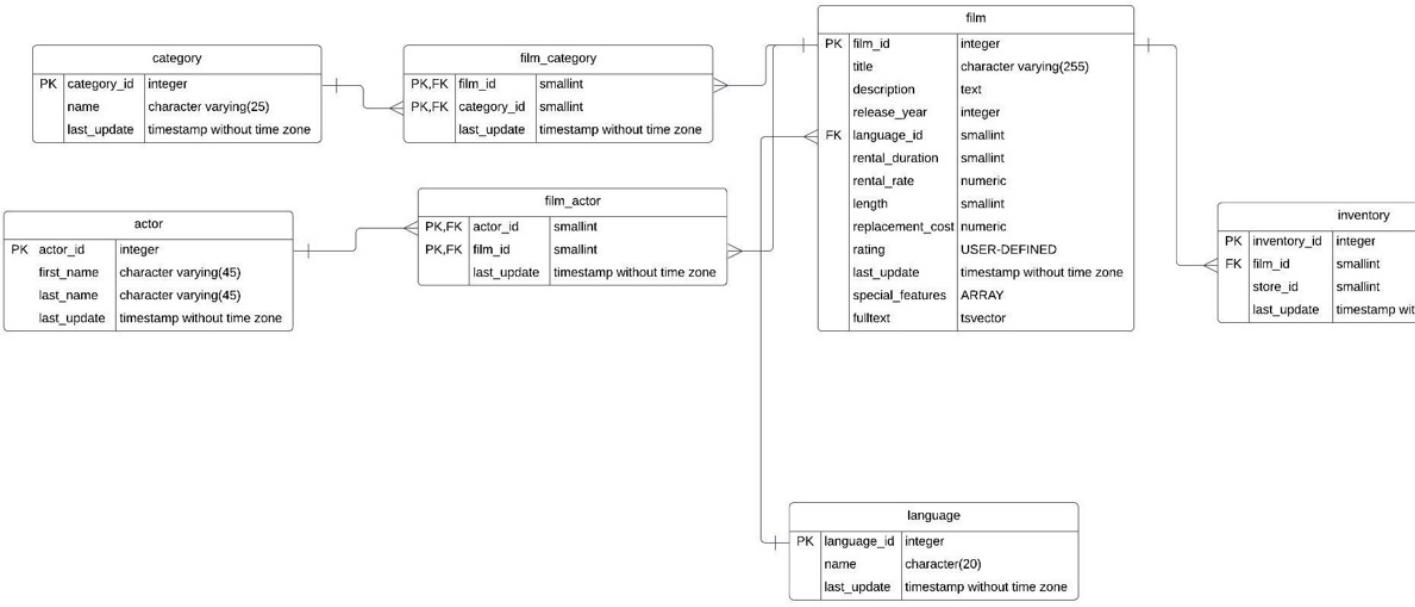
Jeremy Obach
CareerFoundry DA Immersion
Task 3.2

Rockbuster Entity Relationship Diagram (ERD)



Create a first draft of a data dictionary

3a. The Rockbuster database has a snowflake schema, evidenced by the branching tables, from film to film category, to category for example, and by the branching of fact tables, like rental, to sub-dimension tables like category or actor. I was expecting for there to be another type of schema name after Star and Snowflake because both of those sorts of imply one central table,

whereas this Rockbuster ERD has several interconnected tables that make more like a subway system pattern.
The more I think about it though, is the rental table actually the central table? Because all the other tables could be dimension tables that describe various aspects of the central transaction/event, a rental. Maybe it does fit the Snowflake schema better than I initially thought.

3b.
Fact Tables:
rental

Columns	**Data Type**	**Description**
rental_id	Integer	Primary key, unique ID for each rental.
rental_date	Timestamp without time zone	Date/time rental event took place.
Inventory_id	Integer	Foreign key, unique ID for each piece of inventory
Customer_id	smallint	Foreign key, unique Id for each customer
Return_date	Timestamp without time zone	Date/time rental is due back.
Staff_id	smallint	Foreign key, unique ID for each staff member
last_update	timestamp without time zone	Date/time entry last updated

(Question for tutor, is payment an event, and therefore a fact table? It feels like it would be, rather than a dimension table.)
payment

Columns	**Data Type**	**Description**
payment_id	integer	Primary key, unique ID for each payment event
customer_id	smallint	foreign key, unique ID for each customer
staff_id	smallint	foreign key, unique ID for each staffmember
rental_id	integer	foreign key, unique ID fore each rental event
amount	numeric	price of particular payment event.
payment_date	timestamp without time zone	Date/time that payment took place

Dimension Tables:
Category

Columns	**Data Type**	**Description**
category_id	integer	Primary key, unique ID for each film category
name	character varying (25)	name of category

last_update	timestamp without time zone	Date/time of last update

Film_category

Columns	**Data Type**	**Description**
film_id	smallint	PK and FK, this table links film and category tables
category_id	smallint	PK and FK, this table links film and category tables
last_update	timestamp without time zone	Date/time of last update

Actor

Columns	**Data Type**	**Description**
actor_id	integer	Primary key, unique ID for each actor
first_name	character varying (45)	Actor's first name
last_name	character varying (45)	Actor's last name
last_update	timestamp without time zone	Date/time of last update

Film_actor

Columns	**Data Type**	**Description**
actor_id	smallint	PK and FK, this table links film and actor tables
film_id	smallint	PK and FK, this table links film and actor tables
last_update	timestamp without time zone	Date/time of last update

film

Columns	**Data Type**	**Description**
film_id	integer	Primary key, unique for each film
title	character varying (255)	The title of the film
description	text	Description of the film
release_year	integer	Year film was released
language_id	smallint	Foreign key, unique ID for each language
rental_duration	smallint	length of rental (allowed I'm assuming)
rental_rate	numeric	Price for rental for days listed in rental_duration.

length	smallint	length of movie, probably in minutes
replacement_cost	numeric	cost to replace this film
rating	USER-DEFINED	MPAA rating (G thru NC-17)
last_update	timestamp w/o time zone	Date/time of last update
special_features	ARRAY	special features accompanying the film. Bloopers, deleted scenes, etc.
full_text	tsvector	lists keywords associated with the film and its contents

language

Columns	**Data Type**	**Description**
language_id	integer	Primary key, unique ID for each language
name	character(20)	Name of language
last_update	timestamp without time zone	Date/time of last update

inventory

Columns	**Data Type**	**Description**
inventory_id	integer	Primary key, unique ID for each piece of inventory
film_id	smallint	Foreign key, unique ID for each film
store_id	smallint	Unique ID for each store/location
last_update	timestamp without time zone	Date/time of last update

customer

Columns	**Data Type**	**Description**
customer_id	integer	Primary key, unique ID for each customer
store_id	small_int	Unique ID for each store/location
first_name	character varying (45)	customer's first name
last_name	character varying (45)	customer's last name
email	character varying (50)	customer's email
address_id	smallint	Foreign key, unique ID for each address
activebool	Boolean	Is the customer active? Y/N
create_date	date	date customer account was created
last_update	timestamp without time zone	Date/time of last update
active	integer	Assuming it's how many rentals this customer has are active. Looked up the table and it's mostly 1's with some 0's but the zero's aren't collocated with "activebool" falses.

staff

Columns	Data Type	Description
staff_id	integer	Primary key, unique ID for each staff member
first_name	character varying (45)	staff member's first name
last_name	character varying (45)	staff member's last name
address_id	smallint	Foreign key, unique ID for each address
email	character varying (50)	staff member's email
store_id	smallint	Unique ID for each store/location
active	Boolean	is this staff member active? Y/N
username	character varying (16)	staff member's username
password	character varying (40)	staff member's password
last_update	timestamp without time zone	Date/time of last update
picture	bytea	photo of staff member – so there's only two employees, one has [binary data] in this slot, the other has [null]

address

Columns	Data Type	Description
address_id	integer	Primary key, unique ID for each address
address	character varying (50)	Line one of physical address
address2	character varying (50)	Line two of physical address
district	character varying (20)	State or province, likely
city_id	smallint	foreign key, unique ID for each city
postal_code	character varying (10)	address's postal code
phone	character varying (20)	phone number associated with address
last_update	timestamp without time zone	Date/time of last update

store

Columns	Data Type	Description
store_id	integer	Primary key, unique ID for each store
manager_staff_id	smallint	Foreign key, staff_id of branch/store manager
address_id	smallint	Foreign key, unique ID for each address
last_update	timestamp without time zone	Date/time of last update

city

Columns	Data Type	Description
city_id	integer	Primary key, unique ID for each city
city	character varying (50)	name of city

country_id	smallint	Foreign key, unique ID for each country
last_update	timestamp without time zone	Date/time of last update

Columns	**Data Type**	**Description**
country_id	integer	Primary key, unique ID for each country
country	character varying (50)	name of country
last_update	timestamp without time zone	Date/time of last update

Find Information

Use Data Dictionary to figure out which tables you'd need to answer the questions below.
4a. Which actors brought Rockbuster the most revenue?
Tables: actor, film_actor, film, inventory, rental. Possibly payment, but could probably tabulate revenue from # of rentals from a film a certain actor is in by the rental_rate in film table.

4b. What language are the majority of movies in the collection?
Tables: language, film. If you were trying to find the majority language by physical copies then add inventory. Otherwise those two should be enough.