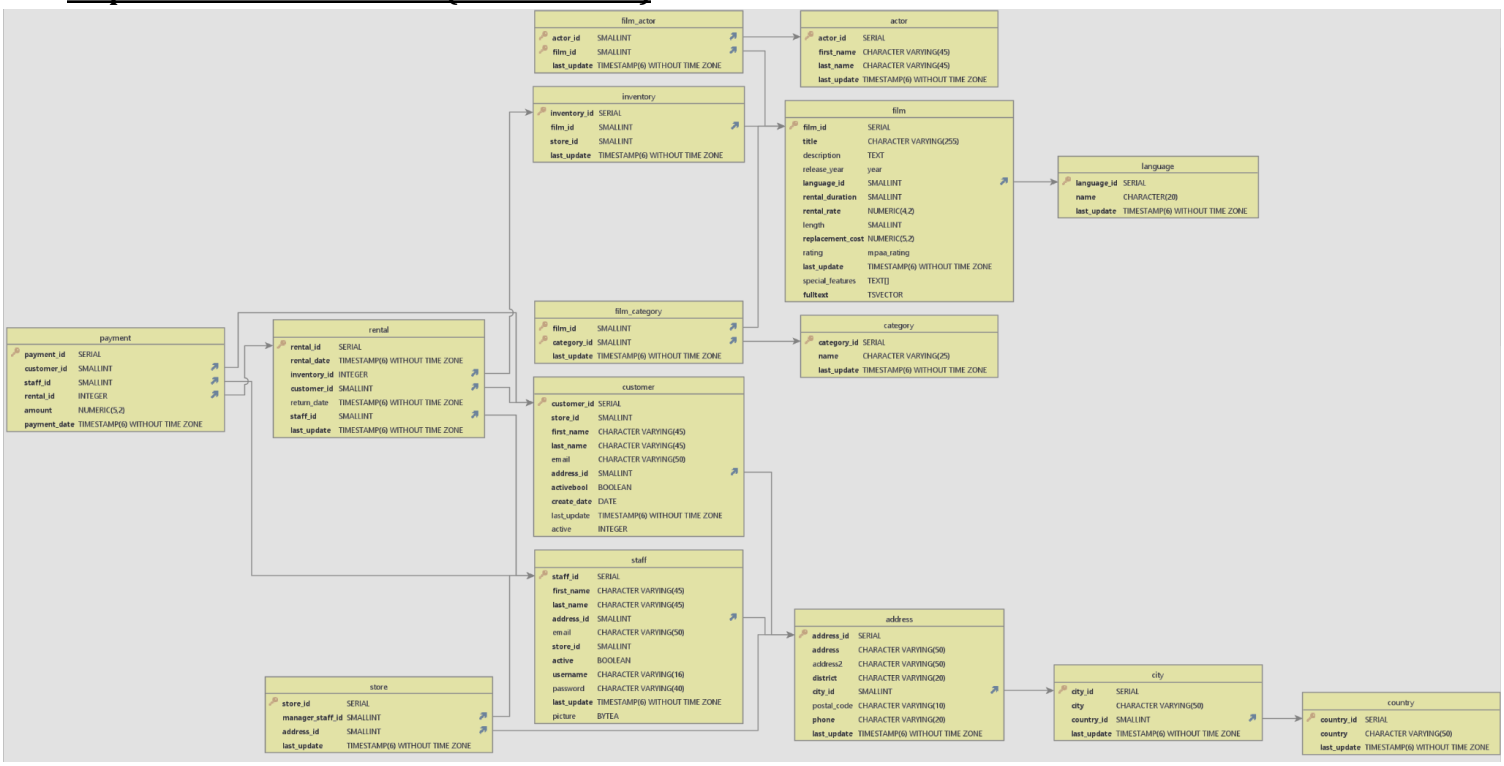
## Step 2: Rockbuster- Schema (Visualizer DB)




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

- The schema exhibits a snowflake configuration because it has multiple sub-dimensional tables that branch off of the dimension tables that branch off the fact table.

## Schema Breakdown

**Fact Tables:**

- rental

| Column | Data Format | Description |
|---|---|---|
| **rental_id** | Serial | Unique identifier for a rental (number) |
| **rental_date** | Timestamp (w/o timezone) | The date a rental was checked out |
| **inventory_id** | Integer | Unique identifier for a specific piece of inventory (i.e. the specific VHS or DVD) (number) |
| **customer_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for a customer (number) |

| | | |
|---|---|---|
| **return_date** | Timestamp (w/o timezone) | The date a rental was returned |
| **staff_id** | Small Integer (2-byte, values between -32,768 and 32,767) | Unique identifier for the staff member involved in the rental (number) |
| **last_update** | Timestamp (w/o timezone) | The date a rental was returned |

- payment

| Column | Data Format | Description |
|---|---|---|
| **payment_id** | Serial | Unique identifier for a payment transaction (number) |
| **customer_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for a customer in the transaction (number) |
| **staff_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for a staff member involved in the transaction (number) |
| **rental_id** | Integer | Unique identifier for a rental transaction (number) |
| **amount** | Numeric (5,2) | Payment value. |
| **payment_date** | Timestamp (w/o timezone) | Date payment was rendered. |

**Dimension Tables:**

- inventory

| Column | Data Type | Description |
|---|---|---|
| **inventory_id** | Serial | Unique identifier of a specific piece of inventory (in this case, VHS tapes or DVDs.) (number) |
| **film_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for a specific film title (i.e. *Gone With the Wind, Up, etc.*) (number) |
| **store_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier of a specific store related to the inventory in question (number) |
| **last_update** | Timestamp (w/o timezone) | Date when inventory status |

| | | |
|---|---|---|
| | | was last updated. |

- film

| Column | Data Type | Description |
|---|---|---|
| **film_id** | Serial | Unique identifier for a specific film title (i.e. *Gone With the Wind, Up, etc.*) (number) |
| **title** | Character Variable (Max 255 characters) | Unique film title. |
| **description** | Text | Description of film. |
| **release_year** | year | Year specific film was released. |
| **language_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier of a film's language (number) |
| **rental_duration** | Small Integer(2-byte, values between -32,768 and 32,767) | Length of time for a particular film was rented- cumulative, I presume. |
| **rental_rate** | Numeric (4,2) | Price for renting expressed as a rate. |
| **length** | Small Integer(2-byte, values between -32,768 and 32,767) | Length of time for a specific film. |
| **replacement_cost** | Numeric (5,2) | Replacement cost for a specific film. |
| **rating** | mpaa_rating | Unique MPAA rating for a specific film. |
| **last_update** | Timestamp (w/o timezone) | Date a specific film was last updated in the database. |
| **special_features** | Text[] | A film's special features. |
| **fulltext** | TSVector (breaks up a sentence or phrase into distinct words/lexemes and provides each letter's respective order as a number next to it) | Organizes text fields in the entry into an easily-searchable format. |

- language

| Column | Data Type | Description |
|---|---|---|
| **language_id** | Serial | Unique identifier for a language (number) |
| **name** | Character (Max: 20 characters) | Full name for the language |
| **last_update** | Timestamp (w/o timezone) | Last update of the language entry in the database |

- customer

| Column | Data Type | Description |
|---|---|---|
| **customer_id** | Serial | Unique identifier for the customer (number) |
| **store_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for the store (number) |
| **first_name** | Character Variable (Max 45 characters) | First name of the customer |
| **last_name** | Character Variable (Max 45 characters) | Last name of the customer |
| **email** | Character Variable (Max 50 characters) | Email of the customer |
| **address_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for a customer address (number) |
| **activebool** | Boolean | True, False, or Null, is the customer active? |
| **create_date** | Date (date only) | Date the entry for the customer was created. |
| **last_update** | Timestamp (w/o timezone) | Date the entry was most recently updated. |
| **active** | Integer | Number value |

- store

| Column | Data Type | Description |
|---|---|---|
| **store_id** | Serial | Unique identifier for a store |

| | | |
|---|---|---|
| | | location (number) |
| **manager_staff_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for a managerial staff member (number) |
| **address_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique address identifier for a particular store (number) |
| **last_update** | Timestamp (w/o timezone) | Date the entry was most recently updated. |

- staff

| Column | Data Type | Description |
|---|---|---|
| **staff_id** | Serial | Unique staff member identifier (number) |
| **first_name** | Character Variable (Max 45 characters) | First name of the staff member |
| **last_name** | Character Variable (Max 45 characters) | Last name of the staff member |
| **address_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique address identifier (number) |
| **email** | Character Variable (Max 50 characters) | Staff member email |
| **store_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique store identifier (number) |
| **active** | Boolean | True, False, or Null- is the staff member active? |
| **username** | Character Variable (Max 16 characters) | Staff member username |
| **password** | Character Variable (Max 40 characters) | Staff member password |
| **last_update** | Timestamp (w/o timezone) | Date the entry was last updated in the database |
| **picture** | Bytea (8-bit bytes, values- a series of binary strings) | Picture of the staff member as a string of binary data 2-bit bytes. |

- address

| Column | Data Type | Description |
|---|---|---|
| **address_id** | Serial | Unique identifier for a store address (number) |
| **address** | Character Variable (Max 50 characters) | First line of store address |
| **address2** | Character Variable (Max 50 characters) | Second line of store address |
| **district** | Character Variable (Max 20 characters) | Store district |
| **city_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for an address' city (number) |
| **postal_code** | Character Variable (Max 10 characters) | Postal code for the specific address |
| **phone** | Character Variable (Max 20 characters) | Phone number associated with the address. |
| **last_update** | Timestamp (w/o timezone) | Date the entry was last updated in the database. |

- city

| Column | Data Type | Description |
|---|---|---|
| **city_id** | Serial | Unique identifier for the city (number) |
| **city** | Character Variable (Max 50 characters) | Full name of the city |
| **country_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for the country of the city (number) |
| **last_update** | Timestamp (w/o timezone) | Date the entry was last updated in the database. |

- country

| Column | Data Type | Description |
|---|---|---|
| **country_id** | Serial | Unique identifier for the country associated with the film (number) |
| **country** | Character Variable (Max 50 characters) | Country name |
| **last_update** | Timestamp (w/o timezone) | Date the entry was last updated in the database. |

- film_actor

| Column | Data Type | Description |
|---|---|---|
| **actor_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for the actor (number) |
| **film_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for the film (number) |
| **last_update** | Timestamp (w/o timezone) | Date the entry was last updated in the database. |

- actor

| Column | Data Type | Description |
|---|---|---|
| **actor_id** | Serial | Unique identifier of the actor (number). |
| **first_name** | Character Variable (Max 45 characters) | Actor's first name. |
| **last_name** | Character Variable (Max 45 characters) | Actor's last name. |
| **last_update** | Timestamp (w/o timezone) | Date the entry was last updated in the database. |

- film_category

| Column | Data Type | Description |
|---|---|---|
| **film_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier for the film (number) |
| **category_id** | Small Integer(2-byte, values between -32,768 and 32,767) | Unique identifier of the category (number) |
| **last_update** | Timestamp (w/o timezone) | Date the entry was last updated in the database. |

- category

| Column | Data Type | Description |
|---|---|---|
| **category_id** | Serial | Unique identifier for the category of the film (number) |
| **name** | Character Variable (Max 45 characters) | Full name of the film category/genre |
| **last_update** | Timestamp (w/o timezone) | Date the entry was last updated in the database. |

**Step 4: Find Information**

- **Which tables you'd need to answer: which actors brought Rockbuster the most revenue?**
  The film_actor, actor, and film tables would provide the actors by revenue generated because the records in the film_actor and actor tables are linked by the actor_id key while the records in the film table, which contains the rental rate and rental duration metrics are linked to the records in the film_actor table by the film_id key. As the film table contains the rental rate and rental duration of each film, the total revenue generated by film can be calculated from the film table and subsequently connected to the actor_id in the film_actor table and the associated actor name in the actor table.

- **Which tables you'd need to answer: What language are the majority of movies in the collection?**
  The film and language table would tell you the language of the majority of movies in the collection. In the film table is the entire count of all films in the collection and the language table is connected to the film table via the language_id key.