Data Engineering Diploma Program

SS
SanyaSyed
sanyashireen@gmail.com
Programs Settings
Sign Out

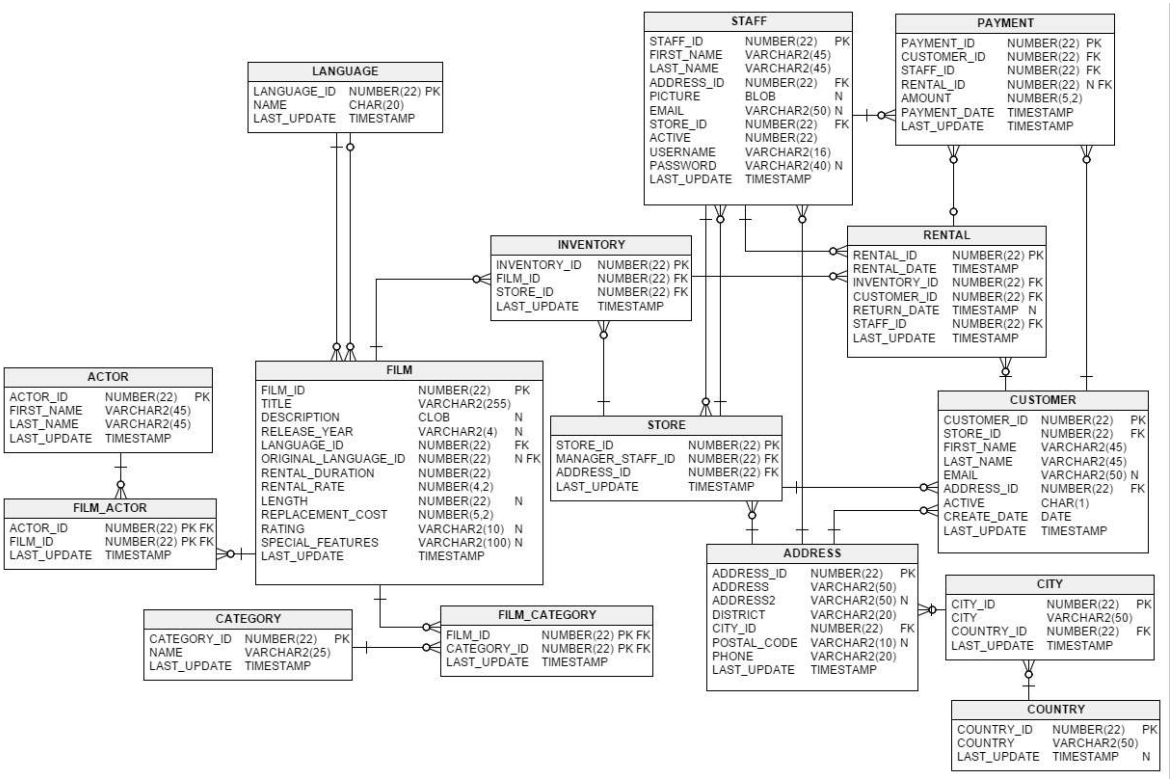
1. Project Description

In the project, you are going:
Step 1: create a data model for a data warehouse.
Step 2: write the ETL script to transform the data from the original tables to the data model you create.

(Be aware that this dataset maybe not the best dataset for the requirements, but the main purpose of this project is to practice the idea of data modeling and create DDL and DML.)

2. About the data

The dataset is the Sakila database about online DVD store. The database relationship in the original database will be like this:



This is a typical OLTP dataset for operational system. The dataset contains several tables, you first need to tell which tables can be used for dimension tables and which tables can be used for fact table.

3. Business Requirements

In this project, you need to make a data model to meet the following requirements from the managemental team:

1. List the total revenue of each store everyday.
2. List the total revenue of totally everyday.
3. List the top store according to their weekly revenue every week.
4. List top sales clerk who have the most sales each day/week/month.
5. Which film is the top film each week/month in each store/totally?
6. Who are our top 10 customers each month/year?
7. Is there any store the sales is in a decline trend (within the recent 4 weeks the avg sales of each week is declining)

Our target is to create a data model make end user can easily query simply with SELECT, GROUP BY, JOIN

4. Project Steps

1. **Load original data**: You will use this script to load the dataset into snowflake data warehouse.(it is better to use DBeaver to load data than putting the script into Snowflake console).
2. **Analyze the Business Requirements and translate the requirements into the technical requirements (rough formulas).**
3. **Consider the Grain of the data model based on your analyzation of the above Requirements.** Consider the atomic row of a fact table.
4. **Decide what dimension tables you will use in the data model.** Also consider what columns you will include into the dimensions where are these columns original from (which original tables).
5. **Define your fact table.** Consider what columns will be included in your fact table and how to get them.

6. **Create the dimension and fact tables in the Snowflake data warehouse.** Make sure your data model is in a different schema.

7. **Write ETL script to transform data from the original tables to the target tables in the data model**. If you have problem in this step, please finish it after the **Lecture -- ETL**.

5 Help

If you have big challenge to finish such project, you can find help from HERE. In your data modelling, you may need the calendar dimension table, this script is the the script for you to create calendar table.