



Models and Systems for Big Data Management

Relational Databases & Structured Query Language

The purpose of this practical work is to work with a relational database. We will use a dataset that consists of a collection of movies we will store in a relational database managed by PostgreSQL¹ server. SQL Developer² will be used as a client to connect to PostgreSQL server and to submit sql queries.

1 POSTGRESQL AND SQL DEVELOPER

1. The Postgresql server is already running and waiting for client connection requests on the port number 5432.
2. Launch SQL Developer client from a terminal using the command `sqldeveloper`, open preferences in ORACLE SQL Developer tools bar, then in database section choose third party JDBC driver, click “Add Entry” button, pick the PostgreSQL JDBC Driver (to download first `postgresql-42.2.2.jar`),³ then click OK to make it enabled.
3. Now you can create a connection to Postgresql server to access to postgres database by specifying:
 - a) a name for the connection,
 - b) `localhost` as the location of the service,
 - c) `postgres` as user and password.

The work sheet enables you to execute SQL queries by running all the script or running only the selected statement. The script output displays results.

¹<https://www.postgresql.org>

²<http://www.oracle.com/technetwork/developer-tools/sql-developer/overview/index.html>

³PostgreSQL JDBC 4.2 Driver, 42.2.2 <https://jdbc.postgresql.org/download.html>

2 DOWNLOAD DATA, CREATE SCHEMA, INSERT DATA

1. Execute the following command to import data from `artists_movies.sql` file (available on Edunao) in the database in the work sheet window: `@/PATH/artists_movies.sql` where `PATH` is the path to access to the downloaded data files.

2. Which tables are created, **how many tuples** in each table? What is the meaning of each tuple? What about the 1st, 2nd and 3th normal forms.

3. Execute the following SQL queries in the following order to create new tables and to insert data:

- a) Create movies table:

```
CREATE TABLE movies(  
    id VARCHAR(8) PRIMARY KEY,  
    title VARCHAR(35),  
    year INT,  
    genre VARCHAR(15),  
    summary VARCHAR(4406),  
    country VARCHAR(3)  
);
```

- b) Insert movies referenced in movies table from moviesReferences table:

```
INSERT INTO movies (  
    SELECT id, title , year , genre , summary , country  
    FROM moviesreferences  
    GROUP BY id, title , year , genre , summary , country  
);
```

- c) Create movies_directors table:

```
CREATE TABLE movies_directors (  
    director_id VARCHAR(10), movie_id VARCHAR(8) ,  
    PRIMARY KEY(director_id, movie_id),  
    FOREIGN KEY(director_id ) references artists(id),  
    FOREIGN KEY(movie_id ) references movies (id)  
);
```

- d) Insert data in movies_directors table using directors referenced in moviesreferences table:

```
INSERT INTO movies_directors (  
    SELECT distinct director_id, id FROM moviesreferences  
);
```

- e) Create movies_actors table:

```
CREATE TABLE movies_actors (  
    actor_id VARCHAR(10), actor_role VARCHAR(30), movie_id VARCHAR(8) ,
```

```

        unique (actor_id, movie_id),
        foreign key(actor_id ) references artists(id),
        foreign key(movie_id ) references movies (id)
    );

```

- f) Insert data in movies_actors table using actors referenced in moviesReferences table:

```

INSERT INTO movies_actors (
    SELECT DISTINCT actors_id, actors_role, id FROM moviesreferences
);

```

- g) Remove moviesReferences table:

```

DROP TABLE moviesReferences;

```

4. Analyse the final schema/tables and give the underlying Entity Association model. What about the 1st, 2nd and 3th normal forms.

3 QUERYING DATA

Write the following SQL queries:

1. What are the genres of the movies? how many distinct genres of movies?
2. What are the titles and the years of French movies released from 1960 to 2010?
3. What are the names and the roles of the actors who played in 'The Dark Knight Rises' movie? Give at least two different equivalent queries (one of them should use a sub-query).
4. What are the names and the roles of the actors who directed and played in a movie (the same movie)?
5. What is the number of movies by country? by country and year. Retrieve the answers in different kind of orders.
6. What is the number of movies by artist? First, retrieve only the identifier's artist then the first/last name's artist.
7. What is the name of the actor who played in the most of movies?
8. What are the artists who are not associated to any movie as an actor? as a director? as neither?

4 ENRICHING DATA

Insert new movies and artists to enrich your database using the following Wikipedia links:

https://en.wikipedia.org/wiki/The_Notebook

[https://fr.wikipedia.org/wiki/Scarface_\(film,_1983\)](https://fr.wikipedia.org/wiki/Scarface_(film,_1983))

[https://en.wikipedia.org/wiki/Scarface_\(1983_film\)](https://en.wikipedia.org/wiki/Scarface_(1983_film))

5 CHANGING THE SCHEMA

Complete the EA model to take into account users who assign a score to a movie (integer between 1 and 5). Give the corresponding SQL schema.