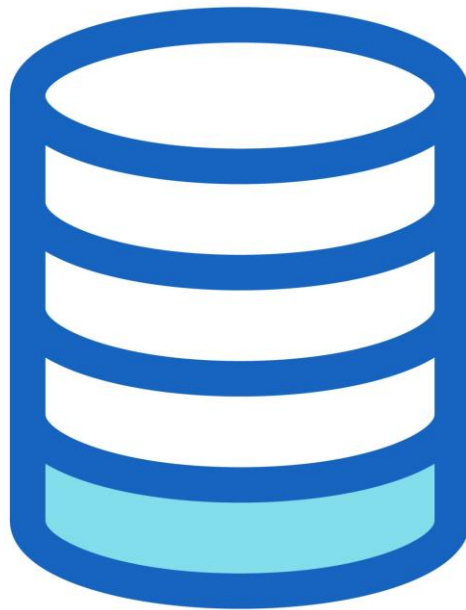


Databases - 5CC507

Anastasios-Nektarios Kefalas

Assignment Project Scenario 2022-2023

Student ID : Reducted



SCENARIO

A research team has asked you to create a database for a project on movie production companies; the project aims to use machine learning, neural networks and other methods to extract information about the situation of movie production companies in Europe and the health of this sector for a set of specific countries, including the UK. The data analytics application resulting from this project – which you DO NOT have to develop; your job is to develop the central, server-side database that underpins it – has been commissioned by a research institute (which shall remain nameless), and it is intended to be open source, and therefore available to anyone.

Basically, it is a machine learning application that would run on a database with the aim to identify the correlation between different aspects of the sector, including funding opportunities and development of new production companies or studios.

The database records every production company in Europe, including the name of the company, the address, ZIP code, city, country, type of the company (e.g., non-profit organisation), number of employees and net worth (calculated as total assets minus total liabilities). Every production company has its name registered with one and only one local government authority (for example, Companies House in the UK) on a specific date; each company can have many shareholders. The authority typically requires information about all the shareholders, including town of birth, mother's maiden name, father's first name, their personal telephone number (only one), national insurance number (each country in Europe has a similar unique ID), and passport number. Also, the registration procedure has a cost associated with it (e.g., 12£ in the UK).

The database also records the employees' data for each company: each employee is assumed to work for a single production company. Due to the complex structure of movie production companies and the need for various skills and professions, employees are categorised into crew and staff. The crew consists of three main groups: the actors, the director(s) and those who work on other jobs relevant to the filming (producers, editors, production designers, costume designers, composer, etc.). All other employees belong to the staff group, including those responsible for HR, advertising, etc. Employees are identified by an employee ID, first name, last name and an optional middle name, date of birth and start date. Also, each employee has their contact details recorded, whether it is a single phone number or multiple, with a description associated with each of them. Each employee has a single email address, too.

Members of the crew are paid hourly, and this is recorded in the database as well as a bonus that depends on their contract. Actors get a bonus for each day of work and another bonus for each scene completed; directors get a bonus at the end of the shooting; crew members that work in other jobs relevant to the filming get a bonus at the end of the shooting, and they have their role recorded as well (e.g., producer or costume designer).

Staff members have the monthly salary and the working hours (e.g., full time 9-5). Furthermore, each staff member belongs to a specific department (e.g., advertising), which is located in a given building at a given address (both recorded in the database). The database records all movies from each production company. More specifically, for each movie the following information is recorded: a universal unique

movie code (similar to the ISBN for books), the title of the movie, the year and the first release date (different release dates are not important and should NOT be recorded).

Also, the database records each member of the crew that is part of the movie, and the role they have in the movie: each crew member can play a single role or multiple roles in the same movie, and each role has a description associated with it. For example, in each movie there can be a single protagonist or more than one, the same actor can play one or several roles, or even have a cameo.

One of the aims of the project is to provide insights on the impact of funding and grants within the movie industry. To this end, the database should be able to record all the funding that each production company receives. This must include the name of the grant, the funding body (e.g., the government of a given country or European Union grants such as the ERDF), the maximum amount for that grant and the deadline to submit a proposal.

Then, for each company the database must record the date of the application to a given grant, the amount requested, the outcome (successful/unsuccessful).

A grant can be given to a single production company or shared among several. Finally, once the database is ready, the project will run a set of machine learning algorithms to perform high level data analysis based on the different grants and their corresponding impact with the aim to investigate the impacts of such funding against a list of criteria. No additional information is provided at this stage from the project.

REQUIREMENTS

In the spec, the requirements are numerated from 1 to 5, as the scenario was not given at that time. The details of each requirement are provided in the following:

1. Each production company may have received one or multiple grants, and grants can be shared by more than one company.
2. It is possible for each employee to have more than one telephone number. Each telephone number has a description associated with it (e.g., personal, or work).
3. Each production company is registered only once but can have many shareholders.
4. Each employee can either be a member of the crew OR a staff member. Each crew member can be an actor OR a director OR have another role. Each staff member belongs to a department. No duplication of data is allowed.
5. Each crew member may be part of one or more movies in a single role or many.

Task 1 : ER Diagram

Entity Relationship Diagram (before Normalization)

The entity relationship diagram (Figure 1) describes the entities *Production Company*, *Employee*, *Shareholders*, *Grant*, *Government Authority*, *Application*, *Staff Member*, *Crew Member* and *Movie*.

The primary and foreign keys of each entity are underlined(dotted for foreign keys).

And the relationship types are also included.

Please refer to the “*Figure1-ERD Before Normalization.png*” picture I have in my assignment files for a better resolution.

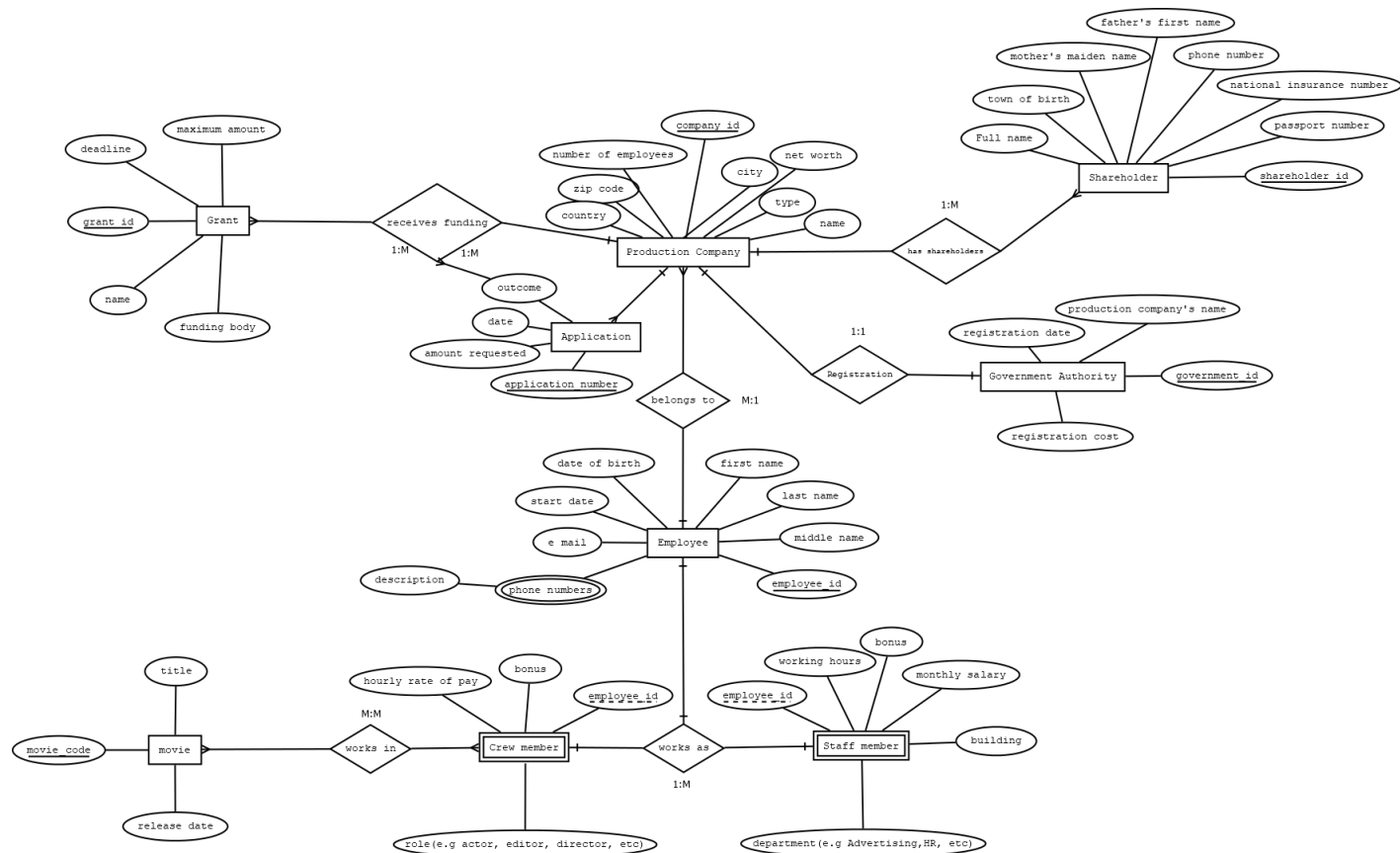


Figure1-ERD Before Normalizationigure.png : E/R Diagram Before Normalization

Relationship Between Entities

- Registration **1:1** , Connects a production company to the local government authority that is registered with.
- Has Shareholders **1:M** , Connects a production company to its shareholders.
- A company receives funding through the Application entity, so if the outcome is successful, we can continue with receiving the grant. The relationship between The Production Company and the Application **1:M** because a lot of applications can be made to receive funding. The Production Company also has a **1:M** relationship through receives funding with the Grant entity because a Production Company can receive multiple grants.
- Belongs to **M:1** , Connects an employee to the production company they work for
- Works as **1:M** , Connects Crew members and Staff members to the Employee entity
- Works in **M:M** , Connects Crew Members to the Movie entity.

Normalization Process

1NF Normalization

Step 1 : Attribute Identification

Production Company(company_id, name, country, city, zip code, number of employees, type, net worth)

Shareholder(shareholder_id, full name, father's first name, mother's maiden name, town of birth, phone number, passport number, national insurance number)

Application(application_number, amount requested, date, outcome)

Government Authority(government_id, registration date, registration cost, production company's name)

Grant(grant_id, name, funding body, maximum amount, deadline)

Movie(movie_code, title, release date)

Employee*(employee_id, first name, last name, middle name, date of birth, start date, phones numbers with description, e-mail)

Crew Member*(employee_id, role, hourly rate of pay, bonus)

Staff Member*(employee_id, department, building, monthly salary, working hours, bonus)

Step 2 : Primary Key Identification

- I created a company_id Primary Key to identify the Production Companies.
- I created a shareholder_id Primary Key to identify the Shareholders.
- I created a government_id Primary Key to identify Government Authorities.
- I created a grant_id Primary Key to identify the Grants.
- I create an application_number Primary Key to identify the Applications.
- I used employee_id as specified in the scenario we were given as Primary key for Employees and used it as Foreign Key for Crew members and Staff members.
- I used the movie_code as specified in the scenario we were given as Primary Key for Movies.

Step 3 : Removal of repeating groups

The name attribute was repeated so I changed all name attributes accordingly and made various changes on attribute names to make them less prone to repetition.

Step 4 : Made the attributes atomic

As you can see in Figure 2 on the next page, the ER Diagram is now 1NF Normalized. Please refer to the “*Figure2 - ERD After 1NF Normalization.png*” picture I have in my assignment files for a better resolution.

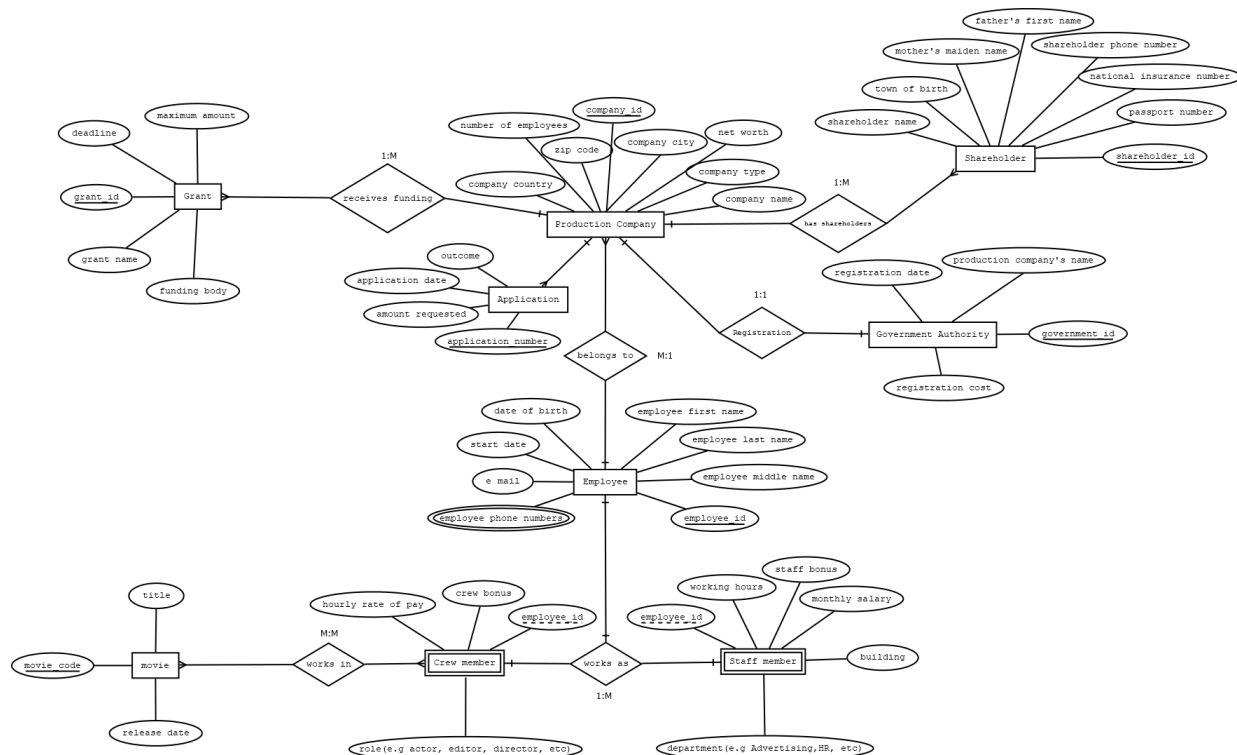


Figure2 - ERD After 1NF Normalization.png : E/R Diagram After 1NF Normalization

2NF & 3NF Normalization

To be in second normal form, a relation must be in first normal form and relation must not contain any partial dependency.

After resolving all of the multi-valued attributes we need to resolve the partial dependencies i.e.

To make all of the entities into 2nd NF I applied single keyed prime attributes.

Then in order to make all of the entities into 3rd NF I had to remove the transitive dependencies.

To remove all such transitive dependencies, I had to dissolve all the tables and after normalization I created the Final ER Diagram(Figure 3). Please refer to the “Figure3 - ERD After 2NF&3NF Normalization.png” picture I have in my assignment files for a better resolution.

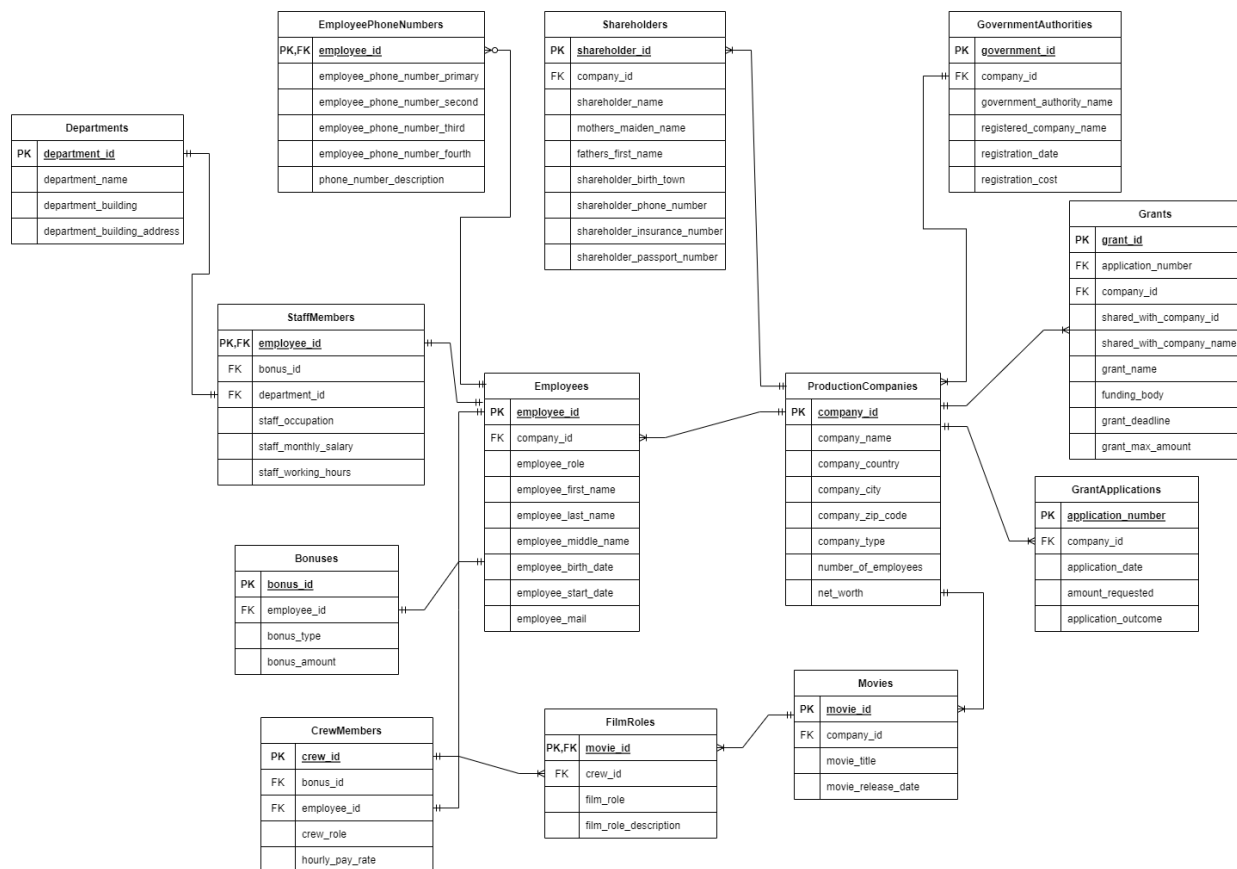
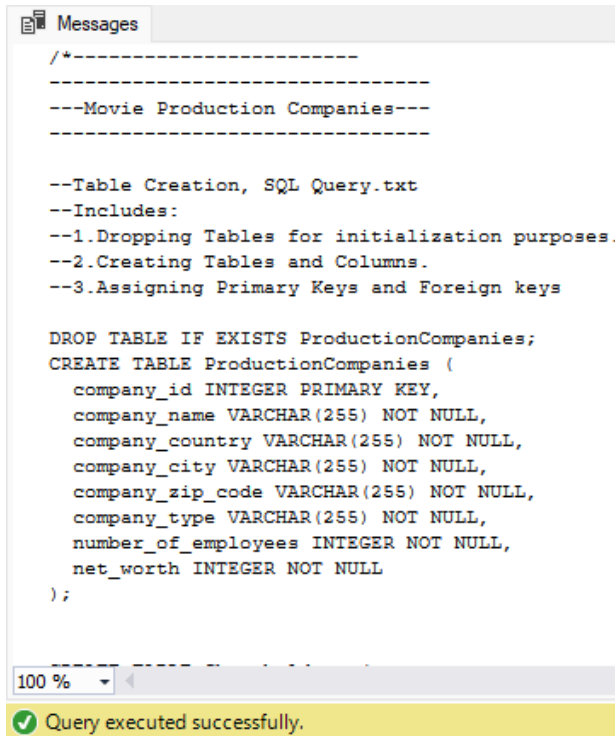


Figure3 - ERD After 2NF&3NF Normalization.png : Final E/R Diagram After 2NF & 3NF Normalization

Task 2 : SQL Script

- After creating a “MovieProductionCompanies” database, I created the tables. The query can be found in the *Task 2 [SQL Script]* folder.

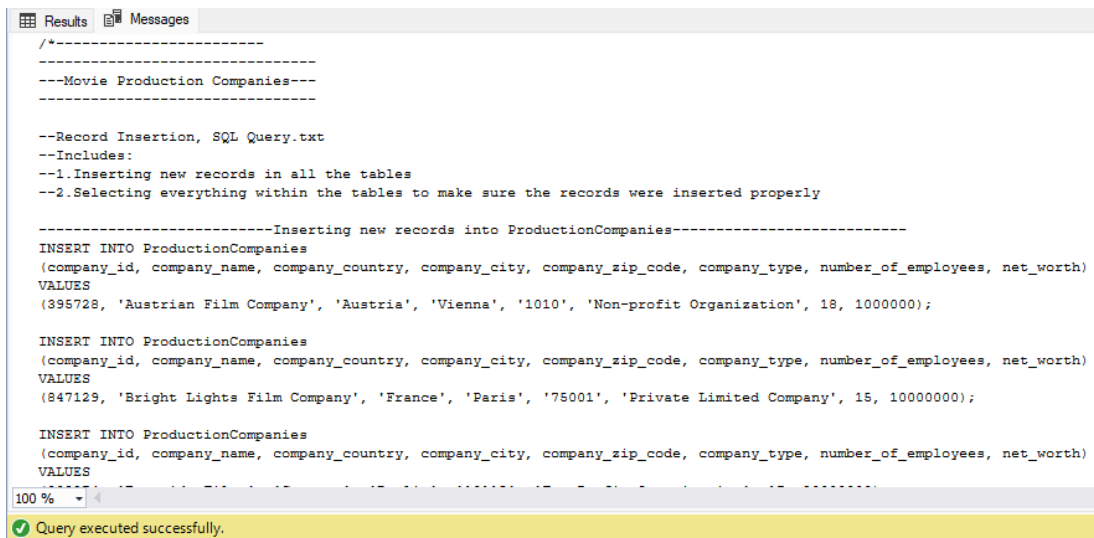


The screenshot shows the SQL Server Enterprise Manager interface. The 'Messages' pane is active, displaying the execution of a SQL script. The script includes comments and a CREATE TABLE statement for 'ProductionCompanies'.

```
/*-----  
-----  
---Movie Production Companies---  
-----  
  
--Table Creation, SQL Query.txt  
--Includes:  
--1.Dropping Tables for initialization purposes.  
--2.Creating Tables and Columns.  
--3.Assigning Primary Keys and Foreign keys  
  
DROP TABLE IF EXISTS ProductionCompanies;  
CREATE TABLE ProductionCompanies (  
    company_id INTEGER PRIMARY KEY,  
    company_name VARCHAR(255) NOT NULL,  
    company_country VARCHAR(255) NOT NULL,  
    company_city VARCHAR(255) NOT NULL,  
    company_zip_code VARCHAR(255) NOT NULL,  
    company_type VARCHAR(255) NOT NULL,  
    number_of_employees INTEGER NOT NULL,  
    net_worth INTEGER NOT NULL  
);
```

100 %
✓ Query executed successfully.

- I then started to insert records in the tables. The query can be found in the *Task 2 [SQL Script]* folder.



The screenshot shows the SQL Server Enterprise Manager interface. The 'Messages' pane is active, displaying the execution of a SQL script. The script includes comments and INSERT statements for the 'ProductionCompanies' table.

```
/*-----  
-----  
---Movie Production Companies---  
-----  
  
--Record Insertion, SQL Query.txt  
--Includes:  
--1.Inserting new records in all the tables  
--2.Selecting everything within the tables to make sure the records were inserted properly  
  
-----Inserting new records into ProductionCompanies-----  
INSERT INTO ProductionCompanies  
(company_id, company_name, company_country, company_city, company_zip_code, company_type, number_of_employees, net_worth)  
VALUES  
(395728, 'Austrian Film Company', 'Austria', 'Vienna', '1010', 'Non-profit Organization', 18, 1000000);  
  
INSERT INTO ProductionCompanies  
(company_id, company_name, company_country, company_city, company_zip_code, company_type, number_of_employees, net_worth)  
VALUES  
(847129, 'Bright Lights Film Company', 'France', 'Paris', '75001', 'Private Limited Company', 15, 1000000);  
  
INSERT INTO ProductionCompanies  
(company_id, company_name, company_country, company_city, company_zip_code, company_type, number_of_employees, net_worth)  
VALUES  
(395728, 'Austrian Film Company', 'Austria', 'Vienna', '1010', 'Non-profit Organization', 18, 1000000);
```

100 %
✓ Query executed successfully.

Task 3 : Meeting the requirements

Requirement	What I Did
1.Each production company may have received one or multiple grants, and grants can be shared by more than one company.	In the Grants table that I created, there are multiple companies that received more than one grant, in the same table I made one grant shared between two companies by creating two columns, one for an extra company id and a name that corresponds to the extra company id.
2.It is possible for each employee to have more than one telephone number. Each telephone number has a description associated with it (e.g., personal, or work).	I created a separate EmployeePhoneNumbers table that connects to the Employees table with employee_id as the Foreign Key. That allowed me to assign multiple phone numbers to one person and add a description for each phone number as well.
3.Each production company is registered only once but can have many shareholders.	I created a Shareholders table that connects to the ProductionCompanies table with company_id as the Foreign Key, which allowed me to link multiple shareholders to one company. I also created a GovernmentAuthorities table that connects to the ProductionCompanies table with company_id as the Foreign Key. Each Company is registered once under its government authority.
4.Each employee can either be a member of the crew OR a staff member. Each crew member can be an actor OR a director OR have another role. Each staff member belongs to a department. No duplication of data is allowed.	I created a Staff Members table and a Crew Members table that connect to the Employees table with employee_id as the Foreign Key, which allowed me to separate them and assign roles to the crew. I break the Staff role further to HR,Advertising and Janitor in the Departments table.
5.Each crew member may be part of one or more movies in a single role or many.	I created a FilmRoles table that connects to the Movies table with movie_id as the Foreign Key in order to give actors their role and a description of that role. I gave one of the movies a sequel so I can make some actors part of more than one movie.

1.

```
SELECT company_id, shared_with_company_id, shared_with_company_name  
FROM Grants;
```

2.

```
SELECT *  
FROM EmployeePhoneNumbers;
```

3.

```
SELECT GovernmentAuthorities.government_id,  
GovernmentAuthorities.government_authority_name,  
GovernmentAuthorities.registered_company_name,  
Shareholders.shareholder_id, Shareholders.shareholder_name  
FROM GovernmentAuthorities  
INNER JOIN Shareholders ON Shareholders.company_id =  
GovernmentAuthorities.company_id;
```

4.

```
SELECT employee_role  
FROM Employees;
```

```
SELECT *  
FROM Departments;
```

5.

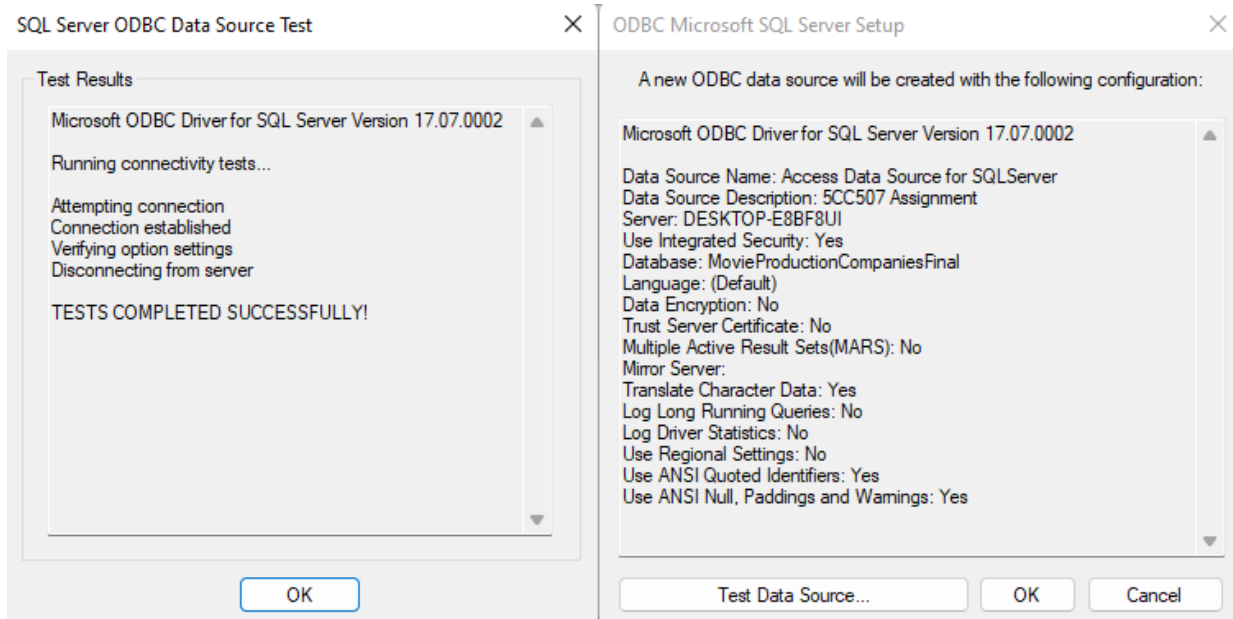
```
SELECT *  
FROM FilmRoles;
```

SQL Query that tests the database's functionality.txt Contains the above statements to test the database's functionality.

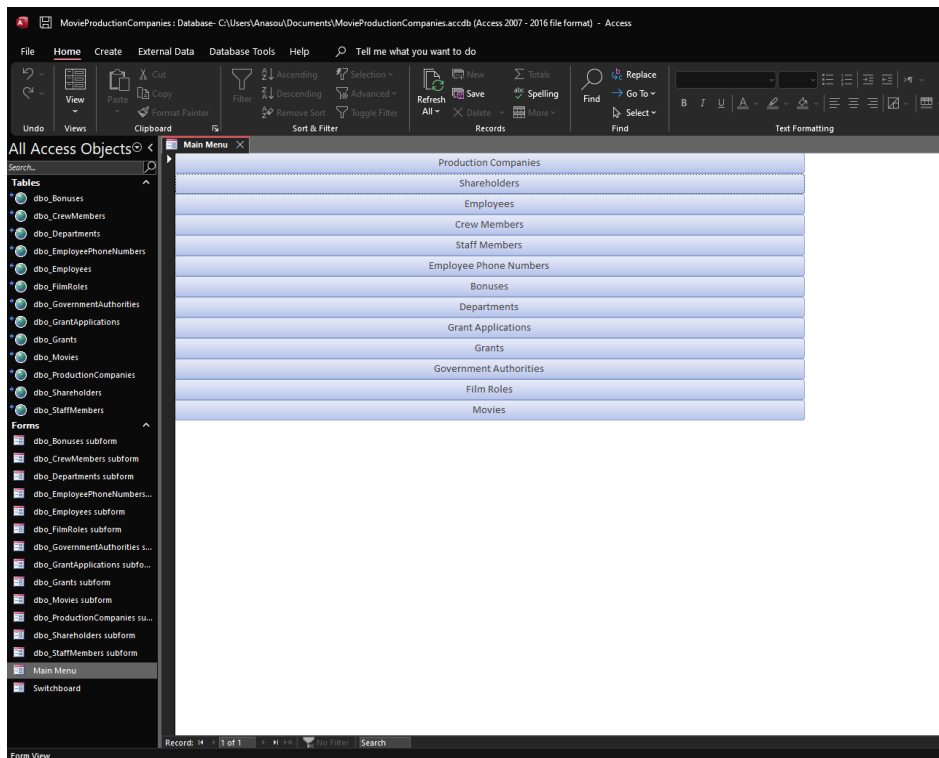
5CC507-Databases

Task 4 : Front-End in MS Access

ODBC :



Main Menu :



A Few Table Examples :

All Access Objects < Main Menu × dbo_ProductionCompanies subform ×

Search...

Tables

- dbo_Bonuses
- dbo_CrewMembers
- dbo_Departments
- dbo_EmployeePhoneNumbers
- dbo_Employees
- dbo_FilmRoles
- dbo_GovernmentAuthorities
- dbo_GrantApplications
- dbo_Grants
- dbo_Movies
- dbo_ProductionCompanies
- dbo_Shareholders
- dbo_StaffMembers

Forms

- dbo_Bonuses subform
- dbo_CrewMembers subform
- dbo_Departments subform
- dbo_EmployeePhoneNumbers...
- dbo_Employees subform

Company ID : 217435

Name : Midnight Movies

Country : Spain

City : Madrid

Zip Code : 28001

Type : Private Limited Company

Number of Employees : 16

Net Worth : 40000000

All Access Objects < Main Menu × dbo_Employees subform ×

Search...

Tables

- dbo_Bonuses
- dbo_CrewMembers
- dbo_Departments
- dbo_EmployeePhoneNumbers
- dbo_Employees
- dbo_FilmRoles
- dbo_GovernmentAuthorities
- dbo_GrantApplications
- dbo_Grants
- dbo_Movies
- dbo_ProductionCompanies
- dbo_Shareholders
- dbo_StaffMembers

Forms

- dbo_Bonuses subform
- dbo_CrewMembers subform
- dbo_Departments subform

Employee ID : 3301

Company ID : 847129

Role : Producer

First Name : Amelie

Last Name : Rousseau

Middle Name : Lucie

Date of Birth : 1/29/1987

Start Date : 7/4/2018

E-mail Address : amelie.rousseau@brightlights.com