



# MODERN BIG DATA ANALYSIS WITH SQL SPECIALIZATION

CSE 443 - SEMINAR ON  
SUMMER TRAINNG

QAZI MAAZ ARSHAD

B.TECH CSE

11906424

# TABLE OF CONTENTS

## 01 | My Course

I will discuss my summer training MOOC. The tasks, assignments, timeline, etc.

## 03 | Learning Outcomes

I will talk about the experience and skills I have gained from the course and project.

## 02 | Project

I will explain my project in detail. The aim, problems solved, approach, tech stack, etc.

## 04 | Conclusion

I will give an overview of my training and talk over future scope of BIG DATA.

# OVERVIEW OF COURSE

# Modern Big Data Analysis wth SQL Specialization

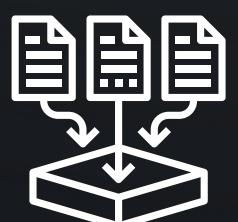
“Modern Big Data Analysis with SQL Specialization” is an online specialization course offered by CLOUDERA consisting of 3 courses “Foundations for Big Data Analysis with SQL”, “Analyzing Big Data with SQL”, and “Managing Big Data in Clusters and Cloud Storage”. This specialization teaches the essential skills for working with large-scale data using SQL.



# Foundations for Big Data Analysis with SQL



## Analyzing Big Data with SQL



## Managing Big Data in Clusters and Cloud Storage



# TIMELINE

Completed the Summer Training before the deadline -  
10th August 2021

---

Each course comprised of multiple quizzes and assignments. A minimum of 80% marks were required in most quizzes and assignments to pass the course.

---



Enrolled in MOOC  
13 May 2021



Completed the  
Specialization  
12 July 2021

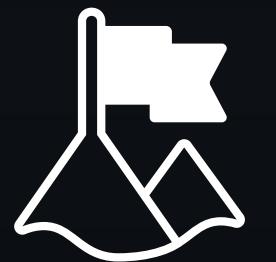
---

Will discuss project later

It took approximately 1 week to complete the project.

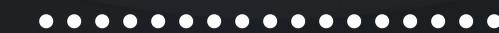


Commenced Project  
19 July 2021



Finished Project Work  
25 July 2021

THE  
**PROJECT**



# MOVIES DATA ANALYSIS

What is my project all about

## PROJECT OVERVIEW

Data analysis of a very large films database.

## TECH STACK INVOLVED

Created this project using SQL and MySQL Workbench.

## APPLICATION

This project can be used as a demo for managing other data sets.

Created a database using 4 different data sets and calculated significant results using SQL queries.

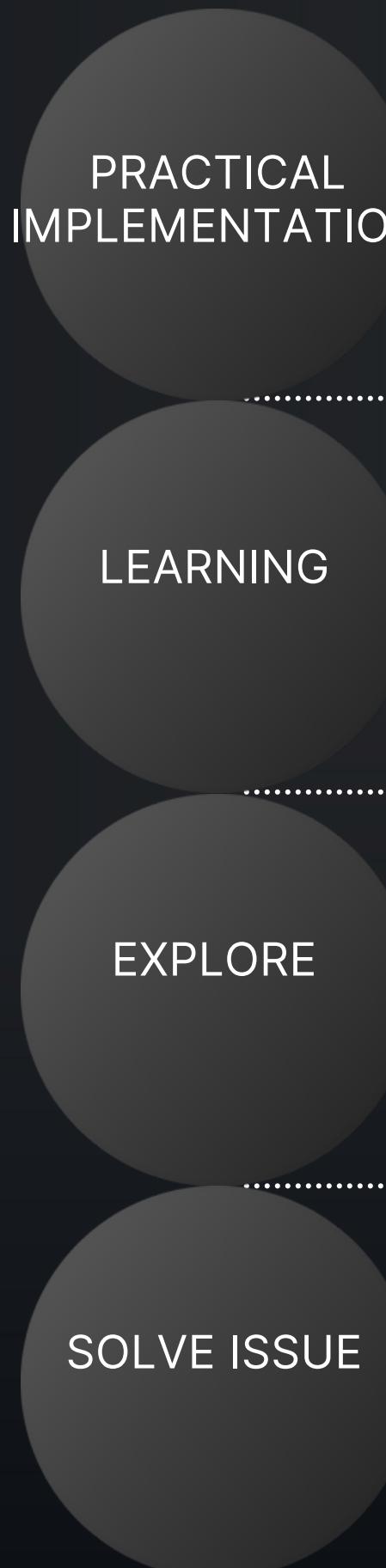
Results calculated from the database is done using SQL query operations. I have used MySQL Workbench to design and manage the database.

This project is useful in calculating results from past movies to analyze trends in the movie industries across the globe.

# WHY THIS PROJECT

---

The main aim was to implement all the knowledge and skills gained in summer training on a real project.

- 
- 1 PRACTICAL IMPLEMENTATION**

I did this project to get practical exposure of managing a large data set.
  - 2 LEARNING**

I learned a lot from this project, and revised all the learnings from my summer training.
  - 3 EXPLORE**

I got the opportunity to explore more about this domain, learn new things, and enhance existing skills.
  - 4 SOLVE ISSUE**

I learned how to manage large datasets. This project calculates important results from unstructured data.

# WHICH PROBLEMS THIS PROJECT SOLVES

---

The movie's database contains information regarding the name of the movies, the year when they were released, country of origin, duration of films, the language of films, the budget of films, etc.



ANALYZES  
UNSTRUCTURED  
DATA



INSIGHTS  
ABOUT PAST  
TRENDS IN  
FILMS  
INDUSTRY



DEDUCE  
IMPORTANT  
FACTS

---

This project will also serve as a prototype to design, manage, and study any other datasets in future.

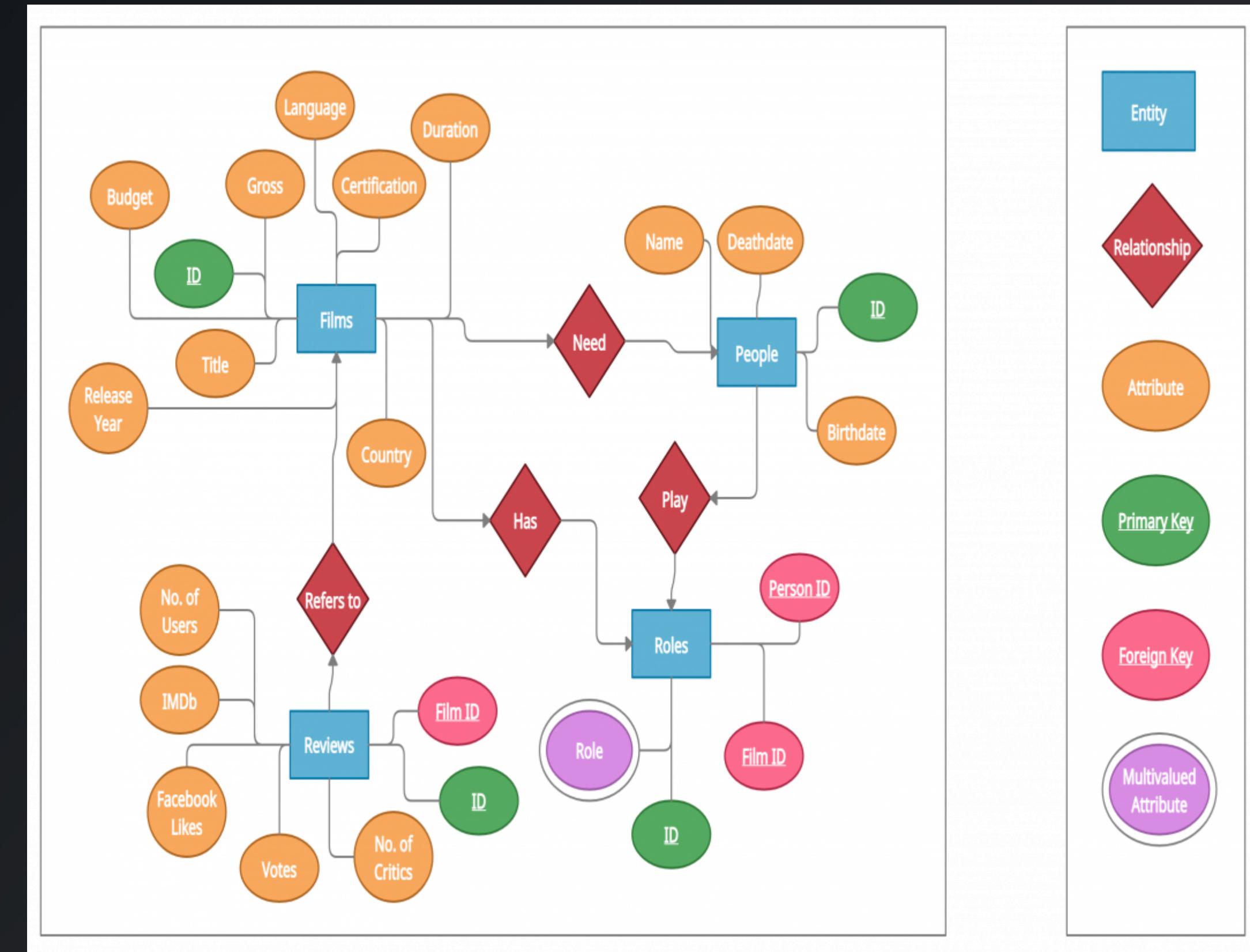
---

# PROJECT DESIGN

## ER DIAGRAM

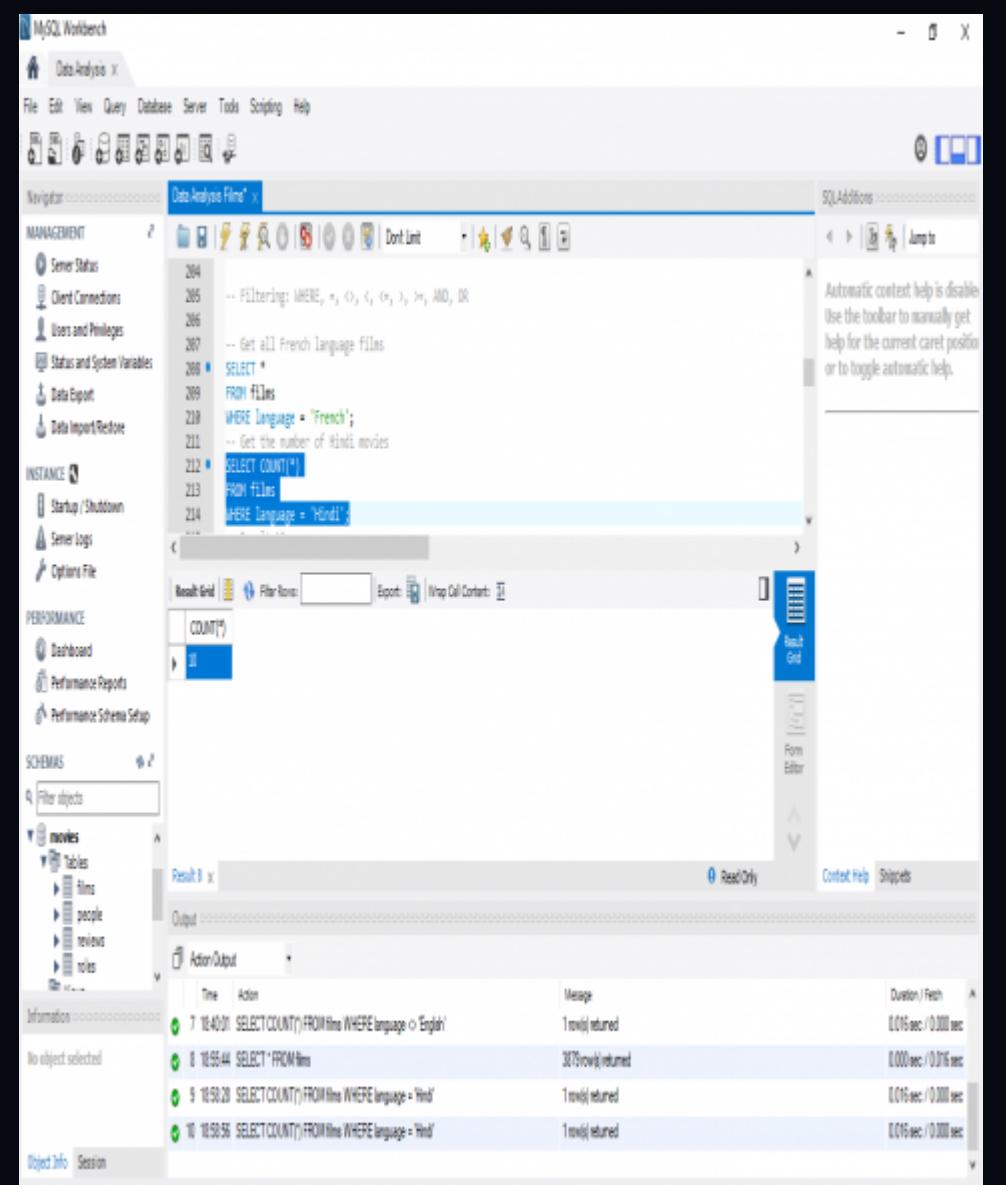
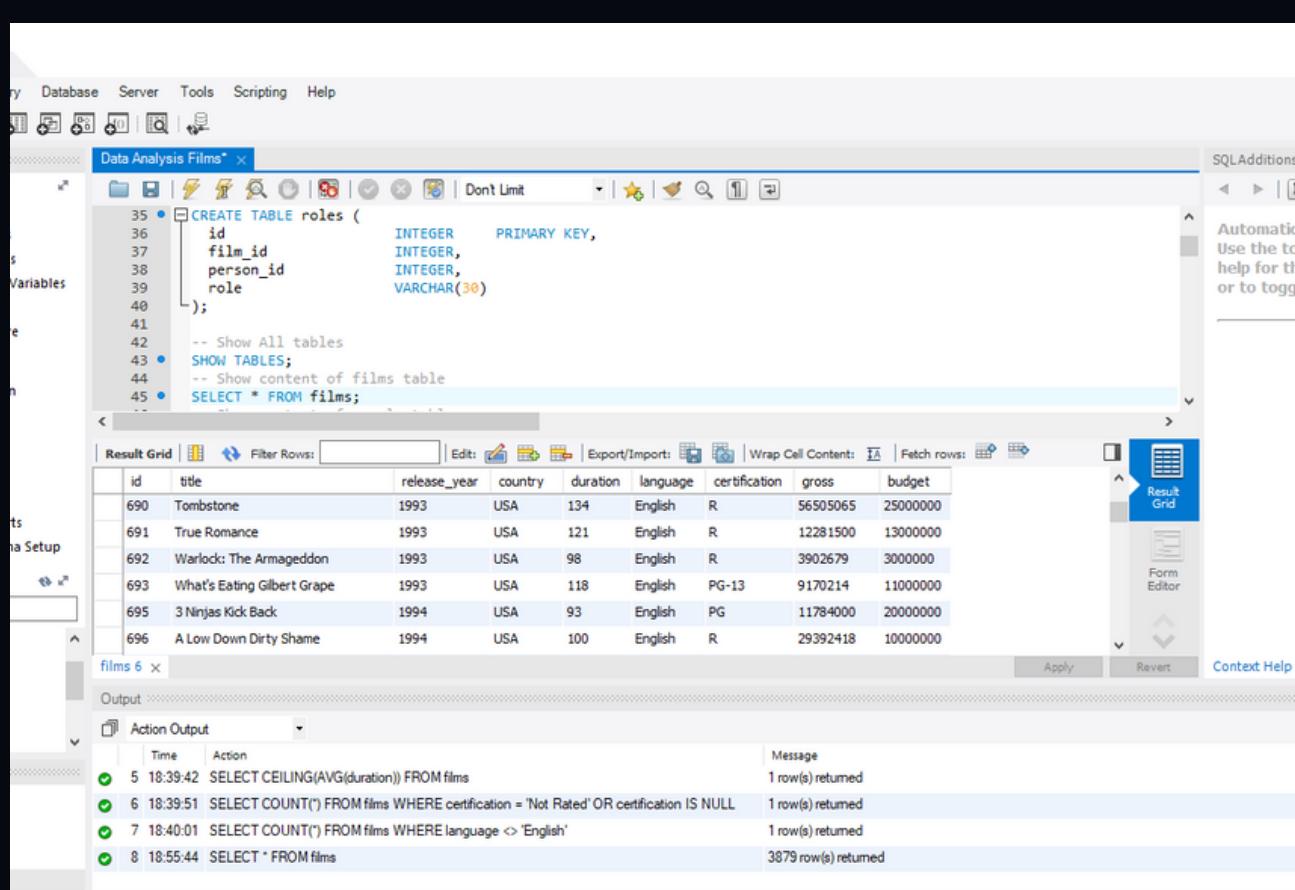
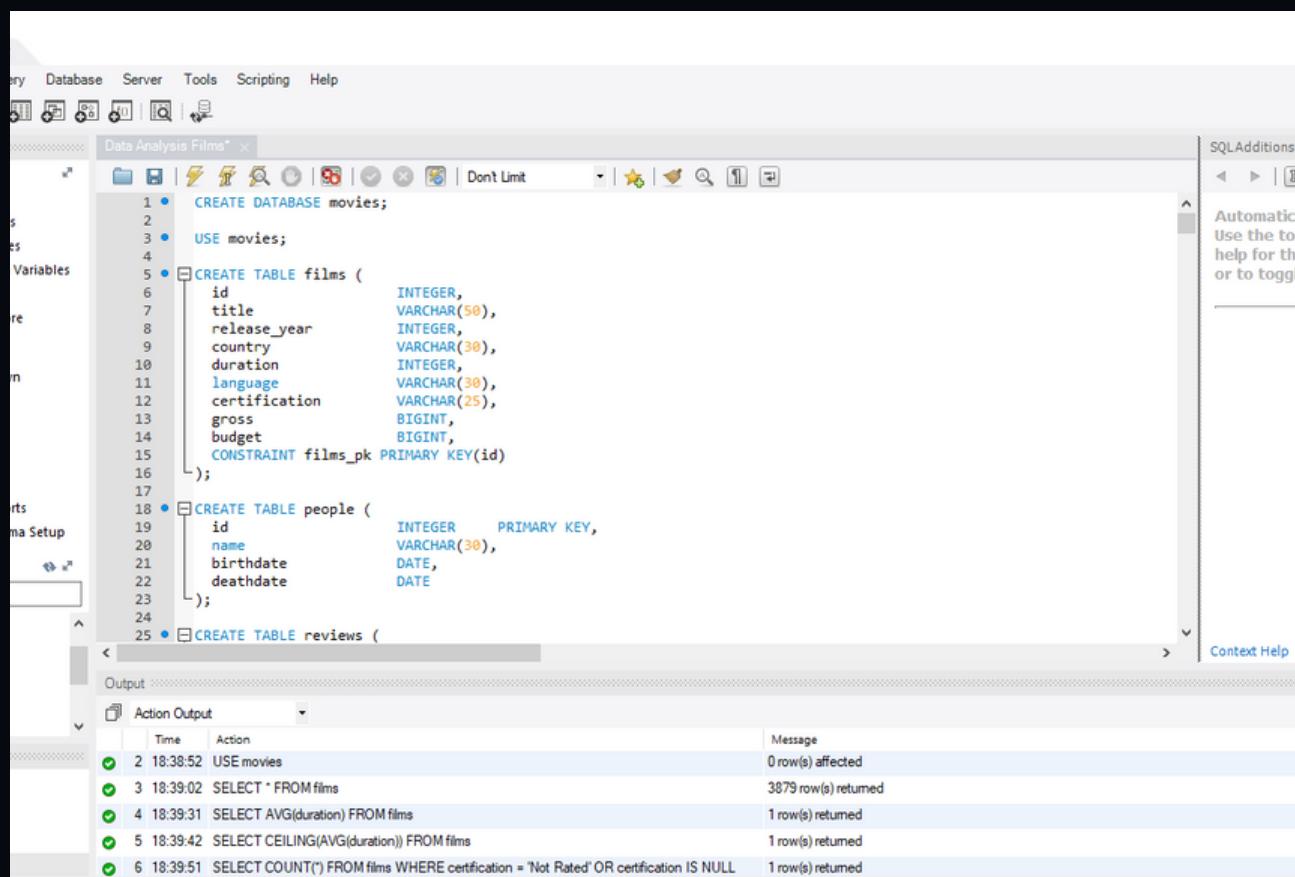
An Entity Relationship (ER) Diagram is a type of flowchart that illustrates how “entities” such as people, objects or concepts relate to each other within a system.

This ER Diagram is created using CREATELY.



# SCREEN SHOTS

These are a few screenshots of data sets, code and query results in MySQL Workbench and from GitHub repository.



37	It's a Wonderful Life	1946	USA	118	English	PG		3100000
38	The Best Years of Our Lives	1946	USA	172	English	Not Rated	2365000	2100000
39	Gentleman's Agreement	1947	USA	118	English	Approved		2000000
40	The Lady from Shanghai	1947	USA	92	English	Not Rated	7827	2300000
41	Tycoon	1947	USA	95	English	Approved		3209000
42	Open Secret	1948	USA	68	English	Approved		
43	Red River	1948	USA	126	English	Approved		3000000
44	The Pirate	1948	USA	102	English	Approved	2956000	3700000
45	Sands of Ilima	1949	USA	109	English	Approved		1000000
46	She Wore a Yellow Ribbon	1949	USA	103	English	Unrated		1600000
47	Annie Get Your Gun	1950	USA	107	English	Passed	8000000	3768765
48	A Streetcar Named Desire	1951	USA	125	English	PG		1800000
49	Quo Vadis	1951	USA	171	English	Passed		7823000
50	Show Boat	1951	USA	108	English	Approved		2295429
51	Deadline - U.S.A.	1952	USA	87	English			
52	High Noon	1952	USA	85	English	PG		750000
53	Singin' in the Rain	1952	USA	103	English	Approved		2540000
54	The Greatest Show on Earth	1952	USA	152	English	Not Rated	3600000	4000000
55	From Here to Eternity	1953	USA	118	English	Not Rated		1450000

# SWOT ANALYSIS

SWOT analysis of my project (In general of SQL) will assess the four important aspects of the project (SQL).

S

## Strength

Simple and easy in handling large data. Data analysis can be performed easily using basic commands.

W

## Weakness

SQL is good for fetching data but it cannot be used alone for visualizing data.

O

## Opportunity

We can improve project by representing results in graphs by exporting results to other tools.

T

## Threat

Python and tools like tableau will be preferred over SQL when the requirement is to visualize the data into figures.

# LEARNING OUTCOMES

---

Skills and technologies learned from the summer training and experience gained from working on real projects.



SQL



DBMS



Data Analysis



MySQL  
Workbench

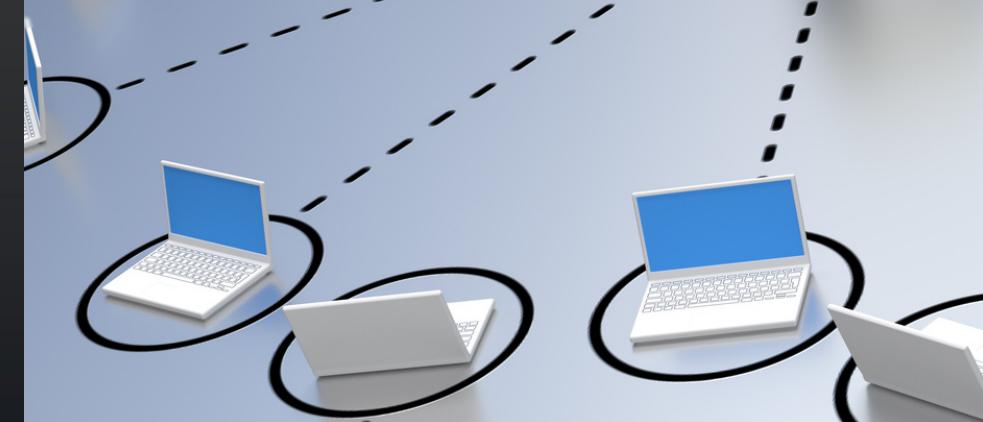
# FUTURE OF SQL DATABASES

The future of SQL Server will depend on the future of the use of SQL as a query language.



## SQL In the Past

SQL started as an IBM project in 1974 to implement a “relational model of data”.



## SQL in the Present

SQL, being the ANSI and ISO standard for relational databases, is adapting to the changing world of data transforming into big data.



## SQL in the Future

Data is increasing exponentially, and data is not going to reduce, so there will always be a need to manage data.

---

# THANK YOU

---