LOVELY PROFESSIONAL UNIVERSITY

MODERN BIG DATA ANALYSIS WITH SQL SPECIALIZATION

CSE 443 - SEMINAR ON
SUMMER TRAINNG

QAZI MAAZ ARSHAD
B.TECH CSE
11906424

TABLE OF CONTENTS

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

Enrolled in MOOC

13 May 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.

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.

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

TECH STACK INVOLVED

Created this project using SQL and MySQL Workbench.

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

APPLICATION

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

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.

PRACTICAL IMPLEMENTATION

1

I did this project to get practical exposure of managing a large data set.

LEARNING

2

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

EXPLORE

3

I got the opportunity to explore more about this domain, learn new things, and enhance existing skills.

SOLVE ISSUE

4

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

DEDUCE IMPORTANT FACTS

INSIGHTS ABOUT PAST TRENDS IN FILMS INDUSTRY

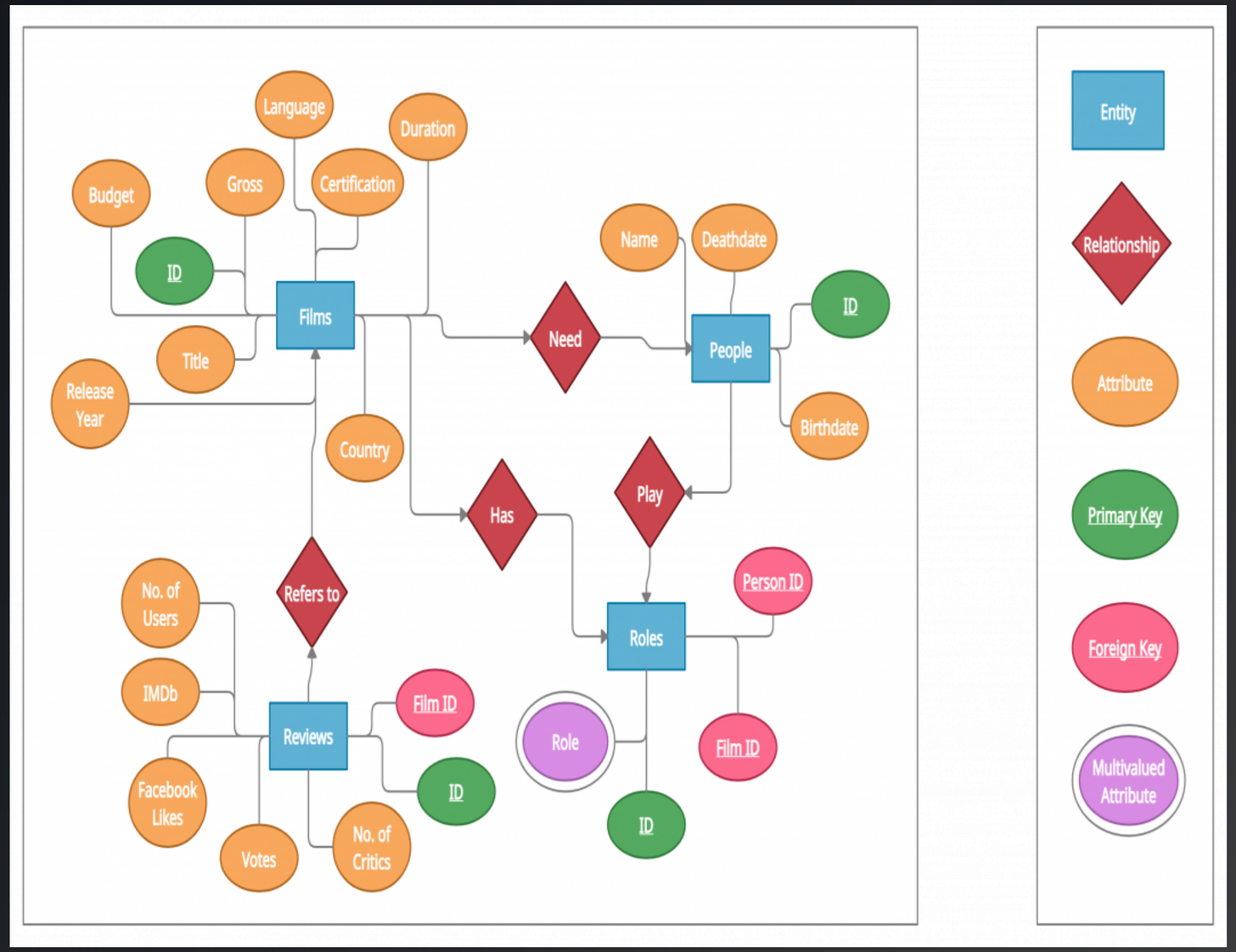
This project will also serve as a prototype to design, manage, and study any other datasets in future.

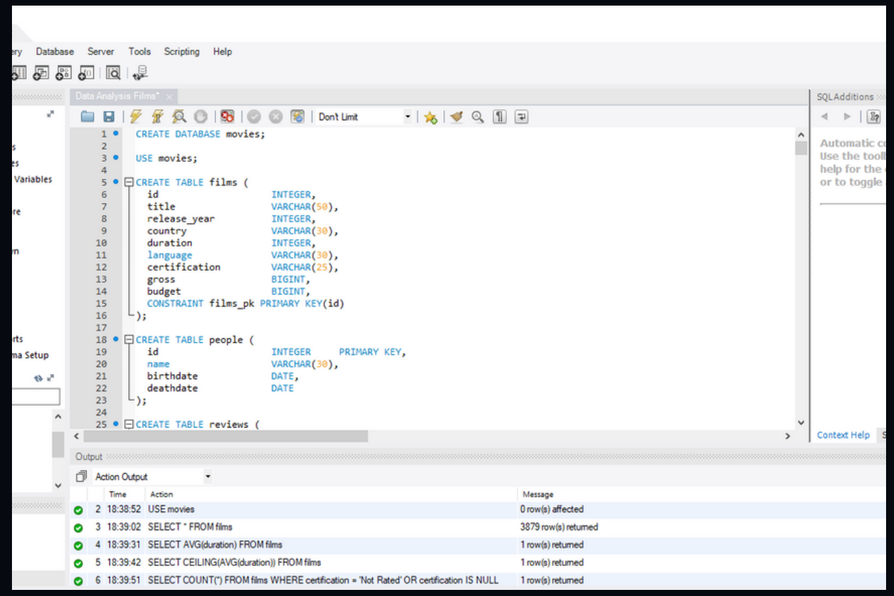
PROJECT DESIGN

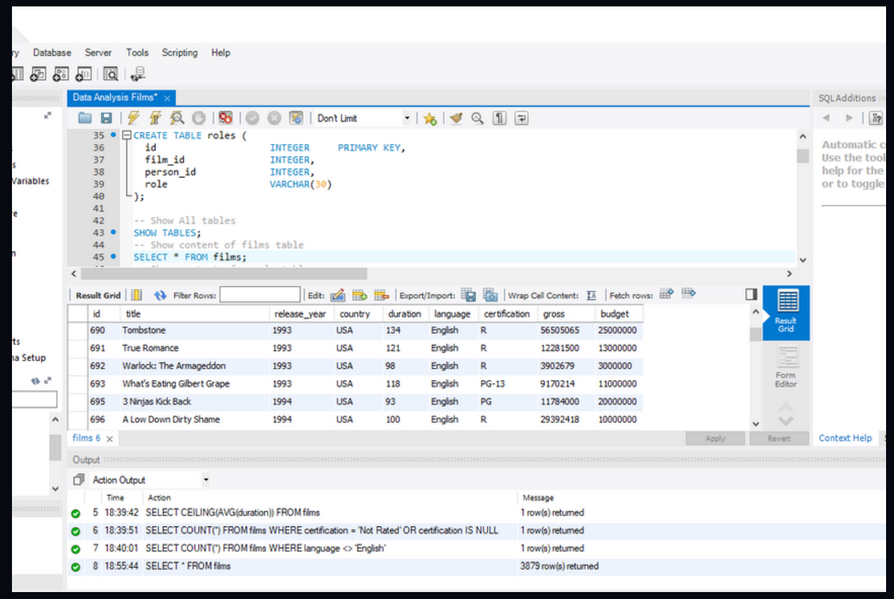
ER DIAGRAM

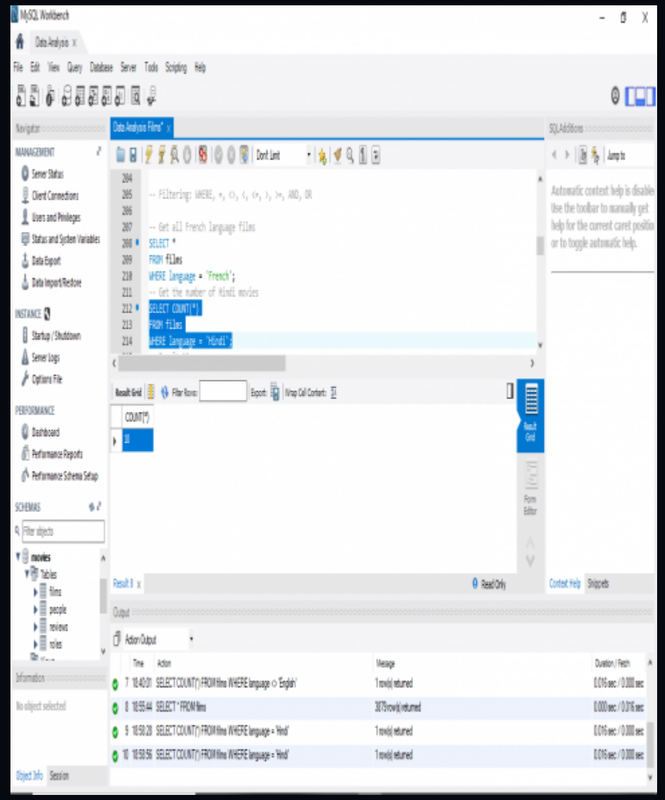
An Entity Relationship (ER) Diagram is a type of flowchart that illustrates how “entities” such as people, objects or concepts relate to each other within a system.

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.

SWOT ANALYSIS

SWOT analysis of my project(In genral 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

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