BT2102 – LAB SESSION 2

Kaustubh Jagtap – A0168820B

Schema:

CREATE DATABASE lab2;

USE lab2;

CREATE TABLE movies(serial INT, budget BIGINT, company VARCHAR(100), country VARCHAR(50), director VARCHAR(100), genre VARCHAR(50), gross BIGINT, name VARCHAR(100), rating VARCHAR(15), runtime SMALLINT, score DECIMAL(4,2), star VARCHAR (50), votes MEDIUMINT, writer VARCHAR(50), year SMALLINT, PRIMARY KEY (serial));

CREATE TABLE awards(serial INT, year SMALLINT, award VARCHAR(200), winner TINYINT, film VARCHAR(100), names VARCHAR(250), PRIMARY KEY (serial));

Note: I used the serial number as primary key for both tables, because other combinations of composite primary keys were causing lines to be skipped in the database. No information was lost in doing so, hence I felt it was fine.

Load Data:

LOAD DATA LOCAL INFILE 'C:/Users/ksjag/Documents/NUS/Sem 2/BT2102/Labs/Lab 2/movies.csv'
INTO TABLE movies
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

LOAD DATA LOCAL INFILE 'C:/Users/ksjag/Documents/NUS/Sem 2/BT2102/Labs/Lab 2/awards.csv'
INTO TABLE awards
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

Note: I had to convert the csv files to utf-8 format, since mysql was throwing an error.

Queries:

A) <u>Which awards did the movie 'The Lion in Winter' win?</u>

SELECT award FROM awards WHERE (film='The Lion in Winter' AND winner=1)

B) <u>How many awards did 'The Grand Budapest Hotel' win?</u>

SELECT count(award) FROM awards WHERE (film='The Grand Budapest Hotel' AND winner=1);

C) <u>How much did 'The Grand Budapest Hotel' cost to make?</u>

SELECT budget FROM movies WHERE (name='The Grand Budapest Hotel');

Di) <u>How many movies has Denzel Washington starred in?</u>

SELECT COUNT(*) FROM movies WHERE (star='Denzel Washington');

Dii) <u>How much did the most expensive cost to make?</u>

SELECT MAX(budget) FROM movies WHERE(star='Denzel Washington');

Dii) <u>What was its title?</u>

SELECT name FROM movies WHERE (star='Denzel Washington') AND budget = (SELECT MAX(budget) FROM movies WHERE star='Denzel Washington');

E) <u>Which movies in the movies table have won awards?</u>

SELECT DISTINCT (movies.name)
FROM movies inner join awards
ON movies.name = awards.film
WHERE awards.winner = 1;

Advanced Queries on some interesting questions:

1) <u>What's the most expensive award winning film?</u>

SELECT name FROM movies WHERE budget =
(SELECT max(movies.budget)
FROM movies INNER JOIN awards
ON movies.name = awards.film
WHERE awards.winner = 1);

2) <u>Who directed the cheapest movie to win the 'Best Picture' award?</u>

SELECT movies.director
FROM movies INNER JOIN awards
ON movies.name = awards.film
WHERE award = 'best picture' AND winner = 1 AND budget =
(SELECT MIN(budget)
FROM movies inner join awards
ON movies.name = awards.film
WHERE award = 'best picture' AND winner = 1);

3) <u>What is the most successful movie company in terms of profits?</u>

SELECT profits.company, profits.profit
FROM (SELECT company, sum(gross) – sum(budget) AS profit
FROM movies GROUP BY company) AS profits
WHERE profits.profit = (SELECT MAX(profits.profit)
FROM (SELECT sum(gross) – sum(budget) AS profit, company
FROM movies GROUP BY company) AS profits);

4) <u>What is the amount spent on movies in each country?</u>

SELECT country, sum(budget) AS amt_spent
FROM movies GROUP BY country;

5) <u>What film in the movies table won the most awards?</u>

SELECT movies.name, winners.count FROM movies INNER JOIN
(SELECT film, COUNT(film) AS count FROM awards WHERE winner=1 GROUP BY film) AS winners
ON movies.name = winners.film WHERE winners.count =
(SELECT MAX(winners.count) FROM movies INNER JOIN
(SELECT film, COUNT(film) AS count FROM awards WHERE winner=1 GROUP BY film) AS winners
ON movies.name = winners.film);