DSCI550: Data Science at Scale

Homework 4, Fall 2023

Due: 11:59 pm, Oct. 27, 2023

1. (40 pts) Draw a full ER diagram for the following requirements:

   The university database stores details about university students, courses, the semester a student took a particular course (and his mark and grade if he completed it), and what degree program each student is enrolled in. Consider the following requirements list:

   - The university offers one or more programs.
   - A program is made up of one or more courses.
   - A student must enroll in a program.
   - A student takes the courses that are part of her program.
   - A program has a name, a program identifier, the total credit points required to graduate, and the year it commenced.
   - A course has a name, a course identifier, a credit point value, and the year it commenced.
   - Students have one or more given names, a surname, a student identifier, a date of birth, and the year they first enrolled. We can treat all given names as a single object—for example, "John Paul."
   - When a student takes a course, the year and semester he attempted it are recorded. When he finishes the course, a grade (such as A or B) and a mark (such as 60 percent) are recorded.
   - Each course in a program is sequenced into a year (for example, year 1) and a semester (for example, semester 1).

2. (30 pts) In the Store_Sales table, the Item_Code and Store_ID are key attributes.

Store_ID	Street	City	ZipCode	Item_Code	Item_Description	Quantity	Unit_Price
S-100	Main St	LA	90010	I-200	Something Good	30	25
S-100	Main St	LA	90010	I-201	Something Better	22	35
S-101	3rd St	LA	90011	I-200	Something Good	54	25

   1) (5 pts) Explain if this is in 1st Normal Form.
   2) (10 pts) Explain if this is in 2nd Normal Form. Use specific attribute names in your explanation.
   3) (15 pts) If it is not in 2nd Normal Form, create tables in 2nd Normal Form.

3. (30 pts) Using the following tables in the movie database, write SQL queries. Suppose that tuples are already inserted into the tables.

- *Actor (id, fname, lname, age, gender, nationality)*
- *Movie (id, name, year, rank, revenue, studio, director_id)*
- *Director (id, fname, lname, gender, income)*

id column in ACTOR, MOVIE, and DIRECTOR tables is the key for the respective table.

1) (3 pts) List all the male actors (i.e., gender = 'M').
2) (3 pts) Find the actor whose first name is 'John' and from 'Spain'.
3) (3 pts) List first name and last name of all the actors whose nationality is 'USA'.
4) (3 pts) List the name and revenue of movies made by "Universal Studio".
5) (3 pts) List the total number of movies released in 2018.
6) (3 pts) List max income of all directors.
7) (4 pts) List the name of movies in descending order of revenue, i.e., the highest first.
8) (4 pts) List the average income of female directors.
9) (4 pts) List the names and years of the movies directed by "Jeniffer" (fname).