Homework 3: SQL

## DSCI 551 – Fall 2023

## Due: 11:59pm, Friday, October 20, 2023

## Points: 100

Consider the following tables in a course management database.

Course(number, title, semester)

Student(id, name, program)

Instructor(id, name, department)

Take(sid, cno, semester) //sid is student id, cno is course number

Teach(rid, cno, semester) // rid is instructor id

TA(sid, hours) // where hours are TA’s work hours per week, we assume that hours of a

TA are the same in different semesters

Assist(sid, cno, semester)

**Assumptions:**

1. All attributes in the above tables are of string types (char/varchar/text) except for hours of TA which are integers.
2. The homework assumes that you have created a database dsci551 on your EC2 MySQL, a user dsci551@localhost (with password Dsci-551), and grant all privileges on dsci551.\* (i.e., objects in dsci551) to the user. You can log in as root and execute the following to satisfy the assumption:

create database dsci551;

create user dsci551@localhost identified by "Dsci-551";

grant all privileges on dsci551.\* to dsci551@localhost;

1. Your code should run without error and satisfy all requirements stated in the question. It’s only required for you to define PK, FK and data type for each attribute. It’s up to you

whether to add “unique”/ “NOT NULL”/ “CHECK”/ FK CASCADE, etc.

Create your own sample database with the above schema and insert some data into the tables for your own testing. Assume that student ids are in the form of "sxxx" where xxx are 3 digits, e.g., s100, s101, etc.

Write one SQL script “create.sql” that creates all 7 tables in a MySQL

database dsci551. The script may assume that the database dsci551 already exists.

If the above tables already exist in the database, your script should be able to recreate them.

1. [70 points] Write an SQL query for each of the following questions:
   1. Find out which courses were offered in Spring 2023 but not in Fall 2023. Report course number and title.
   2. Find out titles of which courses in Fall 2023 contain “data” and are taught by Professor “John Smith”.
   3. Find out which students have taken both DSCI 351 and DSCI 250 (both are course numbers). Report student id and name.
   4. Find out which students have taken DSCI 351 but not DSCI 250. Report student id and name.
   5. Find out which students did not take any courses in Fall 2023.
   6. Find out which instructors teach the largest number of courses in Fall 2023. Report instructor id and name.
   7. Find out which instructors teach only one course in Fall 2023 without using aggregate functions in SQL. Report instructor names and course numbers.
   8. Find out for each instructor, the average number of students in his/her courses offered in Fall 2023. For example, if an instructor teaches two courses, course X with 100 students and course Y with 200 students, then the average number of students in his/her course would be 150. Report the instructor id and the average number.
   9. Find out which courses offered in Fall 2023 have only one TA. Report the course titles.
   10. Find out which TAs work for more than 15 hours a week and take 3 courses in Fall 2023. Report names of such TAs.
2. [30 points] Write a Python script "find\_courses.py" that takes a student id and output student name, program, titles and semesters of all courses taken by the student.

Execution format: python3 find\_courses.py "s100"

Note that your find\_courses.py should access the **dsci551** database (with **user = dsci551** and

**password = Dsci-551**, as mentioned earlier) in your MySQL on your localhost. You can

test your code by first uploading it to your EC2 instance and running it there.

**Permitted libraries:** You can use sqlalchemy, pymysql, pandas.

**Submission Instructions:**

1. Submit only 3 files   
   - one **.sql file** for create table queries called create.sql  
   - one **.sql file** for Q1 called Q1.sql  
   - one **.py file** for Q2 called find\_courses.py
2. Please name your files exactly as mentioned above
3. **DO NOT** submit zip file for submission (strict validation).
4. Points for each query will only be awarded if the query runs on test data or else 0 points.
5. All the 10 queries of Q1 should be present in Q1.sql file. Any code other than queries in Q1.sql should be in comments.