# Lab 3: MongoDB

## DSCI 551– Fall 2023

## Due: 11:59pm, October 27, 2023

## Points: 20 (note this lab is worth twice as much).

Create a course management database in MongoDB with the following collections.

Course(number, title, semester)

Student(id, name, program, age)

Instructor(id, name, department)

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

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

Assume that all attributes/fields in the above collections are of string type. Note that keys are underlined, and you should make the key the primary key in the collection. For example, the Course collection should use the course number as its \_id, while \_id of the Take collection takes the form: {sid: "s100", cno: "DSCI 551"}.

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

1. [12 points] Write a MongoDB function for each of the following questions:
   1. Find the name and program of students who are between 25 and 30 years old (inclusive).
   2. Find out which courses are offered in Fall 2023. Report course number and title.
   3. Find out titles of courses in Fall 2023 that contain “data”. (“data” is case insensitive, e.g., “Data” or “DATA” is also considered as “data”)
   4. Find out which students have taken DSCI 551 in Fall 2023. Report both student ids and names.
   5. Find out which students did not take any courses in Fall 2023. Report only the student ids.
   6. Find out for each instructor, the number of courses (i.e., how many courses) he/she teaches in Fall 2023.
2. [8 points] Write a Python script that uses the library pymongo for each of the following questions. Name your script q2\_a.py, q2\_b.py, etc.
   1. Find out which courses are offered in Fall 2023. Report course number and title.
   2. Find out which students have taken both DSCI 551 and DSCI 552 (both are course numbers). Report only student ids.
   3. Find out which students have taken DSCI 551 but not DSCI 552. Report only student ids.
   4. Find out which instructors teach the largest number of courses in Fall 2023. Report only instructor ids. You can assume that there is only one such instructor.

Note: an example of Python script that uses pymongo are posted Blackboard under MongoDB lecture folder. You may install pymongo on EC2 by executing "pip3 install pymongo".

More details about the library can be found at: <https://pymongo.readthedocs.io/en/stable/tutorial.html>

**Submission Instructions:**

1. **5 files**: q1.txt, q2\_a.py, q2\_b.py, q2\_c.py, q2\_d.py
2. Do NOT submit zip files.
3. Make sure to double check the final queries in q1.txt after copying/pasting operation.
4. If queries/code did not run, then 0 points.