# CS348: Introduction to Database Systems

(Spring 2018)

Assignment 2 (due by 5pm on Friday, June 8th)

Overview: For this assignment, you must use your Unix accounts and DB2 to compose and evaluate a number of SQL queries over the same enrollment database as was used in your first assignment. The visualization of the database schema for this database is reproduced below. All submissions must also use the SQL DDL code for this database given in the contents of a file downloadable from the course web page.

   As with the first assignment, you are given a requirement for each query in English, and your task is to write source code in the SQL query language that implements the requirement.

Assignment submission: By the assignment due date, you must have used the `submit` command to submit a file containing SQL queries that implement each of the requests for information. In particular, put *all* your queries (in numerical order) in one file named "a2.sql". We should be able to run the queries using the command "`db2 -f a2.sql`". Assume a database connection already exists (i.e., you do not need `connect` or `disconnect` statements in your submitted file). To submit your assignment, use the following: "`submit cs348 a2 .`".
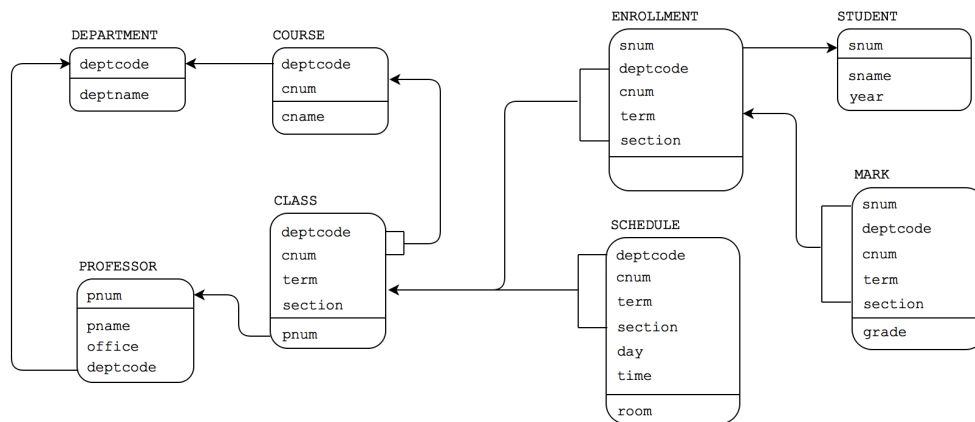
Note that some of the requirements stipulate conditions on what features of SQL may be used in your source code, e.g., that `group by` clauses and aggregate functions may not be used. Part of the grading for your answers in these

cases relate to these conditions. The rest of grading will be based on two additional criteria: (1) correctness, the query implements the requirement, and (2) readability. Consequently, the efficiency of your source code, e.g., as determined by DB2, will not be a factor in any grading.

Queries that may *not* use aggregation in SQL

1. The student number and name of second year students who have obtained a grade lower than 65 in at least two courses in a department with the name "computer science".

2. The number and name of professors who are not in the pure math (PM) department, and who are teaching CS245 for the first time.

3. The number, name and year of each student who has obtained a grade in CS240 that is within 3 marks of the highest ever grade recorded for that course.

4. The number and name of students who have completed two years, who have a final grade of at least 85 in every computer science course that they have taken, and who have always been taught by a professor in the computer science (CS) department.

5. A sorted list of all departments who do not have a professor currently teaching a course offered by a different department.

6. For each pair of classes for the same course that were taught in the same term, and that where also taught by different professors: the minimum grades obtained and the maximum grades obtained. In addition to these four values, each result should include the number and name of each professor, as well as the identifying attributes for each class.

7. Pairs of distinct professors such that whenever the first one teaches a class in a particular term the second also teaches a class for the same course in the same term. Report a oprofessor number and name for both the professors.

Queries that may use aggregation in SQL

8. The course number and total enrollment count for all of its classes of each course. Also, include only those course numbers for courses with a total enrollment count among the three lowest such counts. (Note that one possible result could be:

$$\{[CS, 348, 120], [CS, 446, 120], [CS, 341, 105], [CS, 245, 120], [CS, 234, 121]\}.$$

Also note that all classes, past and ongoing, need to be considered.)

9. The percentage of professors in pure math who have always taught no more than a single course in any given term. (Note that a percentage should be a number between 0 and 100.)

10. The number of different third or fourth year students in each section of each course taught by a pure math professor in past terms. The result should include the professor number, professor name, course number and section, and should also be sorted first by the name of the professor, then by the professor number, third by the course number, and finally by section. (Note that a section is identified by a term and a section number. Also assume that sorting by section means sorting by term and then by section number. The result will therefore have a total of six columns.)

11. The ratio of professors in pure math (PM) to professors in applied math (AM) who have taught a class in which the average grade obtained in the class was greater than 77.

12. For the current term, report how many courses there are in the schedule with a particular number of classes. For example an output

$$\{[5, 1], [4, 2], [1, 5]\}$$

indicates that there are 5 courses with a single class (section), 4 courses with 2 classes, and 1 course with 5 classes scheduled in the curent term.