Homework 2: Data Management

Introduction to Data Science
Fall 2019
Tsinghua University

1 Introduction

In this homework, you are given an **Excel** file with redundant information which describes the relationships between instructors, students and departments in a university, as shown in Figure 1. The definition and type for each field are described as follow:

1. **dept_name**: department name, *varchar (20)*
2. **building**: the building of each department, *varchar (20)*
3. **budget**: the budget of each department, *decimal(10,2)*
4. **i_ID**: the ID of each instructor, *varchar (20)*
5. **i_name**: the name of each instructor, *varchar (20)*
6. **salary**: the salary of each instructor, *decimal(8,2)*
7. **s_ID**: the ID of each student, *varchar (20)*
8. **s_name**: the name of each student, *varchar (20)*
9. **tot_cred**: the total credit of each student, *int*

dept_name	building	budget	i_ID	i_name	salary	s_ID	s_name	tot_cred
Biology	Watson	90000.00	76766	Crick	72000.00	98988	Tanaka	120
Comp. Sci.	Taylor	100000.00	10101	Srinivasan	65000.00	12345	Shankar	32
Comp. Sci.	Taylor	100000.00	45565	Katz	75000.00	76543	Brown	58
Comp. Sci.	Taylor	100000.00	45565	Katz	75000.00	00128	Zhang	102
Comp. Sci.	Taylor	100000.00				54321	Williams	54
Comp. Sci.	Taylor	100000.00	83821	Brandt	92000.00			
Elec. Eng.	Taylor	85000.00	98345	Kim	80000.00	98765	Bourikas	98
Elec. Eng.	Taylor	85000.00	98345	Kim	80000.00	76653	Aoi	60
Finance	Painter	120000.00	76543	Singh	80000.00	23121	Chavez	110
Finance	Painter	120000.00	12121	Wu	90000.00			
History	Painter	50000.00				19991	Brandt	80
History	Painter	50000.00	32343	El Said	60000.00			
History	Painter	50000.00	58583	Califieri	62000.00			
Music	Packard	80000.00				55739	Sanchez	38
Music	Packard	80000.00	15151	Mozart	40000.00			
Physics	Watson	70000.00	22222	Einstein	95000.00	44553	Peltier	56
Physics	Watson	70000.00	22222	Einstein	95000.00	45678	Levy	46
Physics	Watson	70000.00				70557	Snow	0
Physics	Watson	70000.00	33456	Gold	87000.00			

Figure 1: Example data in a university.

2 Tasks

1. Install MySQL Server and MySQL WorkBench in your laptop.
2. **Create**: Create a database named ***University*** in MySQL WorkBench.
3. **Create**: Convert the given Entity-Relationship (ER) models as shown in Figure 2 to tables as we have learned in class and then create them using SQL in the ***University*** database.
4. **Insert**: Insert all the values in the file named ***Data-University.xlsx*** into the ***University*** database using SQL.
5. **Single-Table Queries**:
   (a) Select those students whose total credit are higher than 100 (including 100).
   (b) Select those instructors whose salary are lower than 70000.00 (including 70000.00).
   (c) Select those departments whose budget are higher than 80000.00 (including 80000.00) and then sort them by budget in a descended order.
   (d) Calculate the average salary of those instructors whose salary are between 50000.00 and 100000.00 (including 50000.00 and 100000.00).
   (e) Calculate the student number and teacher number of those departments with more than 2 students (including 2).
6. **Multi-Table Queries:**
   (a) Join table *student* with table *advise* conditioning on **student ID** under the following join types.
      i. natural join
      ii. inner join
      iii. left outer join
      iv. right outer join
      v. full outer join
   (b) List all instructors along with the number (including 0) of students they advise using a proper join type.
7. **Update**: Double the salary of the instructors who advise more than 2 students (including 2).
8. **Delete**: Delete the departments whose budget are lower than 50000.00 (including 50000.00).

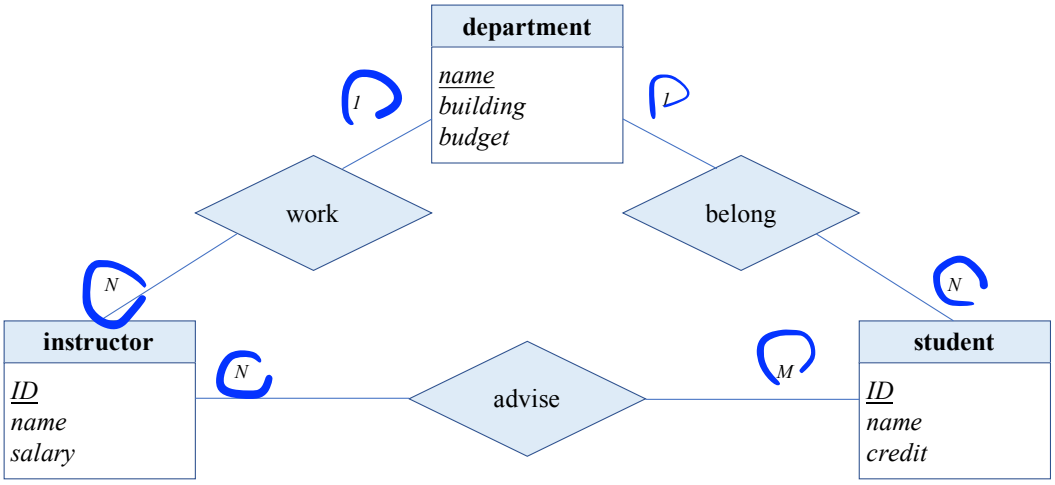

Figure 2: Entity-Relationship (ER) models for describing a university.

3 Submissions

You should submit a **SQL** file as well as a report that including results. Since it is also a simple homework, **NO BONUS** will be given.