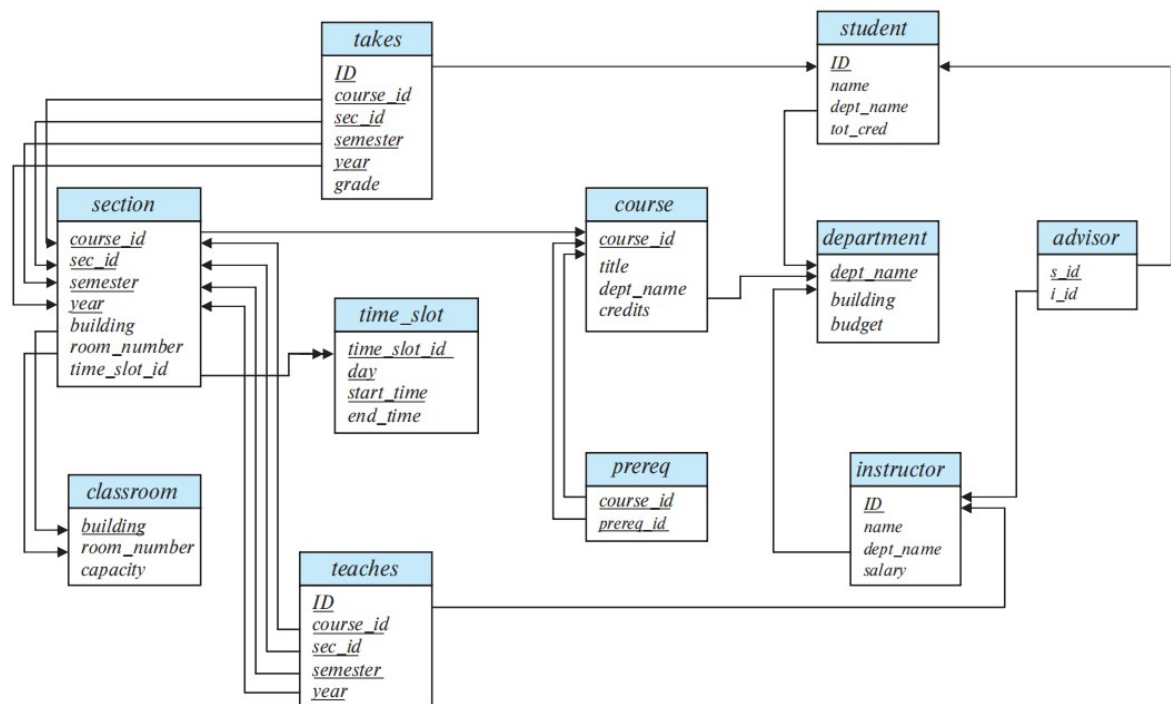# Quiz 2

Q1. Consider the University database as follows:

(1) Suppose Bob writes the following SQL query to find all courses taught in Summer 2019 together with the name of the instructor.

> **select** *name, title*
> **from** *instructor* **natural join** *teaches* **natural join** *section* **natural join** *course*
> **where** *semester* = *'Summer'* **and** *year* = *'2019'*;

Does Bob write the correct query? (If the answer is NO, explain why.)

Q2. Window functions in SQL performs a "sliding" calculation across a set of tuples that are related. They are like an aggregation but tuples (rows) are not grouped into a single output tuples.

RANK() is special window function determines the rank of a value in a group of values, based on the **ORDER BY** expression in the **OVER** clause. If the optional **PARTITION BY** clause is present, the rankings are reset for each group of rows. Rows with equal values for the ranking criteria receive the same rank.

Example: Suppose we have a table student_GPA(ID, GPA) giving the grade-point average of each student. The following query gives the rank of each student:

> **select** *ID*, **rank() over (order by** (*GPA*) **desc) as** *s_rank*
> **from** *student_GPA*;

Note that the order of tuples in the output is not defined, so they may not be sorted by rank. An extra **order by** clause is needed to get them in sorted order, as follows:

> **select** *ID*, **rank () over (order by** (*GPA*) **desc) as** *s_rank*
> **from** *student_GPA*
> **order by** *s_rank*;

Write SQL queries using RANK() for the following:

(1) Suppose we have a table about sales of a bookstore as follows:

month	year	num_books_sold	sales_volume
9	2021	1905	37,551
9	2022	1811	36,002
10	2022	2033	59,990

List each month in descending order of sales volume, and show each month's rank.

Q3. **ON DELETE CASCADE** - When we create a foreign key using this option, it deletes the referencing rows in the child table when the referenced row is deleted in the parent table which has a primary key. If you do not specify cascading deletes, the default behavior of the database server prevents you from deleting data in a parent table if other tables reference it. SQL allows foreign key constraints to refer to the same table.

A table peer_review is created for annual performance appraisal review among instructors in the university. Each instructor's performance appraisal will be reviewed by at most one of his colleagues. The foreign-key clause ensures that every peer reviewer is also an instructor.

> **create table** *peer_review*
> (
> *instructor_ID*　　　　**char**(5),
> *peer_reviewer_ID*　　**char**(5),
> **primary key** *instructor_ID*,
> **foreign key** (*peer_reviewer_ID*) **references** *peer_review*(*instructor_ID*)
> 　　　　　　　　　　　　**on delete cascade**
>);

The table peer_review contains the following tuples.

instructor_ID	*peer_reviewer_ID*
10001	20001
10002	20001
30001	30011
20001	30010
30010	30011
30011	30010

List all the tuples in table *peer_review* after the following deletion. Each deletion shall be considered dependently.

(1) Tuple (20001, 30010) is deleted.

instructor_ID	*peer_reviewer_ID*

(2) Tuple (30011, 30010) is deleted.

instructor_ID	*peer_reviewer_ID*

Q4. Consider the following tables about staffs, meeting rooms and bookings, which are bookings of meeting rooms by staffs.

staff

sid	sname	level	age
1	Fred	7	22
2	Jim	2	39
3	Nancy	8	27

room

rid	rname	nseat
101	Marina	50
102	Teh-C	5
103	Laksa	10

booking

sid	rid	day
1	102	9/12/2021
2	102	9/13/2021
1	101	3/10/2022
1	103	5/10/2022

(1) Write the SQL query to create this table *booking*. Specify the proper constraints and domain types for all attributes.

(2) How many rows are there in the result of the following query?

　　SELECT *　FROM *staff, booking*;

(3) Write SQL query for "Find staffs who have booked all rooms."

Q5. Consider the following table about staffs.

staff

sid	sname	level	age
1	Fred	11	22
2	Jim	2	39
3	Nancy	8	27
4	Bob	*null*	33

What is the result of the following queries?

(1) Is Bob in the result of query "SELECT * FROM staff WHERE level > 7;"?

A. Yes
B. No
C. None of the above

(2) Is Bob in the result of query "SELECT * FROM staff WHERE level = null;"?

A. Yes
B. No
C. None of the above

(3) Is Bob in the result of query "SELECT * FROM staff WHERE NOT (level > 7);"?

 A. Yes
 B. No
 C. None of the above

(4) SELECT count(*) FROM staff;

A. 3
B. 4
C. null
D. None of the above

(5) SELECT sum(level) FROM staff;

A. 21
B. 33
C. 0
D. null
E. None of the above

(6) Suppose table staff now has 150 rows (staffs). The level values of the first 100 rows are unique while the rest, null values. The sum of the 100 unique level values is 9000.

(6.1) SELECT count(*) FROM staff;

A. 100
B. 101
C. 150
D. null
E. None of the above

(6.2) SELECT count(level) FROM staff;

A. 100
B. 101
C. 150
D. null
E. None of the above

Q6. Suppose a relation R has five attributes JKXYZ with the following given functional dependencies (FDs):

$J \rightarrow K$

$KX \rightarrow Z$

$ZY \rightarrow J$

(1) List all candidate keys for R and explain why.

(2) Write an SQL query to check whether FD $J \rightarrow K$ holds and explain.

Q7. Consider the attribute set R = ABCDEGH and the FD set:
F = {AB → C, AC → B, AD → E, B → D, BC → A, E → G}.

(1) For the following attribute sets X, do the following:

    (a) Compute F_X +, i.e., the closure of F_X, where FX is the the projection of F on attribute set X. Trivial FDs could be omitted, e.g., A → A. FDs by simply applying augmentation could also be omitted, e.g., BC → DC by augmenting B → D.

    (b) Identify if X is in BCNF. If X is not in BCNF, decompose it into a set of BCNF relations.

X = ACEH

(2) Consider the following decompositions of R = $ABCDEG$, with the same set of dependencies F, is (a) dependency-preserving? (b) lossless? (Briefly explain why)

{AB, BC, ABDE, EG }

Q8. Consider the database of the NewWear clothes shop. NewWear sells three types of clothes: women's clothes, men's clothes, and children's clothes. Men can choose between polo shirts and T-shirts. Each polo shirt has a list of available colors, sizes and a uniform price. Each T-shirt has a price, a list of available colors, and a list of available sizes. Women have the same choice of polo shirts and T-shirts as men. In addition, women can choose among three types of jeans: slim fit, regular fit, and relaxed fit jeans. Each pair of jeans has a list of possible waist sizes and possible lengths. The price of a pair of jeans only depends on its type. Children can choose between T-shirts and baseball caps. Each T-shirt has a price, a list of available colors and a list of available patterns. T-shirts for children all have the same size. Baseball caps come in three different sizes: small, medium and large. Each item has an optional sales price that is offered on special occasions.

(1) Complete the XML DTD design below for NewWear shop so that it can publish its catalog on the Web. (Read the following queries before design the DTD.)

<!DOCTYPE NewWear[
        <!ELEMENT Item (...)>
]>

(2) Write the following queries in XQuery.

Find the most expensive item sold by NewWear. (Use max() function which returns the maximum value from a sequence of atomic values.)

Q9. Answer the following questions about Spark.

(1) Which of the following parts is in charge of creating RDDs?

A. Local CPU
B. Spark Executor
C. Worker Node
D. Storage
E. Driver Program

(2) Which of the following about laze evaluation is correct in Spark?

A. Transformations are queued and executed at a certain threshold.
B. Actions are not executed until the transformation stage.
C. Actions are queued and executed at a certain threshold.
D. Transformations are not executed until the action stage.

(3) What are the consequences of lazy evaluation?

A. There are no consequences.
B. Errors sometimes do not show up until the action stage.
C. Hiccups within the system during queue execution.