Student Number: _____

HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

COMP3311: Database Management Systems

Spring Semester, 2019

Midterm Examination

March 16th, 2019

2:30pm to 4:30pm

Name: _____ Student Number: _____

Email: _____

Instructions:

1. This is an open-note examination. You are allowed to bring and use the printed course slides from lectures, tutorials and labs in the examination. No electronic device is allowed throughout the exam.
2. This examination paper consists of 15 pages and **4** questions.
3. Please write your name, student ID and Email on this page.
4. For each subsequent page, please write your student ID at the top of the page in the space provided.
5. Please answer **all** the questions within the space provided on the examination paper. You may use the back of the pages for your rough work and afterwards drawn a diagonal line through it to show that it is not part of your answer.
6. Please read each question very carefully and answer the question clearly and to the point. Make sure that your answers are neatly written, readable and legible.
7. Leave all pages stapled together.
8. The examination period will last for **2 hours**.
9. Stop writing immediately when the time is up.

Questions	Marks	Scores
1	20	
2	30	
3	15	
4	15	
5	20	
Total	100	

Student Number: _____

Q1 Multiple-choice Questions (20 marks)

For each question, there is only one correct answer. Please put your answers in the following boxes. Only answers in the boxes will be graded.

1	**B**
2	**A**
3	**C**
4	**B**
5	**C**

6	**C**
7	**A**
8	**C**
9	**D**
10	**A**

11	**A**
12	**D**
13	**C**
14	**A**
15	**B**

16	**B**
17	**A**
18	**D**
19	**B**
20	**C**

1. Which of the following statement(s) is/are true about weak entity?

   I. Weak entity is an entity that has only ID relationship.
   II. Weak entity inherits part of its key from another entity in its ID relationship.
   III. Weak entity is never mapped to a table when converting the ERD into relational schemas.

   A. I only.
   B. II only.
   C. II and III only.
   D. All of them.

2. Which of the following constraint is *not* possible to represent as a relational schema in general?

   A. A total participation constraint in a many-to-many relationship.
   B. A partial participation constraint in a one-to-one relationship.
   C. A cardinality constraint 0..* in a many-to-many relationship.
   D. A cardinality constraint 1..* in a one-to-one relationship.

3. Which of the following three schema design is *not* implemented in a single relational schema? (The relation names are bold. The underlined attributes are to indicate the key. The description after the colon is to explain the meaning of an attribute.)

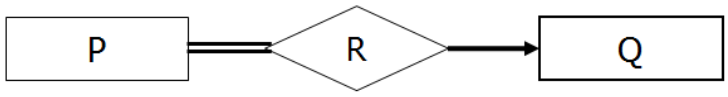
   I. **Student** (<u>name</u>: name of the student, birthdate: when the student was born, from: where the student lives, clubs: a set of the student's club memberships (references **Club** that contains each club's information)).
   II. **Marriage** (<u>wife</u>: name of the wife, <u>husband</u>: name of the husband, marriagedate: the date when they were married).
   III. **Book** (<u>title</u>: the title of the book, <u>year</u>: when the book was published, authors: a list of author names who wrote the book (references **Authors** that contains each author's information), publisher: which published the book, text: the text of the entire book).

   A. I and II only.
   B. II and III only.
   C. I and III only.

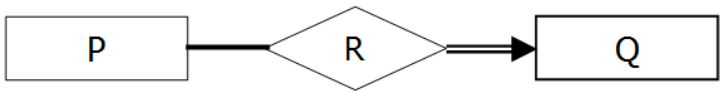
D. All.

4. Consider the following two ER diagrams labelled as ER1 and ER2:

ER1



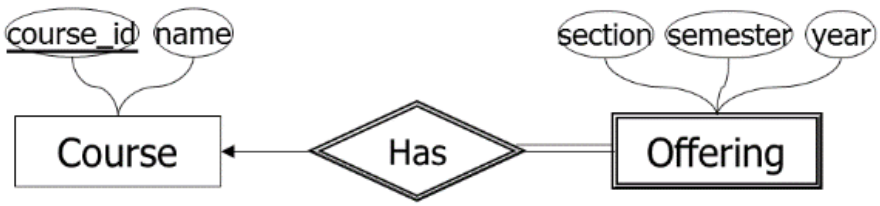
ER2



The statement that the number of entities in P must be greater than or equal to the number of entities in Q holds for:

A. ER1 but not ER2
B. ER2 but not ER1
C. Both ER1 and ER2
D. Neither ER1 nor ER2

5. Considering the following ERD, where the attributes section, semester and year are together forming the partial key of Offering, and course_id is the key of Course:



Which of the following FDs are true for the ERD?

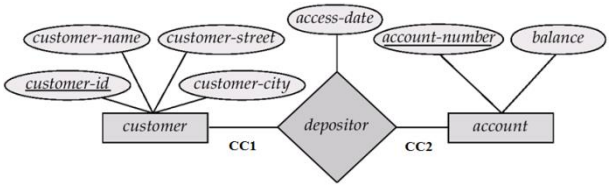
I. course_id→name
II. course_id→section, semester, year
III. section, semester, year→section

A. I and II only.
B. II and III only.
C. I and III only.
D. All.

6. Which of the following is false for a recursive relationship in an ER diagram?

   A. All entities participated in the relationship must have a key.
   B. All entities participated in the relationship must be in the same entity set.
   C. The relationship must be either one-to-many or many-to-one.
   D. The edges connecting entity sets in the relationship must have different role name.

7. Which statement about the following ER diagram is true?



   I. The cardinality constraints CC1 and CC2 on the edge can be 0..*
   II. Access-date is a single-valued attribute.
   III. We can have more than one access-date for a given pair of customer-id and account-number.

   A. I and II only.
   B. II and III only.
   C. I and III only.
   D. All.

8. Suppose we have the following table T(A,B):

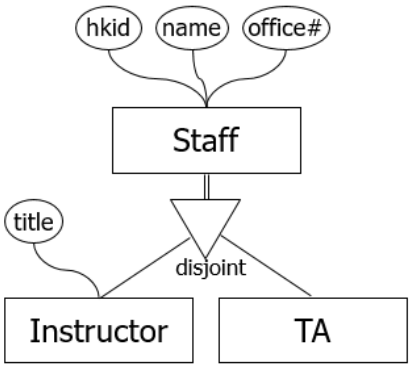
A	B
1	
2	a
2	b
3	
3	C
3	D

```
SELECT A, MAX(B), COUNT(*), COUNT(B)
FROM T
GROUP BY A
ORDER BY A;
```

   Which of the following statement about the result of the above SQL query is false?

   A. The SQL result has only three records in total.
   B. The first record contains null values.
   C. The last record contains null values.
   D. The values in columns COUNT(*) and COUNT(B) are not exactly the same.

9. Which of the following statement about the ER diagram is true?



A. Some staff members can be both Instructor and TA.
B. Some staff members can be neither Instructor nor TA.
C. It is not possible that all staff members are Instructor.
D. It is not possible that a TA member has a title attribute.

10. Consider a relation $R(a, b)$ having n tuples, where a and b are the attributes, and the RA expression $\pi_{S.a}(R \times \rho(S, R))$. Which of the following statement is true?

A. The number of rows returned is the same as that of $\pi_{R.a}(R)$.
B. The number of rows returned is the same as that of $\pi_{R.b}(R)$.
C. The number of rows returned is n^2.
D. The number of rows returned is $2n$.

11. Suppose we have following relations R1 and R2.

R1

A	**C**
3	3
6	4
2	3
3	5
7	1

R2

B	**C**	**D**
5	1	6
1	5	8
4	3	9

Which of the following (A, B, C, D) values are in a tuple of the result (R1 Natural Join R2)?

A. (7, 5, 1, 6)
B. (3, 3, 5, 1)
C. (3, 5, 5, 8)
D. (2, 3, 3, 9)

12. Which of the following statements about using SQL view is generally true?

    I. You can always insert a record in an existing view. (assuming the record is compatible to the view schema.)
    II. You can always delete a record in an existing view. (assuming the record exists in the view.)
    III. You can always modify a record in an existing view. (assuming the record exists in the view.)

    A. I only.
    B. II only.
    C. III only.
    D. None.

13. Which of the following SQL statements can be used to remove the entire table called "ABC" in a database?

    I. DELETE * FROM ABC;
    II. DELETE TABLE ABC;
    III. DROP TABLE ABC;

    A. I only
    B. II only
    C. III only
    D. None.

14. Consider the following TRANSCRIPT table.

NAME	COURSE	SCORE
Alice	Java	89
Alice	C++	91
Bob	Java	89
Carl	Ruby	91
Dave	Python	95

    What is the result when running the following SQL statement?

```
SELECT NAME, AVG(SCORE)
FROM TRANSCRIPT
GROUP BY NAME
HAVING COUNT(*)>1;
```

    A. Only one record is displayed, and it is (Alice, 90).
    B. Only two records are displayed, and they are (Alice, 89) and (Alice, 91).
    C. Five records are displayed.
    D. No record is displayed.

15. Consider the relations A(a,c,d), B(a,b), which of the following operation is *not* permissible?

    A. A × B
    B. A ∪ B
    C. A NATURAL JOIN B
    D. $B \,/\, \pi_a(A)$

16. Consider the relation schema R = (A, B, C, D) with a set of FDs {A→B, B→C, C→A}. Which one of the following is a super key but not a candidate key?

    A. AB
    B. ADC
    C. BD
    D. None of the above

17. Consider the relation schema R = (A, B, C, D) with a set of FDs {D→C, A→B, D→A, B→D}. Which one of the following is the attribute closure of D?

    A. DCAB
    B. DC
    C. DCA
    D. DAB

18. Consider the relation schema R = (A, B, C, D, E) with a set of FDs F = {D→BA, A→AB, AD→EC, A→D}. Which one/ones of the following sets is/are a canonical cover of F?

    I. {D→AB, A→AB, AD→CE, A→D}
    II. {D→ABCE, A→D}
    III. {A→BCDE, D→A}

    A. I Only.
    B. II Only.
    C. III Only.
    D. II and III Only.

19. Consider the relation schema R = (A, B, C, D) with a set of FDs F = {A→B, C→D, D→B}. Which one of the following statements about decomposing R into R1(A, B, D) and R2(C, D) is true?

    A. The decomposition R1, R2 are in BCNF.
    B. The decomposition R1, R2 are dependency preserving.
    C. The decomposition R1, R2 are lossless.
    D. The decomposition R1, R2 are in 3NF.

20. Given R(A, B, C, D, E) and a set of FDs {B→CD, D→E}. Which of the following is the output result of running BCNF decomposition algorithm?

    A. R(A, B, C, D, E) is already in BCNF so R is the final output.
    B. R1(B,C,D), R2(A,B,E) are the final output.
    C. R1(B,C,D), R2(A,B), R3(D,E) are the final output.
    D. None of the above.

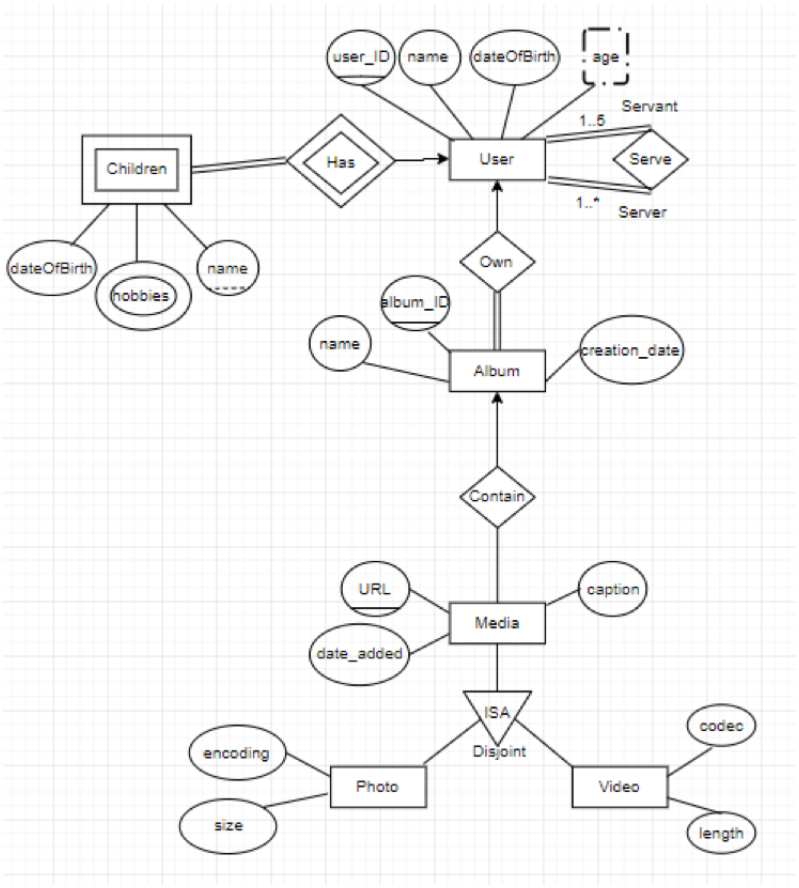
Q2. ER Diagram and Database Schema (30 marks)

a. Consider the following requirements for designing a database that stores information for a social networking website. The database should have the following properties:

- Every **user** has a unique user ID, a name, a date of birth and an age computed from the date of birth.
- Every user (as a server) serves at least one client. Every user (as a client) is served by at least one server and at most five servers. In other words, all users are servers and clients.
- A user has zero or more **children** who are not users**.** A child has a name, a date of birth and a list of hobbies. Some children's names are the same. However, a user ID together with a child's name is sufficient for us to differentiate a child from other children.
- Users are allowed to create zero or more albums. An **album** has a unique album ID (integer), a creation date, and a name. Every album is owned by exactly one user who created the album.
- We store media files in the system. A **media** file may belong to at most one album or to no album.
- An album can contain zero or more media files. For every media file, we record its unique URL, the date when the file was added to the album, and a caption if one exists.
- Users may add zero or more **photos** to albums. Photos are a type of media file, and we also track the encoding (e.g., JPEG, PNG, etc.) and the size of the photo (in bytes).
- Users may add zero or more **videos** to the album. Videos are a type of media file, and we also track the codec used to encode the video (e.g., MPEG-4), the length of the video (in seconds).

You are required to create an Entity Relationship Diagram (ERD) for this website. The ERD should contain all the objects highlighted in bold in the above requirement. **(15 marks)**

Suggested Solution:



Grading criteria: (15pts)
Entities: 0.5pt for each (**3pts**)
Relationship: 0.5pt for each (**2pts**)
ISA with disjoint: **1pt**
Weak entity: double rectangle **1pt**
Identifying relationship notation: **1pt**
3 primary key: 0.5pt for each (**1.5pts**)
1 discriminator: 0.5pt
8 constraints: 0.5 pt for each (**4pts**)
Multivalued attribute: **0.5pt**
Derived attribute: **0.5pt**
Attributes: -0.5pt for each missing, -1pt in total (**1pt**)
Any wrong, missing, redundant relationships, entities, attributes: **-0.5pt for each**

b. Convert the ERD in Part (a) into its corresponding relational schemas. Identify all the primary keys (underline them) and foreign keys (underline them with clear dashed lines) of the relational schemas. Any explicit integrity constraints on the relation schemas (e.g. foreign key, insert and delete constraints) should also be clearly specified. **(15 marks)**

Suggested solution:

User(user_id, name, dateOfBirth)

Serve(server_id, client_id)
Server_id references user_id on delete cascade
Client_id references user_id on delete cascade

Children(user_id, name, dateOfBirth)
user_id references User on delete cascade

ChildHobbies(user_id, name, hobby)
(user_id, name) references Children on delete cascade

Album(album_id, name, create_date, user_id)
User_id not null
user_id references User on delete cascades

Media(URL, data_added, caption, album_id)
Album_id not null
Album_id references Album on deleted set null

Photo(URL, size, encoding)
URL references Media on delete cascade

Video(URL, codec, length)
URL reference Media on delete cascade

8 Table : 8 * 0.5 = 4
8 Primary key : 8 * 0.5 = 4
8 Reference constraint: 8 * 0.5 =4
7 Foreign key: 7 * 0.5 = 3.5

Q3 Structured Query Language (15 marks)

Consider the following database having three relational schemas with the primary key underlined as follows:

- employee(employee-name, college, major)
- works(employee-name, company-name, position, salary)
- company(company-name, company-type)

The database stores the information of the current working details. You should assume that all people work for *at most one company* and that you are using the Oracle system introduced in the labs for running the queries. Write the following queries in SQL using the database. State your necessary assumptions, if any.

a. Find the college and major of all employees who work as "Software Engineer" and earn more than $100,000 **(3 marks)**

```
SELECT EMPLOYEE.college, EMPLOYEE.major FROM EMPLOYEE, WORKS
WHERE EMPLOYEE.employee_name=works.employee_name
AND position = 'Software Engineer'
AND salary > 100000;
```

b. Find the employee names, company name and position of all employees who graduated from HKUST and earn more than other HKUST alumni working in the same company. **(4 marks)**
Hint: HKUST alumni means college = 'HKUST' in the employee table.

```
SELECT W1.employee_name, W1.company_name,W1.position
FROM WORKS W1,EMPLOYEE
WHERE EMPLOYEE.employee_name=W1.employee_name AND college = 'HKUST'
AND W1.salary >= ALL(SELECT W2.SALARY FROM WORKS W2, EMPLOYEE
    WHERE EMPLOYEE.employee_name=W2.employee_name AND college = 'HKUST' AND W1.company_name = W2.company_name);
```

c. Find the names and major and position of all employees who earn not less than the average salary (greater than or equal to the average salary) of all employees of their company. **(3 marks)**

```
SELECT T.employee_name
FROM WORKS T
WHERE SALARY >=(SELECT AVG(SALARY) FROM WORKS S WHERE T.company_name = S.company_name);
```

d. Find the type of the company that gives the smallest payroll (total salaries of employees) for "Software Engineer". **(4 marks)**

```
SELECT company_type FROM company
WHERE company.company_name in(SELECT company_name
FROM works
WHERE position = 'Software Engineer'
GROUP BY company_name
HAVING SUM(SALARY) <= ALL (SELECT SUM(SALARY) FROM WORKS
WHERE position = 'Software Engineer'
GROUP BY company_name));
```

Q4 Relational Algebra (15 marks)

Consider the following database having four relational schemas with the primary key underlined as follows. The names are self-explanatory and have usual meaning.

- Engineer (EngineerID, Name, Email, City)
- Repair (EngineerID, EquipmentID, Date)
- Equipment (EquipmentID, Name, CompanyID, Price)
- Company (CompanyID, Name, City)

Write the following queries in Relational Algebra. State your necessarily assumptions, if any.

a. Find all the equipment IDs that have never been repaired. **(3 marks)**

$$\pi_{EquipmentID}(Equipment) - \pi_{EquipmentID}(Repair)$$

b. Find the equipment items having the same names but they are from two companies named Apple and Samsung.**(3 marks)**

$$\pi_{Equipment.Name}(Equipment\ JOIN(\sigma_{Name='Apple'}(Company)) \cap \pi_{Equipment.Name}(Equipment\ JOIN(\sigma_{Name='Samsung'}(Company))$$

c. Find all employee IDs who have repaired all the equipment produced by company named DELL. **(4 marks)**

$$\pi_{Engineer.EngineerID,Repair.EquipmentID}(Engineer\ JOIN\ Repair) / \pi_{Equipment.EquipmentID}(Equipment\ JOIN(\sigma_{Name='Dell'}(Company))$$

d. Find all the equipment names and their prices, which are more expensive than $3000 and they were repaired at least once. **(4 marks)**

$$\pi_{Equipment.Name,\ Equipment.Price}(Repair\ JOIN(\sigma_{price>3000}(Equipment))$$

Q5 Functional Dependencies and Normalization (20 marks)

Consider the relation schema R = (A, B, C, D, E, H). The following set ***F*** of functional dependencies holds on R:

$$
\begin{aligned}
AB &\rightarrow E \\
C &\rightarrow A \\
BC &\rightarrow EA \\
EA &\rightarrow D \\
HE &\rightarrow AC \\
A &\rightarrow BC
\end{aligned}
$$

Answer the following questions and explain how you derived the solution clearly.

a. Compute the *attribute closure* of C (i.e., C^+). Show all the steps during the computation. **(5 marks)**

{C} → {A} (using {C} → {A}),
{C, A} → {A, B, C} (using A→ BC)
{A, B,C} → {A,B,C,E} (using AB→ E)
{A, B,C, E} → {A, B,C, E, D} (Using EA→ D)
C += ABCDE

b. Compute the *canonical cover* of F (i.e., F_c). Show whether in each step you apply the *union rule* or you remove an *extraneous* attribute. In the second case, explain why the attribute is extraneous. **(5 marks)**

1) Remove BC→EA(because C → A implies C→EA)
2) Remove B from AB→E (because A→BC)
3) Remove E from EA→D (because A→E)
4) Apply union A → BCDE
The canonical cover Fc is {A→ BCDE, C→A, HE→A}

c. Write down all the candidate keys for R. (Any incorrect answer gets mark deducted) **(5 marks)**

1) HC
2) HE
3) HA

Grading: Only mark the first three written by the student if they have written more

d. Is the schema R in 3NF? Explain why or why not. If not, show how to obtain a 3NF by a decomposition of R. **(5 marks)**

The schema is not in 3NF since in A →BCDE violates the rules (neither A is a candidate key nor B, E are prime attributes).

A 3NF decomposition is R1(A,B,C,D,E), R2(B,F), R3(A,H,E).

- End of the Exam Paper-