Student Number: _____

HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

COMP3311: Database Management Systems

Spring Semester, 2018

Midterm Examination

March 10th, 2018

9:30am to 11:30am

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 14 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	30	
4	20	
Total	100	

Student Number: _____

Q1 Multiple-choice Questions (30 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	**C**
2	**C**
3	**A**
4	**B**
5	**D**

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

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

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

1. The result of the following SQL SELECT statement is _____ .

```
SELECT COUNT(A)
FROM R;
```

Table R:

A	**B**
NULL	NULL
NULL	1

A. 2
B. 1
C. 0
D. NULL

2. Which of the following statements is/are true about using aggregate functions in SQL?

I. We can apply aggregate functions on the attributes of a view.
II. COUNT(A) also count the occurrence of duplicated values in the column of A.
III. Both MAX(A) and MIN(A) return NULL if all values in the column A are NULL.

A. I only.
B. I and II only.
C. All of the above.
D. None of the above.

3. Consider the following tables (WORKS and COMPANY) in a database.

WORKS:

Person_name	**Company_name**	**Salary**
Jack	SCB	50000
Tom	SCB	40000
Lee	HSBC	40000
Nick	HSBC	60000
Jim	Tencent	110000

COMPANY:

Company_name	**City**
SCB	Hong Kong
HSBC	Hong Kong
Tencent	Shen Zhen

What is/are the company name(s) obtained from executing the following SQL statement?

```
select company.company_name
from works, company
where
    works.company_name = company.company_name
   and company.city = “Hong Kong”
group by company.company_name
having sum(works.salary) < 100,000
```

A. SCB
B. SCB, HSBC
C. SCB, HSBC, Tencent
D. Tencent

4. Given a relation r, the result of r ⋈ r (i.e. r natural join r) is equal to:

I. $r \cup r$
II. $r \cap r$
III. $r \times r$

A. I only.
B. I and II only.
C. All of the above.
D. None of the above.

5. Which of the following is/are true about weak entity sets?

I. A weak entity always has an existence dependency as a relationship.
II. The discriminator of a weak entity set is the set of attributes that distinguishes the weak entities.
III. A weak entity set should be depicted as a doublelined rectangle in the ERD.

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

6. What of the following is true about keys in the relational model?

I. A superkey is any set of attributes that can identify a unique tuple in a relation.
II. A relation may have more than one candidate key.
III. The set of all the attributes in a relation schema are a superkey.

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

D. All of the above.

7. Given the following two tables R(A, B, C) and S(A,C) in a database:

R:

A	**B**	**C**
1	3	6
4	6	8
4	7	6

S:

A	**C**
1	3
4	6

How many rows in the result of R ⋈ S (i.e. natural join of R and S)?

A. 0
B. 1
C. 2
D. 3

8. Given the following two tables R(A, B) and S(B,C) in a database:

R:

A	**B**
1	3
2	4

S:

B	**C**
4	7
5	8
6	9

Which of the following statement(s) is/are true?

I. There are 2 rows in (R left outerjoin S).
II. There are 3 rows in (R right outerjoin S).
III. There are 4 rows in (R full outerjoin S).

A. I only.
B. II only.
C. III only.
D. All of the above.

9. Given a relation R(A, B, C, D, E) with a set of functional dependencies F = {A→BC, B→AC, AD→E, E→D}, which of the following is a candidate key for R?

I. AD
II. AE
III. BD
IV. BE

A. I and II only
B. III and IV only
C. I, II, III only
D. All of the above

10. Which of the statement(s) about the division R/S, where R and S are two tables, is/are correct?

I. The schema of the result of (R/S) should be a subset of the schema R.
II. The schema of R should be a subset of the schema of S.
III. We can use only the three operators π, - and $\times$ on R and S to perform the division.

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

11. Given two tables: R1 has m rows and R2 has n rows. What is true about maximum and minimum size of the result R1 $\bowtie$ R2 (i.e. R1 natural join R2)?

A. Maximum m+n and minimum 0
B. Maximum m+n and minimum m
C. Maximum m*n and minimum 0
D. Maximum m*n and minimum m+n

12. When inserting records into the tables that represent a "One to Many" relationship and its related entities, which of following is true?

A. Records on the table of "Many" side should be inserted first.
B. Records on the table of "One" side should be inserted first.
C. Order of record insertion does not matter.
D. Order of record insertion depends on the primary key of the table of "Many" side.

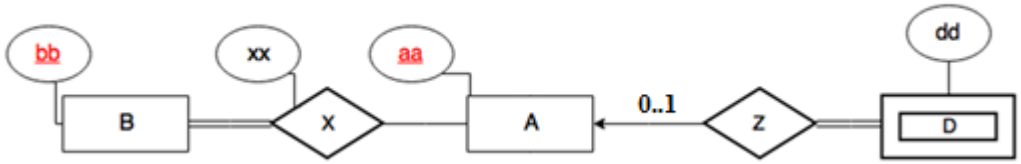
13. To store a multivalued attribute A in an entity E having the primary key K, the following statement is true:

A. The attribute A should be one of the attributes in the table of E.
B. A separate table, other than the table of E, having attributes A and K should be created, where A should not be null but K can be null.
C. A separate table, other than the table of E, having attributes A and K should be created, where both A and K should not be null.
D. A separate table, other than the table of E, having attributes A and K should be created, where both A and K can be null.

14. A role name is required for which type of relationship?

A. Ternary relationships
B. Unary relationships
C. Binary one to many relationships
D. Binary many to many relationships that have relationship attributes

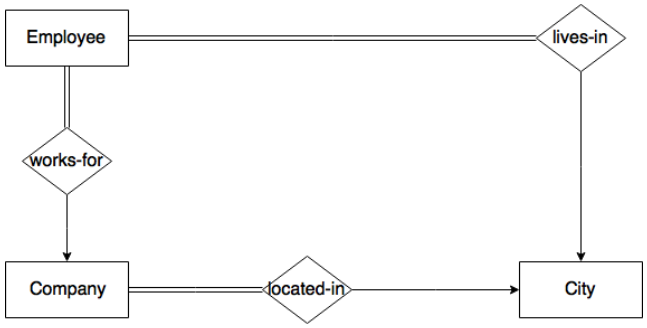
15. Which FD(s) is/are present in the following ERD?



I. aa → dd
II. aa → xx
III. aa → bb

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

16. Can the relationship "Lives-in" be represented by the two relationships "Works-for" and "Located-in" in the following ERD?



A. Yes, if we have the constraint that every employee works in the city in which he lives.
B. Yes, if we have the constraint that every employee works for only one company.
C. Yes, if we have the constraint that every company is located in different city.
D. No, it is impossible.

17. Given a relation schema R(A, B, C, D, E) with the following set of functional dependencies F = {A → DE, D → C, BD → EA , CE→ B }.

Which of the following is **NOT** a key of R?

A. BD
B. BCE
C. A
D. ABC

18. Consider the relation R(A, B, C ,D, E) with the following set of functional dependencies F = {A → E, B → DE, C → A , CB→ D }.

    Which of the following is the *attribute closure* of {A, C} (i.e. X^+ where X = {A, C})?

    A. ABCDE
    B. ABCE
    C. ACE
    D. ACDE

19. Consider the following set of functional dependencies F = { A→ CE, B → E, B → D , AB→ DE }.

    Which of the following is the *canonical cover* of F (i.e. F_C)?

    A. { A→ DE, B → D }
    B. { A→ CE, B → DE }
    C. { B→ D, B → E , A → C }
    D. { A → CE , AB→ DE }

20. Consider the relation schema R(A, B, C, D, E) with the following set of functional dependencies F = { ABC → D, AB → D, A → D}. Suppose that we decompose R into R1(A, D) and R2(A, B, C, E).

    Which of the following is true for the given decomposition?

    A. The decomposition is in 3NF, but not in BCNF.
    B. The decomposition is neither in 3NF nor in BCNF.
    C. The decomposition is both in 3NF and BCNF.
    D. The decomposition is in BCNF, but not in 3NF.

Student Number: _____

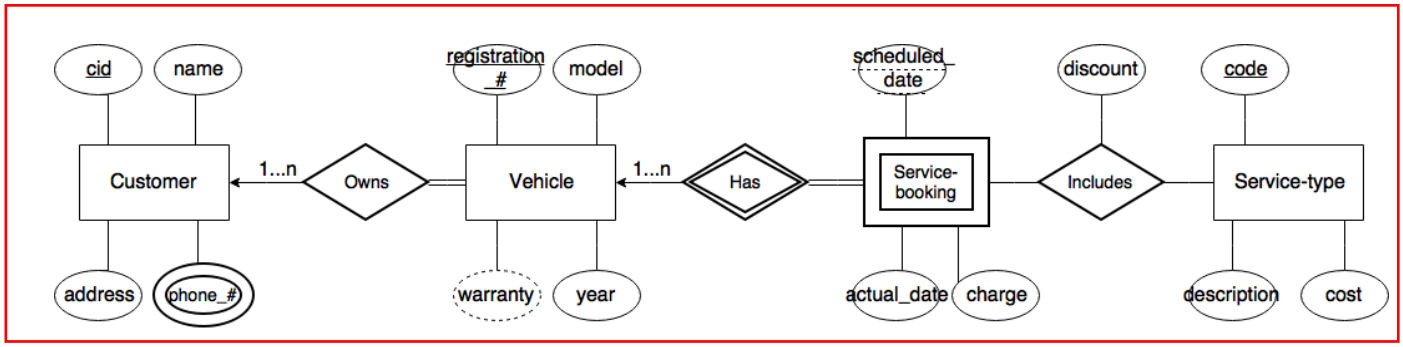
Q2. ER Diagram and database schema (30 marks)

1. You are required to create an Entity-Relationship Diagram (ERD) for a small automobile repair and service company. The company keeps information on its customers, their vehicles and the service booking for each vehicle. The details of the requirements are as follows:

- For each **customer**, the company records a unique id, name, address and a list of phone numbers.
- For each **vehicle**, the company records its unique registration number, the model name, and the year of manufacture of the vehicle. There is a warranty for a vehicle. The warranty of a vehicle depends on its model and the manufacture year.
- A customer may own more than one vehicle, but each vehicle is owned by only one customer.
- The company provides several **service types** for which it records a unique code, a description and the cost of the service.
- For each vehicle, the company records the **service bookings**. Each service booking contains *only* the following information: scheduled service date, the actual service date and the total service charge. Each service booking includes an arbitrary number of service types that are performed for the vehicle.
- For some service bookings, the company may offer a discount on some of the service types, and the discount can be different for different service types and for different bookings.

Your ERD should use the usual notation to show all the entities (in **bold**), relationships, attributes, keys (underline the primary keys with solid lines, underline the discriminator(s) with dashed line(s)). Assume all the relationships are binary. Your ERD should also show the cardinality constraints and the participation constraints. **(15 marks)**

Answer:



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

Answer & Grading criteria (**15pt**):

1. Customer (cid, name, address) **1pt**
2. Phone_number (cid, number) **1pt**
   cid references Customer, on delete cascade **1pt**
3. Vehicle (registration_number, cid, model, year) **1pt**

cid not null, references customer, on delete cascade **1.5pt**

4. Service-booking (scheduled_date, registration_number, actual date, charge) **1pt**
   registration_number references Vehicle, on delete cascade **1pt**
5. Service-type (code, description, cost) **1pt**
6. Service (code, scheduled_date, registration_number, discount) **1pt**
   code references Service-type, on delete cascade **1pt**
   scheduled_date, registration_number reference Service-booking, on delete cascade **1pt**
   (if missing partial: -0.5pt)

6 primary keys: 0.5pt for each (**3pts**) (partial answer: 0pt)
1 foreign key: **0.5pt**

Q3 Relational Algebra and Structured Query Language (30 marks)

Consider the following database having four relational schemas:

- Professor(pid, pname, salary, pdept)
- Course(cid, cname, dept)
- Teaches(pid, cid, sfq_score)

pid is the primary key of Professor table.
cid is the primary key of Course table.
pid and cid combined is the primary key for Teaches table.

Where "pname" in the "Professor" table refers to the name of the Professor; "pdept" in the "Professor" table refers to the department that the professor is affiliated to; "cname" in the "Course" table refers to the name of the course and "dept" in the "Course" table refers to the department that offers the course. The meaning of other attributes is self-explanatory.

Express the following queries in BOTH Relational Algebra and SQL. Assume that you are using SQL for the Oracle system introduced in the labs.

Query 1. Find the name and pid of all the professors who have a salary greater than 1000 (i.e. salary > 1000), and obtain sfq_score 100 in "COMP3311" course (i.e. cid = 'COMP3311'). The course is offered in the same department as the professor's department.

Relation Algebra: **(5 marks)**

$\pi_{\text{Professor.pid, Professor.pname}}((\sigma_{\text{salary} > 1000}(\text{Professor})) \bowtie_{\text{pid}} (\sigma_{\text{Teaches.sfq\_score= 100} \wedge \text{Teaches.cid = "COMP3311"}}(\text{Teaches})) \bowtie_{\text{professor.pdept = course.dept}} (\sigma_{\text{cid=COMP3311}}\text{Course}))$

SQL: **(5 marks)**

```
Select pname, S.pid
From Professor S, Teaches E, Course C
Where S.pid = E.pid AND S.salary>1000 AND sfq_score = 100 and E.cid= C.cid AND C.dept =
S.pdept AND C.cid = 'COMP3311';
```

Query 2. Find the name and salary of all professors who have the 3rd lowest salary among all professors.

Relation Algebra: **(5 marks)**

$$Temp \leftarrow \pi_{s1.pid,\ s1.pname,\ s1.salary}\left(\sigma_{s1.salary > s2.salary}\left(\rho\ (s1, Professor) \times \rho(s2, Professor)\right)\right)$$
$$Inter \leftarrow \pi_{t1.pid,\ t1.pname,\ t1.salary}\left(\sigma_{t1.salary > t2.salary}\left(\rho\ (t1, Temp) \times \rho(t2, Temp)\right)\right)$$

$$\Pi_{pname,\ salary}\left(Inter \leftarrow \pi_{I1.pid,\ I1.pname,\ I1.salary}\left(\sigma_{I1.salary > I2.salary}\left(\rho\ (I1, Inter) \times \rho(I2, Inter)\right)\right)\right)$$

SQL: **(5 marks)**

```
SELECT pname,  salary
FROM PROFESSOR
WHERE salary = SELECT MIN(salary)
FROM PROFESSOR
WHERE salary > (SELECT MIN(salary)
                FROM PROFESSOR
                Where salary >  (SELECT MIN(salary)
                FROM PROFESSOR))
```

Alternative:

```
Select pname, salary
From Professor s1
Where 3 = (select count(distinct salary) from Professor s2 where s1.salary>=s2.salary);
```

Query 3. Find the pid and name of the Professors who teaches all the courses.

Relation Algebra: **(5 marks)**

$$\pi_{pid,pname}\ (\ (\ \pi_{pid,\ cid}\ (Teaches)\ /\ \pi_{cid}\ (Course)\ )\ \bowtie_{pid}\ Professors\ )$$

SQL: **(5 marks)**

```
Select S.pid, pname
From Professor S, Teaches E
Where S.pid = E.pid AND NOT EXISTS
(
 (select cid from Course)
EXCEPT
(select cid from Teaches E2 where E2.pid = S.pid)
)
```

Q4 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:

A→ BE
D → CE
B→ H
BD → C
H→ AD

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

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

{H} → {A,D,H} (using H → AD)
{A,D,H} → {A, C, D, E, H} (using D → CE)
{A,C,D,E,H} → {A,B,C,D,E,H} (using A →BE)

H^+= ABCDEH

(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 the extraneous attribute B from BD → C, since we can infer from D → CE that D → C and thus BD → C. So, we get D → C
2) Apply the union rule for D→C, D→CE and get D→CE
3) Remove E as an extraneous attribute from A→BE, since if A→B then B→H and since H^+= ABCDEH, then also A→E and A→BE

Finally, the canonical cover F_c is {A→B, B→H, D→CE, H→AD}

(c) List all the candidate keys for R. Explain your answer(s). **(5 marks)**

1) H
2) B
3) A

(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 D →CE violates the rules (neither D is a candidate key nor C, E are prime attributes).

A 3NF decomposition is R1(A,B), R2(B,F), R3(C,D,E), R4(A,D,F).

- End of the Exam Paper-