NATIONAL UNIVERSITY OF SINGAPORE

CS2102 - DATABASE SYSTEMS

(Semester 2: AY2016/17)

Time Allowed: 2 Hours

INSTRUCTIONS TO STUDENTS

1. Please write your Student Number only. Do not write your name.
2. This assessment paper contains **EIGHT** questions and comprises **THIRTEEN** printed pages, including this page.
3. Questions 1 and 5 consist of a total of 8 multiple-choice questions (OCR1 to OCR8) which are to be answered on the provided **OCR form**; answer all the remaining questions within the space in this booklet.
4. This is a **CLOSED BOOK** assessment.
5. You are allowed to refer to a single, double-sided A4-sized sheet of notes.

STUDENT NO: | | | | | | | | | |

EXAMINER'S USE ONLY

Question:	1	2	3	4	5	6	7	8	Total
Points:	6	7	7	7	10	7	6	10	60
Score:									

Part I: Relational Calculus & SQL

This part consists of four questions based on the following schema.

staff (sid, sname)
project (pid, pdesc, pfrom, pto)
asset (aid, acat, adesc)
workfor (sid, pid, wfrom, wto)
assignment (aid, sid, pid, afrom, ato)

The primary key of each table is underlined.

The `staff` table stores the identifiers, `sid`, and names, `sname`, of staffs. The `project` table stores the identifiers, `pid`, descriptions , `pdesc`, start and end dates, `pfrom` and `pto`, of projects. The `asset` table stores the identifiers, `aid`, categories , `acat`, and descriptions, `adesc`, of assets. The `workfor` table stores the staffs, `sid`, working for projects, `pid`, with the start and end dates of the work, `wfrom` and `wto`. The `assignment` table stores the assets, `aid`, assigned to staffs working for projects, `sid` and `pid`, and the start and end dates of the assignment, `afrom` and `ato`.

Staff members are identified by their identifier, `sid`. Projects are identified by their identifier, `pid`. A project starts on date `pfrom` and ends on date `pto`. Assets are identified by their identifier, `aid`. A staff member works for a project for a period of time starting on date `wfrom` and ending on date `wto`. An asset is assigned to a staff working for a project for a period of time starting on date `afrom` and ending on date `ato`.

There is a foreign key from `workfor` to `staff`. There is a foreign key from `workfor` to `project`. There is a foreign from `assignment` to `asset`. There is a foreign key from `assignment` to `workfor`. There is no other foreign key.

Assets and Projects are considered related to Python when their description, `adesc` and `pdesc`, respectively, contain the word "Python". Use the LIKE operator for such conditions in all query languages. Assets have a category (`acat`), for instance "Computer". Dates (`pto`, `pfrom`, `wto`, `wfrom`, `ato` and `afrom`) are in the format `''DD/MM/YYYY''`.

Question 1 consists of three multiple-choice questions (OCR1 to OCR3) which are to be answered on the OCR form. Questions 2 to 4 are to be answered in the query language indicated. Use integrity constraints to simplify your answers. Correct answers with unnecessary components may not get full marks. Subqueries, aggregates and algebraïc operators may be considered unnecessary if they can be avoided.

1. (6 points) This question consists of three multiple-choice questions (OCR1 to OCR3) which are to be answered on the **OCR form**.

   The schema for Part I is repeated below.

   staff (sid, sname)
   project (pid, pdesc, pfrom, pto)
   asset (aid, acat, adesc)
   workfor (sid, pid, wfrom, wto)
   assignment (aid, sid, pid, afrom, ato)

OCR 1 Which of the following queries finds the identifiers and descriptions of all the projects related to Python?

(a) $\{< X1, X2 > \mid \exists X3\ \exists X4$
$(project(X1, X2, X3, X4)\ \wedge\ X2\ LIKE\ '\%Python\%')\}$

(b) $\{< X1, X2 > \mid \forall X3\ \forall X4$
$(project(X1, X2, X3, X4)\ \Rightarrow\ X2\ LIKE\ '\%Python\%')\}$

(c) $\{< X1, X2 > \mid \forall X3\ \forall X4$
$(project(X1, X2, X3, X4)\ \wedge\ X2\ LIKE\ '\%Python\%')\}$

(d) All of the above.

(e) None of the above.

OCR 2 Which of the following queries finds the identifiers of the assets that have never been assigned?

(a) $\{< X1 > \mid \exists X2\ \exists X3\ \forall Y1\ \forall Y2\ \forall Y3\ \forall Y4\ \forall Y5$
$(asset(X1, X2, X3)\ \wedge\ (assignment(Y1, Y2, Y3, Y4, Y5) \Rightarrow X1 \neq Y1))\}$

(b) $\{< X1 > \mid \exists X2\ \exists X3\ \forall Y1\ \forall Y2\ \forall Y3\ \forall Y4\ \forall Y5$
$(asset(X1, X2, X3)\ \wedge\ (\neg(assignment(Y1, Y2, Y3, Y4, Y5)) \vee X1 \neq Y1))\}$

(c) $\{< X1 > \mid \exists X2\ \exists X3\ \forall Y1\ \forall Y2\ \forall Y3\ \forall Y4\ \forall Y5$
$(asset(X1, X2, X3)\ \wedge\ \neg(assignment(Y1, Y2, Y3, Y4, Y5) \wedge X1 = Y1))\}$

(d) All of the above.

(e) None of the above.

OCR 3 Which of the following queries finds the identifiers of the assets that have never been assigned?

(a) $\{< X1 > \mid \exists X2\ \exists X3\ \forall Y2\ \forall Y3\ \forall Y4\ \forall Y5$
$(asset(X1, X2, X3)\ \wedge\ \neg(assignment(X1, Y2, Y3, Y4, Y5)))\}$

(b) $\{< X1 > \mid \forall Y2\ \forall Y3\ \forall Y4\ \forall Y5$
$(\neg(assignment(X1, Y2, Y3, Y4, Y5))\}$

(c) $\{< X1 > \mid \forall X2\ \forall X3\ \forall Y1\ \forall Y2\ \forall Y3\ \forall Y4\ \forall Y5$
$(asset(X1, X2, X3)\ \wedge\ (assignment(Y1, Y2, Y3, Y4, Y5) \Rightarrow X1 \neq Y1))\}$

(d) All of the above.

(e) None of the above.

The schema for Part I is repeated below.

staff (sid, sname)
project (pid, pdesc, pfrom, pto)
asset (aid, acat, adesc)
workfor (sid, pid, wfrom, wto)
assignment (aid, sid, pid, afrom, ato)

2. (7 points) (SQL) Find the names of the different staffs who worked or are working for more than 20 projects that started in March 2017 (each of the 20 projects started in March 2017). Print the names of these different staffs and the corresponding numbers of projects (numbers of projects refers to the quantity of projects).

The schema for Part I is repeated below.

staff (sid, sname)
project (pid, pdesc, pfrom, pto)
asset (aid, acat, adesc)
workfor (sid, pid, wfrom, wto)
assignment (aid, sid, pid, afrom, ato)

3. (7 points) (TRC) Find the names of the staffs who worked or are working for more than one project.

The schema for Part I is repeated below.

staff (sid, sname)
project (pid, pdesc, pfrom, pto)
asset (aid, acat, adesc)
workfor (sid, pid, wfrom, wto)
assignment (aid, sid, pid, afrom, ato)

4. (7 points) (TRC) Find the identifiers and categories of the assets related to Python that have been assigned to all projects also related to Python.

Part II: Relational Algebra & Database Design

5. (10 points) **This question consists of 5 multiple-choice questions (OCR4 to OCR8) which are to be answered on the OCR form.**

OCR 4 Consider the schema $R(A, B, C, D, E)$. Which of the following statements is the most appropriate about R?

(a) If $A \to BC$ does not hold on R, then $A \to B$ also does not hold on R.

(b) If $AB \to C$ holds on R and $CD \to E$ holds on R, then $ABD \to E$ must also hold on R.

(c) Both (a) and (b) are true.

(d) Both (a) and (b) are false.

OCR 5 Consider the following two sets of FDs on $R(A, B, C, D, X, Y)$:

$$F = \{A \to Y,\ AB \to CD,\ BC \to XY,\ C \to Y,\ Y \to B\}$$

$$G = \{A \to C,\ A \to D,\ A \to Y,\ C \to X,\ C \to Y,\ Y \to B\}$$

Which of the following statements is the most appropriate about F and G?

(a) $F \models G$.

(b) $G \models F$.

(c) $F \equiv G$.

(d) G is a minimal cover for F.

(e) None of the above.

OCR 6 Consider the schema $R(A, B, C, D, E, X, Y)$ with FDs

$$F = \{A \to BC,\ C \to XY,\ E \to DY,\ Y \to A\}$$

Let G be the set containing all the non-trivial FDs in F_{AEX} that are of the form $\alpha \to I$, where $\alpha \subseteq R$ and $I \in R$. How many FDs are there in G?

(a) 2

(b) 3

(c) 4

(d) 5

(e) None of the above.

CONTINUATION OF QUESTION 5

OCR 7 Consider the schema $R(A, B, C, D, E)$ with FDs F (not shown) and the decomposition of R (w.r.t. F) into $\{R_1(A, B, C),\ R_2(C, D, E)\}$. The following two relations, r_1 and r_2, are legal instances of R_1 and R_2, respectively.

r_1

A	B	C
0	1	2
1	3	2
3	2	0

r_2

C	D	E
2	2	3
0	1	1
0	2	4

Based on the given information, which of the following statements is the most appropriate about the decomposition of R into $\{R_1, R_2\}$?

(a) The decomposition is a lossless-join decomposition (w.r.t. F).

(b) The decomposition is not a lossless-join decomposition (w.r.t. F).

(c) The given information is not sufficient to determine whether or not the decomposition is a lossless-join (w.r.t. F).

OCR 8 Consider the following database schema where the primary key of each table is underlined.

`Restaurants` (rname, area)
`Contains` (pizza, ingredient)
`Sells` (rname, pizza, price)

Restaurants indicates the name and location of each restaurant. **Contains** indicates the ingredients used in each pizza. **Sells** indicates the pizzas sold by restaurants and their prices.

Assume the following additional constraints:

- $\pi_{rname}(Sells) \subseteq \pi_{rname}(Restaurants)$
- $\pi_{pizza}(Sells) \subseteq \pi_{pizza}(Contains)$

Consider the following three queries on the schema:

Q_1: $\pi_{pizza}(Sells \ltimes \sigma_{area='East'}(Restaurants)) - \pi_{pizza}(Sells \ltimes \sigma_{area\neq'East'}(Restaurants))$
Q_2: $\pi_{pizza}(Sells \ltimes \sigma_{area='East'}(Restaurants)) \triangleright \pi_{pizza}(Sells \ltimes \sigma_{area\neq'East'}(Restaurants))$
Q_3: $\pi_{pizza}(Sells \triangleright \sigma_{area\neq'East'}(Restaurants))$

Which of the following statements is the most appropriate about the queries?

(a) $Q_1 \equiv Q_2$ and $Q_1 \equiv Q_3$.

(b) $Q_1 \equiv Q_2$ and $Q_1 \not\equiv Q_3$.

(c) $Q_1 \not\equiv Q_2$ and $Q_1 \equiv Q_3$.

(d) $Q_1 \not\equiv Q_2$, $Q_1 \not\equiv Q_3$, and $Q_2 \equiv Q_3$.

(e) None of the above.

6. (7 points) This question is based on the same database schema in OCR8 which is repeated below:

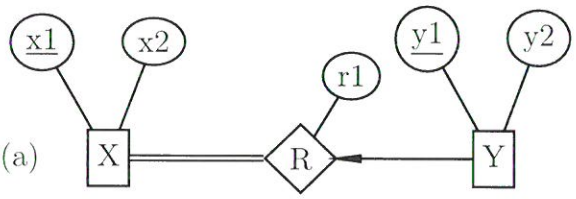
   `Restaurants` (rname, area)
   `Contains` (pizza, ingredient)
   `Sells` (rname, pizza, price)

   Given two restaurants R1 and R2, we say that R1 is more expensive than R2 if for every pizza P that is sold by both R1 and R2, R1's selling price for P is higher than R2's selling price for P. Write a relational algebra query to find all pairs of restaurants (R1,R2) where R1 is more expensive than R2. Exclude restaurant pairs that do not sell any common pizza.

7. (6 points) This question consists of two parts. Each part shows an Entity-Relationship diagram (ERD) and its corresponding relational mapping (RM). Assume that the domain of each attribute is INTEGER.

   For each part, write down all the inaccuracies (if any) of the RM. The inaccuracies could be missing constraints (i.e., constraints specified in the ERD that are not captured by the RM) or extraneous constraints (i.e., constraints that are specified by the RM that are not specified in the ERD).

(a)

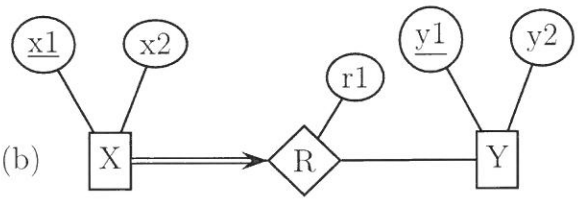



```
CREATE TABLE Y (
  y1        integer,
  y2        integer,
  PRIMARY KEY (y1));


CREATE TABLE X (
  x1        integer,
  x2        integer,
  y1        integer UNIQUE NOT NULL,
  r1        integer,
  PRIMARY KEY (x1),
  FOREIGN KEY (y1) REFERENCES  Y);
```

CONTINUATION OF QUESTION 7

(b)




```
CREATE TABLE Y (
  y1        integer,
  y2        integer,
  PRIMARY KEY (y1));

CREATE TABLE R (
  x1        integer,
  y1        integer NOT NULL,
  r1        integer,
  PRIMARY KEY (x1),
  FOREIGN KEY (y1) REFERENCES Y);

CREATE TABLE X (
  x1        integer,
  x2        integer,
  PRIMARY KEY (x1),
  FOREIGN KEY (x1) REFERENCES R);
```

8. (10 points) Consider the schema $R(A, B, C, D, E, X, Y)$ with FDs

$$F = \{AB \to CD,\ CDE \to XY\}$$

Determine whether or not there exists a dependency-preserving, lossless-join BCNF decomposition of R. If it exists, compute this decomposition. If it does not exists, compute a lossless-join BCNF decomposition of R.

PAGE LEFT BLANK FOR QUESTION 8

END OF PAPER