# NATIONAL UNIVERSITY OF SINGAPORE

## CS2102 - DATABASE SYSTEMS

(Semester 2: AY2017/18)

Time Allowed: 2 Hours

## INSTRUCTIONS TO STUDENTS

1. This assessment paper contains **TWELVE** questions and comprises **SIXTEEN** printed pages, including this page.
2. This is a **CLOSED BOOK** assessment. You are allowed to refer to a single, double-sided A4-sized sheet of notes.
3. **Questions 1 and 7** consists of a total of 8 multiple-choice questions (OCR1, OCR2, $\cdots$, OCR8) to be answered on the **OCR form** using only 2B pencil. No mark is deducted for wrong answers.
4. Except for Questions 1 and 7, all questions must be answered within the space provided in this question paper. You may use pen or pencil to write your answers, but please write legibly. Marks may be deducted for illegible handwriting.
5. Write your **Student Number** below. Do not write your name.
6. You are to submit both this **question paper** and the **OCR form**.

STUDENT NUMBER: | | | | | | | | | |

**EXAMINER'S USE ONLY**

| Question: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Points: | 6 | 4 | 4 | 5 | 2 | 5 | 2 | 4 | 4 | 5 | 5 | 4 | 50 |
| Score: | | | | | | | | | | | | | |

2. (4 points) Prove, using the Armstrong axioms only, that $\{A, C, D\}$ is a superkey of $R$ with $F$.

3. (4 points) What are the candidate keys of $R$ with $F$? (Do not explain.)

4. (5 points) Apply the BCNF Decomposition Algorithm from the lecture to decompose $R$ with $F$ into $R1 = \{B, C\}, R2 = \{D, E\}$, and $R3 = \{A, C, E\}$. For each step, indicate the functional dependency used to decompose a relation. Explain how you construct the fragments by indicating the projected functional dependencies on the fragments and the candidate keys of the fragment with the set of projected functional dependencies.

PAGE LEFT BLANK FOR QUESTION 4

5. (2 points) Is the decomposition of $R$ with $F$ into $R1 = \{B, C\}, R2 = \{D, E\}, R3 = \{A, C, E\}$ dependency preserving? Justify your answer very briefly.

6. (5 points) Apply the 3NF Synthesis Algorithm from the lecture to decompose $R$ with $F$. Show the steps (including the computation of the extended minimal cover).

PAGE LEFT BLANK FOR QUESTION 6

PAGE LEFT BLANK FOR QUESTION 6

7. (2 points) This question consists of 2 multiple-choice questions (OCR7 and OCR8) to be answered on the **OCR form**.

Consider a database consisting of a single relation $R$ and the following four transactions $T_1$, $T_2$, $T_3$, and $T_4$. Assume that each transaction is executing using some unknown transaction isolation level.

R

| A | B |
|---|---|
| 1 | 10 |
| 2 | 20 |

Transaction $T_1$

```
begin transaction;
update R set B = 100 where A = 1;
update R set B = 200 where A = 2;
commit;
```

Transaction $T_2$

```
begin transaction;
select B from R where A = 1;
select B from R where A = 2;
commit;
```

Transaction $T_3$

```
begin transaction;
select B from R where A = 2;
update R set B = 1000 where A = 1;
commit;
```

Transaction $T_4$

```
begin transaction;
select B from R where A = 1;
update R set B = 2000 where A = 2;
commit;
```

Let $B_{2,1}$ and $B_{2,2}$ denote the values returned by the first and second **select** statements in $T_2$, respectively. Let $B_3$ and $B_4$ denote the values returned by the **select** statement in $T_3$ and $T_4$, respectively.

Answer each of the two questions independently.

OCR 7 Consider the concurrent execution of only transactions $T_1$ and $T_2$. If both $T_1$ and $T_2$ successfully commit, which of the following statements is the most appropriate about the concurrent execution of $T_1$ and $T_2$?

(a) If $B_{2,1} = 10$ and $B_{2,2} = 20$, then the execution is serializable.
(b) If $B_{2,1} = 100$ and $B_{2,2} = 200$, then the execution is serializable.
(c) If $B_{2,1} = 10$ and $B_{2,2} = 200$, then the execution is serializable.
(d) Only (a) and (b) are true.
(e) None of the above.

OCR 8 Consider the concurrent execution of only transactions $T_3$ and $T_4$. If both $T_3$ and $T_4$ successfully commit, which of the following statements is the most appropriate about the concurrent execution execution of $T_3$ and $T_4$?

(a) If $B_3 = 20$, and $B_4 = 10$, then the execution is serializable.
(b) If $B_3 = 20$, and $B_4 = 1000$, then the execution is serializable.
(c) If $B_3 = 2000$, and $B_4 = 10$, then the execution is serializable.
(d) All of the above.
(e) None of the above.

**Questions 8 to 12 are based on the following database schema.**

```
create table Students (
   sid integer primary key,
   sname varchar(50) not null,
   major varchar(50) not null
);

create table Courses (
   cid integer primary key,
   cname varchar(50) not null,
   cdept varchar(50) not null
);

create table Profs (
   pid integer primary key,
   pname varchar(50) not null,
   pdept varchar(50) not null
);

create table Offerings (
   semester integer check (semester in (1,2,3,4,5,6,7,8)),
   cid integer,
   pid integer not null,
   primary key (semester, cid),
   foreign key (pid) references Profs (pid),
   foreign key (cid) references Courses (cid)
);

create table Enrolls (
   sid integer,
   cid integer,
   semester integer,
   score integer not null,
   primary key (sid, cid, semester),
   foreign key (sid) references Students (sid),
   foreign key (semester,cid) references Offerings (semester,cid)
);
```

A tuple $(s, n, m)$ in **Students** means that there is a student with identifier $s$ named $n$ whose major is $m$. A tuple $(c, n, d)$ in **Courses** means that there is a course with identifier $c$ named $n$ offered by department $d$. A tuple $(p, n, d)$ in **Profs** means that there is a professor with identifier $p$ named $n$ in department $d$. A tuple $(x, c, p)$ in **Offerings** means that the course with identifier $c$ was offered in semester $x$ and taught by the professor with identifier $p$. A tuple $(s, c, x, m)$ in **Enrolls** means that the student with identifier $s$ obtained a score of $m$ for the course with identifier $c$ offered in semester $x$.

8. (4 points) Write a relational algebra query that is equivalent to the following SQL query.

```
select sid
from Enrolls
group by sid
having max(score) > 70
```

9. (4 points) Write a SQL query to find the identifiers of all professors who have taught some course that was offered by a department different from the professor's department. Remove all duplicate tuples from the output.

10. (5 points) We say that a pair of courses $(c1, c2)$ is *unpopular* in semester $x$ if (a) both $c1$ and $c2$ are offered in semester $x$, and (b) the number of students who are enrolled in both $c1$ and $c2$ in semester $x$ is at least 1 and at most 5.

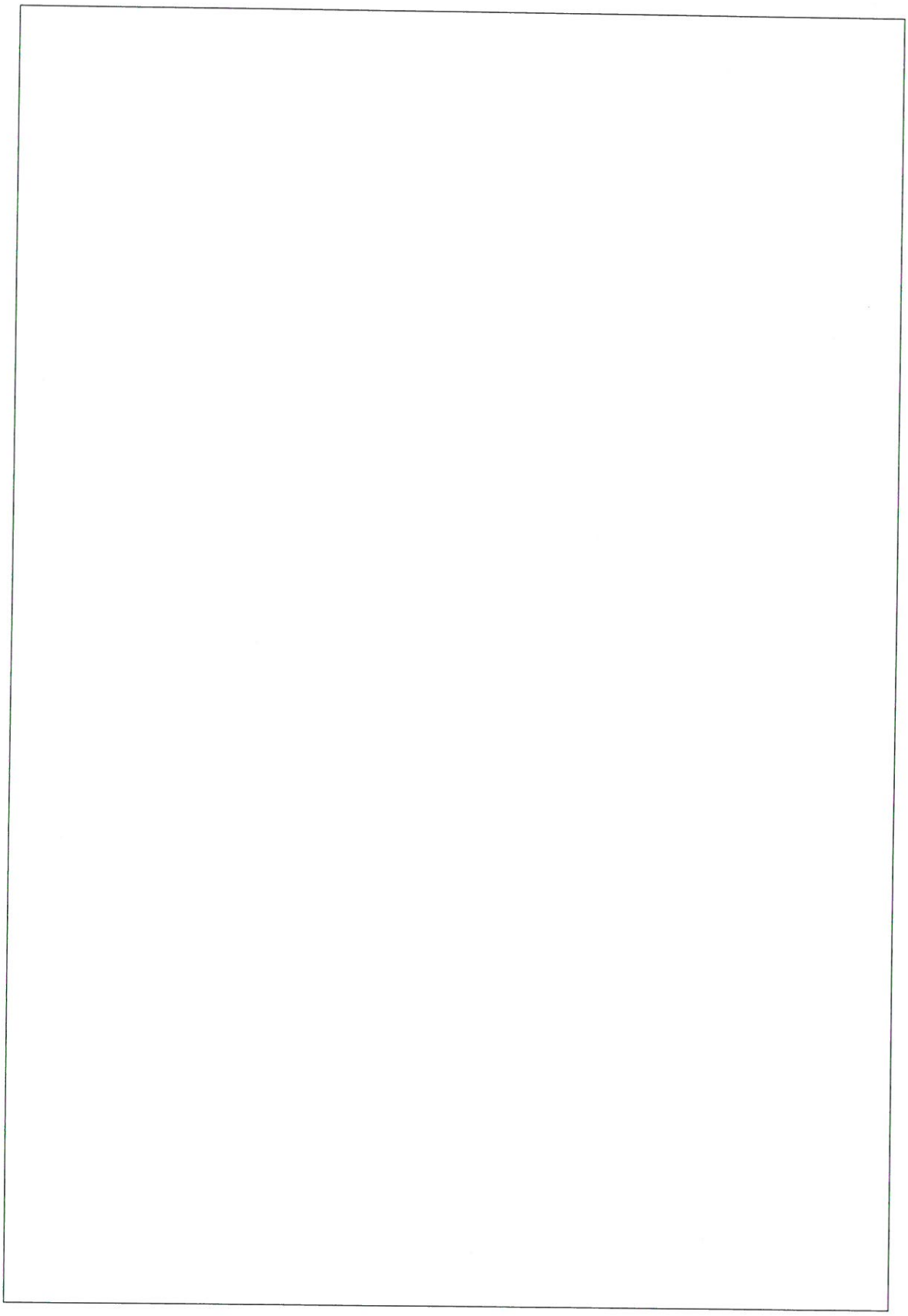
    Write a SQL query to find all $(c1, c2, x)$ tuples where $(c1, c2)$ is a unpopular pair of courses in semester $x$ and $c1 < c2$. Remove all duplicate tuples from the output.

    Note that the number of students enrolled in a pair of courses $(c1, c2)$ is not to be confused with the sum of the number of students enrolled in $c1$ and the number of students enrolled in $c2$.

11. (5 points) Consider a student with identifier $s$ who has obtained a score of $m$ for a course with identifier $c$ offered in semester $x$. We say that $s$ has performed well for course $c$ in semester $x$ if for every semester $y$ that the course $c$ was offered, $m$ is higher than the average score obtained by the students enrolled for course $c$ in semester $y$.

    Write a SQL query to find all $(s, c, x)$ tuples where the student with identifier $s$ has performed well in the course with identifier $c$ offered in semester $x$.

12. (4 points) Write a tuple relational calculus query to find the names of the students who are not enrolled in any courses in semester 2. Do not use $\Rightarrow$.



**END OF PAPER**