School of Computer Science
The University of Adelaide

Distributed Databases and Data Mining
Assignment 2

Due 23:59 pm, Friday 1 October 2021

● **Introduction**

Assignment 2 has four questions. Both UG and PG students need to finish all four questions. The first two questions are the same for UG and PG students, while the last two questions have different requirements for UG and PG students.

Given the following relations:

    STUDENT (SID, SNAME, SAGE, SSEX)
    COURSE (CID, CNAME, CCONTEXT)
    MARK (SID, CID, SCORE)

Please answer the questions in the format of descriptions, equations, query graphs and the operator trees in question 1-4. Only the query graph and the operator tree can be handwritten; others such as **descriptions and equations must be in electronic version**. For the handwritten parts, you could save them as jpg. or png., and insert them into office word. Please submit the **electronic version** and do not submit the handwritten version of descriptions and equations. Handwritten descriptions and equations will not be marked.

You are expected to provide your answers for the all four questions in **one pdf** and submit the pdf in MyUni, Assignment 2.

If you have any questions about assignment 2, please feel free to post the questions on the discussion board or contact Congbo Ma (congbo.ma@adelaide.edu.au).

● **Questions**

Question 1 (4 marks):

Given three relations STUDENT, COURSE and MARK, a user query is: "Find the student name and Course name of students whose scores are more than 85 points."

(1) Please give the clasulus query (SQL) (1 mark);

(2) Please give the corresponging query graph (2 mark);

(3) Please map the SQL query into an **original operator tree** (Not optimized) (1 mark).

Question 2 (4 marks):

For the following query:

```
SELECT   SNAME, CNAME
FROM     STUDENT, COURSE, MARK
WHERE    (SAGE < 16
```

```
OR        MARK.CID < "C2")
AND       (CCONTEXT = "Science"
OR        SCORE > 85)
AND       STUDENT.SID = MARK.SID
AND       (CCONTEXT = "Science"
OR        SCORE ≤ 85)
AND       COURSE.CID = MARK.CID
```

(1) Please simplify the query (2 mark);

(2) Please transform the SQL expression into an **optimized operator tree** using the restructuring algorithm where select and project operations are applied as soon as possible to reduce the size of intermediate relations (2 mark).

Question 3 (5 marks):

Assume that relation STUDENT of the sample database is horizontally fragmented:

$$STUDENT_1 = \sigma_{SID \geq "S_3"}(STUDENT)$$
$$STUDENT_2 = \sigma_{"S_1" \leq SID < "S_3"}(STUDENT)$$
$$STUDENT_3 = \sigma_{SID < "S_1"}(STUDENT)$$

relation MARK is indirectly fragmented as：

$$MARK_1 = MARK \ltimes_{SID} (STUDENT_2)$$
$$MARK_2 = MARK \ltimes_{SID} (STUDENT_1)$$
$$MARK_3 = MARK \ltimes_{SID} (STUDENT_3)$$

and relation COURSE is vertically fragmented as：

$$COURSE_1 = \prod_{CID,CNAME}(COURSE)$$
$$COURSE_2 = \prod_{CID,CCONTEXT}(COURSE)$$

Please describe the detailed processes of how to transform the following query into a reduced query on fragments:

For **undergraduate** students, the query is (5 marks):

```
SELECT  CNAME
FROM    STUDENT, COURSE, MARK
WHERE   STUDENT.SID = MARK.SID
AND     COURSE.CID = MARK.CID
AND     SNAME = "James"
```

For **postgraduate** students, the query is (5 marks):

```
SELECT  CNAME
FROM    STUDENT, COURSE, MARK
WHERE   STUDENT.SID = MARK.SID
AND     COURSE.CID = MARK.CID
AND     SNAME = "James"
AND     SCORE > 85
```
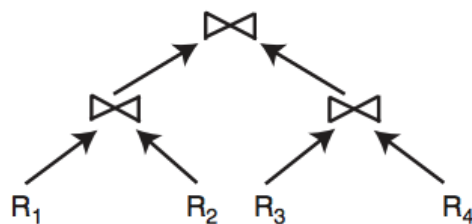
Question 4 (7 marks):

The first table shows the placement of the copies of the 4 relations at the 4 sites and the site loads:

sites	load	R_1	R_2	R_3	R_4
s_1	1	R_{11}		R_{31}	R_{41}
s_2	2		R_{22}		
s_3	2	R_{13}		R_{33}	
s_4	2	R_{14}	R_{24}		

Please describe the detailed processes of how to extend the **SQAllocation algorithm** to the following bushy join trees using the data placement and site loads shown in the table.

For **undergraduate** students, you only need to provide **one** solution (7 marks).
For **postgraduate** students, you need to provide **two** solutions (4 marks for the first solution and 3 marks for the second solution).