MANDATORY ASSIGNMENT #1

IN 3020 / IN 4020

Due Friday, 12 March 2021

Exercises in this mandatory assignment must be solved and submitted individually.

Completion and submission of the mandatory assignment must take place in accordance with current guidelines of the Department of Informatics (Institutt for Informatikk), see:

www.uio.no/studier/admin/obligatoriske-aktiviteter/mn-ifi-oblig.html

Submission of the mandatory assignment will be taken as the confirmation that guidelines have been read and understood.

Deadline: Friday, 12 March 2021 at 23:59 (by midnight).

The deadline is absolute, and deferral will only be granted in accordance with the mandatory assignment rules (obligreglement).

All questions must be answered to get an approval.

Exercise 1 Some SQL operations

Look into the four relation schemas with some example tuples presented below and express in SQL the tasks (a–f) which follow. (Please note that your SQL statements are not dependent on the tuples.)

Course

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

Section

Section_Identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	04	King
92	CS1310	Fall	04	Anderson
102	CS3320	Spring	05	Knuth
112	MATH2410	Fall	05	Chang
119	CS1310	Fall	05	Anderson
135	CS3380	Fall	05	Stone

Grade_Report

Student_number	Section_Identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

Student

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

a) For each section taught by professor King, retrieve the course number, semester, year, and number of students who took the section.

b) Retrieve the names and major departments of all straight-A students (students who have a grade of A in all their courses).

c) Insert a new student < ‘Johnson’, 25, 1, ‘MATH’> into the database.

d) Change the class of student ‘Smith’ to 2.

e) Delete records for the students whose name is ‘Smith’ and student number is 17.

Exercise 2 Introduction to DBMS Architecture

2.1 Show, schematically, the locations of the typical management systems in the Query and Transaction Execution (QTE) part of a DBMS architecture and briefly explain the functions of each management system. (Please show the schematic of the entire architecture; you do not need to draw all the other components, which are not part of the QTE; instead, show those in one bigger box with appropriate label and briefly mention in 2–3 sentences how QTE is interfaced/connected to those parts.)

2.2 What are the Central components and the Higher-level components in a DBMS? Briefly mention the interactions between each pair of the Central components. (Note: this part does not ask for detailed descriptions of the interactions.)

Exercise 3 Indices and index structures

3.1 List up the data structures for indices that we have seen and explain their respective advantages and disadvantages, as well as their uses.

3.2 Explain what an inverted index is and mention at least one example of the uses of an inverted index.

3.3 Can you think of techniques other than an unordered overflow file that can be used to make insertion in an ordered file more efficient?

Exercise 4 Relational Algebra

4.1 Consider the following two schemas – Purchase and Supply – with the attributes given in the parentheses:

Purchase(A, B, C)
Supply(A, D, F)

a) Rewrite the following SQL query into a relational algebra (RA) expression. Show the steps that you have followed.

```
SELECT P.C
FROM Purchase P
WHERE P.A IN
     (SELECT S.A FROM Supply S WHERE S.D > 0)
```

b) Translate the RA expression in a) into an equivalent RA expression that you expect to be evaluated more efficiently. Write down the RA laws you have applied in the translation.

c) Explain why your optimized expression is better than the original expression.

4.2 Consider the following three schemas – P, Q, and R – with the attributes given in the parentheses:

P(A, B, C)
Q(A, D, H)
R(F, H, K, M)

Let $<c>$ be the condition that attribute C has value equal to 1 and $<d>$ be the condition that attribute D has value equal to 0.

a) Give a sequence of equivalence steps for the following RA transformation:

$\pi_M(\sigma_{<c>\wedge<d>}(\text{P} \bowtie \text{Q} \bowtie \text{R})) \rightarrow \pi_M[\{\pi_A(\sigma_{<c>}(\text{P})) \bowtie \pi_{A,H}(\sigma_{<d>}(\text{Q}))\} \bowtie \text{R}]$

b) Which of these two expressions you think might perform better in terms of performance efficiency? Justify your answer.

Exercise 5 – Query Compilation

From 2020 exam

5.1 We went through the steps of query processing as an introduction to query compilation. Sketch the steps with the inputs/outputs for each step and the relations between them. Describe their roles/functions briefly in your own words.

5.2 Consider the four schemas in Exercise 1.1 and do the following:

a) Write an SQL query that retrieves the course name, semester, year, and instructor for the course "Database Systems" since (year) 2010.

b) Translate the SQL query from a) into an RA expression.

c) Draw the RA tree of the RA expression you give in b).

d) Translate the expression from b) into an equivalent but improved RA expression and draw the corresponding RA tree.

e) Explain with numerical calculations why your improved tree is likely to be evaluated more efficiently than the original tree. (Make necessary and relevant assumptions on the size of the tuples, the no. of tuples, etc., when you perform the calculations.)

End of the assignment paper

Good luck!