Lab week 10 (ISYS2120 sem2 2024)

This week, lab is again one hour only (except for those with Monday class, where there is no lab this week, and the material will instead be covered during a 2 hour session in week 11). This week we are going to focus on practice with the concepts of index structure and its impact on query performance.

Every group should have a 20-minute meeting with the tutor this week. If your meeting is usually on Monday, be in touch with the tutor to arrange a different time! As described in asst3 instructions, each student should show the SQL commands they will use, at the heart of their code, for each of the required functionalities. The group can also use some time to describe what they have considered about security goals and possible attacks to consider, for the application. The tutor can provide feedback and suggestions.

Reminder.
Tutors will manage time in their class, aiming to cover the most crucial material. Should students have any additional questions or need further clarification, we strongly encourage them to utilize the ED platform for further assistance. Additionally, we will be releasing sample answers on Canvas at the end of the week. This will give students a chance to review the material and check their understanding even after the tutorial session.

Before the lab
Read through the instructions for this lab! If anything is unclear, post on Ed about it.

During the lab
For week 10, there are two activities, with details below:

- Tutor presents an overview of the concepts and skills of index structure and choice,
- Questions on using these concepts and skills

A. Tutor presents an overview
The tutor will give a quick refresh of some key aspects of the material needed for this lab, and in particular, they will try to fix any gaps or misunderstanding they saw commonly in the hand-written summaries.

B. Index Impact

Consider executing the following commands, and then inserting a large amount of data into the table.

```
CREATE TABLE UnitOfStudy (
  uoSCode       CHAR(8),
  deptId        CHAR(3)  NOT NULL,
  uoSName       VARCHAR(40) NOT NULL,
  credits       INTEGER NOT NULL,
  fee           INTEGER NOT NULL,
  PRIMARY KEY (uoSCode),
  UNIQUE (deptId, uoSName)
);

CREATE INDEX UoS_Name_IX ON UnitOfStudy (uoSName);
```

Notice that the schema indicates that different uoSCode could have the same uoSName, as long as the units are from different departments.

With this CREATE TABLE statement, a typical system will structure this table with a primary clustering tree-structured index on uoSCode (the primary key), and then the second command creates a secondary unclustered tree-structured index on the attribute uoSName.

B(i) Describe how the system could calculate the result of the following query

```
SELECT uoSName
FROM UnitOfStudy
WHERE uoSCode = 'ISYS2120';
```

B(ii) Describe how the system could calculate the result of the following query

```
SELECT uoSCode
FROM UnitOfStudy
WHERE uoSName = 'Data and Information Management' AND
deptId = 'SCS';
```

B(iii) For each of the queries in B(i) or B(ii), would it be an improvement if the indices were done as hash-based rather than tree-based?

B(iv) Explain why the system will not be able to calculate the result of the query below *efficiently* (that is, with substantially less disk accesses, than simply scanning through the whole table)

```
SELECT uoSCode
FROM UnitOfStudy
WHERE (fee BETWEEN 5000 and 10000) AND deptId = 'SCS';
```

B(v) Declare an index to allow the query from B(iii) to be answered efficiently. Can you declare an index that covers the query?

B(vi). Consider the fact about likely data at this university, where most units have credits = 6, and most of the others have values of 2, 12 or 24. What does this suggest about an index to create for speeding the query

```
SELECT uoSName
FROM UnitOfStudy
WHERE (credits BETWEEN 6 AND 12);
```

After the lab session

- Before the end of **Friday** (October 11) look over the week 11 lecture slides on Canvas, and use them to prepare your hand-written half-page summary11 and upload it to Canvas
- Before the end of **Sunday** October 13, complete SQL Tasks 15 and 16 [these are the last SQL Tasks for the semester]
- To check your basic knowledge from the slides, you are encouraged to do the multichoice quiz8 on Canvas (this can be repeated as often as you wish, but it does not count to your grade).