DBMS Final Exam (Take Home)

2021 Spring

NOTICES

※ The exam **due time is 7:00 pm, today**. Make your answer sheet into a **pdf file** and upload it to the designated place in Moodle. The **file name** of your answer sheet shall be "**student#_name.pdf**", e.g., P12345678_張三.pdf.

※ To avoid network congestion, we strongly suggest that you DONOT wait until the last minute to upload your answer file. Do **upload as early as possible**. If there is a network problem encountered while uploading your answer file and **the time is close to 7:00 pm**, you may email your answer file to ta_@dblab.csie.ncku.edu.tw. However, the emailing time must be before 7:00pm. **LATE SUBMISSION IS NOT ACCEPTABLE IN ANY CASE**.

※ If your answer file is successfully uploaded to Moodle, then please DONOT email your answer file to TAs to increase their workload.

※ If we find an answer file of yours in Moodle and also receive an emailing answer file from you, **we will use the file in Moodle as your answer** and will ignore the answer file from email.

QUESTIONS

1. (20%) Given a database schema as follows.

   Student(SID, SName, SAge, Address)
   Teacher(TID, TName, TAge, Office#)
   Course(CID, CTitle, Credit)
   Take(CID, TID, SID, Score)

   Answer the following queries in SQL.

   (a) (5%) For those students who are younger than 18 (i.e., SAge<18), list each student's SID and the average score of the courses that a student takes.

   (b) (5%) List the course title and the name of the teacher who lectures the course for the course that has more than 5 (i.e., > 5) students taking the course whose age is greater than the teacher's age.

   (c) (10%) For each course that is 3 credit hours (i.e., Credit=3) and more than 50 students take the course, list the course title, the name of the teacher who lectures the course, and the number of students taking this course whose score is above 95. Assume that there won't be two teachers teaching the same course.

2. (10%) An inference rule known as *the transitivity of functional dependency* says if X→Y and Y→Z then X→Z, where X, Y, and Z are sets of attributes of a relation. Based on this rule, answer the following questions for relation R(A, B, C, D, E, F, G, H, I, J) in which there exist functional dependencies AB→C, A→DE, B→F, F→GH, and D→IJ.
   (a) Give the result of the second normal form, underline the key attribute(s) of each relation, and state your reasons. /* Unnecessary normalization will be considered as wrong answer. */
   (b) Give the result of the third normal form, underline the key attribute(s) of each relation, and state your reasons.

3. (10%) Answer the following questions.
   (a) Explain why the two-phase locking protocol can guarantee serializability in transaction execution.
   (b) What's the major disadvantage of the two-phase locking protocol?

4. (10%) The time-stamping protocol does not require locking anything when two transactions are competing for the same data item.
   (a) So how can the time-stamping protocol avoid two transactions that access the same data item? What is the essential idea of this protocol in arranging the conflicting transactions?
   (b) The major advantage of the time-stamping protocol is that it relieves the need of locks in transaction processing so that deadlocks will never occur. Then, why hasn't it been widely used in commercial DBMS products? What is the major problem of this protocol?

5. (15%) Answer the following questions.
   (a) What is "write-ahead-log"?
   (b) In the deferred update recovery mechanism, if a transaction begins after the last checkpoint and finishes before the coming crash, what will the system do to this transaction? Explain why.
   (c) What's the major advantage and the major disadvantage of the shadow paging recovery protocol? /* Just give ONE key advantage and ONE key disadvantage of this protocol. */

6. (10%) There are some schemes to avoid deadlock in concurrent transaction processing. One is the “wound-wait” protocol and the other the “wait-die” protocol. Which one is more aggressive than the other? Explain your reasons.

7. (10%) In relational database systems, view can be implemented in one of the two ways: materialized or non-materialized. Materialized view means the tuples that meet the view definition are actually stored in the database. Discuss the pros and cons of having a materialized and a non-materialized view.
   /* NOTE: If you simply cut-and-paste any material found on the web without a clear and reasonable explanation, you will get no score in this question. */

8. (15%) When multiple joins are involved in a query, the joins can be performed in different orders in the execution. For example, to perform a query of three joins over four relations such as A ⋈ B ⋈ C ⋈ D, this query can be executed in the following different ways ① ((A ⋈ B) ⋈ C) ⋈ D, ② (A ⋈ B) ⋈ (C ⋈ D), ③ (A ⋈ (B ⋈ C)) ⋈ D, ④ A ⋈ ((B ⋈ C) ⋈ D), and ⑤ A ⋈ (B ⋈ (C ⋈ D)), etc.
   (a) (5%) Will the execution time of these different ways of execution be the same? Explain your reasons.
   (b) (10%) Let |X| be the cardinality of relation X (i.e., |X| gives the number of tuples of X). The join-selectivity factor of X⋈Y is defined as $\frac{|X \bowtie Y|}{|X| * |Y|}$.
   Assume that the join-selectivity factors of all joins are known before query execution. How can this information be utilized in optimizing the execution of this query? Explain your idea in detail.