Instructions:

- Read through the complete exam and note any unclear directives before you start solving the questions. The following guidelines hold:
  - Write clear and neat answers! Answers that cannot be read can obviously not result in any points and unclear formulations can be misunderstood.
  - Assumptions outside of what is stated in the question must be explained. Any assumptions made should not alter the given question.
  - Write your answer on only one side of the paper and use a new paper for each new question to simplify the correction process and to avoid possible misunderstandings.

- A passing grade requires about 50% of the maximum number of points.
1. **Database terminology:**

   Explain the following database concepts:
   
   (a) meta data
   (b) primary index
   (c) referential integrity (sv. referensintegritet)
   (d) full functional dependency (sv. fullt-funktionellt beroende)

2. **Data models - three schema architecture:**

   Explain and give examples on what is meant by the concepts *physical* and *logical data independence*.

3. **Physical database design - indexing:**

   (a) Describe the structure of B+ trees
   (b) Describe the principle for retrieval of data records by using a B+ tree index.

4. **SQL and relational algebra:**

   Assume that we have a literature database where there are two relations (tables) with the following schemas:

   - BOOK(BID, BNAME)
   - CHAPTER(CID, CNAME, LENGTH, BOID)

   , where xID’s represents keys.

   (a) Formulate a query in relational algebra that retrieves book id, book name, chapter id, chapter name and the length of the chapters for the book “Guide Uppsala”. (2p)

   (b) Formulate an SQL query that retrieves the book id, book name, and the number of chapters for each book, i.e. how many chapters each book consists of. (2p)

5. **Concurrency control (sv. samtidighetskontroll):**

   (a) Describe the principles for lock management (sv. läshantering) for transactions controlled by a *two-phase locking protocol* (sv. två-fas läsningsprotokoll).

   (b) What important properties can be stated by transaction schedules controlled by two-phase locking protocols?
6. **Object-oriented and object-relational databases:** 4 pts

(a) What is the purposes of the three main kinds of user-definable database extensibility mechanisms available in an *object-relational* database system (sv. 'objektrelationella databaser')?

(b) Which of the above extensibility mechanisms are lacking or weak in a first generation object-oriented database system, i.e. in *object stores* (sv. 'objektlager')?

---

7. **Query optimization:** 4 pts

(a) What is *selectivity* (sv. 'selektivitet') and how is it used in cost-based query optimization?

(b) What is the worst case complexity of cost-based query optimization and how does one avoid this cost?

(c) Why does cost-based query optimization pay off despite its complexity?

(d) How are operators in execution plans different from relational algebra operators? Give examples.

---

8. **ODBC:** 4 pts

(a) What is ODBC?

(b) What is the difference between JDBC and ODBC?

(c) What is an ODBC driver (sv. 'drivrutin')?

(d) What is the purpose of the ODBC driver manager?

---

Good luck and a Merry, Merry Christmas!

/ Kjell och Tore