Uppsala University
Department of Information Technology
Kjell Orsborn, Tore Risch

Tentamen 2007-08-22
DATABASE TECHNOLOGY - 1DL116, 1MB025, 1DL124

Date .... Wednesday 22 Aug, 2007
Time 09:00-14:00
Teacher on duty .... Kjell Orsborn, phone 471 11 54 or 070 425 06 91

Exam aids .... calculator

Instructions:

- Read through the complete exam and note any unclear directives before you start solving the questions. The following guide lines 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 simplyfy the correction process and to avoid possible misunderstandings.

- A passing grade requires about 50% of the maximum number of points.

1. **Database terminology:** 4 pts

   Explain the following database concepts:

   (a) transaction (sv. transaktion)
   (b) secondary index (sv. sekundärindex)
   (c) recovery (sv. "återhämtning")
   (d) BCNF

2. **Data models and the Three-schema architecture:** 4 pts

   Explain and give examples of what is meant by the two concepts:

   (a) *physical* data independence (sv. fysiskt dataoberoende) (2 pts)
   (b) *logical* data independence (sv. logiskt dataoberoende) (2 pts)

3. **Data integrity (sv. dataintegritet):** 4 pts

   (a) Explain the term referential integrity (sv. referensintegritet) within the relational data model (2 pts)
   (b) Let $r1(R1)$ and $r2(R2)$ be relations with primary keys (sv. primärnycklar) $K1$ and $K2$ respectively. The attribute fk of the relational schema $R2$ is a foreign key (sv. främmade nyckel) referencing $K1$ of the relational schema $R1$. What tests must be made in order to preserve the referential integrity constraint (sv. bivillkor) during an UPDATE operation. (2pts)

4. **SQL:** 4 pts

   Assume that we have a product database consisting of two relations (tables) with the following schemas:

   ```
   PRODUCT(PID,PNAME)
   COMPONENT(CID,CNAME,WEIGHT,COST,PID)
   ```

   , where xID's represent keys.

   (a) Formulate a query in relational algebra that retrieves the product id and name, the component id and name, and the weight of the product named "Transporter v0.9b". (2pts)

   (b) Formulate an SQL query that retrieves the product id, name and the number of components that *each* product consists of. (2pts)

5. **Physical database design:** 4 pts

   Explain the organization and functionality of hash-files (hash-filer). The answer should include how to retrieve a data record (sv. datapost) with regard to a specific search key (sv. söknyckel) of the hash-file.

6. **Query optimization:**   4pts

  (a) How is selectivity (sv. selektivitet) measured? Why is it very important in cost-based query optimization (sv. kostnadsbaserad frågeoptimering)? (1 pt)

  (b) What is the worst case complexity of a cost-based query optimizer? (1 pt)

  (c) In what language are optimized SQL queries, i.e. execution plans (sv. exekveringsplaner), expressed? (1 pt)

  (d) How does the query interpreter (sv. frågeinterpretator) handle very large intermediate results (sv. mellanresultat) produced in an execution plan? (1 pt)

7. **Database APIs:**   4pts

  (a) What is the difference between JDBC and ODBC? (1 pt)

  (b) What does the "O" in ODBC abbreviate? (1 pt)

  (c) How does JDBC handle very large query results? (1 pt)

  (d) How can one avoid the high cost of query optimization when using JDBC (1 pt)

8. **Data Warehouses:**   4pts

  (a) What does OLAP stand for? (1 pt)

  (b) What is a star schema (sv. stjärnschema) and when do they occur? (1 pt)

  (c) What is the data cube operator in modern SQL and what is it used for? (1 pt)

  (d) Why are data warehouses normally stored in a separate DBMS from an operational database (sv. produktionsdatabas) ? (1 pt)

Good luck and have a great summer!

/ Kjell och Tore