

Data Management – exam of 11/02/2011

LAST NAME:

FIRST NAME:

STUDENT CODE:

I allow the publication of the grade I will get in this exam on the web page <http://www.dis.uniroma1.it/~lenzerini>, according to the “decreto legislativo 196/2003” (regarding the rules for the privacy of personal data) which I hereby confirm to know. Sincerely,

(Signature)

Problem 1 Consider the schedule $S = w_4(W) r_3(Z) r_2(X) r_1(Y) w_3(W) c_3 w_2(Z) w_1(X) c_1 w_4(Y) c_4 c_2$

1. Tell whether S is recoverable or not, and explain the answer.
2. Tell whether S is conflict-serializable. If the answer is yes, then show a serial schedule that is conflict-equivalent to S . If the answer is no, then explain the answer.
3. Tell whether S is view-serializable, and explain the answer.
4. Tell whether S is accepted by the 2PL scheduler with exclusive and shared locks. If the answer is yes, then show the schedule obtained from S by adding suitable lock and unlock commands. If the answer is no, then explain the answer.

Problem 2 Let S be a sequence of read/write actions, and let S' the schedule obtained from S by adding a new read or write action a_i of transaction T_i . Suppose that, just before action a_i , the “wait-for graph” associated to the sequence of actions S is acyclic, while, after processing action a_i , the “wait-for graph” associated to S' has a cycle of length 2 (and, therefore, action a_i is not executed, since transaction T_i is suspended and put in a waiting state). Prove or disprove the following statements (illustrating in detail the proof):

1. the schedule corresponding to the sequence S is a 2PL schedule with shared and exclusive locks;
2. the schedule corresponding to the sequence S' is conflict-serializable.

Problem 3 Consider a data base with relation `Company(code, countryname, city)`, and relation `Country(countryname, capital, area, population, continent)`. Relation `Company` contains 50.000 tuples, and the size of each tuple is $N/50$, where N is the size of a page. Relation `Country` contains 10.000 tuples, and the size of each tuple is $N/25$. Tell which is the cost (in milliseconds) of the natural join between `Company` and `Country`, assuming that

- the buffer has 7 free slots for computing the join;
- the system uses the “Block Nested Loop” algorithm for computing the join;
- the average time for accessing one page is 5 milliseconds;
- we ignore the time for main memory operations,

and motivating the answer in detail.

Problem 4 Illustrate both the “packed” and the “unpacked” organizations for representing pages with fixed length records, and point out the differences between the two organizations.

Problem 5 Consider the relation `Team(name, country, city, sport, president, budget)`, storing information about the teams of the various countries, with their city, sport, president and budget. We know that the relation `Team` has 500.000 tuples, that every page has room for 5 tuples of the relation, that each value of all attributes of the relation requires M bytes, that M bytes are also needed to store record ids, and that, in the average, each country has 100 teams. The most important queries on `Team` are the following:

1. Given a certain country, find the names of all teams of that country.
2. Find the name of all teams, ordered by country.

Which is the file organization you would choose for the relation? Assuming for the relation the chosen file organization, which is the best algorithm for answering the two queries, and which is the cost of the two algorithms in terms of page accesses? Motivate both answers in detail.