Data Management – exam of 26/01/2010

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 following schedule

$$S = r_1(A)\, r_2(A)\, r_2(B)\, w_1(A)\, w_2(D)\, r_3(C)\, r_1(C)\, w_3(B)\, c_2\, r_4(A)\, c_1\, c_4\, w_3(C)\, c_3.$$

1. 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.
2. Tell whether S is strict or not, and explain the answer.
3. Tell whether S is recoverable or not, and explain the answer.
4. 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.

Problem 2 Provide the definition of "monotone class of schedules". Using only the definition of monotone class of schedules, and the definition of 2PL schedules (with exclusive and shared locks), prove or disprove the following statement: the class of 2PL schedules (with exclusive and shared locks) is monotone.

Problem 3 Consider the relation `PRODUCT(prodcode,size,year)`, and the relation `SOLD(prodcode,shopcode,cost)`, where `SOLD` stores information about products sold in shops, with the corresponding cost. We want to compute the equi-join of `PRODUCT` and `SOLD` on the attribute `prodcode`. We know that

- the products are 500.000,
- in every page used for the relation `PRODUCT` we have 10 tuples,
- in the average, every product is sold 20 times,
- we have a B^+-tree index with search key `prodcode` on `SOLD`, using alternative 1, with fan-out 10, and such that every leaf page contains 50 data entries.

If we use the index-nested loop join algorithm for computing the join, which is the cost of the computation in terms of the number of page accesses (ignoring the cost of writing the result)? Explain your answer in detail.

Problem 4 Provide the definition of the "immediate effect" method for writing values in secondary storage.

Problem 5 Suppose that our DBMS uses the "immediate effect" method for writing values in secondary storage, and a failure occurs when the log contains the following records (note that by "CKP" we denote a checkpoint record, and observe that we have *not* shown the active transactions in checkpoint records)

$B(T1); D(T1; O1; A1); B(T2); I(T2; O2; B2); B(T3); B(T4); D(T4; O3; B3); U(T1; O4; B4; A4);$
$C(T3); CKP; B(T5); U(T5; O5; B5; A5); B(T6); CKP; B(T7); C(T1); C(T4);$
$U(T7; O6; B6; A6); U(T6; O3; B7; A7).$

Describe in detail all the actions performed by the recovery manager to deal with the failure.