Data Management – exam of 24/02/2012

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

$$S = r_2(X)\, w_1(X)\, r_4(X)\, r_2(Y)\, w_3(X)\, w_2(Z)\, w_3(Y)\, r_1(Z)\, w_4(Y)$$

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 conflict-serializable or not, and explain the answer. If the answer is yes, then exhibit a serial schedule that is conflict-equivalent to S.
3. Tell whether S is view-serializable or not, and explain the answer. If the answer is yes, then exhibit a serial schedule that is view-equivalent to S.

Problem 2 Provide the definitions of "rigorous schedule", "strict schedule", and "ACR schedule". Then, consider again the schedule S shown in Problem 1, and tell whether such schedule can be made rigorous, i.e., whether is it possible to insert into S the commit operations of the transactions $T_1, \ldots, T_4$ in such a way that the resulting schedule is rigorous. Then, do the same for checking both whether the schedule can be made strict, and whether the schedule can be made ACR.

Problem 3 Let S' be a complete schedule that contains exactly one write action and such that no element of the database is both read and written by the same transaction. Prove or disprove the following statements:

1. S' follows the 2PL locking protocol;
2. S' follows the strong strict 2PL locking protocol.

Problem 4 Assume that a system failure occurs when the log is as follows (where B means "begin", C means "commit", A means "abort", and CK means "checkpoint"):
$B(T_1), I(T1, o_1, a_1), B(T_2), U(T_2, o_1, b_1, a_2), CK(T_1, T_2), C(T_2), B(T_3), D(T_1, o_2, b_2), I(T_3, o_3, a_3),$
$CK(T_1, T_3), B(T_4), I(T_4, o_4, a_5), C(T_3), D(T_4, o_5, b_5), A(T_4), I(T_1, o_6, a_6), B(T_5), U(T_5, o_6, b_6, a_7)$
Illustrate all the actions performed by the recovery subsystem when the failure occurs, assuming that a mixed effect strategy is adopted.

Problem 5 Consider the relation `Apartment(code,area,city)`, where `code` is the primary key of the relation, and for which the relevant queries are the following:

1. Query **Q1**: given a value c for city and a range r of areas, return all apartments whose city is c, and whose area falls into range r;
2. Query **Q2**: given a code d, return the city of the apartment whose code is d.

We know that the size of each attribute value is 4 bytes, the size of each page-id is also 4 bytes, the size of each page is 1.200 bytes, and the number of records of the relation `Apartment` is 66.666.600.

1. Tell which is the method you would choose for representing the relation `Apartment`, explaining your decision.
2. On the basis of the method chosen for representing the relation `Apartment`, tell which is the cost of the execution of query **Q1**, expressed in terms of page accesses.