Data Management – exam of 31/01/2012

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

$$S = r_1(A)\, r_2(B)\, w_1(A)\, r_5(D)\, w_2(D)\, r_3(C)\, r_1(C)\, w_3(B)\, c_2\, r_4(A)\, c_1\, c_4\, w_3(C)\, c_3$$

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. If the answer is no, then tell which is the minimal change to S that makes it strict.
3. Tell whether S is recoverable or not, and explain the answer. If the answer is yes, then tell which is the minimal change to S that makes it non-recoverable.

Problem 2 Consider the following schedule

$$S = w_1(B)\, r_4(A)\, r_3(D)\, w_2(C)\, w_2(D)\, r_5(B)\, w_4(D)\, r_3(A)\, r_1(E)$$

1. 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.
2. Tell whether it is possible to insert into S the commit operations of the transactions $T_1, \ldots, T_5$ in such a way that the resulting schedule is ACR. If the answer is yes, then show such a resulting schedule, otherwise explain the answer.

Problem 3 Let S be a schedule that has no aborted transaction and that is accepted by the timestamp-based scheduler without ever using the Thomas rule. Prove or disprove that S is conflict-serializable.

Problem 4 Let R be a relation of 30.000 pages stored in a heap represented through directory. Each directory page contains 3000 bytes, each entry in every directory page requires 30 bytes, and no page among the ones used to store R has more than 40 bytes free. What is the number of page accesses required for inserting in R a record of size 50 bytes? Explain the answer.

Problem 5 Consider the relation `PROFESSOR(pcode,lastname,age)`, and the relation `EVALUATION(pcode,coursecode,grade)`, where `EVALUATION` stores information about the grades that students gave to professors for the various courses they taught. We want to compute the equi-join of `PROFESSOR` and `EVALUATION` on the attribute `pcode`. We know that

- the professors are 50.000,
- every page used for the relation `PROFESSOR` has 10 tuples,
- in the average, each professor gets 20 grades,
- we have a B^+-tree index with search key `pcode` on `EVALUATION`, 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.