Data Management – AA 2013/14 – exam of 3/6/2014

Problem 1
Let S, S' be two complete schedules on transactions T_1, T_2, T_3, T_4, T_5, and let $r_5(x)$ precedes both $w_3(x)$ and $r_4(x)$ in S. Prove or disprove each of the following two statements:

1. if S' is conflict-equivalent to S, then $r_5(x)$ precedes $w_3(x)$ in S';
2. if S' is conflict-equivalent to S, then $r_5(x)$ precedes $r_4(x)$ in S'.

Problem 2
Let S be the following schedule:

$$w_3(y)\, r_3(t)\, r_1(x)\, r_1(z)\, w_2(x)\, r_3(z)\, w_1(y)\, r_4(x)$$

- 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.
- Tell whether S is accepted by the strong strict 2PL scheduler, explaining the answer.
- Tell whether S is accepted by the strict 2PL scheduler, explaining the answer.

Problem 3
Consider the following schedule S:

$$w_3(y)\, r_3(t)\, r_1(x)\, w_1(z)\, w_2(x)\, r_3(z)\, w_2(y)\, w_4(x)$$

and tell whether S is accepted by the timestamp-based scheduler.

Problem 4
A relation R has 100.000 pages, and our DBMS has 100 buffer frames free. Describe in detail the algorithm you would use to sort R under this condition, and tell which is the cost of the algorithm in terms of page accesses.

Problem 5
Consider the relations `Person(lastname, firstname, age, cityofbirth)` and `City(code, nation)`, which store information about persons and cities. The most frequent queries posed to such relations are:

- find all people with a given last name;
- produce the list of all persons (showing their last name, first name and age), sorted by ⟨last name, name, age⟩;
- for every person, list her/his last name and the nation of the city of birth.

We know that no two people have the same combination of last name, first name and age. We also know that relation `Person` has 1.000.000 raws, the size of every memory page is 1.2 KB, and the size of each field (relation attribute, or pointer) is 30 B.

Tell which method would you choose for representing the relations in secondary storage, taking into account that your goal is to execute the above queries efficiently. Also, for each query, tell which algorithm would you choose for executing the query, and how many page accesses would be needed for computing the answer to the query, given the chosen algorithm.