Data Management – exam of 26/10/2019

Problem 1
Let S be the following schedule:

$$r_1(v)\, r_1(y)\, w_2(y)\, w_3(v)\, w_3(y)\, r_3(x)\, r_2(x)\, r_1(x)\, r_1(z)\, w_2(z)\, r_1(t)\, w_2(t)$$

- Tell whether S is accepted by the 2PL scheduler with exclusive and shared locks. If the answer is yes, then show the schedule resulting from adding to S the corresponding lock and unlock commands. If the answer is no, then explain why, and tell if there is a single action whose removal from S makes the resulting schedule a 2PL schedule with exclusive and shared locks, explaining the answer in detail.
- Tell whether S is ACR or not, explaining the answer in detail.
- Tell whether S is recoverable or not, explaining the answer in detail.

Problem 2
Suppose that in a DBMS D every transaction T is associated with an integer $\sigma(T)$ such that no two different transactions have the same associated integer. Also, suppose that the scheduler of D is based on the concurrency control stategy defined as follows: when a transaction T_2 wants to read from or to write on a transaction T_1, if $\sigma(T_2) > \sigma(T_1)$, then the system allows T_2 to do it, otherwise the systems aborts T_2. Prove or disprove each of the following two statements.

1. Every schedule that is accepted by the scheduler of D is in the class of 2PL schedules.
2. Every schedule that is accepted by the scheduler of D is view serializable.

Problem 3
Given a binary relation R(A,B), and a unary relation S(C), the *non*-S-*portion* of R is the unary relation defined as follows:

$$\{\ a\ \in\ \mathsf{R[A]}\ \mid\ \forall b\ \langle a, b\rangle\ \in\ \mathsf{R}\ \rightarrow\ b\ \notin\ \mathsf{S[C]}\ \}$$

In other words, a value a is in the *non*-S-*portion* of R if a appears in the projection R[A] of R on attribute A, and every value b related to a by means of R does not appear in the only attribute C of S. Assuming that R is stored in a file sorted on $\langle \mathsf{A}, \mathsf{B}\rangle$ with 70.000 tuples in 7.000 pages, S is stored in a heap file with 9.000 tuples in 600 pages, and the buffer has 3 frames available,

1. describe an algorithm that, given relations R and S as specified above, computes the *non*-S-*portion* of R;
2. tell which is the cost of the algorithm written for item 1 in terms of number of page accesses.

Problem 4

Consider the relations R(A,B) with 115.000 pages, and the relation S(C,D,E) with 15.000 pages stored in a file sorted on C,D, and consider the query Q shown on the right. Assuming that 370 buffer frames are available, (*i*) illustrate the logical query plan associated to the query code, (*ii*) describe both the logical and the physical query plan you would select for evaluating the query, and (*iii*) tell which is the cost (in terms of number of page accesses) of executing the query according to the selected physical query plan.

Query Q:

```
select A
from R
where A > 10 and
      (A,B) not in (select C,D
                    from S
                    where E ≥ 0)
```

Problem 5
Consider the relation Bank(<u>code</u>,name,budget,nation) (with primary key code) stored in a heap with 1.350 pages, where each page contains 10 tuples, and the relation FinancialInstitute(<u>code</u>,name,budget,nation) (with primary key code) with 3.500 pages stored in a heap, where each page contains 10 tuples. Assuming that there is a hash index on FinancialInstitute with search key code, and the buffer has 120 free frames, describe at least two different algorithms for computing the set union between Bank and FinancialInstitute, and for each of them tell the cost in terms of number of page accesses.