Data Management – exam of 5/09/2019

Problem 1
Let S be the following schedule:

$$w_3(z)\, r_1(y)\, w_2(x)\, w_3(y)\, r_3(x)\, r_2(z)\, r_3(z)\, w_1(z)\, w_3(t)$$

- 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
Prove or disprove each of the following statements.

1. Every schedule that is not view serializable is not accepted by the timestamp-based scheduler.
2. Every schedule that is not view serializable is not strict.
3. Every schedule that is not view serializable is not rigorous.

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

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

In other words, a value a is in the 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 appears in the only attribute C of S. Assuming that R is stored in a file sorted on $\langle \mathsf{A}, \mathsf{B}\rangle$ with 50.000 tuples in 5.000 pages, S is stored in a heap file with 6.000 tuples in 400 pages, and the buffer has 3 frames available,

1. describe an algorithm that, given relations R and S as specified above, computes the 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,<u>B</u>) with 12.000 pages stored in a file sorted on the primary key B, S(C,D) with 130.000 pages stored in a heap file, and T(E,F,G) with 15.000 pages stored in a file sorted on E,F, 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
      B in (select C
            from S
            where (C,D) not in (select E,F
                                from T
                                where F ≥ 0)
```

Problem 5
Consider the relation R(<u>A</u>,B,C,D) (with primary key A) stored in a heap with 500 pages, where each page contains 10 tuples, and the relation S(<u>E</u>,F,G,H) (with primary key E) with 1.500 pages stored in a heap, where each page contains 10 tuples. Assuming that there is a hash index on S with search key E, and the buffer has 80 free frames, describe the possible algorithms for computing the difference between R and S, and for each of them tell the cost in terms of number of page accesses.