Data Management – exam of 13/07/2018

Problem 1 Let $\mathcal{T}_2$ be the class of schedules that are defined on two transactions, and that are accepted by the timestamp-based scheduler. Prove or disprove the following two statements:

1. $\mathcal{T}_2$ coincides with the class of conflict-serializable schedules defined on two transactions.
2. $\mathcal{T}_2$ coincides with the class of strict schedules defined on two transactions.

Problem 2 By referring to the following schedule S:

$$r_1(x)\, r_1(y)\, w_2(z)\, w_1(x)\, c_1\, w_2(x)\, c_2\, r_3(x)\, w_4(x)\, w_5(y)\, c_5\, w_3(y)\, c_3\, w_4(y)\, c_4$$

you are asked to:

2.1 Illustrate the various steps carried out by the timestamp-based scheduler when analyzing S, assuming that, initially, rts(α)=wts(α)=wts-c(α)=cb(α)=`true` for each element α of the database, and assuming that the timestamp of each transaction T_i is i.

2.2 Tell whether S is conflict serializable.

2.3 Tell whether S is in 2PL with shared and exclusive locks.

Problem 3 Suppose that there are 3 available frames in the buffer, and that relations R and S (possibly with duplicates) are stored in memory sorted according to all the attributes. Under these conditions, is it possible to compute the bag difference between R and S in one pass? If the answer is positive, then describe the algorithm you would choose. If the answer is negative, then explain the answer in detail.

Problem 4 Consider the relations `R(A,B,D,E)` and `S(C,F,G,H,L,M)`, and the query:

```
select R.A, R.B, S.C, S.F
from R, S
where R.B >= 50 and R.A = S.C
```

We know that `R` has 5.000.000 tuples, `S` has 150.000 tuples, each page contains 100 tuples of `S`, all values and pointers occupy the same space, and the number of available frames is 510.

4.1 Describe the logical query plan associated to the query code, and illustrate both the logical and the physical query plan you would select, motivating the choices.

4.2 Tell which is the cost (in terms of number of page accesses) of executing the query according to the selected physical query plan.

Problem 5 Consider the relations `R(A,B,C,D,E,F)` and `S(G,H,L,M)`, where `A,B` is the primary key of `R`, and `G` is the primary key of `S`. We know that `R` is stored in 400.000 pages, `S` is stored in 200.000 pages, both relations have a B^+-tree index with the primary key as search key, all values and pointers occupy the same space, and the number of available frames is 50. Consider the query:

```
select R.B, S.G
from R join S on R.A = S.G
where B >100
```

5.1 Describe the logical query plan associated to the query code, and illustrate both the logical and the physical query plan you would select, motivating the choices.

5.2 Tell which is the cost (in terms of number of page accesses) of executing the query according to the selected physical query plan.