

Data Management – AA 2015/16 – exam of 01/2/2016

Problem 1

Consider the following schedule $S = w_4(x) r_3(z) r_2(x) w_3(y) r_2(z) w_2(y) r_4(z) w_1(y)$, and answer the following questions:

- Tell whether there exists a serial schedule on the same transactions of S that is view equivalent to S but is not conflict equivalent to S .
- Tell whether S is conflict serializable, and, if so, exhibit one serial schedule which is conflict-equivalent to S .
- Tell whether S is in 2PL with shared and exclusive locks, motivating the answer.
- Tell whether S follows the strict 2PL protocol (with shared and exclusive locks), motivating the answer.

Problem 2

Prove or disprove the following claim: if S is a conflict serializable schedule on two transactions, then S is in 2PL with both shared and exclusive locks (or, in other words, S is accepted by the 2PL scheduler with both shared and exclusive locks).

Problem 3

We have a relation R , stored in a heap file, with 810.000 tuples and with the primary key constituted by the attributes A and B . Suppose that each attribute and each pointer occupies 10 bytes, and the size of each page in our system is 300 bytes. Consider the following query, where a and b are constants:

```
select * from R where A=a and B=b
```

and tell how many page accesses do we need to answer the query in these two cases: (i) if we use a sorted index on R with search key $\langle A, B \rangle$; (ii) if we use a B^+ -tree index on R with the same search key.

Problem 4

Let R be a relation stored in a heap file with 12.250 pages, consider the operation O of duplicate elimination on R , and answer the following questions about executing O in the two cases described below (as usual, the cost must be expressed in terms of number of page accesses).

- *Case (1): the buffer has $M = 13.000$ free frames.* Can we use the block nested loop algorithm? If yes, which is the cost? Can we use the one pass algorithm? If yes, which is the cost?
- *Case (2): the buffer has $M = 115$ free frames.* Can we use the block nested loop algorithm? If yes, which is the cost? Can we use the two pass algorithm? If yes, which is the cost?

Problem 5

Consider the relation $Club(\text{code}, \text{city})$, whose key is code , with 90.000 tuples, and the relation $Member(\text{clubCode}, \text{number}, \text{year}, \text{country}, \text{job})$, whose key is $(\text{clubCode}, \text{number})$, with 2.000.000 tuples. We know that the various clubs have approximately the same number of members, that every attribute and every pointer occupies 10 bytes, and that the size of each page is 1.800 bytes. The most important queries on such relations are the following:

Query 1	Query 2	Query 3
<pre>select * from Club order by code</pre>	<pre>select clubCode, number from Member where clubCode >=1 and clubCode <= 10</pre>	<pre>select number, clubCode, city from Club, Member where code = clubCode</pre>

Answer the following questions:

1. Which is the method you would choose for representing the two relations so as to make the execution of the above queries efficient?
2. Taking into account the chosen method, for each of the above queries, which is the algorithm you would choose for answering the query, and which is the cost of the algorithm in terms of number of page accesses?