Data Management, AA 2015/16 – exam of 05/04/2016

Problem 1
Let `R` be a relation with 10 attributes and 500.000 tuples. The size of each attribute is 100 bytes, and the size of each page in our system is 5000 bytes. Tell which is the minimum number of pages that the buffer must have in order to be able to sort `R` in two passes, and explain your answer.

Problem 2
Consider the following schedule $S = w_2(y)\, w_3(x)\, r_4(y)\, w_4(x)\, w_2(z)\, w_1(x)\, r_3(z)\, r_1(y)\, r_1(z)$, and answer the following questions (motivating each answer):

1. 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.
2. Tell whether there exists a serial schedule on the same transactions of S that is both view equivalent and conflict-equivalent to S.
3. Tell whether there exists a serial schedule on the same transactions of S that is conflict-equivalent to S, but is not view equivalent to S.
4. Tell whether S is in 2PL with shared and exclusive locks.

Problem 3
Prove or disprove the following claim: if S is a view serializable schedule on two transactions, then S is also conflict serializable.

Problem 4
We have a relation `R`, stored in a heap file, with 900.000 tuples and with the primary key constituted by the attributes `A`, `B` and `C`. Suppose that each attribute value and each pointer occupies 12 bytes, and the size of each page in our system is 480 bytes. Consider the following query, where `a`, `b` and `c` are constants:

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

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 <`A`,`B`,`C`>; (*ii*) if we use a B^+-tree index on `R` with the same search key.

Problem 5
Let `R` be a table stored in a heap file with 11.250 pages, and let `Q` be a table stored in a heap file with 15.000 pages. Consider the operation of computing the bag intersection between `R` and `Q`, and answer the following questions about executing such operation (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 one pass algorithm?
  - If yes, describe the algorithm in detail, and the tell which is its cost.
  - If we cannot use the one pass algorithm, can we use the two pass algorithm? If yes, describe the algorithm in detail, and the tell which is its cost. If not, explain the answer.
- *Case* (2)*: the buffer has* $M = 120$ *free frames.* Can we use the one pass algorithm?
  - If yes, describe the algorithm in detail, and the tell which is its cost.
  - If we cannot use the one pass algorithm, can we use the two pass algorithm? If yes, describe the algorithm in detail, and the tell which is its cost. If not, explain the answer.