Data Management – AA 2016/17 (06/07/2017)

Problem 1
We remind the reader that $r_i(E, z)$ indicates that transaction T_i reads the database element E from the database and assing its value to the local variable z, while $w_i(E, z)$ indicates that transaction T_i writes the value of the local variable z into the database element E. Consider the following schedule S

$$\begin{array}{l} r_1(A, v_1),\ v_1 := v_1 + 10,\ w_1(A, v_1),\ r_2(B, v_2),\ v_2 := v_2 - 20,\ w_2(B, v_2), \\ r_1(C, v_3),\ v_3 := v_3 * 2,\ w_1(C, v_3),\ r_2(D, v_4),\ v_4 := v_4 * 4,\ w_2(D, v_4) \\ r_1(B, v_5),\ v_5 := v_5 + 20,\ w_1(B, v_5),\ r_2(A, v_6),\ v_6 := v_6 - 10,\ w_2(A, v_6) \end{array}$$

and answer the following questions, with a detailed motivation for each answer:
(1.1) Tell whether S is serializable.
(1.2) Tell whether S is conflict serializable.
(1.3) Tell whether S is view serializable.

Problem 2
Given the following schedule S

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

(2.1) tell whether S is a 2PL schedule with both shared and exclusive locks, motivating the answer;
(2.2) tell what a timestamp-based scheduler would do when presented with the schedule S.

Problem 3
Consider the relation `REGION(name,nation,area,year)`, built one year ago with information about the regions in the various nations, when it had 24.576 tuples, distributed in 4.096 pages. Associated to the relation `REGION`, we also built a primary, unclustering ISAM index with search key ⟨`name,nation`⟩ using alternative 2. We know that in the last year 20 new tuples have been inserted in the relation, and no tuple has been deleted. Tell which is the cost (in terms of page accesses) of searching for the area and the year of a region, given its name and its nation. Explain in detail the answer.

Problem 4
Consider the relation `DRONE(type,company,number,cost)`, storing information about a set of drones, with 400.000 tuples. We assume that all values (pointers and attribute values) have the same length, the size of each page is such that it contains 160 values, and we have a tree index with search key ⟨`type,company`⟩ using alternative 1. Describe in detail which algorithm you would use to answer the query

```
select type,company,cost from DRONE where cost > 1000 order by type,company
```

and the cost of the execution of such algorithm in terms of number of page accesses.

Problem 5
Assume we have the relation `WORK(employeeNo,lastName,firstName,birthYear,company,salary)` with 1.000.000 tuples, where the size of each attribute value is 10 Bytes, the size of each page is 600 Bytes, and the number of different companies is no more that 900.

(5.1) Consider the query

```
select max(salary) from WORK group by company
```

computing, for each company, the maximum salary given to its employees, and suppose we have 350 free buffer frames available. Describe in detail the algorithm you would use to compute the answer to the query and the cost of the execution of such algorithm in terms of number of page accesses.

(5.2) Consider the query

```
select * from WORK order by employeeNo
```

and describe in detail the algorithm you would use to compute the answer to the query and the cost of the execution of such algorithm in terms of number of page accesses.