Data Management – exam of 11/9/2015

Problem 1 Consider the following schedule

$$S = r_1(V)\, w_3(X)\, w_3(Y)\, r_1(X)\, r_2(Y)\, w_2(X)\, w_3(Z)\, r_2(Z)\, r_2(V)\, r_3(V)\, w_4(V)\, w_4(Z)\, w_2(Z)$$

1. Tell whether S is accepted by the 2PL scheduler with exclusive and shared locks. If the answer is yes, then show the 2PL schedule obtained from S by adding suitable lock and unlock commands. If the answer is no, then explain the answer.
2. Tell whether S is view-serializable. If the answer is yes, illustrate a serial schedule which is view-equivalent to S. If the answer is no, then explain the answer.
3. Answer all the following questions, motivating the answers: (*i*) Is S recoverable? (*ii*) Is S ACR? (*iii*) Is S strict? (*iv*) Is S rigorous?

Problem 2
Let $\mathcal{C}$ be the class of all the schedules S that, when given as input to a timestamp-based scheduler, is such that the scheduler accepts the schedule S and, when processing S, uses the Thomas rule at least once. Prove or disprove that every schedule in the class $\mathcal{C}$ is view-serializable.

Problem 3
Define what a dense, clustering, sorted, primary index is. Describe a picture of an example of such index, and illustrate the algorithm that, given a value K, find the data record with the value K in the search key.

Problem 4
Consider the relation `Park(code,area,nation,director,year)` that stores information about 10.000.000 parks in the world, with the code of the park (which is a key of the relation), the area, the nation, the director and the opening year. We assume that the size of all fields in every record is of 20 bytes, independently of the field type, and the size of each page is 1000 bytes. The following queries are the most frequent and important ones among those accessing the relation `Park`:

1. `select * from Park order by code`
2. `select * from Park where code >= 30 and code <= 50`

Tell which representation method would you choose for the relation `Park`, and tell how many page accesses are needed for executing the two queries above, if the relation is represented according to the chosen method.

Problem 5
Suppose `R` is a relation that is not sorted.

1. Describe in detail the one-pass algorithm for eliminating the duplicates from `R`, and tell which is the condition under which such algorithm can be applied.
2. Describe in detail the two-pass algorithm for eliminating the duplicates from `R` using sorting, and tell which is the condition under which such algorithm can be applied.