Data Management – AA 2016/17 (13/06/2017)

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

$$r_1(A,t),\ t := t-50,\ w_1(A,t),\ r_2(B,s),\ s := s-10,\ w_2(B,s),$$
$$r_1(B,v),\ v := v+50,\ w_1(B,v)\ r_2(A,u),\ u := u+10,\ w_2(A,u)$$

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
Consider the following schedule S

$$r_1(X),\ r_3(X),\ w_2(X),\ r_3(y),\ r_1(y),\ r_4(y),\ w_3(z)$$

and tell whether S is a 2PL schedule with both shared and exclusive locks. Also, tell whether S is a strict 2PL schedule with both shared and exclusive locks. In both cases, motivate the answer.

Problem 3
Describe what a 2-level, primary, sparse, clustering sorted index is, and draw a picture of an example of such index. If relation `R` is stored in 250.000 pages (sorted according to its primary key), tell how many pages are needed to store a 2-level, primary, sparse, clustering sorted index for `R`, knowing that 100 data entries of the index fit in one page. Also, tell how many page accesses do we need to find the tuple in `R` with a given value for the key of `R`.

Problem 4
Consider the relation `BUILDING(code,nation,year,cost,currency)`, in which every record represents the data about a building, with code, nation where the building is located in, year of construction, and cost of construction together with the currency in which the cost is expressed. The relation has 4.000.000 tuples, stored in 400.000 pages, and has 240.000 different pairs of values in the attribute pair $\langle$`cost,currency`$\rangle$. We assume that all pointers and all attributes in every record have the same length, independently of the attribute, and that there is a dense, non-clustering B+-tree index on `BUILDING` with search key $\langle$`cost,currency`$\rangle$, using alternative 2. Consider the query that asks for `code`, `nation` and `year`, of all records of `BUILDING` with a given value for the pair $\langle$`cost,currency`$\rangle$, and tell how many page accesses we need for computing the answer to the query.

Problem 5
A relation `R` with key `A` has 3.000.000 pages, and our DBMS has 150 free buffer frames. Describe in detail the algorithm you would use to sort `R` with respect to `A` under this condition, and tell which is the cost of the algorithm in terms of page accesses.

Problem 6
Assume we have relation `S(A,B,C,D)` with 10.000 tuples, and relation `R(E,F,G)` with 300.000 tuples. Also, assume that every attribute needs 10 bytes, the size of each page is 500 bytes, and 410 buffer frames are free. We need to compute that bag difference between the projection of `S` on `A,B` and the projection of `R` on `E,F`. Describe in detail the algorithm you would use to compute the result, and the cost of the algorihtm in terms of page accesses.