Data Management (A.A. 2024/25) – exam B of 05/06/2025

Problem 1 Consider the relations `T(A,B)` and `V(A,F,G,H)`, where (*i*) both have `A` as key, (*ii*) `T` is stored in a heap with 120 pages (each page with 40 tuples) (*iii*) `V` is stored in a heap with 1.800 pages (each page with 20 tuples) with an associated hash index whose search key is `A` and (*iv*) the buffer has 62 frames available. If your goal is to compute the natural join (equi-join on `A`) between `T` and `V` as efficiently as possible in terms of number of page accesses, which algorithm would you choose among:

1.1 block-nested loop,
1.2 multi-pass based on sorting,
1.3 index-based.

Explain your answer in detail so as to convince that you choice is the right one.

Problem 2 Consider a scheduler D that behaves as follows when processing an input schedule S: D lets S proceed, dynamically building the precedence graph $P(S)$ by adding nodes and edges when needed, and never deleting nodes or edges and acting only whenever it processes the commit action of a transaction T_i. When processing such action, it executes the commit action if T_i is not involved in any cycle in $P(S)$, otherwise it aborts and rollbacks T_i. Let S be any complete schedule on transactions $T_1, \ldots, T_n$, where the last action of each T_i is the commit action c_i, let S' be the schedule produced in output by D when processing S, and let S'' be the schedule obtained from S' by ignoring the actions of the transactions aborted by D.

2.1 Prove or disprove the following two statements: (1.1.1) if $S = S''$, then S is view-serializable; (1.1.2) if S is view serializable, then $S = S''$.
2.2 Prove or disprove that S'' is ACR (Avoiding Cascading Rollback), and in case you disproved that S'' is ACR, tell how you would modify D in order to ensure that S'' is ACR.

Problem 3 Consider the following schedule S:

$$w_0(x)\ r_1(x)\ w_2(y)\ w_3(x)\ r_4(x)\ w_1(y)\ w_2(z)$$

3.1 Is S a 2PL schedule with both shared and exclusive locks? Motivate your answers in detail.
3.2 Is S a recoverable schedule? Motivate your answers in detail.
3.3 Describe the behavior of the timestamp-based scheduler when processing S, assuming that, initially, for each element α of the database, we have rts(α)=wts(α)=wts-c(α)=0, and cb(α)=`true`, and assuming that the subscript of each action denotes the timestamp of the transaction executing such action.

Problem 4 Consider the relation `CONSTRUCTION(code,type,region,cost,year)`, with 800.000 tuples stored in a sorted file with search key `code` (which is also the key of the relation), and with an associated sorted index with search key `region`. We know that no more than 200 constructions are allowed in the same region, that every attribute and pointer in our system occupies 10 Bytes, and that the size of each page in our system is 1.000 Bytes. Consider the following operations (1) given a region, compute the code of all constructions in that region, together with the corresponding type; (2) insert a new construction. For each of the two operations, tell which is the worst-case cost of its execution in terms of number of page accesses.

Problem 5 (only for students who opted for **option 1**, i.e., who do **not** do the project)
Let B be a relational database with relations `TaxiDriver(id,country)`, `Drives(driverid,taxi,since)`, `Taxi(tcode,type)`, `Own(ccode,tcode)`, `Company(ccode,budget)`, `Director(dcode,ccode,salary)`, where (*i*) each driver can drive many taxis (each one since a certain year) and each taxi can be driven by many drivers, (*ii*) each company can own several taxis and each taxi can be owned by several companies, (*ii*) each person can be the director of many companies (each one wirth a certain salary) and each company may have several directors.

5.1 Describe how you would organize a property graph database G in order to represent the relational database B. In particular, (*i*) specify how nodes, edges, labels, etc. of G are used in order to capture the information stored in the tables of B and (*ii*) choose a few tuples for the relations in B, and show the specific property graph database G obtained by applying the chosen representation method.
5.2 Describe how you would organize a document database D in order to represent the relational database B. In particular, (*i*) specify how collections, documents, etc. of D are used in order to capture the information stored in the tables of B and (*ii*) choose a few tuples for the relations in B, and show the specific D obtained by applying the chosen representation method.