CS 441: Midsemester Exam, Fall 1998

1. When would you

   (a) *pin* a record?

   (b) use static hashing (as opposed to extendable hashing)?

   (c) prefer a dense index to a sparse index?

   (d) construct two primary indices on the same relation for different search keys?

   (e) maintain $V(A, r)$ and $SC(A, r)$ for sets of attributes (instead of just for individual attributes)?

   (f) *not* adopt the "perform projections early" heuristic?

   Give examples with short explanations. ...18

2. One of your friends is hired as a consultant by a startup that wants to build YATS - Yet Another Transaction System. He suggests that they achieve durability simply by committing a transaction only after writing – to disk – all the changes done by the transaction. The company runs into performance problems. Why? What suggestions would you make to rectify the problem? Be as precise as possible. ...4

3. Suppose an attribute of a relation is set-valued. Suggest two alternative ways to store records belonging to this relation. State when you would prefer one over the other. ...4

4. What is the (a) maximum and (b) minimum number of keys in a B$^+$-tree with height h and a maximum fanout of k. ...4

5. Assume that records written only partially (say, due to a power failure) can be detected. Suggest an approach to achieve *atomic* block writes under block interleaved distributed parity RAID. ...5

6. Describe how to compute full outer join by using an extension of the hash join algorithm. ...5

7. Show the B^+ tree that results, with six pointers fitting in each node, when key values 2, 3, 5, 7, 11, 17, 19, 23, 29, 31 are inserted in that order.

   Show the tree after 23 is deleted. ...5

8. Give an example of a relational algebra expression and a query processing strategy where (a) MRU is preferable to LRU and (b) LRU is preferable to MRU. ...5

9. Give conditions under which ${}_A\mathcal{G}_B(E_1 \bowtie E_2)$ is equivalent to $E_2 \bowtie ({}_A\mathcal{G}_B(E_1))$. ...5

10. Given R=(A,B,C) and S=(D,E,F) and relations r(R) and s(S),

    (a) write an SQL expression for $\pi_{A,R}(\sigma_{(C=D)}(r \text{ X } s))$.

    (b) show an evaluation plan for the above expression, exploiting pipelining. ...5

11. When the cost of query processing according to a certain plan is computed, it is typically assumed that the needed data is on disk. However, quite often the data may be found in the database buffer.

    Suggest a practical method to estimate the cost that takes into account the presence of the database buffer.

    If it will help, give your answer with respect to a specific relational algebra operator. ...8

12. Is the natural left outer join operation associative? Justify.

    (Hint: Show that the following equivalence does not hold:

    $$(R \mathbin{⟕} S) \mathbin{⟕} T \equiv R \mathbin{⟕} (S \mathbin{⟕} T)$$

    Assume that the schemas of the three relations are $R(A,B), S(A,C)$, and $T(A,D)$, respectively.) ...8

13. (a) Suppose the view $v = r_1 \bowtie r_2$ has been materialized (computed and stored), and now a set of tuples i_1 is inserted into r_1, and i_2 into r_2. Assume you have the sets i_1 and i_2, as well as the relations r_1 and r_2 after the insertion. Give an expression to compute the new tuples to be inserted into the view result computed in order to keep the materialized view up to date.

    (b) Suppose the view $v = {}_A\,\mathcal{G}_{sum\ B}(r)$ has been materialized and then a single tuple t is added to r. Describe how to keep the view up to date.

    (c) Suppose now that the view is $v_1 = {}_A\,\mathcal{G}_{avg\ B}(r)$. What extra do you have to store to allow the view result to be updated efficiently?

    ...12

14. A bank offers its account holders 7% interest if their balance is above Rs. 10000, else 6%. To achieve this, each of the following SQL statements is executed as a separate transaction.

    update *account* set *balance* = *balance* * 1.6 where *balance* <= 10000
    update *account* set *balance* = *balance* * 1.7 where *balance* > 10000

    a) At the end of every month, the two transactions are invoked concurrently. Argue whether the bank achieves its desired goal. If it does not, how would you change things?

    b) What type of index structure would you recommend to execute these transactions efficiently? Why?

    c) How will your answer to (b) change if the interest rate is based not only on the balance but also on the number of accounts held by a customer? ...12

Total Marks = 100