CS 632: Advanced DBMS Mid-semester Exam, Spring 2003

Duration: 2 hours   Total Marks: 50

1. (Short Answers)
   (a) Explain how execution skew can arise when we use local sort followed by merge. Explain also how to avoid this situation. ...1+ 2
   (b) With shared nothing architectures, data must be replicated at more than one site to ensure the system can function in the event of site failure. Explain how replication can be done to minimize skew during failure situations. ...2
   (c) In a version of 2PC, called "early-prepare", as soon as a site completes the work given to it by the coordinator, it sends its "prepared" message along with the result produced. Under what circumstances is this a good idea, and why? ...3
   (d) Give an example of an aggregate function that would be hard to implement on streams. ...1
2. Suppose you had a cross-organization transaction that required messaging. This may require undo of committed transactions.
   (a) For the case of funds transfer from one account to another, suggest how long duration locks can be held till the transaction is completed at the other site, to ensure rollback can happen. ...2
   (b) Now consider a more complicated scenario, where the funds transfer is only the first step, and there is a product delivery step later that may require a completed funds transfer to be rolled back, i.e. funds returned.

       If you use long duration read/write locks at the funds receiving site, what would be the effect on concurrency at the receivers account? Suggest how logical locking can done to minimize the effect on concurrency. ...6
3. In the materialized view maintenance paper,
   (a) Why is the sharability (across queries) condition used for delta expressions (those with delta relations) but not for full expressions. ...2
   (b) How are delta relations handled when being considered for materialization? List a major benefit and a possible drawback of this choice? ...3
4. Consider the paper on computing capabilities of mediators. In the capabilities $c[S]$ the site guarantees that if you give a value from S it will give the correct, complete answer, if you give a value not from S it may not give you a correct/complete answer. Suppose we create an adornment $f[S]$ which is like $c[S]$, but for values outside S, the answer is guaranteed to be empty.

   Give the (new parts of) the tables for union, and join with passing of bindings, to deal with the new adornment. ...7
5. (Continuous queries)
   (a) Assume that expressions are on one or both of attributes A and B, and may involve conjunction of conditions. Explain clearly how to use bitmap indices to index expressions. ...6
   (b) List a few benefits/drawbacks of using R-trees vs. bitmap indices for the above problem. ...2
   (c) What does the NiagraCQ paper do w.r.t. indexing of expressions, and list one significant drawback of the scheme? ...2

6. (XML publication) Consider publishing of relational data as XML.

   (a) Are outerjoin plans relationally equivalent to outerunion plans? If not in what sense are they equivalent? ...2

   (b) Using standard cost models which would be cheaper: $r ⟕ s$ or $r \cup r \bowtie s$, where $\cup$ here is outerunion? Explain why? ...2

   (c) Explain how you can use shared reads to make the costs above comparable. ...2

   (d) Would the greedy-integrated algorithm of Dalvi et al. using pessimistic MQO handle the above case? And with optimistic MQO? Explain why/not and if not explain how optimistic MQO can be modified. ...5

Total Marks = 50