University of Wisconsin-Madison
Computer Sciences Department

Database Qualifying Exam
Fall 2000

Answer any four (4) of the following five (5) questions. Before beginning to answer a question make sure that you read it carefully. If you are confused about what the question means, state any assumptions that you have made in formulating your answer. Good luck!

1. **Query Containment.**

Describe the tableau construct proposed by Aho at al., and explain how it can be used to decide containments and equivalences among relational expressions. Use examples.

2. **ARIES.**

During the REDO pass, when we are deciding whether to redo the operation in a log record l for a page p, the claim is that we don't even have to check the pageLSN for p if p is not in the Dirty Page Table or if the page's recoveryLSN in the DPT is greater than the LSN of l.

a)   Why is this true?
b)   How is it even possible that we are considering an update log record for a page p but p is not in the DPT?
c)   Consider the three points in the log:
i)   beginning of analysis phase,
ii)   beginning of the REDO phase,
iii)   earliest log record considered by the UNDO phase.
Is it possible that they could appear, from earliest to latest, in the order (i), (ii), (iii)? Explain your answer.

3. **Are DBMS Really Necessary?**

Except for certain types of business data, very few people use database systems on a daily basis. Even scientists whose experiments produce vast quantities of data avoid the use of database systems for storing their experimental data, relying instead on standard OS files for the task.   This trend seems to actually be getting worse.

What database system factors do you believe most limit the use of database systems outside the commercial domain?

Do you believe that the availability of object-relational database system technology will significantly increase the use of database technology? Why or why not?

What problems were the designers of object-oriented database systems trying to solve? Would the widespread adoption of such systems 10 years ago have changed the situation we are in today? Why or why not?

What changes to how database system are designed, implemented, and used would have the most impact on expanding the base of database system users.

4. **Join Algorithms**

In Shapiro's paper on join processing algorithms, he does not distinguish between sequential and random I/O. (Here by "sequential I/O" we mean that consecutive reads/writes go to adjacent disk pages, whereas by "random I/O" we mean consecutive reads/writes that likely go to different parts of the disk.) Assume for simplicity that the unit of reading/writing is a single page (that is, reading or writing a run of k pages will take k I/Os.)

a) For the hybrid hash and sort-merge join algorithms give approximate formulas for the number of sequential and random I/Os for each algorithm. Break down this cost by each phase of the algorithms, assuming that you have only one disk available. (Assume that the join is of relations R and S, and that both R and S are smaller than M^2, where M is the number of pages of memory.)

b) Now assume that you have two disks. How would you use them in the join algorithms, and what are the new formulas for random and sequential I/Os?

5. **ObjectStore**

Describe how the ObjectStore system uses virtual memory techniques to handle dereferencing of pointers to objects that may be in memory or on disk. In your description, include a discussion of pointer relocation (when it is necessary, and what data structures are used in its implementation.)