SPRING 1993
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN—MADISON
PH. D. QUALIFYING EXAMINATION

Database
Breadth Examination

Monday, February 8, 1993
3:00 – 5:00 PM
113 Psychology

GENERAL INSTRUCTIONS:

1. Answer each question in a separate book.
2. Indicate on the cover of *each* book the area of the exam, your code number, and the question answered in that book. On *one* of your books list the numbers of *all* the questions answered. *Do not write your name on any answer book.*
3. Return all answer books in the folder provided. Additional answer books are available if needed.

SPECIFIC INSTRUCTIONS:

Answer any *four* of the following *five* questions. Before beginning to answer a question, be sure to read it carefully. If you are confused about what the question means, state any assumptions you have made in formulating your answer. Good Luck!

POLICY ON MISPRINTS AND AMBIGUITIES:

The Exam Committee tires to proofread the exam as carefully as possible. Nevertheless, the exam sometimes contains misprints and ambiguities. If you are convinced a problem has been stated incorrectly, mention this to the proctor. If necessary, the proctor can contact a representative of the area to resolve problems during the *first hour* of the exam. In any case, you should indicate your interpretation of the problem in your written answer. Your interpretation should be such that the problem is nontrivial.

UNIVERSITY OF WISCONSIN-MADISON
COMPUTER SCIENCES DEPARTMENT

DATABASE QUALIFYING EXAMS
SPRING 1993

BREADTH EXAM

Answer any *four* (4) of the following *five* (5) questions. You may *not* select questions from the Depth Exam. Before beginning to answer a question, be sure to read it carefully. If you are confused about what the question means, state any assumptions you have made in formulating your answer. Good Luck!

1. Query Languages

Consider the following schema used by "The Generic Company" to keep track of various information about projects, departments, and employees:

PROJECT (*prno*, prname, fundingagency, cost, dno)
DEPT (*dno*, dname, mgr-id, budget)
EMPLOYEE (*eid*, ename, phone)
WORKSON(*eid, prno*, hoursperweek)

In the above, primary keys are in italics. Write SQL statements to do the following tasks. (In addition to the standard parts of the language, assume an SQL version that supports the keywords **union**, **minus**, **intersect**, **contains**, **not contains**, **exists**, and **not exists**.)

(a) Find the names of all projects managed by Gunga Din or funded by DGFA at a cost of over five million dollars. (DGFA is De Generic Funding Agency.)

(b) Find the names of all projects in which at least 10 employees work for more than 20 hrs a week.

(c) Find the names of all managers whose departments only have DGFA funded projects. (i.e., if the department has a project, the project is funded by DGFA.)

(d) Find the names of all employees who work on *every* DGFA funded project.

(e) Fire (i.e., remove) all employees who no longer work on any project.

2. Query Processing

The Sybase database system uses only the nested-loops join algorithm. In the event that an index does not exist on the join attribute of the inner relation, one is constructed "on the fly". They claim that this strategy offers equivalent performance to the sort-merge join algorithm.

(a) Are they correct? Explain by describing the I/O behavior of the nested-loops and sort-merge join algorithms. Assume that the index has to be created dynamically. Assume, further, that neither relation fits in memory.

(b) Under what conditions will the two algorithms have essentially the same I/O behavior — and hence equivalent performance?

3. Integrity Constraints

Consider the important concept of referential integrity in relational databases.

(a) Define the terms ‘key’ and ‘foreign key’. Give two examples, one that involves two relations and one that involves only a single relation.

(b) What events require referential integrity maintenance actions to be taken? For each one, what set of useful alternative actions should be provided for database designers to choose among?

(c) Should referential integrity constraints be enforced on a per-event basis, or should enforcement be deferred until end-of-transaction? Briefly discuss the issues and make a recommendation.

4. Database Design

Dean Martin of The Polytechnic Institute for Thespians (affectionately known as The PITS) has decided to put some of his data in a relational database. He has to keep track of the hours (H) and rooms (R) where courses (C) are taught, the students (S) who are enrolled in each course, and student grades (G) in each enrolled course. The following constraints are given:

- only one course can meet in a room at one time,
- each student has just one grade in each course,
- a student can only be in one room at one time.

Further, the dean is frequently interested in queries of the following form:

- What was John Smith’s grade in Comparative Histrionics?
- What are the courses that Jane Smythe is enrolled in?

(a) List the functional dependencies.

(b) Suggest a ‘good’ BCNF schema.

(c) If you use the BCNF schema suggested in part (b), can you efficiently check that the FDs given to you are maintained when the database is updated? Explain briefly.

(d) What are the conditions under which a 3NF schema is preferable to a BCNF schema?

(e) In this example, is there a 3NF (but not BCNF) schema that is better than the BCNF schema that you suggested in part (b)?

5. Update Problems

Consider the following database schema, where primary keys in italics:

EMP (*name*, sal, dname)

DEPT(*dname*, floor, mgr)

In what follows we use a hypothetical, SQL-like, syntax for updates. Consider the following update:

```
update EMP
set sal = (sal * 1.2)                                   (U1)
where sal < 100K and sal > 50K
```

(a) Assume that a secondary index exists on the sal field. What problem may occur if this update is not handled properly? (Keep in mind that updates also change the index.)

(b) Describe an algorithm that can be used to avoid the problem. Comment on the positive and negative aspects of your algorithm.

Consider now the following two updates:

```
update EMP
set sal =
        select f.sal                                    (U2)
        from EMP f, DEPT d
        where dname = d.dname and d.mgr = f.name

update EMP
set sal =
        select f.sal                                    (U3)
        from EMP f, DEPT d
        where name = d.mgr and d.dname = f.dname
```

(c) One of the above two updates presents a problem when executed. Identify the problem and the update. Explain the reason why the other update does not share the problem.

(d) Based on your explanation above, suggest how the problem may be detected without executing the problematic query.