1996 Systems Qualifying Exam – Database Area
Three pages, five problems, 30 points total

1. **(10 points)**

Suppose we have the following three entity sets:

Faculty	with attributes Name, Address, and Dept
TAs	with attributes Name and Address
Courses	with attributes Title and Dept

We also have a ternary relationship STAFF among all three attributes. The intent is that STAFF consists of those triples (f, t, c) such that c is a course, f is a faculty member teaching the course, and t is a TA assigned to the course. In general, there are no constraints about these triples: faculty can teach more than one course, TA's can be assigned to more than one course, courses can have several faculty and TA's, and not every TA assigned to a course is associated with every faculty member assigned to that course. For example, course c might have two sections with different faculty and TA's. However, one TA might be assigned to both sections and others assigned to only one. To put it more succinctly but technically: any set of (f, t, c) triples whatsoever could be the value of the relationship STAFF at some time.

(a) Draw an Entity/Relationship diagram for the entity sets and relationship described above.

(b) Sometimes, when we put some constraints on how faculty, TA's and students interact, we can replace the ternary relationship STAFF by two or more binary relationships, and yet preserve the information contained in the ternary relationship. Suppose in particular that for every course there is a unique instructor. Give an E/R diagram representing the faculty-TA-course information, using two binary relationships rather than a single ternary relationship. Be sure to indicate, with arrows on the edges, when a relationship is many-one.

(c) Suppose that courses can have more than one faculty as in the original problem, but when a TA is assigned to a course he/she is assigned to a unique faculty member teaching that course. No other constraints hold; e.g., faculty or TA's can be assigned to more than one course, courses can have multiple TA's or faculty, etc. Someone claims that in this case it is possible to represent the STAFF relationship by two binary relationships, one between Courses and TAs and the other between TAs and Faculty. Show that is not the case by giving specific values for the entity sets and the relationship STAFF so that the two binary relationships do not allow us to determine the true STAFF relationship (i.e., show there are two possible values for the STAFF relationship that are consistent with the same pair of values for the binary relationships).

2. (**6 points**)

Consider the following relational database schema:

```
Student(ID, name, dept, status)  // status = 'grad' or 'undergrad'
RA(ID, advisor)
TA(ID, course)
```

and the following query, expressed in English:

> *Find the name and department of all graduate students who are neither an RA nor a TA.*

Express this query in:

(a) SQL

(b) Relational algebra (excluding the division operator)

3. (**4 points**)

There are two well-known types of indexing methods for relational database systems: the *B-tree* family of indexing methods, and the *hash-based* family of indexing methods. To be concrete, consider specifically B+-tree indexes and indexes based on extendible hashing.

(a) In 1–2 sentences, state a significant advantage of a B+-tree index over an extendible hashing index for an attribute A of a relation R.

(b) In 1–2 sentences, state a significant advantage of an extendible hashing index over a B+-tree index for an attribute A of a relation R.

4. (**4 points**)

(a) Consider a schedule S_1 with the following arcs in its dependency graph: $T_3 \rightarrow T_1$, $T_3 \rightarrow T_4$, $T_1 \rightarrow T_2$, $T_1 \rightarrow T_4$, $T_4 \rightarrow T_2$. Is S_1 serializable? If it is, what schedule is S_1 equivalent to?

(b) Consider a schedule S_2 with the following arcs in its dependency graph: $T_3 \rightarrow T_1$, $T_3 \rightarrow T_4$, $T_1 \rightarrow T_2$, $T_2 \rightarrow T_3$, $T_4 \rightarrow T_2$. Is S_2 serializable? If it is, what schedule is S_2 equivalent to?//

5. **(6 points)**

Consider a program P that in executing a transaction T executes the following actions, at the indicated times:

Time 1:	P issues "write object x"
Time 2:	Log entry for x is written into log tail
Time 3:	Write action completes, control returns to P
Time 4:	P issues "read object y"
Time 5:	Read action completes, control returns to P
Time 6:	P issues "end of transaction T"
Time 7:	Log entry for commit action written to log tail
Time 8:	Commit record is flushed to disk
Time 9:	Control returns to P

Assume that the database system uses a two-phase locking strategy and that it avoids cascading rollbacks. Also assume that the database system (not the user transaction) issues all lock and unlock requests. For each of the following questions, give the *largest* range of times where the specified action can take place, given the information you have.

(a) In what range of times can the lock for x be obtained for T?

(b) In what range of times can the lock for y be obtained for T?

(c) In what range of times can the lock for x be released?

(d) In what range of times can the lock for y be released?