

Fall, 2003 CIS 550

## Database and Information Systems

### Final Examination

Take-home, 10 questions, 100 pts.  
Each question is worth 10 pts.

Your name: \_\_\_\_\_

1. Of the various join algorithms discussed in this course, which one **MUST** be implemented by *every* DBMS that hopes to support the fully generality of SQL or XQuery? Why?
2. What is a serializable scheduling? Why is it important and why does SQL support multiple isolation levels?

- What are the trade-offs that one must consider in choosing between **local-as-view** and **global-as-view** mediation? Which is generally more expressive? Using what you know from our discussion of containment, view unfolding, and query rewriting (answering queries using views), argue for what subset of SQL (instead of Datalog) should be possible to implement in (a) local-as-view and (b) global-as-view mappings.
- Why is dynamic programming (a) applicable to query optimization, and (b) beneficial when “searching the space of possible plans”?

5. Compare SQL, XQuery, and XSLT: what basic features are in common? How are the languages different (at a high level)?
6. What are the qualities of (relational and XML) databases that distinguish this field from other related areas (e.g., web search, using Java serialization, writing COBOL programs, hyperlinks)?

7. Explain some of the major differences between XML and the relational model as ways of ***representing*** structured data. Relate your discussion to issues of normalization, as well as entity sets and relationship sets.

8. Suppose you are given the following information about two tables:

$R(x,y,z,w)$	has 1,000 tuples and occupies 10 pages
$S(a,b,c)$	has 5,000 tuples and occupies 8 pages

- There is an index over R on attribute x that occupies a single page.
- We only have 2 pages in our buffer pool, and we output data to the user (not to disk).

What is the cost, **in terms of the number of disk I/Os**, of performing the ***most efficient nested loops join*** with outer relation S and inner relation R?

What is the cost, in terms of disk I/Os, of performing the ***most efficient nested loops join*** with outer relation R and inner relation S?

9. Given the query:

```
SELECT x, y
FROM R, S
WHERE R.x = S.x AND z = 10 AND w < z
```

over schemas  $R(x,y,w)$  and  $S(x,z)$

Write a relational algebra expression for this query. Assume that the join only is done on the  $R.x = S.x$  equality predicate, and all other predicates must be applied using selection operators. **Push selections and projections to the lowest possible levels, as a query optimizer would do.** (You do not need to write a full query plan with physical algorithms.)

10. Translate the Datalog query:

```
q(x,y) :- r1(x,z), r2(z,w), r3(z,w)
q(x,y) :- r4(x,y,z)
```

to SQL.