

# Notes on *A Relational Model of Data...*

COS 480/580: Database Management Systems

Sudarshan S. Chawathe  
University of Maine

Fall 2005

These notes, cunningly disguised as questions, relate to E. F. Codd's classic paper: *A relational model of data for large shared data banks* (*Communications of the ACM*, 13(6):377–387, June 1970). Some questions are very simple, requiring only a basic understanding of the terms introduced by the paper. Others are more difficult, requiring a deeper understanding of both the paper and current database systems. Some questions require familiarity with material that we have not yet covered in class; these are marked with ◦ and may be skipped for now.

1. The second sentence in Section 1.2.2 claims “[An index] tends to improve response to queries and updates and, at the same time, slow down response to insertions and deletions.” Is it possible for an index to slow down the response to a query? Justify your answer.
2. ◦ Provide XML DTDs corresponding to each of the five structures depicted on page 378. (See Section 4.7 of the textbook.)
3. ◦ For each of the five XML representations in your answer to Question 2, provide XPath or XQuery versions of the query described in the first sentence on page 379.
4. Briefly differentiate between the terms relation and relationship, as used in the paper.
5. Provide a relational-algebra expression for the *active domain* of a database that consists of a single relation  $R(A, B, C)$ , with all attributes of integer type.
6. Provide an E-R diagram that best captures the model outlined in Figure 3(a).
7. Map the E-R model in your answer to Question 6 to a relational schema by using the procedure described in the textbook. Highlight and explain any differences between the resulting schema and the schema of Figure 3(b).
8. Is it possible to define an E-R model that does not satisfy one or both of the two conditions at the top of the right column of page 381? Justify your answer and reconcile it with the claim made below those conditions in the paper.
9. Compare the general name for a data item as described in the paper with the fully qualified names used in a current database system, such as PostgreSQL.
10. Argue for or against the following statement from the first paragraph of Section 1.5: A first-order predicate calculus suffices if the collection of relations is in normal form.
11. Indicate how the actions described in the fifth paragraph of Section 1.5 may be performed in standard SQL.
12. Does UMaine's *WebDSIS* system permit *symmetric exploitation* (page 382)? Justify your answer.
13. Prove or disprove: The number of directed paths needed to support symmetric exploitation of an  $n$ -ary relation is  $n!$ .
14. What are the analogs of the *named* and *stored* sets of relations (Section 1.6)?
15. Justify the claim made by Footnote 6 (page 382). Provide examples in SQL to support your answer.

16. The definition of projection in Section 2.1.2 includes the restriction  $n \geq k$ . Comment on its significance.
17. Compare the definition of a join used in the paper with the definition in the textbook.
18. Exhibit two relations that are not joinable although they have a common domain.
19. Provide a method to test whether two relations are joinable.
20. Prove or disprove the claim made near the bottom of the left column on page 384 (rephrased): If binary relations  $R$ ,  $S$ , and  $T$  possess points of ambiguity  $x$ ,  $y$ , and  $z$  for joining the pairs  $(R, S)$ ,  $(S, T)$ , and  $(T, R)$ , respectively, such that  $xSy$ ,  $yTz$ , and  $zRx$ , then there is a plurality of cyclic 3-joins of  $R$ ,  $S$ , and  $T$ . (We use the notation  $aXb$  to mean  $(a, b) \in X$ .)
21. Is the *tie* operator  $\gamma$ , defined on page 384, expressible using the six basic relational-algebra operators discussed in class? Justify your answer.
22. Exhibit two distinct joins of the relations  $R$  and  $S$  of Figure 12 (page 385).
23. Prove the claim made in the caption of Figure 12.
24. Provide two relations that have multiple compositions. Exhibit two distinct compositions.
25. What is a *connection trap*? Explain the term in your own words and provide a concrete example from a domain other than the one in the paper.
26. How may the operation of restriction (Section 2.1.5) be expressed succinctly in current relational algebra?
27. What are modern analogs of the operator sets  $\theta_1$  and  $\theta_2$  of Section 2.2?
28. Explain how current relational design theory addresses the the redundancy in the schema of the *employee* relation described in Section 2.2.1.
29. Explain how current relational design theory addresses the redundancy in the schema outlined in the second example of Section 2.2.1. Repeat for the weak redundancy outlined in Section 2.2.2.
30. Comment on the validity of the ideas in the last paragraph of Section 2.2.1 in the context of current database management systems.
31. Explain how a database system may induce redundancies as suggested by the first paragraph of Section 2.3.
32. ◦ Write the simplest standard-SQL statement that expresses the constraint described in the second paragraph of Section 2.3. (See Chapter 7 of the textbook.)
33. Comment on the expressive power of an algebraic query language consisting of only the operators described in this paper.