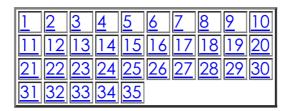# CS145 Final Exam Solutions

Index

Problem 1: (c)

Suppose element x appears r, s, and t times in R, S, and T, respectively. Then Q1 produces $\max(r+s-t,0)$ x's, while Q2 produces $\max(r+s-2t,0)$ x's, which cannot be more and could be fewer.

Problem 2: (b)

Outer joins include the join and sometimes more.

Problem 3: (b)

Remember that UNION removes duplicates.

Problem 4: (a)

Each query produces each a from an (a,b) tuple such that b appears anywhere in the a-column.

Problem 5: (a)

Both produce one copy of every (a,b) pair that appears in one or more tuples of R.

Problem 6: (a)

Sorting tuples doesn't affect the bag of tuples that they represent.

Problem 7: (d)

Q1 can produce a tuple that is not in R. For example, if R = {(1,2), (2,1)}, Q1 produces (1,1). On the other hand, Q2 can surely produce several tuples, while Q1 never produces more than one.

Problem 8: (c)

Q2 is always contained in Q1 because whatever tuple (a,b,c) from R produces a (b,c) pair for Q2 can always be chosen as both the tuples (a,b,c) and (a,b1,c1) in Q2.

Problem 9: (b)

If x is produced by Q1, then there are more tuples in R that begin with x than there are in S. Thus, one of these tuples will survive in R-S and produce x in the answer to Q2.

Problem 10: (d)

If R has nulls, COUNT(a) can be different from COUNT(*). Note that "contained in" is a set or bag concept, and has nothing to do with whether one number is less than another.

Problem 11: (d)

The assertion says you can't have tuples from R and S that agree on a but have different b (resp. c) values.

Problem 12: (d)

Many people picked (c), but this trigger does not handle the case of an update to R.b or any deletion.

Problem 13: (a)

Simulate it, but don't forget that as R gets more tuples, not only does the tuple that you insert get doubled, but any other tuple with the same value also doubles. By the time (8) is inserted, the tuples average exactly 10, so the trigger is not executed to insert (9).

Problem 14: (c)

The keys are ABD and ACD. There are six sets that are supersets of one or both of these, including ABD and ACD themselves.

Problem 15: (a)

T2 must appear to run either before T1 (in which case 1 is produced) or after (in which case 5 is produced).

Problem 16: (c)

Here, T2 can execute at any time during the execution of T1, so it can produce not only 1 or 5, but 6 (if T2 executes just before the DELETE) or 3 (if it executes after the first INSERT).

Problem 17: (b)

Dave retains the full privilege because of the route Alice -> Carol -> Bob -*gt; Dave.

Problem 18: (d)

X=1, Y=2, Z=10 yields TRUE; X=2, Y=1, Z=3 yields FALSE, and X=Y=Z=NULL yields UNKNOWN.

Problem 19: (a)

The keys are AD and BD. B->C is a 3NF violation, because B is not a superkey, and C is not prime.

Problem 20: (c)

The only key is AB. Thus, each of the given FD's has a superkey on the left. But the MVD has a non-superkey on the left, so it violates 4NF.

Problem 21: (c)

The first and third books have isbn's; the second doesn't. So 4 AUTHOR nodes are produced: the first Stevens, plus Abiteboul, Buneman, and Suciu.

Problem 22: (c)

The where-clause forces $b to refer only to the book "Data on the Web." But each of its three authors is returned.

Problem 23: (a)

(b) is violated by the second book, (c) by the second and third, and (d) by the third.

Problem 24: (d)

Pretty much everyone was able to simulate these four sequences correctly.

Problem 25: (b)

C stands for "Consistency."

Problem 26: (c)

Remove all elements with ? and *, and replace a + by a single occurrence. You then get one a with b and d children. The b has c and d children, and the c has a d child. The total is one a, one b, one c, and three d's.

Problem 27: (c)

The marginals are those cells of the cube that have at least one "don't care" coordinate. The easiest way to count them is to imagine a * value in each dimension, and then ask how many cells have the * in at least one dimension. The size of the outer cube is $(8+1)(6+1)(10+1) = 693$, and the size of the inner cube (those cells with no *) is $8*6*10 = 480$. The difference, the number of marginal cells, is 693-480 = 213.

Problem 28: (b)

Sales is clearly the "fact table."

Problem 29: (d)

The entity has to be at the root. Each of the two 3-node subtrees of the root can be in any of five "states": (1) entity not present in any (2) entity at root only (3) entity at root and left child (4) entity at root and right child (5) entity at all three. Thus, there are 5*5 = 25 possible ways to place the entity at a legal subset of the subclasses.

Problem 30: (d)

When we convert to relations, a supporting relationship does not need a relation unless it has attributes of its own.

Problem 31: (a)

The relation for A needs the full key of A, which is a and b, plus other attributes of entity set A --- that is, d. Note: this relation may later be combined with the relation for R, but that is not what the question asked.

Problem 32: (c)

Many people picked (d), but (d) doesn't do anything. After executing the loop, you are no further along toward the goal stated in the problem than you were before. Choice (c) at least makes some progress, in that you can next test whether x is TRUE, and if so, access the first integer in R by a statement like (a).

Problem 33: (d)

This was a reward for people who read the book, or at least had skimmed it enough to know that Section 8.4.4, which we did not cover in class, existed and had an example that shows what you have to do.

Problem 34: (b)

Cursors and Statement Handles are used as ways to access tuples of a query result. The JDBC equivalent is the ResultSet Class.

Problem 35: (a)

There is only one column of R, which has T2 objects as its values. If you doubt this point, think of using the Oracle system to insert into R using INSERT...VALUES(...). How many commas would there be at the top level inside the VALUES(...). Answer: 1. Other commas would be nested inside a type constructor, but that doesn't imply the existence of more than one component, any more than an insertion like `VALUES('Hello, World')` would.