Assignment – 1
CS540 – Database Management Systems

Arash Termehchy

Question-1
Briefly describe the concepts of order independence and index independence and explain why relational queries are order and index independent.

Answer-
Order Independence- The Application programs running on databases should not get affected when the order in which data is stored is altered.

Index Independence- When we query on database, The application should display results correctly irrespective of what index we used and whether indexing is done or not.

Relational Queries are Order Independent and Index Independent because:

1. If we query on relational databases, we can get results in any order we want by specifying in the query (Asc/Desc) and while inserting the values, we are not inserting them in sequential order. So, Irrespective of what order we stored the results, we get same results by specifying in the query.
2. In Relational queries Indexes only affect the Query time but not the output. When querying on relational database, we are not specifying which particular index to use. If we run same query on same table with different indexes, they give same output.

Question-2
Briefly explain a disadvantage of relational model.

Answer-
The Main Disadvantages of Relational Databases are:

1. Require lot of storage space because decomposing into multiple tables.
2. We are decomposing the tables and while querying it we are joining all tables again. So, It's complex.
3. Not feasible for storing heterogeneous multimedia data.

Question-3

Consider schema S1 that contains relation R(A, B, C, D, E) and schema S2 that contains relations P1(A, B, C, D) and P2(A, B, E). Explain whether or not S2 includes S1. If it is not, explain which integrity constraints should be added to S2 and/or S1 so S2 includes S1. Also explain which integrity constraints we should add to S2 and/or S1 so they become equivalent. You should explain your answers using the concepts of schema inclusion and schema equivalence proposed in 'P. Atzeni, et. al., Inclusion and Equivalence Between Relational Database Schemata, TCS, 1982'.
You do not need to use a lot of mathematical notations in your answers. It is sufficient to clearly explain the reasons.

Answer-

S2 is included in S1 If we have a FD A,B→E and A,B→C,D in S2, which means A and B both together should uniquely identify other attributes. Now we can get S2 by applying projection to S1 and get back S2 by joining P1 and P2. Here S1 is included in S2 and S2 is included in S1.So S1 and S2 are equivalent.

Question-4

Provide an example of horizontal decomposition where the source schema (original scheme) is not included in the refined (target) schema and explain why. You do not need to use a lot of mathematical notations in your answer. It is sufficient to clearly explain the reason.

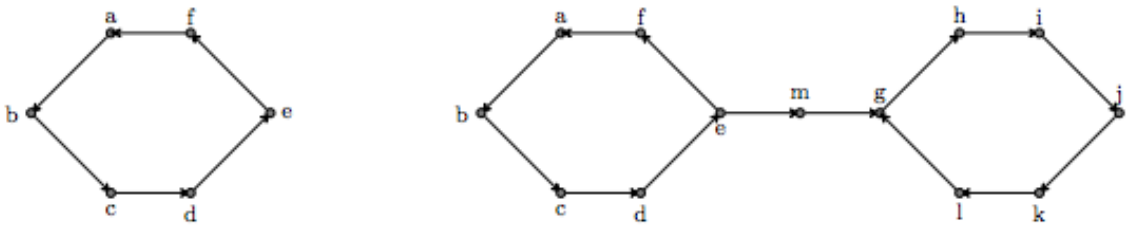
Answer-

For example, consider two schemas S1 (emp. No., emp name, dept.) which is horizontally decomposed into S2(emp no., emp name, dept) where dept = 'test' or dept = 'dev'. Here S2 contains only employee details who are in either testing or development. In this case, S2 is included in S1 because S2 is subset of S1 but not vice versa. That means if we apply a Restriction on Dept. for S1 we can get S2. But we can't get back to S1 from S2.

We can get back from S2 to S1 by applying join if we have another relation with rest of the departments.

Question-5

Assume the following graphs depict parts of the Web, where nodes represent pages and edges show hyper-links. Find out the pages whose PageRank values are greater than zero and their relative PageRank values in both graphs. You do not need to perform the fix point computation to determine the PageRank values. Instead, you should guess the PageRank values based on your understanding of the PageRank algorithm and explain why you think they are correct. If it is no possible to make any Educated guess for some page(s), you should explain why.



(a) Graph 1 (b) Graph 2

Figure 1: Graphs for Problem 5

Answer-

In Graph 1, All pages will have same page rank of 1/6 because the edges in the graph are pointed cyclic forming a loop called as rank sink.
Each node is pointed exactly once by other node. i.e. each page has one back link pointing towards it. So, All pages will have equal page rank.

In Graph 2, Node G has higher page rank because the edges from nodes L and M are pointing to G which means G has many back links pointing towards it.

Nodes M and F will have less page rank because the page rank of E is shared between M and F.

The Other Nodes A, B, C, D, E will have same page ranks.

The Nodes H, I, J, K, L also have same page ranks. But If we consider the second case of page rank, which says the page which is pointed by a popular page will also have high page rank Then H can have high page rank because the popular page (page with High page rank) G is pointing to it.

Submitted By:
Lakshman Madhav Kollipara
Student ID. - 932 655 504