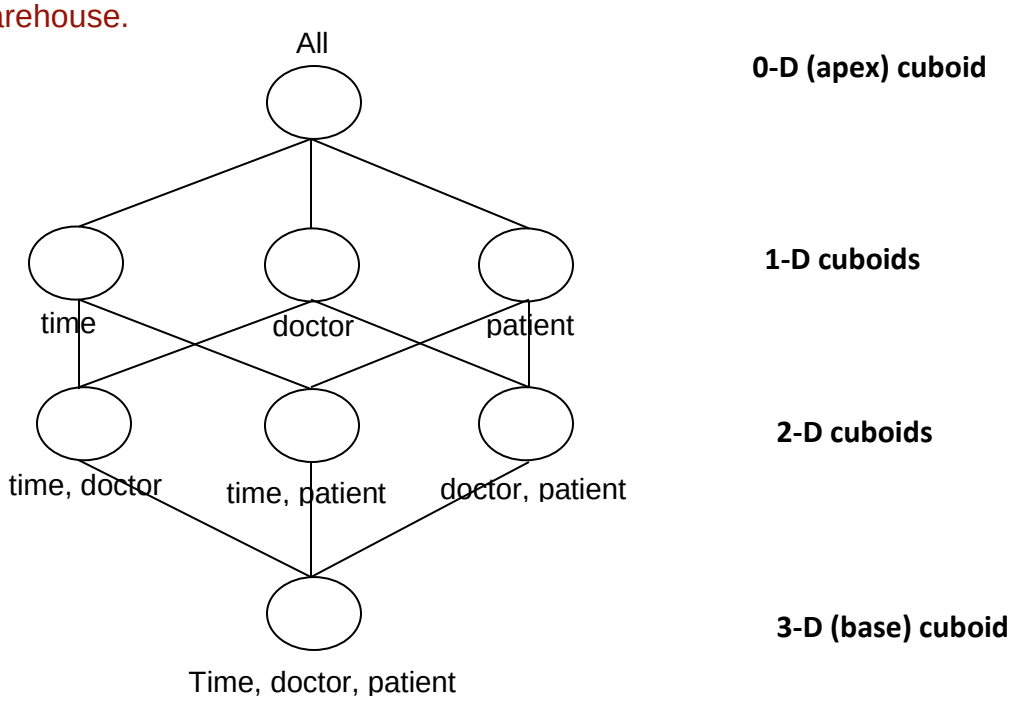
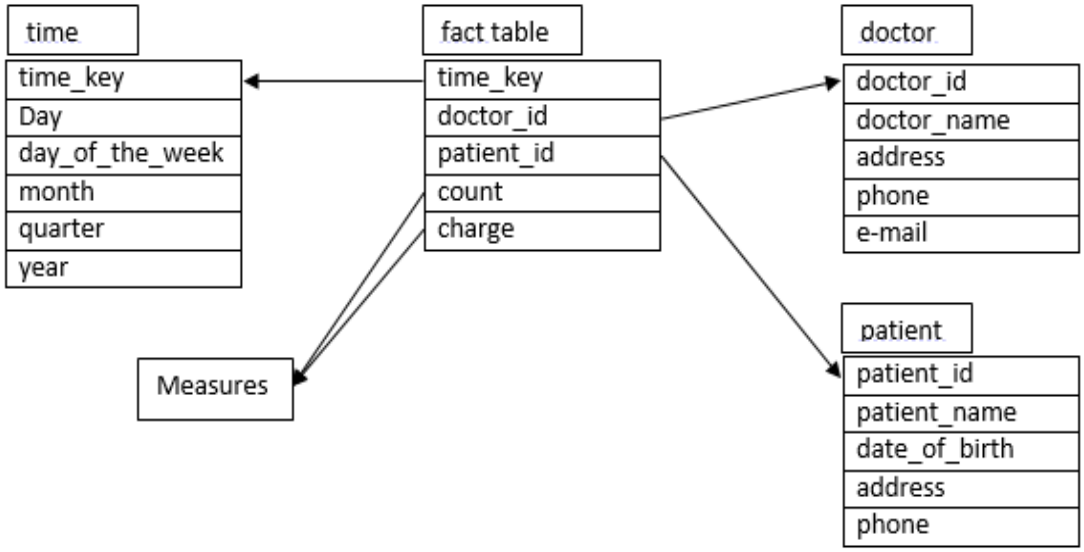
Assignment 2:

❖ Q1. Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit:

a) Draw the lattice of cuboids (from apex to base cuboid) for the above data warehouse.



b) Draw a star schema diagram for the above data warehouse. For each dimension, include the appropriate attributes (conceptual hierarchies).



c) Starting with the base cuboid [day; doctor; patient], what specific OLAP operations (roll-up, drill down, dice, slice) should be performed (based on your schema) in order to list the total fee collected by each doctor in 2004?

- First, we should use roll-up operation to get the year 2004(rolling-up from day then month to year). After getting that, we need to use slice operation to select (2004). Second, we should use roll-up operation again to get all patients. Then, we need to use slice operation to select (all). Finally, we get list the total fee collected by each doctor in 2004.

d) Assume that each dimension has four levels. How many cuboids will this cube contain? Use the equation in chapter 4.

$$Total\ number\ of\ cuboids = \prod_{i=1}^{n}(L_i + 1),$$

- Since the cube has 3 dimensions and each dimension has 4 levels (including all), the total number of cuboids will this cube contain → $4^3 = 64\ cuboids$

==========

❖ Q2. Describe shortly in your own words the general idea of:

a) Prediction cube.

- A technique for multidimensional data mining in which the cube space is explored for prediction tasks. A prediction cube is a cube structure in which all prediction models are stored in multidimensional data space and also it supports prediction in an OLAP manner. In this kind of cubes, each cell value is computed by looking to the predictive model which built on the data subset on that cell, and evaluate it, which means representing its predictive behavior. So, prediction cubes use these prediction models as building blocks to determine the interestingness of subsets, which have been identified to indicate accurate prediction.[1]

[1]: Han, J., & Kamber, M. (2012). Data mining concepts and techniques, third edition (3rd ed., p. 228). Waltham, Mass.: Morgan Kaufmann.

b) Ranking cube.

- This cube is used for efficient processing of top-k queries. It doesn't return a large set of indistinctive answers to a query, but instead the ranking query returns the best k

results which related to the user specified preferences. This preferences consist of a selection condition and a ranking function. The results will be returned in a specific order which rank the results by presenting the best at the top.

This method of Top-k queries are common in some applications related to searching such as searching web databases, and similarity queries in multimedia databases.[2]

[2]: Han, J., & Kamber, M. (2012). Data mining concepts and techniques, third edition (3rd ed., p. 225). Waltham, Mass.: Morgan Kaufmann.

==========