

Tutorial-1

1. A car dealer plans to build a data warehouse that would help them in analyzing the growth of the business in various segments namely low, middle, high in different quarters of the years. They want to analyze if there is any relation between the average sales of a particular quarter and the time of the year. They would also like to know if there were some model of cars offered but did not have any buyers for them for last two quarters. Relative performance (in terms of profit) among the three segments and average sales of cars in various states and cities of the country for each segment must be analyzed.
 - (a) Design a star schema for such a data warehouse clearly identifying the fact table(s) and dimension tables(s), their primary key(s) and foreign key(s). Your schema should at least be able to satisfy the above mentioned analysis requirements. Also give suitable schema definitions for the above design. You may consider other suitable attributes for the dimension table(s).
 - (b) Write an SQL query that runs on your schema and returns the average sales of “middle” segment from the state of Kerala for each fourth quarter during the years 2014-2018.
 - (c) Write an SQL query that runs on your schema and return the overall national profit (sale price- cost price) of “high” segment for each quarter during the year 2016.
 - (d) Draw cuboid of the above query.
 - (e) Write an OLAP operation on the above cuboid to get the national profit during every quarter of the year 2016.
2. Assume a university course database includes the following dimensions: *Address*, *Status*, *Major* and *CGPA*. The *Address* contains information about *Street*, *City*, *State* and *Country*. The status of a student can be *Undergraduate* or *Graduate*. *Graduate* courses are *MSC*, *MBA* and *MA* whereas the *Undergraduate* courses are *BSC*, *BBA*, and *BA*. A student can do major in *Arts*, *Business* and *Science*. *Arts* department offers *English*, *History* and *Philosophy*; *Business* offers *Accounting*, *Finance* and *Marketing*; *Science* offers *Physics*, *Chemistry* and *Biology*. The *CGPA* of a student is calculated in 10 point scale. CGPA of less than 5.0 is considered as *Fail*, more than 5.0 but less than 6.0 is *Pass*, more than 6.0 but less than 7.0 is *Satisfactory*, more than 7.0 but less than 8.0 is *Good*, more than 8.0 but less than 9.0 is *Very Good*, more than 9.0 is *Excellent*. Propose a concept hierarchy for the attributes *Address*, *Status*, *Major* and *CGPA*.
3. The following question is **optional** (not part of your syllabus). For solving the question, you may read the paper which you can download from the link: [Implementing data cubes efficiently](#)

Consider the following lattice of cuboids (A-H) along with a representation of the number of rows in each cuboid where A is the base cuboid. If you have to choose 3 views to materialize apart from the base cuboid, which of the views you choose using a greedy algorithm and how (show in a tabular form your choices sequentially) ? Assume that the cost of running a query is the same as the number of rows in the view from which it is derived.

