

EXPERIMENT 2

Vineet Rao

161070001

AIM:

To create star schema for given problem statement

THEORY:

Every data warehouse must be implemented in terms of lower level structures. Regardless of whether they are implemented with RDBMS, OLAP systems, file systems or specialized warehouse technology (like Teradata), the first step in making a warehouse is to make a schema.

There are two general ways to make a warehouse:

- The top-down approach: Create the model schema first and create the warehouse based on it. Useful for a mature, well-defined and distributed organization, with ready data stores or data marts that can be rapidly joined
- The bottom-up approach: Experiment with various data marts, finding different types of data to discover, and choose the preferable models and structure. This is useful for newer, less defined organizations, where immediate data discovery and mining is important, and structure and architecture can be defined secondarily
- The hybrid approach: Here, simultaneously an overarching schema is created, while experimenting with data marts and information, to get a best fit. This ensures some levels of structure from the start, and in the direction of org needs. At the same time it allows lot of discovery, which is needed if the objectives and requirements are not well-defined from the start.

Schemas for warehouses differ from other paradigms, like ER. In warehouses, there are two types of tables:

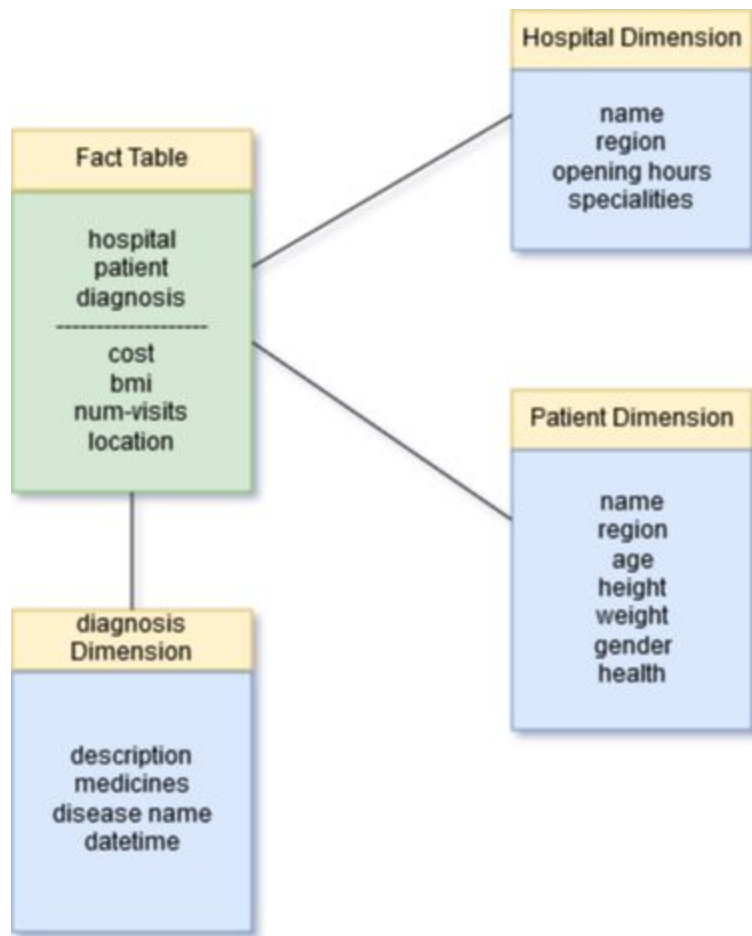
- Fact tables:
 - This contains the majority of the data
 - It connects various points with measures
 - It is often the (or one of the) central tables
- Dimension tables:

- Contains various dimensions and attributes
- Expansion of dimensions from fact tables
- May or may not contain measures

There are 3 major types of schemas:

- Star schema:
 - It consists of a central fact tables, and dimension tables
 - All dimension tables are directly associated with the fact table
 - Thus, the topology is like a star
 - It is the simplest schema, which is also very efficient to run queries on
 - However, it can lead to redundancy, and sometimes difficult to show complicated facts
- Snowflake schema:
 - It consists of a central fact table, and multiple dimension tables
 - However, dimension tables need not be directly connected to fact table
 - This allows for normalization, saving a lot of space
 - Joins become more expensive, taking longer
- Fact constellation:
 - It consists of multiple fact tables and multiple dimension tables
 - A complicated schema, due to there being lot of normalization and multiple tables for analysis
 - Useful when there are many departments that need different data
 - Can help distribute warehouse

STAR SCHEMA:



CONCLUSION:

Thus, a star schema has been created