

Multidimensional Data Modelling Concepts

Multidimensional data modelling is used for the designing Data warehouse, which will help in surround facts with as much relevant context dimensions. Multidimensionality is a design technique that separates the relational data into facts and dimensions. Multidimensional model present information to the end-user in a way that corresponds to his normal understanding of his business, key figures or facts from the different perspectives that influence them.



**KEZUNGULO
MEDIKHRU**
State Informatics Officer
NIC Nagaland
sio-ngl@nic.in



AJAY PK TATAPUDI
Scientific Officer-SB
NIC Nagaland
prasasta.tatapudi@nic.in

Data Warehouse is a subject-oriented, integrated, time variant and non-volatile collection of data in support of management's decision making process.

Business Intelligence are the tools and processes that enable users to query, report, analyze, monitor, and mine integrated information to make decisions, develop plans, or take action.

Why we need a separate DW?

DW involves the computation of large groups of data at summarized levels and may require the use of special data organization, access and implementation methods based on multi-dimensional views. The separation of operational database from DW is based on the different structures, contents and uses of the data.

MULTIDIMENSIONAL DATA MODEL

Dimensional modelling has two basic concepts - Facts & Dimension.

Facts: A business performance measurement, typically numeric or additive that is stored in a fact table. Fact table is a table with numeric performance measurements characterized by a composite key, each of whose elements is a foreign key drawn from a dimension table.

Fact Types are:

- **Additive Fact:** Measurements in a fact table excluding ratios and unit prices can be added across all the dimensions.

- **Semi-Additive Fact:** Numeric facts that can be added along some dimensions in a fact table.
- **Non-Additive Fact:** A fact that cannot logically be added between rows. May be numeric and therefore usually must be combined in a computation with other facts before being added across rows.

Dimension: An independent entity in a dimensional model that serves as an entry point or as a mechanism for slicing and dicing the additive measures located in the fact table of the dimension.

Dimension table: A table in dimensional model with a single -part primary key and descriptive attribute columns.

Slowly Changing Dimensions (SCD): is the tendency of dimension rows to change gradually or occasionally over time. There are three types of SCD such as:

- A type1 SCD is a dimension whose attributes are over written when the value of attribute changes.
- A type2 SCD is a dimension, a new row is created when the values of an attribute changes with new surrogate key.
- A type3 SCD is a dimension, an alternate old column is created. When an attribute changes a new column is added to dimension table to capture the change.

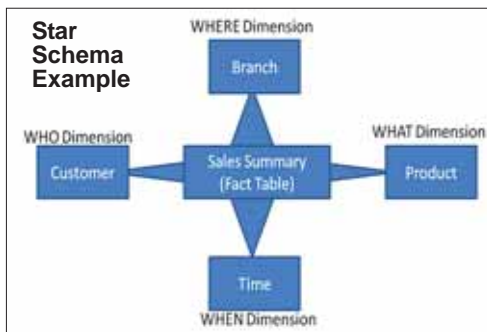
Surrogate Key: is an Integer key that are sequentially assigned. It is required in many data warehouse situations to handle Slowly Changing Dimension & missing or inapplicable data.

Data warehouse and OLAP tools are based on multidimensional data

model. This model view data in the form of data cube. It is defined by dimension and facts. Dimensions are the perspectives or entities with respect to which organizations want to keep records. A data warehouse requires a concise, subject oriented schema that facilitates on-line data analysis.

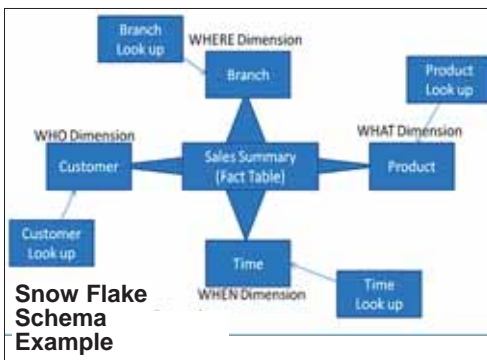
Types of multidimensional model:

- **Star Schema:** The most common modelling design is star Schema in which the data warehouse contains
 - A large central table (fact table) containing bulk of data with no redundancy.
 - A set of smaller attendant tables (dimension table) for each dimension.
 - The Schema resembles a star burst with dimension table display in a radi-



al pattern around the central fact table.

- **Snow Flake Schema:** is a variant of star schema model where some dimension tables are normalized, splitting the dimensional table into additional tables resulting schema graph forms a shape similar to a snowflake.
- **Fact Constellation:** some applica-



tion may require multiple fact tables to share dimension tables. This kind of schema can be viewed as a collection of stars, hence called galaxy schema or fact constellation.

Measures Categorization: A data cube measure is a numerical function that can be evaluated at each point in the data cube space. Its value is computed for given point by aggregating the data corresponding to the respective dimension-value pairs defining the giving point.

Concept Hierarchies: A concept hierarchy defines a sequence of mappings from a set of low-level concepts to higher-level concepts. It exists in dimension table.

OLAP operations in the multidimensional data model: 1.Roll-Up/Drill-Up 2.Drill down 3.Slice and Dice 4. Pivot (rotate) 5. Drill-across 6. Drill-through.

DATA WAREHOUSE/BI ASSESSMENT

Data Warehouse Assessment is the latter one that has been the biggest driving force. Within the organization, all departments and entities need to access information from a central repository that has integrated data. It is not a question of whether we need the data warehouse or not, but rather how efficiently can we build it. So, it is a good idea to assess the organization's readiness, expectations and acceptance criteria before you start building your data warehouse application. The goals and objectives for an assessment/strategy project include:

- **Assessment of Organization Reporting Strategy and Business Requirements:** Collect business needs for a data warehouse, enterprise goals, vision and initiatives define the benefits a data warehouse application will provide to the company.
- **Assessment of Data Warehouse Roadmap:** defines how organization requirements will be implemented and how legacy data will be

converted to information and subsequently to knowledge.

- **Assessment of Technical Architecture:** looks into existing legacy systems, database servers, scalability, database sizing, performance, web servers, job schedulers, and backup and recovery management. Identify type of Process and data flow, Initial and incremental data loads and Changed data capture.
- **Assessment of Databases and Tools:** determines the type of application needs to be implemented to meet the reporting requirement. The assessment of the tools concludes tools for extracting, transporting, and loading the data, building reports and queries, and accessing the reports.
- **Assessment of Data Availability, Data Access and Reports:** defines what data sources will be used and what will be the data acquisition strategy, types of reports will be generated Report will be Ad-hoc, Operational, or analytical and How users will access these reports.

GOVERNMENT AGENCY EXAMPLE

Government Agency is an organization that faces substantial business challenges over the next decade, due largely to projected demand increases, expected staff retirements, heightened service expectations on the part of the citizens served, and a tight Federal budget climate. Accordingly, the Agency is challenged, like many government organizations, to "do more with less" while maintaining or improving service delivery.

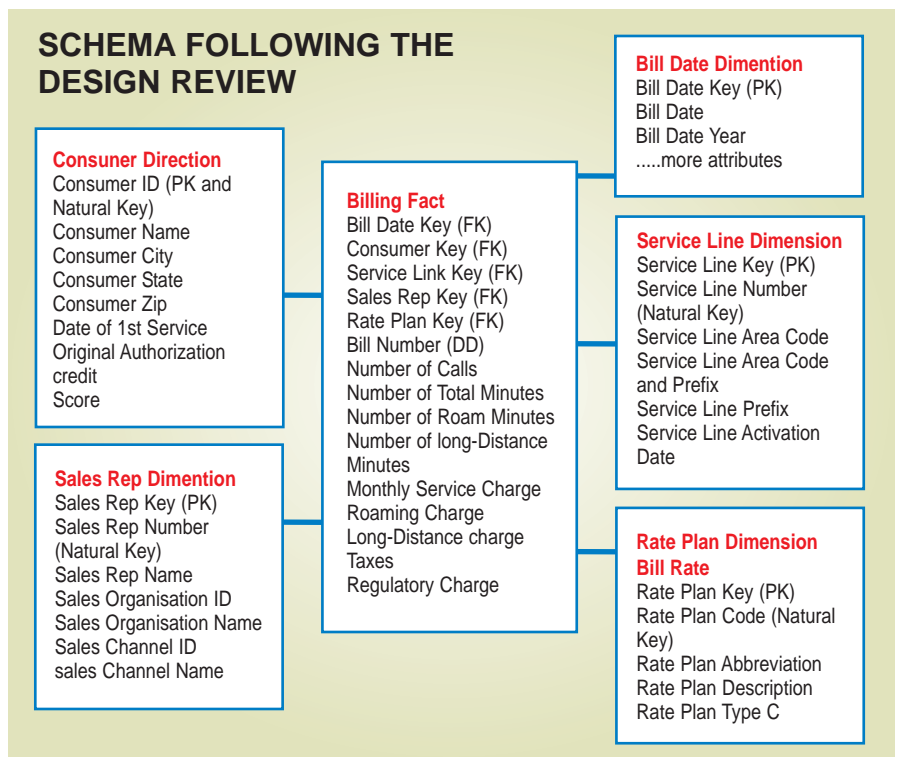
- The Agency funded a comprehensive project to ensure strategic alignment between its strategies, goals, and objectives, its key service delivery processes, and the BI applications required for making those processes more productive while improving service.
- The Agency has a record of continu-

ous process improvement and a culture that supports the goal of operational excellence. The company has come to realize the business value of business intelligence, and has budgeted funds for business process reengineering to capitalize on its BI investments.

- The Agency recognizes the need to change the culture within its operating units & embrace the use of enhanced information, modern analytical tools & advanced optimization models.
- The Agency is in the process of formally assessing its BI and DW technical readiness so that it can take the necessary steps to enhance its overall capabilities to acquire, cleanse, integrate, store, and deliver high quality information to feed the full spectrum of BI applications that will be required to run the agency productively and with high levels of service to citizens.
- The Agency employs effective IT governance mechanisms that promote an effective business/IT partnership, including executive level and working level steering committees, regular off site planning meetings, and web-based program communication mechanisms.

PILOT PROJECT TELECOMMUNICATIONS

Management now wants to analyze monthly usage and billing metrics (*revenue*) by customer, sales organization, and rate plan to perform sales rep and channel performance analysis and the rate plan analysis. Each month, the operational billing system generates a bill for each phone number, is called *service line*. Each service line is associated with a single customer. However, a customer can have multiple wireless service lines, which appear as separate line items on the same bill; each service line has its own set of billing metrics, such as the number of minutes used and monthly service charge. There is a single rate plan associated with each service line on a given bill; this plan can change



as customers' usage habits evolve. Lastly, a sales rep is associated with each service line in order to evaluate the ongoing billing revenue stream generated by each rep and channel partner.

GENERAL DESIGN CONSIDERATIONS

Fact Granularity: In an effort to improve performance or reduce query complexity, aggregated facts totals sometimes sneak into the fact row. These types of fact totals are dangerous because they are not perfectly additive. While these types of facts reduces the complexity and run time of a few specific queries, having it in the fact table invites a query to double count.

Dimension Granularity: If we design the snow-flaking or normalization of dimension tables, the demerit is query performance will be degraded.

Design Exercise: In our case service line has its own set of billing metrics; grain declaration would be one row per service line per bill. Move the service line key into the fact table as a foreign key to the service line dimension. Now every time a bill row is loaded into the fact table, a row also would be loaded

into the bill number dimension table. Define the bill number as a degenerate dimension in fact table. Define bill date dimension table and move the bill date into the fact table and join it to date dimension, which plays the role of a bill date in this schema. If you define Sales rep and Sales organization as separate dimensions then there will be a double joins on the sales rep organization dimension table. To avoid double joins including the sales rep organization and channel identifiers as additional attributes in the sales rep dimension table and define sales rep foreign key in the fact table. The other dimensions are Customer dimension, & Rate plan Dimension. Implement surrogate keys for the entire dimension as primary keys in dimensional table and as foreign keys in fact table.

Data warehousing and Data Mining can be implemented in areas such as data mining for Financial Data Analysis, Retail Industry, Biological Data Analysis, Telecommunication Industry, Scientific Applications and Intrusion Detection.