Principles of Dimensional Modeling

Unit:2

Dr. Shalini Gambhir

What is Dimensional Modeling?

- Dimensional modeling represents data with a cube operation, making more suitable logical data representation with OLAP data management.

- The perception of Dimensional Modeling was developed by **Ralph Kimball** and is consist of **"fact"** and **"dimension"** tables.

- In dimensional modeling, the transaction record is divided into either **"facts,"** which are **frequently numerical transaction data**, or **"dimensions,"** which are the **reference information that gives context to the facts.**

- Let's consider a retail sales database as an example of dimensional modeling:
- **Facts:**
  - Sales Amount
  - Quantity Sold
  - Discount Applied
  - Profit Margin
  - Timestamp (Date and Time of Sale)
  - Customer ID
  - Product ID
  - Store ID
- **Dimensions:**
  - Customer Dimension:
    - Customer ID
    - Customer Name
    - Customer Address
    - Customer Age
    - Customer Gender
  - Product Dimension:
    - Product ID
    - Product Name
    - Product Category
    - Product Brand
    - Product Price
  - Store Dimension:
    - Store ID
    - Store Name
    - Store Location
    - Store Manager
    - Store Size

In this example, each sales transaction record would contain facts such as the sales amount, quantity sold, etc. These facts are associated with various dimensions such as the customer, product, and store. The dimensions provide context and descriptive information about the facts. For instance, the sales amount is associated with a specific customer who made the purchase, a particular product that was sold, and the store where the transaction occurred.

Objectives of Dimensional Modeling

- The purposes of dimensional modeling are:
- To produce database architecture that is easy for end-clients to understand and write queries.
- To maximize the efficiency of queries. It achieves these goals by minimizing the number of tables and relationships between them.

Advantages of Dimensional Modeling

- Following are the benefits of dimensional modeling are:

- **Dimensional modeling is simple:** Dimensional modeling methods make it possible for warehouse designers to create database schemas that business customers can easily hold and comprehend. There is no need for vast training on how to read diagrams, and there is no complicated relationship between different data elements.
- **Dimensional modeling promotes data quality:** The star schema enable warehouse administrators to enforce referential integrity checks on the data warehouse. Since the fact information key is a concatenation of the essentials of its associated dimensions, a factual record is actively loaded if the corresponding dimensions records are duly described and also exist in the database.

- By enforcing foreign key constraints as a form of referential integrity check, data warehouse DBAs add a line of defense against corrupted warehouses data.
- **Performance optimization is possible through aggregates:** As the size of the data warehouse increases, performance optimization develops into a pressing concern. Customers who have to wait for hours to get a response to a query will quickly become discouraged with the warehouses. Aggregates are one of the easiest methods by which query performance can be optimized.

Disadvantages of Dimensional Modeling

- To maintain the integrity of fact and dimensions, loading the data warehouses with a record from various operational systems is complicated.
- It is severe to modify the data warehouse operation if the organization adopting the dimensional technique changes the method in which it does business.

Elements of Dimensional Modeling

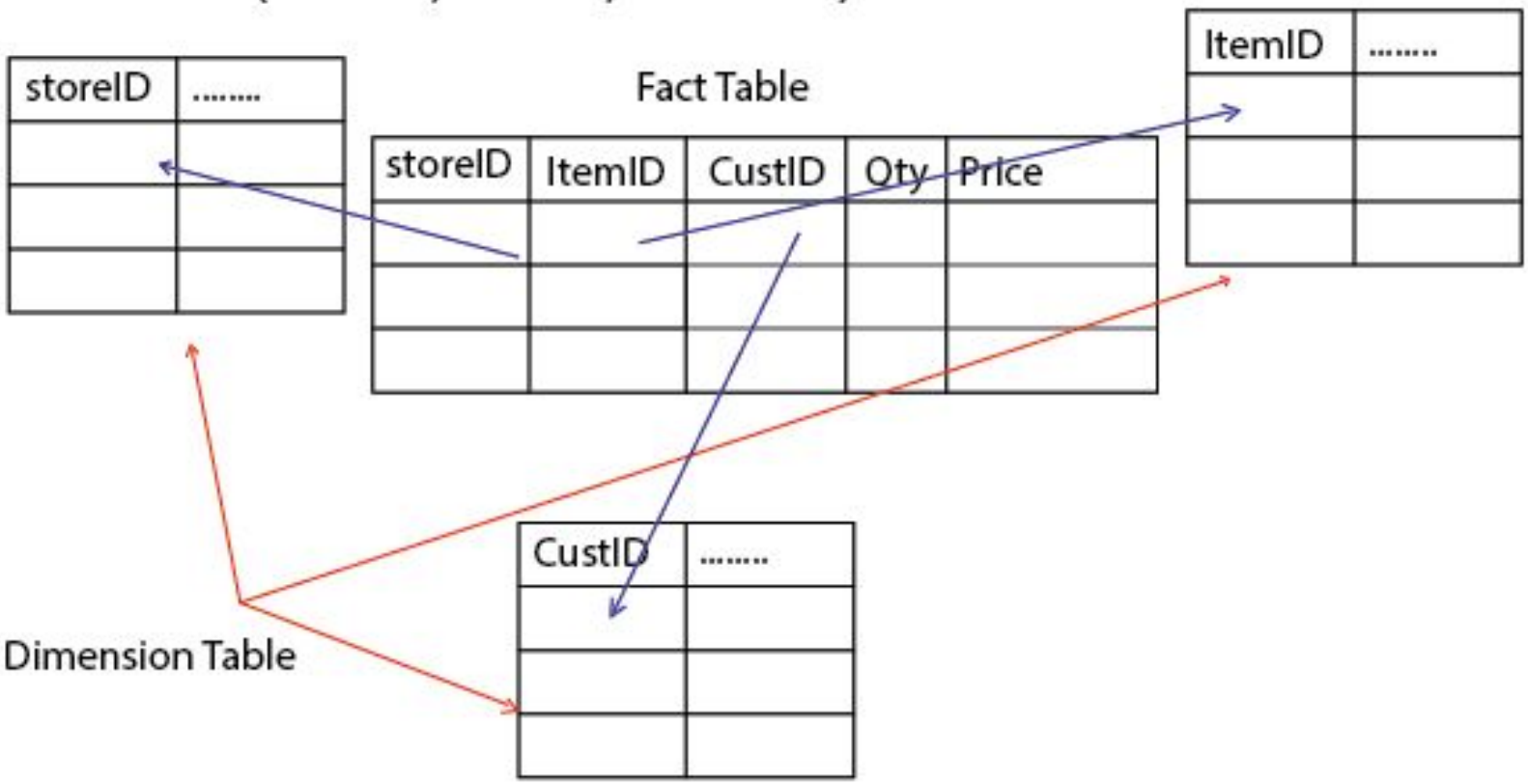
- Fact
- It is a collection of associated data items, consisting of measures and context data. It typically represents business items or business transactions.
- Dimensions
- It is a collection of data which describe one business dimension. Dimensions decide the contextual background for the facts, and they are the framework over which OLAP is performed.
- Measure
- It is a numeric attribute of a fact, representing the performance or behavior of the business relative to the dimensions.

- Considering the relational context, there are two basic models which are used in dimensional modeling:
- Star Model
- Snowflake Model
- The star model is the underlying structure for a dimensional model. It has one broad central table (fact table) and a set of smaller tables (dimensions) arranged in a radial design around the primary table. The snowflake model is the conclusion of decomposing one or more of the dimensions.

- Fact Table
- Fact tables are used to data facts or measures in the business. Facts are the numeric data elements that are of interest to the company.
- **Characteristics of the Fact table**
- The fact table includes numerical values of what we measure. For example, a fact value of 20 might means that 20 widgets have been sold.
- Each fact table includes the keys to associated dimension tables. These are known as foreign keys in the fact table.
- Fact tables typically include a small number of columns.
- When it is compared to dimension tables, fact tables have a large number of rows.

- Dimension Table
- Dimension tables establish the context of the facts. Dimensional tables store fields that describe the facts.
- The dimension tables include descriptive data about the numerical values in the fact table. That is, they contain the attributes of the facts. For example, the dimension tables for a marketing analysis function might include attributes such as time, marketing region, and product type.
- Since the record in a dimension table is denormalized, it usually has a large number of columns. The dimension tables include significantly fewer rows of information than the fact table.
- The attributes in a dimension table are used as row and column headings in a document or query results display.
- **Example:** A city and state can view a store summary in a fact table. Item summary can be viewed by brand, color, etc. Customer information can be viewed by name and address.

Sales (StoreID, ItemID, CustID, qty, price)
StoreID (storeid, city, state)
ItemID (itemid, category, brand, color, size)
CustID (custid, name, address)



In this example, Customer ID column in the facts table is the foreign keys that join with the dimension table. By following the links, we can see that row 2 of the fact table records the fact that customer 3, Gaurav, bought two items on day 8.

Fact Table

Time ID	Product ID	Customer ID	Unit Sold
4	17	2	1
8	21	3	2
8	4	1	1

Dimension Tables

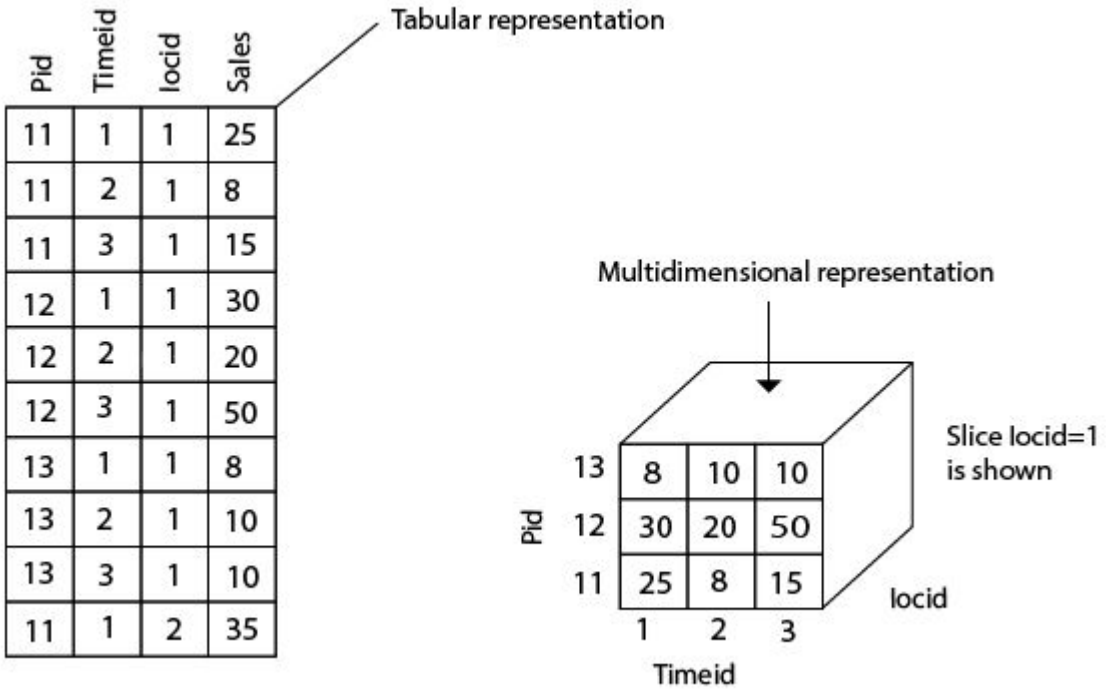
Customer ID	Name	Gender	Income	Education	Region
1	Rohan	Male	2	3	4
2	Sandeep	Male	3	5	1
3	Gaurav	Male	1	7	3

Hierarchy

- A hierarchy is a directed tree whose nodes are dimensional attributes and whose arcs model many to one association between dimensional attributes team. It contains a dimension, positioned at the tree's root, and all of the dimensional attributes that define it.

What is Multi-Dimensional Data Model?

- A multidimensional model views data in the form of a data-cube. A data cube enables data to be modeled and viewed in multiple dimensions. It is defined by dimensions and facts.
- The dimensions are the perspectives or entities concerning which an organization keeps records. For example, a shop may create a sales data warehouse to keep records of the store's sales for the dimension time, item, and location. These dimensions allow the save to keep track of things, for example, monthly sales of items and the locations at which the items were sold. Each dimension has a table related to it, called a dimensional table, which describes the dimension further. For example, a dimensional table for an item may contain the attributes item_name, brand, and type.
- A multidimensional data model is organized around a central theme, for example, sales. This theme is represented by a fact table. Facts are numerical measures. The fact table contains the names of the facts or measures of the related dimensional tables.



A multidimensional data model is organized around a central theme, for example, sales. This theme is represented by a fact table. Facts are numerical measures. The fact table contains the names of the facts or measures of the related dimensional tables.

Consider the data of a shop for items sold per quarter in the city of Delhi. The data is shown in the table. In this 2D representation, the sales for Delhi are shown for the time dimension (organized in quarters) and the item dimension (classified according to the types of an item sold). The fact or measure displayed in rupee_sold (in thousands).

Location="Delhi"				
Time (quarter)	item (type)			
	Egg	Milk	Bread	Biscuit
Q1	260	508	15	60
Q2	390	256	20	90
Q3	436	396	50	40
Q4	528	483	35	50

Now, if we want to view the sales data with a third dimension, For example, suppose the data according to time and item, as well as the location is considered for the cities Chennai, Kolkata, Mumbai, and Delhi. These 3D data are shown in the table. The 3D data of the table are represented as a series of 2D tables.

	Location="Chennai"				Location="Kolkata"				Location="Mumbai"				Location="Delhi"			
	item				item				item				item			
Time	Egg	Milk	Bread	Biscuit	Egg	Milk	Bread	Biscuit	Egg	Milk	Bread	Biscuit	Egg	Milk	Bread	Biscuit
Q1	340	360	20	10	435	460	20	15	390	385	20	39	260	508	15	60
Q2	490	490	16	50	389	385	45	35	463	366	25	48	390	256	20	90
Q3	680	583	46	43	684	490	39	48	568	594	36	39	436	396	50	40
Q4	535	694	39	38	335	365	83	35	338	484	48	80	528	483	35	50

Conceptually, it may also be represented by the same data in the form of a 3D data cube, as shown in fig: