Experiment No. : 3

Title: Study and analyze the dimensional models for different application domains

Objectives:
1. To study and Analyze the ER Model and Dimensional models for different application domains.
2. To identify
3. To identify Grain
4. To identify Fact Tables and types of fact tables
5. To identify Dimension Tables and types of Dimension tables

Key concepts: ER Model, Dimensional models, Business Process, Grain, Fact Tables, Dimension Tables, Star Schema

Theory:
What is Dimensional Model?
A dimensional model is a data structure technique optimized for Data warehousing tools. The concept of Dimensional Modelling was developed by Ralph Kimball and is comprised of "fact" and "dimension" tables.

A Dimensional model is designed to read, summarize, analyze numeric information like values, balances, counts, weights, etc. in a data warehouse. In contrast, relation models are optimized for addition, updating and deletion of data in a real-time Online Transaction System.

These dimensional and relational models have their unique way of data storage that has specific advantages.

For instance, in the relational mode, normalization and ER models reduce redundancy in data. On the contrary, dimensional model arranges data in such a way that it is easier to retrieve information and generate reports.

Hence, Dimensional models are used in data warehouse systems and not a good fit for relational systems.

Steps of Dimensional Modelling
Step 1) Identify the business process
Step 2) Identify the grain
Step 3) Identify the dimensions
Step 4) Identify the Fact
Step 5) Build Schema

Elements of Dimensional Data Model

Fact
Facts are the measurements/metrics or facts from your business process. For a Sales business process, a measurement would be quarterly sales number

Dimension
Dimension provides the context surrounding a business process event. In simple terms, they give who, what, where of a fact. In the Sales business process, for the fact quarterly sales number, dimensions would be

Who – Customer Names
Where – Location
What – Product Name
In other words, a dimension is a window to view information in the facts.

Attributes
The Attributes are the various characteristics of the dimension.

In the Location dimension, the attributes can be
State
Country
Zipcode etc.
Attributes are used to search, filter, or classify facts. Dimension Tables contain Attributes

Fact Table
A fact table is a primary table in a dimensional model.

A Fact Table contains

Measurements/facts
Foreign key to dimension table

Dimension table
A dimension table contains dimensions of a fact.
They are joined to fact table via a foreign key.

Dimension tables are de-normalized tables.
The Dimension Attributes are the various columns in a dimension table
Dimensions offers descriptive characteristics of the facts with the help of their attributes
No set limit set for given for number of dimensions
The dimension can also contain one or more hierarchical relationships

Steps of Dimensional Modelling

The accuracy in creating your Dimensional modeling determines the success of your data warehouse implementation. Here are the steps to create Dimension Model

Identify Business Process
Identify Grain (level of detail)
Identify Dimensions
Identify Facts
Build Star

The model should describe the Why, How much, When/Where/Who and What of your business process

Step 1) Identify the business process
Identifying the actual business process a datawarehouse should cover. This could be Marketing, Sales, HR, etc. as per the data analysis needs of the organization. The selection of the Business process also depends on the quality of data available for that process. It is the most important step of the Data Modelling process, and a failure here would have cascading and irreparable defects.

Step 2) Identify the grain
The Grain describes the level of detail for the business problem/solution. It is the process of identifying the lowest level of information for any table in your data warehouse. If a table contains sales data for every day, then it should be daily granularity. If a table contains total sales data for each month, then it has monthly granularity.

During this stage, you answer questions like

Do we store the product sale information on a monthly, weekly, daily or hourly basis? This decision depends on the nature of reports requested by executives
How do the above two choices affect the database size?
Example of Grain:

The CEO at an MNC wants to find the sales for specific products in different locations on a daily basis.

So, the grain is "product sale information by location by the day."

Step 3) Identify the dimensions
Dimensions are nouns like date, store, inventory, etc. For example, the date dimension may contain data like a year, month and weekday.

Example of Dimensions:

The CEO at an MNC wants to find the sales for specific products in different locations on a daily basis.

Dimensions: Product, Location and Time

Attributes: For Product: Product key (Foreign Key), Name, Type, Specifications

Hierarchies: For Location: Country, State, City, Street Address, Name

Step 4) Identify the Fact
This step is co-associated with the business users of the system because this is where they get access to data stored in the data warehouse. Most of the fact table rows are numerical values like price or cost per unit, etc.

Example of Facts:

The CEO at an MNC wants to find the sales for specific products in different locations on a daily basis.

The fact here is Sum of Sales by product by location by time.

Step 5) Build Schema
In this step, you implement the Dimension Model. A schema is nothing but the database structure (arrangement of tables). There are two popular schemas

Star Schema
The star schema architecture is easy to design. It is called a star schema because diagram resembles a star, with points radiating from a center. The center of the star consists of the fact table, and the points of the star is dimension tables.

The fact tables in a star schema which is third normal form whereas dimensional tables are de-normalized.

Snowflake Schema
The snowflake schema is an extension of the star schema. In a snowflake schema, each dimension are normalized and connected to more dimension tables.

Rules for Dimensional Modelling

Load atomic data into dimensional structures.
Build dimensional models around business processes.
Need to ensure that every fact table has an associated date dimension table.
Ensure that all facts in a single fact table are at the same grain or level of detail.

Benefits of dimensional modeling

Standardization of dimensions allows easy reporting across areas of the business.
Dimension tables store the history of the dimensional information.
The dimensional model is very understandable by the business. This model is based on business terms, so that the business knows what each fact, dimension, or attribute means.
Dimensional models are deformalized and optimized for fast data querying. Dimensional modeling creates a schema which is optimized for high performance. It means fewer joins and helps with minimized data redundancy.
The dimensional model also helps to boost query performance. It is more denormalized therefore it is optimized for querying.

There are three types of facts

1. Additive 2. Non-additive 3. Semi- additive.

Types of Dimensions are

Conformed, Outrigger, Shrunken, Role-playing, Junk, Degenerate

Task to be Performed

Refer the pdf “Dimensional_Modelling_by_Example” for ER model and dimension models for different application
Domains and identify business process, identify Grain, identify Fact Tables and types of fact tables, identify Dimension Tables and types of Dimension tables