Data Warehousing Concepts

Lesson 3: Dimensional Modeling

June 17, 2014   Proprietary and Confidential   – 1 –

IGATE
Speed. Agility. Imagination

Lesson Objectives

➢ **In this lesson, you will learn:**
- – What is Dimensional modeling ?
- – Facts and Dimension tables
- – Database schema
- – Schema Design for Modeling

June 12, 2014   Proprietary and Confidential   – 2 –

IGATE
Speed.Agility.Imagination

3.1: Dimensional Modeling
What is Dimensional Modeling?

➢ Dimensional Modeling (DM) is the name of a logical design technique often used for Data Warehouses.
➢ DM is the technique for databases that are designed to support end-user queries in a Data Warehouse.
➢ A Dimension Model is composed of dimension tables and fact tables.
➢ It provides a conceptual framework.
➢ It simplifies the business flow.
➢ It is structurally classified as fact or dimension.

June 17, 2014   Proprietary and Confidential   - 3 -   IGATE

Dimensional Modeling:
➢ **Dimensional Modeling** has the characteristic for organizing data roughly into base facts and dimensions of those facts.
➢ Dimensional Modeling provides the Conceptual Framework. It is basically used for faster query performance for the business users. **Facts** are basically organization's business processes. They are usually numeric values. **Dimension** is a context that describes the fact.
➢ Every organization has Dimensional Modeling for its business processes, and it consists of **fact tables** and **dimensional tables**. It helps business users in easily understanding the typical system model.
➢ Dimensional Modeling represents the complexities of the business process in a simple manner. **Understandability** and **Query performance** are two major reasons for which dimensional modeling is accepted widely in the industry.
➢ Dimensional Modeling is a logical design technique that allows to retrieve the data with high-performance.

3.2: Fact and Dimension Tables
Concepts of Fact and Dimension Tables

➤ Fact tables and Dimension tables are the two types of objects that are commonly used in designing database schemas.
➤ Fact table contains two columns, namely numeric facts and foreign keys of dimension tables.
➤ Dimension tables contain the attributes that describe fact records.

IGATE

June 17, 2014   Proprietary and Confidential   - 4 -

Fact and Dimension Tables:
➤ A **fact table** has two types of columns.
  - The first column type contains numeric facts (often called measurements).
  - The other column type contains the foreign keys of dimension tables.
➤ A **fact table** contains multiple foreign keys.
➤ Each pair of primary key in dimension and foreign key of fact table contains the measurements.
➤ A **Dimension table** contains the attributes that describe fact records. Some dimension table attributes provide descriptive information and other attributes (primary key) are used to join with fact tables.
   Example: A customer dimension table contains two attributes, namely customer id (Primary key) and customer description. So we will use the primary key attribute customer id to join with fact tables.
➤ However, **dimensional** and **fact modeling** is not of the highest Normal Form, but makes use of a key of performance indicators. Dimensions can strive to be in Boyce Codd (BCNF) 3rd Normal Form. Whereas Fact tables may be in 1st Normal Form, having only a primary key being unique.

Multidimensional Data

- ➤ Designed to resolve complex business queries
- ➤ Helps to analyze data from different dimensions
- ➤ Different dimensions form a cube
- ➤ Every edge represents a dimension

Dimensions: Product, Region, Time
Hierarchical summarization paths

Product	Region	Time
Industry	Country	Year
Category	Region	Quarter
Product Week	City	Month
	Office	Day

IGATE
Speed Agility Imagination

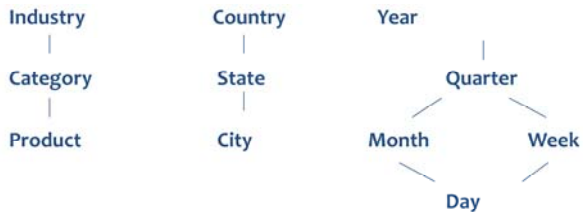June 17, 2014   Proprietary and Confidential   - 5 -

Multidimensional Data :

The multidimensional data model is the integral part of On line analytical Processing. The multidimensional data model is designed to resolve the complex queries.

In the logical multidimensional model, a cube represents the measures with same shape. In a cube every edge represents a dimension. Members of Dimension are aligned on the edges and divide the cube shape into cells in which stored the data values. It is basically used for developing data mart.

In above cube three edges represent the three dimension table Product, Region and time.

Multidimensional Data Analysis

➤ Analysis based on multiple dimensions
➤ Result varies with the dimension change across analysis
➤ Customers (city, state, country)
➤ Time (day, week, month, quarter, year)
➤ Products (product, category, industry)
➤ Hierarchies on dimensions:

Industry	Country	Year
Category	State	Quarter
Product	City	Month / Week
		Day

IGATE
Speed Agility Imagination

June 17, 2014   Proprietary and Confidential   - 6 -

Multidimensional Data Analysis is the analysis of data based on dimensions. It includes analysis of a particular data with respect to different and multiple dimensions. The value varies when there is a change in the dimensions across the analysis. It changes in terms of context one wishes to analyze data.

For E.g. Analysis of Product by City, Transactions for last 3 years.

3.3: Database Schema
Concept of Database Schema

➤ **Database schema includes various elements to store data.**
   Example: facts, dimensions, attributes, hierarchy, cube
➤ **Facts are numeric values to be stored in the database.**
➤ **Dimensions are description about facts.**
➤ **Attributes are characteristics of dimensions.**
➤ **Hierarchy is a logical representation of the order of the entities.**

June 17, 2014   Proprietary and Confidential   - 7 -

IGATE
Speed Agility Imagination

Fact and Dimension Tables:
Database Schema:
➤ **Database schema** is a set of facts in multi-dimensional data. A fact has a measure dimension quantity that is analyzed, for example, number of visas.
➤ It has a set of dimensions on which data is analyzed, for example, country, consulate, date of issue for a visa. Each dimension has a set of attributes
➤ **Example:** "Visa" dimension has visa date, visa type, and visa category
➤ Attributes of a dimension may be related by partial order, or Hierarchy: for example, post > county > region.

Fact and Dimension Tables:

Schema Design for Modeling:

➢ **Schema design** is the organization of database for modeling.

➢ The design shows how the model will be implemented in a system. It must be kept simple and familiar with the business context. It must be easily understood by business user. It should be designed in such a way that the business users can fully understand it in terms of facts, measures, dimensions, and hierarchies.

Schema Types:

➢ **Star Schema-Fact and Dimension tables:** Star schema has all multi-leveled dimensions that are flattened.

➢ **Snowflake Schema:** It has dimensional hierarchy directly by normalizing tables. In Snowflake schema, at least one multi-leveled dimension is kept separate.

Star Schema

> Star Schema consists of a central fact table surrounded by dimension tables.
> The measures of interest for OLAP are stored in the fact table (for example: Dollar Amount, Units in the table SALES).

IGATE

June 17, 2014   Proprietary and Confidential   - 9 -
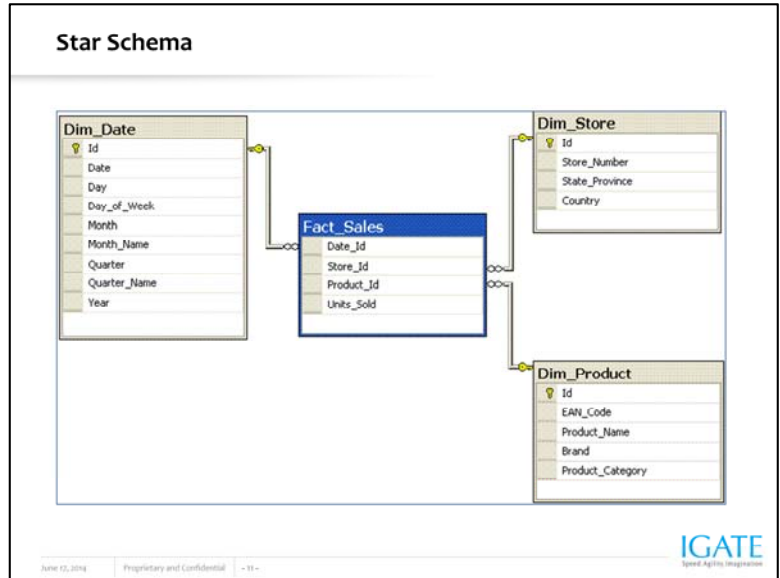
Fact and Dimension Tables:
Schema Types (contd.):

Star Schema (contd.):
> For each dimension of the multidimensional model there exists a dimension table (for example: Geography, Product, Time, Account) with all the levels of aggregation and the extra properties of these levels.
> It consists of Fact table.
> It consists Compound primary key.
> Star schema focuses on two major advantages, namely:
   -    Ease of use
   -    Efficient performance
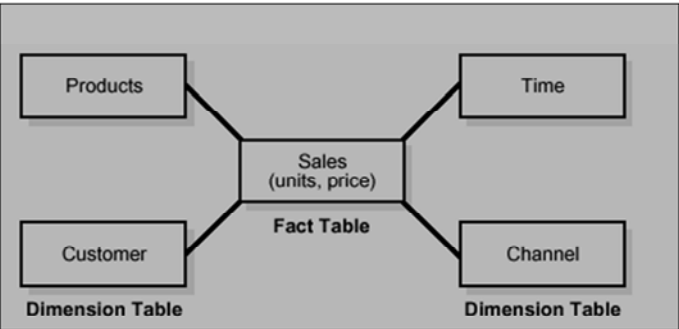
Fact and Dimension Tables:
Schema Types (contd.):

Star Schema:
➢ **Star schema** is commonly used by relational databases. The performance can be improved by using this design rather than traditional join operations. A **Star schema** is a database design that contains a central table, called a **fact table**, which is in relationship with many tables called **dimension tables**. This schema design resembles a star, thus the name is Star Schema. It is a very simple programmatic approach. It is very similar way in which a user thinks about a system, hence it is simple. It is easier to use. It is very efficient in the performance. It is best suited for **MOLAP application tools**. Typically, most of the fact tables in a star schema are in database Third Normal Form, while dimensional tables are de-normalized (Second Normal Form). Despite the fact that the Star schema is the simplest Data warehouse architecture, it is most commonly used in the Data warehouse implementations about 90-95%, across the world today.
   Example:
   - Fact Table: Fact_Sales table
   - Dimension table: Dim_Date, Dim_Store, Dim_Product

Snowflake Schema

➤ Snowflake Schema represents dimensional hierarchy directly by normalizing tables.
➤ It is a variation on the star schema.
➤ It is eazy to maintain and saves storage, very large dimension tables.
➤ They have improved query performance.

June 17, 2014   Proprietary and Confidential   - 12 -

IGATE
Speed Agility Imagination

Fact and Dimension Tables:
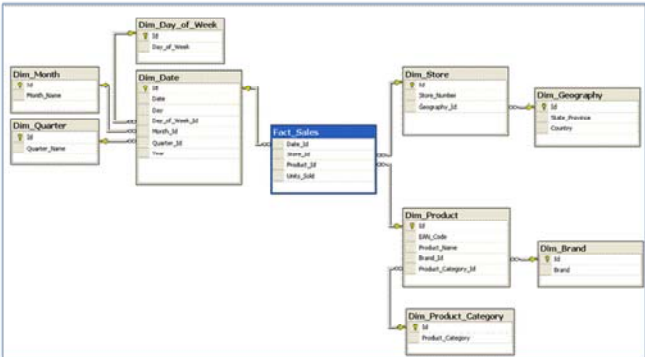**Schema Types (contd.):**
Snowflake Schema:
➤ It is more complex than Star schema design. The main difference is that dimensional tables in a snowflake schema are normalized, so they have a typical relational database design.
➤ Snowflake schemas are generally used when a dimensional table becomes very big and when a Star schema cannot represent the complexity of a data structure. For example, if a PRODUCT dimension table contains millions of rows, then the use of Snowflake schemas should significantly improve performance by moving out some data to other table (with REGION for instance). The data redundancy is eliminated. The problem is that the more normalized the dimension table is, the more complicated SQL joins must be issued to query them. This is because in order for a query to be answered, many tables need to be joined.
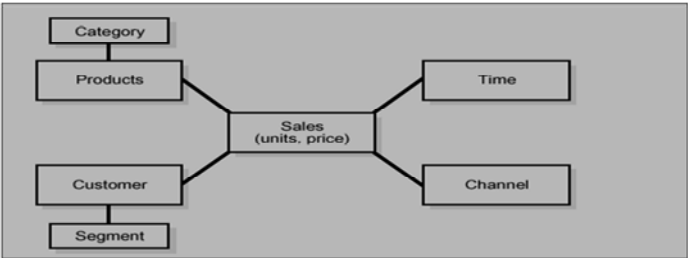
Summary

➤ **In this lesson, you have learnt:**
  – Dimensional Modeling represents the complexities of the business process in a simple manner.
  – The schema types are star schema and snowflake schema

IGATE
Speed.Agility.Imagination

June 17, 2014   Proprietary and Confidential   – 15 –

Summary

➤ Database schema has various elements, such as:
  – Fact
  – Dimension
  – Attributes
  – Hierarchy
  – Cube
➤ Schema design is the organization of database.

June 17, 2014   Proprietary and Confidential   – 16 –

IGATE
Speed Agility Imagination

Review Question

➢ Question 1: _____are description about facts.
➢ Question 2: _____ in Snowflake Shema are normalized into multiple tables.
➢ Question 3: ___ is the name of a logical design technique often used for Data Warehouses.

June 12, 2014   Proprietary and Confidential   – 17 –

IGATE
Speed Agility Imagination