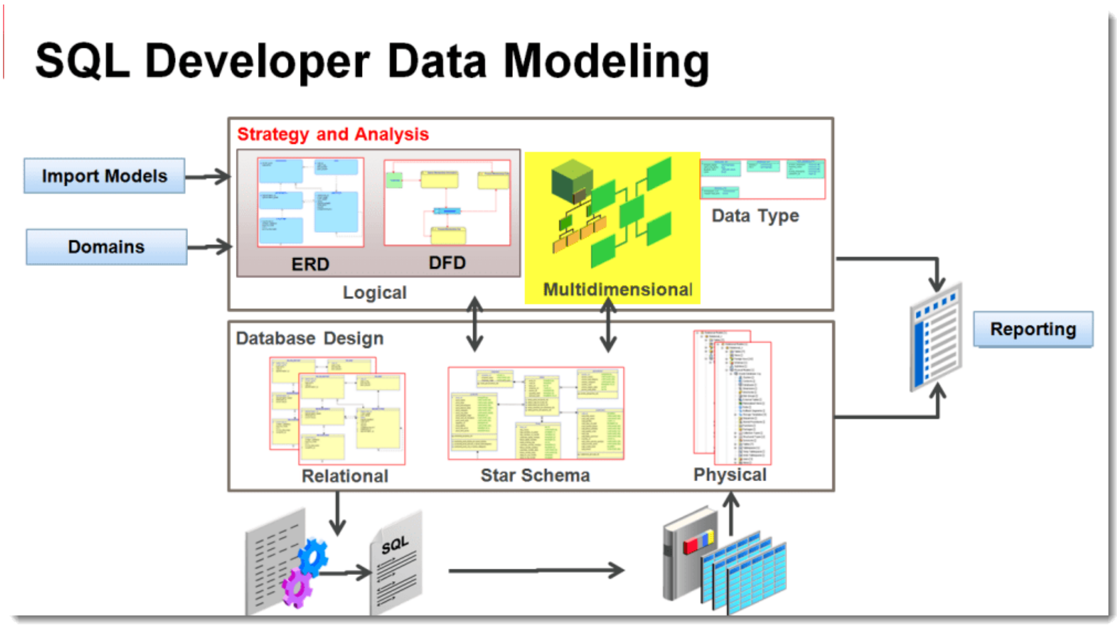
Lecture 10

Dimensional Modeling

Introduction to Data Dimensional Modelling

Data Dimensional Modelling (DDM) is a technique that uses Dimensions and Facts to store the data in a Data Warehouse efficiently. It optimises the database for faster retrieval of the data. Dimensional Models have a specific structure and organise the data to generate reports that improve performance.

It stores the data in the most optimised way to ensure there is no redundancy of the data and to improve performance. The Data Dimensional Model for an SQL Developer looks as follows:



https://www.thatjeffsmith.com/archive/2014/01/dimensional-modeling-in-oracle-sql-developer-data-modeler/

A Dimension Table usually contains the Dimension of the business, and Facts are usually transactions.

For example, in an E-Commerce use case, Dimensions can be products, customer, order items, departments, categories, etc. At the same time, Facts can be order details which link to the dimensions via foreign keys.

Key Features of Dimensional Data Modelling

Data Dimensional Modelling has gained popularity because of its unique way of analysing data present in different Data Warehouses. The 3 main features of DDM are as follows:

Easy to Understand: DDM helps developers create and design databases and Schemas easily interpreted by business users. The relationship between Dimensions and Facts are pretty simple to read and understand.

Promote Data Quality: DDM schemas enforce data quality before loading into Facts and Dimensions. Dimension and Fact are tied up by foreign keys that act as a constraint for referential integrity check to prevent fraudulent data from being loaded onto Schemas.

Optimise Performance: DDM breaks the data into Dimensions and Facts and links them with foreign keys, thereby reducing the data redundancy. The data is stored in the optimized form and hence occupies less storage and can be retrieved faster.

Components of Dimensional Data Modelling

There are 5 main components of any DDM. They are given below.

1) Dimension

Dimensions are the assortment of information that contain data around one or more business measurements. It may be topographical information, item data, contacts, and so on. Dimensions give the context to the Fact creation.

2) Facts

Facts are the collection of measurements, metrics, transactions, etc., from different business processes. It typically contains business transactions and measure values.

3) Attributes/Measures

Attributes are the elements of the Dimension Table. For example, in an account Dimension, the attributes can be:

First Name

Last Name

Phone, etc.

4) Fact Tables

Fact tables are utilized to store measures or transactions in the business. For example, in Internet business, a Fact can store the requested amount of items. Fact Tables, as a rule, have huge rows and fewer columns. The Fact Tables are related to Dimension Tables with the keys known as foreign keys.

5) Dimension Tables

Dimension Tables store the Dimensions from the business and establish the context for the Facts. They contain descriptive data that is linked to the Fact Table. Dimension Tables are usually optimized tables and hence contain large columns and fewer rows. For example:

Contact information can be viewed by name, address and phone dimension.

Product information can be viewed by product-code, brand, color, etc.

City, state, etc. can view store information.

Types of Dimensions in Dimensional Data Modelling

There are 9 types of Dimensions/metrics when dealing with Dimensional Data Modelling. They are given below:

1) Conformed Dimension

A Conformed Dimension is a type of Dimension that has the same meaning to all the Facts it relates to. This type of Dimension allows both Dimensions and Facts to be categorised across the Data Warehouse.

2) Outrigger Dimension

An Outrigger Dimension is a type of Dimension that represents a connection between different Dimension Tables.

3) Shrunken Dimension

A Shrunken Dimension is a perfect subset of a more general data entity. In this Dimension, the attributes that are common to both the subset and the general set are represented in the same manner.

4) Role-Playing Dimension

A Role-Playing Dimension is a type of table that has multiple valid relationships between itself and various other tables. Common examples of Role-Playing Dimensions are time and customers. They can be utilised in areas where certain Facts do not share the same concepts.

5) Dimension to Dimension Table

This type of table is a table in the Star Schema of a Data Warehouse. In a Star Schema, one Fact Table is surrounded by multiple Dimension Tables. Each Dimension corresponds to a single Dimension Table.

6) Junk Dimension

A Junk Dimension is a type of Dimension that is used to combine 2 or more related low cardinality Facts into one Dimension. They are also used to reduce the Dimensions of Dimension Tables and the columns from Fact Tables.

7) Degenerate Dimension

A Degenerate Dimension is also known as a Fact Dimension. They are standard Dimensions that are built from the attribute columns of Fact Tables. Sometimes data are stored in Fact Tables to avoid duplication.

8) Swappable Dimension

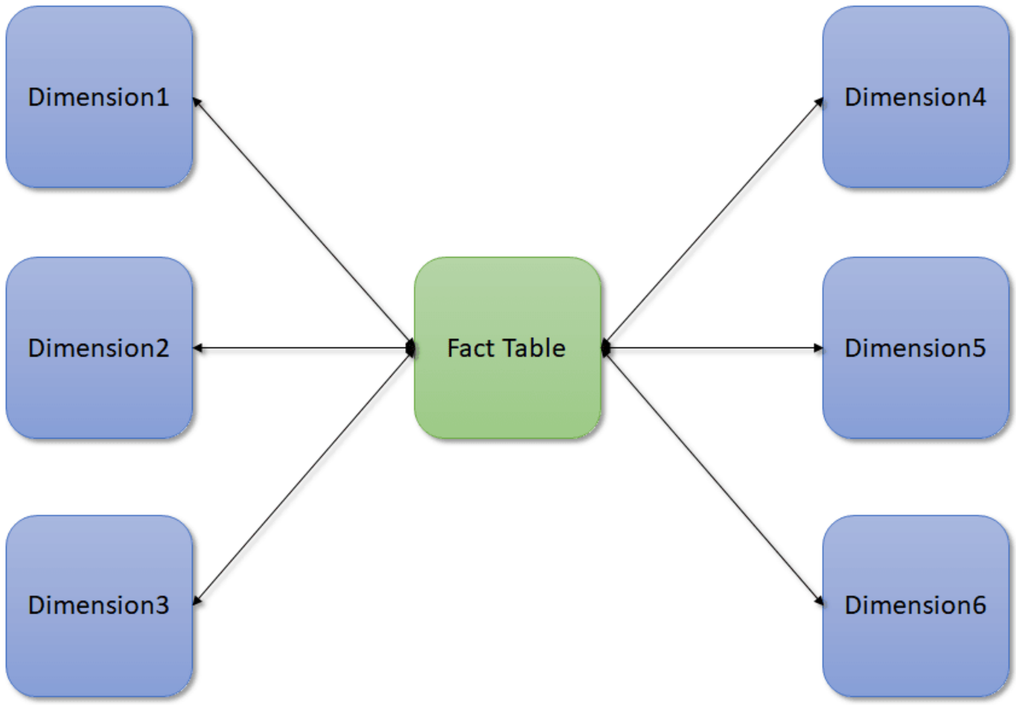
A Swappable Dimension is a type of Dimension that has multiple similar versions of itself which can get swapped at query time. The structure of this Dimension is also different and it has fewer data when compared to the original Dimension. The input and output are also different for this Dimension.

9) Step Dimension

This is a type of Dimension that explains where a particular step fits into the process. Each step is assigned a step number and how many steps are required by that step to complete the process.

Steps to Carry Out Dimensional Data Modelling

Dimensional Data Modelling requires certain analysis on the data to understand data behaviour and domain. The main goal a Dimensional Data Model tries to address is that it tries to describe the Why, How, When, What, Who, Where of the business process in detail. The typical architecture of a DDM is shown below:



The major steps to start Dimensional Data Modelling are:

1) Identify the Business Process

A Business Process is a very important aspect when dealing with Dimensional Data Modelling. The business process helps to identify what sort of Dimension and Facts are needed and maintain the quality of data. To describe business processes, you can use Business Process Modelling Notation (BPMN) or Unified Modelling Language (UML).

2) Identify Grain

Identification of Grain is the process of identifying how much normalisation (lowest level of information) can be achieved within the data. It is the stage to decide the incoming frequency of data (i.e.daily, weekly, monthly, yearly), how much data we want to store in the database (one day, one month, one year, ten years), and how much the storage will cost.

3) Identify the Dimensions

Dimensions are the key components in the Dimensional Data Modelling process. It contains detailed information about the objects like date, store, name, address, contacts, etc. For example, in an E-Commerce use case, a Dimension can be:

Product

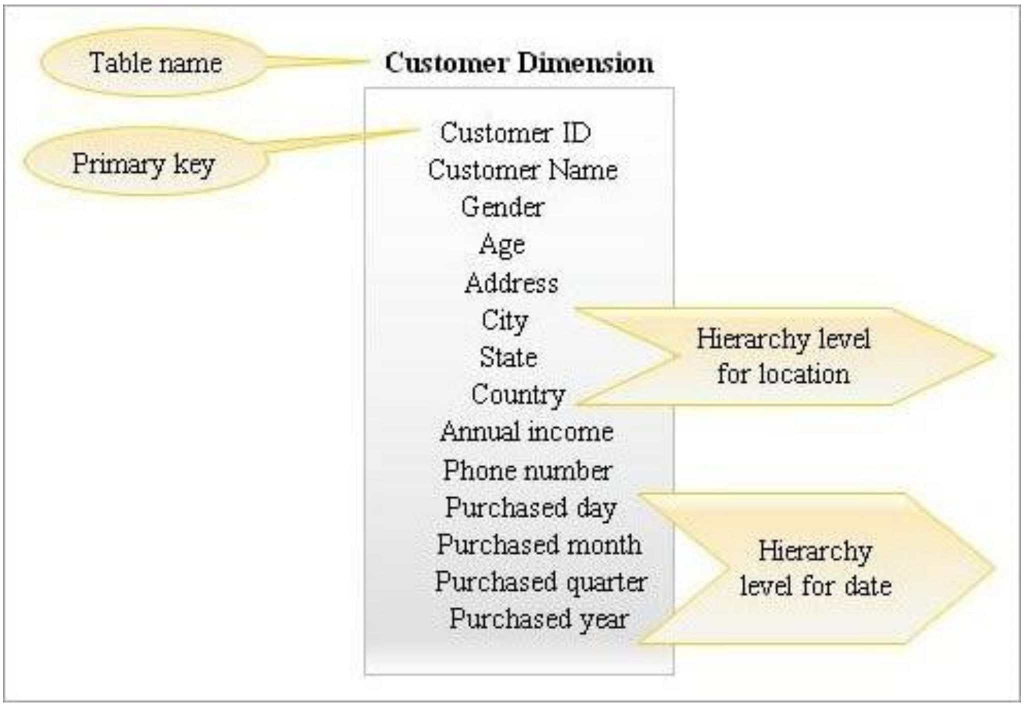
Order Details

Order Items

Departments

Customers (etc).

The Dimensional Model for a customer conducting an E-Commerce transaction is shown below:



4) Identify the Facts

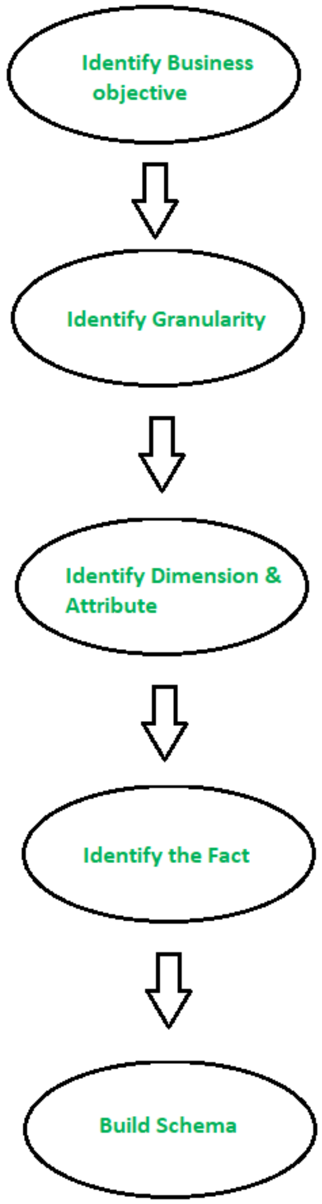
Once the Dimensions are created, the measures/transactions are supposed to be linked with the associated Dimensions. The Fact Tables hold measures and are linked to Dimensions via foreign keys. Usually, Facts contain fewer columns and huge rows.

For example, in an E-Commerce use case, one of the Fact Tables can be of orders, which holds the products' daily ordered quantity. Facts may contain more than one foreign key to build relationships with different Dimensions.

5) Build the Schema

The next step is to tie Dimensions and Facts into the Schema. Schemas are the table structure, and they align the tables within the database.

The steps to design a Dimensional Data Model are given in a visual manner in the figure below:



https://www.geeksforgeeks.org/dimensional-data-modeling/

Benefits of Dimensional Data Modelling

As now you understand the process of Dimensional Data Modelling, you can imagine why it is so important and how many benefits DDM provides for the company. Some of those benefits are given below:

The Dimension Table stores the history information and a standard Dimension Table holds good quality data and allows easy access across the business.

You can introduce new Dimensions without affecting other Dimensions and Facts in the Schema.

Dimension and Fact Tables are easier to read and understand as compared to a normal table.

Dimensional Models are built based on business terms, and hence it is quite understandable by the business.

Dimensional Data Modelling in a Data Warehouse creates a Schema which is optimised for high performance. It means fewer joins between tables and it also helps with minimised data redundancy.

The Dimensional Data Model also helps to boost query performance. It is more denormalized; therefore, it is optimized for querying.

Dimensional Data Models can comfortably accommodate the change. Dimension Tables can have more columns added to them without affecting existing Business Intelligence applications using these tables.

Limitations of Dimensional Data Modelling

Although Dimensional Data Modelling is very crucial to any organisation, it has a few limitations that companies need to take care of when incorporating the concept into their applications. Some of those limitations are given below:

Designing and creating Schemas require domain knowledge about the data.

To maintain the integrity of Facts and Dimensions, loading the Data Warehouses with a record from various operational systems is complicated.

It is severe to modify the Data Warehouse operations if the organisation adopts the Dimensional technique and changes the method in which they do business.

Despite these limitations, the DDM technique has proved to be one of the simplest and efficient techniques to handle data in Data Warehouses till date.

Reference

https://hevodata.com/learn/dimensional-data-modelling/