

Dimensional Data Modeling :

I have written the articles on [data modeling in OBIEE](#) which will give you the information about the data modeling techniques used in OBIEE. This article will give you idea about the Dimensional Data Modeling in Business Intelligence. Dimensional Data Modeling is nothing but database design technique which is useful to fetch different types of complex sql queries in DWH environment. It provides set of methods and concepts useful in DWH design. Dimensional data model is most often used in data warehousing systems. The Dimensional Data modeling is totally different from Online transactional systems (OLTP). In this complex world of data, Dimensional Data Modeling is very important to simple and precise data navigation and database performance as well. Data models are very much useful when there are very complex BI requirements.

What you will find in this article?

1. What is Dimensional Data Modeling?
2. Advantages of Dimensional Data Model
3. Multiple Real life examples.

Advantages of Dimensional Data Modeling :

Let us start with advantages of dimensional data modeling,

1. Handle Complexity of Data :

The data modeling is important to handle the [complex data](#) requirements. In big organizations the data is in complex fashion. If you want to handle that complex data in simple way data modeling is important. User need to make the data model by considering the relationships between tables.

2. Quick data navigation :

The Data modeling is important for Quick data navigation.

3. Performance of Database :

Dimensional modeling creates the optimized schema which will improve the database performance. This means fewer joins, minimized data redundancy, and operations on numbers instead of text which is almost always a more efficient use of CPU and memory.

4. Flexible:

The dimensional models are flexible and if there is a business change it will not affect the data model. Dimensional models are adaptable to change. It's as simple as adding a new column and creating a new table.

These are advantages of dimensional data modeling. Dimensional Modeling always uses facts and dimensions tables. Facts are the tables which contains numerical value which can be aggregated and analyzed on fact values. Dimension defines hierarchies and description of fact values. When it comes to dimensional modeling, fact tables, dimension tables, star schemas, and foreign and primary keys are important to understand.

Steps Of Dimensional Data Modeling :

I will explain the steps of Dimensional Data Modeling with example.

Step 1 : Identify the Dimension tables in the Business Process.

The very first step of Dimensional Data Modeling is to identify the dimension tables of the Business process. So there is a question in mind that how to identify the dimension tables.

Dimension Table :

Dimension table is table which describes the business entities of an enterprise which describes the objects in a fact table. Dimension table has primary key which uniquely identifies each dimension row. Dimension table is sometimes called as lookup or reference table. The primary key of dimension table is used to associate relationship between fact table which contains foreign key. Dimension tables are normally in de-normalized form because these tables are only used to analyse the data and not used to execute transactions.

The fields in a dimension table is used to complete following 3 important requirement :

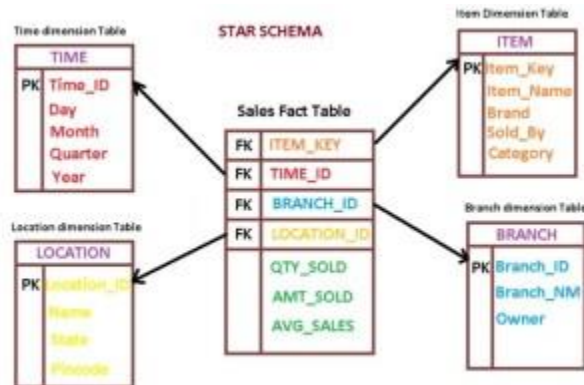
1. Query Constraining
2. Grouping /Filtering
3. Report labeling

Real Life Example :

Consider following table which contains item information. In the following table ITEM KEY is primary key which uniquely identifies the rows in the dimension table. ITEM KEY will be present in Fact table.

ITEM KEY	ITEM NAME	BRAND	SOLD BY	Category
00001	Yellow shirt	Yazaki	Amit	Shirts
00002	Football	Start sports	Rahul	Sports
00003	Blue Shorts	Puma	Amit	Shorts

In the image i have explained which are fact and which are dimension tables. You will able to see there are four dimensions :



- 1.Time
- 2.Location
- 3.Item
- 4.Branch

Step 2 : Choosing correct fact tables with its granularity for dimensional data modeling

This will depend on the total volume of transaction-level data. It's usually a good idea to start with the finest grain of data and store each sales transaction line item.

Fact Tables :

Fact table is central table found in star schema or snowflakes schema which is surrounded by dimension tables. Fact table contains numeric values that are known as measurements. Fact table has two types of columns:

- 1.Facts
- 2.Foreign key of dimension tables.

The measures in a fact table are of three types :

1.Additive :

Measures that can be added across any dimension

2.Non-additive:

Measures that can not be added across any dimension

3.Semi-additive:

Measures that can be added across some dimensions.

Real Example :

Following is a fact table which contains all the primary keys of dimensions table and added measures for ITEM,i.e.Product sold.

ITEM KEY	Time key	Product key	Date key	Product Sold
00001	T001	P001	D001	100
00002	T002	P002	D002	30
00003	T003	P003	D003	15

The fact table contains the foreign keys,time dimensions,product dimension,customer dimension,measurement values.Following are some examples of common facts :

No of unit sold,Margin,Sales revenue and the dimension tables are customer,time and product e.t.c. which is used to analyse data.

The functional difference between facts and dimensions are Fact table holds data we want to analyse and dimension table holds the information required to allow us to query it.

Dimensional data modeling basically includes combination of facts and dimension tables.But there are some situations where user needs to work with the hierarchy ,Aggregate table concept.Adding aggregate table will improve the performance of the database.

Step 3 : Check for use of Aggregate tables in the model.

Aggregate table contains aggregated data which can be calculated by using different aggregated functions like count,avg,min,max.e.t.c.Aggregated tables are most widely used tables in OLAP database.[Aggregate functions](#) are functions where the values of table or column are grouped together and form a single value. Following are some aggregate functions:

1.Average

2.Count

3.MAX

4.MIN

5.Median

6.SUM

Using the above aggregate functions the data will be inserted in aggregate table. The aggregate tables are used for performance optimization and data is coming fast using aggregate table. Aggregations applied on database level improves the performance of the query as the query not hits directly on table it will hit on aggregate table and fetches data.

Real Example:

If table contains the data of year 2016 and 2017 ,User wants a actual count of records in the table monthly,Quarterly and yearly. Then We need to make aggregate table which contains count of records monthly, For Quarterly we need to create other table and push the quarterly count in that table. And we need to use that tables in the reports so that report performance will improve drastically.

Step 4 : Use of Hierarchy at right place

Using of hierarchy at right place is always important. The specification of levels that represents relationship between different attributes within a dimension.

Real Example :

one possible hierarchy in the Time dimension is Year → Quarter → Month → Week → Day.

Reference : <https://www.complexsql.com/dimensional-data-modeling-dimensional-modeling/>