Dimensional Data Model

Shahin Mohammadkhani

06/15/2012

Professor Julie Mannering

Table of Contents

Abstract 2

Introduction 3

Concepts of Dimensional Data Modeling 4

Major Concepts 4

Importance of Data Modeling 5

Key Benefits 5

Kimball's Theory 8

Inmon Theory 8

Dimensional Data Modeling in Business Intelligence 9

Conclusion 10

Bibliography 11

# 

# Abstract

The purpose of this paper is to discuss the concepts of dimensional data models and theories such as Kimball's and Inmon. The paper will discuss the major concepts of dimensional data modeling as well as the key benefits. The paper will also entail the dimensional modeling process such as facts, star schemas, cubes, etc. As data warehousing and the use of business intelligence tools are becoming higher in demand in industries, the paper will touch base on the importance of a data model in an organization and the business values it can bring. Moreover, the benefits and disadvantages of dimensional modeling will be also discussed as it can be against the concepts of the traditional database design techniques.

# Introduction

In today’s organization, every company is facing with the challenge of the quantity of data being captured on many levels. However, with advancements in technology, the issue of how to store this growing volume of data becomes less of a concern than simply how to effectively analyze it. Organizations gather large amount of data, but sometimes they ask themselves how can their users access them. The ability to understand the data in the database can prove a huge advantage to user. However, for many users the ability to create situations to analyze data is limited by the complexity of the underlying data model. Many users either cannot or simple do not want to understand the complexities and discourages users from exploring further solution. One solution to this problem is to encapsulate a relational schema within business definition, which is understood by users and can be change successfully by the model.

# Concepts of Dimensional Data Modeling

## Major Concepts

A data model is a conceptual representation of the data structures that are required by a database. The data structures include the data objects, the associations between data objects and the rules, which govern operations on the objects. As the name data modeling implies, the data model focuses on what data is required and how it should be organized. We are modeling the data and not concentration on what operations the data will perform. A good way to understand what a data model is, is to compare it to an architect’s building plans. Moreover, contrary to many people’s belief, a data model is independent of any software or hardware. Furthermore, instead of representing the data in a way that a database would understand it, the data is presented in a fashion where the user can understand it. The data model presents “real world” data. In fact, the data model is served as the translator, or the bridge between the real work concepts and those represented in a database. (12ht)

## Importance of Data Modeling

In today’s time the concept and practice of data modeling is becoming more and more important. It is become the crucial part of database development. Furthermore, data modeling is arguably the most time consuming and labor intensive part of the development process. Many ask themselves ”so why bother with it if it take so long and we are pressed for time?” Well the simple answer is and always will be that building a database without a data model is like building a corporate structure with absolutely no blueprint.

In addition, because the data model uses natural language that can be understood and reviewed by users, its goal is to make sure that all the data objects are accurately represented in the database.

The data model is also important because its details allow developers to use it for building a physical database. Moreover, the information contained in the data model can be sued to define the relational tables, stored procedures and triggers. Without a data model, a database can be poorly designed and a poorly designed database will require more time to maintain and scale in the long run. Also, without proper planning, we may develop a database that can oversee data can is required for critical reports and is unable to accommodate changes in the future as business and requirements change.

## Key Benefits

We’ve discussed why is dimensional modeling important now lets see what are the differences and benefits of a dimensional model over a traditional E-R diagram. To start, dimensional modeling is different from the OLTP normalized modeling to enable analysis and query through massive and unpredicted queries. Something, which a relational model is, not equipped to handle. (Dimensional Modeling VS. Relational Modeling) Moreover, the dimensional model has many important data warehouse advantages that the ER model simply lacks.

The first advantage of the dimensional model is that all dimensions can be seen as a symmetrically equally entry points into the fact table. Also, the user interfaces are symmetrical, the query strategies are symmetrical and the SQL generated is symmetrical. In other ways, there will never be attributes in fact tables and facts in dimensional tables. (Dimensional Modeling VS. Relational Modeling)

The second benefit of the dimensional model is that is smoothly extensible to accommodate unexpected new data elements and new design decisions.  First, all existing tables (both fact and dimension) can be changed in place by simply adding new data rows in the table. Data should not have to be reloaded. Typically, No query tool OR reporting tool needs to be reprogrammed to accommodate the change. All old applications continue to run without yielding different results. You can, respectively, make the following graceful changes to the design after the data warehouse is up and running by:

* Adding new unanticipated facts (that is, new additive numeric fields in the fact table), as long as they are consistent with the fundamental grain of the existing fact table.
* Adding completely new dimensions, as long as there is a single value of that dimension defined for each existing fact record
* Adding new, unanticipated dimensional attributes.
* Breaking existing dimension records down to a lower level of granularity from a certain point in time forward. (Dimensional Modeling VS. Relational Modeling)

Finally out of many more advantages, the third benefit of dimensional model is that there is a body of standard approaches for handling common modeling situations in the business world. Each of these situations has a well-understood set of alternatives that can be specifically programmed in report writers, query tools, and other user interfaces. These modeling situations include:

* **Slowly changing dimensions**, where a 'constant' dimension such as Product OR Customer actually evolves slowly and asynchronously. Dimensional modeling provides specific techniques for handling slowly changing dimensions depending on the business environment.
* **Heterogeneous products**, where a business such as a bank needs to:
  + Track a number of different lines of business together within a single common set of attributes and facts, but at the same time.
  + It needs to describe and measure the individual lines of business in highly idiosyncratic ways using incompatible measures. (Dimensional Modeling VS. Relational Modeling)

With all these benefits and advantages in the general scheme of dimensional modeling, there are still different theories that people would like to approach.

## Kimball's Theory

Kimball was the first to introduce the concept of dimensional modeling. This bridged he difference between relational database and dimensional data. He designed the data warehouse from (bottom-up) fashion by connecting the data marts with a bus structure. (Kimball, 2002) This structure provides the flexibility to place data marts on different servers across the enterprise while data warehouse existence can be considered as virtual. Furthermore, in the bottom-up approach data is extracted from the existing legacy systems and then consolidated and verified in staging area. The data is fed in to data store and then more data is added or updated. As the data store contains the fresh copy of data it is integrated and transformed to the data mart structure. The data is than aggregated, summarized and available for end user for analysis and strategic decision-making. (Kimball, 2002) This theory expands to a more detailed level, but if we want to summarize Kimball’s theory, he simply says let everybody build what and when he or she wants when they want it, we’ll integrate it all when and if we need to.

## Inmon Theory

Bill Inmon’s approach is considered the (top-down) approach is not really tailored for single departments, which have different standards and integration techniques. The Inmon method is designed for a enterprise data warehouse. He proposed and identified a need to integrate data from various systems to a centralized system where the transaction processing system can be used for strategic decision-making. Furthermore according to Inmon data should be systemized as subject-oriented, integrated, time-variant and non-volatile structures. The data should be attainable at grain level by drilling down or at summarized level by rolling up. The Inmon data warehouse design pattern is dependent on third normalization form, which can afford opportunity for the granularity of data that provides maximum flexibility to the enterprise. (Inmon, 2005) Inmon further mentions that design in the top-down approach begins with the consideration of data extraction from the operations data sources. The data is then feed in to the staging area and cleansing is performed. Data is then transformed, integrated and consolidated and then transferred to operational data store (ODS). This data is then loaded into the data marts, which becomes available to end-users. (Inmon, 2005) To summarize Inmon’s theory, don’t do anything until you have designed everything.

# Dimensional Data Modeling in Business Intelligence

In the business world, the survival of a company relies on how quick it is able to recognize changing business waves and dynamics and respond correctly and swiftly to the change. Furthermore, a company should anticipate and foresee changes and trends and identifies new opportunities to stay ahead of the competition.

We are slowly moving away from an industrial era and have embrace an era of data. Companies’ assets now are data and they have even more access to data than before. They need to be able to transform this data to understandable and actionable information, by storing, consolidating, analyzing and reporting them where it’s easy to access. This is an advantage but also a challenge. These actions and tasks are all the role and goal of business intelligence (BI). BI helps organizations develop knowledge from data to enable better decision making and convert them into action plans.

Business Intelligence can help with the critical issues of a company, such as finding areas with the best growth opportunities, understanding competition, discovering the major profit and loss areas, recognizing trends in customer behavior, determining their key performance indicators, and changing business processes to increase productivity. Ballard mentions that BI analyzes historical business data that is created by business or derived from external sources, such as climatic conditions and demographic data, to study a particular function or line of business. Information is used to understand business trends, strengths, weaknesses, and to analyze competitors

and the market situation. (Chuck Ballard, 2006) I believe that prior to the existence of business intelligence technologies companies used standard conventional methods to transform data into actionable information such as using spreadsheets and charts. This consumed enormous amount of resources and time, not to mention human error factors. However, BI has allowed companies to generate reports based on a centralized data storage, which reduces or eliminated user data entry.

# Conclusion

Organizations know that they have rapidly expanding data. The challenge for the users of the corporation is to transform the knowledge instate the data into competitive advantages. Information is power and the availability to understand the data can make a huge difference. Nevertheless, the process can only take place if users can access data in the correct manner and format. The dimensional model is intended to simplify the structure, requirements change and analysis process for end users. The dimensional data model end process also makes it easier for the interaction of users and the database to be easier and is done using business oriented language and terminology they use in the workplace. I think dimensional data modeling is a tool that has helped and will help organizations become better decision makers and take better action plans. I personally have developed in Pentaho, which is business intelligence too. The new reports and dashboards that our sales department and management is now able to generate and view has allows the company to reach the next level and meet is quarterly goals. Traditional database modeling is slowly becoming obsolete and dimensional data modeling include multi dimensional data modeling is because the hot new way of approaching data.

# Bibliography

(n.d.). Retrieved 06 12, 2012, from http://www.liberty.edu/media/1414/[6330]ERDDataModeling.pdf

*Advantages of Dimensional Model.* (n.d.). Retrieved 6 1, 2012, from http://www.sys-seminar.com/EE/Files/dimensional\_data\_modeling.pdf

Breslin, M. (2004). Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models . *Business Intelligence Journal* , 20.

*Dimensional Modeling VS. Relational Modeling.* (n.d.). Retrieved 06 10, 2012, from Execution - MIH: http://www.executionmih.com/data-warehouse/dimensional-relational-modeling.php

Inmon, W. (2005). *Building the Data Warehouse, Third Edition: John Wiley & Sons.*

Kimball, R. &. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimenstional Modeling.*

Ponniah, P. (2001). *Data Warehouse Fundamentals: A comprehensive guid for IT Professionals.*

Sen, A. &. (2005). A comparison of Data Warehousing Methodologies. *Communications of ACM* .

Stephen J. Smith, A. B. (1997, 12 1). *Components of a Data Warehouse* . Retrieved 05 12, 2012, from www.tdan.com: http://www.tdan.com/view-articles/4213