International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

A Conceptual Model for Multidimensional Data Intended for Decision-making in a Health/Medical Structure

Lobo Minga Bertin

Abstract: ***The purpose of this article is to develop a multidimensional data model that is most appropriate for creating a data warehouse which will be helpful for analysis in taking on medical structure. In today's information systems, the needs of decision-makers require that data be organized around themes or subject of analysis to date and that their formats be standardized despite several sources of data. Hence, the need for multidimensional modeling. The article focuses on the production of the conceptual model, while considering that this is an essential step to the continuity of the other steps of the decision-making process very easily. The choice of the approach used is attached to the users' needs in terms of analysis based on the data saved during a period. The result obtained and presented in this paper is a model of multidimensional data in constellation. Finally, the conceptual model of multidimensional data is behind justified in relation to the expectations of decision makers and other types of existing models in the decision-making systems.***

Keywords: olap, etoile, flocon, constellation, entrepot des données, cube, multidimensionnel, fait, dimension

1. Introduction

Decision-making or decision-making systems in the sense of data warehouses were born of a need expressed by companies that was not satisfied with traditional database systems. By integrating data warehouses, the decision-making process responds to the problem of the continuous growth of data that may be of different formats. In addition, it effectively supports On-Line Analytical Processing (OLAP) processes (Chaudhuri and Dayal, 1997; Chaudhuri et al., 2011).

More years pass, it is noticed an increased volume of data backed up in several companies.
These data are of various kinds and constitute a true source of decision-making for the managers of the companies. There is a problem related to the standardization of different data formats, especially when they come from different sources.

The storage of these data also determines the time and quality of the expected answers. Data stored in disorder without any structuring does not facilitate their interrogation, sometimes occupy space unnecessarily and sometimes pose problems related to their updating.

Relational database systems have addressed many of the data storage and query concerns, but the decision-makers' decision-making needs were still not met by the transactional system that was already in place. The idea for decision-makers was to have more efficient and fast access to the data saved. For this, the information storage structure should be reorganized around the themes to be analyzed. Hence the use of a multidimensional structure.

Multidimensional models aim to provide intuitive and high-performance data access. To do this, the data are organized around the facts we are trying to analyze, characterized by indicators (called measurements), which are normally numerical and additive data, making it possible to measure the modeled activity. These facts are described by a set of axes of analysis, or dimensions, hence the term multidimensional model. This basic model corresponds to the star model (Kimball et al., 2000, Chaudhuri and Dayal, 1997).

In this article, we will start from the needs in terms of analyzing information in a medical structure to come up with a suitable conceptual model of multidimensional data.

This proposed model will serve as a necessary basis for the creation of a data warehouse to be integrated into a decision-making process. This model will only be an abstract representation of a reality that will have to be implemented by referring to it.

With such an organization of information, the medical structure will always be reactive to the demands of its partners and will thus face any eventual competition.

Globalization and the competition that it creates make the management of an organization increasingly complex. This complexity is linked not only to the increase in the number of parameters to be taken into account but also to the need for rapid decision-making in order to be responsive to changes in competition and customer demand. The effectiveness of these decisions depends on the availability of reliable, relevant information and tools to facilitate this task. Traditional systems, dedicated to the day-to-day management of an organization, are unsuited to such an activity [Codd et al. , 1993; Inmon, 1996; Kimball & Ross, 2002]. Faced with this need, the business intelligence sector was born.

The article is organized as follows: First, to recall the basic concepts and theories necessary for the understanding of decision-making systems. Then, present the analytical needs that arise in a medical structure. And finally, use a technique to arrive at creating the conceptual model of the appropriate multidimensional data.

Licensed Under Creative Commons Attribution CC BY

2. Methodology

Multidimensional conceptual models are generally based on the relationship entity model and the object model. They integrate decision-making concepts and adopt a formalism independent of implementation.

To date, researchers recognize three types of approach to developing a decision-making information system. It is:

- The top-down approach that defines the conceptual framework based on the needs of users of the decision-making information system
- The bottom-up approach which defines the conceptual schema from the data source schemas
- The mixed approach that combines the two previous approaches

As part of this article, we will construct our schema from the needs of the users while not neglecting the structure of the data sources. We will therefore use the mixed approach.

Three techniques are available to us to build the conceptual model of multidimensional data. It is:

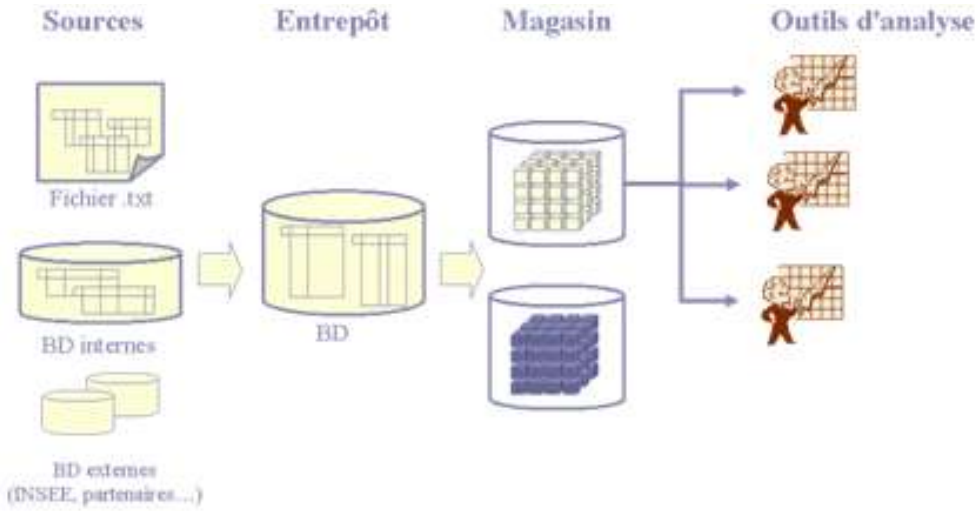
- The Top-down technique which consists of the realization of the entire warehouse as a whole
- The bottom-up technique of building pieces of the warehouse that will have to be unified to give rise to a whole after all, of course eliminating possible redundancies.
- The Middle-Out technique which is a hybrid technique.

Taking into account the simplicity offered, the bottom-up technique will be that which will be retained within the framework of this article.

3. Definition of Some Concepts

3.1 Decision-making project

Unlike some processes based solely on the use of software tools, a decision-making process is a project that is being built. It must be part of a framework that can take into account data, information and knowledge. The data warehousing approach is an important field of research in which many problems remain to be solved. Data warehouses are usually integrated into a decision support system where two storage spaces are distinguished: the data warehouse and the data warehouses. An architecture of the decision-making process is shown in Figure 1 (Bentayeb et al., 2009).



3.2 Data Warehouse

Inmon (1994) defines the data warehouse as "a collection of integrated, subject-oriented, non-volatile, historical, summarized and available data for interrogation and analysis."

3.3 Data Cubes

The multidimensional model makes it possible to represent the data in relation to the dimensions or axes representing essential elements of the activity of a company. Three levels of representation of the data are defined in the decision-making process: the warehouse that gathers data that is transversal to all the company's businesses, the data store which is a vertical representation of the data relating to a particular trade, and finally the cube of data (or hyper cube).

The cube corresponds to a business view where the analyst chooses the measurements to be observed according to certain dimensions. A cube is a collection of aggregated and consolidated data to summarize information and explain the relevance of an observation. The data cube is explored using many operations that allow it to be manipulated. The cube of the data is therefore based on a data warehouse, which consists of tables of facts and dimensions.

3.3 Facts and Dimensions

The analyzed subject, ie the fact, is analyzed from different perspectives. These outcomes correspond to a category used to characterize the activity measures analyzed [Marcel 1998]; we talk about dimensions.

3.3.1 Flake, constellation and star models

From the fact and the dimensions, it is possible to establish a simple data structure that corresponds to the need for

Licensed Under Creative Commons Attribution CC BY

multidimensional modeling. This structure consists of the central fact and the dimensions. This model represents visually a star, we speak of star model [Kimball 1996].

3.3.2 Overview of analysis needs

The analytical needs mainly concern the prescriptions made by the doctors and the deliveries of the products made at the pharmacy of the said medical structure.

The concern for the medical structure manager is how to reorganize the data structure to:

- Analysis of the quantities of products prescribed by doctors by reference to geographical seasons and to the types of products or medicines
- Analysis of the quantities of products delivered by the pharmacy by referring to the types of products
- Analysis of the quantities of products purchased from suppliers by reference to the types of products

These various analyzes are necessary in order to improve the services provided to patients. It will be necessary, for example, to judge what product should not be missed during such a period of time as a result of prior analyzes carried out on historical and aggregated data concerning the prescriptions and deliveries made.

3.3.3 Construction of the conceptual model of multidimensional data

From the illustration of the previous needs, there are three subjects or themes to be analyzed: prescriptions, deliveries and purchases from suppliers. These subjects should be analyzed in relation to two axes or dimensions, namely: products and seasons.

For simplicity, we use the Bottom-up technique proposed by Kimball without neglecting the constraints related to the structure of the data sources.

Considering the prescriptions as the first subject of analysis, there are two dimensions in relation to which they must be analyzed namely: the season (rain or dry for example) and the product. This gives rise to a star model with a fact (prescription) and two dimensions (season and product).

Considering the deliveries as the second subject of analysis, there is a dimension in relation to which they must be analyzed, namely: the product. This gives rise to a star model with a fact (delivery) and a dimension (product).

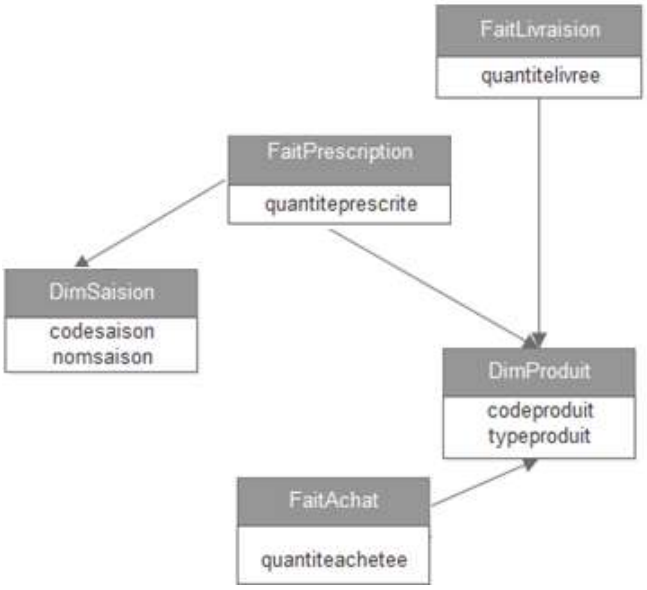
Considering finally the purchases made by the medical structure as the third subject of study, there is a dimension in relation to which they can be analyzed, namely: the product. This gives rise to a star model with a fact (purchasing) and a dimension (product).

From all of the above, it should be noted that there are three different models to build the data warehouse.

Since these three star models all have a common dimension, they should be unified by eliminating redundancy. The unification of these three star models will finally give rise to the constellation model with three facts (prescription, delivery and purchase) and two dimensions (season and product). This way of organizing the data makes it possible to do several synthesis operations on the basis of several axes of analysis very easily and quickly.

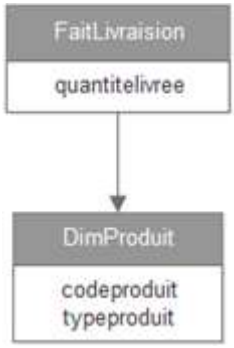
4. Results

After studying the needs and applying the mixed approach to building a decision-making information system, the proposed model is that in constellation, the diagram of which is shown below:

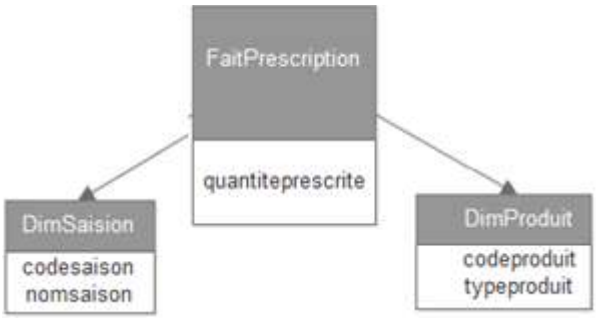


This model was obtained from the fusion of the models below:

First star model

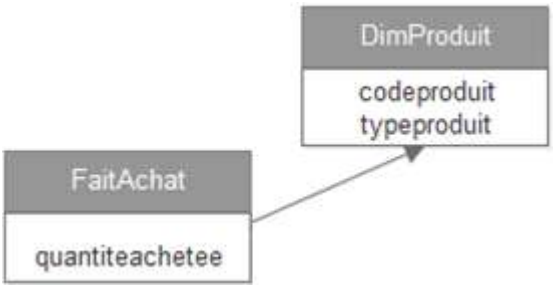


Second star model



Licensed Under Creative Commons Attribution CC BY

Third star model



5. Discussions

Among all the models indicated by the researchers to represent the multidimensional data at the conceptual level, there are no good or bad models, comparing them. Only, there are some who are more suited to this or that situation or circumstance. In our case, the study involved several facts that had common dimensions, so the model that was appropriate for data storage was the constellation model.

On the other hand, if it were a single fact, the star-shaped model would be well suited. For the sake of simplicity, we did not want to explode the dimensions to deal with different levels of hierarchy within the dimensions.

When these dimensions are broken down hierarchically, the model shown is that in flake. The latter is therefore not appropriate for the present case and could be indicated only to the extent that the needs would require a hierarchy within the dimensions. It should be noted, however, that flake modeling induces a denormalization of the dimensions generating greater complexity in terms of readability and management.

6. Conclusion

The main objective of this article was to propose a model of the multi-dimensional data needed to set up a data warehouse within a medical structure. After an analysis of the dimensions and facts, the chosen model is that in Constellation.

After recalling the founding theories and concepts of decision-making systems, we used the mixed approach and the middle out technique to build our model based on the needs expressed by the manager of the medical structure.

The type of model proposed makes it possible to fill in the shortcomings of transactional systems in that it makes it possible to organize the data around the themes or subjects to be analyzed. A poor conceptual model of the proposed data has an adverse effect on the expected quality of the analyzes performed on the data cubes.

Overall, the proposed solution remains appropriate and open to improvements depending on the context and the changes that could occur in time and space.

References

[1] Bentayeb, F., O. Boussaid, C. Favre, F. Ravat, et O. Teste (2009). Personnalisation dans les entrepôts de données : bilan et perspectives. In *5èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2009), Montpellier*, Volume B-5 of *RNTI*, Toulouse, pp. 7–22. Cépaduès.

[2] Chaudhuri, S. et U. Dayal (1997). An Overview of Data Warehousing and OLAP Technology. SIGMOD Rec. 26(1), 65–74.

[3] [Codd et al., 1993] E.F. Codd, S.B. Codd, C.T. Salley, « *Providing OLAP to user analyst : an IT mandate* », Rapport technique, E.F. Codd and associates, 1993.

[4] [Inmon 1994] Inmon W.H., "*Building the Data Warehouse*", John Wiley&Sons, ISBN 0471- 14161-5, 1994.

[5] [Inmon, 1996] W.H. Inmon, « *Building the Data Warehouse* », John Wiley and Sons, New York, NY, deuxième edition, ISBN: 04771-14161-5, 1996.

[6] [Kimball 1996] Kimball R., "*The data warehouse toolkit*", John Wiley and Sons, 1996.

[7] [Kimball & Ross, 2002] R. Kimball, M. Ross, « *The Data Warehouse Toolkit : the complete guide to dimensional modeling* », (deuxieme edition), W iley, 2002.

[8] [Marcel 1998] Marcel P. "*Manipulation de Données Multidimensionnelles et Langages de Règles*", Thèse de Doctorat de l'Institut des Sciences Appliquées de Lyon, 1998

Licensed Under Creative Commons Attribution CC BY