

Business Metadata: Linking Enterprise Goals with Data Warehouse

Ruchi Thakur¹, Anika Gupta²

^{1,2}Department of Computer Science and Engineering, Punjab Institute of Technology Kapurthala, Jalandhar, India

Abstract: Business Organizations use Data Warehouses (DWHs) to analyze their performance. Performance is judged in regard to the goal achievement. Data warehouses and OLAP tools are based on a multidimensional data model. This ideal views data in the form of a data cube. Enterprises often model their goals in terms of formal or semiformal goal models. The problem is that: the Data Warehouse and the Enterprise Goals - are described independently and are not related to each other. We identify a need for combining the two aspects. If their relation is made explicit, it can be used to encourage the way user access and interpret data in the Datawarehouse. To address this limitation, this paper introduces a linking model between enterprise goals and DWH data. Here met models for both aspects as well as the weaving links between them described, which helps to demonstrate the aspects separately but also in combination. It also includes benefits, needs, and a real time applications of business metadata. Also it is demonstrated to how to use the weaving links to create business metadata. Business metadata can be used in the DWH to define the business situations and inferences of the data to the users.

Keywords: Datawarehouse, Metadata, ATLAS, UML, EDM

1. Introduction

A Data Warehouse is an integrated collection of databases rather than a single database. Data Warehouse (DWH) systems represent a single source of information that includes data selection, integration and organization methods to make data easily accessible to the decision making process. Analysts and decision makers examine this information with regard to the goals and the policies of the organization and use that information to identify warning signs and trends, and to decide on future funds. Datawarehouses use powerful tools for analyzing data but business metadata has not been in concern of research. It contains the data ownership information, business description, and changing policies. [2]

This information of goals and context is available in enterprise models or other repositories within the organization. But the link between the organizational information and the DWH is not easily accessible to DWH users and remains mostly unstated knowledge. If knowledge about the business context is left to possibility, data analysis is bound to blunder important points and large part of the potential of the DWH are washed out

There exist many domain-specific modeling approaches that describe the data structures on the conceptual, logical and physical level. Enterprises utilizing a Datawarehouse are aware of these models and are dependent on them to design and describe their DWH.

To describe and examine their objectives and goals, enterprises use different types of goals models. For the goals, metrics are derived to measure the accomplishment of the goals. There exist many modeling approaches for enterprise goals as well as for metrics.

The problem we addressed here is that - the Data Warehouse and the Enterprise Goals - are described separately and are not related to each other. We spot a need for combining

these two aspects. If their relationship is made open, it can be used to develop the way user's access and interpret data in the data warehouse.

The objective of the paper is to address these limitations by

- Making the relationship visible and accessible by showing how the enterprise goals and metrics are mirrored in the DWH data model
- Using this material to support and improve data interpretation
- Eclipse is used as platform for prototypes. ATLAS is used as a weaving model that is used to link enterprise goal with datawarehouse.

2. Need of Analyzing business

The question is how data warehouse can be used for analysis and improvement, possible reason is the lack of methods that can guide business process improvement using warehouse.

Possible highlighted problems can be addressed by taking into account the needs for business process analysis and business performance in accordance to goals. Therefore approach to address the issue is based on business orientation in design and utilization of warehouse. The methods of goal driven approach involves three steps: Build a goal structure, integrate goals into process warehouse, analyze and improve business process.[3]

3. Multidimensional Data Models

The entity-relationship data model is commonly used in the aim of relational databases, where a database schema consists of a set of entity and the associations between them. Such a data model is appropriate for on-line transaction processing. A data warehouse, however, require a brief, subject-oriented schema that facilitates on-line data analysis. The most prevalent data model for a data warehouse is a multi-dimensional model. This sort of model can be in the form of star schema, snowflake schema, or fact constellation

schema. The most common modeling example is the star schema, in which the data warehouse contains (1) a big middle table (fact table) containing the bulk of the data, with no redundancy. It contains data on the subject of analysis. The attributes of the fact table are called measures. and (2) a set of smaller assistant tables (dimension tables), one for each dimension. According to which the data will be analyzed. The dimension can be planned in hierarchies that are useful to aggregate data(e.g. store, city, region, country).

For weaving enterprise goals with the Datawarehouse that, we need is a data model that supports weaving. What we choose is the object-oriented approach and then further developed to a UML profile. A UML Profile is a domain specific extension to the UML modeling language. This profile adapts the UML class dig. for multidimensional modeling, i.e. the support class of the stereotypes is Class. It allow to create complete models of the conceptual characteristics of multidimensional data models.

The snowflake structure can decrease the usefulness of browsing, since more joins will be required to execute a query. So, the system presentation may be unfavorably impacted. Hence, though the snowflake schema reduces redundancy, it is not as popular as the star schema in data warehouse design. Fact constellation schema uses multiple star schemas hence the main schema used in datawarehouse is star schema.[4]

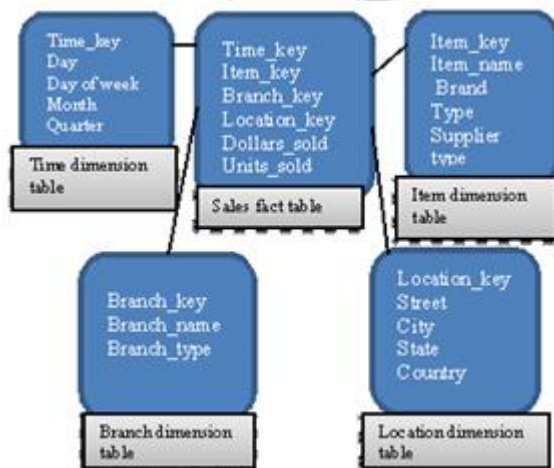


Figure 1: Star schema using uml profile

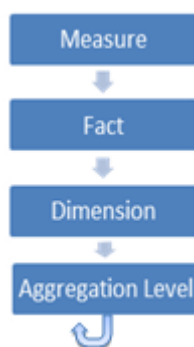


Figure 2: Core elements of the UML profile for modeling multidimensional data models

4. Enterprise Model

A. Modeling the Goals of Enterprise

Enterprise model is used to represent structure, behavior and goals of enterprise. Fig 3 shows the outline of a generic enterprise model, ordered into five aspects:, from a high-level business observation to a generic reasonable data design. It includes links to the physical data designs of individual applications. Through an EDM, one can derive the general scope and perception of the business Requirement

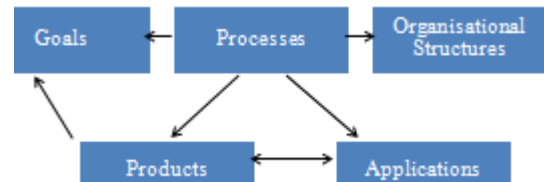


Figure 3: Generic Enterprise Model

B. Enterprise Goals

Increasing market share or reduce operational cost are typical enterprise goals. Goals form the basis for business decision and way a company does business. relevant and significant information for business performance measurement can be read honestly from the enterprise goal model. They govern the design of commerce process and the way in which the institute behaves.

Enterprise goals are long term goals that should remain stable a lot longer than business process, role definition, and operating rules. Therefore, they provide excellent metadata for a DWH.

Based on the explanation of the goals, the enterprise derives metrics that measure the rank of attainment of the goals and point to the performance of the enterprise. These metrics are not identical but are closely related to the measures in the DWH.

For enterprise goals in particular, there is often a distinction between three levels of goals: strategical (long term), tactical (monthly goals) and operational (everyday goal). A goal can be decomposed further and further and thus form a hierarchy.

C. Enterprise Goal Metamodel (EGM)

The Enterprise metamodel is a high-level abstraction of an enterprise. Though it is a metamodel, it is not only an example of a model of a real and available enterprise, e.g. a production facility, a value company or an IT service provider. It is somewhat a model of what more or less most of the models of enterprise have in common: it capture the essential concepts of a wide variety of enterprises and simplifies and abstract them to an extent which allow us to identify their common features.

Important principles of metamodeling: This article obey the listed principles, with the aim of creating a minimal, coherent model that can be extended if required.

- principle of parsimony: Be economical, start small and add
- principle of divide and conquer: separate and get the better of, avoid mega-models[9]

The model uses the notation of the UML 2.0 class diagram. goals can be seen to manipulate each other in various ways. The completion of one goal might be detrimental to another goal, or the goals may be related to each other in a way that if one among them is satisfied, it will also support the other goals. As a result, there are two influencing relationships between goals: support and conflict. Both may occur between any number of goals, e.g. a goal be capable of supporting several goals and conflict with others at the same time.[1]

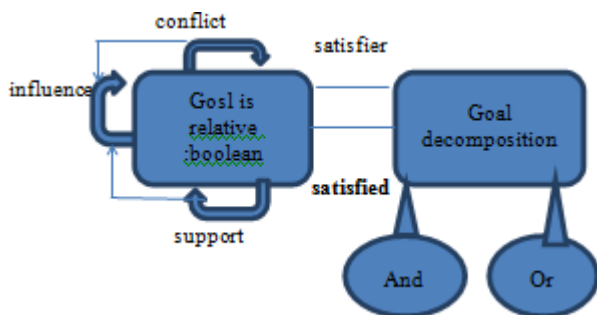


Figure 4: Metamodel for describing goal breakdown hierarchies and relations between goals

5. Common Requirement for Link Management

Four main requirements for supporting link management using weaving models:

- A weaving model should express the idea of links between model elements.
- Different links types must be supported. The link type provides the information on how the essential elements are related. For instance a tie may indicate the parity or combination of model elements.
- It must be possible to define relations with different parities unary, binary, ternary, etc., i.e., a link has countless endpoints. For example an fair link has two endpoints, while a concatenation link may have three endpoints • The weaving model should have an recognition mechanism to uniquely identify the model elements. The link endpoints do not contain the existing metamodel elements, but a pointer that enables to access them in model[5]

6. The Weaving Model

To gain business metadata we need to link enterprise goals with datawarehouse. This link acts as a weaver between datawarehouse and enterprise goals, the adjoining models are called metamodels. This weaving model describes the relationships between two or more model elements from the respective metamodels. Weaving tools like ATLAS support model weaving. The AMW (ATLAS Model Weaver) is a tool for building relationships (i.e., links) between models. The links are saved in a model known as weaving model. It is created conforming to a weaving metamodel. AMW is

part of the AMMA platform. Model weaving superficially resembles techniques used in ETL, but the motive behind it is different. Weaving link does not essentially imply that the two elements it connects should in some way be transformed from one into the other. It indicates that the two elements share some semantic link. But even than weaving can be used as a beginning step to transformation. Here it is employed for interpreting the DWH data with business metadata, and hence does not involve transformation.

Advantages of weaving model include:

- 1) Everything is a model: can capture basically *all* information in requisites of models, also the relationships and correspondences between models. Which further help in analyzing and storing data.
- 2) Weaving avoids having one large meta-model for everything but keeps individual meta models separate, easy to handle and focused on its domain

Fig 5 shows weaving model that is linking datawarehouse and enterprise goals. It consists of three links: two binary and one ternary link. The initial link is among the Parameter describing the focus or scope of the metric in the EGM and an Aggregation Level. The second, most complicated link in it the weaving model connects a Metric by means of a compute and optional Aggregation Level. The third link allows handling the relationships between the Timeframe of a metric's objective value and a Dimension containing time value in the data model.

A timeframe is a time phase indicated by begin and finish point, whereas a time dimension contains single point in time. so, the weaving link connects one timeframe with another on the instance dimensions.[1]

7. Benefits of weaving a model

- It provides goal based navigation of process warehouse, which helps to retrieve data relevant to goals
- And to augment the ways user evaluate and get better business processes.

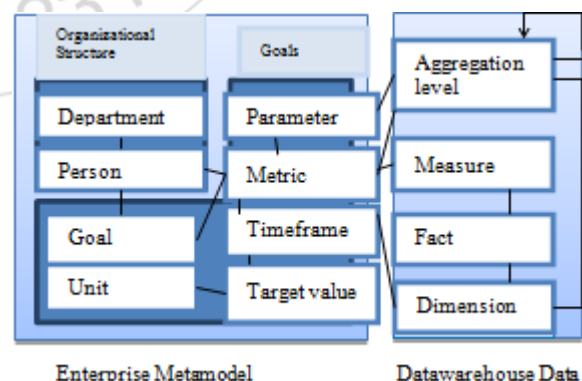


Figure 5: Weaving Model connecting the Goal Model with the Data Model

Modeling how the organization interacts with the DWH and how its structure, behavior and goals are mirrored in the DWH provides: Increased Visibility, Improved Communication. This is useful during development of a DWH, leading to Facilitated Requirements Analysis,

Requirements-driven Design and Streamlined DWH Evolution and Re- Engineering. It also supports Documentation and Maintenance. Business Metadata provides background information directly in the DWH, leading to Improved Data Interpretation as well as Enhanced Usability and User Acceptance of Gathered Data.[6]

8. Real-Time example of Business Metadata

Real-time information about businesses such as, the current occupancy and music level, as well as the kind or exact song playing now, can be important factors in the local search decision process. In this work, we suggest to automatically crowd source such rich, real time business metadata through user check-in events. BingNow 2.0 is a business metadata for real time extraction.

Local search users decide what business to visit exclusively based on generic or stale information such as distance information, and business ratings. It is believed that local search users can merely make an informed decision of where to go next if they know more detailed information about the condition of each business in the search results at the time of query. Here it is proposed to automatically crowd source such wealthy, real time business metadata during real user check-in events. Every time a user checks into a business, the receiver is in user's hands, and the phone's sensors can judge the business environment. In particular, we leverage the phone's microphone through this time to conclude the occupancy level, the music type or exact song playing, as well as the music as well as noise levels in the business. The audio data recorded through the phone's microphone capture all the different aural sources in the business (i.e, music, human chatter, noise). By properly analyzing the recorded audio in the chronological and frequency domains, we dig out a set of features that can capture the unique and finely properties of human speech and music. Then, provided labeled data traces from real business, to train engine learning models, examples include decision tree model, to forecast the occupancy, human chatter, music and noise levels in the business. As users register to businesses all over the day, this type of metadata about the businesses be able to be updated. In this way, a real-time flow of rich business metadata can be automatically extracted. Real-time business metadata can be used to understand the physical world in real-time, and enable the next generation of the local search user experience. [10]

9. Related Work

Weaving is used in aspect ratio programming. Breton and Bézivin relate weaving model to the area of workflow and procedure modeling. The build-time and run-time workflow definitions are weaved together to create a binding among definition and execution of the process. [7]

Giorgini et al. focus on DWH requirement analysis based on goals they originate the data model from the goals, which symbolize a rather narrow software engineering type of goals. In contrast, we integrate enterprise goals and line up the DWH directly with business strategy. Our weaving model is focused on the particulars of enterprise goals and their actions, rather than on all aspects of an enterprise [8]

10. Conclusion

This paper presented an approach to business metadata that is accessible by representing relation between datawarehouse and enterprise goals. Business metadata is created through weaving model. It provides data combination by explaining the application and perspective of data. This paper includes needs and benefits of weaving a model. ATLAS is used as a linking model. The models for describing the association between the DWH and the structure, behavior, and goals of the organization, to increase the visibility of this association and to improve communication by capturing this knowledge. Moreover, business metadata can be added to the DWH that informs users about the context and background of the data, in order to improve data interpretation.

References

- [1] Veronika Stefanov and Beate List. "Business Metadata for the Data Warehouse". In 10th IEEE International Enterprise Distributed Object Computing Conference Workshops, 2006
- [2] Methanias Colaço Júnior, Manoel Mendonça, Francisco Rodrigues, "Data Warehousing in an Industrial Software Development Environment". In 2009 33rd Annual IEEE Software Engineering Workshop
- [3] Witold Abramowicz, Poznań University of Economics and Business, 14th international conference, BIS 2011, Poland
- [4] Chuck Ballard, Dirk Herremans, Don Schau, Rhonda Bell, Eunsaeng Kim, Ann Valencic, "Data Modeling Techniques for Data Warehousing"
- [5] Marcos Didonet Del Fabro, Jean Bézivin, Patrick Valduriez, ATLAS Group, INRIA & LINA University of Nantes, Nantes, France
- [6] Veronika Stefanov, "Bridging the Gap between Data Warehouses and Organizations". In CAiSE'06 Doctoral Consortium
- [7] E. Breton and J. Bézivin, "Weaving Definition and Execution Aspects of Process Meta-models". In 35th Hawaii Int. Conf. on System Sciences, page 290, 2002
- [8] P. Giorgini, S. Rizzi, and M. Garzetti. Goal-oriented requirement analysis for data warehouse design. In Proceedings DOLAP '05, pages 47–56. ACM Press, 2005.
- [9] <http://www.business-prototyping.com/prototyping-the-enterprise/the-enterprise-metamodel>
- [10] <http://research.microsoft.com/enus/projects/bingnow2.0/>