Requirements Driven Data Warehouse Development

Naveen Prakash[1], Anjana Gosain[2]

[1]JIIT A10, Sector 62 NOIDA 201307, India
praknav@hotmail.com
[2] SIT, GGSIPUniversity,Delhi 110006, India
anjana_gosain@hotmail.com

Abstract. Data Warehouses are decisional information artifacts that are embedded in the organizations that create/maintain them. Therefore, Data Warehouse Requirements Engineering must start from goals and work its way to the decisional information needed to take decisions that fulfils these goals. We propose two associations, the goal-decision association and the decision-information association to represent this. These associations form part of a larger model called the goal decisional information (GDI) model.

1 Introduction

In the last decade, great interest has been shown in the development of Data Warehouses (DW). According to Inmon [5], a DW is required for management decision making and is a subject-oriented, integrated, time-variant, non-volatile collection of data. Two different approaches for the development of DWs have been proposed. These are the *data-driven* [5] and the *requirements-driven* [2] approaches. In the former, data is gathered from *operational systems* into DWs. On the other hand, *Requirements-driven approaches* try to identify the information needs to be met by the DW. In these approaches, the real issue has been that of DW design: given data needs what is the logical structure of the DW.

Jarke et al [6] propose to add a **conceptual layer** on top of the logical layer. It is assumed that the conceptual objects can be determined but the question of what are useful conceptual objects for a DW and how these are to be determined is not addressed.

To answer this, we need an explicit Requirements Engineering (RE) phase in DW development. We shall refer to this as Data Warehousing Requirements Engineering, DWRE. The broad outline of our DWRE process is as follows: In the DW domain the information contents must be closely related to the decision-making capability to be supported. Indeed, this capability should drive DW development. However, decision-making itself does not occur in a vacuum but in the larger context of the goals and objectives that an organisation sets for itself. Therefore, to obtain a DW that is a good fit for an organisation, the DWRE process determines the goals of an organisation, uses these to arrive at its decision-making needs, and identifies the information needed for the decisions to be supported.

We organize the process of DW development in four stages. Table 1 shows these stages as well as the output of each stage.

Table 1. The Stages of DW Development.

Stage	Output
Requirements Engineering	Goal and decision hierarchies respectively, decisional information, goal–decision coupling and decision information coupling.
Conceptual Design	Facts, dimensions, aggregations, historisation needs, meta-data
Logical Design	Representation in DDL of the DW package to be used
Physical Design	Data Warehouse physical layout

The aim of the Requirements Engineering (RE) phase is to identify the decisional information to be kept in the Data Warehouse. It deals with the identification of the goals of the organization, decisions that can be taken to achieve these goals and the information needed for decision making. There are two kinds of relationship (a) *is satisfied by* between goals and decisions and (b) *is required for* between decisions and information. The RE stage is completed when the goals, decisions, information and their relationships are all determined.

2 RE for Data Warehouses

Our Goal-Decision-information (GDI) model is shown in Fig.1. In accordance with goal-orientation [1,3], we view a goal as an aim or objective that is to be met. A *goal* is a passive concept and unlike an activity/process/event it cannot perform or cause any action to be performed. A goal is set, and once so defined it needs an active component to realize it. The active component is *decision*. Further to fulfil the decisions appropriate *information* is required.

As shown in Fig.1 a goal can be either simple or complex. A simple goal cannot be decomposed into simpler ones. A complex goal is built out of other goals which may themselves be simple or complex. This makes a goal hierarchy. The component goals of a complex one may be mandatory or optional.

A decision is a specification of an active component that causes goal fulfillment. It is not the active component itself: when a decision is selected for implementation then one or more actions may be performed to give effect to it. ***In other words, a decision is the intention to perform the actions that cause its implementation.*** Decision-making is an activity that results in the selection of the decision to be implemented It is while performing this activity that information to select the right decision is needed. As shown in Fig.1, a decision can be either simple or complex. A simple decision cannot be decomposed into simpler ones whereas a complex decision is built out of other simple or complex decisions. Fig.1 shows that there is an association '*is satisfied by'* between goals and decisions. This association identifies the decisions which when taken can lead to goal satisfaction.

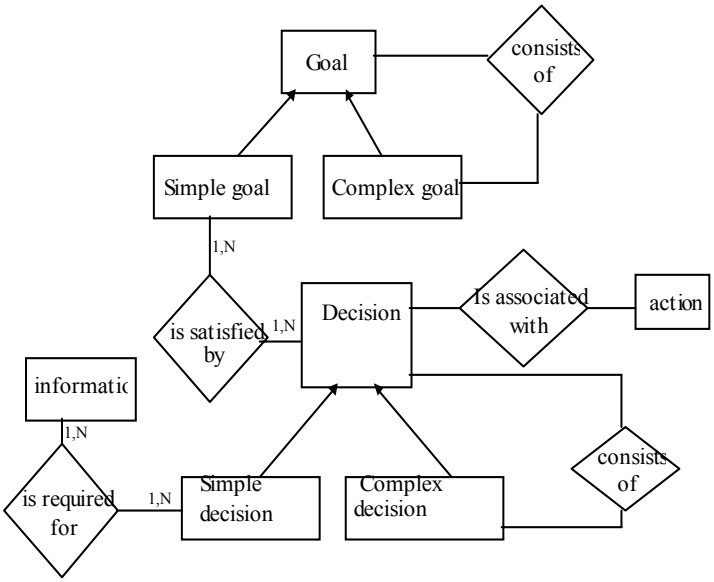

Fig 1 The Goal-Decision-Information (GDI) Model

Knowledge necessary to take decisions is captured in the notion of information shown in Fig.1. This information is a specification of the data that will eventually be stored in the Data Warehouse. Fig.1 shows that there is an association '*is required for'* between decisions and information. This association identifies the information required to take a decision.

A GDI diagram is shown in Fig.2. It shows a goal hierarchy (solid lines between '*Maximize profit'* and *'increase the no. of customers'*, and '*increase sales'*) and a decision hierarchy(solid lines between *'improve the quality of the product'* and '*introduce statistical quality control techniques'* and use '*better quality material'*) for a given set of goals and decisions. The figure shows the '*is satisfied by'* relationship between the goal '*increase sales*' and decisions '*open new sales counter'* and *'improve the quality of the product'* by dashedlines. The '*is required for'* relationship between decisions and associated information is shown by dotted lines

3 Conclusion

In Requirements Engineering goal-orientation was first introduced [1,3], and was then modified to goal-scenario coupling [4]. Goals have also been used in the field of Software Quality to associate metrics with the overall aims of the project and process. This is done through the Goal Question Metric approach. We have proposed that the goal-based approach should be extended to decisional systems through the goal-decision-information process. Goals set up the context within which decisions are taken and decisions, in turn, help in defining the information needed. Similar to

the GQM approach, our proposal is a three stage one: just as goals lead to questions and then to metrics, we start with goals and move to information through decisions.

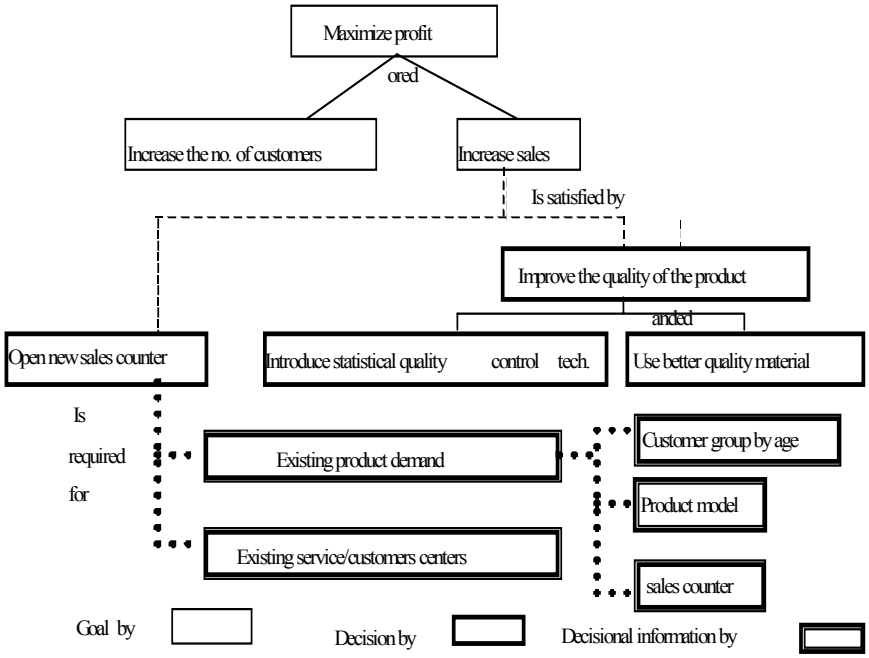

Fig. 2. The GDI Diagram

References

1. A.I. Anton, Goal based requirements analysis. Proceedings of the 2nd International Conference on Requirements Engineering ICRE'96, pp. 136-144, 1996.
2. Ballard C.,Herreman D., Schau D. , Bell R. , Kim E., Valencic A., Data Modeling Techniques for Data Warehousing, redbooks.ibm.com
3. J. Bubenko, C. Rolland, P. Loucopoulos, V De Antonellis, *Facilitating 'fuzzy to formal' requirements modelling*. IEEE 1st Conference on Requirements Engineering, ICRE'94 pp. 154-158, 1994.
4. B. Dano, H. Briand, F. Barbier, *A use case driven requirements engineering process*. Third IEEE International Symposium On Requirements Engineering RE'97, Antapolis, Maryland, IEEE Computer Society Press, 1997.
5. Inmon W.H., Building the Data Warehouse, John Wiley, New York
6. Jarke M., Jeusfeld A., Quix C., Vassiliadis P., Architecture and Quality in Data Warehouses , Proceddings 10th CAiSE Conference, Pernici B., Thanos C. (eds.), Springer, 93-113.