A method for requirements elicitation of a Data Warehouse: An example
JORGE OLIVEIRA E SÁ
Information Systems Department
University of Minho
Azurém Campus, 4800-058 Guimarães
PORTUGAL
jos@dsi.uminho.pt

Abstract: - This paper presents a method with a combination of diverse approaches for requirement elicitation of a Data Warehouse (DW) schema model. This method integrates five approaches: goal-driven, user-driven, process-driven, technology-driven, and data-driven.

This method includes two approaches not usually used in the DW field: process-driven and technology-driven. The role of each approach is: goal-driven produces subjects (DW processes) and KPI’s (Key Performance Indicators) of main business fields; user-driven produces analytical requirements represented by measures and dimensions of each subject; process-driven propose improvements in business processes (by using and creating subject oriented enterprise data model) to satisfy KPI’s, measures and dimensions identified in the previous approaches; technology-driven is an enabler or an obstacle to be considered in a DW; and data-driven is a combination of the results of previous approaches and results in a DW schema model.

By using a method with five approaches, a layered DW schema model will be aligned with business and individual needs. This will be illustrated by using examples.

Key-Words: - Requirement elicitation, Data Warehouse Schema Model, EPC, ADAPT, Data Warehouse Approaches

1 Introduction

Implementation of Data warehousing (DW) in organizations suffers from a high failure rate [1]. Several initiatives to implement DW started with significant expectations, but quickly found that was not guaranteed to satisfy them.

DW departments are usually confronted with a large amount of requirements from several different DW stakeholders, like, for instance, end users (top managers, middle managers and others), DW sponsors, and operational IT departments. The amount of requirements regularly exceeds the available resources for their implementation [2].

Nowadays, researchers, practitioners, consultants, among others, have not yet arrived at a consensus in determining the right method and architecture to implement DW. However, one of the most important issues in DW is how to develop an appropriate DW schema to support analysis, like, querying, exploring, and reporting. Consequently, it is still important to research a method to obtain a better DW schema model in DW field.

The remaining sections are organized as follows: Section 2 describes the problem. Section 3 presents the method step by step. Section 4 presents some examples. Finally in section 5 points out conclusions of this study and future work.

2 Problem Formulation

Existing DW modeling approaches can fall within three basic groups: goal-driven, data-driven and user-driven [3]. Each of the three approaches is usually used separately:

1. Goal-driven approaches consist to analyze the business to obtain a set of visions by interviewing the top and middle management, these set of visions must converge in order to obtain a set of quantifiable Key Performance Indicators (KPI’s) [4]. Goal driven approaches have a disadvantage to depend on the availability of the top and middle management to be involved.

2. Data-driven approaches consist to study the data sources to determine the DW schema model [4,5]. Other authors [6,7] proposed an automatic or semi automatic methods to create DW schema models from Entity Relationship Models (ERM) of data sources. Data-driven approaches have a disadvantage to create a set of information structures for storing information in the DW and
3 Method of DW

In this paper a method based in a multi-driven approach will be described, this approach is a combination of business process management approach (process-driven approach) with goal-driven, user-driven, technology-driven and data-driven approaches, and comprises five stages: KPI’s stage, by identifying user requirements and organizational goals (goal-driven and user-driven approaches); technological stage, by identifying the existing technological capability in the organization to implementing a DW; process stage, by modeling organizational processes (as is); process improvement stage, by modeling organizational processes and its data, based on the requirements and goals identified at stage 1 and its implementation by ensuring that they are aligned with business goals (to-be); and mapping stage; by combining the results of the integrated process-driven, the DW schema model can be obtained from a data-driven approach taking into account the requirements, technological capability and goals identified [11,12,13].

A method for requirements elicitation of a DW schema model is the solution to solve the issues presented above. This method comprises five approaches and five stages. The five stages will be described here:

1. KPI’s stage - started by an identification of organizational issue (or issues), this is called “business pain”; followed by the identification of the business areas and related topics which will be covered by DW, for instance: sales, procurement, production, human resources. This is very important and should be given greater attention, the issues identified allow to set priorities among the various subjects or topics and to make an agenda, which implies setting the DW, in terms of its content and the numbers of iterations to implement - ideally each subject will be an development iteration in the DW; the identification of KPI’s – these KPI’s can be related with one or many subjects or topics; and key users identification, which will be involved in the gathering requirements process. In this stage goal-driven and user-driven will be used. Goal-driven results consist in two documents: one with the business pain description and a list of the subjects / topics and key users; and other with a list of KPI’s. User-driven results are user requirements and a revision of previous KPI’s list, this is achieved by interviewing key users (identified in the goal-driven approach) with the aim of gathering users’ requirements and adjusting the KPI’s obtained at goal-driven approach. As a result one new document and an update of existing document: one with a list of users requirements; and other an update of a list of KPI’s obtained in the goal-driven approach;

2. technological stage – with the aim to identify the technology capability installed in the organization, namely, computers, networks, existing servers hardware characteristics, operating systems, computer applications, database management systems, etc. This result in a diagnosis (snapshot) of the current state of the technology in the company, the level of automation of organizational processes (essentially those related with DW) and the type and size of DW can be implemented and the growth level of DW. There may result in recommendations for hardware and/or software investments which will increase the technology capability.

3. process stage – taking into account the business areas and requirements identified in stage number 1 and the technological capability identified on stage number 2, the goal of this stage was modeling organizational processes (by
using, for example EPC notation [14]) and identify the data necessary to run the process and to feed the needs of stage number 1;

4. process improvement stage – based on processes identified in the previous stage, this stage pretends to optimize organizational processes (by using, for example EPC notation, the EPCs enable an identification of the operational Data Model) and identify new information required in the optimization of business processes; and

5. mapping stage – this is one of most important stage, because DW schema model is obtained based on the mapping KPIs definitions modeled in a top-down fashion with processes and data models from operational perspective – bottom-up fashion, resulting in an analytical model, a ETL process and more operational data required. If operational data model do not include all required data, then the requirements list must be updated as equal the process or processes must be modified (this implies to go to the process improvement stage again). It may also happen that is not possible to obtain such operational data and, if so, the KPI's stage must be repeated in order to adjust the KPI's, or, decide to close the project.

The method with five stages embraces goal-driven, user-driven, technology-driven, process-driven and data-driven approaches. In the stage number 1, the approaches goal-driven and user-driven are used. Technology-driven approach is used in the stage number 2. Process-driven and data-driven approaches are used at stage number 3 and 4. Finally, in the stage number 5 a data-driven approach was used.

The method includes two approaches usually not used in DW initiatives: technology-driven and process-driven approaches. The technology-driven approach is very important because the organizational technological capability installed limit the DW schema model. The process-driven approach allows align business process with the information (KPI’s) need by top and middle management. The others approaches, goal-driven, user-driven and data-driven are usually used to obtain DW schema model, but, normally not combined together.

Now it is possible to answer the research question – How a multi-driven approach with five approaches/stages helps to obtain a better data warehouse model? The paper shows that the DW schema model obtained from a combination of several approaches can improve the satisfaction and trust of organizations and individuals simultaneously, because by following a multi-driven approach that covers, goals and user requirements, and better processes (with right data) and, finally, appropriate technology.

2 Examples
A Portuguese industrial SME was chosen to be an example of implementation.

In the stage number 1 – KPIs stage – the involvement of company management is essential. Thus a set of meetings were conducted with the top and second management levels of the company. As a result of these meetings, a clearly identification of a "business pain" and a set of KPI’s (business goals) were emerged.

A "business pain" is something that needs to be improved within the company and may justify this intervention. In this sense, the company needs to better understand their customers in order to be able to respond to their needs in a most appropriate way. By not knowing their customers, for example, the company cannot determine, upon arrival from customers a request for quote or order the priority of this task. The need arises to classify and rank the customers to conclude who are the most important and to determine their potential. Therefore, the company needs, because the costs are very high, to know how many customer quotes or contacts were made until being materialized in an order.

The highest levels of management of the company identified a large number of KPI’s, but it was found that these KPI’s reflect very tactical concerns rather than strategic, for instance: number of pieces produced, number of minutes spent on mobile phone by vendors; etc.

After these steps, a set of personal interviews with middle management to identify their information needs are conducted. As a result, a transformation of these information needs into KPI’s and an identification and validation of these KPI’s are new or have already been identified are obtained. There are also some KPI’s related with the “business pain” described above. Thus a new KPI’s list is obtained.

The next step was to model the KPI’s in ADAPT™, see Fig.1 [15]. In this step, various analytical models (cubes and reports) necessary to meet the management information needs are modeled. At this point, it is possible to realize what are the operational data needed to satisfy the KPI’s and to realize that the ERP cannot provide all information (through the existing operational data) to meet the needs of the KPI’s, especially those who were
related with "business pain" were identified a lack of information in various levels of management,

![Diagram](image)

**Figure 1 – ADAPT™ KPI Example**

In the stage number 2 – technological stage – There are three ERPs: manufacturing, administrative / financial and human resources, and the three ERPs are not completely integrated. One of these ERPs - manufacturing, because of its importance, was the subject of further detail. ERP manufacturing is not composed by modules and follows an architecture positioned between the client-server and web based architecture. An installation of a web browser and a small component of reporting are enough to work with the ERP. The ERP can be installed in one or more servers, depending on the desired configuration. In the case study, it appears that there was only a single server that also contained the data base management system (DBMS) in this case the choice was based on a database Oracle 10g, but it could be another DBMS. The ERP, includes the following features: budgeting for the production; orders and customer management; work order (interface for production); stocks management/logistics; concept of perpetual and intermittent stocks; supply chain management: integration of orders/receipts/returns; vendor management/quality; invoices management; production data collection - data collection by multi-terminal through sensors installed on production machines; integration with production equipment using standard protocols; and shipping and freight. Regarding to the infrastructure installed: servers, network, computers are quite modern and complies with the essential features needed to manage a DW solution.

In the stage number 3 – process stage – Let us focus on the process showed in Figure 2. The quotation process in industry is significant cost factor, because the sales-man, who calculates and issues the quotations, should be an expert (with a high salary) and the quotation process are quite time consuming. The sales-man has to know about the different production processes, the different raw materials, the different production machines, and so on. On the top of all, he has to advise the customer about some technical aspects, costs and quality of the final product. In order to reduce the unit-price costs, the customer changes some requirements, asking for a 2nd version of the original quotation, and so on.

![Diagram](image)

**Figure 2 – EPC Example.**

The managers want to know the costs variation of all those quotations (original, 2nd version, 3rd version and so on…), until the customer approve the last quotation and submit the order. The final
quotation amount of some "smaller" production is sometimes so high, that the profit is close to null. This is one of the factors, which increases the "business pain". A data model of quoting is described in Figure 3.

In the stage number 4 – process improvement stage – it is possible to realize that a significant number of business processes will be involved, in order to gather the right data, on the right time to implement the KPIs (see Figure 4). To do this is necessary implement a Congruence Relation in Relational Models and related issues [11,16].

The example will be presented with no details for obvious reasons, but it will be conclusive enough to allow the understanding of this method and its goals.

This example shows:
- the connections between two different EPCs, showing schematically how those processes are working and where data is coming from and going to;
- the improvement of the data model by a congruence relation, in order to allow the implementation of the "new" KPI. In this case, the DW ETL process will include the congruence relation, in order to get data of the "new" KPI.

The goal of the congruence relation is to define new relationships and gather data from different sources ("congruent data", data related to each other "in a certain (formal) manner"), in order to allow an improvement of the data model.

The EPCs shed light the dependencies of the involved business processes and the data flow.

Finally the mapping stage - this is the most important point of the multi-driven approach, because the DW schema model is obtained. Based on KPIs modelled with ADAPT™ and operational data models from EPCs and results a data mapping. If this data mapping is not enough to feed the KPIs, there are three possible solutions:

1. adjust the KPIs to the reality of operational processes and data
2. change operational processes to collect the required data
3. close the DW project.

In this example the information was not enough and a new quoting system has been developed. Beyond the technical issues of developing a quoting request process, in terms of: the definition of the workflow; a choice of raw materials; and the production routing; it also records, for a customer, each quotation requested, allowing to know how many requests are made and what changed between each one (through a version based system). If the customer accepts the quotation value, the production orders are automatically created and if the customer refuses the quotation value the CRM allow registering the reject reasons.

As described above, the CRM allows storing several indicators, for instance: the number of quotes to obtain an order and number of quotes refused. In addition, the CRM stores data from various version budgets, this let knowing the changes that customer made in the quoting process to obtain the right product and unit price. All this data is stored in an operational DBMS system.

This operational data can now feed the KPI’s and now a DW schema model can be modeled to store all the data necessary to feed the KPIs.
4 Conclusion

Now it is possible answer the research question – How a method with multi-driven approach helps to obtain a better DW schema model? The paper shows that the DW schema model obtained from a combination of several approaches can improve the satisfaction and trust of organizations and individuals simultaneously, because by following a multi-driven approach that covers, goals and user requirements, better processes (with right data) and, finally, appropriate technology.

The improvement of organizational processes has led to obtain further informational records for operational databases in order to ensure conformity with the organizational processes model. This can be considered a determining success factor for implementing DW in organizations. Ensuring that, business processes are aligned with organizational strategy and identify management indicators (KPI’s) that will measure these business processes in order to feed the organizational goals, is critical to the organizations success and information technologies that support organizations.

The combination of several approaches is becoming a concern of several authors, in order to be able to properly define the DW schema model and this is one of the challenges facing to those who want implementing a DW. It appears that all these approaches were important for defining the DW schema model, however, user-driven approach, despite its importance, helped essentially to validate and identify KPI’s, suffers from the difficulty that users have to think beyond the limits of existing technology or processes. Of course this can be minimized if potential DW users were more experienced in using information analytical tools, even if they are basic, such as pivot tables from spreadsheets.

In conclusion, the method with combination of five stages allow bridging inexperience and difficulties that often occur in DW projects and this guarantee that the DW model is aligned with business strategies.

As future work, it is important implement more DW in organizations to validate the method.

5 Acknowledgements

This work was financed by FCT - Fundação para a Ciência e Tecnologia for the project: PEst-OE/EEI/UI0319/2014

References:


