Data warehouse development with EPC

MARIS KLIMAVICIUS
Faculty of Computer Science and Information Technology
Riga Technical University
Meza 1/3, Riga, LV-1048
LATVIA

Abstract:. The Author introduces the approach to address data warehousing issues at the enterprise level. The approach covers data warehouse requirements specification and initial design stages. Author proposes to extend business process model, to cover the main issues connected with development of data warehouses. In this paper Event-Driven Process Chain diagram is extended, a business process modeling language, which has been developed within the framework of ARIS. The Author considers that the use of business process models is necessary, as they reflect both enterprise information systems and business functions, which are important for data analysis. Proposed approach divides development into three steps with different detailed elaboration of models. ARIS framework is used for development of all models. The Described approach gives possibility to gather requirements and display them to business users in easy manner. As examples to demonstrate the approach business processes of the Treasury of the Republic of the Latvia are used within the paper.

Keywords Data warehouse development, business process modeling, requirements gathering.

1 Introduction

Data warehouse is more than just another big database. It is defined as "a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management's decision-making process" [1]. In modern organizations, data warehouse plays a crucial role, as more and more business processes require information from the data warehouse. Data warehouses integrate data from multiple heterogeneous information sources and transform them into a multidimensional representation for decision-making process. Apart from a complex architecture, involving data sources, the data staging area, operational data stores, the global data warehouse, the client data marts, etc., a data warehouse is also characterized by a complex lifecycle [2].

Enterprises live and die on their decision-making, and specifically on its accuracy, timeliness, and effectiveness. Most enterprise decisions are based at least partly on data analysis. So as the quantity and comprehensiveness of available data keeps expanding, the scope of analytic decision-making expands as well. Data analysis is used for every kind of enterprise decisions – marketing tactics, sales tactics, investment strategy, budget strategy, supplier selection, and others.

Enterprises have increasingly invested in data warehouses, business intelligence tools, and other analytic technology. These investments have for the most part been extremely worthwhile. Studies have shown that 40% of all data warehouse projects never develop, and 85% fail to meet business objectives [17]. On average, data warehouses usually fail as a result of poor communication between IT and business professionals, as well as developers who possess poor project management skills and procedures [18]. But superior data analysis can lead to an almost unlimited range of operational advantages - cheaper supplies, more targeted marketing, wiser investments, more profitable product mix and many more.

To achieve any of these improvements successful data warehouse implementation is essential and it is necessary to reduce the risks related to development process and methodology [18]. In spite of popularity of data warehouse projects there is still no single approach for developing them. There are as many ways to build data warehouses, as there are companies to build them. Author considers that for better support of corporate strategic needs, during the data warehouse development, enterprise view of the business objectives, business processes and supporting data is required. Important step in information system and data warehouse development is requirements gathering, but this is also a weak link in data warehouse development process as studies show [4].

The requirement gathering is very important step in development of information system and data warehouse. And, as studies show [4] it is also a weak link in data warehouse development process.

The author by analyzing failures and issues connected with data warehouse development process has come to conclusion that improvements should be made in the starting process of development. Business requirements, business processes and structures of data warehouse should be integrated through business process models. To prevent the above-mentioned issues the author proposes to extend business process modeling language.

Proposed extension of Event process chain (EPC) diagram creates conceptual view of data warehouse, this view binds enterprise applications with enterprise processes.

As a tool or platform for data warehouse development Architecture of Integrated Information System (ARIS) is used. ARIS is well known tool for modeling enterprise architectures, and author of the paper apply it to data warehouse development process.

The rest of the paper proceeds as follows: 2 section presents description of related work. Process and main problems of data warehouse development are described in section 3. Section 4 provides with description of business process modeling tool used within this paper. Data warehouse development using business process modeling tool examples of business process models are presented in section 5.

2   Related work

Over the years, the scope of business processes and business process modeling has broadened. There are different business processes modeling techniques, which enable organizations to document, model, understand and improve their business processes. Business process modeling today is used for information system development, enterprise application integration but, so far, there are few studies that focuses on the relationship between data warehouse and business process modeling for development purposes at enterprise level, especially requirements gathering.
Studies have shown that business process modeling for development of the specific stage of the data warehouse lifecycle is useful. These data warehouse diagrams are available for multidimensional models [16] and ETL processes [15]. These researches do not address the link to business processes; they solve some particular problem in data warehouse development lifecycle.
Some authors provide design phase of data warehouse development based on entity relationship models. For example Boehnlein and Ende [7] propose to use structured entity relationship model for derivation of data warehouse structures. They

show how this data modeling technique is used for derivitation of initial data warehouse structures from the conceptual schemes of operational sources, but authors also agree that approach is not practical for automatic creation of multidimensional tables because of varying requirements.

V. Stefanov, B. List and J. Schiefer [5] extend Architecture of Integrated Information System (ARIS) with additional business intelligence perspective to show where and how business processes use decision support data.

In this paper author present a business process oriented approach to the development of data warehouse. In order to derive initial data warehouse structures author use business process models instead of operational data models to derive initial data warehouse structures.

3   Data warehouse development

Most techniques that are used by organizations to build a data warehousing system use either a top-down or bottom-up development approach. In the top-down approach [1], an enterprise data warehouse is built in an iterative manner, business area by business area, and underlying dependent data marts are created as required from the enterprise data warehouse contents. In the bottom-up approach [13], independent data marts are created with the view to integrating them into an enterprise data warehouse at some time in the future. The development approach for the data warehouse environment is iterative or a spiral development approach [12].

There is still a great deal of confusion about the similarities and differences among these architectures. In spite of these differences there are two main steps in data warehouse development, which are very closely connected – requirements gathering and information modeling. The author considers that both approaches [12] [13] have such disadvantage as gap between enterprise business processes and technical design of data warehouse. There is a fundamental dilemma in developing information model. Every information model is a compromise of some sort between business requirements, performance requirements and fundamental data warehouse design aspects [8] Figure 1 illustrates the information-modeling dilemma.
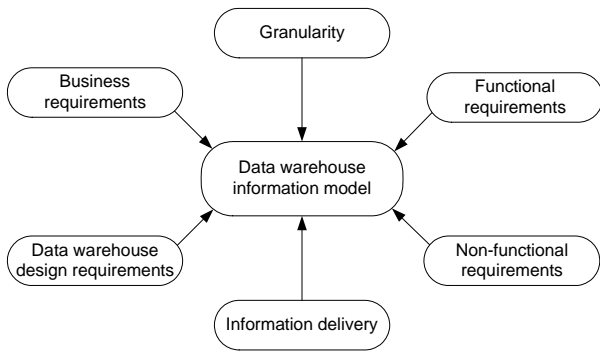
Fig. 1. Information modeling dilemma

Usually, the answer to the question "What information do you need? " is simple: "I need everything!". Business re   quirements usually conflict with non-functional requirements like performance and usability. There is always a trade-off between performance and amount of information available, and the more complex is the information model, the more complex is to use it and more difficult it is for occasional end users to create and execute meaningful queries with consistent and complete result sets.

Data warehouse design requirements are driven from the technical side of data warehouse development, as these people are responsible for actually developing and maintaining the data warehouse system. The art of information modeling is to define an information model that best suits to all these requirements without sacrificing too much of the business side of requirements.

Data modeling is complicated - therefore hard to follow process. The author proposes approach, which raises the modeling to a higher level that is more clear and easy to follow by business users.

4   Business process modeling tool

In this work author address the missing link in data warehouse modeling between data warehouse design and business processes. The author extended a business process modeling language with an additional dimension to close this link. With this perspective, it is possible to create models that show where and how business processes use decision support data and what kind of data are stored in information systems. One of the frameworks is Architecture of Integrated Information System (ARIS). In research as well as in practice, the ARIS is accepted as a standard framework for business process engineering. EPC is a method developed by Scheer, Keller and Nuttgens within the framework of ARIS and is used by many companies for modeling, analyzing, and redesigning business processes.

It is well known fact, that modeling techniques can facilitate communication between representatives of different systems and increase the development speed. Most of business users have a comprehensive understanding of their business processes, which they want to explore and analyze. Business users know how the enterprise works and developers are familiar with information systems running that business. Major problem in building a data warehouse is to identify and consider information needs of potential users.

EPC is applied because of its widespread use in many companies for modeling business processes, and because of its flexible view concept, that allows separating the different aspects of a business process.

The ARIS concept involves dividing complex business process into separate views to reduce the complexity and integrating these views to form a complete view of the whole business process. A starting point to form these views is to understand and model a business process.

5   Business process models driven data warehouse development

The main advantage of the business process oriented development strategy is that the business process model contains a formal description of the users' requirements. Another advantage is possibility to identify requirements that could not be satisfied with the actual information offerings of the source systems. The main improvements by proposed approach:

- Make development of data warehouse clear for business users/decision makers
- Make relationship between data warehouse and business processes
- Identify data needed for analysis.
- Define pseudo data warehouse data structures at beginning
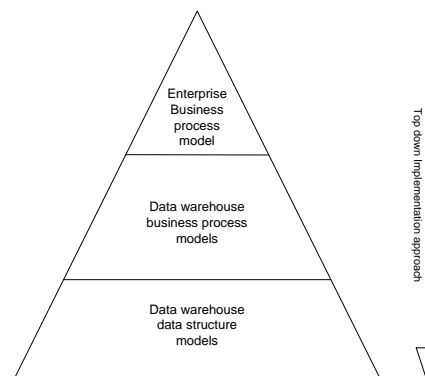
Fig. 2. Data warehouse development stages

Development consists of three phases (Figure 2). The first stage is Enterprise business process model, which describes organization processes at very high level. Next phase is data warehouse business process models. These models are more detailed description of enterprise model, and they give comprehension of business processes involved in data processing. Third phase is detailed design of data warehouse; information model and flows, as well as data types are developed within this phase. Review of these phases is given in next sections.

5.1. Enterprise business process model

Development starts with comprehensive enterprise business process model. This model shows the main functions of the organization it has nothing data warehouse specific. The aim of this model is to define the starting point from which further requirements specification and information modeling has to be done.

For the development of such model EPC diagram is used. It shows the main enterprise's processes and their interconnections. Further developments are done from this diagram. As example we use main business processes (Figure 3) of the Treasury of the Republic of the Latvia.
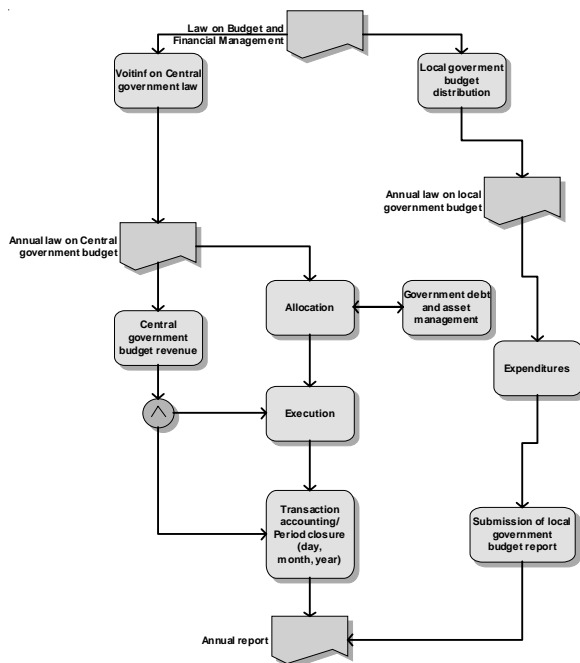
Fig. 3. Budget execution process

The aim of this diagram is to define enterprise view on business processes. Data warehouse development consists of iterations [12]. Function blocks of the diagram are iteration, but it is possible that some functions will be developed in one data warehouse development cycle. During the development of

business processes it is important to analyze whole process not just underlying information systems. Proposed concept is more based on the enterprise business processes, rather then on systems as in Bill Inmon's concept. Proposed approach eliminates issues of long and expensive first iteration of data warehouse development.

5.2. Data warehouse business process models

Author extended a business process modelling language with an additional perspective. This perspective links business processes with underlying data; such approach describes information areas [3] much clearer. A data warehouse business process model describes the business information needed to support the business functions that are encompassed in the baseline business model. It essentially contains information on high-level data areas and their classes. Additionally, it identifies the need for data warehouses and data marts. Such approach allows the developers to understand which source systems and which information is crucial.

Budget execution process (Figure 3) consists of following sub processes for which data warehouse business process models are created:

- Voting on Central government law
- Central government budget revenue
- Allocation
- Government debt and asset management
- Execution
- Transaction accounting/Period closure(day, month, year)
- Local government budget distribution
- Local government expenditures
- Submission of local government budget report

Example of data warehouse business process model is shown on Figure 4. This diagram shows the process of budget allocation (Figure 3) in more detailed view. This diagram consists of standard EPC elements. The author add extra perspective that describes the data warehouse dimensions and facts. The author named these perspectives as Information areas.

Budget allocation is one of the starting points in Budget execution process. The aim of this business process is approval and accounting of financial plans of institutions and ensuring of opening and closing of budgetary book accounts in accordance with structure of accounts.

Information area consists of business objects, which are lowest-level business entities to describe business processes.
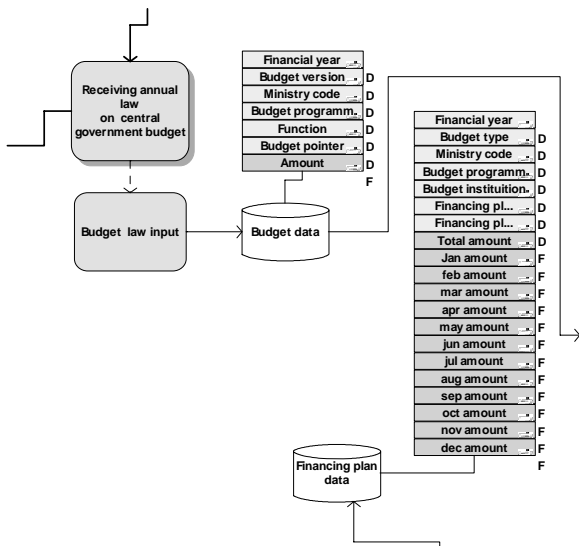
Fig. 4. Data warehouse business process of Budget allocation

When such diagrams are constructed the developers and the customers agree about how the data warehouse is going to be designed. Such type of business process representation is understandable for business users.

Data warehouse business process model of budget allocation combines functions with related data. Figure 4 shows part of such diagram to demonstrate approach presented within this paper. Processes and pseudo data structures are described in understandable manner for business users. Author consider such approach reduce the gap between business process and data warehouse.

During the modeling of Budget allocation process following information areas were identified:
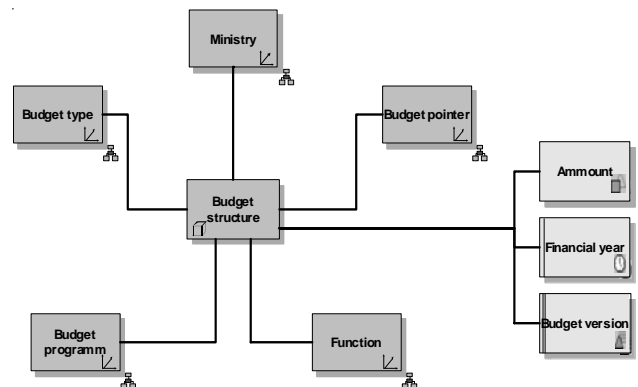
- Budget structure data
- Financing plan data
- Budget account data (master data)
- Account balance data

The next phase is a detailed description of business objects. During this phase the complete data warehouse scope is analyzed, as well as all the organisation structures and existing information systems. A conceptual data model is created and the major systems of reference identified, to ensure the consistency of the future build cycles.

5.3. Data warehouse data structure models

Data warehouse structure is complicated; it is not only dimensions and facts. These diagrams are important in description of business objects and their interactions. In this phase Information areas developed in previous stage are used. The transformation and extraction process is not

described, but the result of the information modeling is the target of the transformation process.

Budget structure derived from data warehouse business process model is shown on Figure 5.


Fig. 5. Dimensions and facts of Budget structure

The data warehouse structure diagrams are used to describe the structure of a data warehouse.

The relations of the information as well as their layout to each other are considered to be the central aspect. The interplay of the dimensions is presented by the Star scheme. The hierarchical relationships of the features with each other are described by tree structures. Finally, the dimensions can be allocated to master data tables with the help of the structure diagram.

Detailed description of Budget structure is created within this phase (Figure 6)
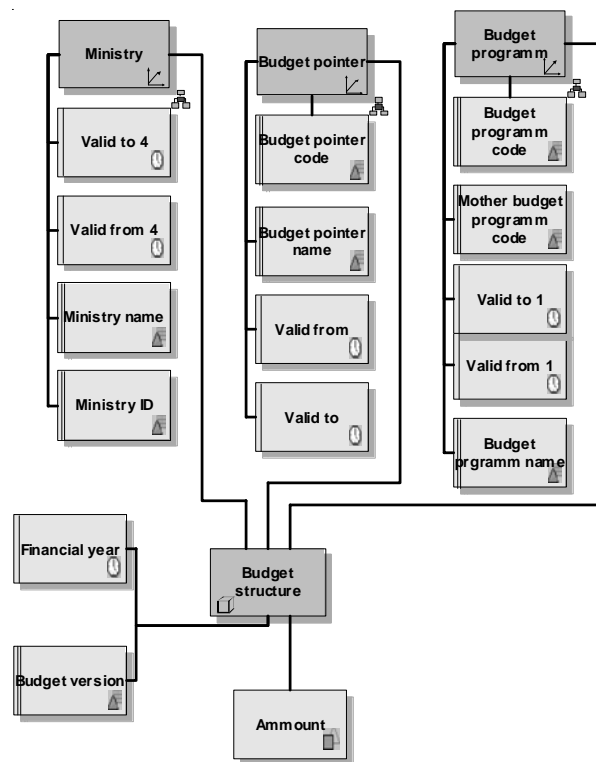
Fig. 6. Data structure model of Budget structure

Implementation of this phase depend on data warehouse solution and architecture, because there should be taken into account software specific requirements.

6. Conclusion

In this work author proposed the approach of business process models oriented development of data warehouses. The business process modeling language author have chosen to extend with the data warehouse perspective is the Event-driven process chain. With the data warehouse perspective, it becomes possible to create models that show business process interaction with data warehouse structures thus avoiding critical requirements specification and design issues. Presented approach can be applied either for top-down or bottom-up data warehouse implementations. Organizing development in such way involves business users much deeper in development, and author consider that such involvement reduces risks of data warehouse failures and help to develop the data warehouse that better reflects business requirements. Visualization of data warehouse information model provides with necessary information to decide which dimensions and facts are important in every business process.

References:
1. W. H Inmon, R. D.Hackethorn, Using the Data Warehouse. New York: John
Wiley & Sons(1994)
2. P. Vassiliadis, C. Quix, Y. Vassiliou, M. Jarke. Data Warehouse Process Management, Information Systems, Vol. 26, No. 3, 2001
3. M.Klimavicius, U.Sukovskis, Business process driven data warehouse development, Scientific Proceedings of Riga Technical University, Computer Science Series, Applied computer Systems, 6th issue, Vol. 22, Riga, Latvia, RTU(2005)
4. M.Klimavicius, U.Sukovskis, Risk management methodology for data warehousing projects, ISD'2005 Proceedings of the Fourteenth International Conference on Information Systems Development: Pre-Conference, Karlstad, Sweden, Karlstad University Studies(2005)
5. V. Stefanov, B. List and J. Schiefer: Bridging the Gap between Data Warehouses and Business Processes - A Business Intelligence Perspective for Event-Driven Process Chains, Proceedings of the 9th International Enterprise Distributed Object Computing Conference (EDOC 2005), Enschede, The Netherlands, IEEE Press(2005)
6. V.Stefanov, B. List and B. Korherr: Extending UML 2 Activity Diagrams with Business Intelligence Objects, Proceedings of the 7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2005), Copenhagen, Denmark, Lecture Notes in Computer Science, Springer Verlag (2005)
7. Michael Boehnlein, Achim Ulbrich-vom Ende, Deriving Initial Data Warehouse Structures from the Conceptual Data Models of the Underlying Operational Information Systems, DOLAP 1999, ACM Second International Workshop on Data Warehousing and OLAP, Kansas City, Missouri, USA, ACM(1999)
8. Kevin McDonald, Andreas Wilmsmeier, David C. Dixon, W.H. Inmon, Mastering the SAP Business Information Warehouse, New York, Wiley Publishing (2002)
9. IDS Scheer AG, ARIS Methods manual, IDS Scheer (2004)
10. Vaclav Repa, Business Processes Based Information Systems Development, BIS99 Conference, Proceedings of the 3rd International Conference on Business Informaiton Systems, Poznan, Poland, Springer-Verlag (1999)
11. Sergio Lujan-Mora, Panos Vassiliadis, Juan Trujillo, Data Mapping Diagrams for Data Warehouse Design with UML, Conceptual Modeling - ER 2004, 23rd International Conference on Conceptual Modeling, Shanghai, China, Lecture Notes in Computer Science, Springer (2004)
12. W. H. Inmon, Building the Data Warehouse, Fourth Edition, Indianapolis, Indiana, USA, Wiley Publishing (2005)
13. Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite - The Data Warehouse Lifecycle Toolkit, John Wiley & Sons (1998)
14. Paulraj Ponniah, Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals, John Wiley & Sons(2001)
15. Vassiliadis, P., Simitsis, A., Skiadopoulos, S.: Conceptual Modeling for ETL Processes. In: Proc. of 5th Intl. Workshop on Data Warehousing and OLAP (DOLAP
2002), McLean, USA (2002)
16. Luján-Mora, S., Trujillo, J., Song, I.: Extending UML for Multidimensional Modeling. In: Proc. of the 5th Intl. Conf. on the Unified Modeling Language (UML'02). Volume 2460 of LNCS., Dresden, Germany (2002)
17. Schneider, A., Wong, B., Wong, K., Maloney, G., Data Warehousing, Bentley College Community (1999).
18. Josef Schiefer, Beate List, Robert M. Bruckner , A holistic approach for managing requirements of data warehouse systems, Eighth Americas Conference on Information Systems,(2002)