ISSN 1870-4069

Agile Dimensional Model for a Data Warehouse Implementation in a Software Developer Company

Kathya E. Mercado[1], Cynthia B. Perez[1], Laura P. Lopez-Arredondo[1], Karina Caro[2], Luis A. Castro[1] and Luis-Felipe Rodriguez[1]

[1] Sonora Institute of Technology (ITSON), Ciudad Obregon, Sonora, Mexico

[2] Drexel University, USA

{kathya.mercado, laurapaolalpz}@hotmail.com, {cynthia.perez, luis.rodriguez}@itson.edu.mx, luis.castro@acm.org, karinacaro@drexel.edu

Abstract. Nowadays, the increasing development of Business Intelligence (BI) solutions in organizations, has enabled executives achieve a better understanding of business information for timely and rapid decision-making in a tremendously dynamic market. Although there is an increasing interest in adopting an agile approach to the software development, the emergent need of using agile methodologies in BI solutions is undeniable. This paper discusses the importance of using agile methods in the design and development of data warehouses taking into account the business processes, requirements analysis, and organizational objectives. Thus, we present a case study derived from a real-world business project where the agile methodology Business Event Analysis and Modeling (BEAM) is used to design the data warehouse. The project is based on a billing system with about one million operations on a daily basis with more than 15,000 clients. Finally, the results of this paper include the design of the dimensional model using an agile approach, the construction of the data warehouse through the ETL processes and an interactive dashboard according to the key performance indicators defined by the business decision-makers.

Keywords: business intelligence, agile dimensional model, data warehouse.

1 Introduction

In the business world we aim to obtain greater profits and a greater competitive advantage, *hence*, appropriate timely decision-making plays a very important role in the fulfillment of the organizational objectives. Thus, the use of Business Intelligence (BI) systems can help meet these needs as it provides a set of methodologies, applications, and practices focused on the information management for accurate, timely decisions making in an organization.

A major component of any BI model is the design and implementation of a data warehouse in organizations, since it provides valuable and strategic information to support decision-making through real-time access to business transactions and advanced business analytics [1]. According to Inmon [2] and Imhoff [3], a data warehouse in organizations is an integrated data collection, non-volatile and variable over time. Hence, it has a complete history of the organization, beyond the transactional and operational information favoring the data analysis for decision-making.

Today, the two most widely used methodologies for the design and implementation of data warehouses are the model of Inmon [2] and Kimball [4]. They consider the data warehouse as the central repository of data for organizations that is used to present business reports. The difference between these two methodologies lies in how to make deliveries of progress (time) and how to manage changes during the process (see Table 1).

Recent studies tend to show that use of agile methodologies for the design and implementation of data warehouse in organizations is playing an important role to obtain value information to help decision-makers and to generate competitive advantage by improving the extraction and processing knowledge. These studies present a structured methodology, inspired from the agile development models as Scrum, XP and AP [5]. Thus, an increasing number of companies are choosing for an agile philosophy in software development due to the constant need to be flexible and adaptable to the technological changes and the new user demands [6-10].

For this reason, it is important to consider agile methods for the design and implementation of a data warehouse in a BI system (see Table 1). In this regard, it has been shown that agile development processes increase the potential for developing the success of a data warehouse by solving many of the typical problems presented in traditional methodologies [11, 8, 12].

In this paper, we use the BEAM methodology in the design, modeling, and implementation of a data warehouse for a management system. This system processes one million operations on a daily basis from more than 15,000 clients. The agile design and modeling of the data warehouse are presented through a case study of a Software Development Company located in Mexico, which offers software solutions focused on meeting the needs of development, implementation and support to any business sector.

2 Agile Dimensional Modeling

Traditional data warehousing projects follow the waterfall structure to perform dimensional modeling [13]. However, its use is increasingly unlikely and alternatives of analyzing and designing similar to those used in software development projects such as agile methods are looked for [7, 12, 14]. In this respect, the agile dimensional modeling is being considered as a solution for BI systems since it allows developers to reduce the risks that the waterfall structure could produce [11, 14]. All this is possible by adopting a highly interactive, incremental and collaborative approach to the whole analysis, design and development activities of a data warehouse, such as the agile BEAM methodology [14] (see Figure 1).

Table 1. Comparison of the Inmon, Kimball, and Corr methodologies.

	Inmonn [2] (Traditonal)	**Kimball [4] (Traditonal)**	**Corr [14] (Agile)**
Business Decisions	Strategic	Tactical	Collaborative
Scope	Product owner	Project manager	Team
Approach	Top – down	Bottom – up	Bottom – up
Objective	Deliver a robust technical solution based on proven methods	Provide a solution that facilitates the end users to consult the data	Responding to change and user needs
Data requirements	Enterprise-wide	Business process	Individual business requirement (KPIs)
Data modeling	Normalized form (3NF)	Dimension model (Star or Snowflake)	Dimension model (Star or Snowflake)
Orientation	Enterprise-wide	Business process	People
Communication	Formal	Formal/Informal	Informal
Time	Longer start-up time	Shorter start-up time	Minimal start-up time
Project schedule risk	High	High	Low
Ability to respond to change	Low	Medium	High
End users involvement	Minimal	Oscillate depending on the project	High
Cost to build	High initial cost	Low initial cost	Minimal initial cost

2.1 BEAM Methodology

Corr [14] proposes the BEAM (Business Event Analysis and Modeling) methodology, an agile data modeling method for the design and development of data warehouses and data marts. This method combines analysis and modeling techniques to meet data requirements related to business events and data modeling for database design that is easy to understand by stakeholders and also, easy to translate into logical/physical models for IT developers. The BEAM methodology involves stakeholders who think beyond their current reporting requirements by describing data stories, that is, narratives that define the dimensional details of business activities necessary to be measured. In order to obtain these data stories, data modelers ask questions to stakeholders using a framework based on the 7Ws (who, what, where, when, how many, why and how) [14]. The way to find these answers of the 7Ws and make sure they inform data warehouse design is to ask end-users about the events that are happening in their business. Therefore, the enhanced Start Schema is used to generate and show schema of physical data bases, where are they involved Data Modelers, DBAs, DBMS, ETL Developers, BI Developers and Testers. This framework is one of the main activities of the BEAM methodology because it allows discovering and modeling data requirements and thus, to construct the table of dimensions and facts of the data warehouse depicted through the star model.

According to Corr [14], the BEAM methodology has several diagrams for the analysis and design of the data warehouse model, such as: BEAM Table, Hierarchy Chart, Timeline, Event Matrix and Enhanced Star Schema.

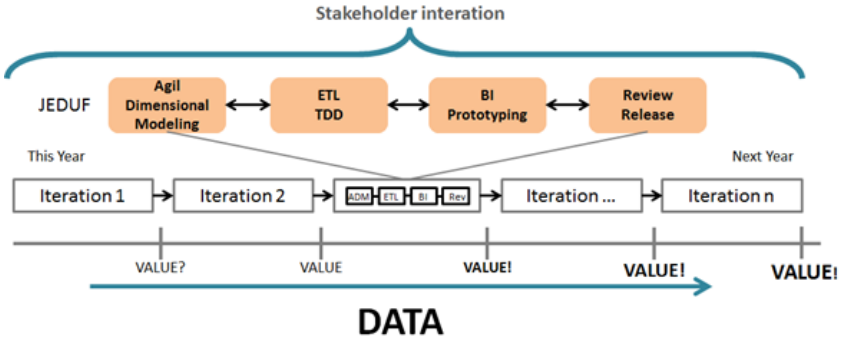

Fig. 1. Agile data warehouse development timeline. Source [14].

For instance, the BEAM Table is used to model business events and dimensions at the same time; people involved in this diagram are data modelers, business analysts, stakeholders and BI users. The Enhanced Star Schema is used to visualize the dimensional model for the implementation of the physical database schemas; the people involved in this diagram are Data Modelers, DBAs, DBMS, ETL Developers, BI Developers and Testers.

3 Agile Data Warehouse: A Case Study of a Billing System

This case study focuses on the design and implementation of a data warehouse using the BEAM methodology for a billing system of a company based in Mexico with operations in software development with around 15,000 active clients in 19 countries and more than a million operations on a daily basis. Despite all the information daily stored on the company's servers, this information is not used or analyzed so far by the working team, identifying an area of opportunity for the design and implementation of BI systems. Therefore, decision-makers could offer their customers significant knowledge through scorecards and thereby provide a competitive advantage. The management system access to a database based on the Entity-Relationship Model (ER) allowing to record, update, delete, and query information from the main business processes. This system has the following modules such as: billing, inventories, clients, payroll, branches, among others. Hence, in this case study, we focus on the billing module. The database used by the system contains around 70 tables using only the most relevant according to the key performance indicators (KPI).

3.1 Analysis, Design and Implementation of the Data Warehouse

Unlike the development of software applications, where the requirements of the organizations are often relatively well defined by the result of the stability of business rules over time; create a data warehouse depends on the reality of the company and its current conditions.

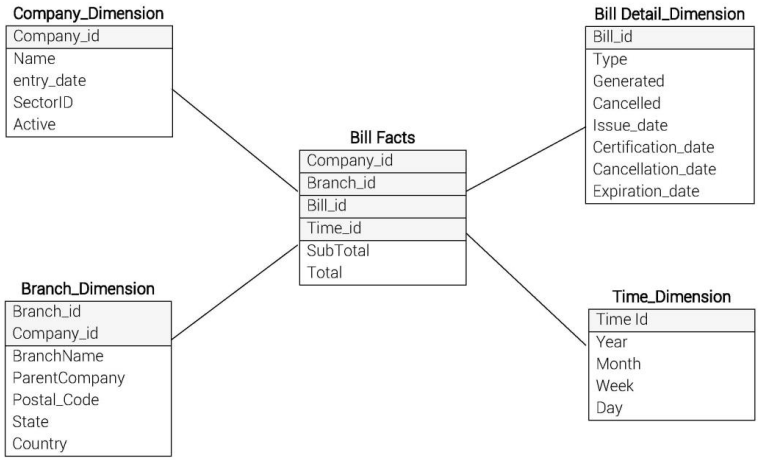

Fig. 2. Star schema for the proposed data warehouse.

Thus, the company requires identifying useful information in order to obtain significant information about its clients. In this way, the following key performance indicators (KPIs) were identified along with the business owners for the design and implementation of the data warehouse: a) Compare monthly growth of registered companies; b) Identify customer loyalty, through the use of the system more than 3 years; c) Identify quantity and list of branches by parent enterprise; d) Visualize the States where there are more than four companies using the management system; e) Measure different types of invoices volumes identifying the invoices variations not only by a time period but also, for company; f) Geographically view the top 10 companies that make the most invoices for a period of time; g) Compare the invoices by branch, time and state.

In order to model the data requirements, the 7Ws framework was used resulting in the identification of facts and dimensions in the star model, presented in Figure 2.

In this regard, the proposed dimensional model contains a fact table where the measurements or metrics of a specific event are recorded; for example, the invoice for a purchase and foreign keys referencing dimensional data tables (Company, Branch, Bill Detail and Time) which contain descriptive information. In order to implement the proposed model (see Figure 2), it is necessary to perform the ETL (Extraction, Transformation, Load) process which enables moving data from multiple sources, transform and load them into the data warehouse to analyze and thereby giving valuable information to organizations. In this way, Microsoft SQL Server Integration Services (SSIS) of Visual Studio 2015 was used to perform the ETL process. Finally, the information contained in the data warehouse was visualized through a scorecard developed using the Microsoft PowerBI tool. Thus, the KPIs were analyzed in order to identify the visual elements corresponding to each key indicator.

Once the visual elements are selected, the information about the clients is then shown in the scorecard by using queries in which decision-makers could use it interactively.

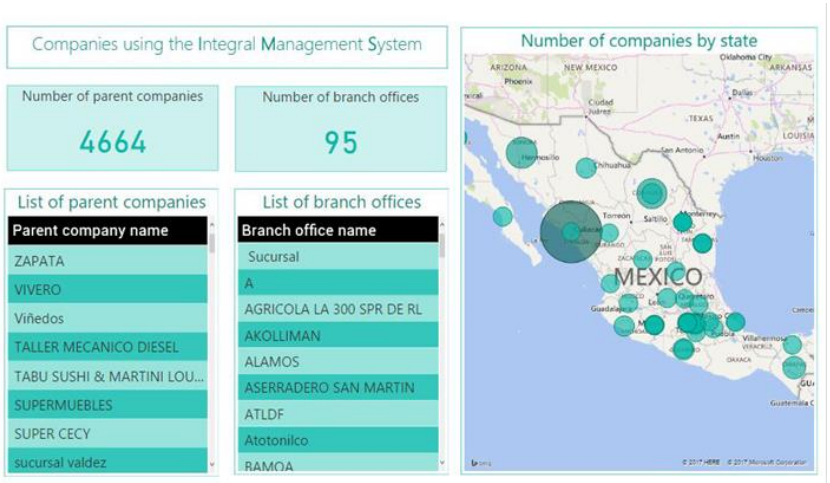

Fig. 3. Scorecard interface: Geographic coverage view of the system.

On the other hand, the software company's CEO was interested in seeing through graphs and trend lines, the situation and invoice generation behavior of those companies who use the management system.

In this way, it could be easily visualized when billing peaks are produced, i.e. sales generated by companies using the system. Figure 3 presents the view of the BI system that shows by state the number of companies who use the system, the bigger the circle the more companies that use the system are in the corresponding state. In this way, it is easy to identify the states where there is little or no presence of the system helping the decision-maker to pay attention in sale strategies. Likewise, Figure 4 presents the view related to invoice analysis where the number of invoices per type is visualized through an interactive list, a trend line to observe the behavior of the number of invoices is generated by each company who use the system and a pie chart showing the top 10 companies with the highest number of invoices issued by the system is shown.

In this way, it is possible to graphically observe the behavior of the number of invoices in a period of time by interacting with the system, allowing the decision-makers not only to know in what years more invoices per company are registered but also, detect those decreases that may indicate a risk in the strategic plan in order to develop an action plan.

3.2 Discussion of Results

The design and implementation of the proposed BI system using BEAM methodology allowed the analysis and design of the data warehouse through an agile method that focused on the users' needs and that easily respond to changes. Therefore, meetings with the working team, i.e., stakeholders, BI users, ETL developers, business analysts, among others, were held in order to compile information requirements during all stages of the project emphasizing uninterrupted communication and collaborative work.

Fig. 4. Number of invoices by time period.

On this basis, it ensures a greater understanding of the data warehouse information and the functionalities of the BI system. Moreover, the agile method maintains a logical data structure, scalable and adaptable to future functionalities such as the integration of other system's modules, predictive analysis, among others.

In this way, a robust and scalable BI system was designed and implemented where decision-makers can count on reliable, fast, flexible and easy-to-understand analyses through the scorecard, thereby facilitating the diagnosis of indicators and decision making. Accordingly, Figure 4 shows a scorecard of the invoices section, thus, the use of the scorecard provides reports of different participants in the decision-making process, representing an opportunity for homogenize and refine business processes. For this reason, it is expected to improve business opportunities through the use of key performance indicators by the extraction, processing and presentation of significant information according to the business strategic objectives. Eventually, the use of the BI system will positively impact the improvement of the company's value chain processes, its competitiveness and thus, the profitability of the business.

4 Conclusions and Future Work

Today, entrepreneurs need to analyze and interact with real-time visual information in order to support decision-making. In this regard, the methodologies used in the design of BI systems should consider the current needs and challenges where business requirements are not static and change constantly. Hence, this paper proposed the use of an agile dimensional model for the design and implementation of a data warehouse based on the BEAM methodology applied to a case study for a Software Development Company. In order to complete the project successfully, the organizational requirements were defined, the star schema was modeled, the ETL process was

executed, the data warehouse was implemented and finally, the KPIs were graphically displayed into the scorecard for decision-making.

The results obtained from the use of an agile methodology are found as a model easy- to-understand for the stakeholders; for this reason, it is mandatory to involve them in the whole process.

By adopting this agile approach, flexibility is ensure, as well as, personal coordination with the stakeholders, consistency and simplicity in the whole process.

As a future work, we plan to incorporate a predictive analysis section into the BI system allowing decision-makers to discover patterns, opportunities and prevent risks by increasing the profitability of the business.

References

1. Chaudhuri, S., Dayal, U.: An overview of data warehousing and OLAP technology. ACM Sigmod record, 26(1), pp. 65–74 (1997)
2. Inmon, W. H.: Building the Data Warehousing. John Wiley & Sons (2002)
3. Imhoff, C., Galemmo, N., Geiger, J. G.: Mastering data warehouse design: relational and dimensional techniques. John Wiley & Sons (2003)
4. Kimball, R., Ross, M.: The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons (2011)
5. Wilson, N. B., Edgar, C. S.: Agile Methodology for Modeling and Design of Data Warehouses-AM4DW. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 9(9), pp. 2132–2137 (2016)
6. Golfarelli, M.: From User Requirements to Conceptual Design in Data Warehouse Design. IGI Global (2010)
7. Golfarelli, M., Rizzi, S., Turricchia, E.: Modern software engineering methodologies meet data warehouse design: 4WD. In: Data Warehousing and Knowledge Discovery, pp. 66–79 (2011)
8. Collier, K.: Agile analytics: A value-driven approach to business intelligence and data warehousing. Addison-Wesley (2012)
9. Chow, T., Cao, D.: A survey study of critical success factors in agile software projects. Journal of systems and software, 81(6), pp. 961–971 (2008)
10. Dyba, T., Dingsoyr, T.: Empirical studies of agile software development: A systematic review. Information and software technology, 50(9), pp. 833–859 (2008)
11. Boutkhoum, O., Hanine, M., Tikniouine, A., Agouti, T.: Integration approach of multicriteria analysis to OLAP systems: Multidimensional model. In: Computer Systems and Applications (AICCSA), ACS International Conference on IEEE, pp. 1–4 (2013)
12. Deshpande, K., Desai, B.: Model for Assessment of Agile Methodology for Implementing Data Warehouse Projects. International Journal of Applied Information Systems, 9(5), pp. 42–49 (2015)
13. Moody, D., Kortink, M.: From enterprise models to dimensional models: a methodology for data warehouse and data mart design. DMDW, pp. 5 (2000)
14. Corr, L., Stagnitto, J.: Agile Data Warehouse Design: Colaborative Dimensional Modeling, from Whiteboard to Star Schema. UK: Decision One Press (2011)