The 4th International Conference on Electrical Engineering and Informatics (IC EEI 2013)

User Requirement Analysis in Data Warehouse Design: A Review

Nur Hani Zulkifli Abai\textsuperscript{a,\ast}, Jamaiah H. Yahaya\textsuperscript{b}, Aziz Deraman\textsuperscript{c}

\textsuperscript{a}Computer Center, University Utara Malaysia, Kedah, Malaysia
\textsuperscript{b}Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Selongor, Malaysia
\textsuperscript{c}Faculty of Science and Technology, University of Malaysia, Terengganu, Terengganu, Malaysia

Abstract

User requirement analysis is crucial in data warehouse design. It effects almost every decision throughout implementation of data warehouse or business intelligence system. Currently, there are no standardization of user requirement analysis approaches had been outlined and this leads to complexity in data warehouse design. Various requirement analysis approaches have been proposed. However, data warehouse designer unable to choose which approach most suitable to be adapted in their project. This paper review various user requirement analysis approaches that been classified into four categories; data-driven approach, user driven approach, goal driven approach and mixed driven approach. It also outlines their strengths and weaknesses within different contexts. The review enables readers to identify appropriate user requirement analysis approach suitable for their projects.

© 2013 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of the Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia.

Keyword: Data Warehouse Design; User Requirement Analysis

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: nurhani.ukm@gmail.com
1. Introduction

The past few decades had shown explosive of products and services related to data warehousing[1][2]. This is as consequences of most large companies worldwide had established Data Warehouse (DW) as one of their main component in information system environment. DW has become the standard tools and essential component of decision support system [3],[4]. In addition, it has been successfully deployed in many industries such as manufacturing, retail business, financial services, transportation, telecommunication, utilities and healthcare [5] to support decision making. DW is built to improve decision making system[6][7] by providing consistent, timely, subject oriented information at the required level of detail [8]. It produces concise analysis to assist decision maker and increases organization corporeal performance.

DW is a collection of decision support technologies, aimed at enabling knowledge worker to make better and faster decision [5]. It is the heart of Business Intelligence (BI) System and contains all the information integrated from heterogeneous source into multidimensional schema to enhance data access for analysis and decision making. Data from various sources go through tremendous Extract, Transform and Loading (ETL) process that clean, transform and modify data so that could match multidimensional schema in DW. High volume of data from multiple sources causes high probabilities of errors and anomalies. Therefore, data needs to be transformed, scrubbed and audited before they are loaded into DWs.

There are unique criteria in DW that differentiate it with database management system (DBMS). DW is a subject oriented, integrated, time varying and non-volatile collection of data[7]. It contains historical data and those stored data will not be either deleted or changed. Analyzing historical, summarized and consolidated data could lead to precious information such as trend and behavioral correlation. The uniqueness of DWs features cause differences in DW modeling and design. For example in DBMS, redundancy should be eliminated to enhance business process while in DW the existence of redundancy is required to enhance DW performance. However, keeping all historical data that had rarely changed is costly since it requires big data space. To encounter this problem, Kimball et al (2008) proposed slowly changing dimension method to increase its efficiency [9].

The unique features of DW had exhorted different DW methodologies and design. In DW lifecycle, one of most critical phase that might influence all other phases in DW development is the Requirement Analysis phase. It is a process of gathering requirement from end user. Since DW end users might vary, it is important to identify accurate end users that will represent requirements in different point of views. The predominant objective of this phase is to identify organization goals and elaborate requirements that could measure organization performance. However, DW projects are inherently risky and some of them failed in the implementation. Studies shows that more that 80% of DW project fail to fulfill business goals [10],[11].

There are few reasons that contribute to its failure; some of them are related to user requirement analysis. Kimball et al.(2008) stated that requirement definition phase is paramount and impact almost every decision in DW project [9]. Most of failed DW projects execute poor requirement definition phase and some of them skip this phase [12]. They tend to focus on technical issues and not paid enough attention to requirement analysis [10]. Groggy method of conducting requirement definition analysis will lead to unstable design process [13] and poor communication between IT people and decision maker [3][10]. With lack of higher-level requirements vision [12], it is difficult to ensure the project will meet business goal. Furthermore, incomplete or inconsistence requirement lead developer to incorrect assumption [14]. In other hand, with tremendous amount of users requirements from different levels and categories cause big challenge to specify important requirement that could lead to business goal [13].

DW itself can be discussed in many aspects, however in this paper we will concentrate on user requirement approach in data warehouse design. This paper had been organized as follows. Section II presents the user requirement analysis approach. Section III discusses the advantage and disadvantage of available approach. It also discusses comparison of proposed mixed approach towards organization environment, technical team skills and analysis outcome. Finally, Section IV present conclusion and future work on user requirement analysis approach for data warehouse.
2. User Requirement Analysis Approach

Requirement analysis is one of the important tasks to ensure successful data warehouse project[9],[10],[15]. It collects and restructures base information that establish DW design and development of front-end application [13]. It also analyzes business requirement to assist estimation of project scope and leads to ensonce project planning. Sorting out requirement analysis phase properly and systematically will ensure effective DW implementation [16]. Basically, user requirement analysis approach can fall within four categories: data-driven, user-driven, goal-driven and mixed-driven approaches.

Data-driven approach, also known as supply-driven approach applies bottom-up technique[1], [7]. It was introduced by Inmon (1996) with focus on underlying operational data source as a basis to established scope and functionality of DW [6],[14]. Data-driven approach starts with identifying all available data within transactional data source and analyses it in order to generate data mapping. Data issues pertaining data overlapping, aligning ambiguous terms, and code mapping will be sorted out to ensure mapped data can be reengineered to build logical data schema. User requirements is determined by selecting portion of data available in data mapping [14]. Since main role in eliciting requirements plays by database administrator, it diminishes business user involvement.

On the other hand, user driven approach determines information requirement with stress involvement of business users[8],[13],[14],[1]. Similar with data driven approach, it applies bottom up technique that collect information from specific resources, compiles and restructures it before validate with business goals. This approach starts with determining requirements of different business user at tactical level. To elicit information requirements, project manager has to deal with approaches that facilitate user participations[1]. All user requirements shall be documented properly before integrate for DW requirements. Integrated information requirements will be mapped to available data source before designing logical data schema.

The third approach is Goal-Driven approach that applies top-down technique. It requires high level of top management involvement and focuses on their needs to align data warehouse with corporate strategy and business objectives [1]. Goal driven approach starts with organizing top management interview. They will specify business goals and determines its priority. Each business goal will trigger business questions on how to measure its achievement. Prioritizing business questions will assist to explicitly define most important business question that will determine data elements and hierarchies. In order to design logical data schema, it is important to identify measurements characteristic of each business questions. This information then could be transformed into data model. Kumar et al.(2009) proposed early requirement model that apply goal driven approach[17]. However they had expand goal driven ideas by include the important of addressing agent in user requirement analysis phase.

These three basic approaches have their strengths and weaknesses that will be discussed in Section III. However, to strengthen user requirement analysis approach, some Mixed-Driven Approaches being proposed. Giorgini et al.(2005) proposed a combination of goal driven and data driven approach called GRAnD[13]. The whole process of defining user requirement is divides into three main phases that are organizational modelling, decisional modelling and mixed design framework. It starts with understanding organizational objectives by interviewing stakeholders and identifying fact and attribute from their objectives. Then, in second phase, decision maker will be interviewed to get picture of business goal from their point of view. These goals will be extended to get fact, dimension and measurement. The final phase is to map both rationale diagrams with data source and enhance it with hierarchy construction. The finding will go through final refinement process that will rearrange conceptual schema to ensure it meet user’s need.

Other approaches had mixed goal driven and user driven approach [3],[10]. Gam & Salinesi(2006) introduces CADWA approach [3] that starts eliciting requirements by studying informal needs expressed by decision makers and summarize it into strategic plan. End user will assist in mapping user requirement and extend it to more detail action plan to achieve macro and micro goals. Final plan will then be used to build conceptual model of Data Mart (DM). On the other hand, Mazon et al.(2005) classify goals into three levels; strategic, decision and information goals[10]. Strategic goals are subjective business goal from broad view of top management. It will be detailed out by end user into decision goals that are more specific and later into information goals that will distinguish which specific information could lead to business goal.

An iteration of mixed approach that combine user driven and data driven approach had been proposed by Jukic & Nicolas(2010)[14]. User requirement phase starts with established end user analytical needs and consider its implication on DW. User requirements model will then be modified with neglecting unsupported requirements by
available data and adding potential requirements that available from data source. The acquisition will then be discussed with end user and this process will be iterated until end user satisfies with the outcome.

Another mixed approached that combined user driven and data driven suggested by Winter & Strauch (2004) [8]. It starts with identifying target user and dominant application that could feed information to DW. The end user will specify their requirements and then arrange it in priority sequence. This process will continue repeatedly until end user satisfied with the outcome. The main difference between this approach and approach proposed by Jukic & Nicolas (2010) [14] is the stage of iteration process. Finally, the requirements will be mapped with available data in data source and been presented for priority review.

Kaldeich & Oliveira e Sá (2004) incorporates business process within his approach with combination of data driven, user driven and goal driven approach [4]. The approach starts with analyzing the source data and produces the Entity Relationship (ER) diagrams. This approach investigates the business process and produces business process for current process (AS-IS model). In this approach end user will be interviewed and will produce a TO-BE business process model. Then the models (AS-IS and TO-BE models) will be integrated to form a new Integrate-Process-Diagram (IPD). The IPD is used as comprehensive DW requirements and will be verified through interview with top management to ensure it could lead to achieve business goals.

Another mixed driven approach that combine all three approaches is triple driven approach[1]. However, this approach proposes to execute all three basic approaches at same time and integrate all the result at final stage. Using goal driven approach as guideline, project manager determines business goals. To ensure successfulness of DW implementation, project manager develops organization corporate strategy and identifies main business field within the organization. Top management then identifies Key Performance Indicator (KPI) that could help measure its achievement. At the same time at data driven stage, database administrator should identify data source that relevant with DW project and analyze it to produce subject oriented data schema. On the other hand, eliciting comprehensive requirements by interviewing end user and analyze business reports should be conducted. All documented information within all three stages is then merged to produce logical data schema for DW.

3. Discussion

It is paramount to choose the right user requirement approach that might be suitable to organization environment. Recognizing the strengths and weaknesses of each approach could assist the process of selecting suitable user requirement analysis that might contribute to successful data warehouse project. There are few advantages in applying data driven approach. First and foremost it could simplify ETL design [13], [14], [18]. Since the requirements selection are based on data available in data source, it is easy to construct data warehouse schema [8] and at the same time ensure data availability. Data source schema is normally stable and does not change repeatedly. This will establish high stability in multidimensional schema [18] compare with requirements based on end user needs. However there are few weaknesses of data driven approach. This approach is high risk of wasting resources to handle unneeded information structure [4],[8]. With abundant of information from data source, it is almost impossible to extract relevant requirements [10].This will diminish end user motivation to participate in design process especially when they need to work with large data model[4]. In addition, it is possible that the result of multidimensional schema did not fit user requirements [18] due to lack of end user involvement.

The user driven approach gives priority to user requirements [14] and encourages high end user participation during DW design process [8]. With this approach, DW is tailored exactly to end user’s need [13] and it is highly appreciated by business user [18]. It minimizes risk of wasting resource [8] as a result of mapping requirements with certain data available in data source. On the other hand, this approach requires large effort in designing ETL [13], [14] and difficult in mapping user needs onto available data source[18]. Failure in mapping user needs will lead to user disappointments. In addition, it is a risk that end user normally not capable to specify requirements that could assist in achieving business goals because they don’t have sufficient knowledge or understanding of whole business process and organizational visions. However, to ensure successfulness of eliciting and integrating user requirements, great effort from project manager with good moderation and leadership skill is required. Alternatively, it is suitable to apply goal driven approach that will focus on goals and strategies of the company for an efficient decision making. This will maximize probability of correct identification of relevant indicator. However, it fully depends on the willingness of top management to involve in DW design process.
Table 1 reports comparison between mixed-driven approaches to organization environment, technical team and analysis output that involve in DW project.

Table 1 : Comparison of mixed driven approaches impact to organization environment, technical team and user requirements analysis output.

| Mixed Approaches | GRAnD [13] | CADWA [3] | Mazon et al. [10] | Jukic & Nicolas [14] | Winter & Strauch [8] | IPD [4] | Triple Driven Approach [1] | DWARF [12] |
|------------------|------------|-----------|-------------------|----------------------|----------------------|--------|---------------------------|-----------|
| Top Management Involvement | High | High | High | Low | Low | Moderate | High | High |
| End User Involvement | Low | High | High | High | High | High | High | Moderate |
| Notation / Diagram | Adopt Tropos notation | Not define | Adapt i* technique | Not Define | Not Define | Adapt ERM Model | Not define | Adapt UML |
| Time Consume | Moderate | High | Moderate | High | Moderate | High | Moderate | Moderate |
| Modeling Technique Complexity | High | High | Low | Moderate | High | High | Low | Moderate |
| Fulfill business goal Schema Quality | High | High | Moderate | Low | Low | Moderate | High | High |
| Quality | High | High | High | Moderate | Moderate | Moderate | High | Moderate |

Organization environment considered in this study are divided into level of involvement among top management and end user to provide user requirements. This is important characteristic to be considered in choosing appropriate approach because it will be the main input to DW project. Another consideration is about time frame of project. Most of DW projects require long term planning. However, it depends on the organization requirements and failure to deliver project on time could draw millions of dollars. Second important consideration is the technical team skill. They should be able to deal with diagram notation and modelling technique complexity. Finally is the output of approach that being classify into DW schema quality and ability to fulfil business goal. The main consideration in choosing suitable approach is the characteristic of organization itself. For example if the organization itself gets high commitment from top management but not from end user, it is suitable to choose GRAnD approach that could lead to high quality of data warehouse schema and high probability in fulfilling business goals.

4. Conclusion

In this paper, we have presented previous works in user requirement analysis in data warehouse design domain. The approaches for analysing user requirement have been surveyed and their strengths and weaknesses have been discussed. As a result, all basic approach has its own strengths and weaknesses. In order to avoid the drawbacks of single approach, few mixed-driven approaches had also been proposed. However, choosing appropriate approach should be done by tailoring it to organization environment and developer’s skill. This paper compares mixed-driven approach in user requirement analysis and this will guide data warehouse designer in choosing the right and appropriate approach so that will enhance the success rate of data warehouse implementation.

References

[1] Y. Guo, Sh. Tang, Y. Tong, and D. Yang. Triple-Driven Data Modeling Methodology in Data Warehousing : A Case Study. DOLAP '06: Proceedings of the 9th ACM international workshop on Data warehousing and OLAP; 2006. p. 59–66.

[2] J. Hammer, M. Schneider, and T. Sellis, “Data Warehousing at the Crossroads,” in Manifesto of a Dagstuhl Perspectives Seminar, 2004.
[3] I. Gam and C. Salinesi. A Requirement-driven Approach for Designing Data Warehouses. In Requirements Engineering: Foundation for Software Quality 2006; i:1–14.
[4] C. Kaldeich and J. Oliveira e Sá. Data Warehouse Methodology: A Process Driven Approach,” Advanced Information Systems Engineering; 2004. p. 1–16.
[5] S. Chaudhuri and U. Dayal. An Overview of Data Warehousing and OLAP Technology SIGMOD Record 1998; 26(1): 65–74.
[6] W. H. Inmon. The Data Warehouse and Data Mining 1996; 39(1): 49–50.
[7] W. H. Inmon. Building the Data Warehouse. 5th Edition. John Wiley & Sons; 2005.
[8] R. Winter and B. Strauch. Information requirements engineering for data warehouse systems. Proceedings of the 2004 ACM symposium on Applied computing ; 2004.
[9] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, and B. Becker. The Data Warehouse Lifecycle Toolkit. Indianapolis: Wiley Publishing, Inc; 2008.
[10] J.-N. Mazon, J. Trujillo, M. Serrano, and M. Piattini. Designing Data Warehouses: From Business Requirement Analysis to Multidimensional Modelling. 1st International Workshop On Requirements Engineering For Business Need And It Alignment; 2005. p. 44–53.
[11] J. Schiefer, R. M. Bruckner, and B. List. A Holistic Approach For Managing Requirements Of Data Warehouse Systems. Eight Americas Conference on Information Systems; 2002. p. 77–87.
[12] F. Ri. S. Paim and J. F. B. de Castro. DWARP: An approach for Requirements Definition and Management of Data Warehouse Systems. Proceedings of the 11th IEEE International Requirements Engineering Conference; 2003. p. 75–84.
[13] P. Giorgini, S. Rizzi, and M. Garzetti. GRaND: A Goal-Oriented Approach to Requirement Analysis in Data Warehouses. DOLAP ’05: Proceedings of the 8th ACM international workshop on Data warehousing and OLAP; 2005. p. 1–31.
[14] N. Jukic and J. Nicholas. A Framework for Requirement Collection and Definition Process for Data Warehousing Projects. Proceeding of the International Conference on Information Technology Interface; 2010. p. 187–192.
[15] S. Rizzi, A. Abell, J. Lechtenb, and J. Trujillo. Research in Data Warehouse Modeling and Design: Dead or Alive?; 2006. p. 3–10.
[16] B. List, R. M. Bruckner, K. Machaczek, and J. Schiefer. A Comparison of Data Warehouse Development Methodologies Case Study of the Process Warehouse. DEKA 2002. p. 203–215.
[17] M. Kumar, A. Gosain, and Y. Singh. Agent Oriented Requirements Engineering for a Data Warehouse ACM SIGSOFT Software Engineering Notes. ACM SIGSOFT Software Engineering Notes 2009; 24(5): 3–6.
[18] M. Golfarelli. From User Requirements to Conceptual Design in Data Warehouse Design – a Survey. Data Warehouse Design and Advanced Engineering Applications: Methods for Complex Construction; 2010: 1