A Comprehensive Analysis of Requirement Engineering Approaches in a Data Warehouse Environment

Geetika Prakash a, Kanika Talwar b * and Neelam Oberoi c *

a Maharishi Markandeshwar University, Sadopur, Ambala - 134007, India
b Panipat Institute of Engineering and Technology, Haryana - 132106, India
c Maharishi Markandeshwar University, Sadopur, Ambala - 134007, India
geetikaparkash@gmail.com, kanika.ncce@gmail.com, neelamoberoi1030@gmail.com

Abstract—Data warehouses contain vast amount of data generated from different sources. In order to maintain data in the data warehouse according to the organization some designing approaches are important. In data warehouse systems, Requirement Engineering (RE) is a crucial part of the software development process. Requirement Engineering is a sub discipline of Software Engineering, which is an important and vital phase in the overall software engineering life cycle. Software as a final product is deemed to fail if it does not fulfil the needs of its users. Several authors have proposed various approaches, which take part in analyzing information. In this paper we have provided a comprehensive comparison of different approaches proposed by various researchers based on certain features.

Index Terms—Data warehouse, requirements analysis, data warehouse design, requirement engineering.

I. INTRODUCTION

Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, and analyst) to make better and faster decisions. A data warehouse is defined as a subject-oriented, integrated, time variant, non-volatile collection of an organization’s digitally stored data that augments the decision making body within the organization [14]. It is a repository of consistent historical data that can be easily accessed and utilized in order to assist reporting and trend analysis.

In general, a Data Warehousing is constructed with the goal of storing and providing all the relevant information that is generated along the different databases of an enterprise.

Besides, requirements stated by the various stakeholders and developers frequently change owing to numerous reasons like ambiguous or insufficient requirements, changes in requirements in later stages of data warehouse environments, generation of new requirements due to technological advances, updates, etc. From a technical point of view, Requirements engineering is the initial step of software development activity in which the requirements from the customer are elicited and documented. According to Zave [18], “Requirements engineering is the branch of software engineering concerned with the real world goals for functions of and constraints on the software systems. It is also concerned with the relationship of these
factors to precise specifications of software behaviour and their evolution over time and across software families.” Requirement must be measurable, testable, related to identified business needs or opportunities, and defined to a level of detail sufficient for system design. The requirement engineering stage is divided into two phases: early requirement engineering phase and late requirement engineering phase [16], [17]. There are various Data Warehouse Requirement Engineering (DWRE) approaches, which have not distinguished the early requirement engineering phase from the late requirement engineering phase. Requirement can be classified as functional or non-functional requirements. In context of a DW, functional requirements specify what data is to be stored in it. Functional requirements are stated by the users in context of reports, views and results. Non-functional requirements specify how the information should be provided to facilitate reporting and analysis efficiently [15]. Non-functional requirements basically focus on scalability, reliability, feasibility, throughput, etc. of the DW system, which are specified by the programmers or developers involved in DW design and evolution process. These requirements also provide information on various constraints under which data is used and analyzed.

In the literature, several authors [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12] have proposed different types of approaches, which take part in analyzing information. In [2], requirement engineering method for data warehouse systems is proposed. The authors using method engineering approach, a methodology that supports the matching information requirements with actual information supply. In [4, 6] OORE approaches are defined using UML Meta model. The goal of this paper is to present a survey of efforts done by various researchers in context of requirement engineering techniques.

The paper is organized as follows. Section 2 discusses various kinds of approaches/techniques proposed by various authors. Section 3 presents the comparative analysis of the related work in a tabular manner based on certain features. Lastly, Section 4 gives conclusion.

II. STATE OF ART

The below sub-sections discuss about various approaches proposed by various authors.

A. Robert winter, Bernhard [2004]

In this paper, requirement engineering method for data warehouse systems is proposed. The authors using method engineering approach, a methodology that supports the matching information requirement with actual information supply (discussed in table 1). Method Engineering approach must be also characterized according to the driving factors or criteria used to swing methods.

<table>
<thead>
<tr>
<th>Method engineering approach</th>
<th>It is the field of information systems is the discipline to construct new methods from existing methods.</th>
</tr>
</thead>
</table>

B. Christopher, Dr. Armin [2001]

This paper [2] presents the RETTM approach to evaluating the applicability of RE methods and tools for use in time-to-market projects. This approach examines the reduction of schedule time not within the RE stage of a project concentrates on the definition of multidimensional hierarchies. The authors present two RE methods (described in table 2), Joint Application Development (JAD) and Quality Function Development (QFD).

C. Naveen, Anjana [2003]

In this paper, using demand driven approach try to identify the information needs to be met by the DW. The authors [3] present Goal-Decision-information (GDI) model. The authors organize the process of DW development in four stages.

The Goal-Decision-Information (GDI) Model shows that there is an association ‘is influenced by’ between goals and decisions. This association identifies the decisions, which when taken can lead to goal satisfaction. The knowledge necessary to take decisions is captured in the notion of decisional information shown in Fig.1.

This GDI model does not explicitly capture the stakeholders of the organization and their dependencies for achieving the goals.
TABLE II: METHODS OF REQUIREMENT ENGINEERING

<table>
<thead>
<tr>
<th>Methods of RE</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Application Development (JAD)</td>
<td>It is a method whereby system stakeholders work together in facilitated group sessions to specify and perform preliminary development of a system.</td>
</tr>
<tr>
<td>Quality Function Development (QFD)</td>
<td>It is the analysis of requirements is explicitly built into the technique. It does not explicitly help the analyst determine the initial requirements.</td>
</tr>
</tbody>
</table>

---

Fig. 1. GDI (Goal-Decision-Information) model for data warehouse requirement engineering [3]

**D. Anandi, Dr. Anurag [2013]**

This paper [4] is a survey paper on various Object Oriented requirement engineering methods. Authors have defined Use-Case modeling of UML and properties of good object oriented model. Also they have discussed about several advantages and disadvantages of various approaches.

**E. Anirban Sarkar [2012]**

In [5] has proposed a Business object based requirement analysis framework for DW system, which is supported with abstraction mechanism and reuse capability. The framework is compromised of three phases namely, early requirement analysis phase, detailed requirement analysis phase and mapping phase. The proposed framework supports abstraction mechanism and reuse of different well defined elements used to realize the different business concepts of the domain.

**F. Rajni, Shweta [2012]**

In this paper, we make a comparative study of various approaches and techniques for DW design and propose an object oriented framework for the conceptual design of a DW. The authors have used UML in the design process as it has become a standard for object modeling during analysis and design steps of software system development.
G. Paolo, Rizzi, Garzetti [2005]
Authors addressed a goal-oriented approach to requirement analysis for data warehouses, based on the Tropos methodology and proposed a requirements analysis approach with two perspectives. They are organizational modeling, centered on stakeholders, and Decisional modeling, focused on decision makers.

H. Manoj, Anjana, Yogesh [2010]
In this paper, the authors propose an AGDI (Agent-Goal-Decision-Information) model (fig. 2) to support early and late requirements for the development of DWs. The proposed model also supports three interrelated modeling activities namely, organization modeling, decision modeling, information modeling. The proposed approach has been demonstrated for capturing requirements of a university data warehouse, as a case study. This AGDI model is used to carry out various modeling activities to capture the early and late requirements for a DW to support the decisional goals of an organization. The Agent may depend on another agent for goals to be achieved, decisions to be suggested and information to be provided. These dependencies among agents are called goal, decision and information dependencies respectively.

I. Paim, Jaelson [2003]
In this paper [9], the authors propose a technique DWARF, a Data WArehouse REquirement deFINition. This technique operated a deep change on the client developer paradigm. The authors present validate and manage DW requirements. This paper has discussed a proposed framework Phase Oriented is supported by OLAP tools and fact & dimension schema.

J. Azman, Abdullah, Norita [2010]
In this paper, the authors [10] have proposed the Requirement Analysis Method for ETL process (RAMEPs) (fig. 3) that utilize ontology with the goal-driven approach. It also includes schema and instance normalization techniques for mapping conceptual design to logical design.
The RAMEPs is based on the Tropos methodology that was developed from the well accepted i*conceptual framework of software development [17]. The aim is to provide the decisional information from the perspective of organizational, decision maker, and developer.

K. Vishakha, Anupam, Amit, Shweta [2011]

In [11], various techniques have been proposed such as GORE, AORE, MORE, SCENARIO based etc. each handling requirements in a different way. In this paper, we have combined scenario based approach and UML (Unified Modeling Language), and proposed a model for requirement engineering of a data warehouse, which will help in its design process. Authors have proposed a model, UREM, based on scenario-based requirements engineering approach and benefits of using such an approach.

L. Jose-Norberto Maz´on, Jes´us Pardillo, and Juan Trujillo [2007]

In this paper, authors proposed a GORE approach for modeling organizational goals that the DW supports and relating them to information requirements. GORE approach is integrated into a model driven architecture (MDA) framework for the development of DW. This framework is based on defining a computation independent model (CIM) which addresses goals and information requirements.

M. Giorgini,P., Rizzi,S., & Garzetti, M (2007)

In this paper, authors proposed GRAnD, a goal-oriented approach to requirement analysis for data warehouses based on the Tropos methodology. Our approach can be employed within both a demand-driven and a mixed supply/demand-driven design framework. GRAnD adopts two different perspectives for requirement analysis: organizational modeling, centered on stakeholders, and decisional modeling, focused on decision makers.

III. COMPARATIVE STUDY

We have analyzed the various research works on several parameters and presented their comparison below in the table 3.

IV. CONCLUSION AND FUTURE WORK

In this paper, we make a comparative study of different approaches used for data warehouse design. In the literature, several authors [1,2,3,4,5,6,7,8,9,10,11,12,13] have proposed various types of approaches, which take part in analyzing information. Our comparative study is based on following criteria: Approach/Technique, Framework / Architecture, Model, Case study, Tool, Objective, Advantages and Disadvantages used.
<table>
<thead>
<tr>
<th>Features</th>
<th>Approach</th>
<th>Framework</th>
<th>Architecture</th>
<th>Model</th>
<th>Case Study</th>
<th>Tool Support</th>
<th>Objective</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>Technique</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robert Winter, Bernhard (2004) [1]</td>
<td>Method engineering Demand-driven</td>
<td>X</td>
<td>Method meta model</td>
<td>four case studies were used to derive all important components of a method.</td>
<td>X</td>
<td>Four case studies represent business practice for I.R.A</td>
<td>Information demand &amp; supply are synchronize in a two step.</td>
<td>Contradiction of proposing a comprehensive method.</td>
<td></td>
</tr>
<tr>
<td>Christopher, Dr. Armin (2007) [2]</td>
<td>RETTM</td>
<td>Two methods JAD, QFD</td>
<td>X</td>
<td>Stadish group CHAOS</td>
<td>X</td>
<td>Performance of an adequate amount of RE while requiring a minimal amount of effort.</td>
<td>Focus stakeholders on achieving acceptable solutions.</td>
<td>Participating in JAD session is limited to a few individual.</td>
<td></td>
</tr>
<tr>
<td>Naveen, Anjana (2003) [3]</td>
<td>Requirement Driven</td>
<td>X</td>
<td>GDI</td>
<td>X</td>
<td>X</td>
<td>Start from goals &amp; work its way to the decisional information.</td>
<td>Two kinds of relationship i.e. is satisfied &amp; is required for.</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Anirban Sarkar (2012) [5]</td>
<td>Business - Object</td>
<td>DWRA</td>
<td>GOOMD (Multi-dimensional)</td>
<td>Retail Organization</td>
<td>OLAP</td>
<td>Capture info. about a real world concepts, Operations &amp; relationship b/w them.</td>
<td>Set of BO can be reusable + Scalable, reliable.</td>
<td>Does not provide any guideline to move from req. model to high level conceptual design.</td>
<td></td>
</tr>
<tr>
<td>Rajni, Shweta (2012) [6]</td>
<td>X</td>
<td>Object Oriented framework</td>
<td>Various DW design approaches</td>
<td>UML</td>
<td>Various approaches on the basis of design criteria &amp; propose Object Oriented framework.</td>
<td>More adaptable as the user req. are constantly changing</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
All the requirement engineering approaches have their own merits and demerits. Our future work includes combining two or more techniques such as RETTM and GRanD approach in order to increase accuracy and reduce risk of failure.
REFERENCES