

Requirements Engineering for A Data Warehouse :

A Study of Managerial & Technical Aspects

Anil Kumar^{*1}, Anil K. Saini²

^{*1}Dravidian University, Andhra Pradesh, India

²Department of Information Technology, GGSIPU, New Delhi, India

ABSTRACT

This research paper focuses on how to prepare an effective data warehouse with the use of Requirements Engineering. Earlier operating systems used for storage of information, facilitated us in making only day-to-day decisions, because they provided us with only the current data, which cannot consider appropriate enough to involve significant conclusions. Thusly, it evoked the need of using Data Warehouse for storing all the information. Construction of efficient Data Warehouse and its maintenance involves extraction, cleaning and filtering of information etc. It is a costly and time-consuming procedure, because this has made in advance of the necessities to be met. To surmount this problem, such a technique should use, which will help in exploring and clarifying each requirement in particular having a meaningful consideration while constructing the Data Warehouse at an early phase. For this we have combined the technique 'Requirement Analysis' of 'Requirement Engineering' with the Requirement gathering techniques of Data Warehouse i.e. 'Interview' and 'Survey Method'.

Keywords: Requirements Engineering, Data Warehouse, Requirement Analysis, Requirement gathering techniques, ETL Process (Data Staging Process)

I. INTRODUCTION

Every organization deals with vast amounts of data, which is stored in different formats or patterns, developed on different platforms by different developers. Which in future becomes a part of different database structures? To form databases analysts or users have to do lot of struggle and hardwork, so that the databases can access easily. In this manner, organizations construct and maintain several projects, which provide valuable, readymade solidified information for the utilization of the clients. This procedure is required to execute and construct the Data Warehouse [1].

DATA WAREHOUSE

A data warehouse is a subject oriented, integrated, non-volatile and time-variant collection of data in support of management's decision.[2]

In other words, it is a system, which collects, integrates data by performing extraction, transformation, loading and storage of easily readable and accessed data for

making strategic decisions [3].

The information in the Data Warehouse is cleaned, changed and listed and is made accessible to administrators and different business experts for their utilization for gathering information in mining, online analytical, market research and decision support[4].

II. PROBLEM STATEMENT

Data warehousing initializes the procedure to get valuable data from different data sources; Thus they can clean, channelize, and change the data; and store the data in a structure otherwise which is difficult to get, understand, and utilize. The data is then utilized for interviews, broadcast, and data analysis etc.

This process is expensive, time consuming, and exceptionally tedious. Therefore, Data warehousing needs still a superior approach.

For storage, Data warehousing uses Decision Support Systems that enables decision making but the drawback in this systems is that they do not focus on the requirements in the early phase. These systems

discover the requirements usually in the later phase [5]. This further creates the issue of determining organizational requirements and this is the key research factor in this paper for Data warehouse development. Data Warehouse must be designed keeping in mind the requirements for decision makers in advance, which can derive data from data sources at the later stage according to their need. Designing of Data Warehouse consists of two categories:[6]

Supply-driven

It is also known as Data-Driven approach for designing Data Warehouse [7]. It deals with acquiring the data with detailed analysis from the data sources. In this, the data warehouse designer selects the portion of data or information, which is essential for analyzing and decision-making, which is still further structure according to the multidimensional model like ETL (Extract, Transform and Load) process, which is the simplest approach, but the user's requirements has not given much importance in ETL approach thus creating further problems [8].

Demand-driven

It is as well recognize as a Requirement Driven approach. It begins with determining the informational requirements of the users of the Data Warehouse. The trouble rises up in the later phase when mapping of these requirements is done in contrast to the available data sources, which makes it designing difficult [9].

Therefore, there needs to further develop the processes, which can help the decision makers to access data according to their needs and get satisfaction.

III. IMPORTANCE OF REQUIREMENTS ENGINEERING IN DATA WAREHOUSE

In Requirement Engineering, "Requirements" signify "what of a system", not "how" [6]. It is "what a system must be able to do. Requirements go on changing over time as the project moves from analysis to design until its implementation. These requirement types are - Functional, nonfunctional, Domain requirements depending upon the demands of client [10].

Requirement Engineering enables us to prepare a document format that describes what the system or the product will do without describing how it will perform. It aims at refining all the requirements of the system under construction wherever problem statements act as an input [11].

Requirements are defined during the early stage of system development. Therefore, Requirement Engineering plays a vital role when the Data Warehouse is under construction [12], that is, we can cater to the requirements first and then construct an effective data warehouse, which was not possible in earlier stages when decision support systems was simply practiced. This is because, while building decision support systems, requirements was taken into consideration at the end, which was a drawback for these systems to predict future requirements of the decision makers. [13]. But with the use of requirement engineering we can prepare a document of the present and future requirements of the decision makers by using the process which is as follows:

2.1 Requirements Elicitation

2.2 Requirements Analysis [14]

2.3 Requirements Documentation

2.4 Requirements Review

From the above process of Requirements Engineering we have considered and combined 'Requirement Analysis' with the Requirement gathering technique of Data Warehouse i.e. 'Interview'

IV. REQUIREMENT GATHERING TECHNIQUES [15]

- ✓ Interview
- ✓ JAD Session [16]

2.1 Interviews

Interviews are used to gather information from key stakeholders of a software project. It can be performed one to one or in small groups and can be formal interview or informal interview. The interview can be performed by anyone but a successful interview is the one where all the essential requirements can be acquired, which might be needed during the whole process and attaining such an interview involves arranging and booking, planning, opening, directing, winding up and pursuing in a timely and cost effective manner [17].

2.2 JAD Sessions

JAD stands for Joint Application Development. With JAD Sessions we can gather a great number of interested users to meet together in group sessions. It is a joint process that enables the users and the IT professionals to develop the application required. JAD is a five-phased approach consisting of: Project

Definition, Research, Preparation, JAD Sessions, and Final Document.

V. APPROACH ADOPTED: INTERVIEW

Interviews are used to gather information. Pre-interview research plays a very important role in moving an interview to a successful interview. Thus preparing an interview involves background domain research. This Pre-interview research includes history and current structure being followed in the business; the primary goal of the business; number of employees with their authorities and responsibilities and the location of the users categorized into- senior executives, departmental managers, analysts, and IT professionals. One can review organization reports to achieve a sight of the project scope, ideas and framework [18].

At this state Interviews can be Structured or Unstructured. In Structured Interviews, the expert creates particular arrangements of inquiries before the interviews and Unstructured Interviews look for an expansive and generally characterized set of data. Structured interviews are regularly a superior approach as they compel the formalization of the interview procedure. The inquiries in the Interview can be-Close Ended inquiries, Open Ended questions or Probing questions. Close ended inquiries require a particular answer, Open Ended questions, leave space for elaboration on the purpose of the interviewee and Probing questions follow up on what has quite recently been talked about so as to take in more.

These Interviews can be one-to-one or two or three persons can be interviewed at a time.



Figure 1. One-On-One [16]



Figure 2. Group Interview

An Interview can follow Top-Down Approach or Bottom up Approach. On Top Down Approach the interviewer begins with expansive, general issues and step by step progresses in the direction of more particular ones though in Bottom up Approach, the interviewer begins with certain inquiries and moves towards the more important issues.

Now if we talk about how Interview-Requirement Gathering Technique helps in the maintenance of a Data Warehouse, we have a very important reason to opt for this technique is that an interview helps in exploring and clarifying each requirement in detail. This can only be obtained by organizing a conversational interview involving a question and answer session with stakeholders to gather information about necessities and specifications. An interview may include one or more stakeholders who help in discovering needs and the high-level requirements. The detailed requirements can be derived from these needs easily. Interviews also help in acquiring the approval from stakeholders in respect to their needs, requirements or any other available information.

Most important things regarding the data warehouse point of view is that, with the help of Interview all the requirements from the stakeholders could discover and obtained at an early stage, that is, when the Data Warehouse is constructed. All the detailed requirements of all the important persons in the organization from top level to lower level could catered in this Interview technique, which ultimately helps in defining the objectives of the organization, and this is the primary job. Though the requirements are gathered in the early phase, still it requires further cleaning, filtering, and transformation of the data which gives the benefit that the data can now be explored and processed with less cost and less time, thus, saving the two most important resources of the organization further helping in achieving the organizational objectives.

Formal Interview Process Steps

1. Distinguishing stakeholders to be interviewed.
2. Acquiring a general comprehension of the client's business.
3. Creating interview questions utilizing open-ended inquiries.
4. Setting for gathering time and area for the interview.
5. Providing an arrangement of questions to interviewees before the interview (So that they should be able to plan the interview).
6. Using at least one Recorder to exactly protect recordings of the interview for analysis.
7. Give the results to interviewees to for confirmation of the content.

Informal Interview Process Steps

1. Recognizing stakeholders to be interviewed.
2. Getting a general comprehension of the client's business.
3. Creating interview questions (for interviewer's utilization just) to ensure that certain questions are replied amid session.
4. Setting up an easygoing gathering or phone discussion time for the interview.
5. Taking manually written notes amid the interview; abstain from utilizing electronic information catch.
6. Giving results to interviewee to confirmation of the content.

Advantages

1. It is a less complex approach with lesser effort.
2. Interviews of people and tiny groups require less effort and preplanned work than expansive workshops.
3. Interviews of people and tiny groups require less partner responsibility than expansive workshops.
4. Interviews give a chance to investigate or specify topics in more detail.
5. With the assistance of Interview, need of the stakeholder can be obtained at the earlier stage, that is, the point at which the Data Warehouse is developed.

VI. SOLUTION APPROACH

While working towards the solution i.e. Building an efficient Data Warehouse we came across many techniques and found Interview technique as best suitable technique for solving this problem. Key facts that should be considered while applying this technique

is that the questions asked in the interview should not reflect the interviewer's pre imagined ideas, as it can influence the responses. During the Interview, closed questions should not be asked and stressed upon because it limits the input with respect to the subject. Closed questions are to the point and can be addressed rapidly by the client without giving any foundation or context to the interviewer. In fact, Open-ended questions are advisable they do not deliver to the point answers and carry fewer constraints upon the subject's response, thus are most useful in identifying the scope of the problem domain. For projects with an expansive number of stakeholders, the interview strategy may not work productively for the variables in particular - kind of data, profundity of data, and broadness of data, joining of data, client contribution, and cost. To overcome this drawback, Interview should not be the sole requirements gathering technique for a project. To make it an efficient tool, it should complement with Survey approach. The survey can constrain clients to choose from decisions, wrote something ("Agree Strongly, Agree... "), or have open-ended inquiries permitting freestyle reactions thus providing qualitative guidance for characterizing the market and achieving the objective appropriately and efficiently [19].

VII. CONCLUSION

This research project focused on the drawbacks of operational systems and the need of Data Warehouses. The project also highlighted the problems that were earlier faced when data warehouses were built without taking into consideration "requirements" during the early stage of the development. Requirements Engineering for the data warehouse aims to identify the informational needs of the decision-makers thus saving the cost, time which was incurred before, in extraction, cleaning and filtration of the data needed, when decision support systems were used. Here, Interview has been used as Requirement gathering technique as it helps in exploring and clarifying each requirement in detail which can be considered while constructing the Data Warehouse in the early phase. With this research, we have tried to give the solution approach to overcome the drawbacks that came in front and also opened the path for all those researchers who might find some new techniques which can result in an efficient building of Data Warehouses.

VIII. REFERENCES

- [1]. "DKMS Briefs No. Six: Data Warehouses, Data Marts, and Data Warehousing" by Joseph M. Firestone, 24 May 2002.
- [2]. Paulraj Ponniah, "Data warehousing fundamentals", By John Wiley & Sons, Inc., 2006.
- [3]. Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM Sigmod record*, 26 (1), 65-74.
- [4]. Mohamed Frendi and Camille Salinesi, "Requirements Engineering for Data Warehousing", France, 2003.
- [5]. The Data Administrator Newsletter-TDAN.com, Robert S. Seiner Publisher
- [6]. Schiefer J., List B., Bruckner, R., "A holistic approach for managing requirements of data warehouse systems". In *Proceeding of Americas Conference on Information Systems*, 2002.
- [7]. A Model-Driven Goal-Oriented Requirement Engineering Approach for Data Warehouses, Dept. of Software and Computing Systems, University of Alicante, Spain
- [8]. Salinesi, C., & Gam, I. (2006, June). A requirement-driven approach to designing data warehouses. In *Requirements Engineering Foundations for Software Quality (REFSQ'06)* (p. 1).
- [9]. Salinesi, C., & Gam, I. (2006, June). A requirement-driven approach to designing data warehouses. In *Requirements Engineering: Foundations for Software Quality (REFSQ'06)* (p. 1).
- [10]. Winter, R., & Strauch, B. (2004, March). Information requirements engineering for data warehouse systems. In *Proceedings of the 2004 ACM symposium on Applied computing* (pp. 1359-1365). ACM.
- [11]. Mercier-Laurent, E., Owoc, M. L., & Boulanger, D. (Eds.). (2016).
- [12]. Inmon, W.: *Building the Data Warehouse*. Wiley & Sons, Chichester (2002) [11] Gam, I., Salinesi, C.: A requirement-driven approach to designing data warehouses. In: *REFSQ (2006)*
- [13]. Ines Gam, Camille Salinesi, "A Requirement-Driven Approach for Designing Data Warehouses, France, 2006.
- [14]. Giorgini, P.; Rizzi, S.; Gazette, M. A. Siena, "Grand: A Goal-Oriented Approach to Requirement Analysis in Data Warehouses", 2008.
- [15]. "Software Requirements Engineering: Practices and Techniques", Ronald Kirk Kandt, JPL, California Institute of Technology, November 7, 2003.
- [16]. "Improvement to Requirement Engineering (JAD process) for Data Warehouses", arjesh Kochar; Nidhi Khurana; Akash Agarwal, INDIACOM, National Conference 2008.
- [17]. Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: interviews and focus groups. *British dental journal*, 204 (6), 291.
- [18]. Paetsch, F., Eberlein, A., & Maurer, F. (2003, June). Requirements engineering and agile software development. In *Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2003. WET ICE 2003. *Proceedings. Twelfth IEEE International Workshops on* (pp. 308-313). IEEE.
- [19]. Inmon, W. H. (2005). *Building the data warehouse*. John Wiley & sons.