International Journal Of Engineering Research & Management Technology ISSN: 2348-4039

IJERMT :mail: editor@ijermt.org March 2014 Volume 1, Issue 2 Website: ijermt.org

AGPODI: A New Conceptual Model For Designing Of Data Warehouses

Amit Kumar Goyal, Girish Kumar Sharma
KIET Group of Institutions, Ghaziabad, BPIBS, Govt. Of NCT, Delhi

ABSTRACT-
Creation and adoption of corporate policies requires significant vow of inadequate senior management for strategic decisions to be made. The absence of appropriate and timely strategic decision leads to the convergence upon a policy that may not be achieving the organizational goal in a timely manner. The appropriate and timely decisions for organizational profits to be made by keeping all business data in a well-situated and interactive way so that business executives and managers can take strategic decisions. Data warehouse is the only feasible solution. Data warehouses are decisional information systems that organizations embed for a strategic decision to be made. A Data warehouse is required for decision making and is a subject-oriented, integrated, time-variant, non-volatile collection of data. Building a data warehouse is a very challenging task. Various data warehousing design methodologies are there to support the growing market need. Development of a data warehouse is different from developing transaction systems because in case of data warehouse requirements keep on changing. While designing the data warehouse, the Conceptual design and requirement analysis are two of the key steps. In this paper, we have surveyed the literature that is available for designing the data warehouse and compared methodologies based on the various parameters and have proposed a new conceptual model AGPODI. This model can be used in early as well as late requirement elicitation.

Keywords: - Data Warehouse (DW), software development life cycle (SDLC), Requirement Engineering (RE), Goal decision information model (GDI), Agent Goal decision information model (AGDI).

I. INTRODUCTION

To extend market share and improve profitability, Companies are frantic for the strategic information to defy fiercer competition. For the decisions to formulate business strategies, the executives and managers, who are responsible for making the business competitive, need the information so as to take proper establish organizational goals, set objectives, and also monitor the results. All organizations have vast amounts of data but as the time passes it has been found increasingly difficult to access it and make use of it. The reasons behind this are that data is stored in many different formats, exists on many different platforms, and resides in many different file and moreover, the database structures developed by different vendors too. One has to write and maintain perhaps hundreds of programs that are used to extract, prepare, and consolidate data for use by many different applications for analysis and reporting. Also once initial findings are made, decision makers often want to dig deeper into the data. This would leads to the modification of the previously written programs or development of new ones as a when required. Both these approaches are costly, inefficient, and very time consuming. Data warehousing offers a better approach. The concept of data warehousing has evolved out of the need for easy access to a structured repository of quality data that can be used for decision-making. Basically, Data warehouse (DW) refers to technologies for collecting, refining, integrating, analyzing large volume of data to provide information that is required for managers to make better decisions. To support decision making organizations uses data Warehouses (DW) as a standard tool. DW is required for management decision making and is a subject-oriented, integrated, time-variant, non-volatile collection of data. Initially, DWs was there for

numerical facts and textual dimensions [1]. In recent years, however, Data Warehouses have been proposed for domains of image data [2] and voice data [3]. DWs are large amounts of data integrated from heterogeneous sources into multidimensional schemata which are optimized for data access so that the analysis of data is easier. A multidimensional schema is made up of facts, measures and dimensions. Facts are required for the decision-making process (such as sales, orders) and monitored through measures and dimensions. Measures are actual numerical KPIs (such as quantity of product sold, price of the products, etc.), while the dimensions are responsible for analyzing these measures (such as time, customer, product, etc.). Due to their specificities, Development of a data warehouse is different from developing transaction systems because in case of data warehouse requirements are always keep on changing. Most of the existing DW development approaches deal with how data should be structured, stored, and managed in DW. Thus the development of DWs is particularly complex and requires ad-hoc methodologies and an appropriate life-cycle. Requirement analysis and Conceptual design are the key steps within the DW design process. Earlier designer of data warehouses ignored these two, but they have received attention recently. In this paper we reviewed and compared several prominent data warehousing methodologies based on a common set of attributes in order to help the designer make crucial choices more consciously. Rest of the paper is organized as follows: related literature is discussed in section II and we have compared the various methodologies for designing the data warehouse based on various parameters. We have proposed a new conceptual model (AGPODI) in Section III, followed by our conclusions in section IV.

II.LITERARY REVIEW

A DW is required for management decision making and is a subject-oriented, integrated, time-variant, non-volatile collection of data. The classical SDLC does not work in the world of the DSS (decision support system) analyst as SDLC assumes that requirements are known at the start of the design whereas in the world of the DSS analyst, requirements are usually the last thing to be discovered in the DSS development life cycle. Development of a data warehouse is different from developing transaction systems, because in case of data warehouse requirements are always changing. Thus the design of DW is characterized by several complex factors which are when determined in the early stages of the design causes a high percentage of DW project failures. The complexity in designing the DW along with the past attempts give rise to the different design methodologies and thus, selection of appropriate methodology can increase the probability of completing the project successfully well in time along with fulfilling all the desired organizational goals. According to [6], the appropriate life-cycle for the DW can be selected by considering various points such as the DWs depend solely on the data present in the operational databases, While designing a DW, it is difficult to find out the user requirements; moreover, they usually keep on changing during the project development, the large and complex nature of DW projects, the average time for their construction is 12 to 36 months and their average is too high, and the requirements of top executives and managers keep on changing in an ad-hoc manner and they want the reliable solution for that. Based on these considerations the main phases for the DW life-cycle [4] can be summarized in “table I”.

Table I: life-cycle of a data warehouse

PHASE	DESCRIPTION
DW planning	Scope and the goals of the DW are identified, and how many data marts are to be implemented along with their order which depends on the business priorities and the technical constraints [4]. During the project planning phase the staffing of the project is also being carried out.

DW design and implementation	This phase will be repeated for each data mart to be implemented.
DW maintenance and evolution	The performance optimization must be periodically carried out due to change in user requirements. On the other hand, DW evolution [6] concerns keeping the DW schema up-to-date with respect to the business domain and the business requirement changes.

Earlier, DW design basically concern with the phases in "table II".

Table II: phases in designing the DW

PHASES	DESCRIPTION
Requirement analysis	In this phase, information that is relevant to the decisional process is identified either by gathering the user requirements or the availability of data in the operational data sources.
Conceptual design	This phase concentrate on designing the conceptual schema using the conceptual model.
Logical design	With the conceptual schema ready, the logical design phase creates a corresponding logical schema on the chosen logical model. most of the DW systems are based on the relational logical model (ROLAP), most software vendors are proposing also pure or mixed multidimensional solutions (MOLAP/HOLAP)
ETL process design	In this phase, the mappings and all the data transformations necessary to load into the logical schema of the DW are designed.
Physical design	This phase addresses all the issues related to the selection of tools for implementation – such as indexing and allocation

In past few years considerable efforts been made to define the Systems Development Life Cycle for data warehouse development [7] [8]. An approach based on the SOM (Semantic Object Model) process modelling technique [9] [13]. Requirements for data warehouse can be elicited using business process requirements and strategic decision processes [10]. They had suggested that models for data warehouse are produced by combining data warehouse requirements and as-is data models [5]. Further, data models thus created may used to elicit new requirements. According to [11], the development of data warehouse should not be started by analyzing the business processes only. Rather main concern should be to develop a DW system that exclusively supports decision processes. A different approach in which the problem is to extract data marts from the enterprise wide information system [12]. Goals are determined by a top down approach using Goal Question-Metric approach and they are aggregated and refined then in an abstraction sheets. These sheets are inputs for star schemata. Bottom-up approach is used for ER schemata of operational databases and extracts candidate star schemata. The ideal star schemata thus created are matched with these and ranking of candidate star schemata is done according to the metrics for selection. These goals and objectives identify the decision-making capability. The Supply-driven approach (also called data-driven) is a bottom-up technique that starts with an analysis of operational data sources in order to identify all the available data [7] [15]. Here user involvement is limited to select which chunks of the available data are relevant for the decision-making process. The user-driven approach is a bottom-up technique that starts from determining the information requirements of different business users [11] [14]. Their points of view are then integrated and made consistent in order to obtain a unique set of multidimensional schemata. The emphasis is on the requirement analysis process and on the

approaches for facilitating user participations. The problem of mapping these requirements onto the available data sources is faced only a posteriori, and may fail, thus determining the users' disappointment. Although this approach is highly appreciated by business users that feel involved in the design and can understand the rationale of the choices, it is usually time expansive since the business users at the tactical level rarely have a clear and shared understanding of the business goals, processes and organization. Consequently, this approach usually requires great effort by the project manager that must have very good moderation and leadership skills, in order to integrate the different points of view. The goal-driven approach focuses on the business strategy that is extrapolated by interviewing the top-management. Different visions are then analyzed and merged in order to obtain a consistent picture and finally translated into quantifiable KPIs [9]. This approach is typically top-down since by starting from the analysis of a few key business processes, their characterizing measurements are derived first and than transformed into a data model that includes a wider set of KPIs that characterize such processes at all the organizational levels. The applicability of this approach strictly depends on the willingness of the top management to participate to the design process and usually require the capability of the project staff in translating the collected high-level requirements into quantifiable KPIs. By reviewing the literature for designing the data warehouse, we tried to compare the design methodologies based on the parameters and try to find out the strengths and weaknesses of the methodologies as shown in "table III".

Table III: Comparison between various data warehouse design methodologies based on parameters

Parameters/ Approach	Data Driven / supply driven	Process driven	Goal driven	User driven
Prerequisite	Only Appropriate data must be available in the data base	Organization must have clear understanding of the need of DW and most importantly must be willing to participate in the design process	stakeholders external / internal must have clear goals for organization as well as must have a good knowledge of the processes	Project manager must have very good moderation and leadership skills, in order to integrate the different points of view of business users.
Role of Requirement Engineering	Very low	low	Very high	high
Permanence of Data model	A few	To model the new requirements	Permanent throughout the DW design	To model early requirements
Design Methodologies used	Bottom-up approach	Top-Down approach	Top -down approach	bottom-up
Participation of Controlling	Very Low; include only Data Base	moderate; includes from	High; mainly stakeholders	High; different business users'

stakeholders	Administrators as stakeholders	Top Management i.e. internal stakeholders to Users	external/internal to organization	requirements are identified and their point of views are integrated.
Risk associated	Implemented Data warehouse doesn't fulfil organizational goals and the business user's requirements.	Poor communication between business people and design team may lead to inappropriate design	External stakeholders and/or users may change so as to change their requirements thus changing the design.	Tactical users rarely have a clear and shared understanding of the business goals, processes and organization
Categories / levels	Only at Operational level	Conceptual and Strategic	Depends on categories and level of stakeholders participated in design process	business users
Technical knowledge Required	Low as it includes only Data warehouse designers	Restrained as it involves Business Analysts, Data Warehouse Designers	Highly technical skills as it involves Requirement Engineering Team, Business analyst, Data Warehouse designers	Low as only the business users having tactical knowledge are involved
Number of Tools Required	Very few	Restrained	Variety of Tools	Variety of tools such as questionnaire, interviews etc.
Degree of Devolution	Very Low	Low	High	Very high
Time Required	Less Time Required	Restrained	Requires considerable amount of time	time expansive
Cost Involved	Less	High Cost	High Cost	High cost

Therefore, it is proposed in that the identification of goals suggests the decisions that Influence the satisfaction of these goals. Implementation of decisions is done by actions associated with them [16]. Finally, from knowledge of the decisions, it is possible to identify the information that is needed to make the decision. Here, the product of requirements engineering is represented as a schema of the GDI model. The easy requirements modelling technique for a DW system represents good practices as requirements management [17]. This approach focuses on the communication between the stakeholders and users of a new data DW system and those who are building it. Furthermore, it helps to apply an effective requirements engineering method by the use of different perspectives for

representing DW requirements. The traditional requirements engineering process was used and proposed DWARF, a Data WArehouse Requirements deFinition technique [18] methodological approach for requirements definition and management of data warehouse systems. An AGDI model [19] for the requirements engineering of data warehouses was proposed and used to capture the early and late requirements of a data warehouse from stakeholders' perspective. The combination of requirement engineering techniques should be applied for efficient and successful requirement engineering process [20].

III. Agent Goal Process Oriented Decision Information Model (AGPODI)

The GDI and AGDI models have been emphasizing on considering agents, goals, decisions in the requirement elicitation process to reduce the risk of failure and make the developing data ware house as per user requirement, but both the models are not considering the DW requirements that can be elicited using business process requirements. DW development approach that takes into account both business process requirements, requirements from strategic decision processes, and operational data models of existing systems along with the external/internal stakeholders. Besides, we propose to combine DW requirements to DW models. On one hand, DW requirements can be elicited using both business process requirements (i.e. by analysing the usage of the future DW in the decisions made in To-Be business processes) and strategic decision processes and goals (i.e. by analysing the usage of the DW to make strategic decisions about change. The proposed AGPODI model is shown in "fig. I".

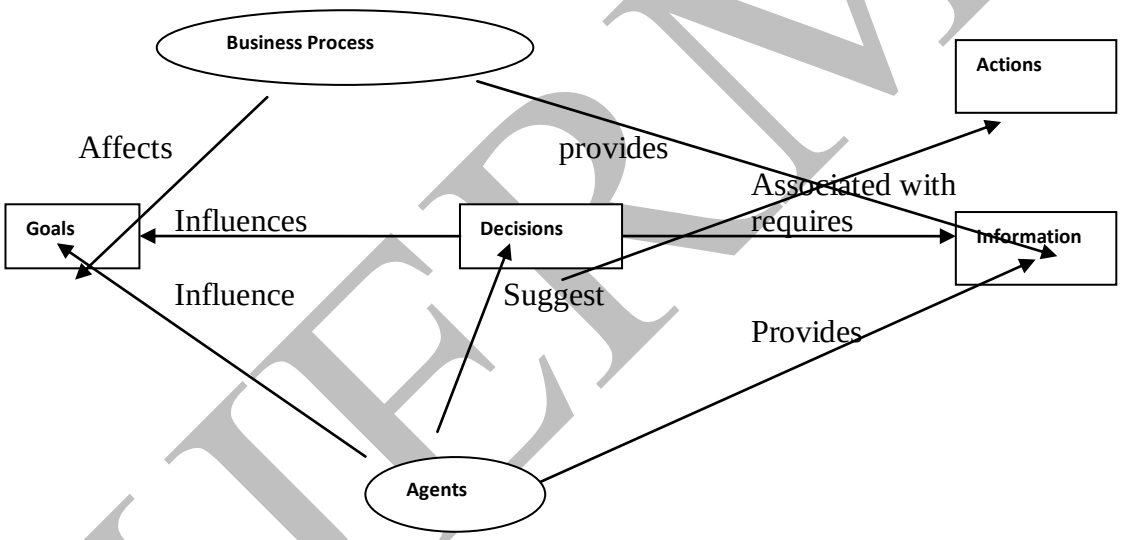

Figure I: Agent Goal Process Oriented Decision Information Model (AGPODI)

The GDI and AGDI models have been emphasizing on considering agents, goals, decisions and information in the requirement elicitation process but we are also considering the DW requirements that can be elicited using business process requirements.

IV. Conclusion

All large organizations have DWs to support their strategic and business decisions and now the development of Data Warehouse is not the new process. Our goal in this paper was double: (i) to compare the existing methodologies based on various parameters and (ii) to layout an innovative requirements-based DW development model. Our review says that although data driven approach is easier and cheaper to implement but risk of not meeting the business and user's requirement is there; Process driven method may not have that risk but it suffers from the unwillingness of top management to implement a data warehouse and poor communication between design team and

business team is there; though goal driven approach require highly technical skilled persons, variety of tools and time taking approach but we can say that the data warehouse design thus produced will meet the organization objectives and goals more precisely. The data-driven and the goal-driven development approach are not mutually exclusive. Rather they pursue different purposes so they both may co-exist for the development of data warehouse. One major issue that we met in all approaches is the lack of guidance of the requirements engineering part of the DW development process. We aim at an approach that is utilizing the RE in full to produce data warehouse indistinctively.

V. REFERENCES

1. Inmon W.H. ,"Building the Data Warehouse", John Wiley, New York (1996).
2. Wong, S.T., Hoo, K.S. Jr., Knowlton, R.C., Laxer, K.D., Cao, X., Hawkins, R.A., Dillon, W.P., Arenson, R.L. , "Design and applications of a multimodality image data warehouse framework", Journal of the American Medical Information Association, vol 9, issue3, page no 239–254, 2002.
3. Miller, L.L., Honavar, V., Barta, T, "Warehousing Structured and Unstructured Data for Data Mining",Proceedings of the ASIS Annual Meeting, vol 34, page no -215-24 , 1997.
4. Kimbal, R., Reeves, L., Ross, M., & Thornthwaite, W,' The Data Warehouse Lifecycle Toolkit". John Wiley and Sons, Inc., N.Y. 1998.
5. Ballard, C., Herreman, D., Schau, D., Bell, R., Kim, E., Valencic, A., "Data modeling techniques for data warehousing", http://www.redbooks.ibm.com/redbooks/pdfs/sg242238.pdf.
6. Giorgini, P., Rizzi, S., & Garzetti, M. "GRAnD: A goal-oriented approach to requirement analysis in data warehouses". Decision Support System, vol. 45(1), 4-21, 2007.
7. Golfarelli, M., Rizzi, S.,"Designing the data warehouse: key steps and crucial issues", Journal of Computer Science and Information Management, vol 2, 1999.
8. Husemann, B., Lechtenborger, J., Vossen, G., "Conceptual data warehouse design", proceedings of the international workshop on design and management of data warehouses, DMDW, page no- 3-9 , 2000.
9. Boehnlein, M., Ulbrich vom Ende A., "Deriving initial data warehouse structures from the conceptual data models of the underlying operational information systems", Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP DOLAP 99, ISBN: 1581132204, Page no: 15-21, 1999.
10. Frendi, M., Salinesi, C. "Requirements engineering for data warehousing", proceedings of REFSQ Workshop, 2003.
11. Winter, R., Strauch, B., "Information requirements engineering for data warehouse systems", proceedings of ACM symposium on applied computing, ISBN: 1-58113-812-1, 2003.
12. Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., Paraboschi, S.,"Designing data marts for data warehouses". ACM Transactions on Software Engineering and Methodology (TOSEM), vol 10, issue 4, page no: 452-483, 2001.
13. Boehnlein, M., & Ulbrich vom Ende, A, "Business process oriented development of data warehouse structures". In Proceeding Data Warehousing 2000, Physica Verlag, 2000.
14. Winter, R., & Strauch, B., "A method for demand-driven information requirements analysis in data warehousing". In Proceedings of Hawaii International Conference on System Sciences, Hawaii, 1359-1365, 2003.
15. Jensen, M., Holmgren, T., & Pedersen, T., "Discovering Multidimensional Structure in Relational Data". In Proceedings of International Conference on Data Warehousing and Knowledge Discovery, Zaragoza, Spain, 138-148, 2004.
16. Prakash, N., Gosain, A., "Requirements driven data warehouse development", proceedings of CAiSE Short Paper, page no-13-17, 2003.
17. Shiefer, J., List, B., Bruckner, R.M., "A holistic approach for managing requirements of data warehouse systems", proceedings of 8th Americas Conference on information systems, AIS, page no: 77-87, 2002.
18. Rilson, F., and Freire, J., "DWARF: AN Approach for Requirements Definition and Management of Data Warehouse Systems",Proceeding of the 11th IEEE International Requirements Engineering Conference 1090-9, 2003.
19. Manoj Kumar, Anjana Gosain, Yogesh Singh, " Stakeholders Driven Requirements Engineering Approach for Data Warehouse Development", Journal of Information Processing Systems, Vol.6,No.3, 2010.
20. Deepti Mishra, Alok Mishra, Ali Yazici, "successful requirement elicitation by combining requirement engineering techniques", First International Conference on the Applications of Digital Information and Web Technologies, ICADIWT, ISBN: 978-1-4244-2623-2, page: 258–263, 2008.