# Goal Structured Requirement Engineering and Traceability Model for Data Warehouses

**Vinay Kumar**
Professor, Department of IT, VIPS, GGSIPU, New Delhi 110 088, India
*E-mail: vinay5861@gmail.com*


**Reema Thareja\***
Assistant Professor SPM Delhi University, Author of Oxford University Press, India
*\*Corresponding Author E-mail: reema_thareja@yahoo.com*

*Abstract*— Data warehouses are decision support systems that are specifically designed for the business managers and executives for reporting and business analysis. Data warehouse is a database that stores enterprise-wide data that can be used to deduce useful information. Business organizations can achieve a great level of competitive advantage by analyzing its historical data and learning from it. However data warehouse concept is still maturing as a technology. In order to effectively design and implement a data warehouse for an organization, its goal needs to be understood and requirement must be analyzed in the perspective of the identified goal. In this paper we present a goal structured model for requirements engineering that also enables its users to manage traceability between the goals, decisions, business strategy and the corresponding business model.

*Index Terms*— Business Model, Data Centric, Data Driven, Data Warehouse, Requirements Engineering, Traceability

## I. Introduction

Data warehousing is a new paradigm that provides strategic information to its users whenever required and therefore it is becoming an integral part in any MIS implementation. In the 1990s, many organizations began to achieve competitive advantage by moving into this technology. Basically, data warehousing is a comprehensive term which includes the various activities involved in the construction, maintenance, and use of the *information oriented architecture.*

Business organizations can achieve a great level of competitive advantage by analyzing its historical data and learning from it. This analysis can reveal certain unusual trends in the concerned business activities that in turn can indicate opportunities for new business. For instance, analysis of past customer demands can help in forecasting production needs. A data warehouse is thus an integrated collection of *enterprise-wide data* that is organized and managed to support enterprise related decision making process. Data warehousing systems facilitate the business executives and managers to acquire and integrate information from heterogeneous sources and to query very large databases efficiently.

There are broadly two approaches to develop data warehouses. One approach is data driven [9] and the other is requirements driven approach [6]. In the data driven approach, data from operational systems are collected, cleaned and then stored in a data warehouse. In the requirements driven approach, the data needs of the users are first identified and then the data warehouse is designed in such a way that all the data needs of the users are satisfied. It experienced that the requirements driven approach to build a data warehouse is better suited than the data driven approach [12]. For the requirements driven approach to succeed, an explicit phase of requirements engineering in the data warehouse development is required. In a data warehouse, the information in warehouse must drive the business model. The model in turn is a part of business strategy which is formulated based on one or more decisions [2].

Strategy formulation does not take place in vacuum rather it depends on the goals and objectives of the business. In order to develop a data warehouse that best meets the user's needs, the first step must therefore be to identify the goals of the business. A g*oal* is the objectives of using the data warehouse. A goal can be either simple or complex. While a simple goal cannot be decomposed into simpler ones, a complex goal, on the other hand, is decomposed in smaller goals that may itself be simple or complex. This makes a goal hierarchy.

A *stakeholder* is a person or a group of people having interest in the project. They are actively involved in the project as they may affect the project. Stakeholders may own more than one goal. But a simple goal may not be shared between different stakeholders. For instance customer, user group, etc are stake holders. *Strategy* is defined as a plan to accomplish the goal. *Decision* is a choice that is made between *n* numbers of possible

actions. For a decision to be effective and in tune with the time, the information base should be of quality. *Quality Information* is the Information that is accurate, specific and organized for a purpose, presented within a context that gives it meaning and relevance to enhance understanding and decrease uncertainty. The information is valuable only if it is effective enough to affect a decision. Strategy is associated with a *justification* that explains how the framed strategy can help to achieve business goal. Justification may be given using formal proofs.

The paper is organized in five sections. Section 2 deals with the concept of goal structured requirement engineering. In Section 3, we have proposed a traceability model for data warehousing and its implementation is discussed in the Section 4. The paper is concluded in the section 5.

## II. Goal Structured Requirement Engineering

For the requirements driven approach to develop a data warehouse, Jarke et al [10] had proposed to add a Conceptual Design Phase before the Logical Design Phase that could identify the conceptual objects like facts, dimensions, hierarchies, etc. But how these objects can be identified has not been discussed. So another phase of Requirements Engineering was proposed to be added before the Conceptual Design Phase [4]. The Data Warehouse Requirements Engineering process must be closely related to the strategy formulation process because this is the main reason for using and deploying a Data Warehouse in an organization. The data warehouse development process can be better understood by breaking it down into three phases which are given as follows:

*Conceptual Design Phase* that gives a Conceptual Model that identifies facts, dimensions, dimension hierarchies, aggregations

*Logical Design Phase* that gives a Logical model to represent data in DDL which depends on the DW package to be used

*Physical Design Phase* that gives a Physical Model to define the physical layout of how data will be organized in the Data Warehouse

A data warehouse is used by user to extract strategic information. Thus, a data warehousing solution is not only about the technology but also about solving user's problems and providing them useful information. In the requirements engineering phase of data warehouse project, expectations from data warehouse is determined. A data warehouse is an information delivery system. So, while collecting information for data warehouse, approach should be based information delivery model rather than simply data capture model. The requirements engineering phase guides the whole process of system design and development. The

accuracy of this phase is thus, critical for the success of the system.

A user may like to access the data warehouse to search for strategic information, execute queries, get results, manipulate it to view the results along several dimensions in graphical format and then perform analysis. The user may not like to take help from others for such tasks. It is therefore extremely important to develop the system in user friendly manner so that all the needed elements of information is in optimal formats and it may be used by the user with little or no training and independently without any help from IT professionals.
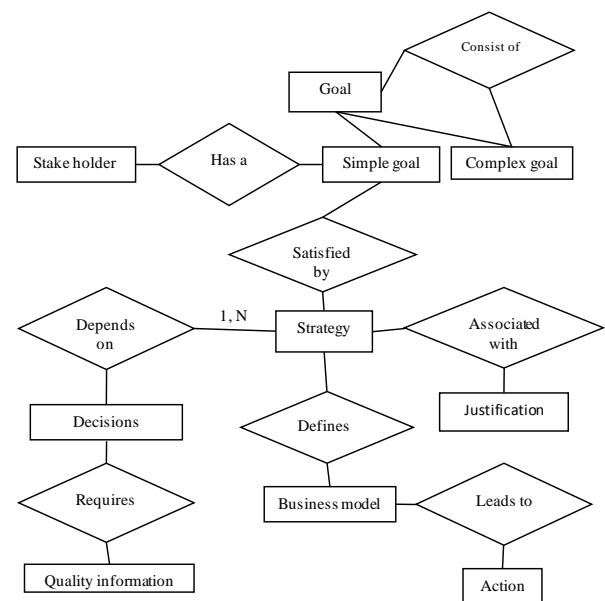


Fig. 1: Goal Structured Requirement Engineering and Traceability Model

In a data warehouse environment, the requirements of the users are driven by their goal. It is important to understand this basic goal which is the most crucial elements for success of any data warehouse. Task performed in different phases of development of the data warehouse is guided by the requirements and hence goals of the users- be it data design, architectural design, infrastructure planning or planning of information delivery mechanisms. The model shown in Fig 1 can be used for requirements engineering and maintaining traceability in data warehouse systems. This model starts with identification of goals of the users.

## III. Traceability Model

Traceability refers to having complete information about every step in the project. It enables users to chronologically inter-relate uniquely identifiable item, verify its history, location, or application through proper documentation. Existing traceability models [7] focus on tracking relationships between requirements

specifications and design. This approach is not enough for a system like that of a data warehouse. Here, we need to assign proper importance to the goals of the users. The proposed model is able to identify a number of traceability relationships between goals, existing business model and stakeholders.

In this model, we have a strategy for each simple goal that is owned by a stakeholder. The strategy leads to a business model. So we can easily trace the relationship between a goal, its stakeholder and the corresponding business model that was designed. This helps in gathering information about the decisions that were taken, assumptions that were made and justifications that were given to approve or reject a particular decision. Since we can have a hierarchy of goals, the model helps to trace the relationships between goals and strategies through-out the hierarchy. This leads to traceability from abstract statements to documented requirements and then to its design level. Whenever certain decisions turn out to be impractical, the effect of undoing a decision or modifying the strategy can be understood. This may be because of the change in user's requirement or due to additional requirements of information or data processing tools.

A business model is a framework that describes core aspects of a business, including purpose, offerings, strategies, infrastructure, organizational structures, etc. Therefore, it gives a complete picture of an organization from a high-level perspective. According to the proposed model, the first and foremost step in the Requirements Engineering process is to identify goals of the users. To accomplish these goals, formulate a sound strategy. However, in order to finalize a strategy a number of decisions may have to be taken. These decisions can be made only with the help of quality information present in the data warehouse. Moreover, every strategy is associated with a justification that explains as how the strategy can accomplish the identified goals. Let us take an example to realize the proposed model.

**GOAL:** Increase market share by 15%

**SUB-GOAL 1:** Increase sales

**SUB-GOAL 2:** Retain Customer Loyalty

**STAKEHOLDER:** CEO of the business

**DECISIONS:**

a. Launch a new product in the market

b. Improve the quality of the existing products

c. Open new Retail Outlets

**INFORMATION:**

a. Information about existing products of the business

b. Information about competitor's products

c. Information about user's preferences and usage patterns

d. Information about sales trends in the market

e. Information about area where there is no or less number of stores

**STRATEGY:** To launch a new product with better quality in the next two years and open a new retail outlet in the north-west region within a year

**JUSTIFICATION:** To launch a new product requires R&D in the area which takes more time, so it can be accomplished within two years. However, opening a new area is a region where there are no/few stores can be done in less time. This strategy will enable the business to enhance its market share by 15%. Justification also includes factual data to give a formal proof. This factual data can be obtained only from the information stored in the data warehouse.

In this model, when a user has a new goal, then the existing business model and the new goal are inputted in the proposed model to formulate a new strategy and thus a revised business model. This is shown in the Fig 2.
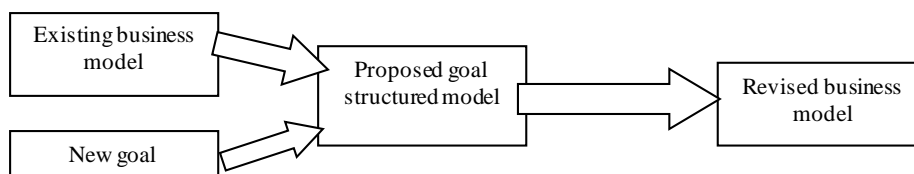


Fig. 2: Revising an existing business strategy using Traceability Model

Thus, we see that backbone of this model is the goals of the user and quality information stored in the data warehouse. With this model in place, the data warehouse development team can easily get an insight into what information is actually required to take the decisions.

**IV. Implementation Strategy**

In the data warehouse environment, metadata occupies a key position. It is only through metadata that the communication among various applications and processes is made possible. It will not be wrong to say that the metadata acts like a nerve center in the data warehouse. Data warehouse users retrieve information

by creating ad hoc queries and running them under the data warehouse environment under implementation. Since report format is decided by the user, user needs to know about data in the data warehouse before creating their own reports and queries. Metadata comes into picture here.

The success of the proposed model lies in appropriate storage and timely retrieval of useful data about goals, strategies and their performance parameters. For this we need to record all this information required in the metadata repository. As of now, we know that we have three types of metadata: end user metadata, operational metadata and extraction and transformation metadata, as summarized in Table 1. The proposed model defines a new perspective called the goal's perspective to store information in the metadata repository. Look at the figure which shows a detailed explanation of a goal.

Table 1: Types of metadata in data warehouse systems

| Type of Metadata | Usage |
|---|---|
| Operational | It contains information about the operational data sources. |
| Extraction and Transformation | It contains information about data extraction from source systems and various transformation techniques that were applied to the data before storing it in the data warehouse. |
| End User | It acts as a navigational map of the data warehouse by enabling end-users to find information using their own terminologies. End-user metadata translates a cryptic name code of a data element into a meaningful description so that end-users can understand and use that data |

The metadata in data warehouse can be made to store additional data that is centered on goals. Doing this an organization can reap the following benefits:

- Make relationship between the data in data warehouse and user's goals visible and accessible

- show how user's goals and metrics are reflected in the data warehouse model

- enable the development team to use this information to support and improve data interpretation

- Helps the decision makers to better interpret the performance of the organization and understand the implications of strategies formulated.

- Involves a comparatively small investment for valuable metadata but gives useful information to data warehouse users

- Goals of the organization acts as a single driving force of information

- Very useful for requirements analysis and design or re-design of data warehouse as it helps the development team to extract more information as compared to that extracted using user interviews

According to the proposed metadata model, the data warehouse metadata must store detailed information about goals. Here, every simple goal is owned by a stake holder who belongs to a particular department and has a metric associated with it. The metric has a target value which is specified in some units. In addition to this, scope of the metric is defined by dimensions and may be limited by a specific time frame.

For example, a simple goal "Increase the sales of Product xyz by 5% in north region by the end of this year" has a timeframe (end of the current year), the goal may be owned by an employee Mr. A of department B. Here, scope of the goal is constrained by dimensions product and region. Finally the value of the metric with units is 5%.

**Conceptual Schema for the Proposed Model**

We know that in data warehouse, data is modeled based on star schema. In star schema dimesnion tables surounds the fact table thereby forming a star type formation. Like fact and dimension tables, we can also arrange the Goal table and dimension tables. Let us now analyze how the details about user's goal can be represented in the form of a star schema.

According to Fig 4, Goal table is at the centre and stores details about user's goal. These details include a goal id, goal name. In addition to this, it stores name of the stakeholder who owns that goal, metric related to the goal, the region where the goal is applicable, the product for which the goal is meant, time at which it entered the system and the remarks that explains whether the goal was accomplished and the technical details of the process carried out to satisfy the goal.

Note that Goal table has a measure- frequency of use, which is a counter that is automatically incremented whenever a particular goal is accessed and refereed for making decisions. Such a construct gives an insight into the effectiveness, efficiency, satisfaction, confidence and acceptance level of the decisions.

All type of information is made accessible by organizing the details in separate tables and then using keys of those tables in the Goal table. Advantages of the Goal-Dimension Tables include:

- Helps to organize the details of user's goals, the strategies formulated to satisfy those goals and other related information in separate tables.

- Helps to know which strategies were applied in which situation and whether the chosen strategy was successful or not

- Having a historical data on goals and strategies helps to make better, more informed, quality decisions
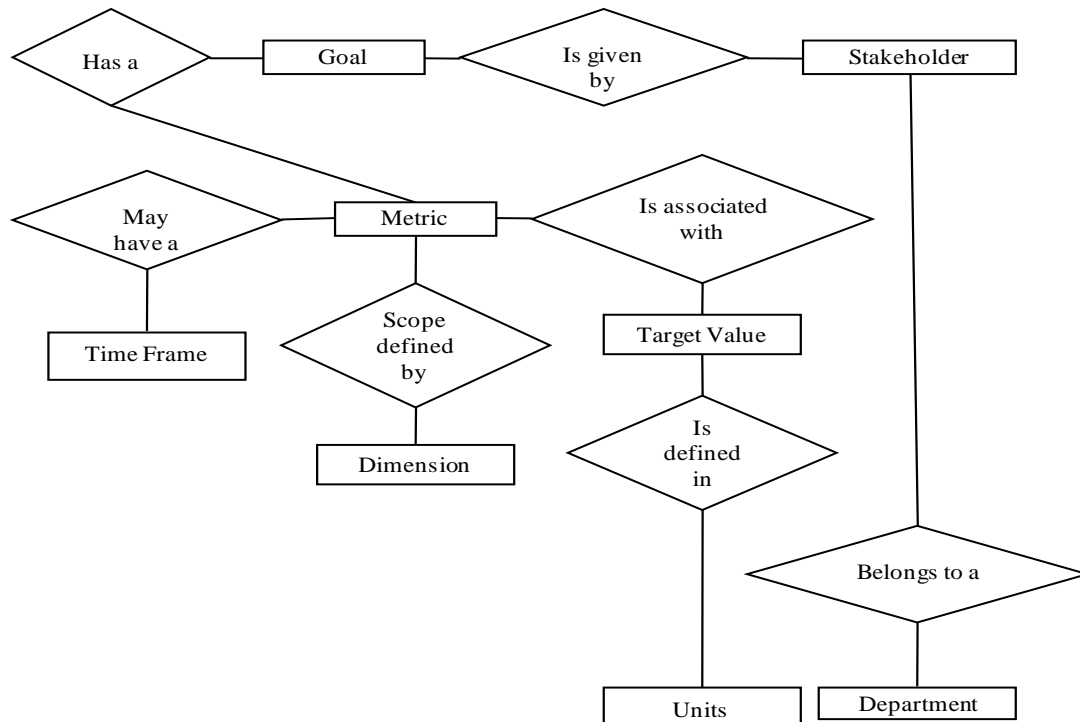


Fig. 3: goal perspective data to be stored in the metadata

The metadata in data warehouse can be made to store additional data that is centered on goals. Doing this an organization can reap the following benefits:

- Make relationship between the data in data warehouse and user's goals visible and accessible

- show how user's goals and metrics are reflected in the data warehouse model

- enable the development team to use this information to support and improve data interpretation

- Helps the decision makers to better interpret the performance of the organization and understand the implications of strategies formulated.

- Involves a comparatively small investment for valuable metadata but gives useful information to data warehouse users

- Goals of the organization acts as a single driving force of information

- Very useful for requirements analysis and design or re-design of data warehouse as it helps the development team to extract more information as compared to that extracted using user interviews

According to the proposed metadata model, the data warehouse metadata must store detailed information about goals. Here, every simple goal is owned by a stake holder who belongs to a particular department and has a metric associated with it. The metric has a target value which is specified in some units. In addition to this, scope of the metric is defined by dimensions and may be limited by a specific time frame.

For example, a simple goal "Increase the sales of Product xyz by 5% in north region by the end of this year" has a timeframe (end of the current year), the goal may be owned by an employee Mr. A of department B. Here, scope of the goal is constrained by dimensions product and region. Finally the value of the metric with units is 5%.

**Conceptual Schema for the Proposed Model**

We know that in data warehouse, data is modeled based on star schema. In star schema dimesnion tables surounds the fact table thereby forming a star type formation. Like fact and dimension tables, we can also arrange the Goal table and dimension tables. Let us now analyze how the details about user's goal can be represented in the form of a star schema.

    

According to fig 4, Goal table is at the centre and stores details about user's goal. These details include a goal id, goal name. In addition to this, it stores name of the stakeholder who owns that goal, metric related to the goal, the region where the goal is applicable, the product for which the goal is meant, time at which it entered the system and the remarks that explains whether the goal was accomplished and the technical details of the process carried out to satisfy the goal.

Note that Goal table has a measure- frequency of use, which is a counter that is automatically incremented whenever a particular goal is accessed and refereed for making decisions. Such a construct gives an insight into the effectiveness, efficiency, satisfaction, confidence and acceptance level of the decisions.

All type of information is made accessible by organizing the details in separate tables and then using keys of those tables in the Goal table. Advantages of the Goal-Dimension Tables include:

- Helps to organize the details of user's goals, the strategies formulated to satisfy those goals and other related information in separate tables.

- Helps to know which strategies were applied in which situation and whether the chosen strategy was successful or not

- Having a historical data on goals and strategies helps to make better, more informed, quality decisions
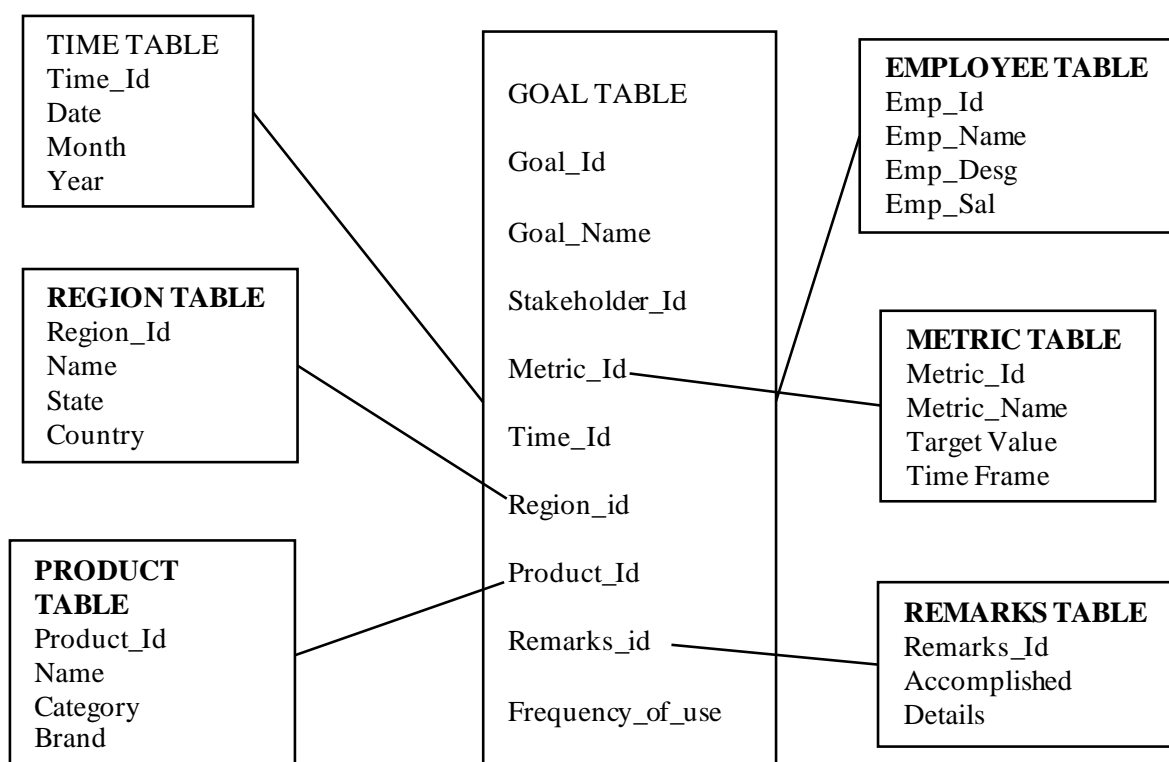


Fig. 4: Goal and Dimension Tables

However, to make the proposed model a success, data warehouse team should monitor the usage pattern of the data and its impact on the strategic decision toward fulfilling the goal. Continuous monitoring would bring forward user's concerns. A performance metric used by organization to measure the performance of its decision can very well be used to measure the effectiveness of data warehouse. Hess and Wells [11] observed that although metadata is heavily used by the data analysts, data warehouse team gives very less or no attention to its maintenance. All the more, access to the metadata is often not convenient, timely and centralized. Poor quality of metadata resulted due to lack of maintenance, affected the effectiveness of the decisions that were made. Many a times, data warehouse users appeared not to use strategies that were recommended.

Therefore, metadata maintenance and ease of access must be of utmost concern for the data warehouse team.

## V. Conclusion

We have proposed a goal structured model in which the data about goals are organized in a table called the *goal table*. This table is linked with corresponding *dimension tables* to access relevant data in an organized manner. D*ata*-centric approach to building the *data* warehouse has not been successful. The traceability model extends beyond tracing the relationships between requirement specifications and design. It is structured around user's goals and helps in the following ways.

- identifying the requirements that would be affected when the designer of the system wants to undo a specific design path,

- analyzing the impact on existing requirements when a user's changes the requirements of an ongoing project, and

- re-using the existing business model as the decisions, justifications and assumptions if any can be easily understood.

Requirements driven approach to build a data warehouse is a better approach, but it requires an explicit phase of requirements engineering to gather information from users. The information collected from the users gives an insight into the data that must be stored in the data warehouse. Formulation of effective strategies depends on the goals of the users. The proposed model keeps detailed record of goals and helps in evaluation of impact of decision taken based on the data from the data warehouse.

The model helps in identifying dormant data in the data warehouse. Dormant data is the one that is never accessed by the users. Keeping this data in the data warehouse is a useless overhead and therefore such data may be removed from data warehouse. Goal structured metadata stores data pertaining to the goals of the stakeholders who own that goal, the strategies that were formulated to satisfy them, alternative paths of actions that were proposed, reasons for not choosing the alternate options and performance of strategic decisions. Thereby, this approach makes the relationship between the data and user's goals visible and accessible. It shows how user's goals and metrics are reflected in the data warehouse model. The model also helps in evaluating implications of strategies formulated.

## Acknowledgement

## References

[1] A. I. Anton, (1996). Goal based requirements analysis. Proceedings of the 2nd International Conference on Requirements Engineering ICRE'96, pp. 136-144.

[2] Anjana Gosain, Navin Prakash. Requirements Driven Data Warehouse Development. Proceedings CAiSE Conference, Klagenfurt/Velden, Austria, 16-20 June, 2003.

[3] B. Dano, H. Briand, F. Barbier. *A use case driven requirements engineering process*. Third IEEE International Symposium On Requirements Engineering RE'97, Antapolis, Maryland, IEEE Computer Society Press, 1997.

[4] Ballard C.,Herreman D., Schau D. , Bell R. , Kim E., Valencic A. Data Modeling Techniques for Data Warehousing, redbooks.ibm.com

[5] Davies, A., Overmyer, S. Identifying and Measuring Quality in a Software Requirements Specification. *First International Software Metrics Symposium*, 1993.

[6] Foder, J., Higgins, C., McDermied, J. & Storrs, G. SAM. A Tool to Support the Construction, Review and Evaluation of Safety Arguments in F. Redmill & T. Anderson [eds] *Directions in Safety, Critical Systems*. Springer Verlag, 1993

[7] Gotel, O.C.Z & Finklestein, A.C.W. *An Analysis of the Requirements Traceability Problem*. Imperial College of Science and Technology, 1993.

[8] Inmon W.H., Building the Data Warehouse, John Wiley, New York.

[9] J. Bubenko, C. Rolland, P. Loucopoulos, V De Antonellis. *Facilitating 'fuzzy to formal' requirements modelling*. IEEE 1st Conference on Requirements Engineering, ICRE'94 pp. 154-158, 1994.

[10] Jarke M., Jeusfeld A., Quix C., Vassiliadis P. Architecture and Quality in Data Warehouses, Proceedings 10th CAiSE Conference, Pernici B., Thanos C. (eds.), Springer, 93-113.

[11] Robert Winter, Bernhard Strauch (2003). Demand-driven Information Requirements Analysis in Data Warehousing. Proceedings of the Hawai International Conference on Systems Sciences, IEEE- 2003.

[12] Traci J. Hess, John D. Well. Understanding How Metadata and Explanations Can Better Support Data Warehousing and Related Decision Support Systems: An Exploratory Case Study. Proceedings of the 35th Hawaii International Conference on System Sciences – 2002.

## Authors' profiles

**Vinay Kumar** is a Professor in Vivekananda Institute of Professional Studies, Delhi. Earlier he worked as Scientist in National Informatics Centre, MoCIT, Government of India. He completed his Ph.D. in Computer Science from University of Delhi and MCA from JNU, Delhi. He has authored a book on Discrete Mathematics and has contributed many research papers to refereed journals and conferences. His area of interest is graph

algorithm, steganography, data security, data mining and e-governance. He is member of CSI and ACM.

**Reema Thareja** is working as an Assistant Professor in Department of Computer Science, Shyama Prasad Mukherjee College for Women, University of Delhi. She is author of Programming in C, Data Structures, Data Warehousing, Data and File Structures (GTU), Computer Fundamentals and Programming in C, Introduction to C Programming and co-author Computer Programming and Data Structures (JNTU) all published by Oxford University Press.