PREFACE

Data warehousing is the process of implementing and using a data warehouse. A data warehouse is built by integrating data from multiple heterogeneous sources to support decision making, analytical reporting and executing structured and/or ad hoc queries on the data.

However, building a data warehouse is a challenging task as it involves many organizational units of an organization. A data warehouse is modeled as a multidimensional data store of historicized, cleansed, validated, internal as well as external data. It supports a comprehensive and flexible analysis of data by all its stakeholders where each stakeholder has their own way of viewing and analyzing data. Hence, it is necessary to elicit requirements from these stakeholders. Very often data warehouses fail because they are built without correctly understanding the user’s requirements.

Although during the requirement definition process system analysts and stakeholders work together to describe the requirements for the data warehouse system, the analysts find it difficult to understand the business terminology. Moreover, the requirements specification document is too technical for the end users to understand. They somehow accept it in order to move forward.

Thus, we see that there is a big communication gap between users and analysts. Our research work is therefore focused to minimize such problems so that a user friendly data warehouse can be developed. We have also worked to ensure how the quality of data can be improved by taking some measures in the requirements definition phase. Stress has been laid on the first step of data warehouse development because sooner the sources of data pollution are detected and corrected the better it is. It would cost less time, money and effort to develop a useful data warehouse.

Chapters in the thesis are included to highlight the following contributions.
a. We have proposed a goal structured model traceability model that keeps detailed record of goals and helps in evaluation of impact of decision taken based on the data from the data warehouse.

b. Identified potential source of errors causing quality compromises

c. Proposed a meta-data based quality model to enforce quality in the data warehouse.

d. Given a set of quality goals and their corresponding parameters

e. Given a technique to quantify quality parameters

f. Proposed a technique to improve the quality of a component if it fails to meet the specified quality level

g. Set of activities to implement traceability in data warehouse systems

All this is given in chapters 2, 3 and 4

h. Given a methodological support for information requirements analysis, elicitation and management that is based on demand-oriented approach for an iterative, priority oriented approach for requirements engineering. (chapter 7)

i. To incorporate spatial information, we have given a simple star schema that organizes spatial data in dimension table. We have also proposed a framework for collecting requirements for spatial analysis. (chapter 5)

j. We have given a layered security mechanism so that even if one layer fails to provide security, the other layer do not allow users with malicious intentions to breech the security
rules implemented for the data warehouse. The encryption algorithm is simple in computational sense to retain the response time for any query. (Chapter 6)