

# Preface

Database practices, techniques and technologies continue to play a significant role in riding the increasingly turbulent wave of technological advances. The Internet is having a major impact on information sharing in a global marketplace. Advances in science and engineering fields have resulted in an explosion of information that must be effectively modeled and managed. Legacy software systems have vast amounts of data ready for analysis and interpretation. These are just a few examples of the opportunities and challenges associated with today's database systems.

Effective mechanisms for collection, storage, retrieval, analysis and dissemination of information are needed to take advantage of technological breakthroughs. Data complexity issues need to be addressed such as, security, maintainability, completeness and correctness, in order to minimize the risks associated with these new technologies. Innovations in database practices, techniques and technology are needed to meet the increasing challenges of this new decade.

The objective of this book is to share innovative and groundbreaking database concepts as proposed by the contributing authors. The concepts proposed in these chapters provide a foundation for continuous improvements in the database field. The diversity of coverage includes quality and organizational issues, measurement systems, design and implementation methods, data warehousing and mining techniques, data modeling and reengineering techniques, security and enhanced query capabilities. Each of these chapters is briefly described.

In the chapter entitled, "Organizational Concepts and Measures for the Evaluation of Data Modeling," Ronald Maier presents an organizational perspective on data modeling. His discussion of the evaluation of data modeling is based on existing theoretical approaches and his empirical studies. The chapter proposes that data modeling should include both project-driven activities and enterprise-wide activities inclusive of long-term goals. It also points out the need for focusing on organizational issues associated with data modeling inclusive of process and product perspectives of quality.

In "Dimensions of Database Quality," John Hoxmeier reminds us that data quality is a critical issue because databases are part of virtually all conventional and e-business applications. He points out the need for a comprehensive set of quality dimensions in order to be successful in the development of high-quality database systems. In addition to process and data factors, it is proposed that quality dimensions include model and behavioral factors, as these are not typically part of data quality assessment. A framework is presented, which allows for an assessment of process, data, model and behavior quality dimensions. A test case is used to illustrate the application of the proposed framework.

Coral Calero, Mario Piattini and Marcela Genero address an important but mostly overlooked issue of metrics for assessing database complexity. They propose internal

measures in their chapter, “Metrics for Controlling Database Complexity,” for assessing the quality of database systems. These measures characterize the complexity of a database in order to promote data quality. The focus of the chapter is on the measurement of complexity that affects the maintainability of the relational, object-relational and active database schemas. The authors are expanding upon their work in collaboration with industry and public organizations.

In “Integrating Hypermedia Functionality into Database Applications,” Bhaumik et al. describe their research efforts at exploring all aspects of hypermedia support for database applications. They propose that a dynamic hypermedia engine (DHE) be used to automate features associated with database systems inclusive of Web technology. One such feature is the automatic generation of links based on the database’s conceptual schema with its original specification. This technology allows for the developer to specify which kinds of database elements are related to diverse elements in the same application, other database applications or other software systems. It is proposed that data warehousing and data mining be incorporated into this technology.

Many of today’s database systems require the processing of large volumes of data in order to support the discovery of new knowledge. In “Optimization of the Knowledge Discovery Process in Very Large Databases,” Mehdi Owrang discusses how today’s vast amount of data makes the discovery process computationally expensive. It is proposed that domain knowledge be used to reduce the size of the database and to optimize the hypothesis, thus eliminating implied, unnecessary and redundant conditions. The resulting benefits include greater efficiency and the discovery of meaningful, non-redundant and consistent rules.

Data warehousing is the focus of Ladjel Bellatreche, Kamalakara Karlapalem and Mukesh Mohania’s chapter on, “Some Issues in Design of Data Warehousing Systems.” The authors tell us that data warehousing can assist intelligent decision-making in order to improve the functioning of an organization. Several trade-offs associated with data warehousing designs are discussed inclusive of materialized views, partitioning a data warehouse and index selection to efficiently execute queries. A study is described which addresses these issues, and the findings are shared.

“Data Mining for Supply Chain Management in Complex Networks: Concepts, Methodology and Applications” describes data mining concepts, its methodology and its application in the context of supply chain management of complex networks. Manoj Singh and Mahesh Raisinghani help us make sense of data mining as a set of techniques used to uncover previously obscure or unknown patterns and relationships in very large databases. The authors describe supply chain management with its data complexity, which can be facilitated by data mining.

Relational databases and the current SQL standard are poorly suited to support hierarchical data. Ido Millet, in “Accommodating Hierarchies in Relational Databases,” describes techniques to address the data retrieval and maintenance problems posed by hierarchical data structures. Data denormalization is proposed as a means of addressing these problems. But there are costs associated with this approach. The chapter describes how these costs can be avoided using processing logic in the form of triggers.

The chapter by Esperanza Marcos and Paloma Caceres is appropriately entitled, “Object-Oriented Database Design,” because it reviews the state-of-the-art in database design for object-relational and object database technologies. The SQL: 1999 and Object Database

Management Group (ODMG) object models are summarized in terms of their advantages and disadvantages. An example illustrates the difficulty of using the relational model to represent complex objects and relationships. It is also used to show how the object-relational and object databases address this complexity.

In “INTECoM: An Integrated Approach to the Specification and Design of Information Requirements,” a framework is presented, which was developed to utilize the strength of existing data modeling approaches while compensating for their weaknesses. Clare Atkins presents this framework as a means of using existing techniques as matched by specific requirements of analysis and design. Though it needs further study in an organizational setting, it encompasses an already-recognized framework for database development, doesn’t require much in the way of new skills, methods and techniques, and may prove extremely useful for less experienced practitioners, Atkins points out.

The chapter entitled, “Inclusion Dependencies,” offers insight into reengineering techniques for quality improvement of existing database systems. Laura Rivero, Jorge Doorn and Viviana Ferragine point out that many of today’s database systems have been poorly designed or may have become flawed when physical objects are removed. As a result, database quality suffers until it is re-engineered. The authors provide heuristics to redesign the conceptual schema, which are based on the identification of hidden business rules and the conversion of non-key inclusion dependencies into key-based ones.

Cheryl Dunn and Severin Grabski present a normative semantic model for enterprise information systems in their chapter on “Semantically Modeled Databases in Integrated Enterprise Information Systems.” This work focuses on the fact that huge investments have been made in enterprise resource planning (ERP) systems, and more value could be realized if databases semantically reflected the underlying reality of organizations. They review empirical research on semantically modeled information systems and then provide an example of a semantic model as proof of concept. They expand this discussion to include the model’s application to ERP and inter-organizational systems.

Mohammad Dadashzadeh presents a generalized approach to formulating set comparison queries in SQL, in “Set Comparison Queries in SQL.” This generalized approach can be used to teach advanced users how to formulate complex set comparison queries in on-line analytical processing scenarios.

In “Toward a Framework for Advanced Query Processing,” Suk-Chung Yoon presents a method for advanced query processing. The author’s approach provides a simple and reasonable way of incorporating user’s needs and preferences into query processing. Instead of relying on conventional queries, Yoon proposes that flexible and intensional query processing techniques be used in order include concepts at different levels of abstraction.

In the chapter, “Security in Database Systems: A State of the Art,” Eduardo Fernandez-Medina and Mario Piattini address the very important security issue associated with today’s technological advances. The Internet, electronic business and data warehousing are a few of the technologies that have shown the need for security considerations early in the development lifecycle. The authors provide a synthesis of the principal aspects that affect confidentiality in the design of databases: control of access, modeling of security requirements and a retrospective view of the important generations of methods for the development of security techniques. They introduce a methodology through which multilevel databases can be designed, while taking into consideration confidentiality requisites and risk management factors.

“A Case Study of the Military Utility of Telemedicine” focuses on the need for

implementation of telemedicine at a strategic level in the military. This success of telemedicine efforts requires systematic analysis and effective technology management inclusive of database systems. This chapter differs from the others in that it summarizes a case study with insights into the complexities associated with telemedicine and supporting technologies. David Paper, James Rodger and Parag Pendharkar present methodologies for investigating military utility of telemedicine, and summarize initial lessons learned.

In summary, this book offers an exciting opportunity to find out about current practices, techniques and technologies to meet the needs of today's software and information systems.

## **Acknowledgement**

The editor would like to acknowledge the help of all persons involved in the collation and review process of this book. The authors not only did a remarkable job of ensuring that high quality chapters were created, but many served as referees for chapters written by other authors. Thanks to all of you who have provided constructive and comprehensive reviews. A note of thanks to Mehdi Khosrowpour, who saw a need for this book, and to the staff at Idea Group Publishing for their guidance and professional support. A special thanks goes to David and Marissa Fassino for their unfailing encouragement and patience.