# Preface

Computer-based information technologies have been extensively used to help industries manage their processes, and information systems hereby become their nervous center. More specifically, databases are designed to support the data storage, processing, and retrieval activities related to data management in information systems. Database management systems provide efficient task support, and database systems are the key to implementing industrial data management. Industrial data management requires database technical support. Industrial applications, however, are typically data- and knowledge-intensive and have some unique characteristics (e.g., large volumes of data with complex structures) that make them difficult to manage. Some new techniques such as the Web, artificial intelligence, and so forth have been introduced into industrial applications. These unique characteristics and the usage of new technologies have put many potential requirements on industrial data management, which challenges today's database systems and promotes their evolvement.

Viewed from database technology, information modeling in databases (database modeling for short) can be identified at two levels: conceptual data modeling and database modeling. This results in conceptual (semantic) data model and logical database model. Generally, a conceptual data model is designed, then the designed conceptual data model will be transformed into a chosen logical database schema. Database systems based on logical database models are used to build information systems for data management. Much attention has been directed at conceptual data modeling of industrial information systems. Product data models, for example, can be viewed as a class of semantic data models (i.e., conceptual data models) that take into account the needs of engineering data. Recently, conceptual data modeling of enterprises has received increasing attention. Generally speaking, traditional ER/EER or

UML models in database areas can be used for industrial data modeling at the conceptual level. But, limited by their power in industrial data modeling, some new conceptual data models such as IDEF1X and STEP/EXPRESS have been developed. In particular, to implement share and exchange of industrial data, the Standard for the Exchange of Product Model Data (STEP) is being developed by the International Organization for Standardization (ISO). EXPRESS is the description methods of STEP and a conceptual schema language, which can model product design, manufacturing, and production data. EXPRESS model hereby becomes a major one of conceptual data models for industrial data modeling. Many research works have been reported on the database implementation of the EXPRESS model in context of STEP, and some software packages and tools are available in the marketplace. For industrial data modeling in database systems, the generic logical database models such as relational, nested relational, and object-oriented databases have been used. However, these generic logical database models do not always satisfy the requirements of industrial data management. In non-transaction processing such as CAD/CAM, knowledge-based system, multimedia and Internet systems, for example, most of these data-intensive application systems suffer from the same limitations of relational databases. Some non-traditional database models based on special, hybrid, and/or the extended database models above have been proposed accordingly.

Database technology is typically application-oriented. With advances and in-depth applications of computer technologies in industry, database modeling for industrial data management is emerging as a new discipline. The research and development of industrial databases is receiving increasing attention. By means of database technology, large volumes of industrial data with complex structures can be modeled in conceptual data models and further stored in databases. Industrial information systems based the databases can handle and retrieve these data to support various industrial activities. Therefore, database modeling for industrial data management is a field which must be investigated by academic researchers, together with developers and users both from database and industry areas.

# Introduction

This book, which consists of 11 chapters, is organized into two major sections. The first section discusses the issues of industrial databases and applications in the first nine chapters. The next two chapters covering the data modeling issue in generic databases comprise the second section.

First of all, we take a look at the problems of the industrial databases and applications.

Databases are designed to support data storage, processing, and retrieval activities related to data management, and database systems are the key to implementing engineering information modeling. But some engineering requirements challenge current mainstream databases, which are mainly used for business applications, and promote their evolvement. Ma tries to identify the requirements for engineering information modeling and then investigates the satisfactions of current database models to these requirements at two levels: conceptual data models and logical database models. Also, the relationships among the conceptual data models and the logical database models for engineering information modeling are presented as viewed from database conceptual design.

ASSO is a database design methodology defined for achieving conceptual schema consistency, logical schema correctness, flexibility in reflecting the real-life changes on the schema, and efficiency in accessing and storing information. B is an industrial formal method for specifying, designing, and coding software systems. Locuratolo investigates the integration of the ASSO features in B. Starting from a B specification of the data structure and of the transactions allowed on a database, two model transformations are designed: The resulting model *Structured Database Schema* integrates static and dynamics, exploiting the novel concepts of *Class-Machines* and *Specialized Class-Machines*. Formal details which must be specified if the conceptual model of ASSO is directly constructed in B are avoided; the costs of the consistency obligations are minimized. Class-Machines supported by semantic data models can be correctly linked with Class-Machines supported by object models.

Carnduff and Goonetillake present research aimed at determining the requirements of a database software tool that supports integrity validation of versioned design artifacts through effective management of evolving constraints. It results in the design and development of a constraint management model, which allows constraint evolution through representing constraints within versioned objects called Constraint Versions Objects (CVOs). This model operates around a version model that uses a well-defined configuration management strategy to manage the versions of complex artifacts. Internal and interdependency constraints are modeled in CVOs. They develop a model which has been implemented in a prototype database tool with an intuitive user interface.

The user interface allows designers to manage design constraints without the need to program. Also, they introduce the innovative concepts developed using an ongoing example of a simple bicycle design.

Similarity search in database systems is an important task in modern application domains such as multimedia, molecular biology, medical imaging and many others. Especially for CAD (Computer-Aided Design), suitable similarity models and a clear representation of the results can help to reduce the cost of developing and producing new parts by maximizing the reuse of existing parts. Kriegel, Kröger, Pfeifle, Brecheisen, Pötke, Schubert, and Seidl present different similarity models for voxelized CAD data based on space partitioning and data partitioning. Based on these similarity models, they introduce an industrial prototype, called BOSS, which helps the user to get an overview over a set of CAD objects. BOSS allows the user to easily browse large data collections by graphically displaying the results of a hierarchical clustering algorithm.

STEP-NC is an emerging ISO standard, which defines a new generation of NC programming language and is fully compliant with STEP. There is a whole suite of implementation methods one may utilize for development purposes. STEP-NC brings richer information to the numerically-controlled machine tools; hence intelligent machining and control are made possible. Its Web-enabled feature gives itself an additional dimension in that e-manufacturing can be readily supported. Xu addresses the issue of product development chain from the perspective of data modeling and streamlining. The focus is on STEP-NC, and how it may close the gap between design and manufacturing for a complete, integrated product development environment. A case study is given to demonstrate a STEP compliant, Web-enabled manufacturing system.

Yuan shares his experience of enabling semantic-based dynamic information integration across multiple heterogeneous information sources. While data is physically stored in existing legacy data systems across the networks, the information is integrated based upon its semantic meanings. Ontology is used to describe the semantics of global information content, and semantic enhancement is achieved by mapping the local metadata onto the ontology. For better system reliability, a unique mechanism is introduced to perform appropriate adjustments upon detecting environmental changes.

Panagis, Sakkopoulos, Sioutas, and Tsakalidis present the Web Service architecture and propose Web Service integration and management strategies for large-scale datasets. They mainly present the elements of Web Service architecture, the challenges in implementing Web Services whenever large-scale data are involved, and the design decisions and business process reengineering steps to integrate Web Services in an enterprise information system. Then they provide a case study involving the largest private-sector telephony provider in Greece, where the provider's billing system datasets is utilized. Moreover, they present the scientific work on Web Service discovery along with experiments on implementing an elaborate discovery strategy over real-world, large-scale data.

Bose, Chun, Yue, Ines, and Helen describe the planning and implementation of the Wal-Mart data warehouse and discuss its integration with the operational systems. They also highlight some of the problems encountered in the developmental process of the data warehouse. The implications of the recent advances in technologies such as RFID, which is likely to play an important role in the Wal-Mart data warehouse in future, is also detailed.

Content-based image retrieval (CBIR) can be used to locate medical images in large databases using image features, such as color and texture, to index images with minimal human intervention. Wei, Li, and Wilson introduce a content-based approach to medical image retrieval. First, they introduce the fundamentals of the key components of content-based image retrieval systems are to give an overview of this area. Then they present a case study, which describes the methodology of a CBIR system for retrieving digital mammogram database.

In the second section, we see the generic database modeling.

A strong design phase is involved in most current application development processes (e.g., ER design for relational databases). But conceptual design for XML has not been explored significantly in literature or in practice. Most XML design processes start by directly marking up data in XML, and the metadata is typically designed at the time of encoding the documents. So Mohan and Sengupta introduce the existing methodologies for modeling XML. A discussion is presented comparing and contrasting their capabilities and deficiencies, and delineating the future trend in conceptual design for XML applications.

Ravat, Teste, and Zurfluh focus on constraint-based multi-dimensional modeling. The defined model integrates a constellation of facts and dimensions. Along each dimension, various hierarchies are possibly defined and the model supports multiple instantiations of dimensions. To facilitate data querying, they also define a multi-dimensional query algebra, which integrates the main multi-dimensional operators. These operators support the constraint-based multi-dimensional modeling. Finally, they present two implementations of this algebra, which are OLAP-SQL and a graphical query language. The former is a textual language integrating multi-dimensional concepts (fact, dimension, hierarchy), but it is based on classical SQL syntax. This language is dedicated to specialists such as multi-dimensional database administrators. The latter consists in a graphical representation of multi-dimensional databases and users specify directly their queries over this graph. This approach is dedicated to non-computer scientist users.