Preface

This is the third volume of the *Advances in Data Warehousing and Mining* (ADWM) book series. ADWM publishes books in the areas of data warehousing and mining. The topic of this volume is data warehousing and OLAP. This volume consists of 16 chapters in 4 sections, contributed by researchers in data warehousing.

Section I on “Conceptual Model and Development” consists of five chapters covering various conceptual modeling, data cleaning, production process, and development.

Chapter I, “Development of Data Warehouse Conceptual Models: Method Engineering Approach” by Laila Niedrite, Maris Treimanis, Darja Solodovnikova, and Liga Grundmane, from University of Latvia, discusses the usage of method engineering approach for the development of conceptual models of data warehouses. They describe three methods, including (a) user-driven, (b) data-driven, and (c) goal-driven methods.

Chapter II, “Conceptual Modeling Solutions for the Data Warehouse” by Stefano Rizzi, University of Bologna, is a reprint from *Data Warehouses and OLAP: Concepts, Architectures and Solutions*, edited by R. Wrembel and C. Koncilia (2007). The chapter thoroughly discusses dimensional fact modeling. Several approaches to conceptual design, such as data-driven, requirement-driven, and mixed approaches, are described.

Chapter III, “A Machine Learning Approach to Data Cleaning in Databases and Data Warehouses” by Hamid Haidarian Shahri, University of Maryland, is also a reprint. It is initially published in *Handbook of Research on Fuzzy Information Processing in Databases*, edited by J. Galindo (2008). This chapter introduces entity resolution (or duplicate elimination) in data cleaning process. It also exploits neutofuzzy modeling in the context of entity resolution.

Chapter IV, “Interactive Quality-Oriented Data Warehouse Development” by Maurizio Pighin and Lucio Ieronutti, both from University of Udine, Italy, proposes quantitative and qualitative phases in data warehousing design and evaluation. They also present a tool that they have developed, called ELDA (EvalLuation DAta warehouse quality) to support data warehouse design and evaluation.

Chapter V, “Integrated Business and Production Process Data Warehousing” by Dirk Draheim, Software Competence Center Hagenberg, Austria, and Oscar Mangisengi, BWIN Interactive Entertainment and SMS Data System, Austria, is a chapter contributed by practitioners in industry. They focus on production process data based on business activity monitoring requirements.

Section II on “OLAP and Pattern” consists of 4 chapters covering multi-node OLAP systems, multi-dimensional patterns, and XML OLAP.

Chapter VI, “Selecting and Allocating Cubes in Multi-Node OLAP Systems: An Evolutionary Approach” by Jorge Loureiro, Instituto Politécnico de Viseu, Portugal, and Orlando Belo, Universidade do Minho, Portugal, focuses on multi-node distributed OLAP systems. They propose three algorithms: M-OLAP Greedy, M-OLAP Genetic, and M-OLAP Co-Evol-GA; the last two are based on genetic algorithm and evolutionary approach.
Chapter VII, “Swarm Quant’ Intelligence for Optimizing Multi-Node OLAP Systems”, also by Jorge Loureiro and Orlando Belo, also discusses multi-node OLAP systems. But in this chapter, they propose distributed cube selection algorithms based on discrete particle swarm optimizers to solve the distributed OLAP selection problem. They propose M-OLAP Discrete Particle Swarm Optimization (M-OLAP Di-PSO), M-OLAP Discrete Cooperative Particle Swarm Optimization (M-OLAP Di-CPSO), and M-OLAP Discrete Multi-Phase Particle Swarm Optimization (M-OLAP Di-MPSO).

Chapter VIII, “Multidimensional Analysis of XML Document Contents with OLAP Dimensions” by Franck Ravat, Olivier Teste, and Ronan Tournier, IRIT, Universite Toulouse, France, focuses on XML documents, where they present a framework for multidimensional OLAP analysis of textual data. They describe this using the conceptual and logical model.

Chapter IX, “A Multidimensional Pattern Based Approach for the Design of Data Marts” by Hanene Ben-Abdallah, Jamel Feki, and Mounira Ben Abdallah, from University of Sfax, Tunisia, concentrates on multi-dimensional patterns. In particular the authors describe multi-dimensional pattern from the logical and physical levels.

Section III on “Spatio-Temporal Data Warehousing” consists of 5 chapters covering various issues of spatial and spatio-temporal data warehousing.

Chapter X, “A Multidimensional Methodology with Support for Spatio-Temporal Multigranularity in the Conceptual and Logical Phases” by Concepción M. Gascueña, Carlos III de Madrid University, and Rafael Guadalupe, Politécnica de Madrid University, presents a methodology to design multi-dimensional database to support spatio-temporal granularities. This includes conceptual and logical phases which allow multiple representations of the same spatial data interacting with other spatial and thematic data.

Chapter XI, “Methodology for Improving Data Warehouse Design using Data Sources Temporal Metadata” by Francisco Araque, Alberto Salguero, and Cecilia Delgado, all from University of Granada, focuses on temporal data. They also discuss properties of temporal integration and data integration process.

Chapter XII, “Using Active Rules to Maintain Data Consistency in Data Warehouse Systems” by Shi-Ming Huang, National Chung Cheng University, Taiwan, John Tait, Sunderland University, UK, Chun-Hao Su, National Chung Cheng University, Taiwan, and Chih-Fong Tsai, National Chung Cheng University, Taiwan, focuses on data consistency, particularly from the temporal data aspects.

Chapter XIII, “Distributed Approach to Continuous Queries with kNN Join Processing in Spatial Telemetric Data Warehouse” by Marcin Gorawski and Wojciech Gębczyk, from Silesian Technical University, Poland, concentrates on continuous kNN join query processing, the context of spatial telemetric data warehouse, which is relevant to geospatial and mobile information systems. They also discuss spatial location and telemetric data warehouse and distributed systems.

Chapter XIV, “Spatial Data Warehouse Modelling” by Maria Luisa Damiani, Università di Milano, and Stefano Spaccapietra, Ecole Polytechnique Fédérale de Lausanne, is a reprint from Processing and Managing Complex Data for Decision Support, edited by Jérôme Darmont and Omar Boussaid (2006). The chapter presents multi-dimensional data models for spatial data warehouses. This includes a model for multi-granular spatial data warehouse and spatial OLAP.

The final section of this volume, Section IV on “Benchmarking and Evaluation”, consists of two chapters, one on benchmarking data warehouses and the other on evaluation of slowly changing dimensions.

Chapter XV, “Data Warehouse Benchmarking with DWEB” by Jérôme Darmont, University of Lyon, focuses on the performance evaluation of data warehouses, in which it presents a data warehouse engineering benchmark, called DWEB. The benchmark also generates synthetic data and workloads.
Finally, Chapter XVI, “Analyses and Evaluation of Responses to Slowly Changing Dimensions in Data Warehouses” by Lars Frank and Christian Frank, from the Copenhagen Business School, focuses on dynamic data warehouses, where the dimensions are changing slowly. They particularly discuss different types of dynamicity, and responses to slowly changing dimensions.

Overall, this volume covers important foundations to researches and applications in data warehousing, covering modeling, OLAP and patterns, as well as new directions in benchmarking and evaluating data warehousing. Issues and applications, particularly in spatio-temporal, shows a full spectrum of the coverage of important and emerging topics in data warehousing.

David Taniar
Editor-in-Chief