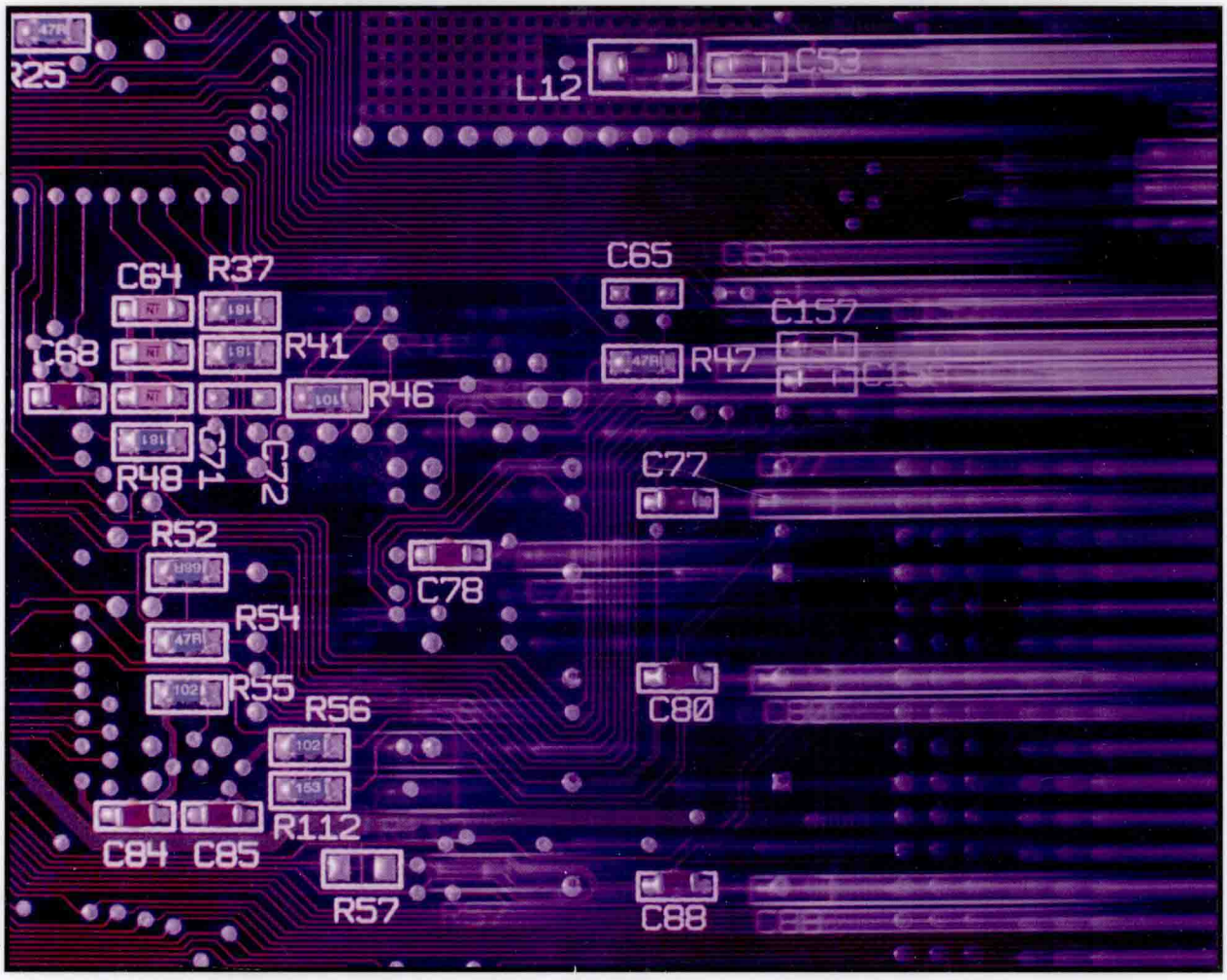
PREMIER REFERENCE SOURCE

Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development

Innovative Methods and Applications



Tho Manh Nguyen

Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development:
Innovative Methods and Applications

Tho Manh Nguyen
Institute of Software Technology and Interactive Systems,
Vienna University of Technology, Austria

常州大学山书馆
藏书章

Information Science REFERENCE
INFORMATION SCIENCE REFERENCE
Hershey · New York

Director of Editorial Content: Kristin Klinger
Senior Managing Editor: Jamie Snavely
Managing Editor: Jeff Ash
Assistant Managing Editor: Michael Brehm
Publishing Assistant: Sean Woznicki
Typesetter: Jeff Ash
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com/reference

Copyright © 2010 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher.

Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Complex data warehousing and knowledge discovery for advanced retrieval development : innovative methods and applications / Tho Manh Nguyen, editor.
p. cm.

Includes bibliographical references and index.
Summary: "This book provides a comprehensive analysis on current issues and trends in retrieval expansion"--Provided by publisher.

ISBN 978-1-60566-748-5 (hardcover) -- ISBN 978-1-60566-749-2 (ebook) 1. Data warehousing. 2. Database searching. I. Nguyen, Tho Manh.

QA76.9.D37C67 2009
005.74--dc22

2009005271

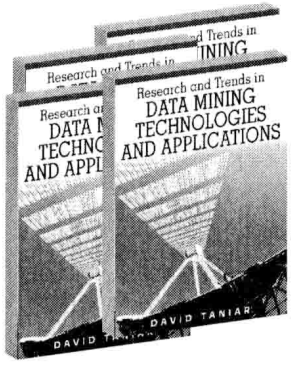
British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Advances in Data Warehousing and Mining Series (ADWM)

ISBN: 1935-2646

Editor-in-Chief: David Taniar, Monash Univerisy, Australia



Research and Trends in Data Mining Technologies and Applications
David Taniar, Monash University, Australia

IGI Publishing • copyright 2007 • 340 pp • H/C (ISBN: 1-59904-271-1) • US $85.46 (our price) • E-Book (ISBN: 1-59904-273-8) • US $63.96 (our price)

Activities in data warehousing and mining are constantly emerging. Data mining methods, algorithms, online analytical processes, data mart and practical issues consistently evolve, providing a challenge for professionals in the field. Research and Trends in Data Mining Technologies and Applications focuses on the integration between the fields of data warehousing and data mining, with emphasis on the applicability to real-world problems. This book provides an international perspective, highlighting solutions to some of researchers' toughest challenges. Developments in the knowledge discovery process, data models, structures, and design serve as answers and solutions to these emerging challenges.

The Advances in Data Warehousing and Mining (ADWM) Book Series aims to publish and disseminate knowledge on an international basis in the areas of data warehousing and data mining. The book series provides a highly regarded outlet for the most emerging research in the field and seeks to bridge underrepresented themes within the data warehousing and mining discipline.
The Advances in Data Warehousing and Mining (ADWM) Book Series serves to provide a continuous forum for state-of-the-art developments and research, as well as current innovative activities in data warehousing and mining. In contrast to other book series, the ADWM focuses on the integration between the fields of data warehousing and data mining, with emphasize on the applicability to real world problems. ADWM is targeted at both academic researchers and practicing IT professionals.

IGI GLOBAL
DISSEMINATOR OF KNOWLEDGE

Hershey • New York

Order online at www.igi-global.com or call 717-533-8845 x 10 –
Mon-Fri 8:30 am - 5:00 pm (est) or fax 24 hours a day 717-533-8661

Editorial Advisory Board

Torben Bach Pedersen, *Aalborg University, Denmark*
Stephane Bressan, *National University of Singapore (NUS) Singapore*
Johann Eder, *University of Klagenfurt, Austria*
Mukesh Mohania, *IBM India Research Lab, India*
Zbigniew W. Ras, *University of North Carolina, USA*
Il-Yeol Song, *Drexel University, USA*
David Taniar, *Monash University, Australia*
Dimitri Theodoratos, *New Jersey Institute of Technology, USA*
A Min Tjoa, *Vienna University of Technology, Austria*
Karen Davis, *University of Cincinnati, USA*
Mohammed J. Zaki, *Rensselaer Polytechnic Institute, USA*
Stefano Rizzi, *University of Bologna, Italy*
Sourav S. Bhowmick, *Nanyang Technological University, Singapore*
Panos Vassiliadis, *University of Ioannina, USA*
Juan Trujilo, *University of Alicante, Spain*

List of Reviewers

Alberto Abello, *Universitat Politecnica de Catalunya, Spain*
Elena Baralis, *Politecnico di Torino, Italy*
Ladjel Bellatreche, *LISI/ENSMA, France*
Bettina Berendt, *Humboldt University Berlin, Germany*
Peter Brezany, *University of Vienna, Austria*
Henrik Bostrom, *University of Skovde, Sweden*
Amilcar Cardoso, *University of Coimbra, Portugal*
Ezeife Christie, *University of Windsor, Canada*
Sunil Choenni, *University of Twente and Dutch Ministry of Justice, Netherlands*
Alfredo Cuzzocrea, *University of Calabria, Italy*
Vladimir Estivill-Castro, *Griffith University, Australia*
Eduardo Fernandez-Medina, *University of Castilla-La Mancha, Spain*
Matteo Golfarelli, *DEIS - University of Bologna, Italy*
Gunter Grieser, *Technical University Darmstadt, Germany*
Eyke Hermeier, *Otto-von-Guericke-University Magdeburg, Germany*

此为试读,需要完整PDF请访问: www.ertongbook.com

Jaakko Hollmen, *Helsinki University of Technology, Finland*
Andreas Hotho, *University of Kassel, Germany*
Se June Hong, *RSM emeritus, IBM TJ Watson Research Center, USA*
Jimmy Huang ,*York University, Canada*
Ryutaro Ichise, *National Institute of Informatics, Japan*
Fan Jianping, *University of North Carolina, USA*
Jinyan Li, *Institute for Infocomm Research, Singapore*
Guanling Lee, *National Dong Hwa University, Taiwan*
Giuseppe Manco, *ICAR-CNR, National Research Council, Italy*
Michael May, *Fraunhofer Institut Autonome Intelligente Systeme, Germany*
Rosa Meo, *University of Turin, Italy*
Alexandros Nanopoulos, *Aristotle University of Thessaloniki, Greece*
Jaakko Peltonen, *Helsinki University of Technology, Finland*
Andrew Rau-Chaplin, *Dalhousie University, Canada*
Monica Scannapieco, *University of Rome, Italy*
Michael Schrefl, *University Linz, Austria*
Giovanni Semeraro, *University of Bari, Italy*
Alkis Simitsis, *Stanford University, USA*
Min Song, *New Jersey Institute of Technology, USA*
Jerzy Stefanowski, *Poznan University of Technology, Poland*
Ah-Hwee Tan, *Nanyang Technological University, Singapore*
Il-hoi Yoo, *University of Missouri, Columbia, USA*
Jenny Zhang, *RMIT University, Australia*
Marek Wojciechowski, *Poznan University of Technology, Poland*
Wolfram Wöß, *University of Linz, Austria*

Preface

Data warehousing and knowledge discovery are established key technologies in many application domains. Enterprises and organizations improve their abilities in data analysis, decision support, and the automatic extraction of knowledge from data; for scientific applications to analyze collected data, for medical applications for quality assurance and for steps to individualized medicine, to mention just a few examples. With the exponentially growing amount of information to be included in the decision making process, the data to be processed become more and more complex in both structure and semantics. Consequently, the process of retrieval and knowledge discovery from this huge amount of heterogeneous complex data constitutes the reality check for research in the area.

The primary objective of the book *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications* is to give readers comprehensive knowledge on the current issues as well as the development trends of data warehousing and knowledge discovery. The book presents the latest research issues and experiences in developing and deploying data warehousing and knowledge discovery systems, applications, and solutions. The 16 book chapters, organized in 5 Sections, were contributed by authors and editorial board members from the *International Journal of Data Warehousing and Mining*, as well as invited authors who are experts in the data warehousing and knowledge discovery field.

Three chapters in **Section 1, Data Warehouse Architectures & Fundamentals**, present the current trends of research on data warehouse architecture, storage and implementations towards improving performance and response time.

In Chapter 1, "The LBF R-tree: Scalable Indexing and Storage for Data Warehousing Systems" by Todd Eavis and David Cueva (Concordia University, Canada), the authors propose an LBF R-tree framework for effective indexing mechanisms in multi-dimensional database environments. The proposed framework addressed not only improves performance on common user-defined range queries, but also gracefully degrades to a linear scan of the data on pathologically large queries. Experimental results demonstrating both efficient disk access on the LBF R-tree, as well as impressive compression ratios for data and indexes.

Chapter 2, "Dynamic Workload for Schema Evolution in Data Warehouses: A Performance Issue," by Cécile Favre, Fadila Bentayeb and Omar Boussaid (ERIC Laboratory, University of Lyon, France), addresses the issues related to the workload's evolution and maintenance in data warehouse systems in response to new requirements modeling resulting from users' personalized analysis needs. The proposed workload management system assists the administrator to maintain and adapt dynamically the workload according to changes arising on the data warehouse schema by improving two types of workload updates: (1) maintaining existing queries consistent with respect to the new data warehouse schema and (2) creating new queries based on the new dimension hierarchy levels.

Songmei Yu, Vijayalakshmi Atluri, and Nabil R. Adam (MSIS Department and CIMIC, Rutgers University, USA) in Chapter 3, "Preview: Optimizing View Materialization Cost in Spatial Data Warehouses," present an optimization approach for materialized view implementation in spatial data warehouse. Due to the fact that spatial data are larger in size and spatial operations are more complex than the traditional relational operations, both the view materialization cost and the on-the-fly computation cost are often extremely high. The authors propose a new notion, called preview, for which both the materialization and on-the-fly costs are significantly smaller than those of the traditional views, so that the total cost is optimized.

Section 2, Multidimensional Data & OLAP, consists of three chapters discussing related issues and challenges in multidimensional database and online analytical processing (OLAP) environment.

Chapter 4, "Decisional Annotations: Integrating and Preserving Decision-Makers' Expertise in Multidimensional Systems," by Guillaume Cabanac, Max Chevalier, Franck Ravat and Olivier Teste (Université de Toulouse, France), deals with an annotation-based decisional system. The decisional system is based on multidimensional databases, which are composed of facts and dimensions. The expertise of decision-makers is modeled, shared and stored through annotations. Every piece of multidimensional data can be associated with zero or more annotations which allow decision-makers to carry on active analysis and to collaborate with other decision-makers on a common analysis.

In Chapter 5, "Federated Data Warehouses," Stefan Berger and Michael Schrefl (University of Linz, Austria) discuss federated data warehouse systems which consist of a collection of data marts provided by different enterprises or public organizations and widen the knowledge base for business analysts, thus enabling better founded strategic decisions. The authors argue that the integration of heterogeneous data marts at the logical schema level is preferable to the migration of data into a physically new system if the involved organizations remain autonomous. They present a federated architecture that provides a global multi-dimensional schema to which the data mart schemas are semantically mapped, repairing all heterogeneities.

Chapter 6, "Decisional Annotations: Integrating and Preserving Decision-Makers' Expertise in Multidimensional Systems," by Jérôme Cubillé, Christian Derquenne, Sabine Goutier, Françoise Guisnel, Henri Klajnmic (EDF Research and Development, France) and Véronique Cariou (ENITIAA, France), describes algorithms to support business intelligence in OLAP databases applying data mining methods to the multidimensional environment. Those methods help end-users' analysis in two ways. First, they identify the most interesting dimensions to expand in order to explore the data. Then, they automatically detect interesting cells among a user selected ones. The solution is based on a tight coupling between OLAP tools and statistical methods, based upon the built-in indicators computed instantaneously during the end-users' exploration of the data cube.

The next three chapters in **Section 3, DWH & OLAP Applications**, present some typical applications using data warehouse and OLAP technology as well as the challenges and issues facing in the real practice.

Chapter 7, "Conceptual Data Warehouse Design Methodology for Business Process Intelligence," by Svetlana Mansmann, Matthias Röger and Marc H. Scholl (University of Konstanz, Konstanz, Germany), Neumuth and Oliver Burgert (Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany), presents a conceptual framework for adopting the data warehousing technology for business process analysis, with surgical workflows analysis as a challenging real-world application. Deficiencies of the conventional OLAP approach are overcome by proposing an extended multidimensional data model, which enables adequate capturing of flow-oriented data. The model supports a number of advanced properties, such as non-quantitative and heterogeneous facts, many-to-many relationships be-

tween facts and dimensions, full and partial dimension sharing, dynamic specification of new measures, and interchangeability of fact and dimension roles.

Deployment of a federated data warehouse approach for the integration of the wide range of different medical data sources and for distribution of evidence-based clinical knowledge, to support clinical decision makers, primarily clinicians at the point of care is the main topic of Chapter 8, "Data Warehouse Facilitating Evidence-Based Medicine," authored by Nevena Stolba, Tho Manh Nguyen, A Min Tjoa (Vienna University of Technology, Austria). A real-world scenario is used to illustrate the possible application field in the area of emergency and intensive care in which the evidence-based medicine merges data originating in a pharmacy database, a social insurance company database and diverse clinical DWHs with the minimized administration effort.

In Chapter 9, "Deploying Data Warehouses in Grids with Efficiency and Availability," Rogério Luís de Carvalho Costa and Pedro Furtado (Departamento de Engenharia Informática, University of Coimbra, Portugal) discuss the deployment of data warehouses over Grids. The authors present the Grid-NPDW architecture, which aims at providing high throughput and data availability in Grid-based warehouses. High efficiency in situations with site failure is also achieved with the use of on-demand query scheduling and data partitioning and replication. The chapter also describes the main components of the Grid-NPDW scheduler and presents some experimental results of proposed strategies.

Section 4, Data Mining Techniques, consists of three chapters discussing a variety of traditional data mining techniques such as clustering, ranking, classification but towards the efficiency and performance improvement.

Chapter 10, "MOSAIC: Agglomerative Clustering with Gabriel Graphs," by Rachsuda Jiamthapthaksin, Jiyeon Choo, Chun-sheng Chen, Oner Ulvi Celepcikay, Christoph F. Eick (Computer Science Department, University of Houston, USA) and Christian Giusti (Department of Mathematics and Computer Science, University of Udine, Italy), introduces MOSAIC, a post-processing technique that has been designed to overcome these disadvantages. MOSAIC is an agglomerative technique that greedily merges neighboring clusters maximizing an externally given fitness function; Gabriel graphs are used to determine which clusters are neighboring, and non-convex shapes are approximated as the unions of small convex clusters. Experimental results are presented that show that using MOSAIC leads to clusters of higher quality compared to running a representative clustering algorithm stand-alone.

In Chapter 11, "Ranking Gradients in Multi-Dimensional Spaces," Ronnie Alves, Joel Ribeiro, Orlando Belo (Department of Informatics, University of Minho, Portugal) and Jiawei Han (Department of Computer Science, University of Illinois at Urbana-Champaign, USA) investigate how to mine and rank the most interesting changes in a multi-dimensional space applying a promising TOP-K gradient strategy. Interesting changes in customer behavior are usually discovered by gradient queries which are particular cases of multi-dimensional data analysis on large data warehouses. The main problem, however, arises from the fact that more interesting changes should be those ones having more dimensions in the gradient query (the curse-of-dimensionality dilemma). Besides, the number of interesting changes should be of a large amount (the preference selection criteria).

In Chapter 12, "Simultaneous Feature Selection and Tuple Selection for Efficient Classification," by Manoranjan Dash & Vivekanand Gopalkrishnan (Nanyang Technological University, Singapore), a method is proposed to combine feature selection and tuple selection to improve classification accuracy. Although feature selection and tuple selection have been studied earlier in various research areas such as machine learning, data mining, and so on, they have rarely been studied together. Feature selection and tuple selection help the classifier to focus better. The method is based on the principle that a representative subset has similar histogram as the full set. The proposed method uses this principle both to choose

a subset of features and also to choose a subset of tuples. The empirical tests show that the proposed method performs better than several existing feature selection methods.

The last four chapters in **Section 5, Advanced Mining Applications** introduces innovative algorithms and applications in some emerging application fields in data mining and knowledge discovery, especially continuous data stream mining, which could not be solved by traditional mining technology.

Alberto Freitas, Altamiro Costa-Pereira, Pavel Brazdil (Faculty of Medicine, University of Porto, Portugal) discuss in Chapter 13, "Learning cost-sensitive decision trees to support medical diagnosis," a cost-sensitive learning method. The chapter aims to enhance the understand of cost-sensitive learning problems in medicine and presents a strategy for learning and testing cost-sensitive decision trees, while considering several types of costs associated with problems in medicine. It begins with a contextualization and a discussion of the main types of costs. Then, reviews related work and presents a discussion about the evaluation of classifiers as well as explains a cost-sensitive decision tree strategy and presents some experimental results.

Chapter 14, "An Approximate Approach for Maintaining Recent Occurrences of Itemsets in a Sliding Window over Data Streams," by Jia-Ling Koh, Shu-Ning Shin, and Yuan-Bin Don (Department of Information Science and Computer Engineering National Taiwan Normal University Taipei, Taiwan), discusses about catching the recent trend of data when mining frequent itemsets over data streams. A data representation method, named frequency changing point (FCP), is introduced for monitoring the recent occurrence of itemsets over a data stream to prevent from storing the whole transaction data within a sliding window. The effect of old transactions on the mining result of recently frequent itemsets is diminished by performing adjusting rules on the monitoring data structure. Accordingly, the recently frequent itemsets or representative patterns are discovered from the maintained structure approximately. Experimental studies demonstrate that the proposed algorithms achieve high true positive rates and guarantees no false dismissal to the results yielded.

In Chapter 15, "Protocol Identification of Encrypted Network Streams," Matthew Gebski, Alex Penev, Raymond K. Wong (National ICT Australia and University of New South Wales, Australia) propose a simple machine learning approach for protocol identification in network streams that have been encrypted, such that the only information available for identifying the underlying protocol of a connection was the size, timing and direction of packets. With very little information available from the network stream, it is possible to pinpoint potentially inappropriate activities for a workplace, institution or research center, such as using BitTorrent, GMail or MSN, and not confuse them with other common protocols such as HTTP and SSL.

Finally, Chapter 16, "Exploring Calendar-based Pattern Mining in Data Streams," by Rodrigo Salvador Monteiro, Geraldo Zimbrão, Jano Moreira de Souza (COPPE / UFRJ - Brazil), Holger Schwarz, Bernhard Mitschang (IPVS - University of Stuttgart - Germany) introduces a calendar-based pattern mining aims at identifying patterns on specific calendar partitions in continuous data streams. The authors present how a data warehouse approach can be applied to leverage calendar-based pattern mining in data streams and how the framework of the DWFIST approach can cope with tight time constraints imposed by data streams, keep storage requirements at a manageable level and, at the same time, support calendar-based frequent itemset mining. The minimum granularity of analysis, parameters of the data warehouse (e.g., mining minimum support) and parameters of the database (e.g., extent size) provide ways to tune the load performance.

Overall, the sixteen book chapters, contributed by recognition experts and researchers in data warehousing and knowledge discovery, cover broad research areas on both theoretical and practical aspects of the fields. They present innovative principles, methods, algorithms and solutions to challenging

problems faced in the real applications with experimental results. Future trends, research challenges and opportunities are also mentioned and discussed.

Tho Manh Nguyen, Editor
Vienna University of Technology, Austria
January 2009

此为试读,需要完整PDF请访问: www.ertongbook.com

Section 1
DWH Architectures and Fundamentals

Table of Contents

Preface xvii

Section 1
DWH Architectures and Fundamentals

Chapter 1
The LBF R-Tree: Scalable Indexing and Storage for Data Warehousing Systems 1
Todd Eavis, Concordia University, Canada
David Cueva, Concordia University, Canada

Chapter 2
Dynamic Workload for Schema Evolution in Data Warehouses: A Performance Issue 28
Fadila Bentayeb, University of Lyon – ERIC Laboratory, France
Cécile Favre, University of Lyon – ERIC Laboratory, France
Omar Boussaid, University of Lyon – ERIC Laboratory, France

Chapter 3
Preview: Optimizing View Materialization Cost in Spatial Data Warehouses 47
Songmei Yu, Felician College, USA
Vijayalakshmi Atluri, Rutgers University, USA
Nabil Adam, Rutgers University, USA

Section 2
Multidimensional Data and OLAP

Chapter 4
Decisional Annotations: Integrating and Preserving Decision-Makers' Expertise in Multidimensional Systems 65
Guillaume Cabanac, Université de Toulouse, France
Max Chevalier, Université de Toulouse, France
Franck Ravat, Université de Toulouse, France
Olivier Teste, Université de Toulouse, France

Chapter 5
Federated Data Warehouses 82
Stefan Berger, University of Linz, Data & Knowledge Engineering Group, Austria
Michael Schrefl, University of Linz, Data & Knowledge Engineering Group, Austria

Chapter 6
Built-In Indicators to Support Business Intelligence in OLAP Databases 108
Jérôme Cubillé, EDF R&D, France
Christian Derquenne, EDF R&D, France
Sabine Goutier, EDF R&D, France
Françoise Guisnel, EDF R&D, France
Henri Klajnmic, EDF R&D, France
Véronique Cariou, ENITIAA, France

Section 3
DWH and OLAP Applications

Chapter 7
Conceptual Data Warehouse Design Methodology for Business Process Intelligence 129
Svetlana Mansmann, University of Konstanz, Konstanz, Germany
Thomas Neumuth, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany
Oliver Burgert, Innovation Center Computer Assisted Surgery (ICCAS), Leipzig, Germany
Matthias Röger, University of Konstanz, Konstanz, Germany
Marc H. Scholl, University of Konstanz, Konstanz, Germany

Chapter 8
Data Warehouse Facilitating Evidence-Based Medicine 174
Nevena Stolba, Vienna University of Technology, Austria
Tho Manh Nguyen, Vienna University of Technology, Austria
A Min Tjoa, Vienna University of Technology, Austria

Chapter 9
Deploying Data Warehouses in Grids with Efficiency and Availability 208
Rogério Luís de Carvalho Costa, University of Coimbra, Portugal
Pedro Furtado, University of Coimbra, Portugal

Section 4
Data Mining Techniques

Chapter 10
MOSAIC: Agglomerative Clustering with Gabriel Graphs 231
Rachsuda Jiamthapthaksin, University of Houston, USA
Jiyeon Choo, University of Houston, USA
Chun-sheng Chen, University of Houston, USA
Oner Ulvi Celepcikay, University of Houston, USA
Christian Giusti, University of Udine, Italy
Christoph F. Eick, University of Houston, USA

Chapter 11
Ranking Gradients in Multi-Dimensional Spaces 251
Ronnie Alves, University of Nice Sophia-Antipolis, France
Joel Ribeiro, University of Minho, Portugal
Orlando Belo, University of Minho, Portugal
Jiawei Han, University of Illinois at Urbana-Champaign, USA

Chapter 12
Simultaneous Feature Selection and Tuple Selection for Efficient Classification 270
Manoranjan Dash, Nanyang Technological University, Singapore
Vivekanand Gopalkrishnan, Nanyang Technological University, Singapore

Section 5
Advanced Mining Applications

Chapter 13
Learning Cost-Sensitive Decision Trees to Support Medical Diagnosis 287
Alberto Freitas, CINTESIS – Center for Research in Health Information Systems and Technologies, Portugal and University of Porto, Portugal
Altamiro Costa-Pereira, CINTESIS – Center for Research in Health Information Systems and Technologies, Portugal and University of Porto, Portugal
Pavel Brazdil, LIAAD INESC Porto L.A. – Laboratory of Artificial Intelligence and Decision Support, Portugal and University of Porto, Portugal

Chapter 14
An Approximate Approach for Maintaining Recent Occurrences of Itemsets in a Sliding Window over Data Streams 308
Jia-Ling Koh, National Taiwan Normal University, Taiwan
Shu-Ning Shin, National Taiwan Normal University, Taiwan
Yuan-Bin Don, National Taiwan Normal University, Taiwan

Chapter 15
Protocol Identification of Encrypted Network Streams 328
Matthew Gebski, National ICT Australia and University of New South Wales, Australia
Alex Penev, National ICT Australia and University of New South Wales, Australia
Raymond K. Wong, National ICT Australia and University of New South Wales, Australia

Chapter 16
Exploring Calendar-Based Pattern Mining in Data Streams 342
Rodrigo Salvador Monteiro, COPPE / UFRJ, Brazil
Geraldo Zimbrão, COPPE / UFRJ, Brazil
Holger Schwarz, IPVS - University of Stuttgart, Germany
Bernhard Mitschang, IPVS - University of Stuttgart, Germany
Jano Moreira de Souza, COPPE / UFRJ, Brazil

Compilation of References 361

About the Contributors 389

Index 400

Detailed Table of Contents

Preface xvii

Section 1
DWH Architectures and Fundamentals

Three chapters in Section 1 present the current trends of research on data warehouse architecture, storage and implementations which towards on improving performance and response time.

Chapter 1
The LBF R-Tree: Scalable Indexing and Storage for Data Warehousing Systems 1
Todd Eavis, Concordia University, Canada
David Cueva, Concordia University, Canada

In Chapter 1, the authors propose an LBF R-tree framework for effective indexing mechanisms in multi-dimensional database environment. The proposed framework addressed not only improves performance on common user-defined range queries, but also gracefully degrades to a linear scan of the data on pathologically large queries. Experimental results demonstrating both efficient disk access on the LBF R-tree, as well as impressive compression ratios for data and indexes.

Chapter 2
Dynamic Workload for Schema Evolution in Data Warehouses: A Performance Issue 28
Fadila Bentayeb, University of Lyon – ERIC Laboratory, France
Cécile Favre, University of Lyon – ERIC Laboratory, France
Omar Boussaid, University of Lyon – ERIC Laboratory, France

Chapter 2 addresses the issues related to the workload's evolution and maintenance in data warehouse systems in response to new requirements modeling resulting from users' personalized analysis needs. The proposed workload management system assists the administrator to maintain and adapt dynamically the workload according to changes arising on the data warehouse schema by improving two types of workload updates: (1) maintaining existing queries consistent with respect to the new data warehouse schema and (2) creating new queries based on the new dimension hierarchy levels.

此为试读,需要完整PDF请访问: www.ertongbook.com