Skylines and Other Dominance-Based Queries

Synthesis Lectures on Data Management

Synthesis Lectures on Data Management is edited by H.V. Jagadish of the University of Michigan. The series publishes 80–150 page publications on topics pertaining to data management. Topics include query languages, database system architectures, transaction management, data warehousing, XML and databases, data stream systems, wide scale data distribution, multimedia data management, data mining, and related subjects.

Transaction Processing on Modern Hardware
Mohammad Sadoghi and Spyros Blanas
2019

Data Management in Machine Learning Systems
Matthias Boehm, Arun Kumar, and Jun Yang
2019

Non-Volatile Memory Database Management Systems
Joy Arulraj and Andrew Pavlo
2019

Scalable Processing of Spatial-Keyword Queries
Ahmed R. Mahmood and Walid G. Aref
2019

Data Exploration Using Example-Based Methods
Matteo Lissandrini, Davide Mottin, Themis Palpanas, and Yannis Velegrakis
2018

Data Profiling
Ziawasch Abedjan, Lukasz Golab, Felix Naumann, and Thorsten Papenbrock
2018

Querying Graphs
Angela Bonifati, George Fletcher, Hannes Voigt, and Nikolay Yakovets
2018

Query Processing over Incomplete Databases
Yunjun Gao and Xiaoye Miao
2018

Natural Language Data Management and Interfaces
Yunyao Li and Davood Rafiei
2018

Human Interaction with Graphs: A Visual Querying Perspective
Sourav S. Bhowmick, Byron Choi, and Chengkai Li
2018

On Uncertain Graphs
Arijit Khan, Yuan Ye, and Lei Chen
2018

Answering Queries Using Views
Foto Afrati and Rada Chirkova
2017

Databases on Modern Hardware: How to Stop Underutilization and Love Multicores
Anatasia Ailamaki, Erieta Liarou, Pınar Tözün, Danica Porobic, and Iraklis Psaroudakis
2017

Instant Recovery with Write-Ahead Logging: Page Repair, System Restart, Media Restore, and System Failover, Second Edition
Goetz Graefe, Wey Guy, and Caetano Sauer
2016

Generating Plans from Proofs: The Interpolation-based Approach to Query Reformulation
Michael Benedikt, Julien Leblay, Balder ten Cate, and Efthymia Tsamoura
2016

Veracity of Data: From Truth Discovery Computation Algorithms to Models of Misinformation Dynamics
Laure Berti-Équille and Javier Borge-Holthoefer
2015

Datalog and Logic Databases
Sergio Greco and Cristina Molinaro
2015

Big Data Integration
Xin Luna Dong and Divesh Srivastava
2015

Instant Recovery with Write-Ahead Logging: Page Repair, System Restart, and Media Restore
Goetz Graefe, Wey Guy, and Caetano Sauer
2014

Similarity Joins in Relational Database Systems
Nikolaus Augsten and Michael H. Böhlen
2013

Information and Influence Propagation in Social Networks
Wei Chen, Laks V.S. Lakshmanan, and Carlos Castillo
2013

Data Cleaning: A Practical Perspective
Venkatesh Ganti and Anish Das Sarma
2013

Data Processing on FPGAs
Jens Teubner and Louis Woods
2013

Perspectives on Business Intelligence
Raymond T. Ng, Patricia C. Arocena, Denilson Barbosa, Giuseppe Carenini, Luiz Gomes, Jr., Stephan Jou, Rock Anthony Leung, Evangelos Milios, Renée J. Miller, John Mylopoulos, Rachel A. Pottinger, Frank Tompa, and Eric Yu
2013

Semantics Empowered Web 3.0: Managing Enterprise, Social, Sensor, and Cloud-based Data and Services for Advanced Applications
Amit Sheth and Krishnaprasad Thirunarayan
2012

Data Management in the Cloud: Challenges and Opportunities
Divyakant Agrawal, Sudipto Das, and Amr El Abbadi
2012

Query Processing over Uncertain Databases
Lei Chen and Xiang Lian
2012

Foundations of Data Quality Management
Wenfei Fan and Floris Geerts
2012

Incomplete Data and Data Dependencies in Relational Databases
Sergio Greco, Cristian Molinaro, and Francesca Spezzano
2012

Business Processes: A Database Perspective
Daniel Deutch and Tova Milo
2012

Data Protection from Insider Threats
Elisa Bertino
2012

Deep Web Query Interface Understanding and Integration
Eduard C. Dragut, Weiyi Meng, and Clement T. Yu
2012

P2P Techniques for Decentralized Applications
Esther Pacitti, Reza Akbarinia, and Manal El-Dick
2012

Query Answer Authentication
HweeHwa Pang and Kian-Lee Tan
2012

Declarative Networking
Boon Thau Loo and Wenchao Zhou
2012

Full-Text (Substring) Indexes in External Memory
Marina Barsky, Ulrike Stege, and Alex Thomo
2011

Spatial Data Management
Nikos Mamoulis
2011

Database Repairing and Consistent Query Answering
Leopoldo Bertossi
2011

Managing Event Information: Modeling, Retrieval, and Applications
Amarnath Gupta and Ramesh Jain
2011

Fundamentals of Physical Design and Query Compilation
David Toman and Grant Weddell
2011

Methods for Mining and Summarizing Text Conversations
Giuseppe Carenini, Gabriel Murray, and Raymond Ng
2011

Probabilistic Databases
Dan Suciu, Dan Olteanu, Christopher Ré, and Christoph Koch
2011

Peer-to-Peer Data Management
Karl Aberer
2011

Probabilistic Ranking Techniques in Relational Databases
Ihab F. Ilyas and Mohamed A. Soliman
2011

Uncertain Schema Matching
Avigdor Gal
2011

Fundamentals of Object Databases: Object-Oriented and Object-Relational Design
Suzanne W. Dietrich and Susan D. Urban
2010

Advanced Metasearch Engine Technology
Weiyi Meng and Clement T. Yu
2010

Web Page Recommendation Models: Theory and Algorithms
Sule Gündüz-Ögüdücü
2010

Multidimensional Databases and Data Warehousing
Christian S. Jensen, Torben Bach Pedersen, and Christian Thomsen
2010

Database Replication
Bettina Kemme, Ricardo Jimenez-Peris, and Marta Patino-Martinez
2010

Relational and XML Data Exchange
Marcelo Arenas, Pablo Barcelo, Leonid Libkin, and Filip Murlak
2010

User-Centered Data Management
Tiziana Catarci, Alan Dix, Stephen Kimani, and Giuseppe Santucci
2010

Data Stream Management
Lukasz Golab and M. Tamer Özsu
2010

Access Control in Data Management Systems
Elena Ferrari
2010

An Introduction to Duplicate Detection
Felix Naumann and Melanie Herschel
2010

Privacy-Preserving Data Publishing: An Overview
Raymond Chi-Wing Wong and Ada Wai-Chee Fu
2010

Keyword Search in Databases
Jeffrey Xu Yu, Lu Qin, and Lijun Chang
2009

Skylines and Other Dominance-Based Queries

Apostolos N. Papadopoulos, Eleftherios Tiakas, Theodoros Tzouramanis, Nikolaos Georgiadis, and Yannis Manolopoulos

ISBN: 978-3-031-00748-4 paperback
ISBN: 978-3-031-01876-3 ebook
ISBN: 978-3-031-00103-1 hardcover

DOI 10.1007/978-3-031-01876-3

A Publication in the Springer series

SYNTHESIS LECTURES ON DATA MANAGEMENT

Lecture #63
Series Editor: H.V. Jagadish, *University of Michigan*
Founding Editor: M. Tamer Özsu, *University of Waterloo*
Series ISSN
Print 2153-5418 Electronic 2153-5426

Skylines and Other Dominance-Based Queries

Apostolos N. Papadopoulos
School of Informatics, Aristotle University of Thessaloniki, Greece

Eleftherios Tiakas
School of Informatics, Aristotle University of Thessaloniki, Greece

Theodoros Tzouramanis
Department of Computer Science & Biomedical Informatics, University of Thessaly, Greece

Nikolaos Georgiadis
School of Informatics, Aristotle University of Thessaloniki, Greece

Yannis Manolopoulos
School of Pure & Applied Sciences, Open University of Cyprus, Cyprus

SYNTHESIS LECTURES ON DATA MANAGEMENT #63

ABSTRACT

This book is a gentle introduction to dominance-based query processing techniques and their applications. The book aims to present fundamental as well as some advanced issues in the area in a precise, but easy-to-follow, manner. Dominance is an intuitive concept that can be used in many different ways in diverse application domains. The concept of dominance is based on the values of the attributes of each object. An object p dominates another object q if p is *better* than q. This goodness criterion may differ from one user to another. However, all decisions boil down to the minimization or maximization of attribute values. In this book, we will explore algorithms and applications related to dominance-based query processing. The concept of dominance has a long history in finance and multi-criteria optimization. However, the introduction of the concept to the database community in 2001 inspired many researchers to contribute to the area. Therefore, many algorithmic techniques have been proposed for the efficient processing of dominance-based queries, such as skyline queries, k-dominant queries, and top-k dominating queries, just to name a few.

KEYWORDS

multi-dimensional data, preference-based queries, dominance, skyline queries, range skylines, skyline cubes, top-k dominating queries, k-dominance, dynamic skylines, spatial skylines, metric-based dominance queries, multi-criteria decision-making, applications

To our families

Contents

List of Figures

List of Tables

Preface

The main objective of this book is to provide an easy-to-follow, self-contained, and concise coverage of dominance-based query-processing techniques. During the last two decades a large corpus of relevant research results has been accumulated. For instance, if someone searches in a digital library, such as Digital Bibliography & Library Project (DBLP), she will realize that more than 1,000 of papers have been published in scientific journals and conference proceedings, besides to technical reports and theses.

It turns out that dominance-based queries have many interesting applications in diverse scientific fields. Therefore, the book may be followed easily by non-computer science students. For example, the application of dominance in finance, healthcare, Internet of Things, scientometrics, and many other fields makes this book relevant to a broader audience.

Most of this book does not require specialized background. However, the content is better suited for students and practitioners that have basic knowledge of data structures and database systems. More specifically, the target audience includes:

- graduate students who want to enhance their knowledge in more advanced query processing techniques. The material of the book may be part of a data management-oriented course;
- M.Sc. students who wish to better understand the concept and the applications of dominance-based query processing;
- Ph.D. students and researchers who wish to master the most important concepts in dominance-based query processing and may want to apply these ideas in more complex problems that involve user preferences; and
- practitioners who wish to learn about dominance-based queries and apply these ideas in data management and mining tasks in the field of their expertise.

We really hope that this book will be a valuable companion toward understanding the concepts and techniques related to dominance-based query processing. The interested reader can focus further on the selected bibliography. It is certain that the research community will devote more human and financial resources in the area in the years to come.

Apostolos N. Papadopoulos, Eleftherios Tiakas, Theodoros Tzouramanis, Nikolaos Georgiadis, and Yannis Manolopoulos
November 2020

Acknowledgments

The authors would like to thank Professors Hosagrahar Visvesvaraya Jagadish and Tamer Özsu for accepting our book in the *Syntesis Lectures in Data Management* series. Also, we are grateful to Diane Cerra and Christine Kiilerich for their great assistance during the preparation of the book. Moreover, the authors would like to acknowledge the assistance of collaborators and friends who were involved in research works cited in this book. More specifically, the authors are grateful to: Dimitrios Gunopulos, Dimitrios Katsaros, Maria Kontaki, Alexandros Nanopoulos, Timos Sellis, Yannis Theodoridis, and Georgios Valkanas.

Apostolos N. Papadopoulos, Eleftherios Tiakas, Theodoros Tzouramanis, Nikolaos Georgiadis, and Yannis Manolopoulos
November 2020