

Skylines and Other Dominance-Based Queries

Synthesis Lectures on Data Management

Editor

H.V. Jagadish, *University of Michigan*

Founding Editor

M. Tamer Özsu, *University of Waterloo*

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.

Skylines and Other Dominance-Based Queries

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

Cloud-Based RDF Data Management

Zoi Kaoudi, Ioana Manolescu, and Stamatis Zampetakis
2020

Community Search over Big Graphs

Xin Huang, Laks V.S. Lakshmanan, and Jianliang Xu
2019

On Transactional Concurrency Control

Goetz Graefe
2019

Data-Intensive Workflow Management: For Clouds and Data-Intensive and Scalable Computing Environments

Daniel C.M. de Oliveira, Ji Liu, and Esther Pacitti
2019

Answering Queries Using Views, Second Edition

Foto Afrati and Rada Chirkova
2019

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, Pinar 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

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2021

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

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	xvii
	List of Tables	xix
	Preface	xxi
	Acknowledgments	xxiii
1	Introduction	1
1.1	Objects and Attributes	1
1.2	The Concept of Dominance	2
1.3	Best Points	4
1.4	Book Roadmap	7
1.5	Summary	7
2	Skyline Queries	9
2.1	Main-Memory Computation	9
2.2	Algorithms for Secondary Memory	13
2.2.1	Index-Free Techniques	14
2.2.2	Index-Based Techniques	20
2.3	Advanced Skyline Processing	27
2.3.1	Distributed and Parallel Techniques	27
2.3.2	Skylines in Dynamic Environments	30
2.4	Skyline Cardinality	32
2.5	Summary	34
3	Variations of Skyline Queries	37
3.1	k -Dominant Skyline Queries	37
3.2	Skycube	40
3.3	Dynamic Skyline Queries	43
3.4	Spatial Skyline Queries	46
3.5	Metric Space Skyline Queries	51

3.6	Range-Based Skyline Queries	54
3.7	Other Variations	58
3.8	Summary	61
4	Top-k Dominating Queries	63
4.1	Problem Definition	63
4.2	A Skyline-Based Algorithm	64
4.3	Methods Based on R-Tree Variants	66
4.3.1	Iterative Top- k Dominating Algorithm (ITD)	68
4.3.2	Simple Counting Guided Algorithm (SCG)	69
4.3.3	Lightweight Counting Guided Algorithm (LCG)	71
4.3.4	Priority- and Upper Bound-Based Traversal Algorithms (PBT, UBT)	71
4.3.5	Cost-Based Traversal Algorithm (CBT)	72
4.4	Index-Free Algorithms	72
4.5	Subspace Progressive Algorithms	73
4.5.1	Basic Scan Algorithm (BSA)	76
4.5.2	Union Algorithm (UA)	78
4.5.3	Reverse Algorithm (RA)	79
4.5.4	Differential Algorithm (DA)	81
4.6	Metric-Based Approaches	83
4.6.1	Skyline-Based Algorithm (SBA)	85
4.6.2	Aggregation-Based Algorithm (ABA)	86
4.6.3	Pruning-Based Algorithms (PBA1, PBA2)	88
4.7	Top- k Dominating Queries in Other Environments	89
4.8	Summary	90
5	Applications of Dominance-Based Queries	91
5.1	Multi-Criteria Decision Making	91
5.1.1	Service Composition	91
5.1.2	Web Service Selection	92
5.1.3	Reliability in Cloud Computing	93
5.2	Machine Learning	93
5.2.1	Intrusion Detection	93
5.2.2	Polarity Classification	94
5.3	Network Analysis	96
5.3.1	Graph Clustering	96

5.3.2	Graph Community Search	98
5.4	Marketing	99
5.4.1	Targeting a General Group of Buyers	99
5.4.2	Product Feature Selection to Maximize Profitability	100
5.4.3	Effective Product Positioning	100
5.5	Healthcare	101
5.5.1	Online Primary Diagnosis	101
5.5.2	Over-Treatment Reduction	103
5.6	Other Interesting Applications	103
5.6.1	Image Retrieval	103
5.6.2	Scientometrics	104
5.6.3	Monitor of Chemical Process	106
5.6.4	Wireless Routing	106
5.6.5	Sensor Selection	106
5.6.6	E-Commerce	107
5.6.7	Indoor Route Search	108
5.7	Database Support for Dominance Queries	109
5.8	Summary	111
	Bibliography	113
	Authors' Biographies	131
	Index	133

List of Figures

1.1	The concept of dominance.	4
1.2	Design space (a) and criterion space (b).	5
2.1	Skyline in one dimension ($d = 1$).	10
2.2	Skyline query processing for $d = 2$	11
2.3	Divide and conquer algorithm example.	18
2.4	Progressive skyline example.	21
2.5	A set of rectangles (left) and an R-tree (right).	22
2.6	NNS algorithm example.	23
2.7	Distance from a point and a rectangle to the origin.	24
2.8	BBS algorithm example.	25
2.9	Examples of grid-based and angle-based partitioning.	29
2.10	Examples of handling insertions and deletions.	31
2.11	Different data distributions (examples for the two-dimensional case).	32
3.1	Skycube.	40
3.2	Dynamic skyline query.	44
3.3	Dynamic skyline query in a graph.	45
3.4	Spatial skyline query.	47
3.5	Nearest and the farthest spatial skyline queries.	48
3.6	Spatio-textual skyline query.	49
3.7	Direction-based spatial skyline query.	50
3.8	Metric skyline query.	53
3.9	Constrained skyline query.	55
3.10	Type of range-skyline query proposed in Rahul and Janardan (2012).	56
3.11	Privacy-preserving range-based skyline query.	57
3.12	Range skyline query.	59

4.1	Top-3 dominating query.	64
4.2	First iteration of STD algorithm.	65
4.3	Second iteration of STD algorithm.	66
4.4	Domination relationship among aR-tree entries.	68
4.5	SCG algorithm—constructing the heap H.	70
4.6	SCG algorithm—computing necessary scores.	70
4.7	Data organization utilized by BSA, UA, RA, and DA.	75
4.8	Detection of terminating objects and other definitions.	76
4.9	BSA algorithm processing.	78
4.10	UA algorithm processing.	80
4.11	RA algorithm processing.	81
4.12	Differential calculation with DA.	82
4.13	Metric-based top-3 dominating query.	84
4.14	Data organization used by metric-based top- k dominating algorithms.	85
4.15	ABA algorithm processing.	87
5.1	Airplane service pruning.	92
5.2	Service workflow model.	92
5.3	Proposed naive Bayesian classifier model.	95
5.4	Classification process.	96
5.5	Skyline of the features.	97
5.6	Skyline community search example for 2-core.	99
5.7	Architecture of CINEMA.	102
5.8	Medical analysis system.	104
5.9	Hash BBS skyline process.	105
5.10	Monitor process.	107
5.11	Dynamic skyline sensor selection.	108

List of Tables

- 1.1 Interesting points in original (design) space and transformed (criterion) space 5
- 2.1 Bitmap encodings example 19
- 3.1 Movie rating dataset 38
- 3.2 Other variations of skyline and other dominance-based queries (*Continues.*) . 60
- 3.2 (*Continued.*) Other variations of skyline and other dominance-based queries . 61
- 5.1 List of words 96

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