Querying Graphs

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.

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

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2018

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.

Querying Graphs

Angela Bonifati, George Fletcher, Hannes Voigt, and Nikolay Yakovets

ISBN: 978-3-031-00736-1 paperback
ISBN: 978-3-031-01864-0 ebook
ISBN: 978-3-031-00091-1 hardcover

DOI 10.1007/978-3-031-01864-0

A Publication in the Springer scries
SYNTHESIS LECTURES ON DATA MANAGEMENT

Lecture #51
Series Editor: H.V. Jagadish, *University of Michigan*
Founding Editor: M. Tamer Özsu, *University of Waterloo*
Series ISSN
Print 2153-5418 Electronic 2153-5426

Querying Graphs

Angela Bonifati
Université Claude Bernard Lyon 1

George Fletcher
Technische Universiteit Eindhoven

Hannes Voigt
Neo4j/Technische Universität Dresden[1]

Nikolay Yakovets
Technische Universiteit Eindhoven

SYNTHESIS LECTURES ON DATA MANAGEMENT #51

[1] Author is now at Neo4j. The book was mainly written while the author was still at Technische Universität Dresden.

ABSTRACT

Graph data modeling and querying arises in many practical application domains such as social and biological networks where the primary focus is on concepts and their relationships and the rich patterns in these complex webs of interconnectivity. In this book, we present a concise unified view on the basic challenges which arise over the complete life cycle of formulating and processing queries on graph databases. To that purpose, we present all major concepts relevant to this life cycle, formulated in terms of a common and unifying ground: the property graph data model—the pre-dominant data model adopted by modern graph database systems.

We aim especially to give a coherent and in-depth perspective on current graph querying and an outlook for future developments. Our presentation is self-contained, covering the relevant topics from: graph data models, graph query languages and graph query specification, graph constraints, and graph query processing. We conclude by indicating major open research challenges towards the next generation of graph data management systems.

KEYWORDS

graph databases, property graphs, graph query languages, graph constraints, graph query specification, graph data representation, graph query processing

Contents

Foreword

The current surge of interest in Graph Data Bases (GDBs) reflects the popularity of their data models based on nodes and edges, which, in many applications, provide a more intuitive conceptualization for entities and relationships than the one offered by Relational Data Bases (RDBs). This has inspired the design and development of many GDB systems and their use in a wide range of applications. Indeed to date, we counted more than 20 GDB systems developed and used in application areas such as Semantic Web, Social Networking, Fraud Detection, Recommendation Systems, Life Science, and Knowledge Bases.

For all their remarkable achievements, GDBs still lack the conceptual coherence that RDBs have been blessed with from the beginning as a result of E.F. Codd's seminal contributions which, combined with the major research advances in theory and systems that followed, provide the subject of numerous textbooks. However, the fast-expanding technology of GDBs is still quite far from achieving similar levels of conceptual unification and this create hurdles for researchers, instructors, and students alike.

This book tackles this problem head on by presenting a comprehensive unified treatment of GDBs, as needed to serve as a reference book for experts and a textbook for graduate students. The book's coverage begins with a formal treatment of the Property Graph Data Model that is common to most GDBs. Then, the book discusses GDB query languages and, moving past their many differences, it proposes a core property graph query language and elucidates its properties both in terms of graph logic and graph algebra. After that, the book covers techniques for efficient GDB implementation, including data structures, indexes, query operators, and processing, for which the presentation underscores how solutions different from those of traditional DBs are often required. Furthermore, the departures from traditional technology are even more dramatic for (i) integrity constraints, which lose their key role in normal-form RDB schema design, but find new important uses in GDBs, and (ii) interactive query specification via examples and counter-examples that have proven to be surprisingly effective with GDBs. The book's comprehensive treatment is further enhanced by extensive references and suggestions on open research problems for further investigation.

Carlo Zaniolo
Computer Science Department
University of California at Los Angeles (UCLA)

Acknowledgments

The authors would like to warmly thank the many people who helped us to make this book a reality. First and foremost, we thank our families and partners for their patience and support throughout the many months dedicated to the writing of this book.

We also give many thanks to the Editor H.V. Jagadish and the Founding Editor M. Tamer Özsu for the opportunity and encouragement to publish this book. During the writing, the staff at Morgan & Claypool were just awesome, especially Diane Cerra. Thank you all for keeping the writing moving forward. We also thank the three reviewers for their critical and insightful feedback.

Our sincere thanks further go to Carlo Zaniolo for kindly writing the Foreword. We are greatly honored by your contribution!

Finally, we give our heartfelt thanks to colleagues for reading early drafts. We especially thank Sourav Bhowmick, Stefania Dumbrava, Jan Hidders, Wilco van Leeuwen, Davide Mottin, Oskar van Rest, and Kaijie Zhu for their detailed proofreading and helpful comments.

The work presented here was supported in part by a donation from Oracle Labs and by the CNRS Mastodons grant MedClean (2016–2018).

Angela Bonifati, George Fletcher, Hannes Voigt, and Nikolay Yakovets
September 2018