Cloud-Based RDF Data Management

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.

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

Cloud-Based RDF Data Management

Zoi Kaoudi, Ioana Manolescu, and Stamatis Zampetakis

ISBN: 978-3-031-00747-7 paperback
ISBN: 978-3-031-01875-6 ebook
ISBN: 978-3-031-00102-4 hardcover

DOI 10.1007/978-3-031-01875-6

A Publication in the Springer series

SYNTHESIS LECTURES ON DATA MANAGEMENT

Lecture #62
Series Editor: H.V. Jagadish, *University of Michigan*
Founding Editor: M. Tamer Özsu, *University of Waterloo*
Series ISSN
Print 2153-5418 Electronic 2153-5426

Cloud-Based RDF Data Management

Zoi Kaoudi
Technische Universität Berlin

Ioana Manolescu
INRIA

Stamatis Zampetakis
TIBCO Orchestra Networks

SYNTHESIS LECTURES ON DATA MANAGEMENT #62

ABSTRACT

Resource Description Framework (or RDF, in short) is set to deliver many of the original semi-structured data promises: flexible structure, optional schema, and rich, flexible Universal Resource Identifiers as a basis for information sharing. Moreover, RDF is uniquely positioned to benefit from the efforts of scientific communities studying databases, knowledge representation, and Web technologies. As a consequence, the RDF data model is used in a variety of applications today for integrating knowledge and information: in open Web or government data via the Linked Open Data initiative, in scientific domains such as bioinformatics, and more recently in search engines and personal assistants of enterprises in the form of knowledge graphs.

Managing such large volumes of RDF data is challenging due to the sheer size, heterogeneity, and complexity brought by RDF reasoning. To tackle the size challenge, distributed architectures are required. Cloud computing is an emerging paradigm massively adopted in many applications requiring distributed architectures for the scalability, fault tolerance, and elasticity features it provides. At the same time, interest in massively parallel processing has been renewed by the MapReduce model and many follow-up works, which aim at simplifying the deployment of massively parallel data management tasks in a cloud environment.

In this book, we study the state-of-the-art RDF data management in cloud environments and parallel/distributed architectures that were not necessarily intended for the cloud, but can easily be deployed therein. After providing a comprehensive background on RDF and cloud technologies, we explore four aspects that are vital in an RDF data management system: data storage, query processing, query optimization, and reasoning. We conclude the book with a discussion on open problems and future directions.

KEYWORDS

RDF, cloud computing, MapReduce, key-value stores, query optimization, reasoning

Contents