Data-Centric Systems and Applications

Series editors

M.J. Carey
S. Ceri

Editorial Board

A. Ailamaki
S. Babu
P. Bernstein
J.C. Freytag
A. Halevy
J. Han
D. Kossmann
I. Manolescu
G. Weikum
K.-Y. Whang
J.X. Yu

More information about this series at
http://www.springer.com/series/5258

Minos Garofalakis • Johannes Gehrke •
Rajeev Rastogi
Editors

Data Stream Management

Processing High-Speed Data Streams

Springer

Editors
Minos Garofalakis
School of Electrical and
Computer Engineering
Technical University of Crete
Chania, Greece

Johannes Gehrke
Microsoft Corporation
Redmond, WA, USA

Rajeev Rastogi
Amazon India
Bangalore, India

ISSN 2197-9723
Data-Centric Systems and Applications
ISBN 978-3-540-28607-3
DOI 10.1007/978-3-540-28608-0

ISSN 2197-974X (electronic)

ISBN 978-3-540-28608-0 (eBook)

Library of Congress Control Number: 2016946344

Springer Heidelberg New York Dordrecht London
© Springer-Verlag Berlin Heidelberg 2016
The fourth chapter in part 4 is published with kind permission of © 2004 Association for Computing Machinery, Inc.. All rights reserved.
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Data Stream Management: A Brave New World 1
Minos Garofalakis, Johannes Gehrke, and Rajeev Rastogi

Part I Foundations and Basic Stream Synopses

Data-Stream Sampling: Basic Techniques and Results 13
Peter J. Haas

Quantiles and Equi-depth Histograms over Streams 45
Michael B. Greenwald and Sanjeev Khanna

Join Sizes, Frequency Moments, and Applications 87
Graham Cormode and Minos Garofalakis

Top-*k* Frequent Item Maintenance over Streams 103
Moses Charikar

Distinct-Values Estimation over Data Streams 121
Phillip B. Gibbons

The Sliding-Window Computation Model and Results 149
Mayur Datar and Rajeev Motwani

Part II Mining Data Streams

Clustering Data Streams 169
Sudipto Guha and Nina Mishra

Mining Decision Trees from Streams 189
Geoff Hulten and Pedro Domingos

Frequent Itemset Mining over Data Streams 209
Gurmeet Singh Manku

Temporal Dynamics of On-Line Information Streams 221
Jon Kleinberg

Part III Advanced Topics

Sketch-Based Multi-Query Processing over Data Streams 241
Alin Dobra, Minos Garofalakis, Johannes Gehrke, and Rajeev Rastogi

Approximate Histogram and Wavelet Summaries of Streaming Data . . 263
S. Muthukrishnan and Martin Strauss

Stable Distributions in Streaming Computations 283
Graham Cormode and Piotr Indyk

Tracking Queries over Distributed Streams 301
Minos Garofalakis

Part IV System Architectures and Languages

STREAM: The Stanford Data Stream Management System 317
Arvind Arasu, Brian Babcock, Shivnath Babu, John Cieslewicz, Mayur Datar, Keith Ito, Rajeev Motwani, Utkarsh Srivastava, and Jennifer Widom

The Aurora and Borealis Stream Processing Engines 337
Uğur Çetintemel, Daniel Abadi, Yanif Ahmad, Hari Balakrishnan, Magdalena Balazinska, Mitch Cherniack, Jeong-Hyon Hwang, Samuel Madden, Anurag Maskey, Alexander Rasin, Esther Ryvkina, Mike Stonebraker, Nesime Tatbul, Ying Xing, and Stan Zdonik

Extending Relational Query Languages for Data Streams 361
N. Laptev, B. Mozafari, H. Mousavi, H. Thakkar, H. Wang, K. Zeng, and Carlo Zaniolo

Hancock: A Language for Analyzing Transactional Data Streams . . . 387
Corinna Cortes, Kathleen Fisher, Daryl Pregibon, Anne Rogers, and Frederick Smith

Sensor Network Integration with Streaming Database Systems 409
Daniel Abadi, Samuel Madden, and Wolfgang Lindner

Part V Applications

Stream Processing Techniques for Network Management 431
Charles D. Cranor, Theodore Johnson, and Oliver Spatscheck

High-Performance XML Message Brokering 451
Yanlei Diao and Michael J. Franklin

Fast Methods for Statistical Arbitrage 473
Eleftherios Soulas and Dennis Shasha

Adaptive, Automatic Stream Mining 499
Spiros Papadimitriou, Anthony Brockwell, and Christos Faloutsos

Conclusions and Looking Forward 529
Minos Garofalakis, Johannes Gehrke, and Rajeev Rastogi