

---

# Advances in Database Technology — EDBT 2014

17th International Conference  
on Extending Database Technology  
Athens, Greece, March 24–28, 2014  
Proceedings

*Editors*

Sihem Amer-Yahia  
Vassilis Christophides  
Anastasios Kementsietsidis  
Minos Garofalakis  
Stratos Idreos  
Vincent Leroy

---



Advances in Database Technology — EDBT 2014  
Proceedings of the 17th International Conference  
on Extending Database Technology  
Athens, Greece, March 24–28, 2014

*Editors*

Sihem Amer-Yahia, CNRS – LIG, France  
Vassilis Christophides, University of Crete, Greece, and R&I Center, Technicolor, Paris, France  
Anastasios Kementsietsidis, IBM Research, USA  
Minos Garofalakis, Technical University of Crete, Greece  
Stratos Idreos, Harvard University, USA  
Vincent Leroy, University of Grenoble & CNRS – LIG, France



OpenProceedings.org  
University of Konstanz  
University Library  
78457 Konstanz, Germany

COPYRIGHT NOTICE: Copyright © 2014 by the authors of the individual papers.

Distribution of all material contained in this volume is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0

OpenProceedings ISBN: 978-3-89318065-3

DOI of this front matter: 10.5441/002/edbt.2014.01

# Foreword

The 17th International Conference on Extending Database Technology (EDBT 2014) was held in Athens, Greece, March 24–28, 2014. Originally biennial, the EDBT conference has been held annually and jointly with ICDT (“Conference on Database Theory”) since 2009.

The proceedings of EDBT 2014 include an overview of a keynote address by Serge Abiteboul (INRIA), an overview of a keynote address by Peter Boncz (CWI), a laudation concerning the EDBT 2014 Test of Time Award, and 49 research papers and 3 vision papers that were selected by the Program Committee from 266 submissions.

The EDBT 2014 Test of Time Award is given to the paper *A Condensation Approach to Privacy Preserving Data Mining* by Charu C. Aggarwal, Philip S. Yu, which originally appeared in the proceedings of EDBT 2009.

I thank Walid Aref, Gustavo Alonso and Torsten Grust for their help in selecting the EDBT 2014 Test of Time Award. Warmest congratulations to the authors of this award winning paper!

I thank all authors who submitted papers to EDBT 2014. I would also like to thank all members of the Program Committee, and the many external reviewers, for the enormous amount of work they have done. The Program Committee did not meet in person, but carried out extensive discussions during the electronic PC meeting. I thank Andrei Voronkov for his EasyChair system, which made it easy to manage and coordinate the discussion.

I thank the EDBT Council members for their advice on issues of policy during the conference; I am particularly grateful to Norman Paton (EDBT 2013 PC Chair) for his advice. Special thanks also go to Vincent Leroy and Grigoris Karvounarakis, the Proceedings and Publicity Chairs of EDBT/ICDT 2014. I thank many colleagues involved in the organisation of the conference for fruitful collaboration, in particular, Vassilis Christophides (EDBT/ICDT 2014 Conference Chair).

Sihem Amer-Yahia  
EDBT 2014 Program Chair

# Program Committee Members and External Reviewers

## Program Committee

Serge Abiteboul (INRIA)  
Chengkai Li (UTA)  
David Gross-Amblard (IRISA)  
Christian Jensen (Aarhus University)  
Stephane Bressan (NUS)  
Jignesh Patel (U Wisconsin)  
Anastasia Ailamaki (EPFL)  
Ioanna Lykourantzou (INRIA)  
Torsten Grust (U Tuebingen)  
Theodore Johnson (at&t labs)  
Joos-Hendrik Böse (SAP)  
Norman Paton (U Manchester)  
Reza Akbarinia (INRIA)  
Stefan Manegold (CWI)  
Melanie Herschel (Paris Sud)  
Alexis Joly (INRIA)  
Andrea Cali (U. London)  
Vagelis Hristidis (UCR)  
Gustavo Alonso (ETH)  
Ioana Manolescu (INRIA)  
Wendy Hui Wang (Stevens)  
Mouna Kacimi (U Bozano)  
Selcuk Candan (ASU)  
Noha Ibrahim (LIG)  
Bernd Amann (LIP6)  
Amélie Marian (Rutgers U)  
Ihab Ilyas (U Waterloo)  
Verena Kantere (U Geneva)  
Bogdan Cautis (Paris Sud)  
Jean-Marc Petit (LIRIS)  
Walid Aref (Purdue)  
Florent Masseglia (INRIA)  
Evaggelia Pitoura (U Ioannina)  
Yaron Kanza (Technion)  
Stefano Ceri (Polytecnico Milano)  
Philippe Pucheral (INRIA)  
Denilson Barbosa (U Alberta)  
Guido Moerkotte (U Mannheim)  
Tran Quoc Trung (NUS)  
Grigoris Karvounarakis (LogicBlox)  
Yi Chen (ASU)  
Maya Ramanath (IIT Delhi)  
Luciano Barbosa (IBM)  
Mohamed Mokbel (U Minesota)  
Rajeev Rastogi (Amazon)  
Bettina Kemme (McGill)  
En Cheng (U Akron)

Philippe Rigaux (CNAM & Internet Mem.)  
Mahmoud Barhamgi (CNRS, LIRIS)  
Hannes Muehleisen (CWI)  
Elke Rudensteiner (PWI)  
Anne-Marie Kermarrec (INRIA)  
Alfredo Cuzzocrea (U Calabria)  
Maria Luisa Sapino (U Torino)  
Senjuti Basu Roy (U. Washington Tacoma)  
Felix Naumann (HPI)  
Ralf Schenkel (MPI)  
Yannis Kotidis (AUEB)  
Theodore Dalamagas (IMIS Athens)  
Pierre Senellart (Télécom ParisTech)  
Srikanta Bedathur (IIIT Delhi)  
Dan Olteanu (Oxford U)  
Dennis Shasha (NYU)  
Nick Koudas (U Toronto)  
Gautam Das (UTA)  
Alkis Simitsis (HP Labs)  
Omar Benjelloun (Google)  
Mourad Ouzzani (QCRI)  
Divesh Srivastava (at&t labs)  
Sreenivasa Kumar (IIT Madras)  
Amr El Abbadi (UCSB)  
Evimaria Terzi (Boston U)  
Francesco Bonchi (Yahoo!)  
Themis Palpanas (Paris Descartes)  
Martin Theobald (U Antwerp)  
Alexandros Labrinidis (U Pittsburgh)  
Irina Fundulaki (FORTH)  
Maurice Van Keulen (U Twente)  
Angela Bonifati (INRIA)  
Dimitris Papadias (HKUST)  
Vasilis Vassalos (AUEB)  
Alberto Laender (UFMG)  
Giorgio Ghelli (U Pisa)  
Yannis Velegarakis (U Trento)  
Philippe Bonnet (ITU)  
Paolo Papotti (QCRI)  
Gottfried Vossen (U Muenster)  
Vincent Leroy (LIG)  
Shantanu Godbole (IBM)  
Gerhard Weikum (Max Planck MPI)  
Luc Bouganim (INRIA)  
Aditya Parameswaran (Stanford U)  
Cong Yu (Google)  
Chengkai Li (UTA)

## External Reviewers

Abujarour, Mohammed  
Alaggan, Mohammad  
Alagiannis, Ioannis  
Alarabi, Louai  
Allard, Tristan  
Aly, Ahmed  
Anciaux, Nicolas  
Arora, Vaibhav  
Asudeh, Abolfazl  
Athanassoulis, Manos  
Avery, James  
Ayat, Naser  
Badieh Habib, Mena  
Bedathur, Srikanta  
Bonaque, Raphael  
Boneva, Iovka  
Butterstein, Dennis  
Bykau, Sjarhei  
Camacho-Rodríguez, Jesús  
Carvalho, Moises  
Chakraborti, Sutanu  
Champ, Julien  
Chasseur, Craig  
Cheng, Shiwen  
Christopoulos, Andreas  
Ciucanu, Radu  
Dallachiesa, Michele  
Das, Mahashweta  
de Graaff, Victor  
Deshmukh, Harshad  
Dinh, Thanh-Cong  
Doulkeridis, Christos  
Du, Xiaofei  
Eldawy, Ahmed  
Elser, Benedikt  
Fan, Qi  
Fekete, David  
Frey, Davide  
Gambis, Sebastien  
Georgiou, Theodore  
Georgoulas, Konstantinos  
Gheorghiu, Roxana  
Giannopoulos, Giorgos  
Haijan, Sara  
Hasan Dalip, Daniel  
Hassan, Naeemul  
Heise, Arvid  
Héman, Sándor  
Jayaram, Nandish

Jayaram, Nandish  
Kaoudi, Zoi  
Karanasos, Konstantinos  
Karoos, Kostis  
Karpathiotakis, Manos  
Karuvath, Sangeetha  
Katsipoulakis, Nick R.  
Kaul, Manohar  
Kenkre, Sreyash  
Kim, Mijung  
Kokou Dedzoe, William  
Kourtellis, Nicolas  
Kyzirakos, Kostis  
Lang, Willis  
Lechtenböcker, Jens  
Li, Lu  
Liarou, Erietta  
Lietard, Ludovic  
Liroz Gistau, Miguel  
Lissandrini, Matteo  
Liu, Yunzhong  
Lydia, Abrouk  
M. Moro, Mirella  
Magdy, Amr  
Maniu, Silviu  
Mathieu, Lafourcade  
Mazzara, Manuel  
Miklos, Zoltan  
Mirylenka, Katsiaryna  
Moise, Diana  
Mueller, Tobias  
Nawab, Faisal  
Nazi, Azade  
Nizar, Abdul  
Nobari, Sadegh  
Papaggelis, Manos  
Parchas, Panos  
Pensa, Ruggero  
Penta, Antonio  
Pflanzl, Nicolas  
Plachouras, Vassilis  
Poppe, Olga  
Provetti, Alessandro  
Psaroudakis, Iraklis  
R. Vieira, Marcos  
Rahman, Habibur  
Raschia, Guillaume  
Raykar, Vikas  
Roatis, Alexandra

Roatis, Alexandra  
Ruiz, Eduardo  
Sacharidis, Dimitris  
Sadah, Shouq  
Sahin, Cetin  
Salihoglu, Semih  
Sandu Popa, Iulian  
Sankaranarayanan, Karthik  
Sans, Virginie  
Santos, Rodrygo L.T.  
Sarwat, Mohamed  
Schifanella, Claudio  
Schomm, Fabian  
Scuturici, Marian  
Shahbazi, Moloud  
Shan, Yi  
Sidirourgos, Lefteris  
Sidlauskas, Darius  
Silva Parreiras, Fernando  
Skoutas, Dimitrios  
Skovsgaard, Anders  
Spyropoulos, Vasilis  
Stahl, Florian  
Stefanidis, Kostas  
Sultana, Afroza  
Tang, Ruiming  
Termier, Alexandre  
Terrovitis, Manolis  
Thirumuruganathan, Saravanan  
Thoma, Cory  
To, Cuong  
Torp, Kristian  
Tzompanaki, Katerina  
Ulrich, Alexander  
Vartak, Manasi  
Venetis, Tassos  
Vergoulis, Thanasis  
Vo, Hoang Tam  
Vouzoukidou, Nelly  
Wang, Chong  
Wang, Guoping  
Wiley, Matthew  
Wu, Yinghui  
Yang, Bin  
Yehia, Taher  
Zhou, Zhuojie  
Zhu, Haohan  
Zoumpatianos, Kostas

# Table of Contents

Foreword .....	i–i
PC and External Reviewers .....	ii–iii
Table of Contents .....	iv–xiii
PStorM: Profile Storage and Matching for Feedback-Based Tuning of MapReduce Jobs Mostafa Ead, Herodotos Herodotou, Ashraf Aboulnaga and Shivnath Babu .....	1–12
Optimization Techniques for “Scaling Down” Hadoop on Multi-Core, Shared-Memory Systems K. Ashwin Kumar, Jonathan Gluck, Amol Deshpande and Jimmy Lin .....	13–24
Redoop: Supporting Recurring Queries in Hadoop Chuan Lei, Elke Rundensteiner and Mohamed Eltabakh .....	25–36
Efficient Skyline Computation in MapReduce Kasper Mullesgaard, Jens Laurits Pedersen, Hua Lu and Yongluan Zhou .....	37–48
Projected Clustering for Huge Data Sets in MapReduce Sergej Fries, Stephan Wels and Thomas Seidl .....	49–60
Efficient and Flexible Index Access in MapReduce Zhao Cao, Shimin Chen, Dongzhe Ma, Jianhua Feng and Min Wang .....	61–72
JISC: Adaptive Stream Processing Using Just-In-Time State Completion Ahmed M. Aly, Walid G. Aref, Mourad Ouzzani and Hosam M. Mahmoud .....	73–84
Adaptive Fault-Tolerance for Dynamic Resource Provisioning in Distributed Stream Processing Systems Paolo Bellavista, Antonio Corradi, Spyros Kotoulas and Andrea Reale .....	85–96
Adapting Tree Structures for Processing with SIMD Instructions Steffen Zeuch, Johann-Christoph Freytag and Frank Huber .....	97–108
Contender: A Resource Modeling Approach for Concurrent Query Performance Prediction Jennie Duggan, Olga Papaemmanouil, Ugur Cetintemel and Eli Upfal .....	109–120
CAQE: A Contract Driven Approach to Processing Concurrent Decision Support Queries Venkatesh Raghavan and Elke Rundensteiner .....	121–132
Multi-Query Diversification in Microblogging Posts Shiwen Cheng, Anastasios Arvanitis, Marek Chrobak and Vagelis Hristidis .....	133–144
Query-Based Why-Not Provenance with NedExplain Nicole Bidoit, Melanie Herschel and Katerina Tzompanaki .....	145–156
Corroborating Facts from Affirmative Statements Minji Wu and Amélie Marian .....	157–168
Overcoming Semantic Drift in Information Extraction Zhixu Li, Hongsong Li, Haixun Wang, Yi Yang, Xiangliang Zhang and Xiaofang Zhou .....	169–180
COLARM: Cost-based Optimization for Localized Association Rule Mining Abhishek Mukherji, Elke Rundensteiner and Matthew Ward .....	181–192
Fast Mining of Interesting Phrases from Subsets of Text Corpora Deepak P, Atreyee Dey and Debapriyo Majumdar .....	193–204

Sequence Pattern Matching over Time-Series Data with Temporal Uncertainty Yongluan Zhou, Chunyang Ma, Qingsong Guo, Lidan Shou and Gang Chen .....	205–216
Spatial Data Management Challenges in the Simulation Sciences Thomas Heinis, Farhan Tauheed and Anastasia Ailamaki .....	217–222
What Can Programming Languages Say About Data Exchange? Michael Johnson, Jorge Pérez and James Terwilliger .....	223–228
Toward Hardware-Sensitive Database Operations David Broneske, Sebastian Breß, Max Heimes and Gunter Saake .....	229–234
Distributed Spatial Keyword Querying on Road Networks Siqiang Luo, Yifeng Luo, Shuigeng Zhou, Gao Cong and Jihong Guan .....	235–246
Continuous Quantile Query Processing in Wireless Sensor Networks Johannes Niedermayer, Mario Nascimento, Matthias Renz, Peer Kröger and Hans-Peter Kriegel ...	247–258
RIPPLE: A Scalable Framework for Distributed Processing of Rank Queries George Tsatsanifos, Dimitris Sacharidis and Timos Sellis .....	259–270
HCS: Hierarchical Cut Selection for Efficiently Processing Queries on Data Columns using Hierarchical Bitmap Indices Parth Nagarkar and K. Selcuk Candan .....	271–282
Adaptive String Dictionary Compression in In-Memory Column-Store Database Systems Ingo Müller, Cornelius Ratsch and Franz Färber .....	283–294
Online Topic-aware Influence Maximization Queries Cigdem Aslay, Nicola Barbieri, Francesco Bonchi and Ricardo Baeza-Yates .....	295–306
Determining Essential Statistics for Cost Based Optimization of an ETL Workflow Ramanujam S Halasipuram, Prasad M Deshpande and Sriram Padmanabhan .....	307–318
CLUDE: An Efficient Algorithm for LU Decomposition Over a Sequence of Evolving Graphs Chenghui Ren, Luyi Mo, Ben Kao, Reynold Cheng and David W. Cheung .....	319–330
Model Selection for Semi-Supervised Clustering Mojgan Pourrajabi, Davoud Moulavi, Ricardo Campello, Arthur Zimek, Jörg Sander and Randy Goebel	331–342
Spatial Partitioning of Large Urban Road Networks Tarique Anwar, Chengfei Liu, Hai L Vu and Christopher Leckie .....	343–354
ENFrame: A Platform for Processing Probabilistic Data Sebastiaan van Schaik, Dan Olteanu and Robert Fink .....	355–366
Diversified Spatial Keyword Search On Road Networks Chengyuan Zhang, Ying Zhang, Wenjie Zhang, Xuemin Lin, Muhammad Cheema and Xiaoyang Wang ...	367–378
Cleaning trajectory data of RFID-monitored objects through conditioning under integrity constraints Bettina Fazzinga, Sergio Flesca, Filippo Furfaro and Francesco Parisi .....	379–390
Multi-Criteria Optimal Location Query with Overlapping Voronoi Diagrams Ji Zhang, Wei-Shinn Ku, Min-Te Sun, Xiao Qin and Hua Lu .....	391–402
Efficient Concept-based Document Ranking Anastasios Arvanitis, Matthew Wiley and Vagelis Hristidis .....	403–414

Metric-Based Top-k Dominating Queries Eleftherios Tiakas, George Valkanas, Apostolos N. Papadopoulos and Yannis Manolopoulos	415–426
A Unified Framework for Efficiently Processing Ranking Related Queries Muhammad Cheema, Zhitao Shen, Xuemin Lin and Wenjie Zhang	427–438
Exploiting the query structure for efficient join ordering in SPARQL queries Andrey Gubichev and Thomas Neumann	439–450
Interactive Inference of Join Queries Angela Bonifati, Radu Ciucanu and Slawek Staworko	451–462
Processing Interval Joins On Map-Reduce Bhupesh Chawda, Himanshu Gupta, Sumit Negi, Tanveer Faruque, L V Subramaniam and Mukesh Mohania	463–474
Differentially Private Synthesization of Multi-Dimensional Data using Copula Functions Haoran Li, Li Xiong and Xiaoqian Jiang	475–486
Privacy-Preserving Query Execution using a Decentralized Architecture and Tamper Resistant Hardware Quoc-Cuong To, Benjamin Nguyen and Philippe Pucheral	487–498
Privacy Preserving Similarity Evaluation of Time Series Data Haohan Zhu, Xianrui Meng and George Kollios	499–510
Reachability Queries in Very Large Graphs: A Fast Refined Online Search Approach Renê R. Veloso, Loic Cerf, Wagner Meira Junior and Mohammed J. Zaki	511–522
Graph Analytics on Massive Collections of Small Graphs Dritan Bleco and Yannis Kotidis	523–534
Fast Reliability Search in Uncertain Graphs Arijit Khan, Francesco Bonchi, Aris Gionis and Francesco Gullo	535–546
Distance oracles in edge-labeled graphs Francesco Bonchi, Aristides Gionis, Francesco Gullo and Antti Ukkonen	547–558
Privacy Preserving Estimation of Social Influence Tamir Tassa and Francesco Bonchi	559–570
A Privacy-Preserving Framework for Personalized, Social Recommendations Zach Jorgensen and Ting Yu	571–582
L-opacity: Linkage-Aware Graph Anonymization Sadegh Nobari, Panagiotis Karras, Hweehwa Pang and Stéphane Bressan	583–594
Privacy Risk in Anonymized Heterogeneous Information Networks Aston Zhang, Xing Xie, Kevin Chen-Chuan Chang, Carl A. Gunter, Jiawei Han and Xiaofeng Wang	595–606
A Condensation Approach to Privacy Preserving Data Mining Charu C. Aggarwal and Philip S. Yu	607–607
Mindmap-Inspired Semantic Personal Information Management Jenny Rompa, Christos Tryfonopoulos, Costas Vassilakis and George Lepouras	608–611
READ: Rapid data Exploration, Analysis and Discovery Udayan Khurana, Srinivasan Parthasarathy and Deepak Turaga	612–615



Demonstrating Self-Learning Algorithm Adaptivity in a Hardware-Oblivious Database Engine Max Heimel, Filip Haase, Martin Meinke, Sebastian Breß, Michael Saecker and Volker Markl	616–619
SECRETA: A System for Evaluating and Comparing RELational and Transaction Anonymization algorithms Giorgos Poulis, Aris Gkoulalas-Divanis, Grigorios Loukides, Spiros Skiadopoulos and Christos Tryfonopoulos	620–623
SIAS-V in Action: Snapshot Isolation Append Storage - Vectors on Flash Robert Gottstein, Thorsten Peter, Iliia Petrov and Alejandro Buchmann	624–627
inWalk: Interactive and Thematic Walks inside the Web of Data Silvana Castano, Alfio Ferrara and Stefano Montanelli	628–631
KIEV: a Tool for Extracting Semantic Relations from the World Wide Web Naimdjon Takhirov, Fabien Duchateau, Trond Aalberg and Ingeborg Torvik Solvberg	632–635
AGGREGO SEARCH: Interactive Keyword Query Construction Gregory Smits, Olivier Pivert, Helene Jaudoin and Francois Paulus	636–639
R2G: a Tool for Migrating Relations to Graphs Roberto De Virgilio, Antonio Maccioni and Riccardo Torlone	640–643
Large-scale Semantic Profile Extraction Michael Gubanov and Michael Stonebraker	644–647
ALIAS: Author Disambiguation in Microsoft Academic Search Engine Dataset Michael Pitts, Swapna Savvana, Senjuti Basu Roy and Vani Mandava	648–651
gst-Store: An Engine for Large RDF Graph Integrating Spatiotemporal Information Dong Wang, Lei Zou and Dongyan Zhao	652–655
Learn2Learn: A Visual Educational System for Study Planning Jishang Wei, Georgia Koutrika and Shanchan Wu	656–659
Helping Teenagers Relieve Psychological Pressures: A Micro-blog Based System Qi Li, Yuanyuan Xue, Jia Jia and Ling Feng	660–663
WePIGE: The WebLab Provenance Information Generator and Explorer Clement Caron, Bernd Amann, Camelia Constantin and Patrick Giroux	664–667
Learning Event Patterns for Gesture Detection Felix Beier, Nedal Alaqrara, Yuting Lai and Kai-Uwe Sattler	668–671
Tutorial: Managing Personal Data with Strong Privacy Guarantees Nicolas Anceaix, Benjamin Nguyen and Iulian Sandu Popa	672–673
Tutorial: An In-Depth Look at Modern Database Systems C. Mohan	674–674
Tutorial: Formal Methods for Event Processing Alexander Artikis and Georgios Paliouras	675–675
Talking to the Database in a Semantically Rich Way Henrietta Dombrovskaya and Richard Lee	676–687
SMILE: A Data Sharing Platform for Mobile Apps in the Cloud Jagan Sankaranarayanan, Hakan Hacigumus, Haopeng Zhang and Mohamed Sarwat	688–699
Diff-Index: Differentiated Index in Distributed Log-Structured Data Stores Wei Tan, Sandeep Tata, Yuzhe Tang and Liana Fong	700–711

Heterogeneous Stream Processing and Crowdsourcing for Urban Traffic Management Alexander Artikis, Matthias Weidlich, Francois Schnitzler, Ioannis Boutsis, Thomas Liebig, Nico Piatkowski, Christian Bockermann, Katharina Morik, Vana Kalogeraki, Jakub Marecek, Avigdor Gal, Shie Mannor, Dimitrios Gunopulos and Dermot Kinane .....	712–723
On Assigning Implicit Reputation Scores in an Online Labor Marketplace Maria Daltayanni, Luca de Alfaro, Panagiotis Papadimitriou and Panayiotis Tsaparas .....	724–725
Annotating the Behavior of Scientific Modules Using Data Examples: A Practical Approach Khalid Belhajjame .....	726–737
Benchmarking Bitemporal Database Systems: Ready for the Future or Stuck in the Past? Martin Kaufmann, Peter Fischer, Norman May and Donald Kossmann .....	738–749
Business-Intelligence Queries with Order Dependencies in DB2 Jaroslaw Szlichta, Parke Godfrey, Jarek Gryz, Wenbin Ma, Weinan Qiu and Calisto Zuzarte .....	750–761
A Tale of Two Graphs: Property Graphs as RDF in Oracle Souripriya Das, Jagannathan Srinivasan, Matthew Perry, Eugene Chong and Jayanta Banerjee .....	762–773