

VLDB2013

39th International Conference on Very Large Data Bases, Riva del Garda, Trento, Italy



Proceedings of the VLDB Endowment

Volume 6, No. 14 – September 2013

**Proceedings of the 39th International Conference on
Very Large Data Bases, Riva del Garda, Trento, Italy**

Editors-in-Chief:

Michael Böhlen, Christoph Koch

Associate Editors – Research Track:

**Ashraf Aboulnaga, Sihem Amer-Yahia, Chee Yong Chan, Yanlei Diao, Ada Waichee Fu, Johannes Gehrke,
Alon Halevy, Jayant Haritsa, Nikos Mamoulis, Thomas Neumann, Dan Olteanu, Divesh Srivastava, Jens Teubner**

Associate Editor – Experiments and Analysis Track:

Stefan Manegold

Proceedings Editors:

Peer Kröger, Stratis D. Viglas

PVLDB – Proceedings of the VLDB Endowment

Volume 6, No. 14, September 2013.

The 39th International Conference on Very Large Data Bases, Riva del Garda, Trento, Italy.

Copyright 2013 VLDB Endowment

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than VLDB Endowment must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists requires prior specific permission and/or a fee. Request permission to republish from PVLDB under email: info@vldb.org.

Volume 6, Number 14, September 2013: VLDB 2013

Pages ii - xii and 1642 - 2025

ISSN 2150-8097

Additional copies only online at: portal.acm.org, arxiv.org/corr, and www.vldb.org

TABLE OF CONTENTS

Front Matter

| | |
|---|-----|
| Copyright Notice | ii |
| Table of Contents | iii |
| VLDB 2013 Organization and Review Board | iv |

Letters

| | | |
|--|---------------------------------------|-----|
| Letter from the Editors-in-Chief | <i>Michael Böhlen, Christoph Koch</i> | xii |
|--|---------------------------------------|-----|

Research Papers

| | |
|--|--|
| Probabilistic Query Rewriting for Efficient and Effective Keyword Search on Graph Data..... | 1642 |
| | <i>Lei Zhang, Thanh Tran, Achim Rettinger</i> |
| QuEval: Beyond high-dimensional indexing a la carte | 1654 |
| | <i>Martin Schäler, Alexander Grebhahn, Reimar Schröter, Sandro Schulze, Veit Köppen, Gunter Saake</i> |
| Discovering Longest-lasting Correlation in Sequence Databases | 1666 |
| | <i>Yuhong Li, Leong Hou U, Man Lung Yiu, Zhiguo Gong</i> |
| PREDICT: Towards Predicting the Runtime of Large Scale Iterative Analytics..... | 1678 |
| | <i>Adrian Daniel Popescu, Andrey Balmin, Vuk Ercegovic, Anastasia Ailamaki</i> |
| On the Embeddability of Random Walk Distances..... | 1690 |
| | <i>Xiaohan Zhao, Adelbert Chang, Atish Das Sarma, Haitao Zheng, Ben Y. Zhao</i> |
| Instant Loading for Main Memory Databases | 1702 |
| | <i>Tobias Mühlbauer, Wolf Rödiger, Robert Seilbeck, Angelika Reiser, Alfons Kemper, Thomas Neumann</i> |
| Adaptive Range Filters for Cold Data: Avoiding Trips to Siberia | 1714 |
| | <i>Karolina Alexiou, Donald Kossmann, Per-Ake Larson</i> |
| Scalable Progressive Analytics on Big Data in the Cloud | 1726 |
| | <i>Badrish Chandramouli, Jonathan Goldstein, Abdul Quamar</i> |
| Scalable XML Query Processing using Parallel Pushdown Transducers..... | 1738 |
| | <i>Peter Ogden, David Thomas, Peter Pietzuch</i> |
| Understanding Insights into the Basic Structure and Essential Issues of Table Placement Methods in Clusters | 1750 |
| | <i>Yin Huai, Siyuan Ma, Rubao Lee, Owen O'Malley, Xiaodong Zhang</i> |
| A Probabilistic Optimization Framework for the Empty-Answer Problem | 1762 |
| | <i>Davide Mottin, Alice Marascu, Senjuti Basu Roy, Gautam Das, Themis Palpanas, Yannis Velegrakis</i> |

| | |
|--|------|
| Summarizing Answer Graphs Induced by Keyword Queries | 1774 |
| <i>Yinghui Wu, Shengqi Yang, Mudhakar Srivatsa, Arun Iyengar, Xifeng Yan</i> | |
| Supporting Keyword Search in Product Database: A Probabilistic Approach | 1786 |
| <i>Huizhong Duan, ChengXiang Zhai, Jinxing Cheng, Abhishek Gattani</i> | |
| A Sampling Algebra for Aggregate Estimation | 1798 |
| <i>Supriya Nirkhiwale, Alin Dobra, Christopher Jermaine</i> | |
| A Temporal-Probabilistic Database Model for Information Extraction..... | 1810 |
| <i>Maximilian Dylla, Iris Miliaraki, Martin Theobald</i> | |
| Counter Strike: Generic Top-Down Join Enumeration for Hypergraphs | 1822 |
| <i>Pit Fender, Guido Moerkotte</i> | |
| Efficient Bulk Updates on Multiversion B-trees..... | 1834 |
| <i>Daniar Achakeev, Bernhard Seeger</i> | |
| Query-Driven Approach to Entity Resolution | 1846 |
| <i>Hotham Altwaijry, Dmitri V. Kalashnikov, Sharad Mehrotra</i> | |
| Expressiveness and Complexity of Order Dependencies | 1858 |
| <i>Jaroslav Szlichta, Parke Godfrey, Jarek Gryz, Calisto Zuzarte</i> | |
| Counting and Sampling Triangles from a Graph Stream | 1870 |
| <i>A. Pavan, Kanat Tangwongsan, Srikanta Tirthapura, Kun-Lung Wu</i> | |
| An Experimental Analysis of Iterated Spatial Joins in Main Memory..... | 1882 |
| <i>Benjamin Sowell, Marcos Vaz Salles, Tuan Cao, Alan Demers, Johannes Gehrke</i> | |
| Scaling Queries over Big RDF Graphs with Semantic Hash Partitioning | 1894 |
| <i>Kisung Lee, Ling Liu</i> | |
| Distributed SocialLite: A Datalog-Based Language for Large-Scale Graph Analysis | 1906 |
| <i>Jiwon Seo, Jongsoo Park, Jaeho Shin, Monica S. Lam</i> | |
| Horton+: A Distributed System for Processing Declarative Reachability Queries over Partitioned Graphs | 1918 |
| <i>Mohamed Sarwat, Sameh Elnikety, Yuxiong He, Mohamed F. Mokbel</i> | |
| Streaming Similarity Search over one Billion Tweets using Parallel Locality-Sensitive Hashing..... | 1930 |
| <i>Narayanan Sundaram, Aizana Turmukhmetova, Nadathur Satish,</i> <i>Todd Mostak, Piotr Indyk, Samuel Madden, Pradeep Dubey</i> | |
| Anti-Caching: A New Approach to Database Management System Architecture..... | 1942 |
| <i>Justin DeBrabant, Andrew Pavlo, Stephen Tu, Michael Stonebraker, Stan Zdonik</i> | |
| Understanding Hierarchical Methods for Differentially Private Histograms..... | 1954 |
| <i>Wahbeh Qardaji, Weining Yang, Ninghui Li</i> | |

| | |
|--|------|
| Towards Social Data Platform: Automatic Topic-focused Monitor for Twitter Stream | 1966 |
| <i>Rui Li, Shengjie Wang, Kevin Chen-Chuan Chang</i> | |
| Simple, Fast, and Scalable Reachability Oracle | 1978 |
| <i>Ruoming Jin, Guan Wang</i> | |
| Aggregation and Ordering in Factorised Databases..... | 1990 |
| <i>Nurzhan Bakibayev, Tomas Kocisky, Dan Olteanu, Jakub Zavodny</i> | |
| Parallel Computation of Skyline and Reverse Skyline Queries Using MapReduce | 2002 |
| <i>Yoonjae Park, Jun-Ki Min, Kyuseok Shim</i> | |
| Fast Iterative Graph Computation with Block Updates | 2014 |
| <i>Wenlei Xie, Guozhang Wang, David Bindel, Alan Demers, Johannes Gehrke</i> | |

VLDB 2013 ORGANIZATION AND REVIEW BOARD

General Chairs

Themis Palpanas, University of Trento

Yannis Velegarakis, University of Trento

Program Chairs

Michael Böhlen, University of Zurich

Christoph Koch, EPFL

Advisory Board

Paolo Atzeni, Università Roma Tre

Stefano Ceri, Politecnico di Milano

John Mylopoulos, University of Trento

Award Committee

Surajit Chaudhuri, Microsoft (Chair)

Mike Carey, University of California, Irvine

Susan Davidson, University of Pennsylvania

Alon Halevy, Google

Sunita Sarawagi, IIT Bombay

Associate Editors

Ada Wai-Chee Fu, Chinese University of Hong Kong

Alon Halevy, Google

Ashraf Aboulnaga, University of Waterloo

Chee-Yong Chan, National University of Singapore

Dan Olteanu, Oxford University

Divesh Srivastava, AT&T Labs

Jayant Haritsa, Indian Institute of Science Bangalore

Jens Teubner, ETH Zurich

Johannes Gehrke, Cornell University

Nikos Mamoulis, University of Hong Kong

Sihem Amer-Yahia, Qatar Computing Research Institute

Stefan Manegold, CWI

Thomas Neumann, Technische Universität München

Yanlei Diao, University of Massachusetts Amherst

Experiments and Analysis Track Associate Editor

Stefan Manegold, CWI

Industrial and Applications Track Associate Editors

Min Wang, Google Research

Cong Yu, Google Research

Demonstration Chairs

Jun Yang, Duke University

Dimitrios Gunopulos, University of Athens

Letizia Tanca, Politecnico di Milano

Reproducibility Chairs

Philippe Bonnet, IT University of Copenhagen

Juliana Freire, New York University

Dennis Shasha, New York University

Research Track Review Board

Karl Aberer, EPFL, Switzerland

Foto Afrati, NTU Athens

Charu Aggarwal, IBM T. J. Watson Research Center

Yanif Ahmad, JHU

Jose-Luis Ambite, University of Southern California

Walid Aref, Purdue University

Magdalena Balazinska, University of Washington

Srikanta Bedathur, IIIT Delhi

Peter Boncz, CWI

Nico Bruno, Microsoft

Randal Burns, JHU

Andrea Cali, University of London, Birkbeck College

Carlos Castillo, Yahoo!

Gang Chen, Zhejiang University

Lei Chen, Hong Kong University of Science and Technology

Shimin Chen, HP Labs China

James Cheng, CUHK

Reynold Cheng, University of Hong Kong

Gao Cong, Nanyang Technological University

Brian Cooper, Google

Bin Cui, Peking University

Carlo Curino, MIT

Sudipto Das, Microsoft Research

Anish Das Sarma, Google Research

Atish Das Sarma, eBay Research Labs

Antonios Deligiannakis, Technical University of Crete

Amol Deshpande, University of Maryland

Xin Luna Dong, AT&T Labs-Research

Sameh Elnikety, Microsoft Research

Mohamed Eltabakh, Worcester Polytechnic Institute

Alan Fekete, University of Sydney

Hakan Ferhatosmanoglu, Bilkent University

Alvaro Fernandes, U. of Manchester

Juliana Freire, New York University

Benjamin C. M. Fung, Concordia University

Fabien Gandon, INRIA

Minos Garofalakis, Technical University of Crete, Greece

Buğra Gedik, Bilkent University

Rainer Gemulla, Max-Planck-Institut Saarbrücken
Gabriel Ghinita, University of Massachusetts Boston
Parke Godfrey, York University
Michaela Goetz, Cornell University
Lukasz Golab, University of Waterloo
Sergio Greco, University of Calabria
Le Gruenwald, University of Oklahoma
Krishna Gummadi, MPI
Haryadi Gunawi, University of California, Berkeley
Rahul Gupta, IIT Bombay
Marios Hadjieleftheriou, AT&T labs
Kuno Harumi, HP Labs
Michael Hay, Cornell
Bingsheng He, NTU Singapore
Sven Helmer, Free University of Bozen-Bolzano
Howard Ho, IBM Almaden Research
Katja Hose, Aalborg University
Bill Howe, University of Washington
Jeong-Hyon Hwang, State University of New York, Albany
Stratos Idreos, CWI
Hans-Arno Jacobsen, University of Toronto
Ricardo Jimenez-Peris, Technical University of Madrid
Ruoming Jin, Kent State University
Ryan Johnson, University of Toronto
Vanja Josifovski, Yahoo Inc.
Panos Kalnis, King Abdullah University of Science and Technology
Vana Kalogeraki, Athens Univ. of Econ. and Business
Carl-Christian Kanne, University of Mannheim
Hillol Kargupta, University of Maryland Baltimore County
Yiping Ke, Institute of High Performance Computing
Anne-Marie Kermarrec, INRIA
Daniel Kifer, PSU
Changkyu Kim, Intel
George Kollios, Boston University
Christian König, Microsoft Research
Laks V. S. Lakshmanan, University of British Columbia

Paul Larson, Microsoft
Mong-Li Lee, National University of Singapore
Wang-Chien Lee, Penn State University
Wolfgang Lehner, Technische Universität Dresden
Chengkai Li, The University of Texas at Arlington
Cuiping Li, Renmin University of China
Feifei Li, University of Utah
Guoliang Li, Tsinghua University
Lipyeow Lim, University of Hawaii at Manoa
Xuemin Lin, University of New South Wales
Eric Lo, The Hong Kong Polytechnic University
Boon Thau Loo, University of Pennsylvania
Qiong Luo, Hong Kong University of Science and Technology
Ashwin Machanavajjhala, Duke University
Sanjay Madria, University of Missouri-Rolla
Amélie Marian, Rutgers University
Frank McSherry, Microsoft
Sharad Mehrotra, University of California, Irvine
Poess Meikel, Oracle
Mohamed Mokbel, University of Minnesota
Bongki Moon, University of Arizona
Kyriakos Mouratidis, Singapore Management University
Gero Muhl, University of Rostock
Karin Murthy, IBM Research
Suman Nath, MSR
Wolfgang Nejdl, University of Hannover
Sylvia Nittel, University of Maine
Beng Chin Ooi, National University of Singapore
Tamer Ozsu, University of Waterloo
Esther Pacitti, University of Montpellier
Ippokratis Pandis, IBM Almaden
Olga Papaemmanouil, Brandeis University
Srinivasan Parthasarathy, The Ohio State University
Jignesh Patel, University of Wisconsin
Peter Pietzuc, Imperial College London
Neoklis Polyzotis, University of California, Santa Cruz
Lucian Popa, IBM Research

Bordawekar Rajesh, IBM T.J. Watson
Vibhor Rastogi, Yahoo
Christopher Re, University of Wisconsin, Madison
Matthias Renz, Ludwig-Maximilians University Munich, Germany
Marie-Christine Rousset, IMAG
Sourav S. Bhowmick, Nayang Technological University
Dimitris Sacharidis, IMIS Athena, Greece
Kenneth Salem, University of Waterloo
Maria Sapino, University of Torino
Monica Scannapieco, Istat
Bernhard Seeger, Philipps-Universität Marburg
Pierre Senellart, Télécom ParisTech
Cyrus Shahabi, USC
Lidan Shou, Zhejiang University
Adam Silberstein, Trifacta
Radu Sion, Stony Brook University
Yannis Sismanis, IBM, USA
Mohamed Soliman, University of Waterloo
Julia Stoyanovich, Drexel University and Skoltech
Yufei Tao, Chinese University of Hong Kong
Sandeep Tata, IBM Research
Nesime Tatbul, ETH Zurich

Demonstration Program Committee

Anastasia Ailamaki, EPFL
Sihem Amer-Yahia, Qatar Computing Research Institute
Leopoldo Bertossi, University of Carleton
Francois Bry, University of Munich
Chee-Yong Chan, National University of Singapore
Kevin Chang, UIUC
Chin-Wan Chung, Korea Advanced Institute of SaT
Gautam Das, University of Texas, Arlington
Aris Gkoulalas-Divanis, IBM Research Ireland
Torsten Grust, Universität Tübingen
Herodotos Herodotou, Microsoft Research
Yoshiharu Ishikawa, Nagoya University
Flip Korn, AT&T Labs

Evimaria Terzi, University of Boston
Martin Theobald, Max Planck Institute, Germany
Anthony Tung, National University of Singapore
Kostas Tzoumas, Technical University of Berlin
Sergei Vassilvitskii, Google
Stratis D. Viglas, University of Edinburgh
Ke Wang, Simon Fraser University
Ingmar Weber, Yahoo!
Raymond Chi-Wing Wong, Hong Kong University of Science and Technology
Xiaokui Xiao, NTU
Dong Xin, Google
Xifeng Yan, University of Santa Barbara
Jiong Yang, Case Western Reserve University
Ke Yi, Hong Kong University of Science and Technology
Man Lung Yiu, Hong Kong Polytechnic University
Cong Yu, Google Research
Ge Yu, Northeastern University, China
Jeffrey Yu, Chinese University of Hong Kong
Wenjie Zhang, UNSW Australia
Baihua Zheng, Singapore Management University
Aoying Zhou, East China Normal University
Xiaofang Zhou, University of Queensland

Nick Koudas, University of Toronto
Nikos Mamoulis, University of Hong Kong
Giansalvatore Mecca, Università della Basilicata
Alexandra Meliou, University of Washington
Rachel Pottinger, University of British Columbia
Rajeev Rastogi, Yahoo! India
Bernhard Seeger, University of Marburg
Ambuj Singh, University of California, Santa Barbara
Jens Teubner, ETH Zurich
Wei Wang, University of New South Wales
Li Xiong, Emory University
Jia Yuan Yu, IBM Research
Demetris Zeinalipour, University of Cyprus
Shuigeng Zhou, Fudan University

Industrial Track Committee

Michael Brodie, Verizon
Alejandro Buchmann, Technische Universität
Darmstadt
Shimin Chen, HP Labs China
Umeshwar Dayal, HP Labs
Shel Finkelstein, SAP
Dieter Gawlick, Oracle
Tasos Kementsietsidis, T.J. Watson Research Center
Tim Kraska, Brown University
Yue Lu, twitter
Arnab Nandi, The Ohio State University

Felix Naumann, University of Potsdam
Fatma Ozcan, IBM Research
Radu Popescu-Zeletin, Fraunhofer-Institut für Offene
Kommunikationssysteme
Raghu Ramakrishnan, Microsoft
Jun Rao, LinkedIn
Len Seligman, MITRE
Eric Simon, SAP
Haixun Wang, Microsoft Research
Fei Wu, Google Research
Jackie Xiang, Foursquare

Reproducibility Committee

Matias Bjørling, IT University of Copenhagen
Wei Cao, Remnin University
Stratos Idreos, Centrum Wiskunde & Informatica
Ryan Johnson, University of Toronto
Martin Kaufmann, ETH Zurich
David Koop, University of Utah
Lucja Kot, Cornell University
Willis Lang, University of Wisconsin

Mian Lu, Hong Kong University of Science and
Technology
Dan Olteanu, University of Oxford
Paolo Papotti, Qatar Computing Research Institute
(QCRI)
Ben Sowell, Cornell University
Radu Stoica, EPFL - Ecole Polytechnique Federale de
Lausanne
Dimitris Tsirogiannis, Microsoft Jim Gray Systems Lab

PhD Workshop Chairs

Angela Bonifati, Icar-CNR
Sanjay Chawla, University of Sydney
Chris Jermaine, Rice University

Tutorial Chairs

Serge Abiteboul, INRIA
Gianni Mecca, Universita della Basilicata
Haixun Wang, Microsoft Research Asia

Panel Chairs

Shivnath Babu, Duke University
Stavros Harizopoulos, Nou Data
Ihab Ilyas, Qatar Computing Research Institute

Sponsorship Chairs

Sam Madden, Massachusetts Institute of Technology
Vassilis Vassalos, Athens Univ. of Econ. and Business
Paolo Merlaldo, Universita Roma Tre

Publicity Chair

Tasos Kementsietsidis, IBM T.J. Watson Research Center

Proceedings Chairs

Peer Kröger, Ludwig-Maximilians University, Munich
Stratis D. Viglas, University of Edinburgh

Web Management Chair

Francesco Guerra, University of Modena and Reggio Emilia

Treasury Chair

Marios Hadjieleftheriou, AT&T Labs Research

Local Administration

Ufficio Convegni and dbTrento Group, University of Trento

Logo Design

Sakis Palpanas

PVLDB Information Director

Gerald Weber, University of Auckland

PVLDB Advisory Committee

Philip Bernstein, Michael Böhlen, Peter Buneman, Susan Davidson, Z. Meral Ozsoyoglu, S. Sudarshan, Gerhard Weikum

LETTER FROM THE PROCEEDINGS EDITORS

This is the last issue of Vol. 6 of the Proceedings of the VLDB Endowment (PVLDB). The last two issues of Vol. 6 include a total of 47 papers that have been accepted by the VLDB 2013 program committee and will be presented at the VLDB 2014 conference in Hangzhou. Thus, a good number of VLDB 2013 papers will be presented in Hangzhou and we look forward to meet you there.

The 127 research papers published in PVLDB Vol. 6 have been selected from a total of 559 submissions. The decision process consisted of a packed schedule of monthly submissions, reviewing, discussions and revision cycles. The competition was fierce with an acceptance rate of 22% for research, 32% for industrial, and 46% for demonstration papers. There were 97 demonstration and 54 industrial paper submissions. 4% of the research papers were accepted directly without revision, and 23% of the papers were selected for a revision. 29 papers were submitted to the experiments and analysis track, and of these 12 papers got a revision decision.

This is the third year that the papers have been accepted only through journal-style reviewing with revisions and published in PVLDB. Much of this process has stabilized during the last few years and is increasingly being considered as a role model by other conferences and communities. It was a great privilege to further consolidate and advance this model in the context of VLDB 2013.

We would like to use this opportunity to thank the many people who contributed and made VLDB 2013 possible. We are grateful to the general chairs for their unwavering support during the past couple of years. They organized an exciting VLDB 2013 conference in a wonderful location. Our warm thanks go to Min Wang and Cong Yu for handling the industrial program; Jun Yang, Dimitrios Gunopulos, and Letizia Tanca for managing the demonstration program; Serge Abiteboul, Gianni Mecca, and Haixun Wang for choosing the tutorials; and Shivnath Babu, Stavros Harizopoulos, and Ihab Ilyas for selecting the panels. We thank all associate editors, PC members, and reviewers for their dedicated and hard work that made VLDB 2013 a big success. We thank the proceedings co-chairs Stratis Viglas and Peer Kröger, for putting the PVLDB issues together, and year around management of the collection and publication of camera-ready copies of accepted papers. And, finally, many thanks to all the authors for submitting their papers to PVLDB.

Michael Böhlen, University of Zurich
Christoph Koch, EPFL
Program Chairs, VLDB 2013