

VLDB2013

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



Proceedings of the VLDB Endowment

Volume 6, No. 10 – August 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

Guest Editors:

Themis Palpanas, Yannis Velegrakis

Proceedings Editors:

Peer Kröger, Stratis D. Viglas

PVLDB – Proceedings of the VLDB Endowment

Volume 6, No. 10, August 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 10, August 2013: VLDB 2013

Pages ii - xii and 781 - 960

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 Guest Editors	xi
-------------------------------------	----

Research Papers

Query Optimization over Crowdsourced Data	781
..... <i>Hyunjung Park, Jennifer Widom</i>	
A Data-adaptive and Dynamic Segmentation Index for Whole Matching on Time Series	793
..... <i>Yang Wang, Peng Wang, Jian Pei, Wei Wang, Sheng Huang</i>	
Extraction and Integration of Partially Overlapping Web Sources	805
..... <i>Mirko Bronzi, Valter Crescenzi, Paolo Merialdo, Paolo Papotti</i>	
The Yin and Yang of Processing Data Warehousing Queries on GPU Devices	817
..... <i>Yuan Yuan, Rubao Lee, Xiaodong Zhang</i>	
Mining and Indexing Graphs for Supergraph Search	829
..... <i>Dayu Yuan, Prasenjit Mitra, C. Lee Giles</i>	
Efficient Recovery of Missing Events	841
..... <i>Jianmin Wang, Shaoxu Song, Xiaochen Zhu, Xuemin Lin</i>	
Hadoop’s Adolescence	853
..... <i>Kai Ren, YongChul Kwon, Magdalena Balazinska, Bill Howe</i>	
RACE: A Scalable and Elastic Parallel System for Discovering Repeats in Very Long Sequences...	865
..... <i>Essam Mansour, Ahmed El-Roby, Panos Kalnis, Aron Ahmadi, Ashraf Aboulnaga</i>	
LLAMA: A Cache/Storage Subsystem for Modern Hardware	877
..... <i>Justin Levandoski, David Lomet, Sudipta Sengupta</i>	
Revisiting Co-Processing for Hash Joins on the Coupled CPU-GPU Architecture	889
..... <i>Jiong He, Mian Lu, Bingsheng He</i>	
Top-K Nearest Keyword Search on Large Graphs	901
..... <i>Miao Qiao, Lu Qin, Hong Cheng, Jeffrey Xu Yu, Wentao Tian</i>	
A General Framework for Geo-Social Query Processing	913
..... <i>Nikos Armenatzoglou, Stavros Papadopoulos, Dimitris Papadias</i>	

Towards Predicting Query Execution Time for Concurrent and Dynamic Database Workloads	925
..... <i>Wentao Wu, Yun Chi, Hakan Hacigumus, Jeffrey F. Naughton</i>	
Sketch-based Geometric Monitoring of Distributed Stream Queries	937
..... <i>Minos Garofalakis, Daniel Keren, Vasilis Samoladas</i>	
Direction-Preserving Trajectory Simplification.....	949
..... <i>Cheng Long, Raymond Chi-Wing Wong, Chenjuan Guo, H. V. Jagadish</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, Università 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, Università 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 Convegna 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 GUEST EDITORS

In its 39th edition, the international conference on Very Large Data Bases (VLDB) meets the Dolomites. VLDB 2013 is hosted by the Autonomous Province of Trento (Italy), home of the Dolomite Mountains, and of the very active research pole of Trento, which includes the University of Trento, the Fondazione Bruno Kessler, one of the six European nodes of the European Institute of Innovation and Technology for Information and Communication Technologies (EIT ICT), the Trento RISE institution, and research labs established by Microsoft, Telecom Italia, and IBM. This is also the home of dbTrento (the Data and Information Management group of the University of Trento), which is organizing the current edition of VLDB.

This feat would not have been possible without the hard work of the organizing and program committees. We are grateful to all members of these committees, which were truly international, representing all regions of the globe: the numerous members of the Review Board and the Demonstration, Industrial, Reproducibility and Best Paper Committees, the Research Track Associate Editors Ashraf Abounaga, 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 and Jens Teubner, the Experiments and Analysis Track Associate Editor Stefan Manegold, the Industrial and Applications Track Associate Editors Min Wang and Cong Yu, the Tutorial Chairs Serge Abiteboul, Gianni Mecca and Haixun Wang, the Panel Chairs, Shivnath Babu, Stavros Harizopoulos and Ihab Ilyas, the Demonstrations Chairs Jun Yang, Dimitrios Gunopoulos and Letizia Tanca, the PhD Workshop Chairs Angela Bonifati, Sanjay Chawla and Chris Jermaine, the Workshop Chairs Tiziana Catarci, AnHai Doan and Tova Milo, and the Reproducibility Chairs Philippe Bonnet, Juliana Freire and Dennis Shasha. We would specifically like to thank Michael Böhlen and Christoph Koch, the Program Chairs, for their excellent cooperation, and Paolo Atzeni, Stefano Ceri and John Mylopoulos, our Advisory Board. We would like to recognize the precious help of our sponsors and congratulate the Sponsorship Chairs Sam Madden, Vasilis Vassalos and Paolo Merialdo for their truly successful mission, the Web Management Chair Francesco Guerra for his continuous availability and responsiveness, the Publicity Chair Tasos Kementsietsidis, and the Treasury Chair Marios Hadjieleftheriou. Special thanks go to the Proceedings Chair Stratis Viglas for the excellent work he did despite the excess load, and the Registrations Chair Periklis Andritsos for the efficient and effective handling of the complex registration task. We acknowledge the help and support of the University of Trento, of the Conference Office and in particular Mara Tomasi, the OKKAM team and Daniel Giacomuzzi in particular, and of the Conference Center staff of Riva del Garda. We are grateful to the dbTrento students and all our volunteers, for their time and efforts: Siarhei Bykau, Alessandro Camera, Cristiana Chitic, Christina Christodoulakis, Alberto Cordioli, Michele Dallachiesa, Zolzaya Dashdorj, Cong Dinh, Vincenzo Falletta, Matteo Lissandrini, Katya Mirylenka, Daniil Mirylenka, Paramita Mirza, Davide Mottin, Dimitra Papadimitriou, Pavlos Paraskevopoulos, Mikalai Tsytsarau and Kostas Zoumpatianos. Last but not least, we would like to thank Costanza Vettori, our administrative assistant that so successfully coordinated everything and made sure that no detail slipped our attention.

In this edition of the VLDB conference, we have an active involvement of the local industry, in an effort to further disseminate and promote the thoughts, ideas and work of the VLDB community, and to cross-fertilize them with those of the local entities. We believe that such efforts can play a significant role in the exchange of results and practical experiences, as well as in the growth of VLDB in the long run.

For this trade of ideas to take place, we could not think of a better setting than the picturesque town of Riva, on the north shore of Lake Garda. In the 17th century, Lake Garda became a popular destination for young central-European nobility, and all towns around the lake are rich in history waiting to be discovered. And those of you, who think better in (literally) thin air, should not miss the opportunity to visit the Dolomite Mountains (part of the Eastern Alps), which are a UNESCO World Heritage Site. These magnificent and impressive mountains offer numerous opportunities for hikes, mountaineering, rock climbing, as well as for brainstorming.

Welcome to VLDB 2013 in Riva del Garda; we hope that you find the technical program, as well as the hosting city and region equally exciting and enticing!

Themis Palpanas, University of Trento
Yannis Velegarakis, University of Trento
[General Chairs, VLDB 2013](#)