

VLDB2013

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



Proceedings of the VLDB Endowment

Volume 6, No. 11 – 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:

Min Wang, Cong Yu

Proceedings Editors:

Peer Kröger, Stratis D. Viglas

PVLDB – Proceedings of the VLDB Endowment

Volume 6, No. 11, 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 11, August 2013: VLDB 2013

Pages ii - xii and 961 - 1197

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	<i>Min Wang, Cong Yu</i> xii
-------------------------------------	------------------------------

Industrial and Applications Track Papers

Continuous Cloud-Scale Query Optimization and Processing	961
..... <i>Nicolas Bruno, Sapna Jain, Jingren Zhou</i>	
Optimization Strategies for A/B Testing on HADOOP.....	973
..... <i>Andrii Cherniak, Huma Zaidi, Vladimir Zadorozhny</i>	
Piranha: Optimizing Short Jobs in Hadoop	985
..... <i>Khaled Elmeleegy</i>	
Making Updates Disk-I/O Friendly Using SSDs.....	997
..... <i>Yuan Yuan, Rubao Lee, Xiaodong Zhang</i>	
Hadoop-GIS: A High Performance Spatial Data Warehousing System over MapReduce	1009
..... <i>Ablimit Aji, Fusheng Wang, Hoang Vo, Rubao Lee, Qiaoling Liu, Xiaodong Zhang, Joel Saltz</i>	
Statistics Collection in Oracle Spatial and Graph: Fast Histogram Construction for Complex Geometry Objects	1021
..... <i>Bhuvan Bamba, Siva Ravada, Ying Hu, Richard Anderson</i>	
MillWheel: Fault-Tolerant Stream Processing at Internet Scale	1033
..... <i>Tyler Akidau, Alex Balikov, Kaya Bekiroglu, Slava Chernyak,</i> <i>Josh Haberman, Reuven Lax, Sam McVeety, Daniel Mills, Paul Nordstrom, Sam Whittle</i>	
Online, Asynchronous Schema Change in F1.....	1045
..... <i>Ian Rae, Eric Rollins, Jeff Shute, Sukhdeep Sodhi, Radek Vingralek</i>	
Scuba: Diving into Data at Facebook	1057
..... <i>Lior Abraham, John Allen, Oleksandr Barykin, Vinayak Borkar,</i> <i>Bhuwan Chopra, Ciprian Gereca, Daniel Merl, Josh Metzler,</i> <i>David Reiss, Subbu Subramanian, Janet L. Wiener, Okay Zed</i>	
F1: A Distributed SQL Database That Scales	1068
..... <i>Jeff Shute, Radek Vingralek, Bart Samwel, Ben Handy, Chad Whipkey,</i> <i>Eric Rollins, Mircea Oancea, Kyle Littlefield, David Menestrina, Stephan Ellner,</i> <i>John Cieslewicz, Ian Rae, Traian Stancescu, Himani Apte</i>	

DB2 with BLU Acceleration: So Much More than Just a Column Store.....	1080
..... Vijayshankar Raman, Gopi Attaluri, Ronald Barber, Naresh Chainani, David Kalmuk, Vincent KulandaiSamy, Jens Leenstra, Sam Lightstone, Shaorong Liu, Guy M. Lohman, Tim Malkemus, Rene Mueller, Ippokratis Pandis, Berni Schiefer, David Sharpe, Richard Sidle, Adam Storm, Liping Zhang	
The Quantcast File System	1092
..... Michael Ovsianikov, Silvius Rus, Damian Reeves, Paul Sutter, Sriram Rao, Jim Kelly	
Adaptive and Big Data Scale Parallel Execution in Oracle	1102
..... Srikanth Bellamkonda, Hua-Gang Li, Unmesh Jagtap, Yali Zhu, Vince Liang, Thierry Cruanes	
WOO: A Scalable and Multi-tenant Platform for Continuous Knowledge Base Synthesis.....	1114
..... Kedar Bellare, Carlo Curino, Ashwin Machanavajihala, Peter Mika, Mandar Rahurkar, Aamod Sane	
Entity Extraction, Linking, Classification, and Tagging for Social Media: A Wikipedia-Based Approach	1126
..... Rohid Kumar, Digvijay S. Lamba, Nikesh Garera, Mitul Tiwari, Xiaoyong Chai, Sanjib Das, Sri Subramaniam, Anand Rajaraman, Venky Harinarayan, AnHai Doan	
Overview of Turn Data Management Platform for Digital Advertising	1138
..... Hazem Elmeleegy, Yinan Li, Yan Qi, Peter Wilmot, Mingxi Wu, Santanu Kolay, Ali Dasdan, Songting Chen	
Unicorn: A System for Searching the Social Graph.....	1150
..... Michael Curtiss, Iain Becker, Tudor Bosman, Sergey Doroshenko, Lucian Grijincu, Tom Jackson, Sandhya Kunnatur, Soren Lassen, Philip Pronin, Sriram Sankar, Guanghao Shen, Gintaras Woss, Chao Yang, Ning Zhang	

Invited Industrial Papers

A New Service for Customer Care Based on the TrentoRise BigData Platform	1162
..... Sergio Ramazzina, Chiara L. Ballari, Daniela Somenzi	
Exploiting the Diversity, Mass and Speed of Territorial Data by TELCO Operator for Better User Services	1164
..... Fabrizio Antonelli, Antonino Casella, Cristiana Chitic, Roberto Larcher, Giovanni Torrisi	
The Trento Big Data Platform for Public Administration and Large Companies: Use cases and Opportunities.....	1166
..... Ivan Bedini, Benedikt Elser, Yannis Velegrakis	
Designing Query Optimizers for Big Data Problems of The Future.....	1168
..... Nga Tran, Sreenath Bodagala, Jaimin Dave	
How to maximize the value of big data with the open source SpagoBI suite through a comprehensive approach	1170
..... Monica Franceschini	

Context-Aware Computing: Opportunities and Open Issues	1172
..... <i>Edward Y. Chang</i>	
Next Generation Data Analytics at IBM Research.....	1174
..... <i>Oktie Hassanzadeh, Anastasios Kementsietsidis, Benny Kimelfeld,</i>	
..... <i>Rajasekar Krishnamurthy, Fatma Özcan, Ippokratis Pandis</i>	
Learning and Intelligent Optimization (LION): One Ring to Rule Them All.	1176
..... <i>Mauro Brunato, Roberto Battiti</i>	
Microsoft SQL Server’s Integrated Database Approach for Modern Applications and Hardware....	1178
..... <i>David Lomet</i>	
Odyssey: A Multi-Store System for Evolutionary Analytics	1180
..... <i>Hakan Hacigumus, Jagan Sankaranarayanan, Junichi Tatemura,</i>	
..... <i>Jeff LeFevre, Neoklis Polyzotis</i>	
A global Entity Name System (ENS) for data ecosystems	1182
..... <i>Paolo Bouquet, Andrea Molinari</i>	
SAP HANA: The Evolution from a Modern Main-Memory Data Platform to an Enterprise Application Platform.....	1184
..... <i>Vishal Sikka, Franz Farber, Anil Goel, Wolfgang Lehner</i>	
Keeping the TPC Relevant!	1186
..... <i>Raghunath Nambiar, Meikel Poess</i>	

Tutorial Papers

Big Data Integration	1188
..... <i>Xin Luna Dong, Divesh Srivastava</i>	
Just-in-time compilation for SQL query processing	1190
..... <i>Stratis D. Viglas</i>	
Toward Scalable Transaction Processing.....	1192
..... <i>Anastasia Ailamaki, Ryan Johnson, Ippokratis Pandis, Pinar Tozun</i>	
Towards Database Virtualization for Database as a Service	1194
..... <i>Aaron J. Elmore, Carlo Curino, Divyakant Agrawal, Amr El Abbadi</i>	
Mobility and Social Networking: A Data Management Perspective.....	1196
..... <i>Mohamed F. Mokbel, Mohamed Sarwat</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 Meriardo, 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 GUEST EDITORS

This is the 11th issue of PVLDB Volume 6, which is produced by Peer Kröger and Stratis Viglas who serve as Proceedings Chairs and Gerald Weber who serves as Information Director for PVLDB. The papers included in this issue will be presented at the industrial track and tutorial sessions at the VLDB conference in Riva del Garda in August 2013.

Innovative works done by industry practitioners on both database systems and database applications have always been an integral part of the database research community. The industrial track is designed to provide a forum to foster discussions on the experience, lessons, and insights learned from those works, which are especially important in the era of big data. The industrial track has been a part of the VLDB conference for over 15 years, and this year, we have contributions from 10 different industrial institutions and a number of academic institutions that have collaborated closely with industrial practitioners. Initiated by the General Chairs, Themis Palpanas and Yannis Velegrakis, we are also experimenting with a new technical session that includes papers contributed by companies participating in VLDB 2013 on what they vision as long term important challenges to address in the field.

Unlike the research track, the industrial track follows the traditional submission process. The review of all papers in this issue is done by the VLDB 2013 industrial track program committee consisting of 19 experienced members from both industry and academia. A warm thanks to all PC members for their tireless efforts and the time they invested into shaping the industrial program of VLDB 2013. They produced many detailed and constructive reviews, followed by active and insightful discussions among reviewers. Our warmest thanks go to the authors who submitted their innovative work on commercial data management systems and their applications, and experience in applying recent research advances to real-world problems.

The industrial program of VLDB 2013 in Riva del Garda is shaping up nicely and we look forward to see you all in Riva del Garda in August. Thanks to everybody, and especially authors and program committee members, for all their support and efforts.

Min Wang, Google Research—Mountain View
Cong Yu, Google Research—New York
Industrial Program Chairs, VLDB 2013