

# VLDB2016

42<sup>nd</sup> International Conference on Very Large Data Bases, New Delhi, India



## Proceedings of the VLDB Endowment

Volume 9, No. 12 – August 2016

**Proceedings of the 42nd International Conference on  
Very Large Data Bases, New Delhi, India**

Program Chairs and Editors in Chief:

**Surajit Chaudhuri and Jayant Haritsa**

Associate Editors – Research Track:

**Daniel Abadi, Peter Boncz, Anhai Doan, Yanlei Diao, Amr El Abbadi, Alan Fekete, Minos Garofalakis, Alon Halevy,  
Ihab Ilyas, Donald Kossman, Sharad Mehrotra, Sunita Sarawagi, Kyuseok Shim, Kian-Lee Tan**

Proceedings Chairs:

**Srikanta Bedathur, Katja Hose**

PVLDB – Proceedings of the VLDB Endowment

Volume 9, No. 12, August 2016.

The 42nd International Conference on Very Large Data Bases, New Delhi, India.

## **Copyright 2015 VLDB Endowment**

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>. Obtain permission prior to any use beyond those covered by the license. Contact copyright holder by emailing [info@vldb.org](mailto:info@vldb.org).

Volume 9, Number 12, August 2016: VLDB 2016

Pages ii – vi and 900 - 1244

ISSN 2150-8097

Additional copies only online at: [portal.acm.org](http://portal.acm.org), [arxiv.org/corr](http://arxiv.org/corr), and [www.vldb.org](http://www.vldb.org)

## TABLE OF CONTENTS

### Front Matter

Copyright Notice .....	ii
Table of Contents .....	iii
VLDB 2016 Organization and Review Board .....	iv

### Research Papers

G-SQL: Fast Query Processing via Graph Exploration .....	900
<i>Hongbin Ma, Bin Shao, Yanghua Xiao, Liang Jeff Chen, Haixun Wang</i>	
Measuring and Optimizing Distributed Array Programs.....	912
<i>Mingxing Zhang, Yongwei Wu, Kang Chen, Teng Ma, Weimin Zheng</i>	
YourSQL: A High-Performance Database System Leveraging In-Storage Computing .....	924
<i>Insoon Jo, Duck-Ho Bae, Andre S. Yoon, Jeong-Uk Kang, Sangyeun Cho, Daniel DG Lee, Jaeheon Jeong</i>	
Lifetime-Based Memory Management for Distributed Data Processing Systems.....	936
<i>Lu Lu, Xuanhua Shi, Yongluan Zhou, Xiong Zhang, Hai Jin, Cheng Pei, Ligang He, Yuanzhen Geng</i>	
ActiveClean: Interactive Data Cleaning for Statistical Modeling .....	948
<i>Sanjay Krishnan, Jiannan Wang, Eugene Wu, Michael J. Franklin, Ken Goldberg</i>	
Compressed Linear Algebra for Large-Scale Machine Learning .....	960
<i>Ahmed Elgohary, Matthias Boehm, Peter J. Haas, Frederick R. Reiss, Berthold Reinwald</i>	
Fast Queries Over Heterogeneous Data Through Engine Customization .....	972
<i>Manos Karpathiotakis, Ioannis Alagiannis, Anastasia Ailamaki</i>	
Data-driven Visual Graph Query Interface Construction and Maintenance: Challenges and Opportunities .....	984
<i>Sourav S. Bhowmick, Byron Choi, Curtis Dyreson</i>	
Detecting Data Errors: Where are we and what needs to be done? .....	993
<i>Ziawasch Abedjan, Xu Chu, Dong Deng, Raul Castro Fernandez, Ihab F. Ilyas, Mourad Ouzzani, Paolo Papotti, Michael Stonebraker, Nan Tang</i>	
Exploiting Soft and Hard Correlations in Big Data Query Optimization .....	1005
<i>Hai Liu, Dongqing Xiao, Pankaj Didwania, Mohamed Y. Eltabakh</i>	
Interactive Browsing and Navigation in Relational Databases .....	1017
<i>Minsuk Kahng, Shamkant B. Navathe, John T. Stasko, Duen Horng Chau</i>	
Cheap Data Analytics using Cold Storage Devices.....	1029
<i>Renata Borovica-Gajić, Raja Appuswamy, Anastasia Ailamaki</i>	
Parallel Local Graph Clustering.....	1041
<i>Julian Shun, Farbod Roosta-Khorasani, Kimon Fountoulakis, Michael W. Mahoney</i>	

Online Minimum Matching in Real-Time Spatial Data: Experiments and Analysis .....	1053
..... <i>Yongxin Tong, Jieying She, Bolin Ding, Lei Chen, Tianyu Wo, Ke Xu</i>	
Index-Assisted Hierarchical Computations in Main-Memory RDBMS.....	1065
..... <i>Robert Brunel, Norman May, Alfons Kemper</i>	
Dynamic Influence Analysis in Evolving Networks.....	1077
..... <i>Naoto Ohsaka, Takuya Akiba, Yuichi Yoshida, Ken-ichi Kawarabayashi</i>	
Distance-based Outlier Detection in Data Streams.....	1089
..... <i>Luan Tran, Liyue Fan, Cyrus Shahabi</i>	
On Measuring the Lattice of Commonalities Among Several Linked Datasets .....	1101
..... <i>Michalis Mountantonakis, Yannis Tzitzikas</i>	
Oblivious RAM: A Dissection and Experimental Evaluation.....	1113
..... <i>Zhao Chang, Dong Xie, Feifei Li</i>	
Optimization of Conjunctive Predicates for Main Memory Column Stores .....	1125
..... <i>Fisnik Kastrati, Guido Moerkotte</i>	
Explaining Outputs in Modern Data Analytics .....	1137
..... <i>Zaheer Chothia, John Liagouris, Frank McSherry, Timothy Roscoe</i>	
RDF Graph Alignment with Bisimulation.....	1149
..... <i>Peter Buneman, Sławek Staworko</i>	
Teaching an RDBMS about ontological constraints.....	1161
..... <i>Damian Bursztyń, François Goasdoué, Ioana Manolescu</i>	
BLAST: a Loosely Schema-aware Meta-blocking Approach for Entity Resolution.....	1173
..... <i>Giovanni Simonini, Sonia Bergamaschi, H. V. Jagadish</i>	
LSH Ensemble: Internet-Scale Domain Search .....	1185
..... <i>Erkang Zhu, Fatemeh Nargesian, Ken Q. Pu, Renée J. Miller</i>	
Magellan: Toward Building Entity Matching Management Systems.....	1197
<i>Pradap Konda, Sanjib Das, Paul Suganthan G. C., AnHai Doan, Andel Ardalan, Jeffrey R. Ballard,</i>	
<i>Han Li, Fatemah Panahi, Haojun Zhang, Jeff Naughton, Shishir Prasad, Ganesh Krishnan,</i>	
<i>Rohit Deep, Vijay Raghavendra</i> .....	
ATHENA: An Ontology-Driven System for Natural Language Querying over Relational Data Stores .....	1209
..... <i>Diptikalyan Saha, Avriila Floratou, Karthik Sankaranarayanan, Umar Farooq Mihans,</i>	
<i>Ashish R. Mittal, Fatma Özcan</i> .....	
Incremental Computation of Common Windowed Holistic Aggregates.....	1221
..... <i>Richard Wesley, Fei Xu</i>	
Effective Community Search for Large Attributed Graphs .....	1233
..... <i>Yixiang Fang, Reynold Cheng, Siqiang Luo, Jiafeng Hu</i>	

## VLDB 2016 ORGANIZATION AND REVIEW BOARD

### GENERAL CHAIRS

Anand Deshpande, Persistent Systems  
T. M. Vijayaraman, Persistent Systems

### Program Chairs and Editors in Chief of PVLDB 9

Surajit Chaudhuri, Microsoft Research  
Jayant Haritsa, I.I.Sc. Bangalore

### Research Track Associate Editors

Daniel Abadi, Yale University  
Peter Boncz, CWI Amsterdam  
Anhai Doan, University of Wisconsin - Madison  
Yanlei Diao, University of Massachusetts - Amherst  
Amr El Abbadi, University of California - Santa Barbara  
Alan Fekete, University of Sydney  
Minos Garofalakis, Tech. University of Crete  
Alon Halevy, Google  
Ihab Ilyas, University of Waterloo  
Donald Kossman, Microsoft Research  
Sharad Mehrotra, University of California - Irvine  
Sunita Sarawagi, I.I.T. Bombay  
Kyuseok Shim, Seoul National University  
Kian-Lee Tan, National University of Singapore

### Industrial Track Associate Editors

Gustavo Alonso, ETH Zurich  
Anant Jhingran, Apigee  
Seshadri Srinivasan, Zettata

### Demonstration Chairs

Mike Cafarella, University of Michigan - Ann Arbor  
Stratos Idreos, Harvard University  
Anthony Tung, National University of Singapore

### Tutorial Chairs

Chris Re, Stanford University  
S. Sudarshan, I.I.T. Bombay  
Gerhard Weikum, Max-Planck Institute for Informatics

### Panel Chairs

Juliana Freire, New York University  
Sam Madden, Massachusetts Institute of Technology

### Workshop Chairs

Sihem Amer-Yahia, CNRS  
Joe Hellerstein, University of California - Berkeley

### PhD Workshop Chairs

Torsten Grust, University of Tübingen  
Kamal Karlapalem, I.I.I.T. Hyderabad  
Andy Pavlo, Carnegie Mellon University

### Proceedings Chairs

Srikanta Bedathur, IBM Research  
Katja Hose, Aalborg University

### Organization Committee Chair

Mukesh Mohania, IBM Research

### Sponsorship Committee Chairs

Michael Carey, University of California – Irvine  
Anand Deshpande, Persistent Systems

### Publicity and Volunteers Committee Chair

Rajkumar Sen, MemSQL

### VLDB Endowment Representative

S. Sudarshan, I.I.T. Bombay

### PVLDB Managing Editor

Divesh Srivastava, AT&T Labs

### PVLDB Information Director

Gerald Weber, University of Auckland

### PVLDB Advisory Committee

Michael Böhlen, Susan Davidson, H. V. Jagadish,  
Renée J. Miller, M. Tamer Özsu,  
S. Sudarshan, Kian-Lee Tan, Gerhard Weikum

## Research Track Review Board

Ashraf Aboulnaga, Qatar Computing Research Institute  
Marcos K. Aguilera, VMware Research  
Yael Amsterdamer, Tel-aviv University  
Walid Aref, Purdue University  
Marcelo Arenas, Pontifical Catholic University  
Shivnath Babu, Duke University  
Magdalena Balazinska, University of Washington  
Zohra Bellahsene, University of Montpellier  
Arnab Bhattacharya, IIT Kanpur  
Francesco Bonchi, Yahoo Labs Barcelona  
Nico Bruno, Google  
Michael Cafarella, University of Michigan  
Andrea Cali, Birkbeck University of London  
Reynold C. K. Cheng, University of Hong Kong  
Venkatesan Chakaravarthy, IBM Research  
Kaushik Chakrabarti, Microsoft Research  
Arbee L P Chen, National Chengchi University  
Lei Chen, Hong Kong University of Science and Technology  
Shimin Chen, Institute of Computing Technology  
Rada Chirkova, North Carolina State University  
Tyson Condie, University of California, Los Angeles  
Isabel Cruz, University of Illinois at Chicago  
Bin Cui, Peking University  
Gautam Das, University of Texas, Arlington  
Mahashweta Das, HP Labs  
Sudipto Das, Microsoft Research  
Antonios Deligiannakis, Technical University of Crete  
Prasad Deshpande, IBM Research  
Daniel Deutch, Tel Aviv University  
Jens Dittrich, Saarland University  
Gillian Dobbie, University of Auckland  
Xin Luna Dong, Google  
Aaron Elmore, University of Chicago  
Wenfei Fan, University of Edinburgh  
Hakan Ferhatosmanoglu, Bilkent University  
Floris Geerts, University of Antwerp  
Johannes Gehrke, Microsoft  
Aristides Gionis, Aalto University  
Sergio Greco, University of Calabria  
Haryadi Gunawi, University of Chicago  
Wook-Shin Han, Pohang University of Science and Tech.  
Stavros Harizopoulos, Amazon  
Herodotos Herodotou, Cyprus University of Technology  
Melanie Herschel, University of Stuttgart  
Zachary Ives, University of Pennsylvania  
Christopher Jermaine, Rice University  
Ryan Johnson, LogicBlox  
Theodore Johnson, AT&T Labs  
Murat Kantarcioglu, University of Texas, Dallas  
Raghav Kaushik, Microsoft Research  
Ken-ichi Kawarabayashi, National Institute of Informatics  
Bettina Kemme, McGill University

Eamonn Keogh, University of California, Riverside  
Florian Kerschbaum, SAP  
Dan Kifer, Pennsylvania State University  
Masaru Kitsuregawa, Tokyo University  
Nick Koudas, University of Toronto  
Tim Kraska, Brown University  
Ravi Kumar, Google  
Laks V.S. Lakshmanan, University of British Columbia  
Justin Levandoski, Microsoft Research  
Feifei Li, University of Utah  
Guoliang Li, Tsinghua University  
Jimmy Lin, University of Maryland  
Qiong Luo, Hong Kong University of Science and Technology  
Ashwin Machanavajjhala, Duke University  
Nikos Mamoulis, University of Hong Kong  
Stefan Manegold, Centrum Wiskunde and Informatica  
Ioana Manolescu, Inria Saclay  
Norman May, SAP AG Walldorf  
Sergey Melnik, Google  
Gerome Miklau, University of Massachusetts, Amherst  
Mohamed Mokbel, University of Minnesota  
Bongki Moon, Seoul National University  
Barzan Mozafari, University of Michigan  
Arnab Nandi, The Ohio State University  
Vivek Narasayya, Microsoft Research  
Mario A. Nascimento, University of Alberta  
Rimma Nehme, Microsoft Gray Systems Lab  
Thomas Neumann, Technical University Munich  
Fatma Ozcan, IBM Research  
Themis Palpanas, Paris Descartes University  
Hweehwa Pang, Singapore Management University  
Olga Papaemmanouil, Brandeis University  
Aditya Parameswaran, University of Illinois, Urbana-Champaign  
Srinivasan Parthasarathy, The Ohio State University  
Jignesh Patel, University of Wisconsin  
Andy Pavlo, Carnegie Mellon University  
Jian Pei, Simon Fraser University  
Neoklis Polyzotis, University of California, Santa Cruz  
Dan Ports, University of Washington  
Rachel Pottinger, University of British Columbia  
Ganesh Ramakrishnan, IIT Bombay  
Vijayshankar Raman, IBM Research  
Rajeev Rastogi, Amazon  
Matthias Renz, Ludwig Maximilians University  
Kenneth Ross, Columbia University  
Pierangela Samarati, University of Milan  
Atish Das Sarma, eBay Research  
Bernhard Seeger, University of Marburg  
Sudipta Sengupta, Microsoft Research  
Cyrus Shahabi, University of Southern California  
Mohamed Soliman, Pivotal  
Torsten Suel, New York University  
Letizia Tanca, Politecnico di Milano  
Sandeep Tata, Google  
Nesime Tatbul, Intel Labs and MIT  
Doug Terry, Samsung Research  
Evmaria Terzi, Boston University

Jens Teubner, TU Dortmund University  
Martin Theobald, Ulm University  
Srikanta Tirthapura, Iowa State University  
Kristin Tufte, Portland State University  
Jeffrey Ullman, Stanford University  
Jaideep Vaidya, Rutgers University  
Vasilis Vassalos, Athens University of Economics and  
Business  
Haixun Wang, Google

Wei Wang, University of New South Wales  
Steven Whang, Google  
Sai Wu, Zhejiang University  
Chuan Xiao, Nagoya University  
Jeffrey Xu Yu, Chinese University of Hong Kong  
Cong Yu, Google  
Nan Zhang, George Washington University  
Xiaofang Zhou, University of Queensland  
Lei Zou, Peking University