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
# 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** ........................................... Hongbin Ma, Bin Shao, Yanghua Xiao, Liang Jeff Chen, Haixun Wang 900

**Measuring and Optimizing Distributed Array Programs** ........................................ Mingxing Zhang, Yongwei Wu, Kang Chen, Teng Ma, Weimin Zheng 912

**YourSQL: A High-Performance Database System Leveraging In-Storage Computing** Insoon Jo, Duck-Ho Bae, Andre S. Yoon, Jeong-Uk Kang, Sangyeun Cho, Daniel DG Lee, Jaeheon Jeong .................................................................................................................. 924

**Lifetime-Based Memory Management for Distributed Data Processing Systems** Lu Lu, Xuanhua Shi, Yongluan Zhou, Xiong Zhang, Hai Jin, Cheng Pei, Ligang He, Yuanzhen Geng 936

**ActiveClean: Interactive Data Cleaning for Statistical Modeling** .......................... Sanjay Krishnan, Jiannan Wang, Eugene Wu, Michael J. Franklin, Ken Goldberg 948

**Compressed Linear Algebra for Large-Scale Machine Learning** ......................... Ahmed Elgohary, Matthias Boehm, Peter J. Haas, Frederick R. Reiss, Berthold Reinwald 960

**Fast Queries Over Heterogeneous Data Through Engine Customization** .......... Manos Karpathiotakis, Ioannis Alagiannis, Anastasia Ailamaki 972

**Data-driven Visual Graph Query Interface Construction and Maintenance: Challenges and Opportunities** ................................................................................................................. 984

**Detecting Data Errors: Where are we and what needs to be done?** ................... Ziawasch Abedjan, Xu Chu, Dong Deng, Raul Castro Fernandez, Ihab F. Ilyas, Mourad Ouzzani, Paolo Papotti, Michael Stonebraker, Nan Tang .................................................................................................................. 993

**Exploiting Soft and Hard Correlations in Big Data Query Optimization** .......... Hai Liu, Dongqing Xiao, Pankaj Didwania, Mohamed Y. Eltabakh 1005

**Interactive Browsing and Navigation in Relational Databases** ............................. Minsuk Kahng, Shamkant B. Navathe, John T. Stasko, Duen Horng Chau 1017

**Cheap Data Analytics using Cold Storage Devices** ............................................. Renata Borovica-Gajić, Raja Appuswamy, Anastasia Ailamaki 1029

**Parallel Local Graph Clustering** ............................................................................. Julian Shun, Farbod Roosta-Khorasani, Kimon Fountoulakis, Michael W. Mahoney 1041
Online Minimum Matching in Real-Time Spatial Data: Experiments and Analysis ................................. Yongxin Tong, Jieying She, Bolin Ding, Lei Chen, Tianyu Wo, Ke Xu 1053

Index-Assisted Hierarchical Computations in Main-Memory RDBMS .......................................................... Robert Brunel, Norman May, Alfons Kemper 1065

Dynamic Influence Analysis in Evolving Networks .................................................................................................. Naoto Ohsaka, Takuya Akiba, Yuichi Yoshida, Ken-ichi Kawarabayashi 1077

Distance-based Outlier Detection in Data Streams .................................................................................................... Luan Tran, Liyue Fan, Cyrus Shahabi 1089

On Measuring the Lattice of Commonalities Among Several Linked Datasets ......................................................... Michalis Mountantonakis, Yannis Tzitzikas 1101

Oblivious RAM: A Dissection and Experimental Evaluation ...................................................................................... Zhao Chang, Dong Xie, Feifei Li 1113

Optimization of Conjunctive Predicates for Main Memory Column Stores ............................................................ Fisnik Kastrati, Guido Moerkotte 1125

Explaining Outputs in Modern Data Analytics .......................................................................................................... Zaheer Chothia, John Liagouris, Frank McSherry, Timothy Roscoe 1137

RDF Graph Alignment with Bisimulation ................................................................................................................. Peter Buneman, Slawek Staworko 1149

Teaching an RDBMS about ontological constraints .................................................................................................. Damian Bursztyn, François Goasdué, Ioana Manolescu 1161

BLAST: a Loosely Schema-aware Meta-blocking Approach for Entity Resolution .................................................. Giovanni Simonini, Sonia Bergamaschi, H. V. Jagadish 1173

LSH Ensemble: Internet-Scale Domain Search ........................................................................................................ Erkang Zhu, Fatemeh Nargesian, Ken Q. Pu, Renée J. Miller 1185


ATHENA: An Ontology-Driven System for Natural Language Querying over Relational Data Stores ........... Diptikalyan Saha, Avrilia Floratou, Karthik Sankaranarayanan, Umar Farooq Mihans, Ashish R. Mittal, Fatma Özcan ............................................................................................................. 1209

Incremental Computation of Common Windowed Holistic Aggregates .................................................................. Richard Wesley, Fei Xu 1221

Effective Community Search for Large Attributed Graphs ....................................................................................... Yixiang Fang, Reynold Cheng, Siqiang Luo, Jiafeng Hu 1233
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
Inab 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.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, Pontificial 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 Chirikova, 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 Kantarcìoglu, 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 Ozcak, 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
Bemhard 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
Sandip Tata, Google
Nesime Tatbul, Intel Labs and MIT
Doug Terry, Samsung Research
Evimaria 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