

VLDB2016

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



Proceedings of the VLDB Endowment

Volume 9, No. 14 – October 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. 14, October 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.

Volume 9, Number 14, October 2016: VLDB 2016

Pages ii – vii and 1623 - 1718

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 2016 Organization and Review Board	iv
Closing Remarks from the Program Chairs	vi

Research Papers

Quill: Efficient, Transferable, and Rich Analytics at Scale	
<i>Badrish Chandramouli, Raul Castro Fernandez, Jonathan Goldstein, Ahmed Eldawy, Abdul Quamar</i>	
.....	1623
Perturbation Analysis of Database Queries.....	
..... <i>Brett Walanz, Jun Yang</i>	1635
HippogriffDB: Balancing I/O and GPU Bandwidth in Big Data Analytics.....	
..... <i>Jing Li, Hung-Wei Tseng, Chunbin Lin, Yannis Papakonstantinou, Steven Swanson</i>	1647
Non-Invasive Progressive Optimization for In-Memory Databases.....	
..... <i>Steffen Zeuch, Holger Pirk, Johann-Christoph Freytag</i>	1659
Dscaler: Synthetically Scaling a Given Relational Database.....	
..... <i>J.W. Zhang, Y. C. Tay</i>	1671
Fast and Adaptive Indexing of Multi-Dimensional Observational Data.....	
..... <i>Sheng Wang, David Maier, Beng Chin Ooi</i>	1683
Price-Optimal Querying with Data APIs.....	
..... <i>Prasang Upadhyaya, Magdalena Balazinska, Dan Suciu</i>	1695
Voodoo – A Vector Algebra for Portable Database Performance on Modern Hardware.....	
..... <i>Holger Pirk, Oscar Moll, Matei Zaharia, Sam Madden</i>	1707

VLDB 2016 ORGANIZATION AND REVIEW BOARD

GENERAL CHAIRS

Anand Deshpande, Persistent Systems
T. M. Vijayarajan, 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
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

CLOSING REMARKS FROM THE PROGRAM CHAIRS

We received 719 research submissions to PVLDB volume 9 during the April 2015–March 2016 period. In the first three quarters of the cycle, a steady stream of around 40 to 50 papers arrived each month in the CMT portal, but in the last quarter, the numbers shot up sharply, culminating in the well-documented “March spike”, wherein more than 200 papers were submitted in the final month. Each manuscript was rigorously reviewed by three reviewers from the Review Board, under the supervision of an Associate Editor. Authors of papers recommended for revision were given up to three months to submit the updated version – our experience was that about 40% were revised within one month, 25% in two months, and the remaining 35% took the full three months.

We accepted 112 papers for publication in this volume, with 104 of them being presented in the research track of the recently concluded 42nd International Conference on Very Large Data Bases (VLDB 2016), held during September 5–9 in New Delhi, India. The remaining few papers, which all appear in this concluding issue, will be presented in the VLDB 2017 conference, scheduled for next August in Munich, Germany.

A hallmark of this year’s reviewing process is that, in spite of the large number of submissions, there were *no postponements* of decisions – all fresh submissions were processed in around six weeks, and revisions in around four weeks, maintaining the advertised schedule. We are grateful to our Associate Editors and the PVLDB Review Board for this signal achievement. A second special feature was that we consciously and carefully evaluated the relevance of each submission to the *scope* of PVLDB, and in a small percentage of the cases, the authors were requested to submit their work to alternative and more appropriate forums. We hope that this filtering policy has resulted in a more pertinent set of papers appearing in this volume.

Drilling down into the content of the submissions, the most popular topics were Graph and Semi-Structured Data, Data Mining, Distributed Data Systems, Query Optimization and Execution, Data Security and Privacy, Data Indexing and Storage, Data Quality, Spatio-Temporal Data, Data Integration and ETL, and Data Streams. The leading categories in which papers were accepted comprised of Graph and Semi-Structured Databases, Query Optimization and Execution, Distributed Data Systems, Data Integration and ETL, and Data Quality.

In addition to the research track, VLDB 2016 featured industrial and demonstration tracks. The Industrial track attracted as many as 53 submissions. The industrial track chairs, assisted by their program committee, accepted 17 papers for presentation, augmented with invited talks that focused on key industrial innovations in the data management area. The Demonstration track received 103 submissions of novel software prototypes, which were evaluated by the demonstration track chairs with the help of a separate program committee. They accepted 37 attractive demonstrations, each of which was presented twice during the conference. All the accepted Industrial and Demonstration papers have been published in Issue 13 of PVLDB volume 9.

The conference also hosted an excellent set of tutorials, hand-picked by the Tutorial Chairs, consisting of six lectures covering important emerging areas in our fast changing world of database technology. An exciting panel was organized by the Panel Chairs to debate the influence of Artificial Intelligence on contemporary data management issues. There were seven thematic workshops chosen by the Workshop Chairs, covering a range of modern topics, as well as a PhD workshop to provide feedback and guidance to young researchers on their ongoing work.

A paper on compressed linear algebra for large-scale machine learning by a team from University of Maryland and IBM Research, which appeared in Issue 12, was chosen as the Best Paper of VLDB 2016. VisDPT, a system for visual exploration of differentially private trajectories developed by a team of researchers from Duke University was recognized as the Best Demonstration. We are grateful to the VLDB 2016 Best Paper and Best Demo Award Committees for making these selections. Further, as in past VLDB conferences, the award session also featured presentations by winners of VLDB Endowment Awards, hand-picked by the Endowment Awards Committee.

Last but not the least, the technical program was rounded off with thought-provoking keynote talks from Professor Ion Stoica of University of California, Berkeley, who shared his insights on the Big Data revolution, and Dr. Anand Rajaraman, a successful serial entrepreneur from Silicon Valley, who provided his perspective on the evolution and future of data-driven innovation.

It has been our privilege to assemble and present the 14 issues of PVLDB volume 9. We are grateful to authors who submitted strong technical papers to PVLDB Volume 9 as well as our Associate Editors and members of PVLDB Board for their dedicated contributions over the twelve reviewing segments. We are also indebted to the industrial, demonstration, tutorial, panel and workshop chairs, and their program committees, for helping us put together a technically strong VLDB 2016. A special mention is due to our Proceedings Chairs for diligently compiling and publicizing this entire volume over the past year. Thank you all very much.

Surajit Chaudhuri and Jayant Haritsa
PVLDB Volume 9 Editors in Chief
VLDB 2016 Program Committee Chairs