



Proceedings of the VLDB Endowment

Volume 8, No. 2 – October 2014

**Proceedings of the 41st International Conference on
Very Large Data Bases, Kohala Coast, Hawaii**

Program Chairs and Editors-in-Chief:

Chen Li and Volker Markl

Associate Editors – Research and Innovative Systems Tracks:

**Kevin Chang, Shivnath Babu, Magdalena Balazinska, Felix Naumann, Stefan Manegold, Yi Chen, Fatma Ozcan,
Jignesh Patel**

Associate Editors – Experiments and Analysis Track:

Rainer Gemulla

Proceedings Chairs:

Tyson Condie, Daisy Zhe Wang

PVLDB – Proceedings of the VLDB Endowment

Volume 8, No. 2, October 2014.

The 41st International Conference on Very Large Data Bases, Kahola Coast, Hawaii.

Copyright 2014 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 8, Number 2, October 2014: VLDB 2015

Pages ii - ix and 101 - 185

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 2015 Organization and Review Board	iv

Letters

Letter from the Associate Editors	<i>Kevin C. Chang</i>	ix
---	-----------------------	----

Research Papers

Selectivity Estimation on Streaming Spatio-Textual Data Using Local Correlations	101
..... <i>Xiaoyang Wang, Ying Zhang, Wenjie Zhang, Xuemin Lin, Wei Wang</i>	
Processing Moving <i>k</i> NN Queries Using Influential Neighbor Sets.....	
..... <i>Chuanwen Li, Yu Gu, Jianzhong Qi, Ge Yu, Rui Zhang, Wang Yi</i>	113
Scaling Up Crowd-Sourcing to Very Large Datasets: A Case for Active Learning	125
..... <i>Barzan Mozafari, Purna Sarkar, Michael Franklin, Michael Jordan, Samuel Madden</i>	
CANDS: Continuous Optimal Navigation via Distributed Stream Processing	137
..... <i>Dingyu Yang, Dongxiang Zhang, Kian-Lee Tan, Jian Cao, Frédéric Le Mouél</i>	
Rare Time Series Motif Discovery from Unbounded Streams.....	149
..... <i>Nurjahan Begum, Eamonn Keogh</i>	
Pregelix: Big(ger) Graph Analytics on a Dataflow Engine	161
..... <i>Yingyi Bu, Vinayak Borkar, Jianfeng Jia, Michael J. Carey, Tyson Condie</i>	
Profiling R on a Contemporary Processor	173
..... <i>Shriram Sridharan, Jignesh M. Patel</i>	

VLDB 2015 ORGANIZATION AND REVIEW BOARD

General Chairs

Michael J. Carey, University of California, Irvine

Program Chairs and Editors-in-Chief of PVLDB 8

Chen Li, University of California, Irvine

Volker Markl, TU Berlin

Research and Innovative Systems Tracks Associate Editors

Kevin Chang, U Illinois

Shivnath Babu, Duke University

Felix Naumann, Hasso Plattner Institute

Stefan Manegold, CWI Amsterdam

Yi Chen, NJIT

Fatma Ozcan, IBM Research Almaden

Jignesh Patel, University of Wisconsin, Madison

Magdalena Balazinska, University of Washington

Experiments and Analysis Track Associate Editors

Rainer Gemulla, MPI Saarbrücken, Germany

Industrial, Applications and Experience Track Associate Editors

Anhai Doan, University of Wisconsin, Madison

Prasan Roy, Sclera

Gregor Hackenbroich, SAP

Demonstration Chair

Alfons Kemper, TU München

Tutorial Chairs

Tova Milo, Tel Aviv University

Pierre Senellart, Telecom Paris Tech, France

Panel Chair

Joseph M. Hellerstein, University of California, Berkeley

PhD Workshop Chairs

Rachel Pottinger (UBC)

Proceedings Chairs

Daisy Zhe Wang, University of Florida

Tyson Condie, University of California, Los Angeles

Sponsorship Chairs

Michael Franklin, University of California, Berkeley

Edward Change, HTC

Patrick Valduriez, INRIA

Local Organization Chair

Lipyeow Lim, University of Hawaii

Conference and Registration Chairs

Ke Chen, Zhejiang University

Cuiping Li, Renmin University

Publicity and Web Management Chair

Rada Chirkova, North Carolina State

Treasury Chair

Malu Castellanos, HP Labs

VLDB Endowment Liaison

Paul Larson, Microsoft Research

PVLDB Managing Editor

Divesh Srivastava, AT&T Labs

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

Research Track Review Board

Daniel Abadi, Yale University
Alberto Abello, UPC Barcelona
Ashraf Aboulmaga, Qatar Computing Research
Institute
Foto Afrati, NTU Athens
Sihem Amer Yahia, CNRS LIG
Aijun An, York University
Arvind Arasu, Microsoft Research
Walid Aref, Purdue University
Paolo Atzeni, Roma Tre University
Denilson Barbosa, University of Alberta
Srikanta Bedathur, IBM Research
Philip Bernstein, Microsoft Research
Michael Böhlen, University of Zürich
Peter Boncz, CWI Amsterdam
Angela Bonifati, Lille 1 U and INRIA
Philippe Bonnet, IT U of Copenhagen
Nico Bruno, Microsoft
Alex Buchmann, TU Darmstadt
Mike Cafarella, University of Michigan
K. Selcuk Candan, Arizona State University
Malu Castellanos, HP Labs
Kaushik Chakrabarti, Microsoft Research
Lei Chen, Hong Kong U of Science and Technology
Fei Chiang, McMaster University
Byron Choi, Hong Kong Baptist University
Philippe Cudre Mauroux, Fribourg University
Mahashweta Das, HP Labs
Sudipto Das, Microsoft Research
Amol Desphande, University of Maryland
Stefan Dessloch, TU Kaiserslautern
Jens Dittrich, Saarland University
Alin Dobra, University of Florida
Xin Luna Dong, Google
Jennie Duggan, MIT
Wenfei Fan, University of Edinburgh
Alan Fekete, University of Sydney Australia

Peter Fischer, Universität Freiburg
Avrilia Floratou, IBM
Avigdor Gal, Technion
Minos Garofalakis, Technical U of Crete
Wolfgang Gatterbauer, Carnegie Mellon U
Tingjian Ge, University of Massachusetts Lowell
Floris Geerts, University of Antwerp
Lukasz Golab, University of Waterloo
Torsten Grust, Universität Tübingen
Jarek Gryz, York University
Dimitrios Gunopulos, University of Athens
Hakan Hacigumus, NEC Labs
Wook Shin Han, POSTECH
Seif Haridi, KTH Stockholm
Oktie Hassanzadeh, IBM Research
Bingsheng He, Nanyang Technological University
Jeffrey Heer, University of Washington
Herodotos Herodotou, Microsoft Research
Katja Hose, Aalborg University
Vagelis Hristidis, University of California, Riverside
Jeong-Hyon Hwang, State University of New York at
Albany
Stratos Idreos, Harvard University
Yannis Ioannidis, University Of Athens
Zachary Ives, University of Pennsylvania
Christopher Jermaine, Rice University
Ruoming Jin, Kent State University
Alekh Jindal, Massachusetts Institute of Technology
Ryan Johnson, University of Toronto
Eser Kandogan, IBM Research Almaden
Gjergji Kasneci, Hasso Plattner Institute
Asterios Katsifodimos, TU Berlin
Yannis Katsis, University of California, San Diego
Daniel Keim, Universität Konstanz
Bettina Kemme, McGill University
Eamonn Keogh, University of California, Riverside
Martin Kersten, CWI Amsterdam
Daniel Kifer, Penn State University
Hideaki Kimura, Hewlett Packard
George Kollios, Boston University

Donald Kossman, ETH Zurich
 Nick Koudas, University of Toronto
 Georgia Koutrika, HP Labs
 Tim Kraska, Brown University
 Harumi Kuno, HP Labs
 Laks Lakshmanan, University of British Columbia
 Paul Larson, Microsoft Research
 Hongrae Lee, Google
 Wolfgang Lehner, TU Dresden
 Alberto Lerner, New York University
 Ulf Leser, Humboldt Universität zu Berlin
 Justin Levandoski, Microsoft Research
 Feifei Li, University of Utah
 Guoliang Li, Tsinghua University
 Jianzhong Li, Harbin Institute of Technology
 Yunyao Li, IBM Research Almaden
 Erietta Liarou, EPF Lausanne
 Xuemin Lin, University of New South Wales
 Ziyang Liu, NEC Labs America
 Eric Lo, Polytecnic University of Hong Kong
 Guy Lohman, IBM Research Almaden
 Jiaheng Lu, Renmin University of China
 Qiong Luo, Hong Kong University of Science and Technology
 Jayant Madhavan, Google
 Ioana Manolescu, INRIA
 Patrick Marcel, University of Tours
 Marta Mattoso, Federal University of Rio de Janeiro
 Alexandra Meliou, University of Massachusetts Amherst
 Sergey Melnik, Google
 Weiyi Meng, Binghamton University
 Sebastian Michel, Saarland University
 Iris Miliaraki, Yahoo Labs Barcelona
 Renee Miller, University of Toronto
 Zhou Minqi, East China Normal University
 Prasenjit Mitra, Penn State University
 Bernhard Mitschang, Universität Stuttgart
 Mohamed Mokbel, Northeastern University
 Barzan Mozafari, University of Michigan

Hannes Mühleisen, CWI Amsterdam
 Arnab Nandi, Ohio State University
 Vivek Narasayya, Microsoft Research
 Jeffrey Naughton, University of Wisconsin Madison
 Rimma Nehme, Microsoft
 Thomas Neumann, TU Munich
 Raymond Ng, University of British Columbia
 Christopher Olston, Google
 Dan Olteanu, Oxford University & LogicBlox
 Beng Chin Ooi, National University of Singapore
 M. Tamer Özsu, University of Waterloo
 Themis Palpanas, Paris Descartes University
 Ippokratis Pandis, IBM Research Almaden
 Dimitris Papadias, Hong Kong U of Science and Technology
 Paolo Papotti, Qatar Computing Research Institute
 Andy Pavlo, Carnegie Mellon University
 Torben Bach Pedersen, Aalborg University
 Jian Pei, Simon Fraser University
 Peter Pietzuch, Imperial College London
 Evaggelia Pitoura, University of Ioannina
 Alkis Polyzotis, Google and University of California, Santa Cruz
 Fabio Porto, LNCC Brazil
 Li Qian, Facebook
 Jorge Arnulfo Quiane Ruiz, Qatar Computing Research Institute
 Tilmann Rabl, University of Toronto
 Erhard Rahm, Universität Leipzig
 Krithi Ramamritham, IIT Bombay
 Ravi Ramamurthy, Microsoft Research
 Vijayshankar Raman, IBM Research Almaden
 Mirek Riedewald, Northeastern University
 Tore Risch, Uppsala Universitet
 Kenneth Ross, Columbia University
 Elke Rundensteiner, Worcester Polytechnic Institute US
 Barna Saha, AT&T Labs Research
 Kenneth Salem, University of Waterloo
 Simonas Saltenis, Aalborg University

Kai Uwe Sattler, TU Ilmenau
Eric Sedlar, Oracle Labs
Bernhard Seeger, Philipps Universität Marburg
Kyuseok Shim, Seoul National University
Jerome Simeon, IBM Research Watson
Alkis Simitsis, HP Labs
Michael Stonebraker, Massachusetts Institute of Technology
Julia Stoyanovich, Drexel University
Nan Tang, Qatar Computing Research Institute
Yufei Tao, The Chinese University of Hong Kong
Sandeep Tata, Google
Nesime Tatbul, Intel Labs and MIT
Arash Termehchy, Oregon State University
Evimaria Terzi, Boston University
Jens Teubner, TU Dortmund
Martin Theobald, University of Antwerp
Andreas Thor, University of Leipzig
Yuanyuan Tian, IBM Research Almaden
Anthony Tung, National University of Singapore
Vasilis Vassalos, Athens University
Yannis Vassiliou, NTU Athens
Yannis Velegrakis, University of Trento

Rares Vernica, HP Labs
Gottfried Vossen, WWU Münster
Florian Waas, Datometry Inc.
Daisy Zhe Wang, University of Florida
Haixun Wang, Google
Wei Wang, University of New South Wales
Gerhard Weikum, Max Planck Institut für Informatik
Till Westmann, Oracle Labs
Steven Whang, Google
Kyu Young Whang, KAIST
Raymond Chi Wing Wong, Hong Kong University of Science and Technology
Xiaokui Xiao, Nanyang Technological University
Xifeng Yan, University of California, Santa Barbara
Jun Yang, Duke University
Xiaochun Yang, Northeastern University China
Cong Yu, Google Research
Ge Yu, Northeastern University China
Jeffrey Xu Yu, Chinese University of Hong Kong
Xiaofang Zhou, The University of Queensland
Esteban Zimanyi, UL Brussels
Marcin Zukowski, Snowflake Computing

LETTER FROM THE ASSOCIATE EDITORS

I am delighted to present to you the second issue of the Proceedings of the VLDB Endowment (PVLDB), Volume 8. PVLDB publishes papers accepted through a journal-style reviewing process from papers submitted on a year-round monthly submission schedule. All PVLDB papers published in this volume will be presented at the VLDB 2015 Conference, to be held in Kohala Coast, Hawaii, August 31 - September 4, 2015.

This issue of PVLDB showcases seven excellent papers. These papers span a broad spectrum of topics, demonstrating the healthy diversity of issues our community is investigating. In particular, the topics in this issue cover selectivity estimation, spatial query processing, time series pattern mining, shortest path queries, graph analytics, R program profiling, and crowd sourcing. I thank all the authors for making high-quality submissions and working with the PC to improve them.

I hope that you enjoy our selection of papers in this issue. We thank the reviewers for their hard work and debates for making the best selection to present to our community-- and I hope that you will continue to submit your best work to PVLDB's future issues!

Kevin C. Chang (PVLDB Associate Editor)