

VLDB2014

40th International Conference on Very Large Data Bases, Hangzhou, China



Proceedings of the VLDB Endowment

Volume 7, No. 4 – December 2013

**Proceedings of the 40th International Conference on
Very Large Data Bases, Hangzhou, China**

Program Chairs and Editors-in-Chief:

H. V. Jagadish, Aoying Zhou

Associate Editors – Research and Innovative Systems Tracks:

**Shivnath Babu, Lei Chen, Graham Cormode, Bin Cui, Wynne Hsu, Martin Kersten,
Donald Kossman, Elke Rundensteiner, Kyuseok Shim, Wang-Chiew Tan, Letizia Tanca, Jeffrey Yu**

Associate Editors – Experiments and Analysis Track:

Gao Cong, Jens Dittrich

Associate Editors – Vision Track:

Zachary Ives

Proceedings Chairs:

Li Xiong, Cong Yu

PVLDB – Proceedings of the VLDB Endowment

Volume 7, No. 4, December 2013.

The 40th International Conference on Very Large Data Bases, Hangzhou, China.

Copyright 2013 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 7, Number 4, December 2013: VLDB 2014

Pages ii - ix and 217 - 328

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 2014 Organization and Review Board	iv

Letters

What Should We Aspire to in Research	<i>Martin Kersten</i>	ix
--	-----------------------	----

Research Papers

Delta: Scalable Data Dissemination under Capacity Constraints	217
..... <i>Konstantinos Karanasos, Asterios Katsifodimos, Ioana Manolescu</i>	
GeoScope: Online Detection of Geo-Correlated Information Trends in Social Networks	229
..... <i>Ceren Budak, Theodore Georgiou, Divyakant Agrawal, Amr El Abbadi</i>	
Optimization for iterative queries on MapReduce	241
..... <i>Makoto Onizuka, Hiroyuki Kato, Soichiro Hidaka, Keisuke Nakano, Zhenjiang Hu</i>	
Willingness Optimization for Social Group Activity	253
..... <i>Hong-Han Shuai, De-Nian Yang, Philip S. Yu, Ming-Syan Chen</i>	
High Performance Stream Query Processing With Correlation-Aware Partitioning	265
..... <i>Lei Cao, Elke A. Rundensteiner</i>	
OLTP-Bench: An Extensible Testbed for Benchmarking Relational Databases	277
..... <i>Djellel Eddine Difallah, Andrew Pavlo, Carlo Curino, Philippe Cudre-Mauroux</i>	
Gestural Query Specification	289
..... <i>Arnab Nandi, Lilong Jiang, Michael Mandel</i>	
Scalable Discovery of Unique Column Combinations	301
..... <i>Arvid Heise, Jorge-Arnulfo Quijano-Ruiz, Ziawasch Abedjan, Anja Jentzsch, Felix Naumann</i>	
Earth Mover's Distance based Similarity Search at Scale	313
..... <i>Yu Tang, Leong Hou U, Yilun Cai, Nikos Mamoulis, Reynold Cheng</i>	
SeeDB: Visualizing Database Queries Efficiently	325
..... <i>Aditya Parameswaran, Neoklis Polyzotis, Hector Garcia-Molina</i>	

VLDB 2014 ORGANIZATION AND REVIEW BOARD

Honorary Chair

Yunhe Pan, Chinese Academy of Engineering

General Chairs

Chun Chen, Zhejiang University

Sharad Mehrotra, University of California, Irvine

Program Chairs and Editors-in-Chief of PVLDB 7

H. V. Jagadish, University of Michigan

Aoying Zhou, East Normal University, China

Research and Innovative Systems Tracks Associate Editors

Shivnath Babu, Duke University

Lei Chen, Hong Kong University of Science and Technology

Graham Cormode, University of Warwick

Bin Cui, Peking University, China

Wynne Hsu, NUS

Martin Kersten, CWI

Donald Kossman, ETH

Elke Rundensteiner, WPI

Kyuseok Shim, Seoul National University

Wang-Chiew Tan, University of California, Santa Cruz

Letizia Tanca, Poli Milano

Jeffrey Yu, Chinese University of Hong Kong

Experiments and Analysis Track Associate Editors

Gao Cong, Nanyang Technology University

Jens Dittrich, Saarland

Visions Track Associate Editor

Zachary Ives, University of Pennsylvania

Industrial and Applications Track Associate Editors

Umeshwar Dayal, HP

C. Mohan, IBM

Ge Yu, Northeastern University, China

Demonstration Chairs

Mong-Li Lee, NUS
Feifei Li, University of Utah
Sunil Prabhakar, Purdue

Tutorial Chairs

Xiaoyong Du, Renmin University
Murat Kantarcioglu, University of Texas, Dallas
Divesh Srivastava, AT&T Labs

Research Track Review Board

Sibel Adali, Rensselaer Polytechnic Institute
Foto Afrati, NTU Athens
Yanif Ahmad, JHU
Jose Luis Ambite, ISI - USC
Walid Aref, Purdue University
Claudia Bauzer Medeiros, University of Campinas
Srikanta Bedathur, IIIT Delhi
Michael Benedikt, Oxford University
Sonia Bergamaschi, Universita Modena
Laure Berti-Equille, IRD, France
Leopoldo Bertossi, Carleton University, Ottawa
Subhash Bhalla, University of Aizu, Japan
Peter Boncz, CWI
Angela Bonifati, University of Lille 1
Rajesh Bordawekar, IBM Watson Research Center
Omar Boucelma, Aix-Marseille University
Nico Bruno, Microsoft Research
Andrea Cali, University of London, Birkbeck College
Malu Castellanos, HP Labs
Badrish Chandramouli, Microsoft Research
Adriane Chapman, Mitre
Gang Chen, Zhejiang University
Yi Chen, New Jersey Institute of Technology
James Cheng, CUHK
Reynold Cheng, University of Hong Kong
Brian Cooper, Google, USA

Workshop Chairs

Anastasia Ailamaki, EPFL
Kaushik Chakrabarti, Microsoft

Panel Chairs

Hakan Hacigumus, NEC Labs
Jignesh Patel, University of Wisconsin
Xiaoyang Sean Wang, Fudan University

Philippe Cudré-Mauroux, University of Fribourg
Carlo Curino, MIT
Gautam Das, UT Arlington and QCRI
Sudipto Das, Microsoft Research
Anish Das Sarma
Atish Das Sarma, eBay Research Labs
Khuzaima Daudjee, University of Waterloo
Antonios Deligiannakis, Technical University of Crete
Daniel Deutch, Ben Gurion University
Yanlei Diao, University of Massachusetts Amherst
Xin (Luna) Dong, Google, USA
Sameh Elnikety, Microsoft Research
Mohamed Eltabakh, Worcester Polytechnic Institute
Ihab F. Ilyas, QCRI
Hakan Ferhatosmanoglu, Bilkent University
Ada Wai-Chee Fu, Chinese University of Hong Kong
Minos Garofalakis, Technical University of Crete
Wolfgang Gatterbauer, Carnegie Mellon University
Tingjian Ge, University of Massachusetts, Lowell
Buğra Gedik, Bilkent University
Rainer Gemulla, Max-Planck-Institut Saarbrücken
Gabriel Ghinita, University of Massachusetts Boston
Parke Godfrey, York University
Lukasz Golab, University of Waterloo
Sergio Greco, University of Calabria
Le Gruenwald, University of Oklahoma

Giovanna Guerrini, Universita Genova
 Krishna Gummadi, MPI-SWS
 Rahul Gupta, Google Research
 Rajeev Gupta, IBM Research
 Shyam Gupta, IIT Delhi
 Marios Hadjieleftheriou, AT&T labs
 Wook-Shin Han, KNU, Korea
 Kuno Harumi, HP Labs
 Bingsheng He, NTU Singapore
 Sven Helmer, Free University of Bozen-Bolzano
 Jan Hidders, TUDelft
 Wei Hong, Cisco System Inc.
 Katja Hose, Aalborg University
 Zi Huang, University of Queensland
 Jeong-Hyon Hwang, SUNY - Albany
 Seung-won Hwang, POSTECH, Korea
 Stratos Idreos, CWI
 Yoshiharu Ishikawa, Nagoya University
 Zachary Ives, University of Pennsylvania
 Ricardo Jimenez-Peris, Technical University of Madrid
 Cheqing Jin, East China Normal University
 Ruoming Jin, Kent State University
 Alekh Jindal, Saarland University/MIT
 Ryan Johnson, University of Toronto
 Dmitri V Kalashnikov, UC Irvine
 Panos Kalnis, KAUST, Saudi Arabia
 Ben Kao, Hong Kong University
 Panagiotis Karras, Rutgers University
 Yiping Ke, Institute of High Performance Computing
 Bettina Kemme, McGill University
 Daniel Kifer, PSU
 Benny Kimelfeld, IBM
 Hideaki Kimura, Microsoft Jim Gray Systems Lab
 George Kollios, Boston University
 Christian König, Microsoft Research
 Tim Kraska, Brown University
 Laks V. S. Lakshmanan, University of British Columbia
 Mounia Lalmas, Yahoo Inc.
 Mong-Li Lee, National University of Singapore
 Wolfgang Lehner, Technische University Dresden
 Justin Levandoski, Microsoft Research
 Chengkai Li, The University of Texas at Arlington
 Cuiping Li, Renmin University of China
 Feifei Li, University of Utah
 Guoliang Li, Tsinghua University
 Jianzhong Li, Harbin Institute of Technology
 Yunyao Li, IBM Almaden
 Zhanhuai Li, Northwestern Polytechnical University
 Dan Lin, Missouri S&T, USA
 Xuemin Lin, University of New South Wales
 Bin Liu, NEC Labs America
 Ziyang Liu, NEC Labs America
 Eric Lo, The Hong Kong Polytechnic University
 Qiong Luo, HKUST
 Shuai Ma, Beihang University
 Ashwin Machanavajjhala, Duke University
 Brad Malin, Duke University
 Nikos Mamoulis, University of Hong Kong
 Stefan Manegold, CWI
 Murali Mani, University of Michigan
 Ioana Manolescu, INRIA, France
 Amélie Marian, Rutgers University
 Volker Markl, TU Berlin
 Marta Mattoso, Federal University of Rio de Janeiro
 Frank McSherry, Microsoft
 Alexandra Meliou, Umass Amherst
 Marco Mesiti, University of Milano
 Dan Miranker, The University of Texas at Austin
 Mohamed Mokbel, University of Minnesota
 Bongki Moon, Seoul National University
 Yasuhiko Morimoto, Hiroshima University
 Mirella Moro, Universidade Federal de Minas Gerais
 Kyriakos Mouratidis, SMU, Singapore
 Karin Murthy, IBM India
 Arnab Nandi, Ohio State University
 Wolfgang Nejdl, University of Hannover

Thomas Neumann, Technology University Munchen
Boris Novikov, St Petersburg University
Dan Olteanu, Oxford University
Gultekin Ozsoyoglu, Case Western Reserve University
Tamer Ozsu, University of Waterloo
Esther Pacitti, University of Montpellier
Themis Palpanas, University of Trento
Ippokratis Pandis, IBM Almaden
Stelios Paparizos, Microsoft Research
Aditya Parameswaran, Stanford University
Srinivasan Parthasarathy, The Ohio State University
Jignesh Patel, University of Wisconsin
Andrew Pavlo, Brown University
Peter Pietzuch, Imperial College London
Neoklis Polyzotis, University of California - Santa Cruz
Cecilia M. Procopiuc, AT&T Labs
Li Qian, University of Michigan
Jorge Quiané-Ruiz, QCRI
Elisa Quintarelli, Politecnico di Milano
Maya Ramanath, IIT Delhi
Louiqa Raschid, University of Maryland
Vibhar Rastogi, Yahoo
Matthias Renz, University of Munich
Kenneth Ross, Columbia University
Sourav S Bhowmick, NTU, Singapore
Dimitris Sacharidis, IMIS Athena, Greece
Kenneth Salem, University of Waterloo
Maria Sapino, University of Torino
Kai-Uwe Sattler, TU Ilmenau
Monica Scannapieco, ISTAT
Bernhard Seeger, University of Marburg
Lidan Shou, Zhejiang University
Adam Silberstein, Trifacta
Lisa Singh, Georgetown University
Radu Sion, Stony Brook University
Yufei Tao, Chinese University of Hong Kong

Nesime Tatbul, ETH Zurich
Arash Termehchy, Oregon State University
Evimaria Terzi, University of Boston
Martin Theobald, Max Planck Institute, Germany
Srikanta Tirthapura, Iowa State University
Riccardo Torlone, Roma Tre University
Anthony Tung, National University of Singapore
Kostas Tzoumas, Technical University of Berlin
Sergei Vassilvitskii, Google Research
Marcos Vaz Salles, University of Copenhagen (DIKU)
Stratis Viglas, University of Edinburgh
Hoang Tam Vo, National University of Singapore
Daisy Zhe Wang, University of Florida
Haixun Wang, Microsoft Research Asia
Ke Wang, Simon Fraser University
Wei Wang, University of New South Wales
Xiaoling Wang, East China Normal University
Ingmar Weber, Yahoo
Raymond Chi Wing Wong, HKUST
Sai Wu, Zhejiang University
Yuqing Wu, Indiana University
Xiaokui Xiao, NTU
Dong Xin, Google
Jianliang Xu, Hong Kong Baptist University
Jun (Jim) Xu, Georgia Institute of Technology
Xifeng Yan, University of Santa Barbara
Xiaoyan Yang, Advanced Digital Science Center
Ke Yi, HKUST
Ge Yu, Northeastern University, China
Hwanjo Yu, POSTECH, Korea
Meihui Zhang, National University of Singapore
Wenjie Zhang, The University of New South Wales
Ying Zhang, The University of New South Wales
Zhenjie Zhang, Advanced Digital Science Center
Wenzhao Zhou, Georgetown University
Xiaofang Zhou, University of Queensland

PhD Workshop Chairs

Erich Neuhold, University of Vienna
Yun Yao Li, IBM

Sponsorship Chairs

Mike Carey, University of California, Irvine
Lizhu Zhou, Tsinghua University

Local Organization Chair

Lidan Shou, Zhejiang University

Web Management Chair

Sai Wu, Zhejiang University

Conference and Registration Chairs

Ke Chen, Zhejiang University
Cuiping Li, Renmin University

Publicity Chairs

Vasilis Vassalos, AUEB, Greece
Dunlu Peng, USST, China

Proceedings Chairs

Li Xiong, Emory University
Cong Yu, Google Research

Treasury Chair

Li (Eric) Qian, University of Michigan

VLDB Endowment Liaison

Kyu-Young Whang, KAIST

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

Logo Design

Guanmin Guo

WHAT SHOULD WE ASPIRE TO IN RESEARCH

PVLDB has been the premiere venue to demonstrate progress in systems design for data management. Often it takes years to mature, assimilate, and productize the solutions proposed for dealing with what is nowadays called Big Data challenges. Seemingly small steps by a crowd of researchers are needed, because building new solutions from the ground up is not affordable for most.

Nevertheless, innovative systems' research also calls for new avenues, high risk taken adventures, and a clear focus on what ultimately has to be achieved. In this respect, the big data challenge reminds me of the famous fable of the invention of Chess. "When the creator of the game of chess (in some tellings an ancient Indian mathematician) showed his invention to the ruler of the country, the ruler was so pleased that he gave him the right to name his prize for the invention. The man, who was very wise, asked the king this: that for the first square of the chess board, he would receive one grain of wheat (in some tellings, rice), two for the second one, four on the third one, ..." [wikipedia]

The Big Data hype resembles the king who fanatically fills the squares of the board. The computer scientists and engineers act as the faithful civil servants, to go after their treasure chest, putting on the board ever more powerful, increasing piles of hardware. Algorithms that work like an army of MR slaves to harvest it for nuggets unknown. Doubling the amount of data and processing power every other year. Scientists dream of selling their new instrument to the king, like the Square Kilometer Array (SKA) telescope. A single instrument with a data production rate several times the total internet traffic nowadays. A futile activity as the fable has clearly identified to the public thousands of years ago.

The corollary of the chess board payment scheme is that not only you can not find enough wheat in the universe, but one should not collect more wheat than one can eat, for it will rot away in storage. This also shows the way out of the data deluge dilemma. We should recognize freshness and deal with rottenness in the core of our Big Data analytics offerings and solutions. Finding the techniques, algorithms and system architectures to bend the slope of the data explosion curve is what I look forward and value highly in papers submitted to PVLDB.

Martin Kersten, CWI, University of Amsterdam, MonetDB
Research Track Associate Editor, VLDB 2014