

# VLDB2014

40<sup>th</sup> International Conference on Very Large Data Bases, Hangzhou, China



## Proceedings of the VLDB Endowment

Volume 7, No. 8 – April 2014

**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. 8, April 2014.

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

## **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](mailto:info@vldb.org).

Volume 7, Number 8, April 2014: VLDB 2014

Pages ii - ix and 601 - 660

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 2014 Organization and Review Board .....	iv

### Letters

PVLDB: An Evolutionary Path Forward .....	<i>Divesh Srivastava</i>	ix
---	--------------------------	----

### Research Papers

A Principled Approach to Bridging the Gap between Graph Data and their Schemas .....	601
..... <i>Marcelo Arenas, Gonzalo Diaz, Achille Fokoue, Anastasios Kementsietsidis, Kavitha Srinivas</i>	
An Efficient Publish/Subscribe Index for ECommerce Databases .....	613
..... <i>Dongxiang Zhang, Chee-Yong Chan, Kian-Lee Tan</i>	
String Similarity Joins: An Experimental Evaluation .....	625
..... <i>Yu Jiang, Guoliang Li, Jianhua Feng, Wen-Syan Li</i>	
Calibrating Data to Sensitivity in Private Data Analysis .....	637
..... <i>Davide Proserpio, Sharon Goldberg, Frank McSherry</i>	
Effective Multi-Modal Retrieval based on Stacked Auto-Encoders .....	649
..... <i>Wei Wang, Beng Chin Ooi, Xiaoyan Yang, Dongxiang Zhang, Yueting Zhuang</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 Pappas, 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



## **PVLDB: AN EVOLUTIONARY PATH FORWARD**

The introduction of the Proceedings of the VLDB, PVLDB for short, in 2008 was in many ways a revolution. The field of Data Management had, and continues to have, many vibrant conferences and prestigious journals. PVLDB was introduced as the first journal + conference hybrid, to combine their strengths and avoid their weaknesses. PVLDB introduced many innovative policies that have been well-liked by the data management community, such as year-around paper submissions, journal-quality reviews of conference-style papers, and rapid online publication of accepted papers followed by presentation at the VLDB conference. But there was also a lack of stability of policies in the early years of PVLDB, which is not unusual in revolutionary times.

In 2012, the VLDB Endowment created the position of PVLDB Managing Editor to oversee an evolutionary path forward for PVLDB, by ensuring stability in PVLDB policy across years, and working with the PVLDB Editors in Chief (who are responsible for the content of PVLDB) on policy innovations. I was appointed as the PVLDB Managing Editor and the chair of the PVLDB Advisory Committee, which includes Phil Bernstein, Michael Boehlen, Peter Buneman, Susan Davidson, H. V. Jagadish, Z. Meral Ozsoyoglu, S. Sudarshan and Gerhard Weikum. Together, we formulated a few principles to help guide PVLDB policy. A useful analogy is that principles should be concise and stable, akin to the constitution, while policies can be verbose and evolve over time, akin to laws.

Our first guiding principle is to develop policies that simultaneously increase PVLDB acceptance rates, improve the quality of PVLDB published papers, and improve the quality of PVLDB reviews. Policies such as year-around paper submissions and journal-quality reviewing have helped make the magic of this seemingly impossible principle possible.

Our second guiding principle is to minimize work for all to the extent possible, while satisfying our first principle; where tradeoffs are required, favor reviewers over associate editors, and associate editors over the Editors in Chief. Recent PVLDB policies such as workload reduction for reviewers are constructive instances of our second principle.

Our third, and final, guiding principle is to encourage innovations in submission, reviewing, and publication processes to enable the first two principles. If an innovation is successful, make it a policy and apply it consistently across years; if it is not successful, we should fail fast. Recent PVLDB innovations include soliciting Innovative Systems papers, and Experiments & Analysis track mini papers.

We hope that our principles and policies will provide an evolutionary path forward for PVLDB, one that ensures stability across years, while enabling the policy innovations that have made PVLDB successful.

---

Divesh Srivastava, AT&T Labs–Research, USA  
PVLDB Managing Editor