

# VLDB2014

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



## Proceedings of the VLDB Endowment

Volume 7, No. 14 – October 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. 14, October 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 14, October 2014: VLDB 2014

Pages ii - x and 1785 - 2028

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 .....	v

### Letters

From The Proceedings Co-Chairs .....	<i>Li Xiong, Cong Yu</i> x
--------------------------------------	----------------------------

### Research Papers

Show Me the Money: Dynamic Recommendations for Revenue Maximization .....	1785
..... <i>Wei Lu, Shanshan Chen, Keqian Li, Laks V.S. Lakshmanan</i>	
ScalaGiST: Scalable Generalized Search Trees for MapReduce Systems .....	1797
[Innovative Systems Paper] .....	
..... <i>Peng Lu, Gang Chen, Beng Chin Ooi, Hoang Tam Vo, Sai Wu</i>	
Finding Patterns in a Knowledge Base using Keywords to Compose Table Answers .....	1809
..... <i>Mohan Yang, Bolin Ding, Surajit Chaudhuri, Kaushik Chakrabarti</i>	
Pregel Algorithms for Graph Connectivity Problems with Performance Guarantees .....	1821
..... <i>Da Yan, James Cheng, Kai Xing, Yi Lu, Wilfred Ng, Yingyi Bu</i>	
Auto-Approximation of Graph Computing .....	1833
..... <i>Zechao Shang, Jeffrey Xu Yu</i>	
DIADEM: Thousands of Websites to a Single Database .....	1845
..... <i>Tim Furche, Georg Gottlob,</i>	
..... <i>Giovanni Grasso, Xiaonan Guo, Giorgio Orsi, Christian Schallhart, Cheng Wang</i>	
Uncertainty Aware Query Execution Time Prediction .....	1857
..... <i>Wentao Wu, Xi Wu, Hakan Hacigumus, Jeffrey F. Naughton</i>	
Optimizing the Chase: Scalable Data Integration under Constraints .....	1869
..... <i>George Konstantinidis, Jose Luis Ambite</i>	
BF-Tree: Approximate Tree Indexing .....	1881
..... <i>Manos Athanassoulis, Anastasia Ailamaki</i>	
ADDICT: Advanced Instruction Chasing for Transactions .....	1893
..... <i>Pinar Tozun, Islam Atta, Anastasia Ailamaki, Andreas Moshovos</i>	

AsterixDB: A Scalable, Open Source BDMS.....	1905
..... <i>Sattam Alsubaiee, Yasser Altowim, Hotham Altwaijry, Alexander Behm, Vinayak Borkar,</i>	
..... <i>Yingyi Bu, Michael Carey, Inci Cetindil, Madhusudan Cheelangi, Khurram Faraaz,</i>	
..... <i>Eugenia Gabrielova, Raman Grover, Zachary Heilbron, Young-Seok Kim, Chen Li,</i>	
..... <i>Guangqiang Li, Ji Mahn Ok, Nicola Onose, Pouria Pirzadeh, Vassilis Tsotras,</i>	
..... <i>Rares Vernica, Jian Wen, Till Westmann</i>	
LogGP: A Log-based Dynamic Graph Partitioning Method.....	1917
..... <i>Ning Xu, Lei Chen, Bin Cui</i>	
Supervised Meta-blocking .....	1929
..... <i>George Papadakis, George Papastefanatos, Georgia Koutrika</i>	
Generating Top-k Packages via Preference Elicitation .....	1941
..... <i>Min Xie, Laks V.S. Lakshmanan, Peter T. Wood</i>	
Fast Range Query Processing with Strong Privacy Protection for Cloud Computing .....	1953
..... <i>Rui Li, Alex X. Liu, Ann L. Wang, Bezawada Bruhadeshwar</i>	
Finish Them!: Pricing Algorithms for Human Computation.....	1965
..... <i>Yihan Gao, Aditya Parameswaran</i>	
TransactiveDB: Tapping into Collective Human Memories .....	1977
..... <i>Michele Catasta, Alberto Tonon, Djellel Eddine Difallah,</i>	
..... <i>Gianluca Demartini, Karl Aberer, and Philippe Cudre-Mauroux</i>	
Blogel: A Block-Centric Framework for Distributed Computation on Real-World Graphs.....	1981
..... <i>Da Yan, James Cheng, Yi Lu, Wilfred Ng</i>	
Efficient Identification of Implicit Facts in Incomplete OWL2-EL Knowledge Bases.....	1993
..... <i>John Liagouris, Manolis Terrovitis</i>	
Where To: Crowd-Aided Path Selection.....	2005
..... <i>Chen Jason Zhang, Yongxin Tong, Lei Chen</i>	
Large Scale Real-time Ridesharing with Service Guarantee on Road Networks.....	2017
..... <i>Yan Huang, Favyen Bastani, Ruoming Jin, Xiaoyang Sean Wang</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

### **Workshop Chairs**

Anastasia Ailamaki, EPFL  
Kaushik Chakrabarti, Microsoft

### **Panel Chairs**

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

Brian Cooper, Google, USA  
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, Università 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 Ozsü, 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

Yunyao 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

## FROM THE PROCEEDINGS CO-CHAIRS

When we were asked to become Proceedings Co-Chairs for PVLDB Volume 7 (a.k.a., VLDB 2014), we knew we shouldn't have taken the job: the notion of handling a non-trivial publication process each month for one whole year (and more) was not appealing at all! But we did not want to refuse. Jag was the driving force behind PVLDB and we suspected he had further innovations in mind this time: we wanted to be the first to know about them! (We were happily rewarded when we were told about the session format reform before most others.) And we couldn't refuse: Jag is the Ph.D. advisor to one of us and that means you really have no choice. One day in March 2013, we came on board.

The workload is every bit as advertised. First, there is the task of producing the initial templates for both the foreword (i.e., this document) and the camera ready copy (i.e., the actual papers). Then, *each month*, the laundry list includes: 1) update the templates to reflect the new issue; 2) communicate with and obtain the camera ready papers from authors of accepted papers in that month; 3) insert additional format bits to the camera ready copies; and 4) generate the table of contents and publish the proceedings on the PVLDB website. Finally, there is the task of publishing the full proceedings in various citation tracking systems such as DBLP.

However, we get to read all the interesting papers and learn the latest trends more than a month before anyone else except the Program Chairs, and we happily consider that as enough compensation! Another equally rewarding experience is to see the excitement of first time student authors and feel the future of database research is in good shape.

We do notice some common issues our experiences. Here are two simple pieces of advice that an author can follow to help out future Proceedings Chairs:

- Follow the instructions carefully and do not assume what has worked before will still work. Some common mistakes include naming the camera ready copy files wrongly, using an old template that has the wrong issue number or publication date, forgetting to remove page numbers, etc. Each such mistake requires us to handle the paper outside of the normal process and significantly adds to the workload.
- Do not attempt to change the metadata about the paper (i.e., paper title and authors) silently. This can easily leave the official paper in an inconsistent state that is very hard to detect. In the near future, such changes will be handled through an official errata process that is being developed, and anyone who is interested in learning about the process early can contact S. Sudarshan and Divesh Srivastava.

Finally, hearty congratulations to the VLDB 2014 organization team led by Chun Chen and Sharad Mehrotra, for a well-organized conference at Hangzhou. We hope all of you who were there had as good a time as we did!

---

Li Xiong, Emory University, Atlanta, GA, USA  
Cong Yu, Google Research, New York, NY, USA  
Proceedings Co-Chairs, VLDB 2014