VLDB2014
40th International Conference on Very Large Data Bases, Hangzhou, China

Proceedings of the VLDB Endowment

Volume 7, No. 12 – August 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. 12, August 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.

Volume 7, Number 12, August 2014: VLDB 2014

Pages ii - x and 1023 - 1318

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 v

Letters

Real Session Chairs Do More than Just Keep Time ..... *H. V. Jagadish* x

Research Papers

Computing Personalized PageRank Quickly by Exploiting Graph Structures 1023
..... *Takanori Maehara, Takuya Akiba, Yoichi Iwata, Ken-ichi Kawarabayashi*

Accordion: Elastic Scalability for Database Systems Supporting Distributed Transactions 1035
..... *Marco Serafini, Essam Mansour, Ashraf Aboulnaga*
..... *Kenneth Salem, Taha Rafiq, Umar Farooq Minhas*

An Experimental Comparison of Pregel-like Graph Processing Systems 1047
.... *Minyang Han, Khuzaima Daudjee, Khaled Ammar, M. Tamer Özsu, Xingfang Wang, Tianqi Jin*

ClusterJoin: A Similarity Joins Framework using Map-Reduce 1059
.....*Akash Das Sarma, Yeye He, Surajit Chaudhuri*

Crowdsourcing Algorithms for Entity Resolution 1071
..... *Norases Vesdapunt, Kedar Bellare, Nilesh Dalvi*

Distributed Graph Simulation: Impossibility and Possibility 1083
..... *Wenfei Fan, Xin Wang, Yinghui Wu, Dong Deng*

Code Generation for Efficient Query Processing in Managed Runtimes 1095
..... *Fabian Nagel, Gavin Bierman, Stratis D. Viglas*

Aggregate Estimation Over Dynamic Hidden Web Databases 1107
..... *Weimo Liu, Saravanan Thirumuruganathan, Nan Zhang, Gautam Das*

Adaptive Query Processing on RAW Data 1119
..... *Manos Karpathiotakis, Miguel Branco, Ioannis Alagiannis, Anastasia Ailamaki*

Storing and Querying Tree-Structured Records in Dremel 1131
..... *Foto N. Afrati, Dan Delorey, Mosha Pasumansky, Jeffrey D. Ullman*

Similarity Search for Scientific Workflows 1143
..... *Johannes Starlinger, Bryan Brancotte, Sarah Cohen-Boulakia, Ulf Leser*

Differentially Private Event Sequences over Infinite Streams 1155
..... *Georgios Kellaris, Stavros Papadopoulos, Xiaokui Xiao, Dimitris Papadias*

Matching Titles with Cross Title Web-Search Enrichment and Community Detection 1167
.....*Nikhil Londhe, Vishrawas Gopalakrishnan, Aidong Zhang, Hung Q. Ngo, Rohini Srihari*

On Concise Set of Relative Candidate Keys 1179
.....*Shaoxu Song, Lei Chen, Hong Cheng*

Reachability Querying: An Independent Permutation Labeling Approach 1191
.....*Hao Wei, Jeffrey Xu Yu, Can Lu, Ruoming Jin*

Hop Doubling Label Indexing for Point-to-Point Distance Querying on Scale-Free Networks 1203
.....*Minhao Jiang, Ada Wai-Chee Fu, Raymond Chi-Wing Wong, Yanyan Xu*

Semantic Culturomics (vision paper)..... 1215
..... *Fabian M. Suchanek, Nicoleta Preda*

Benchmarking Scalability and Elasticity of Distributed Database Systems 1219
.....*Jörn Kuhlenkamp, Markus Klems, Oliver Röss*

Bounded Conjunctive Queries 1231
..... *Yang Cao, Wenfei Fan, Tianyu Wo, Wenyuan Yu*

Optimizing Join Enumeration in Transformation-based Query Optimizers 1243
.....*Anil Shanbhag, S. Sudarshan*

A System for Management and Analysis of Preference Data 1255
.....*Marie Jacob, Benny Kimelfeld, Julia Stoyanovich*

Mesa: Geo-Replicated, Near Real-Time, Scalable Data Warehousing..... 1259
.....*Ashish Gupta, Fan Yang, Jason Govig, Adam Kirsch, Kelvin Chan*
..... *Kevin Lai, Shuo Wu, Sandeep Govind Dhoot, Abhilash Rajesh Kumar, Ankur Agiwal*
.....*Sanjay Bhansali, Mingsheng Hong, Jamie Cameron, Masood Siddiqi, David Jones*
.....*Jeff Shute, Andrey Gubarev, Shivakumar Venkataraman, Divyakant Agrawal*

An Effective Encoding Scheme for Spatial RDF Data 1271
.....*John Liagouris, Nikos Mamoulis, Panagiotis Bouros, Manolis Terrovitis*

DimmWitted: A Study of Main-Memory Statistical Analytics 1283
..... *Ce Zhang, Christopher Re*

SQL-on-Hadoop: Full Circle Back to Shared-Nothing Database Architectures 1295
..... *Avrilia Floratou, Umar Farooq Minhas, Fatma Özcan*

Optimal Security-Aware Query Processing..... 1307
.....*Marco Guarnieri, David Basin*

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

Workshop Chairs

Anastasia Ailamaki, EPFL

Kaushik Chakrabarti, Microsoft

Panel Chairs

Hakan Hacigumus, NEC Labs

Jignesh Patel, University of Wisconsin

Xiaoyang Sean Wang, Fudan University

Research Track Review Board

Sibel Adali, Rensselear 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

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 Massachussets, Lowell

Buğra Gedik, Bilkent University

Rainer Gemulla, Max-Plack-Institut Saarbr點ken

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 Hadjielefhteriou, 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, Univesity 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

REAL SESSION CHAIRS DO MORE THAN JUST KEEP TIME

Paper sessions in conferences are often poorly attended. One frequently hears statements like "Why would I attend the talk when I can read the paper and get much more out of it for the time spent?" Sessions in themselves provide no added value: there is little discussion and no correlation between papers. Instead, each session is a sequence of unilateral presentations, of varying quality.

To address this difficulty, we are trying a system this year in which the session chair is a real "chair" and much more than just a time-keeper. Over time, we hope that being a session chair will be an honor, indicating recognition in the field as a senior researcher with a good broad perspective on the topic of the session.

So, what does a "real chair" do? This person introduces the topic of the session, and places the papers in the session in the context of other work on this topic. To do this, the session chair starts the session with a short "invited talk" providing the background and laying out the issues for the session, along the lines of a good moderator at a typical conference panel session. A "real chair" also facilitates discussion of the presented papers at the end of a session, again in the same manner as a good panel moderator.

One major constraint on the session chairs is that they do not get to choose the papers in their session – rather these papers are selected through a separate (PVLDB) review process and given to them. However, subject to this constraint, session chairs can choose the name of the session, and hence define the topic scope; choose the order in which papers are presented in their session, and choose the (initial) topics for discussion.

A complementary attempt to make sessions more lively is to reduce the time allotted per paper, and to reward good presentation. It is hard to maintain audience interest in a long presentation, and particularly when much of the audience is jet-lagged. By having a common introduction by the session chair, authors have to do less to present background for their work. Similarly, lengthy questions can be held off to the common discussion time at the end. We will have a total time of 14 minutes per paper this year, including 2 minutes for set up and short questions. So no speaker will be on for more than 12 minutes at a time. [I know there are some who would like more time per paper presented – but the ways to make that time have even less support -- to have more parallel sessions, to have more days in the conference, to have conference sessions go longer into the evening, or to invite fewer papers for presentation. On the other hand, there are some who believe a shorter presentation per paper is actually better, even if constraints did not force this].

To reward good presentations, we will empower session chairs to give out certificates for excellence in paper presentation. (We will separately work with session chairs to calibrate award guidelines and limit their number). Particularly for student presenters, who constitute the vast majority of conference presenters, such an award can be a valuable badge on the resume, providing evidence of their presentation skills.

H. V. Jagadish, University of Michigan, Ann Arbor, MI, USA
Program Co-Chair, VLDB 2014