

VLDB2012

38th International Conference on Very Large Data Bases, Istanbul, Turkey



Proceedings of the VLDB Endowment

Volume 5, No. 7 – March 2012

**Proceedings of the 38th International Conference on
Very Large Data Bases, Istanbul, Turkey**

Editor-in-Chief:

Z. Meral Özsoyoğlu

Associate Editors - Research Track:

Uğur Çetintemel, Nilesh Dalvi, Hank Korth, Anthony Tung

Associate Editors - Experiments and Analysis Track:

Gustavo Alonso, Juliana Freire

Proceedings Editors:

Ahmet Saçan, Nesime Tatbul

PVLDB – Proceedings of the VLDB Endowment

Volume 5, No. 7, March 2012.

The 38th International Conference on Very Large Data Bases, Istanbul, Turkey.

Copyright 2012 VLDB Endowment

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than VLDB Endowment must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists requires prior specific permission and/or a fee. Request permission to republish from PVLDB under email: info@vldb.org.

Volume 5, Number 7: VLDB 2012

Pages i - vii and 598 - 691

ISSN 2150-8097, March 2012.

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
PVLDB Review Board	iv

Letters

Letter from the Associate Editors	vii
..... <i>Gustavo Alonso, Juliana Freire</i>	

Research Track Papers

PerfXplain: Debugging MapReduce Job Performance	598
..... <i>Nodira Khoussainova, Magdalena Balazinska, Dan Suciu</i>	
Uncertain Centroid based Partitional Clustering of Uncertain Data	610
..... <i>Francesco Gullo, Andrea Tagarelli</i>	
Scalable K-Means++	622
..... <i>Bahman Bahmani, Benjamin Moseley, Andrea Vattani, Ravi Kumar, Sergei Vassilvitskii</i>	
Querying Schemas With Access Restrictions	634
..... <i>Michael Benedikt, Pierre Bourhis, Clemens Ley</i>	
Definition, Detection, and Recovery of Single-Page Failures, a Fourth Class of Database Failures	646
..... <i>Goetz Graefe, Harumi Kuno</i>	
Concurrency Control for Adaptive Indexing	656
..... <i>Goetz Graefe, Felix Halim, Stratos Idreos, Harumi Kuno, Stefan Manegold</i>	

Experiments and Analysis Track Papers

Comments on "Stack-based Algorithms for Pattern Matching on DAGs"	668
..... <i>Qiang Zeng, Zhuge Hai</i>	
An Analysis of Structured Data on the Web	680
..... <i>Nilesh Dalvi, Ashwin Machanavajjhala, Bo Pang</i>	

PVLDB REVIEW BOARD

VLDB 2012 General PC Chair

Z. Meral Özsoyoğlu, Case Western Reserve University, USA

VLDB 2012 Research Track PC Co-Chairs

Uğur Çetintemel, Brown University, USA

Nilesh Dalvi, Yahoo Research

Hank Korth, Lehigh University, USA

Anthony Tung, National Univ. of Singapore, Singapore

VLDB 2012 Experiments and Analysis Track PC Co-Chairs

Gustavo Alonso, ETH Zurich, Switzerland

Juliana Freire, NYU Poly, USA

VLDB 2012 Research Track Reviewers

Karl Aberer, EPFL, Switzerland

Ashraf Aboulnaga, University of Waterloo

Yanif Ahmad, Johns Hopkins University

Shivnath Babu, Duke University

Magdalena Balazinska, University of Washington

Elisa Bertino, Purdue University

Claudio Bettini, University of Milano

Sourav Bhowmick, Nayang Technological University

Angela Bonifati, Icar-CNR

Randal Burns, Johns Hopkins University

Michael Cafarella, Michigan

Kevin Chang, Univ. of Illinois at Urbana-Champaign

Gang Chen, Zhejiang University, China

Lei Chen, Hong Kong University of SaT, China

Reynold Cheng, University of Hong Kong

Jan Chomicki, University of Buffalo

Chris Clifton, Purdue University

Gao Cong, Nayang Technological University

Brian Cooper, Google, USA

Graham Cormode, AT&T

Bin Cui, Peking University

Carlo Curino, MIT

Anish Das Sharma, Yahoo

Antonios Deligiannakis, Technical University of Crete

Amr El Abbadi, UC Santa Barbara

Alan Fekete, University of Sydney

Hakan Ferhatosmanoğlu, Ohio State University

Alvaro Fernandes, University of Manchester

Johann-Christoph Freytag, Humboldt-Universität zu Berlin

Minos Garofalakis, TU Creta, Greece

Bugra Gedik, IBM Watson
Gabriel Ghinita, Purdue University
Goetz Graefe, HP
Sergio Greco, University of Calabria
Todd Green, University of California, Davis
Le Gruenwald, University of Oklahoma
Haryadi Gunawi, University of California, Berkeley
Rahul Gupta, IIT Bombay
Stavros Harizopoulos, HP
Howard Ho, IBM
Vagelis Hristidis, Florida International University
Meichun Hsu, HP
Jeong-Hyon Hwang, SUNY, Albany
Ihab Ilyas, University of Waterloo
Zachary Ives, University of Pennsylvania
Dean Jacobs, SAP
Hans-Arno Jacobsen, University of Toronto
Christian Jensen, Aalborg University
Chris Jermaine, Rice University
Ricardo Jumenez-Peris, Technical University of Madrid
Theodore Johnson, AT&T
Vana Kalogeraki, Athens U of Econ. and Business
Raghav Kaushik, Microsoft Research
Eamonn Keogh, University of California, Riverside
Martin Kersten, CWI
Benny Kimelfeld, IBM
George Kollios, Boston University
Flip Korn, AT&T Labs Research
Donald Kossmann, ETH Zurich
Laks V.S. Lakshmanan, University of British Columbia
Mong-Li Lee, National University of Singapore
Wang-Chien Lee, Penn State University
Cuiping Li, Renmin University of China
Guoliang Li, Tsinghua University, China
Lipyeow Lim, University of Hawaii at Manoa
David Lomet, Microsoft Research

Boon Thau Loo, University of Pennsylvania
Qiong Luo, Hong Kong University of SaT, China
Ashwin Machanavajjhala, Yahoo
Sanjay Madria, University of Missouri-Rolla
Stefan Manegold, CWI, Netherlands
Sharad Mehrotra, University of California, Irvine
Gerome Miklau, Univ. of Massachusetts, Amherst
Bongki Moon, University of Arizona
Kyriakos Mouratidis, Singapore Management Univ.
Gero Muhl, University of Rostock
Suman Nath, MSR
Sylvia Nittel, University of Maine
Dan Olteanu, Oxford
M. Tamer Özsu, University of Waterloo
Esther Pacitti, University of Montpellier
Olga Papaemmanouil, Brandeis University
Srinivasan Parthasarathy, The Ohio State University
Jian Pei, Simon Fraser University
Peter Pietzuch, Imperial College
Lucian Popa, IBM
Rachel Pottinger, University of British Columbia
Sunil Prabhakar, Purdue University
Vijayshankar Raman, IBM, USA
Vibhar Rastogi, Yahoo
Christopher Re, University of Wisconsin, Madison
Ken Ross, Columbia University
Elke Rundensteiner, Worcester Polytechnic Institute
Maria Sapino, University of Torino
Yücel Saygın, Sabanci University
Kyuseok Shim, Seoul National University
Lidan Shou, Zhejiang University, China
Adam Silberstein, Yahoo!, USA
Yannis Sismanis, IBM, USA
Richard Snodgrass, University of Arizona
Divesh Srivastava, AT&T
Dan Suciu, University of Washington

S Sudarshan, IIT Bombay
Egemen Tanin, University of Melbourne
Yufei Tao, Chinese University of Hong Kong, China
Martin Theobald, Max Planck Institute, Germany
Frank Tompa, University of Waterloo
Stratis Viglas, University of Edinburgh
Wei Wang, University of North Carolina

Gerhard Weikum, Max Planck Institute, Germany
Dong Xin, Google
Jiong Yang, Case Western Reserve University
Jun Yang, Duke
Ge Yu, Northeastern University, China
Aoying Zhou, East China Normal University, China

VLDB 2012 Experiments and Analysis Track Reviewers

Cristiana Amza, U. Toronto, Canada
Phil Bernstein, Microsoft Research, USA
Philippe Bonnet, IT U. of Copenhagen, Denmark
Michael Carey, UC Irvine, USA
Jens Dittrich, Uni. Saarland, Germany
Mary Fernandez, AT&T Research, USA
Bettina Kemme, McGill Uni., Canada
Whang Kyu-Young, KAIST, Korea
Samuel Madden, MIT, USA

Ioana Manolescu, INRIA, France
Dimitris Papadias, Hong Kong University of SaT, China
Hamid Pirahesh, IBM Almaden, USA
Maya Ramanath, IIT Delhi, India
Uwe Roehm, University of Sydney, Australia
Jens Teubner, ETH Zurich, Switzerland
Peter Triantafillou, U. Patras, Greece
Cong Yu, Google, USA

PVLDB Information Director

Gerald Weber, University of Auckland, New Zealand

VLDB 2012 Proceedings Co-Chairs

Ahmet Saçan, Drexel University, USA
Nesime Tatbul, ETH Zurich, Switzerland

Advisory Committee

Philip Bernstein, Michael Boehlen, Peter Buneman, Susan Davidson, Z. Meral Özsoyoğlu, S. Sudarshan, Gerhard Weikum

LETTER FROM THE ASSOCIATE EDITORS

On the occasion of this seventh issue of the Proceedings of the VLDB Endowment (PVLDB, vol. 5), we would like to take the opportunity to comment on the Experimental & Analysis Track we have had the honor to chair during this period.

The Experiments & Analysis Track is a pioneering effort to provide more recognition to papers that make strong, practical contributions to the database community. It was created in 2008 with the goal of providing a more adequate review process for papers covering experimental and systems work. This track also opens up a venue to valuable research that enriches the conference and fosters an ongoing scientific discussion on published work beyond the presentation at the conference. The E&A track, combined with the additional freedom and flexibility afforded by the PVLDB format, enables us to take a new approach to reviewing which we hope will benefit our research community not only by helping establish a complete record of its achievements but also by maintaining an active, ongoing scientific discourse on its results.

This issue contains six papers from the Research Track and two from the Experiments and Analysis Track. One of these latter papers is an excellent example of why a journal-style review process is helpful for identifying valuable contributions that do not fit the traditional model of research papers. The paper elaborates on results published at an earlier database conference and discusses both the validity of the complexity claims as well as the experimental results of the original paper, pointing out that these ideas had been used without questioning in a number of subsequent papers. The review process for such papers is long and involved, including multiple rounds of reviews by the program committee, feedback from the authors of the work being discussed, evaluation of the feedback, and detailed suggestions for authors to improve their paper. The PVLDB format makes this comprehensive review process possible.

We are thankful to the Program Committee of the Experiments and Analysis Track for their responsiveness and constant feedback. We would also like to thank the entire PVLDB editorial board for their continuous support throughout the review cycle, and to authors for working closely with reviewers to produce high-quality contributions.

The program for the VLDB 2012 conference, which includes all the papers that appear in this and previous volumes, is looking increasingly attractive and exciting. We look forward to seeing you all in Istanbul.

Gustavo Alonso, ETH Zürich, Switzerland
Juliana Freire, Polytechnic Institute of New York University, USA
Associate Editors, Proceedings of the VLDB Endowment, Volume 5
Experiments and Analysis Track PC Co-Chairs, VLDB 2012