VLDB2012
38th International Conference on Very Large Data Bases, Istanbul, Turkey

Proceedings of the VLDB Endowment

Volume 5, No. 4 – December 2011

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. 4, December 2011.

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

Copyright 2011 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 4: VLDB 2011

Pages ii - vii and 274 - 393

ISSN 2150-8097, December 2011.

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 Editor ..... *Nilesh Dalvi* vii

Research Track Papers

Towards Cost-Effective Storage Provisioning for DBMSs 274
..... *Ning Zhang, Junichi Tatemura, Jignesh M. Patel, Hakan Hacıgümüş*

B+-tree Index Optimization by Exploiting Internal Parallelism of Flash-based Solid State Drives . 286
..... *Hongchan Roh, Sanghyun Park, Sungho Kim, Mincheol Shin, Sang-Won Lee*

High-Performance Concurrency Control Mechanisms for Main-Memory Databases 298
Per-Åke Larson, Spyros Blanas, Cristian Diaconu, Craig Freedman, Jignesh M. Patel, Mike Zwilling

Capturing Topology in Graph Pattern Matching 310
..... *Shuai Ma, Yang Cao, Wenfei Fan, Jinpeng Huai, Tianyu Wo*

Probabilistic Management of OCR Data using an RDBMS 322
..... *Arun Kumar, Christopher Ré*

RTED: A Robust Algorithm for the Tree Edit Distance 334
..... *Mateusz Pawlik, Nikolaus Augsten*

Putting Lipstick on Pig: Enabling Database-style Workflow Provenance 346
Yael Amsterdamer, Susan B. Davidson, Daniel Deutch, Tova Milo, Julia Stoyanovich, Val Tannen

Relational Approach for Shortest Path Discovery over Large Graphs 358
..... *Jun Gao, Ruoming Jin, Jiashuai Zhou, Jeffrey Xu Yu, Xiao Jiang, Tengjiao Wang*

Mining Flipping Correlations from Large Datasets with Taxonomies 370
..... *Marina Barsky, Sangkyum Kim, Tim Weninger, Jiawei Han*

A Statistical Approach Towards Robust Progress Estimation 382
..... *Arnd Christian König, Bolin Ding, Surajit Chaudhuri, Vivek Narasayya*

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 EDITOR

It is my pleasure to present to you the fourth issue of the Proceedings of the VLDB Endowment (PVLDB), Volume 5, consisting of papers accepted through a year-round, journal-style review process, which will be presented at the VLDB 2012 Conference to be held in Istanbul, Turkey, August 27-31, 2012.

This issue continues to showcase the healthy pipeline of PVLDB with another set of 10 excellent papers. These papers span a broad range of topics in data management, which include storage provisioning, index optimization, concurrency control, graph pattern matching, management of OCR data, algorithms for tree-edit distances, provenance in map-reduce, graph algorithms in relational databases, mining correlations and query progress estimation.

Thanks again to all the authors, the reviewers, and the program committee chairs, whose hard work makes this innovative process possible.

Hope you enjoy these papers as much as we did. I am looking forward to seeing you all in Istanbul.

Nilesh Dalvi, Yahoo Research
Associate Editor, Proceedings of the VLDB Endowment, Volume 5
Research Track PC Co-Chair, VLDB 2012