

VLDB2012

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



Proceedings of the VLDB Endowment

Volume 5, No. 11 – July 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. 11, July 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 11: VLDB 2012

Pages i - x and 1088 - 1695

ISSN 2150-8097, July 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	vii

Letters

Letter from the Editor-in-Chief	x
..... <i>Z. Meral Özsoyoğlu</i>	

Research Track Papers

A Scalable Algorithm for Maximizing Range Sum in Spatial Databases	1088
..... <i>Dong-Wan Choi, Chin-Wan Chung, Yufei Tao</i>	
Spatial Queries with Two kNN Predicates	1100
..... <i>Ahmed M. Aly, Walid G. Aref, Mourad Ouzzani</i>	
Optimal Algorithms for Crawling a Hidden Database in the Web	1112
..... <i>Cheng Sheng, Nan Zhang, Yufei Tao, Xin Jin</i>	
Diversifying Top-K Results	1124
..... <i>Lu Qin, Jeffrey Xu Yu, Lijun Chang</i>	
Keyword-aware Optimal Route Search	1136
..... <i>Xin Cao, Lisi Chen, Gao Cong, Xiaokui Xiao</i>	
Answering Queries using Views over Probabilistic XML: Complexity and Tractability	1148
..... <i>Bogdan Cautis, Evgeny Kharlamov</i>	
Probabilistic Databases with MarkoViews	1160
..... <i>Abhay Jha, Dan Suciu</i>	
The Complexity of Social Coordination	1172
..... <i>Konstantinos Mamouras, Sigal Oren, Lior Seeman, Lucja Kot, Johannes Gehrke</i>	
Efficient Multi-way Theta-Join Processing Using MapReduce	1184
..... <i>Xiaofei Zhang, Lei Chen, Min Wang</i>	
Stubby: A Transformation-based Optimizer for MapReduce Workflows	1196
..... <i>Harold Lim, Herodotos Herodotou, Shivnath Babu</i>	
Labeling Workflow Views with Fine-Grained Dependencies	1208
..... <i>Zhuowei Bao, Susan B. Davidson, Tova Milo</i>	

Fundamentals of Order Dependencies	1220
..... <i>Jaroslav Szlichta, Parke Godfrey, Jarek Gryz</i>	
FDB: A Query Engine for Factorised Relational Databases	1232
..... <i>Nurzhan Bakibayev, Dan Olteanu, Jakub Závodný</i>	
Optimization of Analytic Window Functions	1244
..... <i>Yu Cao, Chee-Yong Chan, Jie Li, Kian-Lee Tan</i>	
Opening the Black Boxes in Data Flow Optimization	1256
<i>Fabian Hueske, Mathias Peters, Matthias Sax, Astrid Rheinländer, Rico Bergmann, Aljoscha Krettek, Kostas Tzoumas</i>	
Spinning Fast Iterative Data Flows	1268
..... <i>Stephan Ewen, Kostas Tzoumas, Moritz Kaufmann, Volker Markl</i>	
REX: Recursive, Delta-Based Data-Centric Computation	1280
..... <i>Svilen R. Mihaylov, Zachary G. Ives, Sudipto Guha</i>	
K-Reach: Who is in Your Small World	1292
..... <i>James Cheng, Zechao Shang, Hong Cheng, Haixun Wang, Jeffrey Xu Yu</i>	
Performance Guarantees for Distributed Reachability Queries	1304
..... <i>Wenfei Fan, Xin Wang, Yinghui Wu</i>	
Efficient Indexing and Querying over Syntactically Annotated Trees	1316
..... <i>Pirooz Chubak, Davood Rafiei</i>	
Queries with Guarded Negation	1328
..... <i>Vince Barany, Balder ten Cate, Martin Otto</i>	
PrivBasis: Frequent Itemset Mining with Differential Privacy	1340
..... <i>Ninghui Li, Wahbeh Qardaji, Dong Su, Jianneng Cao</i>	
Low-Rank Mechanism: Optimizing Batch Queries under Differential Privacy	1352
..... <i>Ganzhao Yuan, Zhenjie Zhang, Marianne Winslett, Xiaokui Xiao, Yin Yang, Zhifeng Hao</i>	
Functional Mechanism: Regression Analysis under Differential Privacy	1364
..... <i>Jun Zhang, Zhenjie Zhang, Xiaokui Xiao, Yin Yang, Marianne Winslett</i>	
Injecting Uncertainty in Graphs for Identity Obfuscation	1376
..... <i>Paolo Boldi, Francesco Bonchi, Aris Gionis, Tamir Tassa</i>	
Publishing Microdata with a Robust Privacy Guarantee	1388
..... <i>Jianneng Cao, Panagiotis Karras</i>	
Measuring Two-Event Structural Correlations on Graphs	1400
..... <i>Ziyu Guan, Xifeng Yan, Lance M. Kaplan</i>	
Ranking Large Temporal Data	1412
..... <i>Jeffrey Jestes, Jeff M. Phillips, Feifei Li, Mingwang Tang</i>	

Compacting Transactional Data in Hybrid OLTP & OLAP Databases	1424
..... <i>Florian Funke, Alfons Kemper, Thomas Neumann</i>	
Processing a Trillion Cells per Mouse Click	1436
..... <i>Alexander Hall, Olaf Bachmann, Robert Büssow, Silviu Gănceanu, Marc Nunkesser</i>	
OLTP on Hardware Islands	1447
..... <i>Danica Porobic, Ippokratis Pandis, Miguel Branco, Pinar Tözün, Anastasia Ailamaki</i>	
Serializability, not Serial: Concurrency Control and Availability in Multi-Datacenter Datastores	1459
..... <i>Stacy Patterson, Aaron J. Elmore, Faisal Nawab, Divyakant Agrawal, Amr El Abbadi</i>	
Automatic Partitioning of Database Applications	1471
..... <i>Alvin Cheung, Owen Arden, Samuel Madden, Andrew C. Myers</i>	
CrowDER: Crowdsourcing Entity Resolution	1483
..... <i>Jiannan Wang, Tim Kraska, Michael J. Franklin, Jianhua Feng</i>	
Whom to Ask? Jury Selection for Decision Making Tasks on Micro-blog Services	1495
..... <i>Caleb Chen CAO, Jieying She, Yongxin Tong, Lei Chen</i>	
ALAE: Accelerating Local Alignment with Affine Gap Exactly in Biosequence Databases	1507
..... <i>Xiaochun Yang, Honglei Liu, Bin Wang</i>	
sDTW: Computing DTW Distances using Locally Relevant Constraints based on Salient Feature Alignments	1519
..... <i>K. Selçuk Candan, Rosaria Rossini, Maria Luisa Sapino, Xiaolan Wang</i>	
SCOUT: Prefetching for Latent Feature Following Queries	1531
..... <i>Farhan Tauheed, Thomas Heinis, Felix Shürmann, Henry Markram, Anastasia Ailamaki</i>	
Accelerating Pathology Image Data Cross-Comparison on CPU-GPU Hybrid Systems	1543
..... <i>Kaibo Wang, Yin Huai, Rubao Lee, Fusheng Wang, Xiaodong Zhang, Joel H. Saltz</i>	
Robust Estimation of Resource Consumption for SQL Queries using Statistical Techniques	1555
..... <i>Jiexing Li, Arnd Christian König, Vivek Narasayya, Surajit Chaudhuri</i>	
Who Tags What? An Analysis Framework	1567
..... <i>Mahashweta Das, Saravanan Thirumuruganathan, Sihem Amer-Yahia, Gautam Das, Cong Yu</i>	
A Generic Framework for Efficient and Effective Subsequence Retrieval	1579
..... <i>Haohan Zhu, George Kollios, Vassilis Athitsos</i>	
Only Aggressive Elephants are Fast Elephants	1591
..... <i>Jens Dittrich, Jorge-Arnulfo Quiané-Ruiz, Stefan Richter, Stefan Schuh, Alekh Jindal, Jörg Schäd</i>	
Multiple Location Profiling for Users and Relationships from Social Network and Content	1603
..... <i>Rui Li, Shengjie Wang, Kevin Chen-Chuan Chang</i>	

Flash-based Extended Cache for Higher Throughput and Faster Recovery	1615
..... <i>Woon-Hak Kang, Sang-Won Lee, Bongki Moon</i>	
Don't Thrash: How to Cache Your Hash on Flash	1627
<i>Michael A. Bender, Martin Farach-Colton, Rob Johnson, Russell Kraner, Bradley C. Kuszmaul, Dzejl</i>	
<i>Medjedovic, Pablo Montes, Pradeep Shetty, Richard P. Spillane, Erez Zadok</i>	
Learning Expressive Linkage Rules using Genetic Programming	1638
..... <i>Robert Isele, Christian Bizer</i>	

Experiments and Analysis Track Papers

Mining Frequent Itemsets over Uncertain Databases	1650
..... <i>Yongxin Tong, Lei Chen, Yurong Cheng, Philip S. Yu</i>	
Uncertain Time-Series Similarity: Return to the Basics	1662
..... <i>Michele Dallachiesa, Besmira Nushi, Katsiaryna Mirylenka, Themis Palpanas</i>	
Statistical Distortion: Consequences of Data Cleaning	1674
..... <i>Tamraparni Dasu, Ji Meng Loh</i>	
Towards Energy-Efficient Database Cluster Design	1684
..... <i>Willis Lang, Stavros Harizopoulos, Jignesh M. Patel, Mehul A. Shah, Dimitris Tsirogiannis</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