

39th International Conference on Very Large Data Bases 2013

(VLDB 2013)

Proceedings of the VLDB Endowment Volume 6 No. 1-14

**Riva del Garda, Trento, Italy
26 – 30 August 2013**

Volume 1 of 3

ISBN: 978-1-63266-058-9

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2013) by the VLDB Endowment
All rights reserved.

Printed by Curran Associates, Inc. (2014)

For permission requests, please contact the VLDB Endowment
at the address below.

VLDB Endowment
Attn: Marianne Siroker
353 Serra Mall – Stanford University
Stanford CA 94305

Phone: (650) 723-0872
Fax: (650) 725-2588

siroker@cs.stanford.edu

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

Research Papers

Spatio-Textual Similarity Joins	1
..... <i>Panagiotis Bouros, Shen Ge, Nikos Mamoulis</i>	
DisC Diversity: Result Diversification based on Dissimilarity and Coverage.....	13
..... <i>Marina Drosou, Evaggelia Pitoura</i>	
On Differentially Private Frequent Itemset Mining.....	25
..... <i>Chen Zeng, Jeffrey F. Naughton, Jin-Yi Cai</i>	

Research Papers

Less is More: Selecting Sources Wisely for Integration.....	37
..... <i>Xin Luna Dong, Barna Saha, Divesh Srivastava</i>	
Distributed Time-aware Provenance.....	49
..... <i>Wenchao Zhou, Suyog Mapara, Yiqing Ren, Yang Li,</i> <i>Andreas Haeberlen, Zachary Ives, Boon Thau Loo, Micah Sherr</i>	
Query Processing under GLAV Mappings for Relational and Graph Databases.....	61
..... <i>Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Moshe Y. Vardi</i>	
Computing Immutable Regions for Subspace Top-k Queries	73
..... <i>Kyriakos Mouratidis, HweeHwa Pang</i>	
Large Scale Cohesive Subgraphs Discovery for Social Network Visual Analysis	85
..... <i>Feng Zhao, Anthony K. H. Tung</i>	
Truth Finding on the Deep Web: Is the Problem Solved?	97
..... <i>Xian Li, Xin Luna Dong, Kenneth Lyons, Weiyi Meng, Divesh Srivastava</i>	
Counting with the Crowd.....	109
..... <i>Adam Marcus, David Karger, Samuel Madden, Robert Miller, Sewoong Oh</i>	
ClouDiA: A Deployment Advisor for Public Clouds.....	121
..... <i>Tao Zou, Ronan Le Bras, Marcos Vaz Salles, Alan Demers, Johannes Gehrke</i>	
An In-depth Comparison of Subgraph Isomorphism Algorithms in Graph Databases	133
..... <i>Jinsoo Lee, Wook-Shin Han, Romans Kasperovics, Jeong-Hoon Lee</i>	
Lightweight Locking for Main Memory Database Systems.....	145
..... <i>Kun Ren, Alexander Thomson, Daniel J. Abadi</i>	

Research Papers

Lightweight Privacy-Preserving Peer-to-Peer Data Integration	157
..... <i>Ye Zhang, Wai-Kit Wong, S. M. Yiu, Nikos Mamoulis, David W. Cheung</i>	
Memory Efficient Minimum Substring Partitioning	169
..... <i>Yang Li, Pegah Kamousi, Fangqiu Han, Shengqi Yang, Xifeng Yan, Subhash Suri</i>	
NeMa: Fast Graph Search with Label Similarity.....	181
..... <i>Arijit Khan, Yinghui Wu, Charu C. Aggarwal, Xifeng Yan</i>	
PARAS: A Parameter Space Framework for Online Association Mining	193
..... <i>Xika Lin, Abhishek Mukherji, Elke A. Rundensteiner, Carolina Ruiz, Matthew O. Ward</i>	
Actively Soliciting Feedback for Query Answers in Keyword Search-Based Data Integration	205
..... <i>Zhepeng Yan, Nan Zheng, Zachary G. Ives, Partha Pratim Talukdar, Cong Yu</i>	
Spatial Keyword Query Processing: An Experimental Evaluation	217
..... <i>Lisi Chen, Gao Cong, Christian S. Jensen, Dingming Wu</i>	

Research Papers

Partitioning and Ranking Tagged Data Sources	229
..... <i>Milad Eftekhari, Nick Koudas</i>	
Efficient Implementation of Generalized Quantification in Relational Query Languages.....	241
..... <i>Bin Cao, Antonio Badia</i>	
DAX: A Widely Distributed Multi-tenant Storage Service for DBMS Hosting	253
..... <i>Rui Liu, Ashraf Aboulnaga, Kenneth Salem</i>	
A Distributed Graph Engine for Web Scale RDF Data	265
..... <i>Kai Zeng, Jiacheng Yang, Haixun Wang, Bin Shao, Zhongyuan Wang</i>	
Upper and Lower Bounds on the Cost of a Map-Reduce Computation	277
..... <i>Foto N. Afrati, Anish Das Sarma, Semih Salihoglu, Jeffrey D. Ullman</i>	

Research Papers

Processing Analytical Queries over Encrypted Data.....	289
..... <i>Stephen Tu, M. Frans Kaashoek, Samuel Madden, Nikolai Zeldovich</i>	
Practical Differential Privacy via Grouping and Smoothing	301
..... <i>Georgios Kellaris, Stavros Papadopoulos</i>	
On Scaling Up Sensitive Data Auditing.....	313
..... <i>Yupeng Fu, Raghav Kaushik, Ravishankar Ramamurthy</i>	
XORing Elephants: Novel Erasure Codes for Big Data	325
..... <i>Maheswaran Sathiamoorthy, Megasthenis Asteris, Dimitris Papailiopoulos,</i> <i>Alexandros G. Dimakis, Ramkumar Vadali, Scott Chen, Dhruva Borthakur</i>	
Scaling Factorization Machines to Relational Data.....	337
..... <i>Steffen Rendle</i>	

Research Papers

Question Selection for Crowd Entity Resolution	349
..... <i>Steven Euijong Whang, Peter Lofgren, Hector Garcia-Molina</i>	
A Comparison of Knives for Bread Slicing.....	361
..... <i>Alekh Jindal, Endre Palatinus, Vladimir Pavlov, Jens Dittrich</i>	
Efficient Error-tolerant Query Autocompletion	373
..... <i>Chuan Xiao, Jianbin Qin, Wei Wang, Yoshiharu Ishikawa, Koji Tsuda, Kunihiko Sadakane</i>	
Top-k Publish-Subscribe for Social Annotation of News.....	385
..... <i>Alexander Shraer, Maxim Gurevich, Marcus Fontoura, Vanja Josifovski</i>	
Efficient Querying of Inconsistent Databases with Binary Integer Programming.....	397
..... <i>Phokion G. Kolaitis, Enela Pema, Wang-Chiew Tan</i>	
Piggybacking on Social Networks	409
..... <i>Aristides Gionis, Flavio Junqueira, Vincent Leroy, Marco Serafini, Ingmar Weber</i>	
Schema Extraction for Tabular Data on the Web	421
..... <i>Marco D. Adelfio, Hanan Samet</i>	
Streaming Algorithms for k-core Decomposition	433
..... <i>Ahmet Erdem Sariyuce, Bugra Gedik, Gabriela Jacques-Silva,</i> <i>Kun-Lung Wu, Umit V. Catalyurek</i>	
Discovering Linkage Points over Web Data	445
..... <i>Oktie Hassanzadeh, Ken Q. Pu, Soheil Hassas Yeganeh, Renee J. Miller,</i> <i>Lucian Popa, Muricio A. Hernandez, Howard Ho</i>	
IS-LABEL: an Independent-Set based Labeling Scheme for Point-to-Point Distance Querying	457
..... <i>Ada Wai-Chee Fu, Huanhuan Wu, James Cheng, Raymond Chi-Wing Wong</i>	
Supporting User-Defined Functions on Uncertain Data.....	469
..... <i>Thanh T. L. Tran, Yanlei Diao, Charles Sutton, Anna Liu</i>	

Incremental and Accuracy-Aware Personalized PageRank through Scheduled Approximation.....	481
..... <i>Fanwei Zhu, Yuan Fang, Kevin Chen-Chuan Chang, Jing Ying</i>	

Research Papers

Efficient SimRank-based Similarity Join Over Large Graphs	493
..... <i>Weiguo Zheng, Lei Zou, Yansong Feng, Lei Chen, Dongyan Zhao</i>	
A Performance Study of Three Disk-based Structures for Indexing and Querying Frequent Itemsets.....	505
..... <i>Guimei Liu, Andre Suchitra, Limsoon Wong</i>	
TripleBit: a Fast and Compact System for Large Scale RDF Data	517
..... <i>Pingpeng Yuan, Pu Liu, Buwen Wu, Hai Jin, Wenya Zhang, Ling Liu</i>	
CorrectDB: SQL Engine with Practical Query Authentication	529
..... <i>Sumeet Bajaj, Radu Sion</i>	

Research Papers

Hybrid Storage Management for Database Systems.....	541
..... <i>Xin Liu, Kenneth Salem</i>	
Scorpion: Explaining Away Outliers in Aggregate Queries.....	553
..... <i>Eugene Wu, Samuel Madden</i>	
Ratio Threshold Queries over Distributed Data Sources	565
..... <i>Rajeev Gupta, Krithi Ramamritham, Mukesh Mohania</i>	
On the Complexity of Query Result Diversification	577
..... <i>Ting Deng, Wenfei Fan</i>	
Streaming Quotient Filter: A Near Optimal Approximate Duplicate Detection Approach for Data Streams	589
..... <i>Sourav Dutta, Ankur Narang, Suman K. Bera</i>	

Research Papers

On Repairing Structural Problems In Semi-structured Data	601
..... <i>Flip Korn, Barna Saha, Divesh Srivastava, Shanshan Ying</i>	
A Distributed Algorithm for Large-Scale Generalized Matching	613
..... <i>Faraz Makari Manshadi, Baruch Awerbuch, Rainer Gemulla,</i> <i>Rohit Khandekar, Julian Mestre, Mauro Sozio</i>	
The LLUNATIC Data-Cleaning Framework	625
..... <i>Floris Geerts, Giansalvatore Mecca, Paolo Papotti, Donatello Santoro</i>	
Sharing Data and Work Across Concurrent Analytical Queries.....	637
..... <i>Iraklis Psaroudakis, Manos Athanassoulis, Anastasia Ailamaki</i>	
Skyline Operator on Anti-correlated Distributions	649
..... <i>Haichuan Shang, Masaru Kitsuregawa</i>	
Low-Latency Multi-Datacenter Databases using Replicated Commit	661
..... <i>Hatem Mahmoud, Faisal Nawab, Alexander Pucher, Divyakant Agrawal, Amr El Abbadi</i>	
Distribution-Based Query Scheduling	673
..... <i>Yun Chi, Hakan Hacigumus, Wang-Pin Hsiung, Jeffrey F. Naughton</i>	
Making Queries Tractable on Big Data with Preprocessing	685
..... <i>Wenfei Fan, Floris Geerts, Frank Neven</i>	
Answering Planning Queries with the Crowd	697
..... <i>Haim Kaplan, Ilija Lotosh, Tova Milo, Slava Novgorodov</i>	
Hardware-Oblivious Parallelism for In-Memory Column-Stores.....	709
..... <i>Max Heimel, Michael Saecker, Holger Pirk, Stefan Manegold, Volker Markl</i>	
Permuting Data on Random-Access Block Storage.....	721
..... <i>Risi Thonangi, Jun Yang</i>	
Improving Flash Write Performance by Using Update Frequency	733
..... <i>Radu Stoica, Anastasia Ailamaki</i>	
Efficient Indexing for Diverse Query Results	745
..... <i>Lu Li, Chee-Yong Chan</i>	
Reducing Uncertainty of Schema Matching via Crowdsourcing.....	757
..... <i>Chen Jason Zhang, Lei Chen, H. V. Jagadish, Chen Caleb Cao</i>	
Travel Cost Inference from Sparse, Spatio-Temporally Correlated Time Series Using Markov Models	769
..... <i>Bin Yang, Chenjuan Guo, Christian S. Jensen</i>	

Research Papers

Query Optimization over Crowdsourced Data	781
..... <i>Hyunjung Park, Jennifer Widom</i>	
A Data-adaptive and Dynamic Segmentation Index for Whole Matching on Time Series	793
..... <i>Yang Wang, Peng Wang, Jian Pei, Wei Wang, Sheng Huang</i>	
Extraction and Integration of Partially Overlapping Web Sources	805
..... <i>Mirko Bronzi, Valter Crescenzi, Paolo Merialdo, Paolo Papotti</i>	
The Yin and Yang of Processing Data Warehousing Queries on GPU Devices.....	817
..... <i>Yuan Yuan, Rubao Lee, Xiaodong Zhang</i>	
Mining and Indexing Graphs for Supergraph Search	829
..... <i>Dayu Yuan, Prasenjit Mitra, C. Lee Giles</i>	
Efficient Recovery of Missing Events.....	841
..... <i>Jianmin Wang, Shaoxu Song, Xiaochen Zhu, Xuemin Lin</i>	
Hadoop's Adolescence.....	853
..... <i>Kai Ren, YongChul Kwon, Magdalena Balazinska, Bill Howe</i>	
RACE: A Scalable and Elastic Parallel System for Discovering Repeats in Very Long Sequences...	865
..... <i>Essam Mansour, Ahmed El-Roby, Panos Kalnis, Aron Ahmadi, Ashraf Aboulnaga</i>	
LLAMA: A Cache/Storage Subsystem for Modern Hardware.....	877
..... <i>Justin Levandoski, David Lomet, Sudipta Sengupta</i>	
Revisiting Co-Processing for Hash Joins on the Coupled CPU-GPU Architecture.....	889
..... <i>Jiong He, Mian Lu, Bingsheng He</i>	
Top-K Nearest Keyword Search on Large Graphs	901
..... <i>Miao Qiao, Lu Qin, Hong Cheng, Jeffrey Xu Yu, Wentao Tian</i>	
A General Framework for Geo-Social Query Processing.....	913
..... <i>Nikos Armenatzoglou, Stavros Papadopoulos, Dimitris Papadias</i>	
Towards Predicting Query Execution Time for Concurrent and Dynamic Database Workloads	925
..... <i>Wentao Wu, Yun Chi, Hakan Hacigumus, Jeffrey F. Naughton</i>	
Sketch-based Geometric Monitoring of Distributed Stream Queries.....	937
..... <i>Minos Garofalakis, Daniel Keren, Vasilis Samoladas</i>	
Direction-Preserving Trajectory Simplification.....	949
..... <i>Cheng Long, Raymond Chi-Wing Wong, Chenjuan Guo, H. V. Jagadish</i>	

Industrial and Applications Track Papers

Continuous Cloud-Scale Query Optimization and Processing	961
..... <i>Nicolas Bruno, Sapna Jain, Jingren Zhou</i>	
Optimization Strategies for A/B Testing on HADOOP.....	973
..... <i>Andrii Cherniak, Huma Zaidi, Vladimir Zadorozhny</i>	
Piranha: Optimizing Short Jobs in Hadoop	985
..... <i>Khaled Elmeleegy</i>	
Making Updates Disk-I/O Friendly Using SSDs.....	997
..... <i>Yuan Yuan, Rubao Lee, Xiaodong Zhang</i>	
Hadoop-GIS: A High Performance Spatial Data Warehousing System over MapReduce	1009
..... <i>Ablimit Aji, Fusheng Wang, Hoang Vo, Rubao Lee, Qiaoling Liu, Xiaodong Zhang, Joel Saltz</i>	
Statistics Collection in Oracle Spatial and Graph: Fast Histogram Construction for Complex Geometry Objects	1021
..... <i>Bhuvan Bamba, Siva Ravada, Ying Hu, Richard Anderson</i>	
MillWheel: Fault-Tolerant Stream Processing at Internet Scale	1033
..... <i>Tyler Akidau, Alex Balikov, Kaya Bekiroglu, Slava Chernyak,</i> <i>Josh Haberman, Reuven Lax, Sam McVeety, Daniel Mills, Paul Nordstrom, Sam Whittle</i>	
Online, Asynchronous Schema Change in F1.....	1045
..... <i>Ian Rae, Eric Rollins, Jeff Shute, Sukhdeep Sodhi, Radek Vingralek</i>	
Scuba: Diving into Data at Facebook.....	1057
..... <i>Lior Abraham, John Allen, Oleksandr Barykin, Vinayak Borkar,</i> <i>Bhuwan Chopra, Ciprian Gereca, Daniel Merl, Josh Metzler,</i> <i>David Reiss, Subbu Subramanian, Janet L. Wiener, Okay Zed</i>	
F1: A Distributed SQL Database That Scales	1068
..... <i>Jeff Shute, Radek Vingralek, Bart Samwel, Ben Handy, Chad Whipkey,</i> <i>Eric Rollins, Mircea Oancea, Kyle Littlefield, David Menestrina, Stephan Ellner,</i> <i>John Cieslewicz, Ian Rae, Traian Stancescu, Himani Apte</i>	

DB2 with BLU Acceleration: So Much More than Just a Column Store.....	1080
..... <i>Vijayshankar Raman, Gopi Attaluri, Ronald Barber, Naresh Chainani,</i>	
..... <i>David Kalmuk, Vincent KulandaiSamy, Jens Leenstra, Sam Lightstone,</i>	
..... <i>Shaorong Liu, Guy M. Lohman, Tim Malkemus, Rene Mueller, Ippokratis Pandis,</i>	
..... <i>Berni Schiefer, David Sharpe, Richard Sidle, Adam Storm, Liping Zhang</i>	
The Quantcast File System	1092
..... <i>Michael Ovsianikov, Silvius Rus, Damian Reeves, Paul Sutter, Sriram Rao, Jim Kelly</i>	
Adaptive and Big Data Scale Parallel Execution in Oracle	1102
..... <i>Srikanth Bellamkonda, Hua-Gang Li, Unmesh Jagtap, Yali Zhu, Vince Liang, Thierry Cruanes</i>	
WOO: A Scalable and Multi-tenant Platform for Continuous Knowledge Base Synthesis.....	1114
..... <i>Kedar Bellare, Carlo Curino, Ashwin Machanavajihala,</i>	
..... <i>Peter Mika, Mandar Rahurkar, Aamod Sane</i>	
Entity Extraction, Linking, Classification, and Tagging for Social Media: A Wikipedia-Based Approach	1126
..... <i>Abhishek Gattani, Digvijay S. Lamba, Nikesh Garera, Mitul Tiwari, Xiaoyong Chai,</i>	
..... <i>Sanjib Das, Sri Subramaniam, Anand Rajaraman, Venky Harinarayan, AnHai Doan</i>	
Overview of Turn Data Management Platform for Digital Advertising	1138
..... <i>Hazem Elmeleegy, Yinan Li, Yan Qi, Peter Wilmot,</i>	
..... <i>Mingxi Wu, Santanu Kolay, Ali Dasdan, Songting Chen</i>	
Unicorn: A System for Searching the Social Graph.....	1150
..... <i>Michael Curtiss, Iain Becker, Tudor Bosman, Sergey Doroshenko, Lucian Grijincu,</i>	
..... <i>Tom Jackson, Sandhya Kunnatur, Soren Lassen, Philip Pronin, Sriram Sankar,</i>	
..... <i>Guanghao Shen, Gintaras Woss, Chao Yang, Ning Zhang</i>	

Invited Industrial Papers

A New Service for Customer Care Based on the TrentoRise BigData Platform	1162
..... <i>Sergio Ramazzina, Chiara L. Ballari, Daniela Somenzi</i>	
Exploiting the Diversity, Mass and Speed of Territorial Data by TELCO Operator for Better User Services	1164
..... <i>Fabrizio Antonelli, Antonino Casella, Cristiana Chitic, Roberto Larcher, Giovanni Torrisi</i>	
The Trento Big Data Platform for Public Administration and Large Companies: Use cases and Opportunities.....	1166
..... <i>Ivan Bedini, Benedikt Elser, Yannis Velegrakis</i>	
Designing Query Optimizers for Big Data Problems of The Future.....	1168
..... <i>Nga Tran, Sreenath Bodagala, Jaimin Dave</i>	
How to maximize the value of big data with the open source SpagoBI suite through a comprehensive approach	1170
..... <i>Monica Franceschini</i>	

Context-Aware Computing: Opportunities and Open Issues	1172
..... <i>Edward Y. Chang</i>	
Next Generation Data Analytics at IBM Research.....	1174
..... <i>Oktie Hassanzadeh, Anastasios Kementsietsidis, Benny Kimelfeld,</i>	
..... <i>Rajasekar Krishnamurthy, Fatma Özcan, Ippokratis Pandis</i>	
Learning and Intelligent Optimization (LION): One Ring to Rule Them All.	1176
..... <i>Mauro Brunato, Roberto Battiti</i>	
Microsoft SQL Server’s Integrated Database Approach for Modern Applications and Hardware....	1178
..... <i>David Lomet</i>	
Odyssey: A Multi-Store System for Evolutionary Analytics	1180
..... <i>Hakan Hacigumus, Jagan Sankaranarayanan, Junichi Tatemura,</i>	
..... <i>Jeff LeFevre, Neoklis Polyzotis</i>	
A global Entity Name System (ENS) for data ecosystems	1182
..... <i>Paolo Bouquet, Andrea Molinari</i>	
SAP HANA: The Evolution from a Modern Main-Memory Data Platform to an Enterprise Application Platform.....	1184
..... <i>Vishal Sikka, Franz Farber, Anil Goel, Wolfgang Lehner</i>	
Keeping the TPC Relevant!	1186
..... <i>Raghunath Nambiar, Meikel Poess</i>	

Tutorial Papers

Big Data Integration	1188
..... <i>Xin Luna Dong, Divesh Srivastava</i>	
Just-in-time compilation for SQL query processing	1190
..... <i>Stratis D. Viglas</i>	
Toward Scalable Transaction Processing.....	1192
..... <i>Anastasia Ailamaki, Ryan Johnson, Ippokratis Pandis, Pinar Tozun</i>	
Towards Database Virtualization for Database as a Service	1194
..... <i>Aaron J. Elmore, Carlo Curino, Divyakant Agrawal, Amr El Abbadi</i>	
Mobility and Social Networking: A Data Management Perspective.....	1196
..... <i>Mohamed F. Mokbel, Mohamed Sarwat</i>	

Demonstration Track Papers

DesTeller: A System for Destination Prediction Based on Trajectories with Privacy Protection	1198
..... <i>Andy Yuan Xue, Rui Zhang, Yu Zheng, Xing Xie,</i>	
..... <i>Jianhui Yu, Yong Tang, Sapna Jain, Jingren Zhou</i>	
Senbazuru: A Prototype Spreadsheet Database Management System	1202
..... <i>Zhe Chen, Mike Cafarella, Jun Chen, Daniel Prevo, Junfeng Zhuang</i>	
ReqFlex: Fuzzy Queries for Everyone.....	1206
..... <i>Grégory Smits, Olivier Pivert, Thomas Girault</i>	
Comprehensive and Interactive Temporal Query Processing with SAP HANA	1210
..... <i>Martin Kaufmann, Panagiotis Vagenas, Peter Fischer, Donald Kossmann, Franz Färber</i>	
Functions Are Data Too (Defunctionalization for PL/SQL).....	1214
..... <i>Torsten Grust, Nils Schweinsberg, Alexander Ulrich</i>	
NADEEF: A Generalized Data Cleaning System	1218
..... <i>Amr Ebaid, Ahmed Elmagarmid, Ihab Ilyas, Mourad Ouzzani,</i>	
..... <i>Jorge-Arnulfo Quiane-Ruiz, Nan Tang, Si Yin</i>	
QUEST: A Keyword Search System for Relational Data based on Semantic and Machine Learning Techniques.....	1222
..... <i>Sonia Bergamaschi, Francesco Guerra, Matteo Interlandi,</i>	
..... <i>Raquel Trillo Lado, Yannis Velegrakis</i>	
GroupFinder: A New Approach to Top-K Point-of-Interest Group Retrieval	1226
..... <i>Kenneth Bøgh, Anders Skovsgaard, Christian Jensen</i>	
A Demonstration of SpatialHadoop: An Efficient MapReduce Framework for Spatial Data	1230
..... <i>Ahmed Eldawy, Mohamed Mokbel</i>	
Aggregate Profile Clustering for Telco Analytics.....	1234
..... <i>Mehmet Ali Abbasoğlu, Buğra Gedik, Hakan Ferhatosmanoglu</i>	
ROSeAnn: Reconciling Opinions of Semantic Annotators	1238
..... <i>Luying Chen, Stefano Ortona, Giorgio Orsi, Michael Benedikt</i>	

RecDB in Action: Recommendation Made Easy in Relational Databases.....	1242
..... <i>Mohamed Sarwat, James Avery, Mohamed Mokbel</i>	
POIKILO: A Tool for Evaluating the Results of Diversification Models and Algorithms	1246
..... <i>Marina Drosou, Evaggelia Pitoura</i>	
CrowdMiner: Mining association rules from the crowd	1250
..... <i>Yael Amsterdamer, Yael Grossman, Tova Milo, Pierre Senellart</i>	
TeRec: A Temporal Recommender System Over Tweet Stream	1254
..... <i>Chen Chen, Hongzhi Yin, Junjie Yao, Bin Cui</i>	
Graph Queries in a Next-Generation Datalog System	1258
..... <i>Alexander Shkapsky, KAI ZENG, Carlo Zaniolo</i>	
iRoad: A Framework For Scalable Predictive Query Processing On Road Networks	1262
..... <i>Abdeltawab Hendawi, Jie Bao, Mohamed Mokbel</i>	
SkySuite: A Framework of Skyline-Join Operators for Static and Stream Environments.....	1266
..... <i>Mithila Nagendra, K. selcuk Candan</i>	
Parallel Graph Processing on Graphics Processors Made Easy	1270
..... <i>Jianlong Zhong, Bingsheng He</i>	
Mosquito: Another One Bites the Data Upload Stream	1274
..... <i>Stefan Richter, Jens Dittrich, Stefan Schuh, Tobias Frey</i>	
NoFTL: Database Systems on FTL-less Flash Storage	1278
..... <i>Sergey Hardock, Iliia Petrov, Robert Gottstein, Alejandro Buchmann</i>	
SmartMonitor: Using Smart Devices to Perform Structural Health Monitoring.....	1282
..... <i>Dimitrios Kotsakos, Panos Sakkos, Vana Kalogeraki, Dimitrios Gunopulos</i>	
Lazy ETL in Action: ETL Technology Dates Scientific Data	1286
..... <i>Yağız Kargın, Milena Ivanova, Ying Zhang, Stefan Manegold, Martin Kersten</i>	
EagleTree: Exploring the Design Space of SSD-Based Algorithms	1290
..... <i>Niv Dayan, Martin Kjær Svendsen, Matias Bjørling, Philippe Bonnet, Luc Bouganim</i>	
EnviroMeter: A Platform for Querying Community-Sensed Data	1294
..... <i>Saket Sathe, Arthur Oviedo, Dipanjan Chakraborty, Karl Aberer</i>	
Scolopax: Exploratory Analysis of Scientific Data	1298
..... <i>Alper Okcan, Mirek Riedewald, Biswanath Panda, Daniel Fink</i>	
PROPOLIS: Provisioned Analysis of Data-Centric Processes	1302
..... <i>Daniel Deutch, Yuval Moskovitch, Val Tannen</i>	
Feature Selection in Enterprise Analytics: A Demonstration using an R-based Data Analytics System	1306
..... <i>Pradap Konda, Arun Kumar, Christopher Re, Vaishnavi Sashikanth</i>	

Flexible Query Processor on FPGAs	1310
..... <i>Mohammadreza Najafi, Mohammad Sadoghi, Hans-Arno Jacobsen</i>	
MASTRO STUDIO: Managing Ontology-Based Data Access applications	1314
..... <i>Cristina Civili, Marco Console, Giuseppe De Giacomo, Domenico Lembo,</i>	
..... <i>Maurizio Lenzerini, Lorenzo Lepore , Riccardo Mancini, Antonella Poggi,</i>	
..... <i>Riccardo Rosati, Marco Ruzzi, Valerio Santarelli, Domenico Fabio Savo</i>	
PLASMA-HD: Probing the LAttice Structure and MAKEup of High-dimensional Data	1318
..... <i>David Fuhry, Yang Zhang, Venu Satuluri, Arnab Nandi, Srinivasan Parthasarathy</i>	
A Demonstration of Iterative Parallel Array Processing in Support of Telescope Image Analysis ..	1322
..... <i>Matthew Moyers , Emad Soroush, Spencer Wallace, Simon Krughoff,</i>	
..... <i>Jake Vanderplas, Magdalena Balazinska, Andrew Connolly</i>	
EvenTweet: Online Localized Event Detection from Twitter	1326
..... <i>Hamed Abdelhaq, Christian Sengstock, Michael Gertz</i>	
IBminer: A Text Mining Tool for Constructing and Populating InfoBox Databases and Knowledge Bases.....	1330
..... <i>Hamid Mousavi, Shi Gao, Carlo Zaniolo</i>	
PAQO: A Preference-Aware Query Optimizer for PostgreSQL.....	1334
..... <i>Nicholas Farnan, Adam Lee, Panos Chrysanthis, Ting Yu</i>	
eSkyline: Processing Skyline Queries over Encrypted Data	1338
..... <i>Suvarna Bothe, Panagiotis Karras, Akrivi Vlachou</i>	
GestureQuery: A Multitouch Database Query Interface	1342
..... <i>Lilong Jiang, Michael Mandel, Arnab Nandi</i>	
Mining and Linking Patterns across Live Data Streams and Stream Archives.....	1346
..... <i>Di Yang, Kaiyu Zhao, Maryam Hasan, Hanyuan Lu, Elke Rundensteiner, Matthew Ward</i>	
PhotoStand: A Map Query Interface for a Database of News Photos.....	1350
..... <i>Hanan Samet, Marco Adelfio, Brenden Fruin,</i>	
..... <i>Michael Lieberman, Jagan Sankaranarayanan</i>	
Hone: "Scaling Down" Hadoop on Shared-Memory Systems.....	1354
..... <i>K. Ashwin Kumar, Jonathan Gluck, Amol Deshpande, Jimmy Lin</i>	
Ringtail: A Generalized Nowcasting System.....	1358
..... <i>Dolan Antenucci, Erdong Li, Shaobo Liu, Bochun Zhang, Mike Cafarella, Christopher Re</i>	
IPS: An Interactive Package Configuration System for Trip Planning	1362
..... <i>Min Xie, Laks V. S. Lakshmanan, Peter Wood</i>	
R2-D2: a System to Support Probabilistic Path Prediction in Dynamic Environments via "Semi-Lazy" Learning	1366
..... <i>Jingbo Zhou, Anthony Tung , Wei Wu, Wee Siong Ng</i>	

REEF: Retainable Evaluator Execution Framework	1370
..... <i>Byung-Gon Chun, Tyson Condie, Carlo Curino,</i>	
..... <i>Raghu Ramakrishnan, Russell Sears, Markus Weimer</i>	
OmniDB: Towards Portable and Efficient Query Processing on Parallel CPU/GPU Architectures ...	1374
..... <i>Shuhao Zhang, Jiong He, Bingsheng He, Mian Lu</i>	
Complete Approximations of Incomplete Queries	1378
..... <i>Ognjen Savkovic, Paramita Mirza, Alex Tomasi, Werner Nutt</i>	
User Analytics with UbeOne: Insights into Web Printing	1382
..... <i>Georgia Koutrika, Qian Lin, Jerry Liu</i>	
DiAI: Distributed Streaming Analytics Anywhere, Anytime	1386
..... <i>Ivo Santos, Marcel Tilly, Badrish Chandramouli, Jonathan Goldstein</i>	

Panel Papers

Big and Useful: What's in the Data for Me?	1390
..... <i>Rada Chirkova, Jun Yang</i>	

PhD Workshop Papers

Universal Indexing of Arbitrary Similarity Models	1392
..... <i>Tomas Bartos</i>	
Why it is time for a HyPE: A Hybrid Query Processing Engine for Efficient GPU Coprocessing in DBMS.....	1398
..... <i>Sebastian Bress</i>	
Database Support for Unstructured Meshes	1404
..... <i>Alireza Rezaei Mahdiraji</i>	
Domain Specific Multi-stage Query Language for Medical Document Repositories	1410
..... <i>Aastha Madaan</i>	
Realtime Analysis of Information Diffusion in Social Media	1416
..... <i>Io Taxidou</i>	
Mining Frequent Patterns with Differential Privacy	1422
..... <i>Luca Bonomi</i>	
Automatic ontology-based User Profile Learning from heterogeneous Web Resources in a Big Data Context.....	1428
..... <i>Anett Hoppe</i>	
Scalable Transactions across Heterogeneous NoSQL Key-Value Data Stores	1434
..... <i>Akon Dey</i>	
Getting Unique Solution in Data Exchange	1440
..... <i>Nhung Ngo</i>	

Storing and Processing Temporal Data in a Main Memory Column Store	1444
..... <i>Martin Kaufmann</i>	
Efficiency and Security in Similarity Cloud Services	1450
..... <i>Stepan Kozak</i>	
Fast Cartography for Data Explorers.....	1456
..... <i>Thibault Sellam</i>	

Research Papers

When Speed Has a Price: Fast Information Extraction Using Approximate Algorithms	1462
..... <i>Goncalo Simoes, Helena Galhardas, Luis Gravano</i>	
Design and Evaluation of Storage Organizations for Read-Optimized Main Memory Databases....	1474
..... <i>Craig Chasseur, Jignesh M. Patel</i>	
Aggregating Semantic Annotators	1486
..... <i>Luying Chen, Stefano Ortona, Giorgio Orsi, Michael Benedikt</i>	
Discovering Denial Constraints.....	1498
..... <i>Xu Chu, Ihab F. Ilyas, Paolo Papotti</i>	
Diversified Top-k Graph Pattern Matching.....	1510
..... <i>Wenfei Fan, Xin Wang, Yinghui Wu</i>	
Bitlist: New Full-text Index for Low Space Cost and Efficient Keyword Search.....	1522
..... <i>Weixiong Rao, Lei Chen, Pan Hui, Sasu Tarkoma</i>	
RCSI: Scalable similarity search in thousand(s) of genomes	1534
..... <i>Sebastian Wandelt, Johannes Starlinger, Marc Bux, Ulf Leser</i>	
Approximate MaxRS in Spatial Databases	1546
..... <i>Yufei Tao, Xiaocheng Hu, Dong-Wan Choi, Chin-Wan Chung</i>	
Multi-Tuple Deletion Propagation: Approximations and Complexity	1558
..... <i>Benny Kimelfeld, Jan Vondrak, David P. Woodruff</i>	
Supporting Distributed Feed-Following Apps over Edge Devices	1570
..... <i>Badrish Chandramouli, Suman Nath, Wenchao Zhou</i>	
Rank Discovery From Web Databases	1582
..... <i>Saravanan Thirumuruganathan, Nan Zhang, Gautam Das</i>	
SPARSI: Partitioning Sensitive Data amongst Multiple Adversaries	1594
..... <i>Theodoros Rekatsinas, Amol Deshpande, Ashwin Machanavajjhala</i>	

Scalable Column Concept Determination for Web Tables Using Large Knowledge Bases.....	1606
..... <i>Dong Deng, Yu Jiang, Guoliang Li, Jian Li, Cong Yu</i>	
Top-K Structural Diversity Search in Large Networks	1618
..... <i>Xin Huang, Hong Cheng, Rong-Hua Li, Lu Qin, Jeffrey Xu Yu</i>	
Synthetising Changes in XML Documents as PULs	1630
..... <i>Federico Cavalieri, Alessandro Solimando, Giovanna Guerrini</i>	

Research Papers

Probabilistic Query Rewriting for Efficient and Effective Keyword Search on Graph Data.....	1642
..... <i>Lei Zhang, Thanh Tran, Achim Rettinger</i>	
QuEval: Beyond high-dimensional indexing a la carte	1654
..... <i>Martin Schäler, Alexander Grebhahn, Reimar Schröter,</i> <i>Sandro Schulze, Veit Köppen, Gunter Saake</i>	
Discovering Longest-lasting Correlation in Sequence Databases	1666
..... <i>Yuhong Li, Leong Hou U, Man Lung Yiu, Zhiguo Gong</i>	
PREDICT: Towards Predicting the Runtime of Large Scale Iterative Analytics.....	1678
..... <i>Adrian Daniel Popescu, Andrey Balmin, Vuk Ercegovic, Anastasia Ailamaki</i>	
On the Embeddability of Random Walk Distances.....	1690
..... <i>Xiaohan Zhao, Adelbert Chang, Atish Das Sarma, Haitao Zheng, Ben Y. Zhao</i>	
Instant Loading for Main Memory Databases	1702
..... <i>Tobias Mühlbauer, Wolf Rödiger, Robert Seilbeck,</i> <i>Angelika Reiser, Alfons Kemper, Thomas Neumann</i>	
Adaptive Range Filters for Cold Data: Avoiding Trips to Siberia	1714
..... <i>Karolina Alexiou, Donald Kossmann, Per-Ake Larson</i>	
Scalable Progressive Analytics on Big Data in the Cloud	1726
..... <i>Badrish Chandramouli, Jonathan Goldstein, Abdul Quamar</i>	
Scalable XML Query Processing using Parallel Pushdown Transducers.....	1738
..... <i>Peter Ogden, David Thomas, Peter Pietzuch</i>	
Understanding Insights into the Basic Structure and Essential Issues of Table Placement Methods in Clusters	1750
..... <i>Yin Huai, Siyuan Ma, Rubao Lee, Owen O'Malley, Xiaodong Zhang</i>	
A Probabilistic Optimization Framework for the Empty-Answer Problem	1762
..... <i>Davide Mottin, Alice Marascu, Senjuti Basu Roy,</i> <i>Gautam Das, Themis Palpanas, Yannis Velegrakis</i>	

Summarizing Answer Graphs Induced by Keyword Queries	1774
..... <i>Yinghui Wu, Shengqi Yang, Mudhakar Srivatsa, Arun Iyengar, Xifeng Yan</i>	
Supporting Keyword Search in Product Database: A Probabilistic Approach	1786
..... <i>Huizhong Duan, ChengXiang Zhai, Jinxing Cheng, Abhishek Gattani</i>	
A Sampling Algebra for Aggregate Estimation	1798
..... <i>Supriya Nirkhiwale, Alin Dobra, Christopher Jermaine</i>	
A Temporal-Probabilistic Database Model for Information Extraction.....	1810
..... <i>Maximilian Dylla, Iris Miliaraki, Martin Theobald</i>	
Counter Strike: Generic Top-Down Join Enumeration for Hypergraphs	1822
..... <i>Pit Fender, Guido Moerkotte</i>	
Efficient Bulk Updates on Multiversion B-trees.....	1834
..... <i>Daniar Achakeev, Bernhard Seeger</i>	
Query-Driven Approach to Entity Resolution	1846
..... <i>Hotham Altwaijry, Dmitri V. Kalashnikov, Sharad Mehrotra</i>	
Expressiveness and Complexity of Order Dependencies	1858
..... <i>Jaroslav Szlichta, Parke Godfrey, Jarek Gryz, Calisto Zuzarte</i>	
Counting and Sampling Triangles from a Graph Stream	1870
..... <i>A. Pavan, Kanat Tangwongsan, Srikanta Tirthapura, Kun-Lung Wu</i>	
An Experimental Analysis of Iterated Spatial Joins in Main Memory.....	1882
..... <i>Benjamin Sowell, Marcos Vaz Salles, Tuan Cao, Alan Demers, Johannes Gehrke</i>	
Scaling Queries over Big RDF Graphs with Semantic Hash Partitioning	1894
..... <i>Kisung Lee, Ling Liu</i>	
Distributed SocialLite: A Datalog-Based Language for Large-Scale Graph Analysis	1906
..... <i>Jiwon Seo, Jongsoo Park, Jaeho Shin, Monica S. Lam</i>	
Horton+: A Distributed System for Processing Declarative Reachability Queries over Partitioned Graphs	1918
..... <i>Mohamed Sarwat, Sameh Elnikety, Yuxiong He, Mohamed F. Mokbel</i>	
Streaming Similarity Search over one Billion Tweets using Parallel Locality-Sensitive Hashing.....	1930
..... <i>Narayanan Sundaram, Aizana Turmukhametova, Nadathur Satish,</i> <i>Todd Mostak, Piotr Indyk, Samuel Madden, Pradeep Dubey</i>	
Anti-Caching: A New Approach to Database Management System Architecture.....	1942
..... <i>Justin DeBrabant, Andrew Pavlo, Stephen Tu, Michael Stonebraker, Stan Zdonik</i>	
Understanding Hierarchical Methods for Differentially Private Histograms.....	1954
..... <i>Wahbeh Qardaji, Weining Yang, Ninghui Li</i>	

Towards Social Data Platform: Automatic Topic-focused Monitor for Twitter Stream	1966
..... <i>Rui Li, Shengjie Wang, Kevin Chen-Chuan Chang</i>	
Simple, Fast, and Scalable Reachability Oracle	1978
..... <i>Ruoming Jin, Guan Wang</i>	
Aggregation and Ordering in Factorised Databases.....	1990
..... <i>Nurzhan Bakibayev, Tomas Kocisky, Dan Olteanu, Jakub Zavodny</i>	
Parallel Computation of Skyline and Reverse Skyline Queries Using MapReduce	2002
..... <i>Yoonjae Park, Jun-Ki Min, Kyuseok Shim</i>	
Fast Iterative Graph Computation with Block Updates	2014
..... <i>Wenlei Xie, Guozhang Wang, David Bindel, Alan Demers, Johannes Gehrke</i>	