

# **37th International Conference on Very Large Data Bases 2011**

**(VLDB 2011)**

**Proceedings of the VLDB Endowment Volume 4 No.1-12**

**Seattle, Washington, USA  
29 August - 3 September 2011**

**Volume 1 of 2**

**Editors:**

**H. V. Jagadish**

**ISBN: 978-1-61839-334-0  
ISSN: 2150-8097**

**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© (2011) by the VLDB Endowment  
All rights reserved.

Printed by Curran Associates, Inc. (2012)

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](mailto: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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# TABLE OF CONTENTS

## VOLUME 1

### PVLDB VOL. 4 NO. 1

#### RESEARCH PAPERS

<b>Generating Efficient Execution Plans for Vertically Partitioned XML Databases</b> .....	1
<i>Patrick Kling, M. Tamer Özsu, Khuzaima Daudjee</i>	
<b>A Generic Framework for Handling Uncertain Data with Local Correlations</b> .....	12
<i>Xiang Lian, Lei Chen</i>	
<b>SnipSuggest: Context-Aware Autocompletion for SQL</b> .....	22
<i>Nodira Khoussainova, Yongchul Kwon, Magdalena Balazinska, Dan Suciu</i>	
<b>The Complexity of Causality and Responsibility for Query Answers and non-Answers</b> .....	34
<i>Alexandra Meliou, Wolfgang Gatterbauer, Katherine F. Moore, Dan Suciu</i>	

### PVLDB VOL. 4 NO. 2

<b>Distributed Threshold Querying of General Functions by a Difference of Monotonic Representation</b> .....	46
<i>Guy Sagy, Daniel Keren, Izchak Sharfman, Assaf Schuster</i>	
<b>On Triangulation-based Dense Neighborhood Graphs Discovery</b> .....	58
<i>Nan Wang, Jingbo Zhang, Kian-Lee Tan, Anthony K. H. Tung</i>	
<b>Graph Indexing of Road Networks for Shortest Path Queries with Label Restrictions</b> .....	69
<i>Michael Rice, Vassilis J. Tsotras</i>	
<b>CRIUS: User-Friendly Database Design</b> .....	81
<i>Li Qian, Kristen Lefevre, H. V. Jagadish</i>	
<b>Efficient Processing of Top-k Spatial Preference Queries</b> .....	93
<i>João B. Rocha-Junior, Akrivi Vlachou, Christos Doulkeridis, Kjetil Nørkvåg</i>	
<b>HYRISE – A Main Memory Hybrid Storage Engine</b> .....	105
<i>Martin Grund, Jens Krüger, Hasso Plattner, Alexander Zeier, P. Cudre-Mauroux, Sam Madden</i>	
<b>Update Rewriting and Integrity Constraint Maintenance in a Schema Evolution Support System: PRISM++</b> .....	117
<i>Carlo A. Curino, Hyun Jin Moon, Alin Deutsch, Carlo Zaniolo</i>	
<b>SXPath - Extending XPath towards Spatial Querying on Web Documents</b> .....	129
<i>Ermelinda Oro, Massimo Ruffolo, Steffen Staab</i>	
<b>Personalized Privacy Protection in Social Networks</b> .....	141
<i>Mingxuan Yuan, Lei Chen, Philip S. Yu</i>	

### PVLDB VOL. 4 NO. 3

<b>A Probabilistic Approach for Automatically Filling Form-Based Web Interfaces</b> .....	151
<i>Guilherme A. Toda, Eli Cortez, Altigran S. Da Silva, Edleno De Moura</i>	
<b>Output URL Bidding</b> .....	161
<i>Panagiotis Papadimitriou, Hector Garcia-Molina, Ali Dasdan, Santanu Koley</i>	
<b>Fast Incremental and Personalized PageRank</b> .....	173
<i>Bahman Bahmani, Abdur Chowdhury, Ashish Goel</i>	
<b>QSkycube: Efficient Skycube Computation using Point-Based Space Partitioning</b> .....	185
<i>Jongwuk Lee, Seung-Won Hwang</i>	
<b>ZINC: Efficient Indexing for Skyline Computation</b> .....	197
<i>Bin Liu, Chee-Yong Chan</i>	

### PVLDB VOL. 4 NO. 4

<b>Large-Scale Collective Entity Matching</b> .....	208
<i>Vibhor Rastogi, Nilesh Dalvi, Minos Garofalakis</i>	
<b>Automatic Wrappers for Large Scale Web Extraction</b> .....	219
<i>Nilesh Dalvi, Ravi Kumar, Mohamed Soliman</i>	

<b>Fast Sparse Matrix-Vector Multiplication on GPUs: Implications for Graph Mining</b> .....	231
<i>Xintian Yang, Srinivasan Parthasarathy, P. Sadayappan</i>	
<b>Using Paxos to Build a Scalable, Consistent, and Highly Available Datastore</b> .....	243
<i>Jun Rao, Eugene J. Shekita, Sandeep Tata</i>	
<b>Fast Set Intersection in Memory</b> .....	255
<i>Bolin Ding, Arnd Christian König</i>	

### **PVLDB VOL. 4 NO. 5**

<b>Human-Assisted Graph Search: It's Okay to Ask Questions</b> .....	267
<i>Aditya Parameswaran, Anish Das Sarma, Hector Garcia-Molina, Neoklis Polyzotis, Jennifer Widom</i>	
<b>Guided Data Repair</b> .....	279
<i>Mohamed Yakout, Ahmed K. Elmagarmid, Jennifer Neville, Mourad Ouzzani, Ihab F. Ilyas</i>	
<b>Hyper-Local, Directions-Based Ranking of Places</b> .....	290
<i>Petros Venetis, Hector Gonzalez, Christian S. Jensen, Alon Halevy</i>	
<b>Incrementally Maintaining Classification using an RDBMS</b> .....	302
<i>M. Levent Koc, Christopher Ré</i>	
<b>High-Throughput Transaction Executions on Graphics Processors</b> .....	314
<i>Bingsheng Ye, Jeffrey Xu Yu</i>	
<b>Distributed Inference and Query Processing for RFID Tracking and Monitoring</b> .....	326
<i>Zhao Cao, Charles Sutton, Yanlei Diao, Prashant Shenoy</i>	

### **PVLDB VOL. 4 NO. 6**

<b>Similarity Join Size Estimation using Locality Sensitive Hashing</b> .....	338
<i>Hongrae Lee, Raymond T. Ng, Kyuseok Shim</i>	
<b>Query Expansion Based on Clustered Results</b> .....	350
<i>Zhiang Liu, Sivaramakrishnan Natarajan, Yi Chen</i>	
<b>CoPhy: A Scalable, Portable, and Interactive Index Advisor for Large Workloads</b> .....	362
<i>Debabrata Dash, Neoklis Polyzotis, Anastasia Ailamaki</i>	
<b>Tuffy: Scaling up Statistical Inference in Markov Logic Networks using an RDBMS</b> .....	373
<i>Feng Niu, Christopher Ré, Anhai Doan, Jude Shavlik</i>	
<b>Automatic Optimization for MapReduce Programs</b> .....	385
<i>Eaman Jahani, Michael J. Cafarella, Christopher Ré</i>	
<b>On Social-Temporal Group Query with Acquaintance Constraint</b> .....	397
<i>De-Nian Yang, Yi-Ling Chen, Wang-Chien Lee, Ming-Syan Chen</i>	

### **PVLDB VOL. 4 NO. 7**

<b>Synthesizing Products for Online Catalogs</b> .....	409
<i>Hoa Nguyen, Ariel Fuxman, Stelios Paparizos, Juliana Freire, Rakesh Agrawal</i>	
<b>Column-Oriented Storage Techniques for MapReduce</b> .....	419
<i>Avrilia Floratou, Jignesh M. Patel, Eugene J. Shekita, Sandeep Tata</i>	
<b>Implementing Performance Competitive Logical Recovery</b> .....	430
<i>David Lomet, Kostas Tzoumas, Michael Zwilling</i>	
<b>Personalized Social Recommendations - Accurate or Private?</b> .....	440
<i>Ashwin Machanavajjhala, Aleksandra Korolova, Atish Das Sarma</i>	
<b>Efficient Diversification of Web Search Results</b> .....	451
<i>Gabriele Capannini, Franco Maria Nardini, Raffaele Perego, Fabrizio Silvestri</i>	
<b>Social Content Matching in MapReduce</b> .....	460
<i>Gianmarco De Francisci Morales, Aristides Gionis, Mauro Sozio</i>	

### **PVLDB VOL. 4 NO. 8**

<b>Efficient Parallel Lists Intersection and Index Compression Algorithms using Graphics Processing Units</b> .....	470
<i>Naiyong Ao, Fan Zhang, Douglas S. Stones, Gang Wang, Xiaoguang Liu, Jing Liu, Sheng Lin</i>	
<b>gStore: Answering SPARQL Queries Via Subgraph Matching</b> .....	482
<i>Lei Zou, Jinghui Mo, Lei Chen, M. Tamer Ozsu, Dongyan Zhao</i>	

<b>Albatross: Lightweight Elasticity in Shared Storage Databases for the Cloud using Live Data Migration</b> .....	494
<i>Sudipto Das, Shoji Nishimura, Divyakant Agrawal, Amr El Abbadi</i>	
<b>An Incremental Hausdorff Distance Calculation Algorithm</b> .....	506
<i>Sarana Nutanong, Edwin H. Jacox, Hanan Samet</i>	
<b>Surrogate Parenthood: Protected and Informative Graphs</b> .....	518
<i>Barbara Blaustein, Adriane Chapman, Len Seligman, M. David Allen, Arnon Rosenthal</i>	

### **PVLDB VOL. 4 NO. 9**

<b>Recovering Semantics of Tables on the Web</b> .....	528
<i>Petros Venetis, Alon Halevy, Jayant Madhavan, Marius Pasca, Warren Shen, Fei Wu, Gengxin Miao, Chung Wu</i>	
<b>Efficiently Compiling Efficient Query Plans for Modern Hardware</b> .....	539
<i>Thomas Neumann</i>	
<b>Distance-Constraint Reachability Computation in Uncertain Graphs</b> .....	551
<i>Ruoming Jin, Lin Liu, Bolin Ding, Haixun Wang</i>	
<b>iCBS: Incremental Cost-based Scheduling under Piecewise Linear SLAs</b> .....	563
<i>Yun Chi, Hyun Jin Moon, Hakan Hacigumus</i>	
<b>CoHadoop: Flexible Data Placement and Its Exploitation in Hadoop</b> .....	575
<i>Mohamed Eltabakh, Yuanyuan Tian, Fatma Özcan, Rainer Gemulla, Aljoscha Krettek, John McPherson</i>	
<b>Merging What's Cracked, Cracking What's Merged: Adaptive Indexing in Main-Memory Column-Stores</b> .....	586
<i>Stratos Idreos, Stefan Manegold, Harumi Kuno, Goetz Graefe</i>	

### **PVLDB VOL. 4 NO. 10**

<b>On Pruning for Top-K Ranking in Uncertain Databases</b> .....	598
<i>Chonghai Wang, Li Yan Yuan, Jia-Huai You, Osmar R. Zaiane, Jian Pei</i>	
<b>PLP: Page Latch-free Shared-everything OLTP</b> .....	610
<i>Ippokratis Pandis, Pinar Tozun, Ryan Johnson, Anastasia Ailamaki</i>	
<b>Entity Matching: How Similar Is Similar</b> .....	622
<i>Jiannan Wang, Guoliang Li, Jeffrey Xu Yu, Jianhua Feng</i>	
<b>Active Complex Event Processing over Event Streams</b> .....	634
<i>Di Wang, Elke A. Rundensteiner, Richard T. Ellison III</i>	
<b>Structural Trend Analysis for Online Social Networks</b> .....	646
<i>Ceren Budak, Divyakant Agrawal, Amr El Abbadi</i>	
<b>Compression Aware Physical Database Design</b> .....	657
<i>Hideaki Kimura, Vivek Narasayya, Manoj Syamala</i>	
<b>Efficient Probabilistic Reverse Nearest Neighbor Query Processing on Uncertain Data</b> .....	669
<i>Thomas Bernecker, Tobias Emrich, Hans-Peter Kriegel, Matthias Renz, Stefan Zankl, Andreas Züfle</i>	
<b>Keyword Search in Graphs: Finding r-cliques</b> .....	681
<i>Mehdi Kargar, Aijun An</i>	

### **PVLDB VOL. 4 NO. 11**

#### **KEYNOTES**

<b>Towards a Global Brain</b> .....	693
<i>Tim O'Reilly</i>	
<b>Is It Still "Big Data" If It Fits In My Pocket?</b> .....	694
<i>David Campbell</i>	

#### **10-YEAR BEST PAPERS AWARD**

<b>Generic Schema Matching, Ten Years Later</b> .....	695
<i>Philip A. Bernstein, Jayant Madhavan, Erhard Rahm</i>	

## **RESEARCH PAPERS**

<b>A Framework for Supporting DBMS-like Indexes in the Cloud</b> .....	702
<i>Gang Chen, Hoang Tam Vo, Sai Wu, Beng Chin Ooi, M. Tamer Özsu</i>	
<b>On Link-based Similarity Join</b> .....	714
<i>Liwen Sun, Reynold Cheng, Xiang Li, David W. Cheung, Jiawei Han</i>	
<b>On Querying Historical Evolving Graph Sequences</b> .....	726
<i>Chenghui Ren, Eric Lo, Ben Kao, Xinjie Zhu, Reynold Cheng</i>	
<b>RemusDB: Transparent High Availability for Database Systems</b> .....	738
<i>Umar Minhas, Shriram Rajagopalan, Brendan Cully, Ashraf Aboulnaga, Kenneth Salem, Andrew Warfield</i>	

## **VOLUME 2**

<b>Completeness of Queries over Incomplete Databases</b> .....	749
<i>Simon Razniewski, Werner Nutt</i>	
<b>A Subsequence Matching with Gaps-Range-Tolerances Framework: Query-By-Humming Application</b> .....	761
<i>Alexios Kotsifakos, Panagiotis Papapetrou, Jaakko Hollmén, Dimitris Gunopulos</i>	
<b>Approximate Substring Matching over Uncertain Strings</b> .....	772
<i>Tingjian Ge, Zheng Li</i>	
<b>Serializable Snapshot Isolation for Replicated Databases in High-Update Scenarios</b> .....	783
<i>Hyungsoo Jung, Hyuck Han, Alan Fekete, Uwe Röhm</i>	
<b>PALM: Parallel Architecture-Friendly Latch-Free Modifications to B+ Trees on Many-Core Processors</b> .....	795
<i>Jason Sewall, Jatin Chhugani, Changkyu Kim, Nadathur Satish, Pradeep Dubey</i>	
<b>Mining Top-K Large Structural Patterns in a Massive Network</b> .....	807
<i>Feida Zhu, Qiang Qu, David Lo, Xifeng Yan, Jiawei Han, Philip S. Yu</i>	
<b>Structure-Aware Sampling: Flexible and Accurate Summarization</b> .....	N/A
<i>Edith Cohen, Graham Cormode, Nick Duffield</i>	
<b>Data Coordination: Supporting Contingent Updates</b> .....	831
<i>Michael Lawrence, Rachel Pottinger, Sheryl Staub-French</i>	
<b>Accelerating Queries with Group-By and Join by Groupjoin</b> .....	843
<i>Guido Moerkotte, Thomas Neumann</i>	
<b>Lightweight Graphical Models for Selectivity Estimation Without Independence Assumptions</b> .....	852
<i>Kostas Tzoumas, Amol Deshpande, Christian S. Jensen</i>	
<b>Dissemination of Models over Time-Varying Data</b> .....	864
<i>Yongluan Zhou, Zografoula Vagena, Jonas Haustad</i>	
<b>Efficient Subgraph Search over Large Uncertain Graphs</b> .....	876
<i>Ye Yuan, Guoren Wang, Haixun Wang, Lei Chen</i>	
<b>Entangled Transactions</b> .....	887
<i>Nitin Gupta, Milos Nikolic, Sudip Roy, Gabriel Bender, Lucja Kot, Johannes Gehrke, Christoph Koch</i>	
<b>Summary Graphs for Relational Database Schemas</b> .....	899
<i>Xiaoan Yang, Cecilia M. Procopiuc, Divesh Srivastava</i>	
<b>RecBench: Benchmarks for Evaluating Performance of Recommender System Architectures</b> .....	911
<i>Justin Levandoski, Michael Ekstrand, Michael Ludwig, Ahmed Eldawy, Mohamed Mokbel, John Riedl</i>	
<b>Business Policy Modeling and Enforcement in Databases</b> .....	921
<i>Ahmed A. Atallah, Frank W. Tompa</i>	
<b>Online Data Fusion</b> .....	932
<i>Xuan Liu, Xin Luna Dong, Beng Chin Ooi, Divesh Srivastava</i>	
<b>Optimistic Concurrency Control by Melding Trees</b> .....	944
<i>Philip A. Bernstein, Colin W. Reid, Ming Wu, Xinhao Yuan</i>	
<b>Linking Temporal Records</b> .....	956
<i>Pei Li, Xin Luna Dong, Andrea Maurino, Divesh Srivastava</i>	
<b>Efficient Algorithms for Finding Optimal Meeting Point on Road Networks</b> .....	968
<i>Da Yan, Zhou Zhao, Wilfred Ng</i>	
<b>Optimal Schemes for Robust Web Extraction</b> .....	980
<i>Aditya Parameswaran, Nilesh Dalvi, Hector Garcia-Molina, Rajeev Rastogi</i>	
<b>PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks</b> .....	992
<i>Yizhou Sun, Jiawei Han, Xifeng Yan, Philip Yu, Tianyi Wu</i>	
<b>Optimizing Query Answering under Ontological Constraints</b> .....	1004
<i>Giorgio Orsi, Andreas Pieris</i>	

<b>OXPath: A Language for Scalable, Memory-efficient Data Extraction from Web Applications</b> .....	1016
<i>Tim Furche, Georg Gottlob, Giovanni Grasso, Christian Schallhart, Andrew Sellers</i>	
<b>Optimizing and Parallelizing Ranked Enumeration</b> .....	1028
<i>Konstantin Golenberg, Benny Kimelfeld, Yehoshua Sagiv</i>	
<b>Where in the World is My Data?</b> .....	1040
<i>Sudarshan Kadambi, Jianjun Chen, Brian Cooper, David Lomax, Raghu Ramakrishnan, Adam Silberstein, Erwin Tam, Hector Garcia-Molina</i>	
<b>Queries with Difference on Probabilistic Databases</b> .....	1051
<i>Sanjeev Khanna, Sudeepa Roy, Val Tannen</i>	
<b>MRI: Meaningful Interpretations of Collaborative Ratings</b> .....	1063
<i>Mahashweta Das, Sihem Amer-Yahia, Gautam Das, Cong Yu</i>	
<b>Storing Matrices on Disk: Theory and Practice Revisited</b> .....	1075
<i>Yi Zhang, Kamesh Munagala, Jun Yang</i>	
<b>Publishing Set-Valued Data via Differential Privacy</b> .....	1087
<i>Rui Chen, Noman Mohammed, Benjamin C. M. Fung, Bipin C. Desai, Li Xiong</i>	
<b>Randomized Generalization for Aggregate Suppression Over Hidden Web Databases</b> .....	1099
<i>Xin Jin, Nan Zhang, Aditya Mone, Gautam Das</i>	
<b>Profiling, What-if Analysis, and Cost-based Optimization of MapReduce Programs</b> .....	1111
<i>Herodotos Herodotou, Shivnath Babu</i>	
<b>Scalable SPARQL Querying of Large RDF Graphs</b> .....	1123
<i>Jiewen Huang, Daniel J. Abadi, Kun Ren</i>	
<b>Online Aggregation for Large MapReduce Jobs</b> .....	1135
<i>Niketan Pansare, Vinayak Borkar, Chris Jermaine, Tyson Condie</i>	
<b>Private Analysis of Graph Structure</b> .....	1146
<i>Vishesh Karwa, Sofya Raskhodnikov, Adam Smith, Grigory Yaroslavtsev</i>	
<b>Stratification Criteria and Rewriting Techniques for Checking Chase Termination</b> .....	1158
<i>Sergio Greco, Francesca Spezzano, Irina Trubitsyna</i>	
<b>Optimizing Probabilistic Query Processing on Continuous Uncertain Data</b> .....	1169
<i>Liping Peng, Yanlei Diao, Anna Liu</i>	
<b>Massive Scale-out of Expensive Continuous Queries</b> .....	1181
<i>Erik Zeitler, Tore Risch</i>	
<b>Keyword Search on Form Results</b> .....	1189
<i>Aditya Ramesh, S. Sudarshan, Purva Joshi</i>	
<b>Efficient Rank Join with Aggregation Constraints</b> .....	1201
<i>Min Xie, Laks V. S. Lakshmanan, Peter T. Wood</i>	

## **PVLDB VOL. 4 NO. 12**

### **INDUSTRIAL, APPLICATIONS, AND EXPERIENCE TRACK PAPERS**

<b>Evaluation Strategies for Top-k Queries over Memory-Resident Inverted Indexes</b> .....	1213
<i>Marcus Fontoura, Vanja Josifovski, Jinhui Liu, Srihari Venkatesan, Xiangfei Zhu, Jason Zien</i>	
<b>Consistent Synchronization Schemes for Workload Replay</b> .....	1225
<i>Konstantinos Morfonios, Romain Colle, Leonidas Galanis, Supiti Buranawatanchoke, Benoît Dageville, Karl Dias, Yujun Wang</i>	
<b>Inspector Gadget: A Framework for Custom Monitoring and Debugging of Distributed Dataflows</b> .....	1237
<i>Christopher Olston, Benjamin Reed</i>	
<b>Online Expansion of Large-scale Data Warehouses</b> .....	1249
<i>Jeffrey Cohen, John Eshleman, Brian Hagenbuch, Joy Kent, Christopher Pedrotti, Gavin Sherry, Florian Waas</i>	
<b>HIWAS: Enabling Technology for Analysis of Clinical Data in XML Documents</b> .....	1260
<i>Joshua Hui, Sarah Knoop, Peter Schwarz</i>	
<b>Jaql: A Scripting Language for Large Scale Semi-Structured Data Analysis</b> .....	1272
<i>K. Beyer, V. Ercegovac, R. Gemulla, A. Balmin, M. Eltabakh, C.-C. Kanne, F. Özcan, E. Shekita</i>	
<b>Auto-Grouping Emails for Faster E-Discovery</b> .....	1284
<i>Sachindra Joshi, Danish Contractor, Kenney Ng, Prasad Deshpande, Thomas Hampp</i>	
<b>Web Scale Taxonomy Cleansing</b> .....	1295
<i>Taesung Lee, Zhongyuan Wang, Haixun Wang, Seung-Won Hwang</i>	
<b>Bridging Two Worlds with RICE</b> .....	1307
<i>Philipp Große, Wolfgang Lehner, Thomas Weichert, Franz Färber, Wen-Syan Li</i>	
<b>Tenzing – A SQL Implementation On The MapReduce Framework</b> .....	1318
<i>B. Chattopadhyay, L. Lin, W. Liu, S. Mittal, P. Aragona, V. Lychagina, Y. Kwon, M. Wong</i>	

<b>An Algebraic Approach for Data-Centric Scientific Workflows</b> .....	1328
<i>Eduardo Ogasawara, Jonas Dias, Daniel De Oliveira, Fábio Porto, Patrick Valduriez, Marta Mattoso</i>	
<b>Citrusleaf: A Real-Time NoSQL DB which Preserves ACID</b> .....	1340
<i>V. Srinivasan, Brian Bulkowski</i>	

## **DEMONSTRATION TRACK PAPERS**

<b>RAMP: A System for Capturing and Tracing Provenance in MapReduce Workflows</b> .....	1351
<i>Hyunjung Park, Robert Ikeda, Jennifer Widom</i>	
<b>BROAD: Diversified Keyword Search in Databases</b> .....	1355
<i>Feng Zhao, Xiaolong Zhan, Anthony Tung, Gang Chen</i>	
<b>TrustedDB: A Trusted Hardware based Outsourced Database Engine</b> .....	1359
<i>Sumeet Bajaj, Radu Sion</i>	
<b>IPL-P: In-Page Logging with PCRAM</b> .....	1363
<i>Kang-Nyeon Kim, Sang-Won Lee, Bongki Moon, Chanik Park, Joo-Young Hwang</i>	
<b>HyPer-sonic Combined Transaction AND Query Processing</b> .....	1367
<i>Florian Funke, Alfons Kemper, Thomas Neumann</i>	
<b>GrouPeer: A System for Clustering PDMSs</b> .....	1371
<i>Verena Kantere, Dimos Bousoumis, Timos Sellis</i>	
<b>CerFix: A System for Cleaning Data with Certain Fixes</b> .....	1375
<i>Wenfei Fan, Jianzhong Li, Shuai Ma, Nan Tang, Wenyuan Yu</i>	
<b>Online Visualization of Geospatial Stream Data using the WorldWide Telescope</b> .....	1379
<i>Mohamed Ali, Badrish Chandramouli, Jonathan Fay, Curtis Wong, Steven Drucker, Balan S. Raman</i>	
<b>Debugging Data Exchange with Vagabond</b> .....	1383
<i>Boris Glavic, Jiang Du, Renée J. Miller, Gustavo Alonso, Laura M. Haas</i>	
<b>CrowdDB: Query Processing with the VLDB Crowd</b> .....	1387
<i>Amber Feng, Michael Franklin, Donald Kossmann, Tim Kraska, Samuel Madden, Sukriti Ramesh, Andrew Wang, Reynold Xin</i>	
<b>Analytics for the Real-Time Web</b> .....	1391
<i>Maxim Grinev, Maria Grineva, Martin Hentschel, Donald Kossmann</i>	
<b>DivDB: A System for Diversifying Query Results</b> .....	1395
<i>Marcos R. Vieira, Humberto L. Razente, Maria C. Barioni, Marios Hadjieleftheriou, Divesh Srivastava, Caetano Traina Jr., Vassilis Tsotras</i>	
<b>HOMES: A Higher-Order Mapping Evaluation System</b> .....	1399
<i>Huy Vu, Michael Benedikt</i>	
<b>Proactive Detection and Repair of Data Corruption: Towards a Hassle-free Declarative Approach with Amulet</b> .....	1403
<i>Nedyalko Borisov, Shivnath Babu</i>	
<b>Automatic Workload Driven Index Defragmentation</b> .....	1407
<i>Vivek Narasayya, Hyunjung Park, Manoj Syamala</i>	
<b>Whodunit: An Auditing Tool for Detecting Data Breaches</b> .....	1410
<i>Raghav Kaushik, Ravi Ramamurthy</i>	
<b>EIRENE: Interactive Design and Refinement of Schema Mappings via Data Examples</b> .....	1414
<i>Bogdan Alexe, Balder Ten Cate, Phokion G. Kolaitis, Wang-Chiew Tan</i>	
<b>DataSynth: Generating Synthetic Data using Declarative Constraints</b> .....	1418
<i>Arvind Arasu, Raghav Kaushik, Jian Li</i>	
<b>InfoNetOLAPer: Integrating InfoNetWarehouse and InfoNetCube with InfoNetOLAP</b> .....	1422
<i>Chuan Li, Philip S. Yu, Lei Zhao, Yan Xie, Wangqun Lin</i>	
<b>From SPARQL to MapReduce: The Journey Using a Nested TripleGroup Algebra</b> .....	1426
<i>Hyeongsik Kim, Padmashree Ravindra, Kemafor Anyanwu</i>	
<b>FuDoCS: A Web Service Composition System Based on Fuzzy Dominance for Preference Query Answering</b> .....	1430
<i>Karim Benouaret, Djamel Benslimane, Allel Hadjali, Mahmoud Barhamgi</i>	
<b>A Demonstration of HYRISE – A Main Memory Hybrid Storage Engine</b> .....	1434
<i>Martin Grund, Philippe Cudre-Mauroux, Samuel Madden</i>	
<b>++Spicy: an Open-Source Tool for Second-Generation Schema Mapping and Data Exchange</b> .....	1438
<i>Bruno Marnette, Giansalvatore Mecca, Paolo Papotti, Salvatore Raunich, Donatello Santoro</i>	
<b>UpStream: A Storage-centric Load Management System for Real-time Update Streams</b> .....	1442
<i>Alexandru Moga, Nesime Tatbul</i>	
<b>MapReduce Programming and Cost-based Optimization? Crossing this Chasm with Starfish</b> .....	1446
<i>Herodotos Herodotou, Fei Dong, Shivnath Babu</i>	



<b>AIDA: An Online Tool for Accurate Disambiguation of Named Entities in Text and Tables</b> .....	1450
<i>Mohamed Amir Yosef, Johannes Hoffart, Ilaria Bordino, Marc Spaniol, Gerhard Weikum</i>	
<b>Microsoft Codename “Montego” – Data Import, Transformation, and Publication for Information Workers</b> .....	1454
<i>Stephen J. Maine, Lorenz Prem, Clemens Szyperski, James F. Terwilliger</i>	
<b>SocialSpamGuard: A Data Mining-Based Spam Detection System for Social Media Networks</b> .....	1458
<i>Xin Jin, Cindy Xide Lin, Jiebo Luo, Jiawei Han</i>	

## **CHALLENGE AND VISION TRACK PAPERS**

<b>Resiliency-Aware Data Management</b> .....	1462
<i>Matthias Böhm, Wolfgang Lehner, Christof Fetzer</i>	
<b>Guided Interaction: Rethinking the Query-Result Paradigm</b> .....	1466
<i>Arnab Nandi, H. V. Jagadish</i>	
<b>Data Generation for Application-Specific Benchmarking</b> .....	1470
<i>Y. C. Tay</i>	
<b>The Researcher's Guide to the Data Deluge: Querying a Scientific Database in Just a Few Seconds</b> .....	1474
<i>Martin L. Kersten, Stratos Idreos, Stefan Manegold, Erietta Liarou</i>	
<b>Anthropocentric Data Systems</b> .....	1478
<i>Peter Triantafillou</i>	
<b>Data Markets in the Cloud: An Opportunity for the Database Community</b> .....	1482
<i>Magdalena Balazinska, Bill Howe, Dan Suciu</i>	
<b>Data is Dead... Without What-if Models</b> .....	1486
<i>Peter Haas, Paul P. Maglio, Patricia G. Selinger, Wang-Chiew Tan</i>	
<b>Reverse Data Management</b> .....	1490
<i>Alexandra Meliou, Wolfgang Gatterbauer, Dan Suciu</i>	
<b>Exploring the Coming Repositories of Reproducible Experiments: Challenges and Opportunities</b> .....	1494
<i>Juliana Freire, Philippe Bonnet, Dennis Shasha</i>	
<b>Databases will Visualize Queries too</b> .....	1498
<i>Wolfgang Gatterbauer</i>	

## **TUTORIALS**

<b>New Frontiers in Business Intelligence</b> .....	1502
<i>Surajit Chaudhuri, Vivek Narasayya</i>	
<b>System Co-Design and Data Management for Flash Devices</b> .....	1504
<i>Philippe Bonnet, Luc Bouganim, Ioannis Koltsidas, Stratis D. Viglas</i>	
<b>Exploration of Deep Web Repositories</b> .....	1506
<i>Nan Zhang, Gautam Das</i>	
<b>Crowdsourcing Applications and Platforms: A Data Management Perspective</b> .....	1508
<i>Anhai Doan, Michael J. Franklin, Donald Kossmann, Tim Kraska</i>	
<b>Graph Data Management Systems for New Application Domains</b> .....	1510
<i>Philippe Cudré-Mauroux, Sameh Elnikety</i>	
<b>Information Diffusion in Social Networks: Observing and Influencing Societal Interests</b> .....	1512
<i>Divyakant Agrawal, Ceren Budak, Amr El Abbadi</i>	

## **PANELS**

<b>Data Management for Meeting Global Health Challenges</b> .....	1514
<i>Tapan S. Parikh, Kuang Chen</i>	
<b>Panel Discussion: Maximizing Impact</b> .....	1515
<i>Ed Lazowska</i>	
<b>Author Index</b>	