38th International Conference on Very Large Data Bases 2012

(VLDB 2012)

Proceedings of the VLDB Endowment
Volume 5 Numbers 1-12

Istanbul, Turkey
27-31 August 2012

Volume 1 of 3

Editors:

Z. Meral Özsoyoğlu

ISBN: 978-1-62276-758-8
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© (2012) by the VLDB Endowment
All rights reserved.

Printed by Curran Associates, Inc. (2013)

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

Volume 1

NUMBER 1

RESEARCH TRACK PAPERS

Explanation-Based Auditing 1
Daniel Fabbri, Kristen Lefevre

Human-powered Sorts and Joins 13
Adam Marcus, Eugene Wu, David Karger, Samuel Madden, Robert Miller

Verifying Computations with Streaming Interactive Proofs 25
Graham Cormode, Justin Thaler, Ke Yi

A MovingObject Index for Efficient Query Processing with Peer-Wise Location Privacy 37
Dan Lin, Christian S. Jensen, Rui Zhang, Lu Xiao, Jiaheng Lu

ERA: Efficient Serial and Parallel Suffix Tree Construction for Very Long Strings 49
Essam Mansour, Amin Allam, Spiros Skiadopoulos, Panos Kalnis

Fast Updates on Read-Optimized Databases Using Multi-Core CPUs 61
Jens Krueger, Changkyu Kim, Martin Grund, Nadathur Satish, David Schwalb, Jatin Chhugani, Hasso Plattner, Pradeep Dubey, Alexander Zeier

A Data-Based Approach to Social Influence Maximization 73
Amit Goyal, Francesco Bonchi, Laks V. S. Lakshmanan

NUMBER 2

RESEARCH TRACK PAPERS

On Predictive Modeling for Optimizing Transaction Execution in Parallel OLTP Systems 85
Andrew Pavlo, Evan P. C. Jones, Stanley Zdonik

View Selection in Semantic Web Databases 97
François Goasdoué, Konstantinos Karanasos, Julien Leblay, Ioana Manolescu

Building Wavelet Histograms on Large Data in MapReduce 109
Jeffrey Jestes, Ke Yi, Feifei Li

Summarization and Matching of Density-Based Clusters in Streaming Environments 121
Di Yang, Elke A. Rundensteiner, Matthew O. Ward

Multilingual Schema Matching for Wikipedia Infoboxes 133
Thanh Nguyen, Viviane Moreira, Huong Nguyen, Hoa Nguyen, Juliana Freire

EXPERIMENTS AND ANALYSIS TRACK PAPERS

Controlling False Positives in Association Rule Mining 145
Guimei Liu, Haojun Zhang, Limsoon Wong

NUMBER 3

RESEARCH TRACK PAPERS

PARIS: Probabilistic Alignment of Relations, Instances, and Schema 157
Fabian M. Suchanek, Serge Abiteboul, Pierre Senellart

Answering Top-k Queries Over a Mixture of Attractive and Repulsive Dimensions 169
Sayan Ranu, Ambuj K. Singh

PIQL: Success-Tolerant Query Processing in the Cloud 181
Michael Armbrust, Kristal Curtis, Tim Kraska, Armando Fox, Michael J. Franklin, David A. Patterson

gSketch: On Query Estimation in Graph Streams 193
Peixiang Zhao, Charu C. Aggarwal, Min Wang

Indexing the Earth Mover's Distance Using Normal Distributions 205
Brian E. Ruttenberg, Ambuj K. Singh
Generating Exact- and Ranked Partially-Matched Answers to Questions in Advertisements 217
Rani Qumsiyeh, Maria S. Pera, Yiu-Kai Ng
Size-l Object Summaries for Relational Keyword Search 229
Georgios J. Fakas, Zhi Cai, Nikos Mamoulis
REX: Explaining Relationships between Entity Pairs 241
Lujun Fang, Anish Das Sarma, Cong Yu, Philip Bohannon
PASS-JOIN: A Partition-based Method for Similarity Joins 253
Guoliang Li, Dong Deng, Jiannan Wang, Jianhua Feng
Relative Lempel-Ziv Factorization for Efficient Storage and Retrieval of Web Collections 265
Christopher Hoobin, Simon J. Puglisi, Justin Zobel

NUMBER 4

RESEARCH TRACK PAPERS

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

NUMBER 5

RESEARCH TRACK PAPERS

Relation Strength-Aware Clustering of Heterogeneous Information Networks with Incomplete Attributes 394
Yizhou Sun, Charu C. Aggarwal, Jiawei Han

EXPERIMENTS AND ANALYSIS TRACK PAPERS

Shortest Path and Distance Queries on Road Networks: An Experimental Evaluation 406
Lingkun Wu, Xiaokui Xiao, Dingxiong Deng, Gao Cong, Andy Diwen Zhu, Shuigeng Zhou

RESEARCH TRACK PAPERS

The Filter-Placement Problem and its Application to Minimizing Information Multiplicity 418
Dóra Erdös, Vatche Ishakian, Andrei Lapets, Evimaria Terzi, Azer Bestavros
Bayesian Locality Sensitive Hashing for Fast Similarity Search 430
Venu Satuluri, Srinivasan Parthasarathy

Fast and Exact Top-k Search for Random Walk with Restart 442
Yasuhiro Fujiwara, Makoto Nakatsuji, Makoto Onizuka, Masaru Kitsuregawa
Densest Subgraph in Streaming and MapReduce 454
Bahman Bahmani, Ravi Kumar, Sergei Vassilvitskii
Mining Attribute-Structure Correlated Patterns in Large Attributed Graphs 466
Arlei Silva, Wagner Meira Jr., Mohammed J. Zaki
Semi-Automatic Index Tuning: Keeping DBAs in the Loop 478
Karl Schnaitter, Neoklis Polyzotis
Aggregation in Probabilistic Databases via Knowledge Compilation 490
Robert Fink, Larisa Han, Dan Olteanu

NUMBER 6

RESEARCH TRACK PAPERS

Stochastic Database Cracking: Towards Robust Adaptive Indexing in Main-Memory Column-Stores 502
Felix Halim, Stratos Idreos, Panagiotis Karras, Roland H. C. Yap
An Adaptive Mechanism for Accurate Query Answering under Differential Privacy 514
Chao Li, Gerome Miklau
SharedDB: Killing One Thousand Queries With One Stone 526
Georgios Giannikis, Gustavo Alonso, Donald Kossmann
Pushing the Boundaries of Crowd-enabled Databases with Query-driven Schema Expansion 538
Joachim Selke, Christoph Lofi, Wolf-Tilo Balke
A Bayesian Approach to Discovering Truth from Conflicting Sources for Data Integration 550
Bo Zhao, Benjamin I. P. Rubinstein, Jim Gemmell, Jiawei Han
How to Price Shared Optimizations in the Cloud 562
Prasang Upadhyaya, Magdalena Balazinska, Dan Suciu
Dense Subgraph Maintenance under Streaming Edge Weight Updates for Real-time Story Identification 574
Albert Angel, Nick Koudas, Nikos Sarkas, Divesh Srivastava
ReStore: Reusing Results of MapReduce Jobs 586
Iman Elghandour, Ashraf Aboulnaga

NUMBER 7

RESEARCH TRACK PAPERS

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

Volume 2

Concurrency Control for Adaptive Indexing 656
Goetz Graefe, Felix Halim, Stratos Idreos, Harumi Kuno, Stefan Manegold

EXPERIMENTS AND ANALYSIS TRACK PAPERS

Comments on "Stack-based Algorithms for Pattern Matching on DAGs" 668
Qiang Zeng, Hai Zhuge

An Analysis of Structured Data on the Web 680
Nilesh Dalvi, Ashwin Machanavajjhala, Bo Pang

NUMBER 8

RESEARCH TRACK PAPERS

Shortest Path Computation with No Information Leakage 692
Kyriakos Mouratidis, Man Lung Yiu
V-SMART-Join: A Scalable MapReduce Framework for All-Pair Similarity Joins of Multisets and Vectors 704
Ahmed Metwally, Christos Faloutsos
Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud 716
Yucheng Low, Joseph Gonzalez, Aapo Kyrola, Danny Bickson, Carlos Guestrin, Joseph M. Hellerstein
Adding Logical Operators to Tree Pattern Queries on Graph-Structured Data 728
Qiang Zeng, Xiaorui Jiang, Hai Zhuge
Learning Semantic String Transformations from Examples 740
Rishabh Singh, Sumit Gulwani
Cologne: A Declarative Distributed Constraint Optimization Platform 752
Changbin Liu, Lu Ren, Boon Thau Loo, Yun Mao, Prithwish Basu
Optimizing I/O for Big Array Analytics 764
Yi Zhang, Jun Yang
Probabilistically Bounded Staleness for Practical Partial Quorums 776
Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Heller Stein, Ion Stoica

NUMBER 9

RESEARCH TRACK PAPERS

Efficient Subgraph Matching on Billion Node Graphs 788
Zhao Sun, Hongzhi Wang, Haixun Wang, Bin Shao, Jianzhong Li
Efficient Subgraph Similarity Search on Large Probabilistic Graph Databases 800
Ye Yuan, Guoren Wang, Lei Chen, Haixun Wang
Truss Decomposition in Massive Networks 812
Jia Wang, James Cheng
SEAL: Spatio-Textual Similarity Search 824
Ju Fan, Guoliang Li, Lizhu Zhou, Shanshan Chen, Jun Hu
On The Spatiotemporal Burstiness of Terms 836
Theodoros Lappas, Marcos R. Vieira, Dimitrios Gunopulos, Vassilis J. Tsotras
Efficient Reachability Query Evaluation in Large Spatiotemporal Contact Datasets 848
Houtan Shirani-Mehr, Farnoush Banaei Kashani, Cyrus Shahabi
Boosting Moving Object Indexing through Velocity Partitioning 860
Thi Nguyen, Zhen He, Rui Zhang, Phillip Ward
Type-Based Detection of XML Query-Update Independence 872
Nicole Bidoit-Tollu, Dario Colazzo, Federico Ulliana
Minuet: A Scalable Distributed Multiversion B-Tree 884
Benjamin Sowell, Wojciech Golab, Mehul A. Shah
Challenging the Long Tail Recommendation 896
Hongzhi Yin, Bin Cui, Jing Li, Junjie Yao, Chen Chen

NUMBER 10

RESEARCH TRACK PAPERS

Answering Table Queries on the Web using Column Keywords 908
Rakesh Pimplikar, Sunita Sarawagi
Efficient Verification of Web-Content Searching Through Authenticated Web Crawlers 920
Michael T. Goodrich, Duy Nguyen, Olga Ohrimenko, Charalampos Papamanthou, Roberto Tamassia, Nikos Triandopoulos, Cristina Videira Lopes

SODA: Generating SQL for Business Users 932
Lukas Blunschi, Claudio Jossen, Donald Kossmann, Magdalini Mori, Kurt Stockinger
Privacy Preservation by Disassociation 944
Manolis Terrovitis, John Liagouris, Nikos Mamoulis, Spiros Skiadopoulos
Supercharging Recommender Systems using Taxonomies for Learning User Purchase Behavior 956
Bhargav Kanagal, Amr Ahmed, Sandeep Pandey, Vanja Josifovski, Jeff Yuan, Lluis Garcia-Pueyo
DBToaster: Higher-order Delta Processing for Dynamic, Frequently Fresh Views 968
Yanif Ahmad, Oliver Kennedy, Christoph Koch, Milos Nikolic
Real Time Discovery of Dense Clusters in Highly Dynamic Graphs: Identifying Real World Events in Highly Dynamic Environments 980
Manoj K. Agarwal, Krithi Ramamritham, Manish Bhide
Sketch-based Querying of Distributed Sliding-Window Data Streams 992
Odysseas Papapetrou, Minos Garofalakis, Antonios Deligiannakis
LogBase: A Scalable Log-structured Database System in the Cloud 1004
Hoang Tam Vo, Sheng Wang, Divyakant Agrawal, Gang Chen, Beng Chin Ooi
Efficient Processing of k Nearest Neighbor Joins using MapReduce 1016
Wei Lu, Yanyan Shen, Su Chen, Beng Chin Ooi
Early Accurate Results for Advanced Analytics on MapReduce 1028
Nikolay Laptev, Kai Zeng, Carlo Zaniolo
CDAS: A Crowdsourcing Data Analytics System 1040
Xuan Liu, Meiyu Lu, Beng Chin Ooi, Yanyan Shen, Sai Wu, Meihui Zhang
Mining Statistically Significant Substrings using the Chi-Square Statistic 1052
Mayank Sachan, Arnab Bhattacharya
Massively Parallel Sort-Merge Joins in Main Memory Multi-Core Database Systems 1064
Martina-Cezara Albutiu, Alfons Kemper, Thomas Neumann
hStorage-DB: Heterogeneity-aware Data Management to Exploit the Full Capability of Hybrid Storage Systems 1076
Tian Luo, Rubao Lee, Michael Mesnier, Feng Chen, Xiaodong Zhang

NUMBER 11

RESEARCH TRACK PAPERS

A Scalable Algorithm for Maximizing Range Sum in Spatial Databases 1088
Dong-Wan Choi, Chin-Wan Chung, Yufei Tao
Spatial Queries with Two kNN Predicates 1100
Ahmed M. Aly, Walid G. Aref, Mourad Ouzzani
Optimal Algorithms for Crawling a Hidden Database in the Web 1112
Cheng Sheng, Nan Zhang, Yufei Tao, Xin Jin
Diversifying Top-K Results 1124
Lu Qin, Jeffrey Xu Yu, Lijun Chang
Keyword-aware Optimal Route Search 1136
Xin Cao, Lisi Chen, Gao Cong, Xiaokui Xiao
Answering Queries using Views over Probabilistic XML: Complexity and Tractability 1148
Bogdan Cautis, Evgeny Kharlamov
Probabilistic Databases with MarkoViews 1160
Abhay Jha, Dan Suciu
The Complexity of Social Coordination 1172
Konstantinos Mamouras, Sigal Oren, Lior Seeman, Lucja Kot, Johannes Gehrke
Efficient Multi-way Theta-Join Processing Using MapReduce 1184
Xiaofei Zhang, Lei Chen, Min Wang
Stubby: A Transformation-based Optimizer for MapReduce Workflows 1196
Harold Lim, Herodotos Herodotou, Shivnath Babu
Labeling Workflow Views with Fine-Grained Dependencies 1208
Zhuowei Bao, Susan B. Davidson, Tova Milo
Fundamentals of Order Dependencies 1220
Jaroslaw Szlichta, Parke Godfrey, Jarek Gryz
FDB: A Query Engine for Factorised Relational Databases 1232
Nurzhan Bakibayev, Dan Olteanu, Jakub Závodný
Optimization of Analytic Window Functions 1244
Yu Cao, Chee-Yong Chan, Jie Li, Kian-Lee Tan

Opening the Black Boxes in Data Flow Optimization 1256
Fabian Hueske, Mathias Peters, Matthias Sax, Astrid Rheinländer, Rico Bergmann, Aljoscha Krettek, Kostas Tzoumas
Spinning Fast Iterative Data Flows 1268
Stephan Ewen, Kostas Tzoumas, Moritz Kaufmann, Volker Markl
REX: Recursive, Delta-Based Data-Centric Computation 1280
Svilen R. Mihaylov, Zachary G. Ives, Sudipto Guha
K-Reach: Who is in Your Small World 1292
James Cheng, Zechao Shang, Hong Cheng, Haixun Wang, Jeffrey Xu Yu
Performance Guarantees for Distributed Reachability Queries 1304
Wenfei Fan, Xin Wang, Yinghui Wu
Efficient Indexing and Querying over Syntactically Annotated Trees 1316
Pirooz Chubak, Davood Rafiei
Queries with Guarded Negation 1328
Vince Barany, Balder Ten Cate, Martin Otto

Volume 3

PrivBasis: Frequent Itemset Mining with Differential Privacy 1340
Ninghui Li, Wahbeh Qardaji, Dong Su, Jianneng Cao
Low-Rank Mechanism: Optimizing Batch Queries under Differential Privacy 1352
Ganzhao Yuan, Zhenjie Zhang, Marianne Winslett, Xiaokui Xiao, Yin Yang, Zhifeng Hao
Functional Mechanism: Regression Analysis under Differential Privacy 1364
Jun Zhang, Zhenjie Zhang, Xiaokui Xiao, Yin Yang, Marianne Winslett
Injecting Uncertainty in Graphs for Identity Obfuscation 1376
Paolo Boldi, Francesco Bonchi, Aris Gionis, Tamir Tassa
Publishing Microdata with a Robust Privacy Guarantee 1388
Jianneng Cao, Panagiotis Karras
Measuring Two-Event Structural Correlations on Graphs 1400
Ziyu Guan, Xifeng Yan, Lance M. Kaplan
Ranking Large Temporal Data 1412
Jeffrey Jestes, Jeff M. Phillips, Feifei Li, Mingwang Tang
Compacting Transactional Data in Hybrid OLTP & OLAP Databases 1424
Florian Funke, Alfons Kemper, Thomas Neumann
Processing a Trillion Cells per Mouse Click 1436
Alexander Hall, Olaf Bachmann, Robert Büssow, Silviu Ganceanu, Marc Nunkesser
OLTP on Hardware Islands 1447
Danica Porobic, Ippokratis Pandis, Miguel Branco, Pinar Tözün, Anastasia Ailamaki
Serializability, not Serial: Concurrency Control and Availability in Multi-Datacenter Datastores 1459
Stacy Patterson, Aaron J. Elmore, Faisal Nawab, Divyakant Agrawal, Amr El Abbadi
Automatic Partitioning of Database Applications 1471
Alvin Cheung, Owen Arden, Samuel Madden, Andrew C. Myers
CrowdER: Crowdsourcing Entity Resolution 1483
Jiannan Wang, Tim Kraska, Michael J. Franklin, Jianhua Feng
Whom to Ask? Jury Selection for Decision Making Tasks on Micro-blog Services 1495
Caleb Chen Cao, Jieying She, Yongxin Tong, Lei Chen
ALAE: Accelerating Local Alignment with Affine Gap Exactly in Biosequence Databases 1507
Xiaochun Yang, Honglei Liu, Bin Wang
sDTW: Computing DTW Distances using Locally Relevant Constraints based on Salient Feature Alignments 1519
K. Selçuk Candan, Rosaria Rossini, Maria Luisa Sapino, Xiaolan Wang
SCOUT: Prefetching for Latent Feature Following Queries 1531
Farhan Tauheed, Thomas Heinis, Felix Shürmann, Henry Markram, Anastasia Ailamaki
Accelerating Pathology Image Data Cross-Comparison on CPU-GPU Hybrid Systems 1543
Kaibo Wang, Yin Huai, Rubao Lee, Fusheng Wang, Xiaodong Zhang, Joel H. Saltz
Robust Estimation of Resource Consumption for SQL Queries using Statistical Techniques 1555
Jiexing Li, Arnd Christian König, Vivek Narasayya, Surajit Chaudhuri
Who Tags What? An Analysis Framework 1567
Mahashweta Das, Saravanan Thirumuruganathan, Sihem Amer-Yahia, Gautam Das, Cong Yu
A Generic Framework for Efficient and Effective Subsequence Retrieval 1579
Haohan Zhu, George Kollios, Vassilis Athitsos

Only Aggressive Elephants are Fast Elephants 1591
Jens Dittrich, Jorge-Arnulfo Quiané-Ruiz, Stefan Richter, Stefan Schuh, Alekh Jindal, Jörg Schad
Multiple Location Profiling for Users and Relationships from Social Network and Content 1603
Rui Li, Shengjie Wang, Kevin Chen-Chuan Chang
Flash-based Extended Cache for Higher Throughput and Faster Recovery 1615
Woon-Hak Kang, Sang-Won Lee, Bongki Moon
Don't Thrash: How to Cache Your Hash on Flash 1627
Michael A. Bender, Martin Farach-Colton, Rob Johnson, Russell Kraner, Bradley C. Kuszmaul, Dzejla Medjedovic, Pablo Montes, Pradeep Shetty, Richard P. Spillane, Erez Zadok
Learning Expressive Linkage Rules using Genetic Programming 1638
Robert Isele, Christian Bizer

EXPERIMENTS AND ANALYSIS TRACK PAPERS

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

NUMBER 12

KEYNOTES

Data Management on the Spatial Web 1696
Christian S. Jensen
Data Analytics Opportunities in a Smarter Planet 1697
Brenda Dietrich
Challenges in Economic Massive Content Storage and Management (MCSAM) in the Era of Self-Organizing, Self-Expanding and Self-Linking Data Clusters 1698
Kenan Sahin

10-YEAR BEST PAPER AWARD

Approximate Frequency Counts over Data Streams 1699
Gurmeet Singh Manku, Rajeev Motwani

INDUSTRIAL, APPLICATIONS, AND EXPERIENCE TRACK PAPERS

The MADlib Analytics Library or MAD Skills, the SQL 1700
Joe Hellerstein, Christopher Ré, Florian Schoppmann, Daisy Zhe Wang, Eugene Fratkin, Aleksander Gorajek, Kee Siong Ng, Caleb Welton, Xixuan Feng, Kun Li, Arun Kumar
Can the Elephants Handle the NoSQL Onslaught? 1712
Avrilia Floratou, Nikhil Teletia, David J. Dewitt, Jignesh M. Patel, Donghui Zhang
Solving Big Data Challenges for Enterprise Application Performance Management 1724
Tilmann Rabl, Mohammad Sadoghi, Hans-Arno Jacobsen, Sergio Gómez-Villamor, Victor Muntés-Mulero, Serge Mankowskii
M3R: Increased Performance for In-memory Hadoop Jobs 1736
Avraham Shinnar, David Cunningham, Benjamin Herta, Vijay Saraswat
A Storage Advisor for Hybrid-Store Databases 1748
Philipp Rösch, Lars Dannecker, Gregor Hackenbroich, Franz Faerber
From Cooperative Scans to Predictive Buffer Management 1759
Michal Switakowski, Peter Boncz, Marcin Zukowski
The Unified Logging Infrastructure for Data Analytics at Twitter 1771
George Lee, Jimmy Lin, Chuang Liu, Andrew Lorek, Dmitriy Ryaboy

Transaction Log Based Application Error Recovery and Point In-Time Query 1781
Tomas Talius, Robin Dhamankar, Andrei Dumitrache, Hanuma Kodavalla
The Vertica Analytic Database: C-Store 7 Years Later 1790
Andrew Lamb, Matt Fuller, Ramakrishna Varadarajan, Nga Tran, Ben Vandier, Lyric Doshi, Chuck Bear
Interactive Analytical Processing in Big Data Systems: A Cross-Industry Study of MapReduce Workloads 1802
Yanpei Chen, Sara Alspaugh, Randy Katz
Muppet: MapReduce-Style Processing of Fast Data 1814
Wang Lam, Lu Liu, S. T. S. Prasad, Anand Rajaraman, Zoheb Vacheri, Anhai Doan
Building User-defined Runtime Adaptation Routines for Stream Processing Applications 1826
Gabriela Jacques-Silva, Bugra Gedik, Rohit Wagle, Kun-Lung Wu, Vibhore Kumar
MOIST: A Scalable and Parallel Moving Object Indexer with School Tracking 1838
Junchen Jiang, Hongji Bao, Edward Y. Chang, Yuqian Li
Serializable Snapshot Isolation in PostgreSQL 1850
Dan R. K. Ports, Kevin Grittner
Exploiting Evidence from Unstructured Data to Enhance Master Data Management 1862
Karin Murthy, Prasad M. Deshpande, Atreyee Dey, Ramanujam Halasipuram, Mukesh Mohania, P. Deepak, Jennifer Reed, Scott Schumacher
Avatara: OLAP for Web-scale Analytics Products 1874
Lili Wu, Roshan Sumbaly, Chris Riccomini, Gordon Koo, Hyung Jin Kim, Jay Kreps, Sam Shah

DEMONSTRATION TRACK PAPERS

Dedoop: Efficient Deduplication with Hadoop 1878
Lars Kolb, Andreas Thor, Erhard Rahm
MapReduce-based Dimensional ETL Made Easy 1882
Xiufeng Liu, Christian Thomsen, Torben Bach Pedersen
CloudVista: Interactive and Economical Visual Cluster Analysis for Big Data in the Cloud 1886
Huiqi Xu, Zhen Li, Shumin Guo, Keke Chen
Myriad: Scalable and Expressive Data Generation 1890
Alexander Alexandrov, Kostas Tzoumas, Volker Markl
A Demonstration of DBWipes: Clean as You Query 1894
Eugene Wu, Samuel Madden, Michael Stonebraker
ASTERIX: An Open Source System for "Big Data" Management and Analysis 1898
Sattam Alsubaiee, Yasser Altowim, Hotham Altwaijry, Alexander Behm, Vinayak Borkar, Yingyi Bu, Michael Carey, Raman Grover, Zachary Heilbron, Young-Seok Kim, Chen Li, Nicola Onose, Pouria Pirzadeh, Rares Vernica, Jian Wen
Blink and It's Done: Interactive Queries on Very Large Data 1902
Sameer Agarwal, Aurojit Panda, Barzan Mozafari, Anand P. Iyer, Samuel Madden, Ion Stoica
Massive Genomic Data Processing and Deep Analysis 1906
Abhishek Roy, Yanlei Diao, Evan Mauceli, Yiping Shen, Bai-Lin Wu
MonetDB/DataCell: Online Analytics in a Streaming Column-Store 1910
Erietta Liarou, Stratos Idreos, Stefan Manegold, Martin Kersten
SWORS: A System for the Efficient Retrieval of Relevant Spatial Web Objects 1914
Xin Cao, Gao Cong, Christian S. Jensen, Jun Jie Ng, Beng Chin Ooi, Nhan-Tue Phan, Dingming Wu
CyLog/Crowd4U: A Declarative Platform for Complex Data-centric Crowdsourcing 1918
Atsuyuki Morishima, Norihide Shinagawa, Tomomi Mitsuishi, Hideto Aoki, Shun Fukusumi
Exploiting Database Similarity Joins for Metric Spaces 1922
Yasin N. Silva, Spencer Pearson
Stethoscope: A Platform for Interactive Visual Analysis of Query Execution Plans 1926
Mrunal Gawade, Martin Kersten
Hum-a-song: A Subsequence Matching with Gaps-Range-Tolerances Query-By-Humming System 1930
Alexios Kotsifakos, Panagiotis Papapetrou, Jaakko Hollmén, Dimitrios Gunopulos, Vassilis Athitsos, George Kollios
SkewTune in Action: Mitigating Skew in MapReduce Applications 1934
Yongchul Kwon, Magdalena Balazinska, Bill Howe, Jerome Rolia
Playful Query Specification with DataPlay 1938
Azza Abouzied, Joseph M. Hellerstein, Avi Silberschatz
NoDB in Action: Adaptive Query Processing on Raw Data 1942
Ioannis Alagiannis, Renata Borovica, Miguel Branco, Stratos Idreos, Anastasia Ailamaki
Complex Preference Queries Supporting Spatial Applications for User Groups 1946
Florian Wenzel, Markus Endres, Stefan Mandl, Werner Kießling

Demonstration of the FDB Query Engine for Factorised Databases 1950
Nurzhan Bakibayev, Dan Olteanu, Jakub Závodný
PET: Reducing Database Energy Cost via Query Optimization 1954
Zichen Xu, Yi-Cheng Tu, Xiaorui Wang
SPAM: A SPARQL Analysis and Manipulation Tool 1958
Andrés Letelier, Jorge Pérez, Reinhard Pichler, Sebastian Skritek
QueryMarket Demonstration: Pricing for Online Data Markets 1962
Paraschos Koutris, Prasang Upadhyaya, Magdalena Balazinska, Bill Howe, Dan Suciu
DISKs: A System for Distributed Spatial Group Keyword Search on Road Networks 1966
Iqiang Luo, Yifeng Luo, Shuigeng Zhou, Gao Cong, Jihong Guan
WETSUIT: An Efficient Mashup Tool for Searching and Fusing Web Entities 1970
Stefan Endrullis, Andreas Thor, Erhard Rahm
Model-based Integration of Past & Future in TimeTravel 1974
Mohamed E. Khalefa, Ulrike Fischer, Torben Bach Pedersen, Wolfgang Lehner
DrillBeyond: Enabling Business Analysts to Explore the Web of Open Data 1978
Julian Eberius, Maik Thiele, Katrin Braunschweig, Wolfgang Lehner
Discovering and Exploring Relations on the Web 1982
Ndapandula Nakashole, Gerhard Weikum, Fabian Suchanek
MapRat: Meaningful Explanation, Interactive Exploration and Geo-Visualization of Collaborative Ratings 1986
Saravanan Thirumuruganathan, Mahashweta Das, Shrikant Desai, Sihem Amer-Yahia, Gautam Das, Cong Yu
Deco: A System for Declarative Crowdsourcing 1990
Hyunjung Park, Richard Pang, Aditya Parameswaran, Hector Garcia-Molina, Neoklis Polyzotis, Jennifer Widom
Developing and Analyzing XSDs through BonXai 1994
Wim Martens, Frank Neven, Matthias Niewerth, Thomas Schwentick
InfoPuzzle: Exploring Group Decision Making in Mobile Peer-to-Peer Databases 1998
Aaron J. Elmore, Sudipto Das, Divyakant Agrawal, Amr El Abbadi
Manage and Query Generic Moving Objects in SECONDO 2002
Jianqiu Xu, Ralf Hartmut Güting
Chronos: Facilitating History Discovery by Linking Temporal Records 2006
Pei Li, Haidong Wang, Christina Tziviskou, Xin Luna Dong, Xiaoguang Liu, Andrea Maurino, Divesh Srivastava
TELEIOS: A Database-Powered Virtual Earth Observatory 2010
Manolis Koubarakis, Kostis Kyzirakos, Manos Karpathiotakis, Charalampos Nikolaou, Stavros Vassos, George Garbis, Michael Sioutis, Konstantina Bereta, Dimitrios Michail, Charalampos Kontoes, Ioannis Papoutsis, Themos Herekakis, Stefan Manegold, Martin Kersten, Milena Ivanova, Holger Pirk, Ying Zhang, Mihai Datcu, Gottfried Schwarz, Corneliu Dumitru, Daniela Espinoza Molina, Katrin Molch, Ugo Di Giammatteo, Manuela Sagona, Sergio Perelli, Thorsten Reitz, Eva Klien, Robert Gregor

TUTORIALS

Efficient Big Data Processing in Hadoop MapReduce 2014
Jens Dittrich, Jorge-Arnulfo Quiané-Ruiz
MapReduce Algorithms for Big Data Analysis 2016
Kyuseok Shim
Entity Resolution: Theory, Practice & Open Challenges 2018
Lise Getoor, Ashwin Machanavajjhala
I/O Characteristics of NoSQL Databases 2020
Jiri Schindler
Mining Knowledge from Interconnected Data: A Heterogeneous Information Network Analysis Approach 2022
Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu
Understanding and Managing Cascades on Large Graphs 2024
B. Aditya Prakash, Christos Faloutsos
Interoperability in eHealth Systems (Invited Tutorial) 2026
Asuman Dogac
Secure and Privacy-Preserving Data Services in the Cloud: A Data Centric View 2028
Divyakant Agrawal, Amr El Abbadi, Shiyuan Wang
Graph Synopses, Sketches, and Streams: A Survey 2030
Sudipto Guha, Andrew McGregor

PANELS

Challenges and Opportunities with Big Data 2032
Alexandros Labrinidis, H. V. Jagadish
Social Networks and Mobility in the Cloud 2034
Amr El Abbadi, Mohamed F. Mokbel
Author Index