44th International Conference on Very Large Data Bases (VLDB 2018)

Proceedings of the VLDB Endowment Volume 11 No. 1-13

Rio de Janeiro, Brazil
27-31 August 2018

Part 1 of 3

Editors:

Sihem Amer-Yahia
Jian Pei

ISBN: 978-1-5108-7500-5

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© (2017) by VLDB Endowment
This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-ncnd/4.0/. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org.

Printed by Curran Associates, Inc. (2019)

For permission requests, please contact 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

info@vldb.org.

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

Front Matter

Table of Contents
VLDB 2018 Organization and Review Board

Research Papers

Relaxed Operator Fusion for In-Memory Databases: Making Compilation, Vectorization, and Prefetching Work Together At Last ..... Prashanth Menon, Todd C. Mowry, Andrew Pavlo 1

ProbeSim: Scalable SingleSource and Top-k SimRank Computations on Dynamic Graphs ..... Yu Liu, Bolong Zheng, Xiaodong He, Zhewei Wei, Xiaokui Xiao, Kai Zheng, Jiaheng Lu 14

A Formal Semantics of SQL Queries, Its Validation, and Applications ..... Paolo Guagliardo, Leonid Libkin 27

Efficient Haar$^+$ Synopsis Construction for the Maximum Absolute Error Measure ..... Jinhyun Kim, Jun-Ki Min, Kyuseok Shim 40

Approximate String Joins with Abbreviations ..... Wenbo Tao, Dong Deng, Michael Stonebraker 53

Query-Driven On-The-Fly Knowledge Base Construction ..... Dat Ba Nguyen, Abdalghani Abujabal, Nam Khanh Tran, Martin Theobald, Gerhard Weikum 66

GRETA: Graph-based Real-time Event Trend Aggregation ..... Olga Poppe, Chuan Lei, Elke A. Rundensteiner, David Maier 80

Parallel Personalized PageRank on Dynamic Graphs ..... Wentian Guo, Yuchen Li, Mo Sha, Kian-Lee Tan 93

Accelerating Dynamic Graph Analytics on GPUs ..... Mo Sha, Yuchen Li, Binsheng He, Kian-Lee Tan 107

TABLE OF CONTENTS

Research Papers

Analyzing the Impact of System Architecture on the Scalability of OLTP Engines for High-Contention Workloads ..... Raja Appuswamy, Angelos Anadiotis, Danica Porobic, Mustafa Iman, Anastasia Ailamaki 121

Scalable Database Logging for Multicores ..... Hyungsoo Jung, Hyuck Han, Sooyong Kang 135

An Analytical Study of Large SPARQL Query Logs ..... Angela Bonifati, Wim Martens, Thomas Timm 149

Approximately Counting Triangles in Large Graph Streams Including Edge Duplicates with a Fixed Memory Usage..... Pinghui Wang, Yiyan Qi, Yu Sun, Xiangliang Zhang, Jing Tao, Xiaohong Guan 162

Subgraph Matching: on Compression and Computation ..... Miao Qiao, Hao Zhang, Hong Cheng 176

Synthesizing Entity Matching Rules by Examples . ..... Rohit Singh, Venkata Vamsikrishna Meduri, Ahmed Elmagarmid, Samuel Madden, Paolo Papotti, Jorge-Arnulfo Quiane-Ruiz, Armando Solar-Lezama, Nan Tang 189

Stylus: A Strongly-Typed Store for Serving Massive RDF Data. ..... Liang He, Bin Shao, Yatao Li, Huanhuan Xia, Yanghua Xiao, Enhong Chen, Liang Chen 203

Holistic Query Evaluation over Information Extraction Pipelines. ..... Ekaterini Ioannou, Minos Garofalakis 217

Interleaving with Coroutines: A Practical Approach for Robust Index Joins. ..... Georgios Psaropoulos, Thomas Legler, Norman May, Anastasia Ailamaki 230

TABLE OF CONTENTS

Research Papers

Efficient Structural Graph Clustering: An Index-Based Approach ..... Dong Wen, Lu Qin, Ying Zhang, Lijun Chang, Xuemin Lin 243

An Authorization Model for Multi-Provider Queries ..... Sabrina De Capitani di Vimercati, Sara Foresti, Sushil Jajodia, Giovanni Livraga, Stefano Paraboschi, Pierangela Samarati 256

Snorkel: Rapid Training Data Creation with Weak Supervision ..... Alexander Ratner, Stephen Bach, Henry Ehrenberg, Jason Fries, Sen Wu, Christopher Re 269

VERIFAS: A Practical Verifier for Artifact Systems..... Yuliang Li, Alin Deutsch, Victor Vianu 283

A Distributed Multi-GPU System for Fast Graph Processing ..... Zhihao Jia, Yongkee Kwon, Galen Shipman, Pat McCormick, Mattan Erez, Alex Aiken 297

Efficient Denial Constraint Discovery with Hydra ..... Tobias Bleifuß, Sebastian Kruse, Felix Naumann 311

ReCache: Reactive Caching for Fast Analytics over Heterogeneous Data..... Tahir Azim, Manos Karpathiotakis, Anastasia Ailamaki 324

Effective and Efficient Dynamic Graph Coloring..... Long Yuan, Lu Qin, Xuemin Lin, Lijun Chang, Wenjie Zhang 338

GPU Rasterization for Real-Time Spatial Aggregation over Arbitrary Polygons. ..... Eleni Tzirita Zacharatou, Harish Doraiswamy, Anastasia Ailamaki, Claudio Silva, Juliana Freire 352

Are Key-Foreign Key Joins Safe to Avoid when Learning High-Capacity Classifiers?..... Vraj Shah, Arun Kumar, Xiaojin Zhu 366

Worker Recommendation for Crowdsourced Q&A Services: A Triple-Factor Aware Approach. ..... Zheng Liu, Lei Chen 380

TABLE OF CONTENTS

Research Papers

Clustering Stream Data by Exploring the Evolution of Density Mountain ……
…… Shufeng Gong, Yanfeng Zhang, Ge Yu 393

Query Fresh: Log Shipping on Steroids ……
…… Tianzheng Wang, Ryan Johnson, Ippokratis Pandis 406

The Ubiquity of Large Graphs and Surprising Challenges of Graph Processing ……
…… Siddhartha Sahu, Amine Mhedhbi, Semih Salihoglu, Jimmy Lin, Tamer Özsu 420

Froid: Optimization of Imperative Programs in a Relational Database……
……Karthik Ramachandra, Kwanghyun Park, K. Venkatesh Emani, Alan Halverson, Cesar Galindo-Legaria, Conor Cunningham 432

An Experimental Study on Hub Labeling based Shortest Path Algorithms……
…… Ye Li, Leong Hou U, Man Lung Yiu, Ngai Meng Kou 445

Concurrent Log-Structured Memory for Many-Core Key-Value Stores ……
…… Alexander Merritt, Ada Gavrilovska, Yuan Chen, Dejan Milojicic 458

Clustering Uncertain Graphs. ……
…… Matteo Ceccarello, Carlo Fantozzi, Andrea Pietracaprina, Geppino Pucci, Fabio Vandin 472

Lusail: A System for Querying Linked Data at Scale. ……
…… Ibrahim Abdelaziz, Essam Mansour, Mourad Ouzzani, Ashraf Aboulnaga, Panos Kalnis 485

Cardinality Estimation: An Experimental Survey. ……
…… Hazar Harmouch, Felix Naumann 499

SQL Statement Logging for Making SQLite Truly Lite……
…… Jong-Hyeok Park, Gihwan Oh, Sang Won Lee 513

TABLE OF CONTENTS

Research Papers

Towards Practical Differential Privacy for SQL Queries..... Noah Johnson, Joseph Near, Dawn Song 526

CloudKit: Structured Storage for Mobile Applications..... Alexander Shraer, Alexandre Aybes, Bryan Davis, Christos Chrysafis, Dave Browning, Eric Krugler, Eric Stone, Harrison Chandler, Jacob Farkas, John Quinn, Jonathan Ruben, Michael Ford, Mike McMahon, Nathan Williams, Nicolas Favre-Felix, Nihar Sharma, Ori Herrnstadt, Paul Seligman, Raghav Pisolkar, Scott Dugas, Scott Gray, Shirley Lu, Sytze Harkema, Valentin Kravtsov, Vanessa Hong, Wan Ling Yih, Yizuo Tian..... 540

BzTree: A High-Performance Latch-free Range Index for Non-Volatile Memory..... Joy Arulraj, Justin Levandoski, Umar Farooq Minhas, Per-Ake Larson 553

FlexPS: Flexible Parallelism Control in Parameter Server Architecture..... Yuzhen Huang, Tatiana Jin, Yidi Wu, Zhenkun Cai, Xiao Yan, Fan Yang, Jinfeng Li, Yuying Guo, James Cheng 566

Automated Migration of Hierarchical Data to Relational Tables using Programming-by-Example... Navid Yaghmazadeh, Xinyu Wang, Isil Dillig 580

TOAIN: A Throughput Optimizing Adaptive Index for Answering Dynamic kNN Queries on Road Networks ..... Siqiang Luo, Ben Kao, Guoliang Li, Jiafeng Hu, Reynold Cheng, Yudian Zheng 594

Ease.ml: Towards Multi-tenant Resource Sharing for Machine Learning Workloads..... Tian Li, Jie Zhong, Ji Liu, Wentao Wu, Ce Zhang 607

Theoretically Optimal and Empirically Efficient R-trees with Strong Parallelizability. ..... Jianzhong Qi, Yufei Tao, Yanchuan Chang, Rui Zhang 621

Domain-Aware Multi-Truth Discovery from Conflicting Sources. ..... Xueling Lin, Lei Chen 635

Contention-Aware Lock Scheduling for Transactional Databases ..... Boyu Tian, Jiamin Huang, Barzan Mozafari, Grant Schoenebeck 648

TABLE OF CONTENTS

Research Papers

Quickstep: A Data Platform Based on the Scaling-Up Approach……. ……. Jignesh Patel, Harshad Deshmukh, Jianqiao Zhu, Navneet Potti, Zuyu Zhang, Marc Spehlmann, Hakan Memisoglu, Saket Saurabh 663

Coconut: A Scalable Bottom-Up Approach for Building Data Series Indexes..... .....Haridimos Kondylakis, Niv Dayan, Kostas Zoumpatianos, Themis Palpanas 677

Distributed Evaluation of Subgraph Queries Using Worst-case Optimal and Low-Memory Dataflows ……Khaled Ammar, Frank Mcsherry, Semih Salihoglu, Manas Joglekar 691

Model-free Control for Distributed Stream Data Processing using Deep Reinforcement Learning…. .....Teng Li, Zhiyuan Xu, Jian Tang, Yanzhi Wang 705

Smoke: Fine-grained Lineage at Interactive Speed……. ……Fotis Psallidas, Eugene Wu 719

TABLE OF CONTENTS

Research Papers

Conjunctive Queries with Inequalities Under Updates……..
……Muhammad Idris, Martín Ugarte, Stijn Vansummeren, Hannes Voigt, Wolfgang Lehner 733

Bubble Execution: Resource-aware Reliable Analytics at Cloud Scale…….
…… Zhicheng Yin, Jin Sun, Ming Li, Jaliya Ekanayake, Haibo Lin, Marc Friedman, Jose Blakeley, Clemens Szyperski, Nikhil Devanur 746

Efficient Discovery of Approximate Dependencies……..
……Sebastian Kruse, Felix Naumann 759

RC-Index: Diversifying Answers to Range Queries…….
…… Yue Wang, Alexandra Meliou, Gerome Miklau 773

UlTraMan: A Unified Platform for Big Trajectory Data Management and Analytics……..
……Xin Ding, Lu Chen, Yunjun Gao, Christian Jensen, Hujun Bao 787

Selecting Subexpressions to Materialize at Datacenter Scale……..
…… Alekh Jindal, Konstantinos Karanasos, Sriram Rao, Hiren Patel 800

Table Union Search on Open Data……..
…… Fatemeh Nargesian, Erkang Zhu, Ken Pu, Renée Miller 813

Scalable Training of Hierarchical Topic Model…….. ……
……Jianfei Chen, Jun Zhu, Jie Lu, Shixia Liu 826

TABLE OF CONTENTS

Research Papers

Indexed Fast Network Proximity Querying ..... Mustafa Coskun, Ananth Grama, Mehmet Koyuturk 840

Order Dispatch in Price-aware Ridesharing..... Libin Zheng, Lei Chen 853

Exact Processing of Uncertain Top-k Queries in Multi-criteria Settings ..... Kyriakos Mouratidis, Bo Tang 866

Discovery of Genuine Functional Dependencies from Relational Data with Missing Values..... Laure Berti-Equille, Hazar Harmouch, Felix Naumann, Noel Novelli, Saravanan Thirumuruganathan 880

Effective Temporal Dependence Discovery in Time Series Data ....Qingchao Cai, Zhongle Xie, Gang Chen, H. Jagadish, Beng Chin Ooi, Meihui Zhang 893

HD-Index: Pushing the Scalability-Accuracy Boundary for Approximate kNN Search in High-Dimensional Spaces..... .....Akhil Arora, Piyush Kumar, Sakshi Sinha, Arnab Bhattacharya 906

LA3: A Scalable Link- and Locality-Aware Linear Algebra-Based Graph Analytics System .... Yousuf Ahmad, Omar Khattab, Arsal Malik, Ahmad Musleh, Mohammad Hammoud, Mucahid Kutlu, Mostafa Shehata, Tamer Elsayed 920

TABLE OF CONTENTS

Research Papers

Trajectory Simplification: An Experimental Study and Quality Analysis ..... Dongxiang Zhang, Mengting Ding, Dingyu Yang, Yi Liu, Ju Fan, Heng Tao Shen 934

Constraint-based Explanation and Repair of Filter-Based Transformations ..... Dolan Antenucci, Michael Cafarella 947

Scalable Semantic Querying of Text ..... Xiaolan Wang, Aaron Feng, Behzad Golshan, Alon Halevy, George Mihaila, Hidekazu Oiwa, Wang-Chiew Tan 961

The Vadalog System: Datalog-based Reasoning for Knowledge Graphs ..... Luigi Bellomarini, Emanuel Sallinger, Georg Gottlob 975

Noticeable Network Delay Minimization via Node Upgrades ..... Sourav Medya, Jithin Vachery, Sayan Ranu, Ambuj Singh 988

Evaluating End-to-End Optimization for Data Analytics Applications in Weld ..... Shoumik Palkar, James Thomas, Deepak Narayanan, Pratiksha Thaker, Rahul Palamuttam, Parimarjan Negi, Anil Shanbhag, Malte Schwarzkopf, Holger Pirk, Dr. saman Amarasinghe, Samuel Madden, Matei Zaharia 1002

Improved Selectivity Estimation by Combining Knowledge from Sampling and Synopses ..... Magnus Müller, Guido Moerkotte, Oliver Kolb . 1016

Efficient Algorithms for Adaptive Influence Maximization ..... Kai Han, Keke Huang, Xiaokui Xiao, Jing Tang, Aixin Sun, Xueyan Tang 1029

Morton Filters: Faster, Space-Efficient Cuckoo Filters via Biasing, Compression, and Decoupled Logical Sparsity ..... Alexander Breslow, Nuwan Jayasena 1041

An Optimal and Progressive Approach to Online Search of Top-K Influential Communities ..... Fei Bi, Lijun Chang, Xuemin Lin, Wenjie Zhang 1056

TABLE OF CONTENTS

Research Papers

Errata for "Analysis of two existing and one new dynamic programming algorithm for the generation of optimal bushy join trees without cross products"..... Andreas Meister, Guido Moerkotte, Gunter Saake 1069

Data Synthesis based on Generative Adversarial Networks ..... Noseong Park, Mahmoud Mohammadi, Kshitij Gorde, Sushil Jajodia, Hongkyu Park, Youngmin Kim 1071

CERES: Distantly Supervised Relation Extraction from the Semi-Structured Web ..... Colin Lockard, Xin Luna Dong, Prashant Shiralkar, Arash Einolghozati. 1084

Efficient Estimation of Inclusion Coefficient using HyperLogLog Sketches ..... Azade Nazi, Bolin Ding, Vivek Narasayya, Surajit Chaudhuri 1097

Set Similarity Joins on MapReduce: An Experimental Survey ..... Fabian Fier, Nikolaus Augsten, Panagiotis Bouros, Ulf Leser, Johann-Christoph Freytag 1110

Plan Stitch: Harnessing the Best of Many Plans ..... Bailu Ding, Sudipto Das, Wentao Wu, Surajit Chaudhuri, Vivek Narasayya 1123

ForkBase: An Efficient Storage Engine for Blockchain and Forkable Applications..... Sheng Wang, Tien Tuan Anh Dinh, Qian Lin, Zhongle Xie, Meihui Zhang, Qingchao Cai, Gang Chen, Beng Chin Ooi, Pingcheng Ruan 1137

Experimental Analysis of Distributed Graph Systems ..... Khaled Ammar, Tamer Özsu 1151

Transform-Data-by-Example (TDE): An Extensible Search Engine for Data Transformations ..... Yeye He, Xu Chu, Kris Ganjam, Yudian Zheng, Vivek Narasayya, Surajit Chaudhuri 1165

Frontier: Resilient Edge Processing for the Internet of Things ..... Dan O'keeffe, Theodoros Salonidis, Peter Pietzuch 1178

LightDB: A DBMS for Virtual Reality Video ..... Brandon Haynes, Amrita Mazumdar, Armin Alaghi, Magdalena Balanziska, Luis Ceze, Alvin Cheung 1192

Optimizing error of high-dimensional statistical queries under differential privacy ..... Ryan Mckenna, Gerome Miklau, Michael Hay, Ashwin Machanavajjhala 1206

MLBench: Benchmarking Machine Learning Services Against Human Experts ..... Yu Liu, Hantian Zhang, Luyuan Zeng, Wentao Wu, Ce Zhang 1220

Maximum Co-located Community Search in Large Scale Social Networks ……
…… Lu Chen, Chengfei Liu, Rui Zhou, Jianxin Li, Xiaochun Yang, Bin Wang 1233

ChronosDB: Distributed, File Based, Geospatial Array DBMS ……
…… Ramon Antonio Rodriges Zalipynis. 1247

Adaptive Sampling for Rapidly Matching Histograms ……
…… Stephen Macke, Yiming Zhang, Silu Huang, Aditya Parameswaran 1262

Leveraging Similarity Joins for Signal Reconstruction ……
…… Abolfazl Asudeh, Azade Nazi,
Jees Augustine, Saravanan Thirumuruganathan, Nan Zhang, Gautam Das, Divesh Srivastava 1276

Sundial: Harmonizing Concurrency Control and Caching in a Distributed OLTP Database Management System ……
…… Xiangyao Yu, Yu Xia, Andrew Pavlo, Daniel Sanchez, Larry Rudolph, Srinivas Devadas 1289

Chi: A Scalable and Programmable Control Plane for Distributed Stream Processing Systems …..
…… Luo Mai, Kai Zeng, Rahul Potharaju, Le Xu, Steve Suh, Shivaram Venkataraman, Paolo Costa, Terry Kim, Saravanam Muthukrishnan, Vamsi Kuppa, Sudheer Dhulipalla, Sriram Rao 1303

TABLE OF CONTENTS

Research Papers

In-RDBMS Hardware Acceleration of Advanced Analytics..... .....Divya Mahajan, Joon Kyung Kim, Jacob Sacks, Adel Ardalan, Arun Kumar, Hadi Esmaeilzadeh 1317

Join Query Optimization Techniques for Complex Event Processing Applications ....Ilya Kolchinsky, Assaf Schuster 1332

Efficient Adaptive Detection of Complex Event Patterns. Join Query Optimization Techniques for Complex Event Processing Applications ....Ilya Kolchinsky, Assaf Schuster 1346

Robustness Metrics for Relational Query Execution Plans ..... Florian Wolf, Michael Brendle, Norman May, Paul Willems, Kai-Uwe Sattler, Michael Grossniklaus 1360

Question Answering Over Knowledge Graphs: Question Understanding Via Template Decomposition ..... Weiguo Zheng, Jeffrey Xu Yu, Lei Zou, Hong Cheng 1373

Explaining Repaired Data with CFDs ....Joeri Rammelaere, Floris Geerts 1387

AIDA - Abstraction for Advanced In-Database Analytics .... Joseph D'silva, Florestan De Moor, Bettina Kemme 1400

RHEEM: Enabling Cross-Platform Data Processing -- May The Big Data Be With You! --..... ..... Divy Agrawal, Sanjay Chawla, Bertty Contreras-Rojas, Ahmed Elmagarmid, Yasser Idris, Zoi Kaoudi, Sebastian Kruse, Ji Lucas, Essam Mansour, Mourad Ouzzani, Paolo Papotti, Jorge-Arnulfo Quiane-Ruiz, Nan Tang, Saravanan Thirumuruganathan, Anis Troudi 1414

An Experimental Evaluation of Task Assignment in Spatial Crowdsourcing .... Peng Cheng, Xun Jian, Lei Chen 1428

2SCENT: An Efficient Algorithm to Enumerate All Simple Temporal Cycles .... Rohit Kumar, Toon Calders 1441

Distributed Representations of Tuples for Entity Resolution ..... Muhammad Ebraheem, Saravanan Thirumuruganathan, Shafiq Joty, Mourad Ouzzani, Nan Tang 1454

Efficient Construction of Approximate Ad-Hoc ML models Through Materialization and Reuse .... Sona Hasani, Saravanan Thirumuruganathan, Abolfazl Asudeh, Nick Koudas, Gautam Das 1468

Axiomatic Foundations and Algorithms for Deciding Semantic Equivalences of SQL Queries .... Shumo Chu, Brendan Murphy, Jared Roesch, Alvin Cheung, Dan Suciu 1482

HomeRun: Scalable Sparse-Spectrum Reconstruction of Aggregated Historical Data ..... Faisal Almutairi, Fan Yang, Hyun Ah Song, Christos Faloutsos, Nicholas Sidiropoulos, Vladimir Zadorozhny 1496

Differentially Private Hierarchical Count-of-Counts Histograms ..... Yu-Hsuan Kuo, Cho-Chun Chiu, Daniel Kifer, Michael Hay, Ashwin Machanavajjhala 1509

Efficient Document Analytics on Compressed Data: Method, Challenges, Algorithms, Insights .... Feng Zhang, Jidong Zhai, Xipeng Shen, Onur Mutlu, Wenguang Chen 1522

You Say `What', I Hear `Where' and `Why'? (Mis-)Interpreting SQL to Derive Fine-Grained Provenance ..... Tobias Müller, Benjamin Dietrich, Torsten Grust 1536

An Eight-Dimensional Systematic Evaluation of Optimized Search Algorithms on Modern Processors ..... Lars-Christian Schulz, David Broneske, Gunter Saake 1550

Vocalizing Large Time Series Efficiently ..... Immanuel Trummer, Mark Bryan, Ramya Narasimha 1563

Filter Before You Parse: Faster Analytics on Raw Data with Sparser ..... Shoumik Palkar, Firas Abuzaid, Peter Bailis, Matei Zaharia 1576

Streaming Graph Partitioning: An Experimental Study ..... Zainab Abbas, Vasiliki Kalavri, Paris Carbone, Vladimir Vlassov 1590

Efficient Distributed Memory Management with RDMA and Caching ..... Qingchao Cai, Wentian Guo, Hao Zhang, Divyakant Agrawal, Gang Chen, Beng Chin Ooi, Kian-Lee Tan, Yong Meng Teo, Sheng Wang 1604

Causal Consistency and Latency Optimality: Friend or Foe? ..... Diego Didona, Rachid Guerraoui, Jingjing Wang, Willy Zwaenepoel 1618

A Unified Approach to Route Planning for Shared Mobility ..... Yongxin Tong, Yuxiang Zeng, Zimu Zhou, Lei Chen, Jieping Ye, Ke Xu 1633

Moment-Based Quantile Sketches for Efficient High Cardinality Aggregation Queries ..... Edward Gan, Jialin Ding, Kai Sheng Tai, Vatsal Sharan, Peter Bailis 1647

How Good Are Modern Spatial Analytics Systems? ..... Varun Pandey, Andreas Kipf, Thomas Neumann, Alfons Kemper 1661

Locality-Sensitive Hashing for Earthquake Detection: A Case Study Scaling Data-Driven Science .. Kexin Rong, Clara E. Yoon, Karianne J. Bergen, Hashem Elezabi, Peter Bailis, Philip Levis, Gregory C. Beroza 1674

ModelarDB: Modular Model-Based Time Series Management with Spark and Cassandra ..... Søren Jensen, Torben Pedersen, Christian Thomsen 1688

Exploiting Coroutines to Attack the "Killer Nanoseconds" ..... Christopher Jonathan, Umar Farooq Minhas, James Hunter, Justin Levandoski, Gor Nishanov 1702

The Tao of Inference in Privacy-Protected Databases ..... Vincent Bindschaedler, Paul Grubbs, David Cash, Thomas Ristenpart, Vitaly Shmatikov 1715

Efficient Searchable Encryption Through Compression ……
…… Ioannis Demertzis, Charalampos Papamanthou, Rajdeep Talapatra 1729

TABLE OF CONTENTS

Industrial and Applications

Challenges and Experiences in Building an Efficient Apache Beam Runner For IBM Streams ..... ……Shen Li, Paul Gerver, John Macmillan, Daniel Debrunner, William Marshall, Kun-Lung Wu 1742

On Optimizing Operator Fusion Plans for Large-Scale Machine Learning in SystemML ..... ……. Matthias Boehm, Berthold Reinwald, Dylan Hutchison, Prithviraj Sen, Alexandre Evfimievski, Niketan Pansare 1755

OLTPShare: The Case for Sharing in OLTP Workloads ..... ……Robin Rehrmann, Carsten Binnig, Alexander Boehm, Kihong Kim, Wolfgang Lehner, Amr Rizk 1769

Automating Large-Scale Data Quality Verification ..... ……. Sebastian Schelter, Dustin Lange, Philipp Schmidt, Meltem Celikel, Felix Biessmann, Andreas Grafberger 1781

Taking Omid to the Clouds: Fast, Scalable Transactions for Real-Time Cloud Analytics ..... Ohad Shacham, Yonatan Gottesman, Aran Bergman, Edward Bortnikov, Eshcar Hillel, Idit Keidar 1795

Providing Streaming Joins as a Service at Facebook ..... ……Gabriela Jacques-Silva, Ran Lei, Luwei Cheng, Guoqiang Jerry Chen, Kuen Ching, Tanji Hu,Yuan Mei, Kevin Wilfong, Rithin Shetty, Serhat Yilmaz, Anirban Banerjee, Benjamin Heintz, Shridar Iyer, Anshul Jaiswal 1809

FusionInsight LibrA: Huawei's Enterprise Cloud Data Analytics Platform ..... ……Le Cai, Jianjun Chen, Jun Chen, Yu Chen, Kuorong Chiang, Marko Dimitrijevic, Yonghua Ding, Yu Dong, Ahmad Ghazal, Jacques Hebert, Kamini Jagtiani, Suzhen Lin, Ye Liu, Demai Ni, Chunfeng Pei, Jason Sun, Li Zhang, Mingyi Zhang, Cheng Zhu 1822

F1 Query: Declarative Querying at Scale ..... ……. Bart Samwel, John Cieslewicz, Ben Handy, Jason Govig, Petros Venetis, Chanjun Yang, Keith Peters, Jeff Shute, Daniel Tenedorio, Himani Apte, Felix Weigel, David Wilhite, Jiacheng Yang, Jun Xu, Jiexing Li, Zhan Yuan, Craig Chasseur, Qiang Zeng, Ian Rae, Anurag Biyani, Andrew Harn, Yang Xia, Andrey Gubichev, Amr ElHelw, Orri Erling, Zhepeng Yan, Mohan Yang, Yiqun Wei, Thanh Do, Colin Zheng, Goetz Graefe, Somayeh Sardashti, Ahmed M. Aly, Divy Agrawal, Ashish Gupta, Shiv Venkataraman 1835

PolarFS: An Ultra-low Latency and Failure Resilient Distributed File System for Shared Storage Cloud Database ..... ……. Wei Cao, Zhenjun Liu, Peng Wang, Sen Chen, Caifeng Zhu, Song Zheng, Yuhui Wang, Guoqing Ma 1849

Accordion: Better Memory Organization for LSM Key-Value Stores ..... ……Edward Bortnikov, Anastasia Braginsky, Eshcar Hillel, Idit Keidar, Gali Sheffi 1863

Real-time Constrained Cycle Detection in Large Dynamic Graphs .....
……Xiafei Qiu, Wubin Cen, Zhengping Qian, You Peng, Ying Zhang, Xuemin Lin, Jingren Zhou 1876

BTrim - Hybrid In-Memory Database Architecture for Extreme Transaction Processing in VLDBs ...
……Aditya Gurajada, Dheren Gala, Fei Zhou, Amit Pathak, Zhan-Feng Ma 1889

Demonstrations

Sherlock: A System for Interactive Summarization of Large Text Collections .....
……Avinesh P.V.S., Carsten Binnig, Benjamin Hättasch, Christian Meyer, Orkan Özyurt 1902

DataStorm-FE: A Data- and Decision-Flow and Coordination Engine for Coupled Simulation Ensembles
.....Hans Behrens, K. Selçuk Candan, Xilun Chen, Ashish Gadkari, Yash Garg, Mao-Lin Li 1906

A Demonstration of the OtterTune Automatic Database Management System Tuning Service .....
…… Bohan Zhang, Dana Van Aken,
Justin Wang, Tao Dai, Shuli Jiang, Jacky Lao, Siyuan Sheng, Andrew Pavlo, Geoffrey Gordon 1910

OctopusFS in Action: Tiered Storage Management for Data Intensive Computing .....
…… Elena Kakoulli, Nikolaos Karmiris, Herodotos Herodotou 1914

TRIPS: A System for Translating Raw Indoor Positioning Data into Visual Mobility Semantics .....
…….. Huan Li, Hua Lu, Feichao Shi, Gang Chen, Ke Chen, Lidan Shou 1918

A Demonstration of PERC: Probabilistic Entity Resolution With Crowd Errors .....
……Xiangyu Ke, Michelle Teo, Arijit Khan, Vijaya Krishna Yalavarthi 1922

CDB: A Crowd-Powered Database System .....
…… Guoliang Li, Chengliang Chai, Ju
Fan, Xueping Weng, Jian Li, Yudian Zheng, Yuanbing Li, Xiang Yu, Xiaohang Zhang, Haitao Yuan 1926

FASTER: An Embedded Concurrent Key-Value Store for State Management .....
…… Badrish Chandramouli,
Guna Prasaad, Donald Kossmann, Justin Levandoski, James Hunter, Mike Barnett 1930

Maverick: A System for Discovering Exceptional Facts from Knowledge Graphs .....
……Gensheng Zhang, Chengkai Li 1934

PTRider: A Price-and-Time-Aware Ridesharing System .....
……Lu Chen, Yunjun Gao, Zixian Liu, Xiaokui Xiao, Christian Jensen, Yifan Zhu 1938

CoreKG: a Knowledge Lake Service .....
……Amin Beheshti, Boualem Benatallah, Reza Nouri, Alireza Tabebordbar 1942

RuDiK: Rule Discovery in Knowledge Bases .....
……. Stefano Ortona, Venkata Vamsikrishna Meduri, Paolo Papotti 1946

The return of JedAI: End-to-End Entity Resolution for Structured and Semi-Structured Data .....
…… George Papadakis, Leonidas
Tsekouras, Emmanouil Thanos, George Giannakopoulos, Themis Palpanas, Manolis Koubarakis 1950

Provenance Summaries for Answers and Non-Answers .....
……..Seokki Lee, Bertram Ludäscher, Boris Glavic 1954

Helix: Accelerating Human-in-the-loop Machine Learning ..... Doris Xin, Litian Ma, Jialin Liu, Stephen Macke, Shuchen Song, Aditya Parameswaran 1958

ShapeSearch: Flexible Pattern-based Querying of Trend Line Visualizations ..... Tarique Siddiqui, Paul Luh, Zesheng Wang, Karrie Karahalios, Aditya Parameswaran 1962

PANDA: A System for Partial Topology-based Search on Large Networks ..... Miao Xie, Sourav S. Bhowmick, Hao Su, Gao Cong, Wook-Shin Han 1966

MSQL+: a Plugin Toolkit for Similarity Search under Metric Spaces in Distributed Relational Database Systems ..... Wei Lu, Xinyi Zhang, Zhiyu Shui, Zhe Peng, Xiao Zhang, Xiaoyong Du, Hao Huang, Xiaoyu Wang, Anqun Pan, Haixiang Li 1970

HYDRA: A Dynamic Big Data Regenerator ..... Anupam Sanghi, Raghav Sood, Dharmendra Singh, Jayant Haritsa, Srikanta Tirthapura 1974

A Demonstration of MAGiQ: Matrix Algebra Approach for Solving RDF Graph Queries ..... Fuad Jamour, Ibrahim Abdelaziz, Panos Kalnis 1978

REGAL+: Reverse Engineering SPJA Queries ..... Wei Chit Tan, Meihui Zhang, Hazem Elmeleegy, Divesh Srivastava 1982

NLProveNAns: Natural Language Provenance for Non-Answers ..... Daniel Deutch, Nave Frost, Tomer Haimovich, Amir Gilad 1986

Fault-Tolerance for Distributed Iterative Dataflows in Action ..... Chen Xu, Rudi Poepsel Lemaitre, Juan Soto, Volker Markl 1990

QuestPro: Queries in SPARQL Through Provenance ..... Efrat Abramovitz, Daniel Deutch, Amir Gilad 1994

GOLDRUSH: Rule Sharing System for Fraud Detection ..... Ariel Jarovsky, Tova Milo, Slava Novgorodov, Wang-Chiew Tan 1998

Discovering Diversified Paths in Knowledge Bases ..... Christian Aebeloe, Gabriela Montoya, Vinay Setty, Katja Hose 2002

Declarative and distributed graph analytics with GRADOOP ..... Martin Junghanns, Max Kießling, Niklas Teichmann, Kevin Gómez, André Petermann, Erhard Rahm 2006

A collaborative framework for tweaking properties in a synthetic dataset ..... Jiangwei Zhang, Yu Wang, Y.C. Tay 2010

Tooling Framework for Instantiating Natural Language Querying System ..... Manasa Jammi, Jaydeep Sen, Ashish Mittal, Sagar Verma, Vardaan Pahuja, Rema Ananthanarayanan, Pranay Lohia, Hima Karanam, Diptikalyan Saha, Karthik Sankaranarayanan 2014

Koko: A System for Scalable Semantic Querying of Text ..... Xiaolan Wang, Jiyu Komiya, Yoshihiko Suhara, Aaron Feng, Behzad Golshan, Alon Halevy, Wang-Chiew Tan 2018

GC: A Graph Caching System for Subgraph/Supergraph Queries ..... Jing Wang, Zichen Liu, Shuai Ma, Nikos Ntarmos, Peter Triantafillou 2022

X2Q: Your Personal Example-based Graph Explorer ..... Matteo Lissandrini, Davide Mottin, Themis Palpanas, Yannis Velegrakis 2026

ConnectionLens: Finding Connections Across Heterogeneous Data Sources ..... Camille Chanial, Rédouane Dziri, Helena Galhardas, Julien Leblay, Minh-Huong Le Nguyen, Ioana Manolescu 2030

ProvSQL:Provenance and Probability Management in PostgreSQL ..... Pierre Senellart, Louis Jachiet, Silviu Maniu, Yann Ramusat 2034

CYADB: A Database that Covers Your Ask ..... Zechao Shang, Will Brackenbury, Aaron Elmore, Michael Franklin 2038

CloudMatcher: A Hands-Off Cloud/Crowd Service for Entity Matching ..... Yash Govind, Erik Paulson, Palaniappan Nagarajan, Paul Suganthan G. C., Anhai Doan, Youngchoon Park, Glenn Fung, Devin Conathan, Marshall Carter, Mingju Sun 2042

Collaborative Edge and Cloud Neural Networks for Real-Time Video Processing ..... Philipp Grulich, Faisal Nawab 2046

Dhalion in Action: Automatic Management of Streaming Applications ..... Ashvin Agrawal, Avrilia Floratou 2050

Ease.ml in Action: Towards Multi-tenant Declarative Learning Services ..... Bojan Karlaš, Ji Liu, Wentao Wu, Ce Zhang 2054

MustaCHE: A Multiple Clustering Hierarchies Explorer ..... Antonio Cavalcante Araujo Neto, Mario Nascimento, Joerg Sander, Ricardo Campello 2058

HypDB: A Demonstration of Detecting, Explaining and Resolving Bias in OLAP queries ..... Babak Salimi, Corey Cole, Peter Li, Johannes Gehrke, Dan Suciu 2062

Learning Efficiently Over Heterogeneous Databases ..... Jose Picado, Arash Termehchy, Sudhanshu Pathak 2066

Scalable and Efficient Data Analytics and Mining with Lemonade ..... Walter Santos, Gustavo Avelar, Manoel Horta Ribeiro, Dorgival Guedes, Wagner Meira Jr. 2070

SkinnerDB: Regret-Bounded Query Evaluation via Reinforcement Learning ..... Immanuel Trummer, Samuel Moseley, Deepak Maram, Saehan Jo, Joseph Antonakakis 2074

iSPEED: a Scalable and Distributed In-Memory Based Spatial Query System for Large and Structurally Complex 3D Data ..... Hoang Vo, Yanhui Liang, Jun Kong, Fusheng Wang 2078

DfAnalyzer: Runtime Dataflow Analysis of Scientific Applications using Provenance ..... Vítor Sousa, Daniel Oliveira, Marta Mattoso, Patrick Valduriez 2082

A Demonstration of Sterling: A Privacy-Preserving Data Marketplace ..... Nick Hynes, David Dao, David Yan, Raymond Cheng, Dawn Song 2086

ConTPL: Controlling Temporal Privacy Leakage in Differentially Private Continuous Data Release ..... Yang Cao, Li Xiong, Masatoshi Yoshikawa, Yonghui Xiao, Si Zhang 2090

Tutorials

Data Integration and Machine Learning: A Natural Synergy ..... Luna Dong, Theodoros Rekatsinas 2094

Database and Distributed Computing Fundamentals for Scalable, Fault-tolerant, and Consistent Maintenance of Blockchains ..... Sujaya Maiyya, Victor Zakhary, Divy Agrawal, Amr El Abbadi 2098

Forecasting Big Time Series: Old and New ..... Christos Faloutsos, Jan Gasthaus, Tim Januschowski, Yuyang Wang 2102

Graph Data Models, Query Languages and Programming Paradigms ..... Alin Deutsch, Yannis Papakonstantinou 2106

Computational fact-checking: a content management perspective ..... Sylvie Cazalens, Julien Leblay, Ioana Manolescu, Philippe Lamarre, Xavier Tannier 2110

Award Talks and Panel

Information and Data Management at PUC-Rio and UFMG ..... Antonio Furtado, Nivio Ziviani 2114

Open Data Integration ..... Renee Miller 2130

Ten Years of WebTables ..... Michael Cafarella, Alon Halevy, Hongrae Lee, Jayant Madhavan, Cong Yu, Daisy Zhe Wang, Eugene Wu 2140

Northstar: An Interactive Data Science System ..... Tim Kraska 2150

Panel: A Debate on Data and Algorithmic Ethics ..... Julia Stoyanovich, Bill Howe, H. Jagadish, Gerome Miklau 2165

TABLE OF CONTENTS

Research Papers

A Comparative Evaluation of Systems for Scalable Linear Algebra-based Analytics 2168
……Anthony Thomas, Arun Kumar

A Concave Path to Low-overhead Robust Query Processing 2183
……Srinivas Karthik Venkatesh, Jayant Haritsa, Sreyash Kenkre, Vinayaka Pandit

Interactive Summarization and Exploration of Top Aggregate Query Answers 2196
……Yuhao Wen, Xiaodan Zhu, Sudeepa Roy, Jun Yang

Everything You Always Wanted to Know About Compiled and Vectorized Queries But Were Afraid to Ask 2209
……Timo Kersten, Viktor Leis, Alfons Kemper, Thomas Neumann, Andrew Pavlo, Peter Boncz

Durable Top-k Queries on Temporal Data 2223
……Junyang Gao, Pankaj Agarwal, Jun Yang

Scalable, Variable-Length Similarity Search in Data Series: The ULISSE Approach..... 2236
……Michele Linardi, Themis Palpanas

FineLine: log-structured transactional storage and recovery 2249
……Caetano Sauer, Goetz Graefe, Theo Härder

ICARUS: Minimizing Human Effort in Iterative Data Completion 2263
……Protiva Rahman, Courtney Hebert, Arnab Nandi