

45th International Conference on Very Large Data Bases (VLDB 2019)

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

Los Angeles, California, USA
26-30 August 2019

Part 1 of 3

Editors:

Lei Chen
Fatma Özcan

ISBN: 978-1-5108-9278-1

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© (2019) 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 with permission 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

Research Papers

List Intersection for Web Search: Algorithms, Cost Models, and Optimizations	1
.....Sunghwan Kim, Taesung Lee, Seungwon Hwang, Sameh Elnikety	
Interactive Checks for Coordination Avoidance	14
..... Michael Whittaker, Joseph M. Hellerstein	
Pigeonring: A Principle for Faster Thresholded Similarity Search	28
..... Jianbin Qin, Chuan Xiao	
Local Algorithms for Hierarchical Dense Subgraph Discovery	43
..... Ahmet Erdem Sariyuce, C. Seshadhri, Ali Pinar	
Cost-Effective Data Annotation using Game-Based Crowdsourcing	57
..... Jingru Yang, Ju Fan, Zhewei Wei, Guoliang Li, Tongyu Liu, Xiaoyong Du	
Optimization for Active Learning-based Interactive Database Exploration.....	71
.....Enhui Huang, Liping Peng, Luciano Di Palma, Ahmed Abdelkafi, Anna Liu, Yanlei Diao	

TABLE OF CONTENTS

Research Papers

Exploring Change - A New Dimension of Data Analytics	
Tobias Bleifuß, Leon Bornemann, Theodore Johnson, Dmitri V. Kalashnikov, Felix Naumann, Divesh Srivastava	85
The Flexible Socio Spatial Group Queries	
Bishwamitra Ghosh, Mohammed Eunus Ali, Farhana M. Choudhury, Sajid Hasan, Timos Sellis, Jianxin Li.....	99
The Lernaean Hydra of Data Series Similarity Search: An Experimental Evaluation of the State of the Art.....	
Karima Echihabi, Kostas Zoumpatianos, Themis Palpanas, Houda Benbrahim	112
Rafiki: Machine Learning as an Analytics Service System.....	
Wei Wang, Sheng Wang, Jinyang Gao, Meihui Zhang, Gang Chen, Teck Khim Ng, Beng Chin Ooi, Jie Shao.....	128
Automatic Index Selection for Large-Scale Datalog Computation	
.....Pavle Subotic, Herbert Jordan, Lijun Chang, Alan Fekete, Bernhard Scholz	141
Start Late or Finish Early: A Distributed Graph Processing System with Redundancy Reduction ...	
.....Shuang Song, Xu Liu, Qinzhe Wu, Andreas Gerstlauer, Tao Li, Lizy K. John	154
Improving Optimistic Concurrency Control Through Transaction Batching and Operation Reordering	
..... Bailu Ding, Lucja Kot, Johannes Gehrke	169

TABLE OF CONTENTS

Research Papers

Query Log Compression for Workload Analytics	Ting Xie, Varun Chandola, Oliver Kennedy	183
The Maximum Trajectory Coverage Query in Spatial Databases	Mohammed Eunus Ali, Shadman Saqib Eusuf, Kaysar Abdullah, Farhana M. Choudhury, J. Shane Culpepper, Timos Sellis	197
Towards a Learning Optimizer for Shared Clouds	Chenggang Wu, Alekh Jindal, Saeed Amizadeh, Hiren Patel, Wangchao Le, Shi Qiao, Sriram Rao	210
Snuba: Automating Weak Supervision to Label Training Data	Paroma Varma, Christopher Re	223
On Obtaining Stable Rankings ...	Abolfazl Asudeh, H. Jagadish, Gerome Miklau, Julia Stoyanovich	237
PS-Tree-based Efficient Boolean Expression Matching for High Dimensional and Dense Workloads	Shuping Ji, Hans-Arno Jacobsen	251
SWIFT: Mining Representative Patterns from Large Event Streams ..	Yizhou Yan, Lei Cao, Samuel Madden, Elke Rundensteiner	265
Smurf: Self-Service String Matching Using Random Forests	Paul Suganthan G. C., Adel Ardalan, Anhai Doan, Aditya Akella	278
Chasing Similarity: Distribution-aware Aggregation Scheduling	Feilong Liu, Ario Salmasi, Spyros Blanas, Anastasios Sidiropoulos	292
ShrinkWrap: Efficient SQL Query Processing in Differentially Private Data Federations	Johes Bater, Xi He, William Ehrich, Ashwin Machanavajjhala, Jennie Rogers	307

TABLE OF CONTENTS

Research Papers

A Study of Partitioning Policies for Graph Analytics on Large-scale Distributed Platforms.....	321
.....Gurbinder Gill, Roshan Dathathri, Loc Hoang, Keshav Pingali	
Utility-Driven Graph Summarization	335
..... K. Ashwin Kumar, Petros Efstathopoulos	
ColumnML: Column-Store Machine Learning with On-The-Fly Data Transformation.....	348
..... Kaan Kara, Ken Eguro, Ce Zhang, Gustavo Alonso	
Cost-efficient Data Acquisition on Online Data Marketplaces for Correlation Analysis.....	362
..... Yanying Li, Haipei Sun, Boxiang Dong, Hui (Wendy) Wang	
Cleaning Crowdsourced Labels Using Oracles For Statistical Classification	376
.....Mohamad Dolatshah, Mathew Teoh, Jiannan Wang, Jian Pei	
Beyond Macrobenchmarks: Microbenchmark-based Graph Database Evaluation	390
..... Matteo Lissandrini, Martin Brugnara, Yannis Velegarakis	
IPA: Invariant-preserving Applications for Weakly consistent Replicated Databases	404
..... Valter Balegas, Sérgio Duarte, Carla Ferreira, Rodrigo Rodrigues, Nuno Preguiça	
DIFF: A Relational Interface for Large-Scale Data Explanation.....	419
Firas Abuzaid, Peter Kraft, Sahaana Suri, Edward Gan, Eric Xu, Atul Shenoy, Asvin Anathanaraya, John Sheu, Erik Meijer, Xi Wu, Jeff Naughton, Peter Bailis, Matei Zaharia.....	
Stream Frequency Over Interval Queries.....	433
.....Ran Ben Basat, Roy Friedman, Rana Shahout	
Helix: Holistic Optimization for Accelerating Iterative Machine Learning	446
.....Doris Xin, Stephen Macke, Litian Ma, Jialin Liu, Shuchen Song, Aditya Parameswaran	

TABLE OF CONTENTS

Research Papers

Fast Approximate Nearest Neighbor Search With The Navigating Spreading-out Graph	461
.....Cong Fu, Chao Xiang, Changxu Wang, Deng Cai	
Document Reordering for Faster Intersection	475
.....Qi Wang, Torsten Suel	
Correlation Constraint Shortest Path over Large Multi-Relation Graphs	488
.....Xiaofei Zhang, Tamer Özsu	
Performance-Optimal Filtering: Bloom overtakes Cuckoo at High-Throughput.....	502
.....Harald Lang, Thomas Neumann, Alfons Kemper, Peter Boncz	
Analyzing Efficient Stream Processing on Modern Hardware	516
Steffen Zeuch, Sebastian Breß Tilmann Rabl, Bonaventura Del Monte, Jeyhun Karimov, Clemens Lutz, Manuel Renz, Jonas Traub, Volker Markl	
Efficient Data Ingestion and Query Processing for LSM-Based Storage Systems.....	531
.....Chen Luo, Michael Carey	
HetExchange: Encapsulating heterogeneous CPU–GPU parallelism in JIT compiled engines.....	544
.....Periklis Chrysogelos, Manos Karpathiotakis, Raja Appuswamy, Anastasia Ailamaki	
Meta-Mappings for Schema Mapping Reuse	557
.....Paolo Atzeni, Luigi Bellomarini, Paolo Papotti, Riccardo Torlone	
An Experimental Evaluation of Garbage Collectors on Big Data Applications	570
.....Lijie Xu, Tian Guo, Wensheng Dou, Wei Wang, Jun Wei	
Adaptive Optimistic Concurrency Control for Heterogeneous Workloads	584
.....Jinwei Guo, Peng Cai, Jiahao Wang, Weining Qian, Aoying Zhou	
MgCrab: Transaction Crabbing for Live Migration in Deterministic Database Systems.....	597
.....Yu-Shan Lin, Shao-Kan Pi, Meng-Kai Liao, Ching Tsai, Aaron Elmore, Shan-Hung Wu	
Unifying Consensus and Atomic Commitment for Effective Cloud Data Management	611
.....Sujaya Maiyya, Faisal Nawab, Divy Agrawal, Amr El Abbadi	

TABLE OF CONTENTS

Research Papers

Autoscaling Tiered Cloud Storage in Anna	624
..... Chenggang Wu, Vikram Sreekanti, Joseph Hellerstein	
Snapshot Semantics for Temporal Multiset Relations.....	639
..... Anton Dignć, Boris Glavic, Xing Niu, Johann Gamper, Michael Böhlen	
Certus: An Effective Entity Resolution Approach with Graph Differential Dependencies (GDDs)....	653
..... Selasi Kwashie, Jixue Liu, Jiuyong Li, Lin Liu, Markus Stumptner, Lujing Yang	
Efficient and Effective Algorithms for Clustering Uncertain Graphs	667
..... Kai Han, Fei Gui, Xiaokui Xiao, Jing Tang, Yuntian He, Zongmai Cao, He Huang	
Pangea: Monolithic Distributed Storage for Data Analytics.....	681
..... Jia Zou, Arun Iyengar, Chris Jermaine	
Scaling-Up In-Memory Datalog Processing: Observations and Techniques.....	695
..... Zhiwei Fan, Jianqiao Zhu, Zuyu Zhang, Aws Albarghouthi, Paraschos Koutris, Jignesh Patel	
Cache-aware load balancing of data center applications.....	709
.Aaron Archer, Kevin Aydin, Mohammadhossein Bateni, Vahab Mirrokni, Aaron Schild, Ray Yang, Richard Zhuang	

TABLE OF CONTENTS

Research Papers

Minimizing Cost by Reducing Scaling Operations in Distributed Stream Processing..... Michael Borkowski, Christoph Hochreiner, Stefan Schulte	724
ProvCite: Provenance-based Data Citation..... Yinjun Wu, Abdussalam Alawini, Daniel Deutch, Tova Milo, Susan Davidson	738
Deducing Certain Fixes to Graphs..... Wenfei Fan, Ping Lu, Chao Tian, Jingren Zhou	752
Solving k-center Clustering (with Outliers) in MapReduce and Streaming, almost as Accurately as Sequentially Matteo Ceccarello, Andrea Pietracaprina, Geppino Pucci	766
Explain3D: Explaining Disagreements in Disjoint Datasets..... Xiaolan Wang, Alexandra Meliou	779
DASH: Database Shadowing for Mobile DBMS Youjip Won, Sundoo Kim, Juseong Yun, Damquang Tuan, Jiwon Seo	793
Accelerating Generalized Linear Models with MLWeaving: A One-Size-Fits-All System for Any-precision Learning Zeke Wang, Kaan Kara, Hantian Zhang, Gustavo Alonso, Ce Zhang, Onur Mutlu	807
Declarative Recursive Computation on an RDBMS..... .. Dimitrije Jankov, Shangyu Luo, Binhang Yuan, Zhuhua Cai, Jia Zou, Chris Jermaine, Zekai Gao	822

TABLE OF CONTENTS

Research Papers

Design, Implementation, and Evaluation of Write-Back Policy with Cache Augmented Data Stores <i>Shahram Ghandeharizadeh, Hieu Nguyen</i>	836
User Guidance for Efficient Fact Checking <i>Thanh Tam Nguyen, Hongzhi Yin, Matthias Weidlich, Bolong Zheng, Quoc Viet Hung Nguyen, Bela Stantic</i>	850
An In-Depth Comparison of s-t Reliability Algorithms over Uncertain Graphs <i>Xiangyu Ke, Arijit Khan, Leroy Lim</i>	864
Dynamic Scaling for Parallel Graph Computations <i>Wenfei Fan, Chunming Hu, Muyang Liu, Ping Lu, Qiang Yin, Jingren Zhou</i>	877
TopoX: Topology Refactorization for Efficient Graph Partitioning and Processing <i>Yiming Zhang, Dongsheng Li, Jinyan Wang, Kian-Lee Tan</i>	891
Multi-Dimensional Balanced Graph Partitioning via Projected Gradient Descent..... <i>Dmitrii Avdiukhin, Sergey Pupyrev, Grigory Yaroslavtsev</i>	906
Efficient Discovery of Sequence Outlier Patterns <i>Lei Cao, Yizhou Yan, Samuel Madden, Elke Rundensteiner, Mathan Gopalsamy</i>	920
A Comparative Evaluation of Order-Revealing Encryption Schemes and Secure Range-Query Protocols <i>Dmytro Bogatov, George Kollios, Leo Reyzin</i>	933

TABLE OF CONTENTS

Research Papers

k/2-hop: Fast Mining of Convoy Patterns With Effective Pruning	948
<i>Faisal Orakzai, Toon Calders, Torben Pedersen</i>	
Balance-Aware Distributed String Similarity-Based Query Processing System	961
<i>Ji Sun, Zeyuan Shang, Guoliang Li, Zhifeng Bao, Dong Deng</i>	
Fine-Grained, Secure and Efficient Data Provenance for Blockchain	975
<i>Pingcheng Ruan, Gang Chen, Anh Dinh, Qian Lin, Beng Chin Ooi, Meihui Zhang</i>	
Progressive Top-k Subarray Query Processing in Array Databases	989
<i>Dalsu Choi, Chang-Sup Park, Yon Dohn Chung</i>	
Megaphone: Latency-conscious state migration for distributed streaming dataflows	1002
<i>Moritz Hoffmann, Andrea Lattuada, Frank Mcsherry, Vasiliki Kalavri, John Liagouris, Timothy Roscoe</i>	
From Anomaly Detection to Rumour Detection using Data Streams of Social Platforms	1016
<i>Thanh Tam Nguyen, Matthias Weidlich, Bolong Zheng, Hongzhi Yin, Quoc Viet Hung Nguyen, Bela Stantic</i>	
Obscure: Information-Theoretic Oblivious and Verifiable Aggregation Queries	1030
<i>Peeush Gupta, Yin Li, Sharad Mehrotra, Nisha Panwar, Shantanu Sharma, Sumaya Almanee</i>	
Selectivity Estimation for Range Predicates using Lightweight Models	1044
<i>Anshuman Dutt, Chi Wang, Azade Nazi, Srikanth Kandula, Vivek Narasayya, Surajit Chaudhuri</i>	

TABLE OF CONTENTS

Research Papers

Constrained Shortest Path Query in a Large Time-Dependent Graph.....	1058
<i>Ye Yuan, Xiang Lian, Guoren Wang, Yuliang Ma, Yishu Wang</i>	
Finding Theme Communities from Database Networks.....	1071
<i>Linyang Chu, Zhefeng Wang, Jian Pei, Yanyan Zhang, Yu Yang, Enhong Chen</i>	
Ridesharing: Simulator, Benchmark, and Evaluation.....	1085
<i>James Pan, Guoliang Li, Juntao Hu</i>	
Distributed Subgraph Matching on Timely Dataflow	1099
<i>Longbin Lai, Zhu Qing, Zhengyi Yang, Xin Jin, Zhengmin Lai, Ran Wang, Kongzhang Hao, Xuemin Lin, Lu Qin, Wenjie Zhang, Ying Zhang, Zhengping Qian, Jingren Zhou</i>	
Hyper Dimension Shuffle: Efficient Data Repartition at Petabyte Scale in Scope.....	1113
<i>Shi Qiao, Adrian Nicoara, Jin Sun, Marc Friedman, Hiren Patel, Jaliya Ekanayake</i>	
Answering Range Queries Under Local Differential Privacy	1126
<i>Graham Cormode, Tejas Kulkarni, Divesh Srivastava</i>	
Vertex Priority Based Butterfly Counting for Large-scale Bipartite Networks.....	1139
<i>Kai Wang, Xuemin Lin, Lu Qin, Wenjie Zhang, Ying Zhang</i>	
Block as a Value for SQL over NoSQL.....	1153
<i>Yang Cao, Wenfei Fan, Tengfei Yuan</i>	
Optimal and General Out-of-Order Sliding-Window Aggregation	1167
<i>Kanat Tangwongsan, Martin Hirzel, Scott Schneider</i>	
Creating Top Ranking Options in the Continuous Option and Preference Space	1181
<i>Bo Tang, Kyriakos Mouratidis, Man Lung Yiu, Zhenyu Chen</i>	
Ontology-based Entity Matching in Attributed Graphs	1195
<i>Hanchao Ma, Morteza Alipourlangouri, Yinghui Wu, Fei Chiang, Jiaying Pi</i>	
Real-time Distributed Co-Movement Pattern Detection on Streaming Trajectories	1208
<i>Lu Chen, Yunjun Gao, Ziquan Fang, Xiaoye Miao, Christian Jensen, Chenjuan Guo</i>	
iBTune: Individualized Buffer Tuning for Large-scale Cloud Databases	1221
<i>Jian Tan, Tieying Zhang, Feifei Li, Jie Chen, Qixing Zheng, Ping Zhang, Honglin Qiao, Yue Shi, Wei Cao, Rui Zhang</i>	

TABLE OF CONTENTS

Research Papers

Online Template Induction for Machine-Generated Emails.....	1235
<i>Michael J. Whittaker, Nick Edmonds, Sandeep Tata, James B. Wendt, Marc Najork</i>	
Querying Shortest Paths on Time Dependent Road Networks	1249
<i>Yong Wang, Guoliang Li, Nan Tang</i>	
Example-Driven Query Intent Discovery: Abductive Reasoning using Semantic Similarity	1262
<i>Anna Fariha, Alexandra Meliou</i>	
Automated Verification of Query Equivalence Using Satisfiability Modulo Theories	1276
<i>Qi Zhou, Joy Arulraj, Shamkant Navathe, William Harris, Dong Xu</i>	
Towards a Unified Framework for String Similarity Joins	1289
<i>Pengfei Xu, Jiaheng Lu</i>	
NETS: Extremely Fast Outlier Detection from a Data Stream via Set-Based Processing.....	1303
<i>Susik Yoon, Jae-Gil Lee, Byung Suk Lee</i>	
STAR: Scaling Transactions through Asymmetric Replication	1316
<i>Yi Lu, Xiangyao Yu, Samuel Madden</i>	
Subjective Databases.....	1330
<i>Yuliang Li, Aaron Feng, Jinfeng Li, Saran Mumick, Alon Halevy, Vivian Li, Wang-Chiew Tan</i>	
Fast and Robust Distributed Subgraph Enumeration	1344
<i>Xuguang Ren, Junhu Wang, Wook-Shin Han, Jeffrey Xu Yu</i>	
An Experimental Evaluation of Large Scale GBDT Systems	1357
<i>Fangcheng Fu, Jiawei Jiang, Yingxia Shao, Bin Cui</i>	
PrivateSQL: A Differentially Private SQL Query Engine	1371
<i>Ios Kotsogiannis, Yuchao Tao, Xi He, Maryam Fanaeepour, Ashwin Machanavajjhala, Michael Hay, Jerome Miklau</i>	
CAPER: A Cross-Application Permissioned Blockchain.....	1385
<i>Mohammad Javad Amiri, Divyakant Agrawal, Amr El Abbadi</i>	
Crossbow: Scaling Deep Learning with Small Batch Sizes on Multi-GPU Servers.....	1399
<i>Alexandros Kolios, Pijika Watcharapichat, Matthias Weidlich, Luo Mai, Paolo Costa, Peter Pietzuch</i>	
Finding Attribute-Aware Similar Region for Data Analysis	1414
<i>Kaiyu Feng, Gao Cong, Christian S. Jensen, Tao Guo</i>	
Intermittent Query Processing.....	1427
<i>Dixin Tang, Zechao Shang, Aaron J. Elmore, Sanjay Krishnan, Michael J. Franklin</i>	

Hillview: A trillion-cell spreadsheet for big data	1442
<i>Mihai Budiu, Parikshit Gopalan, Lalith Suresh, Udi Wieder, Han Krueger, Marcos K. Aguilera</i>	
Embedded Functional Dependencies and Data-completeness Tailored Database Design.....	1458
<i>Ziheng Wei, Sebastian Link</i>	
Ocean Vista: Gossip-Based Visibility Control for Speedy Geo-Distributed Transactions.....	1471
<i>Hua Fan, Wojciech Golab</i>	
An IDEA: An Ingestion Framework for Data Enrichment in AsterixDB	1485
<i>Xikui Wang, Michael Carey</i>	
DimmStore: Memory Power Optimization for Database Systems	1499
<i>Alexey Karyakin, Kenneth Salem</i>	
Generating Application-specific Data Layouts for In-memory Databases.....	1513
<i>Cong Yan, Alvin Cheung</i>	
Rewriting of Plain SO Tgds into Nested Tgds	1526
<i>Rihan Hai, Christoph Quix</i>	
Blockchain Meets Database: Design and Implementation of a Blockchain Relational Database.....	1539
<i>Senthil Nathan, Chander Govindarajan, Adarsh Saraf, Manish Sethi, Praveen Jayachandran</i>	
An Intermediate Representation for Optimizing Machine Learning Pipelines.....	1553
<i>Andreas Kunft, Asterios Katsifodimos, Sebastian Schelter, Sebastian Breß Tilmann Rabl, Volker Markl</i>	
Accelerating Raw Data Analysis with the ACCORDA Software and Hardware Architecture.....	1568
<i>Yuanwei Fang, Chen Zou, Andrew Chien</i>	
Comparing Synopsis Techniques for Approximate Spatial Data Analysis	1583
<i>A. B. Siddique, Ahmed Eldawy, Vagelis Hristidis</i>	
BlockchainDB - A Shared Database on Blockchains	1597
<i>Muhammad El-Hindi, Carsten Binnig, Arvind Arasu, Donald Kossmann, Ravi Ramamurthy</i>	
Efficient Task-Specific Data Valuation for Nearest Neighbor Algorithms	1610
<i>Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nezihe Merve Gürel, Bo Li, Ce Zhang, Costas J. Spanos, Dawn Song</i>	
Distributed Implementations of Dependency Discovery Algorithms	1624
<i>Hemant Saxena, Lukasz Golab, Ihab F. Ilyas</i>	
Rethinking Database High Availability with RDMA Networks	1637
<i>Erfan Zamanian, Xiangyao Yu, Michael Stonebraker, Tim Kraska</i>	
Motivo: Fast Motif Counting via Succinct Color Coding and Adaptive Sampling.....	1651
<i>Marco Bressan, Stefano Leucci, Alessandro Panconesi</i>	
Arx: An Encrypted Database using Semantically Secure Encryption.....	1664
<i>Rishabh Poddar, Tobias Boelter, Raluca Ada Popa</i>	
Efficient Knowledge Graph Accuracy Evaluation	1679
<i>Junyang Gao, Xian Li, Yifan Ethan Xu, Bunyamin Sisman, Xin Luna Dong, Jun Yang</i>	
Optimizing Subgraph Queries by Combining Binary and Worst-Case Optimal Joins.....	1692

Amine Mhedhbi, Semih Salihoglu

Neo: A Learned Query Optimizer	1705
<i>Ryan C. Marcus, Parimarjan Negi, Hongzi Mao, Chi Zhang, Mohammad Alizadeh, Tim Kraska, Olga Papaemmanouil, Nesime Tatbul</i>	
Efficient Algorithms for Densest Subgraph Discovery	1719
<i>Yixiang Fang, Kaiqiang Yu, Reynold Cheng, Laks V.s. Lakshmanan, Xuemin Lin</i>	
Plan-Structured Deep Neural Network Models for Query Performance Prediction	1733
<i>Ryan C. Marcus, Olga Papaemmanouil</i>	
SLOG: Serializable, Low-latency, Geo-replicated Transactions.....	1747
<i>Kun Ren, Dennis Li, Daniel J. Abadi</i>	
GRAIL: Efficient Time-Series Representation Learning.....	1762
<i>John Paparrizos, Michael Franklin</i>	

TABLE OF CONTENTS

Demonstrations

GALO: Guided Automated Learning for re-Optimization.....	1778
<i>Guilherme Damasio, Spencer Bryson, Vincent Corvinelli, Parke Godfrey, Piotr Mierzejewski, Jaroslaw Szlichta, Calisto Zuzarte</i>	
Synergistic Graph and SQL Analytics Inside IBM Db2.....	1782
<i>Yuanyuan Tian, Sui Jun Tong, Mir Hamid Pirahesh, Wen Sun, En Liang Xu, Wei Zhao</i>	
Cleanits: A Data Cleaning System for Industrial Time Series.....	1786
<i>Xiaou Ding, Hongzhi Wang, Jiakuan Su, Zijue Li, Jianzhong Li, Hong Gao</i>	
ITAA: An Intelligent Trajectory-driven Outdoor Advertising Deployment Assistant.....	1790
<i>Yipeng Zhang, Zhifeng Bao, Songsong Mo, Yuchen Li, Yanghao Zhou</i>	
SystemER: A Human-in-the-loop System for Explainable Entity Resolution.....	1794
<i>Kun Qian, Lucian Popa, Prithviraj Sen</i>	
Buckle: Evaluating Fact Checking Algorithms Built on Knowledge Bases.....	1798
<i>Viet-Phi Huynh, Paolo Papotti</i>	
A Query System for Efficiently Investigating Complex Attack Behaviors for Enterprise Security	1802
<i>Peng Gao, Xusheng Xiao, Zhichun Li, Kangkook Jee, Fengyuan Xu, Sanjeev R. Kulkarni, Prateek Mittal</i>	
CAPE: Explaining Outliers by Counterbalancing.....	1806
<i>Zhengjie Miao, Qitian Zeng, Chenjie Li, Boris Glavic, Oliver Kennedy, Sudeepa Roy</i>	
BlackMagic: Automatic Inlining of Scalar UDFs into SQL Queries with Froid.....	1810
<i>Karthik Ramachandra, Kwanghyun Park</i>	
ProgressiveDB - Progressive Data Analytics as a Middleware.....	1814
<i>Lukas Berg, Tobias Ziegler, Carsten Binnig, Uwe Röhm</i>	
doppioDB 2.0: Hardware Techniques for Improved Integration of Machine Learning into Databases.....	1818
<i>Kaan Kara, Zeke Wang, Ce Zhang, Gustavo Alonso</i>	
COVIZ: A System for Visual Formation and Exploration of Patient Cohorts.....	1822
<i>Cícero A. L. Pahins, Behrooz Omidvar-Tehrani, Sihem Amer-Yahia, Valérie Siroux, Jean-Louis Pepin, Jean-Christian Borel, João Comba</i>	
PRIMAT: A Toolbox for Fast Privacy-preserving Matching.....	1826
<i>Martin Franke, Ziad Sehili, Erhard Rahm</i>	
NashDB: Fragmentation, Replication, and Provisioning using Economic Methods.....	1830

Ryan Marcus, Chi Zhang, Shuai Yu, Geoffrey Kao, Olga Papaemmanouil

Flash in Action: Scalable Spatial Data Analysis Using Markov Logic Networks.....	1834
<i>Ibrahim Sabek, Mashaal Musleh, Mohamed F. Mokbel</i>	
I Can't Believe It's Not (Only) Software! Bionic Distributed Storage for Parquet Files	1838
<i>Lucas Kuhring, Zsolt István</i>	
WISE: Vehicle Image Search Engine with Traffic Camera	1842
<i>Hyewon Choi, Erkang Zhu, Arsala Bangash, Renée J. Miller</i>	
WiClean: A System for Fixing Wikipedia Interlinks Using Revision History Patterns	1846
<i>WiClean: A System for Fixing Wikipedia Interlinks Using Revision History Patterns</i>	
SparkCruise: Handsfree Computation Reuse in Spark.....	1850
<i>Abhishek Roy, Alekh Jindal, Hiren Patel, Ashit Gosalia, Subru Krishnan, Carlo Curino</i>	
In-database Distributed Machine Learning: Demonstration using Teradata SQL Engine.....	1854
<i>Sandeep Singh Sandha, Wellington Cabrera, Mohammed Al-Kateb, Sanjay Nair, Mani Srivastava</i>	
SHOAL: Large-scale Hierarchical Taxonomy via Graph-based Query Coalition in E-commerce.....	1858
<i>Zhao Li, Xia Chen, Xuming Pan, Pengcheng Zou, Yuchen Li, Guoxian Yu</i>	
DPSaaS: Multi-Dimensional Data Sharing and Analytics as Services under Local Differential Privacy	1862
<i>Min Xu, Tianhao Wang, Bolin Ding, Jingren Zhou, Cheng Hong, Zhicong Huang</i>	
PriSTE: Protecting Spatiotemporal Event Privacy in Continuous Location-Based Services.....	1866
<i>Yang Cao, Yonghui Xiao, Li Xiong, Liqun Bai, Masatoshi Yoshikawa</i>	
Datalignment: Ontology Schema Alignment Through Datalog Containment	1870
<i>Daniel Deutch, Evgeny Marants, Yuval Moskovitch</i>	
IHCS: An Integrated Hybrid Cleaning System.....	1874
<i>Congcong Ge, Yunjun Gao, Xiaoye Miao, Lu Chen, Christian S. Jensen, Ziyuan Zhu</i>	
CAPRIO: Graph-based Integration of Indoor and Outdoor Data for Path Discovery.....	1878
<i>Constantinos Costa, Xiaoyu Ge, Panos K. Chrysanthis</i>	
HERMIT in Action: Succinct Secondary Indexing Mechanism via Correlation Exploration	1882
<i>Yingjun Wu, Jia Yu, Yuanyuan Tian, Richard Sidle, Ronald Barber</i>	
DISPERS: Securing Highly Distributed Queries on Personal Data Management Systems	1886
<i>Julien Loudet, Iulian Sandu-Popa, Luc Bouganim</i>	
Stateful Functions as a Service in Action	1890
<i>Adil Akhter, Marios Fragkoulis, Asterios Katsifodimos</i>	
Demonstration of Krypton: Optimized CNN Inference for Occlusion-based Deep CNN Explanations	1894
<i>Allen Ordookhanians, Xin Li, Supun Nakandala, Arun Kumar</i>	
LensXPlain: Visualizing and Explaining Contributing Subsets for Aggregate Query Answers	1898
<i>Zhengjie Miao, Andrew Lee, Sudeepa Roy</i>	
Juneau: Data Lake Management for Jupyter	1902
<i>Yi Zhang, Zachary G. Ives</i>	

ApproxML: Efficient Approximate Ad-Hoc ML Models Through Materialization and Reuse	1906
<i>Sona Hasani, Faezeh Ghaderi, Shohedul Hasan, Saravanan Thirumuruganathan, Abolfazl Asudeh, Nick Koudas, Gautam Das</i>	
Flare & Lantern: Efficiently Swapping Horses Midstream	1910
<i>Grégory Essertel, Ruby Y. Tahboub, Fei Wang, James Decker, Tiark Rompf</i>	
Trinity: An Extensible Synthesis Framework for Data Science.....	1914
<i>Ruben Martins, Jia Chen, Yanju Chen, Yu Feng, Isil Dillig</i>	
PSynDB: Accurate and Accessible Private Data Generation	1918
<i>Zhiqi Huang, Ryan Mckenna, George Bissias, Gerome Miklau, Michael Hay, Ashwin Machanavajjhala</i>	
FishStore: Fast Ingestion and Indexing of Raw Data.....	1922
<i>Badrish Chandramouli, Dong Xie, Yinan Li, Donald Kossmann</i>	
Spade: A Modular Framework for Analytical Exploration of RDF Graphs	1926
<i>Yanlei Diao, Pawel Guzewicz, Ioana Manolescu, Mirjana Mazuran</i>	
Making an RDBMS Data Scientist Friendly: Advanced In-database Interactive Analytics with Visualization Support	1930
<i>Joseph Vinish D'silva, Florestan De Moor, Bettina Kemme</i>	
UDAO: A Next-Generation Unified Data Analytics Optimizer.....	1934
<i>Khaled Zaouk, Fei Song, Chenghao Lyu, Arnab Sinha, Yanlei Diao, Prashant Shenoy</i>	
AggChecker: A Fact-Checking System for Text Summaries of Relational Data Sets	1938
<i>Saeahan Jo, Immanuel Trummer, Weicheng Yu, Xuezhi Wang, Cong Yu, Daniel Liu, Niyati Mehta</i>	
GRANO: Interactive Graph-based Root Cause Analysis for Cloud-Native Distributed Data Platform	1942
<i>Hanzhang Wang, Phuong Nguyen, Jun Li, Selcuk Kopru, Gene Zhang, Sanjeev Katariya, Sami Ben-Romdhane</i>	
Dietcoin: Hardening Bitcoin Transaction Verification Process For Mobile Devices	1946
<i>Davide Frey, Marc X. Makkes, Pierre-Louis Roman, François Taïani, Spyros Voulgaris</i>	
Raptor: Large Scale Analysis of Big Raster and Vector Data.....	1950
<i>Samridhhi Singla, Ahmed Eldawy, Rami Alghamdi, Mohamed F. Mokbel</i>	
Data Civilizer 2.0: A Holistic Framework for Data Preparation and Analytics.....	1954
<i>El Kindi Rezig, Lei Cao, Michael Stonebraker, Giovanni Simonini, Wenbo Tao, Samuel Madden, Mourad Ouzzani, Nan Tang, Ahmed K. Elmagarmid</i>	
Tuplex: Robust, Efficient Analytics When Python Rules	1958
<i>Leonhard F. Spiegelberg, Tim Kraska</i>	
Ease.ml/ci and Ease.ml/meter in Action: Towards Data Management for Statistical Generalization	1962
<i>Cedric Renggli, Frances Ann Hubis, Bojan Karlaš, Kevin Schawinski, Wentao Wu, Ce Zhang</i>	
PivotE: Revealing and Visualizing the Underlying Entity Structures for Exploration.....	1966
<i>Han Xueran, Jun Chen, Jiaheng Lu, Yueguo Chen, Xiaoyong Du</i>	

Tutorials

Speedup Your Analytics: Automatic Parameter Tuning for Databases and Big Data Systems.....	1970
--	------

Jiaheng Lu, Yuxing Chen, Herodotos Herodotou, Shivnath Babu

TextCube: Automated Construction and Multidimensional Exploration..... 1974
Yu Meng, Jiaxin Huang, Jingbo Shang, Jiawei Han

The Ever Evolving Online Labor Market: Overview, Challenges and Opportunities 1978
Sihem Amer-Yahia, Senjuti Basu Roy

Machine Learning Meets Big Spatial Data 1982
Ibrahim Sabek, Mohamed F. Mokbel

Data Lake Management: Challenges and Opportunities..... 1986
Fatemeh Nargesian, Erkang Zhu, Renée J. Miller, Ken Pu, Patricia C. Arocena

Combating Fake News: A Data Management and Mining Perspective..... 1990
Laks V.s. Lakshmanan, Michael Simpson, Saravanan Thirumuruganathan

Personal Database Security and Trusted Execution Environments: A Tutorial at the Crossroads... 1994
Nicolas Ancaux, Luc Bouganim, Philippe Pucheral, Iulian Sandu Popa, Guillaume Scerri

Industrial and Applications

SAP HANA goes private - From Privacy Research to Privacy Aware Enterprise Analytics 1998
Stephan Kessler, Jens Hoff, Johann-Christoph Freytag

Guided automated learning for query workload re-optimization..... 2010
Guilherme Damasio, Vincent Corvinelli, Parke Godfrey, Piotr Mierzejewski, Alex Mihaylov, Jaroslaw Szlichta, Calisto Zuzarte

Procella: Unifying serving and analytical data at YouTube 2022
Biswapesh Chattopadhyay, Priyam Dutta, Weiran Liu, Ott Tinn, Andrew McCormick, Aniket Mokashi, Paul Harvey, Hector Gonzalez, David Lomax, Sagar Mittal, Roe Ebenstein, Nikita Mikhaylin, Hung-Ching Lee, Xiaoyan Zhao, Tony Xu, Luis Perez, Farhad Shahmohammadi, Tran Bui, Neil Mckay, Selcuk Aya, Vera Lychagina, Brett Elliott

A Lightweight and Efficient Temporal Database Management System in TDSQL..... 2035
Wei Lu, Zhanhao Zhao, Xiaoyu Wang, Haixiang Li, Zhenmiao Zhang, Zhiyu Shui, Sheng Ye, Anqun Pan, Xiaoyong Du

Native Store Extension for SAP HANA 2047
Reza Sherkat, Colin Florendo, Mihnea Andrei, Rolando Blanco, Adrian Dragusanu, Amit Pathak, Pushkar Khadilkar, Neeraj Kulkarni, Christian Lemke, Sebastian Seifert, Sarika Iyer, Sasikanth Gottapu, Robert Schulze, Chaitanya Gottipati, Nirvik Basak, Yanhong Wang, Vivek Kandiyannallur, Santosh Pendap, Dheren Gala, Rajesh Almeida, Prasanta Ghosh

AnalyticDB: Real-time OLAP Database System at Alibaba Cloud 2059
Chaoqun Zhan, Maomeng Su, Chuangxian Wei, Xiaoqiang Peng, Liang Lin, Sheng Wang, Zhe Chen, Feifei Li, Yue Pan, Fang Zheng, Chengliang Chai

Tunable Consistency in MongoDB..... 2071
William Schultz, Tess Avitabile, Alyson Cabral

TitAnt: Online Real-time Transaction Fraud Detection in Ant Financial 2082
Shaosheng Cao, Xinxing Yang, Cen Chen, Jun Zhou, Xiaolong Li, Yuan Qi

AliGraph: A Comprehensive Graph Neural Network Platform.....	2094
<i>Rong Zhu, Kun Zhao, Hongxia Yang, Wei Lin, Chang Zhou, Baole Ai, Yong Li, Jingren Zhou</i>	
Customizable and Scalable Fuzzy Join for Big Data.....	2106
<i>Zhimin Chen, Yue Wang, Vivek Narasayya, Surajit Chaudhuri</i>	
Q Tune: A Query-Aware Database Tuning System with Deep Reinforcement Learning.....	2118
<i>Guoliang Li, Xuanhe Zhou, Shifu Li, Bo Gao</i>	
Experiences with Approximating Queries in Microsoft's Production Big-Data Clusters	2131
<i>Srikanth Kandula, Kukjin Lee, Surajit Chaudhuri, Marc Friedman</i>	
Constant Time Recovery in Azure SQL Database	2143
<i>Panagiotis Antonopoulos, Peter Byrne, Wayne Chen, Cristian Diaconu, Raghavendra Thallam Kodandaramaih, Hanuma Kodavalla, Prashanth Purnananda, Adrian-Leonard Radu, Chaitanya Sreenivas Ravella, Girish Mittur Venkataramanappa</i>	
Yugong: Geo-Distributed Data and Job Placement at Scale.....	2155
<i>Yuzhen Huang, Yingjie Shi, Zheng Zhong, Yihui Feng, James Cheng, Jiwei Li, Haochuan Fan, Chao Li, Tao Guan, Jingren Zhou</i>	
Choosing A Cloud DBMS: Architectures and Tradeoffs.....	2170
<i>Junjay Tan, Thanaa Ghanem, Matthew Perron, Xiangyao Yu, Michael Stonebraker, David Dewitt, Marco Serafini, Ashraf Aboulnaga, Tim Kraska</i>	
S3: A Scalable In-memory Skip-List Index for Key-Value Store.....	2183
<i>Jingtian Zhang, Sai Wu, Zeyuan Tan, Gang Chen, Zhushi Cheng, Wei Cao, Yusong Gao, Xiaojie Feng</i>	
DDSketch: A Fast and Fully-Mergeable Quantile Sketch with Relative-Error Guarantees	2195
<i>Charles Masson, Jee E. Rim, Homin K. Lee</i>	
A Distributed System for Large-scale n-gram Language Models at Tencent	2206
<i>Qiang Long, Wei Wang, Jinfu Deng, Song Liu, Wenhao Huang, Fangying Chen, Sifan Liu</i>	
A Morsel-Driven Query Execution Engine for Heterogeneous Multi-Cores.....	2218
<i>Kayhan Dursun, Carsten Binnig, Ugur Cetintemel, Garret Swart, Weiwei Gong</i>	
Smile: A System to Support Machine Learning on EEG Data at Scale	2230
<i>Lei Cao, Wenbo Tao, Sungtae An, Jing Jin, Yizhou Yan, Xiaoyu Liu, Wendong Ge, Adam Sah, Leilani Battle, Jimeng Sun, Remco Chang, Brandon Westover, Samuel Madden, Michael Stonebraker</i>	
Updating Graph Databases with Cypher	2242
<i>Alastair Green, Paolo Guagliardo, Leonid Libkin, Tobias Lindaaker, Victor Marsault, Stefan Plantikow, Martin Schuster, Petra Selmer, Hannes Voigt</i>	
Adapting TPC-C Benchmark to Measure Performance of Multi-Document Transactions in MongoDB	2254
<i>Asya Kamsky</i>	
Cloud native database systems at Alibaba: Opportunities and Challenges.....	2263
<i>Feifei Li</i>	
In-Memory for the masses: Enabling cost-efficient deployments of in-memory data management platforms for business applications	2273
<i>Alexander Boehm</i>	

Couchbase Analytics: NoETL for Scalable NoSQL Data Analysis 2275
Murtadha Al Hubail, Ali Alsuliman, Michael Blow, Michael Carey, Dmitry Lychagin, Ian Maxon, Till Westmann

Performance in the spotlight..... 2287
Adrian Coyle

Award Talks and Panel

Integration of Large-Scale Data Processing Systems and Traditional Parallel Database Technology 2290
Azza Abouzied, Daniel J. Abadi, Kamil Bajda-Pawlikowski, Avi Silberschatz

PNUTS to Sherpa: Lessons from Yahoo!'s Cloud Database..... 2300
Brian F. Cooper, P.p.s. Narayan, Raghu Ramakrishnan, Utkarsh Srivastava, Adam Silberstein, Philip Bohannon, Hans-Arno Jacobsen, Nick Puz, Daniel Weaver, Ramana Yerneni

What I probably did right and what I think I could have done better 2308
Wang-Chiew Tan

Enabling Data Science for the Majority..... 2309
Aditya Parameswaran

Opportunities for Data Management Research in the Era of Horizontal AI/ML 2323
Theodoros Rekatsinas, Sudeepa Roy, Manasi Vartak, Ce Zhang, Neoklis Polyzotis