

47th International Conference on Very Large Data Bases (VLDB 2021)

Proceedings of the VLDB Endowment Volume 14

Copenhagen, Denmark
16-20 August 2021

Part 1 of 4

Editors:

**Xin Luna Dong
Felix Naumann**

ISBN: 978-1-7138-3522-6

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© (2020-2021) 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. (2021)

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

Benchmarking Learned Indexes	1
<i>Ryan Marcus, Andreas Kipf, Alexander Van Renen, Mihail Stoian, Sanchit Misra, Alfons Kemper, Thomas Neumann, Tim Kraska</i>	
Tempura: A General Cost-Based Optimizer Framework for Incremental Data Processing	14
<i>Zuozhi Wang, Kai Zeng, Botong Huang, Wei Chen, Xiaozong Cui, Bo Wang, Ji Liu, Liya Fan, Dachuan Qu, Zhenyu Hou, Tao Guan, Chen Li, Jingren Zhou</i>	
Inspector Gadget: A Data Programming-based Labeling System for Industrial Images	28
<i>Geon Heo, Yuji Roh, Seonghyeon Hwang, Dayun Lee, Steven Whang</i>	
Scaling Attributed Network Embedding to Massive Graphs	37
<i>Renchi Yang, Jieming Shi, Xiaokui Xiao, Yin Yang, Juncheng Liu, Sourav S. Bhowmick</i>	
Deep Entity Matching with Pre-Trained Language Models	50
<i>Yuliang Li, Jinfeng Li, Yoshihiko Suhara, Anhai Doan, Wang-Chiew Tan</i>	
NeuroCard: One Cardinality Estimator for All Tables	61
<i>Zongheng Yang, Amog Kamsetty, Sifei Luan, Eric Liang, Yan Duan, Peter Chen, Ion Stoica</i>	

TABLE OF CONTENTS

Research Papers

Tsunami: A Learned Multi-dimensional Index for Correlated Data and Skewed Workloads	74
<i>Jialin Ding, Vikram Nathan, Mohammad Alizadeh, Tim Kraska</i>	
Jointly Optimizing Preprocessing and Inference for DNN-based Visual Analytics	87
<i>Daniel Kang, Ankit Mathur, Teja Veeramachineni, Peter Bailis, Matei Zahari</i>	
Permutable Compiled Queries: Dynamically Adapting Compiled Queries without Recompiling	101
<i>Prashanth Menon, Amadou Ngom, Todd Mowry, Andrew Pavlo, Lin Ma</i>	
EMOGI: Efficient Memory-access for Out-of-memory Graph-traversal In GPUs	114
<i>Seung Won Min, Vikram Sharma Mailthody, Zaid Qureshi, Jinjun Xiong, Eiman Ebrahimi, Wen-Mei Hwu</i>	
On-Off Sketch: A Fast and Accurate Sketch on Persistence	128
<i>Yinda Zhang, Jinyang Li, Yutian Lei, Tong Yang, Zhetao Li, Gong Zhang, Bin Cui</i>	
Real-Time Distance-Based Outlier Detection in Data Streams	141
<i>Luan Tran, Min Mun, Cyrus Shahabi</i>	
Seagull: An Infrastructure for Load Prediction and Optimized Resource Allocation	154
<i>Olga Poppe, Tayo Amunike, Dalitso Banda, Aritra De, Ari Green, Manon Knoertzer, Ehi Nosakhare, Karthik Rajendran, Deepak Shankargouda, Meina Wang, Alan Au, Carlo Curino, Qun Guo, Alekh Jindal, Ajay Kalhan, Morgan Oslake, Sonia Parchani, Vijay Ramani, Raj Sellappan, Saikat Sen, Sheetal Shrotri, Soundararajan Srinivasan, Ping Xia, Shize Xu, Alicia Yang, Yiwen Zhu</i>	
On the Efficiency of K-Means Clustering: Evaluation, Optimization, and Algorithm Selection	163
<i>Sheng Wang, Yuan Sun, Zhifeng Bao</i>	
RapidMatch: A Holistic Approach to Subgraph Query Processing	176
<i>Shixuan Sun, Xibo Sun, Yulin Che, Qiong Luo, Bingsheng He</i>	
Taurus: Lightweight Parallel Logging for In-Memory Database Management Systems.....	189
<i>Yu Xia, Xiangyao Yu, Andrew Pavlo, Srinivas Devadas</i>	
Improving Execution Efficiency of Just-in-time Compilation based Query Processing on GPUs.....	202
<i>Johns Paul, Bingsheng He, Shengliang Lu, Chiew Tong Lau</i>	
PPQ-Trajectory: Spatio-temporal Qantization for Qerying in Large Trajectory Repositories	215
<i>Shuang Wang, Hakan Ferhatosmanoglu</i>	

Aggregated Deletion Propagation for Counting Conjunctive Query Answers	228
<i>Xiao Hu, Shouzhuo Sun, Shweta Patwa, Debmalya Panigrahi, Sudeepa Roy</i>	

TABLE OF CONTENTS

Research Papers

Breaking Down Memory Walls: Adaptive Memory Management in LSM-based Storage Systems ... <i>Chen Luo, Michael Carey</i>	241
Nearest Neighbor Classifiers over Incomplete Information: From Certain Answers to Certain Predictions <i>Bojan Karlaš, Peng Li, Renzhi Wu, Nezihe Merve Gürel, Xu Chu, Wentao Wu, Ce Zhang</i>	255
Elle: Inferring Isolation Anomalies from Experimental Observations Peter Alvaro, Kyle Kingsbury	268
Scotch: Generating FPGA-Accelerators for Sketching at Line Rate <i>Martin Kiefer, Ilias Poulakis, Sebastian Bress, Volker Markl</i>	281
ORBITS: Online Recovery of Missing Values in Multiple Time Series Streams <i>Mourad Khayati, Ines Arous, Zakhar Tymchenko, Philippe Cudre-Mauroux</i>	294
TURL: Table Understanding through Representation Learning <i>Xiang Deng, Huan Sun, Alyssa Lees, You Wu, Cong Yu</i>	307
EdgeDIPN: a Unified Deep Intent Prediction Network Deployed at the Edge <i>Long Guo, Lifeng Hua, Rongfei Jia, Fei Fang, Binqiang Zhao, Bin Cui</i>	320
LOCATER: Cleaning WiFi Connectivity Datasets for Semantic Localization <i>Yiming Lin, Daokun Jiang, Roberto Yus, Georgios Bouloukakis, Andrew Chio, Sharad Mehrotra, Nalini Venkatasubramanian</i>	329
Multi-Modal Transportation Recommendation with Unified Route Representation Learning <i>Hao Liu, Jindong Han, Yanjie Fu, Jingbo Zhou, Xinjiang Lu, Hui Xiong</i>	342
DISK: A Distributed Framework for Single-Source SimRank with Accuracy Guarantee <i>Yue Wang, Ruiqi Xu, Zonghao Feng, Yulin Che, Lei Chen, Qiong Luo, Rui Mao</i>	351
Toward a Better Understanding and Evaluation of Tree Structures on Flash SSDs <i>Diego Didona, Nikolas Ioannou, Radu Stoica, Kornilios Kourtis</i>	364
Answering Multi-Dimensional Range Queries under Local Differential Privacy <i>Jianyu Yang, Tianhao Wang, Ninghui Li, Xiang Cheng, Sen Su</i>	378
Ananke: A Streaming Framework for Live Forward Provenance Dimitris Palyvos-Giannas, Bastian Havers, Marina Papatriantafidou, Vincenzo Gulisano	391

RECEIPT: REfine CoarsE-grained IndePendent Tasks for Parallel Tip decomposition of Bipartite Graphs	404
<i>Kartik Lakhotia, Rajgopal Kannan, Viktor Prasanna, César De Rose</i>	
Comprehensive and Efficient Workload Compression	418
<i>Shaleen Deep, Anja Gruenheid, Paraschos Koutris, Jeff Naughton, Stratis Viglas</i>	
CoroBase: Coroutine-Oriented Main-Memory Database Engine	431
<i>Yongjun He, Jiacheng Lu, Tianzheng Wang</i>	
Scalable Querying of Nested Data	445
<i>Jaclyn Smith, Michael Benedikt, Milos Nikolic, Amir Shaikhha</i>	

TABLE OF CONTENTS

Research Papers

Space- and Computationally-Efficient Set Reconciliation via Parity Bitmap Sketch (PBS)	458
<i>Long Gong, Ziheng Liu, Liang Liu, Jun Xu, Mitsunori Ogihara, Tong Yang</i>	
Astrid: Accurate Selectivity Estimation for String Predicates using Deep Learning	471
<i>Suraj Shetiya, Saravanan Thirumuruganathan, Nick Koudas, Gautam Das</i>	
Compact, Tamper-Resistant Archival of Fine-Grained Provenance	485
<i>Nan Zheng, Zack Ives</i>	
Rumble: Data Independence for Large Messy Data Sets	498
<i>Ingo Müller, Ghislain Fourny, Stefan Irimescu, Can Cikis, Gustavo Alonso</i>	
Capturing and querying fine-grained provenance of preprocessing pipelines in data science	507
<i>Adriane Chapman, Paolo Missier, Giulia Simonelli, Riccardo Torlone</i>	
Local Dampening: Differential Privacy for Non-numeric Queries via Local Sensitivity	521
<i>Victor Farias, Felipe Brito, Cheryl Flynn, Javam Machado, Subhabrata Majumdar, Divesh Srivastava</i>	
Mainlining Databases: Supporting Fast Transactional Workloads on Universal Columnar Data File Formats	534
<i>Tianyu Li, Matthew Butrovich, Amadou Ngom, Wan Shen Lim, Wes Mckinney, Andrew Pavlo</i>	
Accelerating Exact Constrained Shortest Paths on GPUs	547
<i>Shengliang Lu, Bingsheng He, Yuchen Li, Hao Fu</i>	
Towards an Efficient Weighted Random Walk Domination	560
<i>Songsong Mo, Zhifeng Bao, Ping Zhang, Zhiyong Peng</i>	
Scalable Mining of Maximal Quasi-Cliques: An Algorithm-System Codesign Approach	573
<i>Guimu Guo, Da Yan, Tamer Özsu, Zhe Jiang, Jalal Khalil</i>	
CALYPSO: Private Data Management for Decentralized Ledgers	586
<i>Eleftherios Kokoris Kogias, Enis Ceyhun Alp, Linus Gasser, Philipp Jovanovic, Ewa Syta, Bryan Ford</i>	
Stacked Filters: Learning to Filter by Structure.....	600
<i>Brian Hentschel, Stratos Idreos, Kyle Deeds</i>	
Maximizing Social Welfare in a Competitive Diffusion Model	613
<i>Prithu Banerjee, Laks V.s. Lakshmanan, Wei Chen</i>	

Understanding the Idiosyncrasies of Real Persistent Memory	626
<i>Shashank Gugnani, Arjun Kashyap, Xiaoyi Lu</i>	
Explaining Ranking Functions	640
<i>Amelie Marian, Abraham Gale</i>	
ConnectIt: A Framework for Static and Incremental Parallel Graph Connectivity Algorithms	653
<i>Laxman Dhulipala, Changwan Hong, Julian Shun</i>	
Quality of Sentiment Analysis Tools: The Reasons of Inconsistency	668
<i>Wissam Mammam Kouadri, Mourad Ouziri, Salima Benbernou, Karima Echihabi, Themis Palpanas, Iheb Benamor</i>	
Hindsight Logging for Model Training	682
<i>Rolando Garcia, Erick Liu, Vikram Sreekanti, Bobby Yan, Anusha Dandamudi, Joseph Gonzalez, Joseph Hellerstein, Koushik Sen</i>	
Scalable Structural Index Construction for JSON Analytics	694
<i>Lin Jiang, Junqiao Qiu, Zhijia Zhao</i>	
Efficient Join Algorithms For Large Database Tables in a Multi-GPU Environment	708
<i>Ran Rui, Hao Li, Yi-Cheng Tu</i>	

TABLE OF CONTENTS

Research Papers

FlashP: An Analytical Pipeline for Real-time Forecasting of Time-Series Relational Data	721
<i>Yan Shuyuan, Bolin Ding, Wei Guo, Jingren Zhou, Zhewei Wei, Xiaowei Jiang, Sheng Xu</i>	
Efficient Streaming Subgraph Isomorphism with Graph Neural Networks.....	730
<i>Chi Thang Duong, Dung Hoang, Hongzhi Yin, Matthias Weidlich, Quoc Viet Hung Nguyen, Karl Aberer</i>	
Epoch-based Commit and Replication in Distributed OLTP Databases	743
<i>Yi Lu, Xiangyao Yu, Lei Cao, Samuel Madden</i>	
Hierarchical Core Maintenance on Large Dynamic Graphs.....	757
<i>Zhe Lin, Fan Zhang, Xuemin Lin, Wenjie Zhang, Zhihong Tian</i>	
Analyzing and Mitigating Data Stalls in DNN Training	771
<i>Jayashree Mohan, Amar Phanishayee, Ashish Raniwala, Vijay Chidambaram</i>	
Persistent Memory Hash Indexes: An Experimental Evaluation.....	785
<i>Daokun Hu, Zhiwen Chen, Jianbing Wu, Jianhua Sun, Hao Chen</i>	
Optimizing An In-memory Database System For AI-powered On-line Decision Augmentation Using Persistent Memory	799
<i>Cheng Chen, Jun Yang, Mian Lu, Taize Wang, Zhao Zheng, Yuqiang Chen, Wenyuan Dai, Bingsheng He, Weng-Fai Wong, Guoan Wu, Yuping Zhao, Andy Rudoff</i>	
DBTagger: Multi-Task Learning for Keyword Mapping in NLIDBs Using Bi-Directional Recurrent Neural Networks.....	813
<i>Arif Usta, Akifhan Karakayali, Özgür Ulusoy</i>	
Improving Information Extraction from Visually Rich Documents using Visual Span Representations	822
<i>Ritesh Sarkhel, Arnab Nandi</i>	
Zen: a High-Throughput Log-Free OLTP Engine for Non-Volatile Main Memory	835
<i>Gang Liu, Leying Chen, Shimin Chen</i>	
Differentially Private Binary- and Matrix-Valued Data Query: An XOR Mechanism.....	849
<i>Tianxi Ji, Pan Li, Emre Yilmaz, Erman Ayday, Yanfang Ye, Jinyuan Sun</i>	

TABLE OF CONTENTS

Research Papers

Errata for "Cerebro: A Data System for Optimized Deep Learning Model Selection"	863
<i>Supun Nakandala, Yuhao Zhang, Arun Kumar</i>	
ParaX: Boosting Deep Learning for Big Data Analytics on Many-Core CPUs.....	864
<i>Lujia Yin, Yiming Zhang, Zhaoning Zhang, Yuxing Peng, Peng Zhao</i>	
Optimization of Threshold Functions over Streams	878
<i>Walter Cai, Philip Bernstein, Wentao Wu, Badrish Chandramouli</i>	
Budget Constrained Interactive Search for Multiple Targets.....	890
<i>Xuliang Zhu, Xin Huang, Byron Choi, Jiaxin Jiang, Zhaonian Zou, Jianliang Xu</i>	
On the String Matching with k Differences in DNA Databases	903
<i>Yangjun Chen, Hoang Hai Nguyen</i>	
Fast Algorithm for Anchor Graph Hashing	916
<i>Yasuhiro Fujiwara, Sekitoshi Kanai, Yasutoshi Ida, Atsutoshi Kumagai, Naonori Ueda</i>	
Adaptive Code Generation for Data-Intensive Analytics	929
<i>Wangda Zhang, Junyoung Kim, Kenneth Ross, Eric Sedlar, Lukas Stadler</i>	
Materializing Knowledge Bases via Trigger Graphs.....	943
<i>Efthymia Tsamoura, David Carral, Enrico Malizia, Jacopo Urbani</i>	
Dealer: An End-to-End Model Marketplace with Differential Privacy.....	957
<i>Jinfei Liu, Jian Lou, Junxu Liu, Li Xiong, Jian Pei, Jimeng Sun</i>	
NOAH: Interactive Spreadsheet Exploration with Dynamic Hierarchical Overviews	970
<i>Sajjadur Rahman, Mangesh Bendre, Yuyang Liu, Shichu Zhu, Zhaoyuan Su, Karrie Karahalios, Aditya Parameswaran</i>	
Efficient Bi-triangle Counting for Large Bipartite Networks.....	984
<i>Yixing Yang, Yixiang Fang, Maria Orlowska, Wenjie Zhang, Xuemin Lin</i>	
Glean: Structured Extractions from Templatic Documents	997
<i>Sandeep Tata, Navneet Potti, James Wendt, Lauro Beltrão Costa, Marc Najork, Beliz Gunel</i>	
ICS-GNN: Lightweight Interactive Community Search via Graph Neural Network	1006
<i>Jun Gao, Jiazun Chen, Zhao Li, Ji Zhang</i>	

Building Enclave-Native Storage Engines for Practical Encrypted Databases.....	1019
<i>Yuanyuan Sun, Sheng Wang, Huorong Li, Feifei Li</i>	
From Natural Language Processing to Neural Databases.....	1033
<i>James Thorne, Majid Yazdani, Marzieh Saeidi, Fabrizio Silvestri, Sebastian Riedel, Alon Halevy</i>	
Randomized Error Removal for Online Spread Estimation in Data Streaming.....	1040
<i>Haibo Wang, Chaoyi Ma, Olufemi O Odegbile, Shigang Chen, Jih-kwon Peir</i>	
Teseo and the Analysis of Structural Dynamic Graphs.....	1053
<i>Dean De Leo, Peter Boncz</i>	
Charting the Design Space of Query Execution using VOILA.....	1067
<i>Tim Gubner, Peter Boncz</i>	
Heracles: An Efficient Storage Model And Data Flushing For Performance Monitoring Timeseries	1080
<i>Zhiqi Wang, Jin Xue, Zili Shao</i>	
Fine-Grained Lineage for Safer Notebook Interactions	1093
<i>Stephen Macke, Aditya Parameswaran, Hongpu Gong, Doris Lee, Doris Xin, Andrew Head</i>	
FREDE: Anytime Graph Embeddings.....	1102
<i>Anton Tsitsulin, Marina Munkhoeva, Davide Mottin, Panagiotis Karras, Ivan Oseledets, Emmanuel Müller</i>	
On Analyzing Graphs with Motif-Paths	1111
<i>Xiaodong Li, Reynold Cheng, Kevin Chen-chuan Chang, Caihua Shan, Chenhao Ma, Hongtai Cao</i>	

TABLE OF CONTENTS

Research Papers

Collective Influence Maximization for Multiple Competing Products with an Awareness-to-Influence Model	1124
<i>Dimitris Tsaras, George Trimponias, Lefteris Ntaflos, Dimitris Papadias</i>	
Finding Group Steiner Trees in Graphs with both Vertex and Edge Weights.....	1137
<i>Yahui Sun, Xiaokui Xiao, Bin Cui, Saman Halgamuge, Theodoros Lappas, Jun Luo</i>	
Optimizing Bipartite Matching in Real-World Applications by Incremental Cost Computation	1150
<i>Tenindra Abeywickrama, Victor Liang, Kian-lee Tan</i>	
The Case for NLP-Enhanced Database Tuning: Towards Tuning Tools that “Read the Manual” ...	1159
<i>Immanuel Trummer</i>	
Errata for “Unifying Consensus and Atomic Commitment for Effective Cloud Data Management”.	1166
<i>Sujaya Maiyya, Faisal Nawab, Divy Agrawal, Amr El Abbadi</i>	
Software-Defined Data Protection: Low Overhead Policy Compliance at the Storage Layer is Within Reach!.....	1167
<i>Zsolt István, Soujanya Ponnappalli, Vijay Chidambaram</i>	
TRACE: Real-time Compression of Streaming Trajectories in Road Networks	1175
<i>Tianyi Li, Lu Chen, Christian Jensen, Torben Bach Pedersen</i>	
Shortest Paths and Centrality in Uncertain Networks.....	1188
<i>Arkaprava Saha, Ruben Brokkelkamp, Yllka Velaj, Arijit Khan, Francesco Bonchi</i>	
Adaptive Data Augmentation for Supervised Learning over Missing Data.....	1202
<i>Tongyu Liu, Yingqing Luo, Ju Fan, Nan Tang, Guoliang Li, Xiaoyong Du</i>	
KLL±: Approximate Quantile Sketches over Dynamic Datasets	1215
<i>Fuheng Zhao, Sujaya Maiyya, Ryan Weiner, Divy Agrawal, Amr El Abbadi</i>	
Distributed Numerical and Machine Learning Computations via Two-Phase Execution of Aggregated Join Trees	1228
<i>Dimitrije Jankov, Binhang Yuan, Shangyu Luo, Chris Jermaine</i>	
An Inquiry into Machine Learning-based Automatic Configuration Tuning Services on Real-World Database Management Systems.....	1241
<i>Dana Van Aken, Dongsheng Yang, Sebastien Brillard, Ari Fiorino, Bohan Zhang, Christian Billian, Andrew Pavlo</i>	

TABLE OF CONTENTS

Research Papers

RPT: Relational Pre-trained Transformer Is Almost All You Need towards Democratizing Data Preparation	1254
<i>Nan Tang, Ju Fan, Fangyi Li, Jianhong Tu, Xiaoyong Du, Guoliang Li, Samuel Madden, Mourad Ouzzani</i>	
Lachesis: Automated Partitioning for UDF-Centric Analytics (Revision of Paper 270)	1262
<i>Jia Zou, Amitabh Das, Pratik Barhate, Arun Iyengar, Binhang Yuan, Dimitrije Jankov, Chris Jermaine</i>	
Updatable Learned Index with Precise Positions	1276
<i>Jiacheng Wu, Yong Zhang, Shimin Chen, Yu Chen, Jin Wang, Chunxiao Xing</i>	
MDTP: A Multi-source Deep Traffic Prediction Framework over Spatio-Temporal Trajectory Data	1289
<i>Ziquan Fang, Lu Pan, Lu Chen, Yuntao Du, Yunjun Gao</i>	
Symmetric Continuous Subgraph Matching with Bidirectional Dynamic Programming.....	1298
<i>Seunghwan Min, Sung Gwan Park, Kunsoo Park, Dora Giammarresi, Giuseppe F. Italiano, Wook-shin Han</i>	
Approaching DRAM performance by using microsecond-latency flash memory for small-sized random read accesses: a new access method and its graph applications	1311
<i>Tomoya Suzuki, Kazuhiro Hiwada, Hirotsugu Kajihara, Shintaro Sano, Shuou Nomura, Tatsuo Shiozawa</i>	
CBench: Towards Better Evaluation of Question Answering Over Knowledge Graphs	1325
<i>Abdelghny Orogat, Isabelle Liu, Ahmed El-Roby</i>	
Tensor Relational Algebra for Distributed Machine Learning System Design	1338
<i>Binhang Yuan, Dimitrije Jankov, Jia Zou, Yuxin Tang, Daniel Bourgeois, Chris Jermaine</i>	
Parallel Discrepancy Detection and Incremental Detection	1351
<i>Wenfei Fan, Chao Tian, Yanghao Wang, Qiang Yin</i>	
Towards Crowd-aware Indoor Path Planning	1365
<i>Tiantian Liu, Huan Li, Hua Lu, Muhammad Aamir Cheema, Lidan Shou</i>	
Procedural Extensions of SQL: Understanding their usage in the wild	1378
<i>Surabhi Gupta, Karthik Ramachandra</i>	
Discovering Related Data At Scale	1392
<i>Sagar Bharadwaj K S, Praveen Gupta, Ranjita Bhagwan, Saikat Guha</i>	

CGPTuner: a Contextual Gaussian Process Bandit Approach for the Automatic Tuning of IT Configurations Under Varying Workload Conditions.....	1401
<i>Stefano Cereda, Stefano Valladares, Paolo Cremonesi, Stefano Doni</i>	
Language-Agnostic Integrated Queries in a Managed Polyglot Runtime	1414
<i>Filippo Schiavio, Daniele Bonetta, Walter Binder</i>	
Achieving High Throughput and Elasticity in a Larger-than-Memory Store.....	1427
<i>Chinmay Kulkarni, Badrish Chandramouli, Ryan Stutsman</i>	
Efficient Size-Bounded Community Search over Large Networks	1441
<i>Kai Yao, Lijun Chang</i>	

TABLE OF CONTENTS

Research Papers

Minimum Vertex Augmentation	1454
<i>Jianwen Zhao, Yufei Tao</i>	
Database Isolation by Scheduling	1467
<i>Kevin P Gaffney, Robert K Claus, Jignesh Patel</i>	
SaS: SSD as SQL Database System.....	1481
<i>Jong-hyeok Park, Soyeon Choi, Gihwan Oh, Sang Won Lee</i>	
FLAT: Fast, Lightweight and Accurate Method for Cardinality Estimation	1489
<i>Rong Zhu, Ziniu Wu, Yuxing Han, Kai Zeng, Andreas Pfadler, Zhengping Qian, Jingren Zhou, Bin Cui</i>	
Fast Augmentation Algorithms for Network Kernel Density Visualization	1503
<i>Tsz Nam Chan, Zhe Li, Leong Hou U, Jianliang Xu, Reynold Cheng</i>	
AutoGR: Automated Geo-Replication with Fast System Performance and Preserved Application Semantics.....	1517
<i>Jiawei Wang, Cheng Li, Kai Ma, Jingze Huo, Feng Yan, Xinyu Feng, Yinlong Xu</i>	
Local Algorithms for Distance-generalized Core Decomposition over Large Dynamic Graphs.....	1531
<i>Qing Liu, Xuliang Zhu, Xin Huang, Jianliang Xu</i>	
Viper: An Efficient Hybrid PMem-DRAM Key-Value Store.....	1544
<i>Lawrence Benson, Hendrik Makait, Tilmann Rabl</i>	
Estimating Spread of Contact-Based Contagions in a Population Through Sub-Sampling	1557
<i>Sepanta Zeighami, Cyrus Shahabi, John Krumm</i>	
Trident: Task Scheduling over Tiered Storage Systems in Big Data Platforms.....	1570
<i>Herodotos Herodotou, Elena Kakoulli</i>	
Comprehensible Counterfactual Explanation on Kolmogorov-Smirnov Test.....	1583
<i>Zicun Cong, Lingyang Chu, Yu Yang, Jian Pei</i>	
Accelerating Large Scale Real-Time GNN Inference using Channel Pruning	1597
<i>Hongkuan Zhou, Ajitesh Srivastava, Hanqing Zeng, Rajgopal Kannan, Viktor K Prasanna</i>	
Towards Cost-Optimal Query Processing in the Cloud	1606
<i>Viktor Leis, Maximilian Kuschewski</i>	

Automating Incremental Graph Processing with Flexible Memoization	1613
<i>Shufeng Gong, Chao Tian, Qiang Yin, Wenyuan Yu, Yanfeng Zhang, Liang Geng, Song Yu, Ge Yu, Jingren Zhou</i>	
In-Network Support for Transaction Triaging	1626
<i>Theo Jepsen, Alberto Lerner, Fernando Pedone, Robert Soule, Philippe Cudre-Mauroux</i>	
Are We Ready for Learned Cardinality Estimation?	1640
<i>Xiaoying Wang, Changbo Qu, Weiyuan Wu, Jiannan Wang, Qingqing Zhou</i>	
On the algebra of data sketches	1655
<i>Jakub Lemiesz</i>	
Massively Parallel Algorithms for Personalized PageRank	1668
<i>Guanhao Hou, Xingguang Chen, Sibor Wang, Zhewei Wei</i>	
GeCo: Quality Counterfactual Explanations in Real Time.....	1681
<i>Maximilian Schleich, Zixuan Geng, Yihong Zhang, Dan Suciu</i>	
Automated Feature Engineering for Algorithmic Fairness	1694
<i>Ricardo Salazar, Felix Neutatz, Ziawasch Abedjan</i>	

TABLE OF CONTENTS

Research Papers

How to Design Robust Algorithms using Noisy Comparison Oracle	1703
<i>Raghavendra Addanki, Sainyam Galhotra, Barna Saha</i>	
SAND: Streaming Subsequence Anomaly Detection	1717
<i>Paul Boniol, John Paparrizos, Themis Palpanas, Michael Franklin</i>	
Optimizing Fitness-For-Use of Differentially Private Linear Queries.....	1730
<i>Yingtai Xiao, Zeyu Ding, Yuxin Wang, Danfeng Zhang, Daniel Kifer</i>	
Cryptanalysis of An Encrypted Database in SIGMOD '14.....	1743
<i>Xinle Cao, Jian Liu, Hao Lu, Kui Ren</i>	
Unconstrained Submodular Maximization with Modular Costs: Tight Approximation and Application to Profit Maximization.....	1756
<i>Tianyuan Jin, Yu Yang, Renchi Yang, Jieming Shi, Keke Huang, Xiaokui Xiao</i>	
Distributed Deep Learning on Data Systems: A Comparative Analysis of Approaches.....	1769
<i>Yuhao Zhang, Frank Mcquillan, Nandish Jayaram, Nikhil Kak, Ekta Khanna, Orhan Kislal, Domino Valdano, Arun Kumar</i>	
PR-Sketch: Monitoring Per-key Aggregation of Streaming Data with Nearly Full Accuracy.....	1783
<i>Siyuan Sheng, Qun Huang, Sa Wang, Yungang Bao</i>	
Tensors: An abstraction for general data processing.....	1797
<i>Dimitrios Koutsoukos, Supun C Nakandala, Konstantinos Karanasos, Karla Saur, Gustavo Alonso, Matteo Interlandi</i>	
Budget Sharing for Multi-Analyst Differential Privacy.....	1805
<i>David A Pujol, Yikai Wu, Brandon T Fain, Ashwin Machanavajjhala</i>	
In the Land of Data Streams where Synopses are Missing, One Framework to Bring Them All.....	1818
<i>Rudi Poepfel-Lemaitre, Martin Kiefer, Joscha Von Hein, Jorge Arnulfo Quiane Ruiz, Volker Markl</i>	
Data Acquisition for Improving Machine Learning Models.....	1832
<i>Yifan Li, Xiaohui Yu, Nick Koudas</i>	
Efficiently Answering Reachability and Path Queries on Temporal Bipartite Graphs	1845
<i>Xiaoshuang Chen, Kai Wang, Xuemin Lin, Wenjie Zhang, Lu Qin, Ying Zhang</i>	

Preference Queries over Taxonomic Domains.....	1859
<i>Paolo Ciaccia, Davide Martinenghi, Riccardo Torlone</i>	
Revisiting the Design of LSM-tree Based OLTP Storage Engine with Persistent Memory.....	1872
<i>Baoyue Yan, Xuntao Cheng, Bo Jiang, Shibin Chen, Canfang Shang, Jianying Wang, Kenry Huang, Xinjun Yang, Wei Cao, Feifei Li</i>	
Kamino: Constraint-Aware Differentially Private Data Synthesis	1886
<i>Chang Ge, Shubhankar Mohapatra, Xi He, Ihab F Ilyas</i>	
Towards Cost-Effective and Elastic Cloud Database Deployment via Memory Disaggregation	1900
<i>Yingqiang Zhang, Chaoyi Ruan, Cheng Li, Jimmy Yang, Wei Cao, Feifei Li, Bo Wang, Jing Fang, Yuhui Wang, Jingze Huo, Chao Bi</i>	
Dual-Objective Fine-Tuning of BERT for Entity Matching	1913
<i>Ralph Peeters, Christian Bizer</i>	

TABLE OF CONTENTS

Research Papers

GraphMineSuite: Enabling High-Performance and Programmable Graph Mining Algorithms with Set Algebra	1922
<i>Maciej Besta, Zur Vonarburg-shmaria, Yannick Schaffner, Leonardo Schwarz, Grzegorz Kwasniewski, Lukas Gianinazzi, Jakub Beranek, Kacper Janda, Tobias Holenstein, Sebastian Leisinger, Peter Tatkowski, Esref Özdemir, Adrian Balla, Marcin Copik, Philipp Lindenberger, Marek Konieczny, Onur Mutlu, Torsten Hoefler</i>	
PATSQL: Efficient Synthesis of SQL Queries from Example Tables with Quick Inference of Projected Columns	1937
<i>Keita Takenouchi, Takashi Ishio, Joji Okada, Yuji Sakata</i>	
Fauce: Fast and Accurate Deep Ensembles with Uncertainty for Cardinality Estimation.....	1950
<i>Jie Liu, Wenqian Dong, Dong Li, Qingqing Zhou</i>	
A Comprehensive Survey and Experimental Comparison of Graph-Based Approximate Nearest Neighbor Search.....	1964
<i>Mengzhao Wang, Xiaoliang Xu, Qiang Yue, Yuxiang Wang</i>	
Towards Plug-and-Play Visual Graph Query Interfaces: Data-driven Canned Pattern Selection for Large Networks.....	1979
<i>Zifeng Yuan, Huey Eng Chua, Sourav S Bhowmick, Zekun Ye, Wook-shin Han, Byron Choi</i>	
ThunderRW: An In-Memory Graph Random Walk Engine	1992
<i>Shixuan Sun, Yuhang Chen, Shengliang Lu, Bingsheng He, Yuchen Li</i>	
Butterfly-Core Community Search over Labeled Graphs	2006
<i>Zheng Dong, Xin Huang, Guorui Yuan, Hengshu Zhu, Hui Xiong</i>	
Flow-Loss: Learning Cardinality Estimates That Matter	2019
<i>Parimarjan Negi, Ryan C Marcus, Andreas Kipf, Hongzi Mao, Nesime Tatbul, Tim Kraska, Mohammad Alizadeh</i>	
On Querying Historical K-Cores	2033
<i>Michael R Yu, Dong Wen, Lu Qin, Ying Zhang, Wenjie Zhang, Xuemin Lin</i>	
Frequency Estimation under Local Differential Privacy.....	2046
<i>Graham Cormode, Sam Maddock, Carsten Maple</i>	
Doing More with Less: Characterizing Dataset Downsampling for AutoML	2059
<i>Fatjon Zogaj, Jose P Cambronero Sanchez, Martin Rinard, Jürgen Cito</i>	

LES3: Learning-based exact set similarity search.....	2073
<i>Yifan Li, Xiaohui Yu, Nick Koudas</i>	
Large Graph Convolutional Network Training with GPU-Oriented Data Communication Architecture	2087
<i>Seung Won Min, Kun Wu, Sitao Huang, Mert Hidayetoglu, Jinjun Xiong, Eiman Ebrahimi, Deming Chen, Wen-mei Hwu</i>	
FlexPushdownDB: Hybrid Pushdown and Caching in a Cloud DBMS	2101
<i>Yifei Yang, Matt Youill, Matthew Woicik, Yizhou Liu, Xiangyao Yu, Marco Serafini, Ashraf Aboulnaga, Michael Stonebraker</i>	
Approximating Median Absolute Deviation with Bounded Error	2114
<i>Zhiwei Chen, Shaoxu Song, Ziheng Wei, Jingyun Fang, Jiang Long</i>	
An Experimental Evaluation and Guideline for Path Finding in Weighted Dynamic Network.....	2127
<i>Mengxuan Zhang, Lei Li, Xiaofang Zhou</i>	
Robustness against Read Committed for Transaction Templates.....	2141
<i>Brecht Vandervoort, Bas Ketsman, Christoph Koch, Frank Neven</i>	
LANCET: Labeling Complex Data at Scale	2154
<i>Huayi Zhang, Lei Cao, Samuel Madden, Elke A Rundensteiner</i>	
VolcanoML: Speeding up End-to-End AutoML via Scalable Search Space Decomposition.....	2167
<i>Yang Li, Yu Shen, Wentao Zhang, Jiawei Jiang, Yaliang Li, Bolin Ding, Jingren Zhou, Zhi Yang, Wentao Wu, Ce Zhang, Bin Cui</i>	
A Queueing-Theoretic Framework for Vehicle Dispatching in Dynamic Car-Hailing.....	2177
<i>Peng Cheng, Jiabao Jin, Lei Chen, Xuemin Lin, Libin Zheng</i>	
Data Synthesis via Differentially Private Markov Random Field.....	2190
<i>Kuntai Cai, Xiaoyu Lei, Jianxin Wei, Xiaokui Xiao</i>	
Scaling Replicated State Machines with Compartmentalization.....	2203
<i>Michael J Whittaker, Ailidani Ailidani, Aleksey Charapko, Murat Demirbas, Neil Giridharan, Joseph M Hellerstein, Heidi Howard, Ion Stoica, Adriana Szekeres</i>	
Constructing and Analyzing the LSM Compaction Design Space	2216
<i>Subhadeep Sarkar, Dimitris Staratzis, Zichen Zhu, Manos Athanassoulis</i>	
ByShard: Sharding in a Byzantine Environment	2230
<i>Jelle Hellings, Mohammad Sadoghi</i>	
SetSketch: Filling the Gap between MinHash and HyperLogLog	2244
<i>Otmaz Ertl</i>	
CGM: An Enhanced Mechanism for Streaming Data Collection with Local Differential Privacy	2258
<i>Ergute Bao, Yin Yang, Xiaokui Xiao, Bolin Ding</i>	
Errata for "Teseo and the Analysis of Structural Dynamic Graph".....	2271
<i>Dean De Leo, Per Fuchs, Peter Boncz</i>	

QARTA: An ML-based System for Accurate Map Services	2273
<i>Mashaal Musleh, Sofiane Abbar, Rade Stanojevic, Mohamed Mokbel</i>	
Real-World Trajectory Sharing with Local Differential Privacy	2283
<i>Teddy Cunningham, Graham Cormode, Hakan Ferhatosmanoglu, Divesh Srivastava</i>	
PolyFrame: A Retargetable Query-based Approach to Scaling Dataframes.....	2296
<i>Phanwadee Sinthong, Michael Carey</i>	
Scalable Community Detection via Parallel Correlation Clustering.....	2305
<i>Jessica Shi, Laxman Dhulipala, David Eisenstat, Jakub Łacki, Vahab Mirrokni</i>	
SlimChain: Scaling Blockchain Transactions through Off-Chain Storage and Parallel Processing...	2314
<i>Cheng Xu, Ce Zhang, Jianliang Xu, Jian Pei</i>	
Towards an Optimized GROUP BY Abstraction for Large-Scale Machine Learning	2327
<i>Side Li, Arun Kumar</i>	
Accelerating Approximate Aggregation Queries with Expensive Predicates.....	2341
<i>Daniel Kang, John Guibas, Peter D Bailis, Tatsunori Hashimoto, Yi Sun, Matei Zaharia</i>	
A four-dimensional Analysis of Partitioned Approximate Filters	2355
<i>Tobias Schmidt, Maximilian Bandle, Jana Giceva</i>	
SKT: A One-Pass Multi-Sketch Data Analytics Accelerator	2369
<i>Monica Chiosa, Thomas Preußner, Gustavo Alonso</i>	
A Practical Approach to Groupjoin and Nested Aggregates	2383
<i>Philipp Fent, Thomas Neumann</i>	
Robust Voice Querying with MUVE: Optimally Visualizing Results of Phonetically Similar Queries.	2397
<i>Ziyun Wei, Immanuel Trummer, Connor Anderson</i>	
CHEF: A Cheap and Fast Pipeline for Iteratively Cleaning Label Uncertainties	2410
<i>Yinjun Wu, James Weimer, Susan B Davidson</i>	
COMPARE: Accelerating Groupwise Comparison in Relational Databases for Data Analytics	2419
<i>Tarique Siddiqui, Surajit Chaudhuri, Vivek Narasayya</i>	
Crystal: A Unified Cache Storage System for Analytical Databases	2432
<i>Dominik Durner, Badrish Chandramouli, Yinan Li</i>	
The Smallest Extraction Problem	2445
<i>Valerio Cettorelli, Paolo Atzeni, Valter Crescenzi, Franco Milicchio</i>	
Deep Learning for Blocking in Entity Matching: A Design Space Exploration.....	2459
<i>Saravanan Thirumuruganathan, Han Li, Nan Tang, Mourad Ouzzani, Yash Govind, Derek Paulsen, Glenn M Fung, Anhai Doan</i>	
Grain: Improving Data Efficiency of Graph Neural Networks via Diversified Influence Maximization	2473
<i>Wentao Zhang, Zhi Yang, Yexin Wang, Yu Shen, Yang Li, Liang Wang, Bin Cui</i>	

Database Technology for the Masses: Sub-Operators as First-Class Entities.....	2483
<i>Maximilian Bandle, Jana Giceva</i>	
Columnar Storage and List-based Processing for Graph Database Management Systems.....	2491
<i>Pranjal Gupta, Amine Mhedhbi, Semih Salihoglu</i>	
Phoebe: A Learning-based Checkpoint Optimizer	2505
<i>Yiwen Zhu, Matteo Interlandi, Abhishek Roy, Krishnadhan Das, Hiren Patel, Malay Bag, Hitesh Sharma, Alekh Jindal</i>	
Tailoring Data Source Distributions for Fairness-aware Data Integration	2519
<i>Fatemeh Nargesian, Abolfazl Asudeh, H. V. Jagadish</i>	
Missing Value Imputation on Multidimensional Time Series	2533
<i>Parikshit Bansal, Prathamesh Deshpande, Sunita Sarawagi</i>	
Horizon: Scalable Dependency-driven Data Cleaning	2546
<i>El Kindi Rezig, Mourad Ouzzani, Walid G. Aref, Ahmed Elmagarmid, Ahmed Mahmood, Michael Stonebraker</i>	
Declarative Data Serving: The Future of Machine Learning Inference on the Edge	2555
<i>Ted Shaowang, Nilesh Jain, Dennis Matthews, Sanjay Krishnan</i>	
Auto-Pipeline: Synthesize Data Pipelines By-Target Using Reinforcement Learning and Search ...	2563
<i>Junwen Yang, Yeye He, Surajit Chaudhuri</i>	
Explaining Inference Queries with Bayesian Optimization	2576
<i>Brandon Lockhart, Jinglin Peng, Weiyuan Wu, Jiannan Wang, Eugene Wu</i>	
Decomposed Bounded Floats for Fast Compression and Queries.....	2586
<i>Chunwei Liu, Hao Jiang, John Paparrizos, Aaron J Elmore</i>	
Beyond Equi-joins: Ranking, Enumeration and Factorization.....	2599
<i>Nikolaos Tziavelis, Wolfgang Gatterbauer, Mirek Riedewald</i>	
Exathlon: A Benchmark for Explainable Anomaly Detection over Time Series.....	2613
<i>Vincent Jacob, Fei Song, Arnaud Stiegler, Bijan Rad, Yanlei Diao, Nesime Tatbul</i>	
Progressive Compressed Records: Taking a Byte out of Deep Learning Data	2627
<i>Michael Kuchnik, George Amvrosiadis, Virginia Smith</i>	
TQEL: Framework for Query-Driven Linking of Top-K Entities in Social Media Blogs	2642
<i>Abdulrahman Alsaudi, Yasser Altowim, Sharad Mehrotra, Yaming Yu</i>	

TABLE OF CONTENTS

Demonstrations

KDV-Explorer: A Near Real-Time Kernel Density Visualization System for Spatial Analysis.....	2655
<i>Tsz Nam Chan, Pak Lon Ip, Leong Hou U, Weng Hou Tong, Shivansh Mittal, Ye Li, Reynold Cheng</i>	
Refiner: A Reliable Incentive-Driven Federated Learning System Powered by Blockchain.....	2659
<i>Zhebin Zhang, Dajie Dong, Yuhang Ma, Yilong Ying, Dawei Jiang, Ke Chen, Lidan Shou, Gang Chen</i>	
MultiCategory: Multi-model Query Processing Meets Category Theory and Functional Programming	2663
<i>Valter Uotila, Jiaheng Lu, Dieter Gawlick, Zhen Hua Liu, Souripriya Das, Gregory Pogossiants</i>	
Qquirrel: Continuous Query Processing over Acyclic Relational Schemas.....	2667
<i>Qichen Wang, Chaoqi Zhang, Danish Alsayed, Ke Yi, Bin Wu, Feifei Li, Chaoqun Zhan</i>	
DeFiHap: Detecting and Fixing HiveQL Anti-Patterns.....	2671
<i>Yuetian Mao, Shuai Yuan, Nan Cui, Tianjiao Du, Beijun Shen, Yuting Chen</i>	
A Demonstration of KGLac: A Data Discovery and Enrichment Platform for Data Science.....	2675
<i>Ahmed Helal, Mossad Helali, Khaled Ammar, Essam Mansour</i>	
Assessing the Existence of a Model in your Data with ADESIT	N/A
<i>Pierre Faure-Giovagnoli, Marie Le Guilly, Vasile-Marian Scuturici, Jean-Marc Petit</i>	
Path Advisor: A Multi-Functional Campus Map Tool for Shortest Path.....	2683
<i>Yinzhaoyan, Raymond Chi-Wing Wong</i>	
Intermittent Human-in-the-Loop Model Selection using Cerebro: A Demonstration.....	2687
<i>Liangde Li, Supun C Nakandala, Arun Kumar</i>	
Low-Latency Compilation of SQL Queries to Machine Code.....	2691
<i>Henning Funke, Jens Teubner</i>	
Sound of Databases: Sonification of a Semantic Web Database Engine	2695
<i>Sven Groppe, Rico Klinckenberg, Benjamin Warnke</i>	
HyMAC: A Hybrid Matrix Computation System	2699
<i>Zihao Chen, Zhizhen Xu, Chen Xu, Juan Soto, Volker Markl, Weining Qian, Aoying Zhou</i>	
GraphScope: A One-Stop Large Graph Processing System.....	2703
<i>Jingbo Xu, Zhanning Bai, Wenfei Fan, Longbin Lai, Xue Li, Zhao Li, Zhengping Qian, Lei Wang, Yanyan Wang, Wenyan Yu, Kai Zeng, Jingren Zhou</i>	

Just Move It! Dynamic Parameter Allocation in Action	2707
<i>Alexander Renz-Wieland, Tobias Drobisch, Zoi Kaoudi, Rainer Gemulla, Volker Markl</i>	
CBench: Demonstrating Comprehensive Evaluation of Question Answering Systems over Knowledge Graphs Through Deep Analysis of Benchmarks	2711
<i>Abdelghny Orogat, Ahmed El-robby</i>	
PostCENN: PostgreSQL with Machine Learning Models for Cardinality Estimation	2715
<i>Lucas Woltmann, Dominik Olwig, Claudio Hartmann, Dirk Habich, Wolfgang Lehner</i>	
DENOUNCER: Detection of Unfairness in Classifiers	2719
<i>Jinyang Li, Yuval Moskovitch, H. V. Jagadish</i>	
A Demonstration of QARTA: An ML-based System for Accurate Map Services	2723
<i>Sofiane Abbar, Rade Stanojevic, Mashaal Musleh, Mohamed M Elshrif, Mohamed Mokbel</i>	
TraNCE: Transforming Nested Collections Efficiently	2727
<i>Jaclyn Smith, Michael Benedikt, Brandon Moore, Milos Nikolic</i>	
Debugging Missing Answers for Spark Queries over Nested Data with Breadcrumb	2731
<i>Ralf Diestelkämper, Seokki Lee, Boris Glavic, Melanie Herschel</i>	
Demonstration of Panda: A Weakly Supervised Entity Matching System	2735
<i>Renzhi Wu, Prem Sakala, Peng Li, Xu Chu, Yeye He</i>	
Automatic Data Acquisition for Deep Learning	2739
<i>Jiabin Liu, Fu Zhu, Chengliang Chai, Yuyu Luo, Nan Tang</i>	
DBMind: A Self-Driving Platform in openGauss	2743
<i>Xuanhe Zhou, Lianyuan Jin, Ji Sun, Xinyang Zhao, Xiang Yu, Shifu Li, Tianqing Wang, Kun Li, Luyang Liu</i>	
Demonstration of Dealer: An End-to-End Model Marketplace with Differential Privacy	2747
<i>Jinfei Liu, Qiongqiong Lin, Jiayao Zhang, Kui Ren, Jian Lou, Junxu Liu, Li Xiong, Jian Pei, Jimeng Sun</i>	
Assassin: an Automatic claSSificAtion system baSed on algorithm SelectIoN.....	2751
<i>Tianyu Mu, Hongzhi Wang, Shenghe Zheng, Shaoqing Zhang, Cheng Liang, Haoyun Tang</i>	
ATLANTIC: Making Database Differentially Private and Faster with Accuracy Guarantee	2755
<i>Lei Cao, Dongqing Xiao, Yizhou Yan, Samuel Madden, Guoliang Li</i>	
Demonstration of Marius: Graph Embeddings with a Single Machine.....	2759
<i>Anders Carlsson, Anze Xie, Jason M Mohoney, Roger Waleffe, Shanan Peters, Theodoros Rekatsinas, Shivaram Venkataraman</i>	
From Papers to Practice: The openclean Open-Source Data Cleaning Library.....	2763
<i>Heiko Mueller, Sonia Castelo, Munaf A Qazi, Juliana Freire</i>	
Demonstration of Apperception: A Database Management System for Geospatial Video Data	2767
<i>Vanessa Lin, Yongming Ge, Maureen Daum, Alvin Cheung, Brandon Haynes, Magdalena Balazinska</i>	

Automated energy consumption forecasting with EnForce.....	2771
<i>Mary Karatzoglidi, Paraskevas Kerasiotis, Verena Kantere</i>	
RealGraph-Web: A Graph Analysis Platform on the Web	2775
<i>Myung-Hwan Jang, Yong-Yeon Jo, Sang-Wook Kim</i>	
Interactive Demonstration of SQLCHECK	2779
<i>Arthita Ghosh, Arpit Narechania, Visweswara Sai Prashanth Dintyala, Su Timurturkan, Joy Arulraj, Deven Bansod</i>	
T-Cove: An exposure tracing System based on Cleaning Wi-Fi Events on Organizational Premises	2783
<i>Yiming Lin, Pramod Khargonekar, Sharad Mehrotra, Nalini Venkatasubramanian</i>	
Demonstration of Generating Explanations for Black-Box Algorithms Using Lewis.....	2787
<i>Paul Y Wang, Sainyam Galhotra, Romila Pradhan, Babak Salimi</i>	
Auctus: A Dataset Search Engine for Data Discovery and Augmentation.....	2791
<i>Sonia Castelo, Remi Rampin, Aécio Santos, Aline Bessa, Fernando Chirigati, Juliana Freire</i>	
A Demonstration of Relic: A System for REtrospective Lineage InfeRence of Data Workflows.....	2795
<i>Mohammed Suhail Rehman, Silu Huang, Aaron J Elmore</i>	
SChain: A Scalable Consortium Blockchain Exploiting Intra- and Inter-Block Concurrency	2799
<i>Zhihao Chen, Haizhen Zhuo, Quanqing Xu, Xiaodong Qi, Chengyu Zhu, Zhao Zhang, Cheqing Jin, Aoying Zhou, Ying Yan, Hui Zhang</i>	
EPICGen: An Experimental Platform for Indoor Congestion Generation and Forecasting	2803
<i>Chrysovalantis Anastasiou, Constantinos Costa, Panos K. Chrysanthis, Cyrus Shahabi</i>	
Wikinegata: a Knowledge Base with Interesting Negative Statements.....	2807
<i>Hiba Arnaout, Simon Razniewski, Gerhard Weikum, Jeff Z. Pan</i>	
Full Encryption: An end to end encryption mechanism in GaussDB.....	2811
<i>Liang Guo, Jinwei Zhu, Jiayang Liu, Kun Cheng</i>	
DatAgent: The Imminent Age of Intelligent Data Assistants	2815
<i>Antonis Mandamadiotis, Georgia Koutrika, Stavroula Eleftherakis, Apostolis Glenis, Dimitrios Skoutas, Yannis Stavarakas</i>	
DICE: Data Discovery by Example	2819
<i>El Kindi Rezig, Anshul Bhandari, Anna Fariha, Benjamin Price, Allan Vanterpool, Vijay Gadepally, Michael Stonebraker</i>	
AnyOLAP: Analytical Processing of Arbitrary Data-Intensive Applications without ETL.....	2823
<i>Felix Schuhknecht, Aaron Priesterroth, Justus Henneberg, Reza Salkhordeh</i>	
A Demonstration of the Exathlon Benchmarking Platform for Explainable Anomaly Detection.....	2827
<i>Vincent Jacob, Fei Song, Arnaud Stiegler, Bijan Rad, Yanlei Diao, Nesime Tatbul</i>	
An Intermediate Representation for Hybrid Database and Machine Learning Workloads	2831
<i>Amir Shaikhha, Maximilian Schleich, Dan Olteanu</i>	

How Divergent Is Your Data?.....	2835
<i>Eliana Pastor, Andrew Gavgavian, Elena Baralis, Luca De Alfaro</i>	
An Extensible and Reusable Pipeline for Automated Utterance Paraphrases.....	2839
<i>Auday Berro, Mohammad-ali Yaghub Zade Fard, Marcos Baez, Boualem Benatallah, Khalid Benabdeslem</i>	
Compliant Geo-distributed Data Processing in Action	2843
<i>Kaustubh Beedkar, David Brekardin, Jorge Arnulfo Quiane Ruiz, Volker Markl</i>	
Query-Driven Video Event Processing for the Internet of Multimedia Things.....	2847
<i>Piyush Yadav, Dhaval Salwala, Felipe Pontes, Praneet Dhingra, Edward Curry</i>	
A Demonstration of NoDA: Unified Access to NoSQL Stores.....	2851
<i>Nikolaos Koutroumanis, Kousathanas Nikolaos, Christos Doulkeridis, Akrivi Vlachou</i>	
AutoExecutor: Predictive Parallelism for Spark SQL Queries.....	2855
<i>Rathijit Sen, Abhishek Roy, Alekh Jindal, Rui Fang, Jeff Zheng, Xiaolei Liu, Ruiping Li</i>	
Catch a Blowfish Alive: A Demonstration of Policy-Aware Differential Privacy for Interactive Data Exploration	2859
<i>Jiaxiang Liu, Karl Knopf, Yiqing Tan, Bolin Ding, Xi He</i>	
RONIN: Data Lake Exploration	2863
<i>Paul Ouellette, Aidan Sciortino, Fatemeh Nargesian, Bahar Ghadiri Bashardoost, Erkang Zhu, Ken Pu, Renée J. Miller</i>	
SAND in Action: Subsequence Anomaly Detection for Streams	2867
<i>Paul Boniol, John Paparrizos, Themis Palpanas, Michael Franklin</i>	
Valentine in Action: Matching Tabular Data at Scale	2871
<i>Christos Koutras, Kyriakos Psarakis, George Siachamis, Andra Ionescu, Marios Fragkoulis, Angela Bonifati, Asterios Katsifodimos</i>	
GEDet: Detecting Erroneous Nodes with A Few Examples	2875
<i>Sheng Guan, Hanchao Ma, Sutanay Choudhury, Yinghui Wu</i>	

Industrial and Applications

GraphScope: A Unified Engine for Big Graph Processing.....	2879
<i>Wenfei Fan, Longbin Lai, Xue Li, Yong Li, Zhao Li, Zhengping Qian, Chao Tian, Lei Wang, Jingbo Xu, Youyang Yao, Qiang Yin, Wenyuan Yu, Kai Zeng, Kun Zhao, Jingren Zhou, Diwen Zhu, Rong Zhu</i>	
Davos: A System for Interactive Data-Driven Decision Making.....	2893
<i>Zeyuan Shang, Emanuel Zraggen, Benedetto Buratti, Philipp Eichmann, Navid Karimeddiny, Charlie Meyer, Wesley Runnels, Tim Kraska</i>	

Mixer: Efficiently Understanding and Retrieving Visual Content at Web-Scale	2906
<i>Mengbai Xiao, An Qin, Yongwei Wu, Xinjie Huang, Xiaodong Zhang</i>	
Towards A Polyglot Framework for Factorized ML.....	2918
<i>David A Justo, Shaoqing Yi, Lukas Stadler, Nadia Polikarpova, Arun Kumar</i>	
The End of Moore's Law and the Rise of The Data Processor.....	2932
<i>Niv Dayan, Yuval Rochman, Iddo Naiss, Shmuel Dashevsky, Noam Rabinovich, Edward Bortnikov, Igal Maly, Ofer Frishman, Itai Ben Zion, Avraham, Moshe Twitto, Uri Beitler, Evgeni Ginzburg, Mark Mokryn</i>	
tf.data: A Machine Learning Data Processing Framework	2945
<i>Derek Murray, Jiri Simsa, Ana Klimovic, Ihor Indyk</i>	
Not Black-Box Anymore! Enabling Analytics-Aware Optimizations in Teradata Vantage.....	2959
<i>Mohamed Eltabakh, Anantha Subramanian, Awny Alomari, Mohammed Al-kateb, Sanjay Nair, Mahbub Hasan, Wellington Cabrera, Charles Zhang, Amit Kishore, Snigdha Prasad</i>	
Fangorn: Adaptive Execution Framework for Heterogeneous Workloads on Shared Clusters	2972
<i>Yingda Chen, Jiamang Wang, Yifeng Lu, Ying Han, Zhiqiang Lv, Xuebin Min, Hua Cai, Wei Zhang, Haochuan Fan, Chao Li, Tao Guan, Wei Lin, Yangqing Jia, Jingren Zhou</i>	
Napa: Powering Scalable Data Warehousing with Robust Query Performance at Google.....	2986
<i>Ankur Agiwal, Kevin Lai, Gokul Nath Babu Manoharan, Indrajit Roy, Jagan Sankaranarayanan, Hao Zhang, Tao Zou, Jim Chen, Min Chen, Ming Dai, Thanh Do, Haoyu Gao, Haoyan Geng, Raman Grover, Bo Huang, Yanlai Huang, Adam Li, Jianyi Liang, Tao Lin, Li Liu, Yao Liu, Xi Mao, Maya Meng, Prashant Mishra, Jay Patel, Rajesh Sr, Vijayshankar Raman, Sourashis Roy, Mayank Singh Shishodia, Tianhang Sun, Justin Tang, Jun Tatemura, Sagar Trehan, Ramkumar Vadali, Prasanna Venkatasubramanian, Joey Zhang, Kefei Zhang, Yupu Zhang, Zeleng Zhuang, Goetz Graefe, Divy Agrawal, Jeff Naughton, Sujata Kosalge, Hakan Hacigumus</i>	
The Art of Balance: A RateupDB Experience of Building a CPU/GPU Hybrid Database Product.....	2999
<i>Rubao Lee, Minghong Zhou, Chi Li, Shenggang Hu, Jianping Teng, Dongyang Li, Xiaodong Zhang</i>	
RAMP-TAO: Layering Atomic Transactions on Facebook's Online TAO Data Store.....	3014
<i>Audrey Cheng, Xiao Shi, Lu Pan, Anthony Simpson, Neil Wheaton, Shilpa Lawande, Nathan Bronson, Peter Bailis, Natacha Crooks, Ion Stoica</i>	
openGauss: An Autonomous Database System.....	3028
<i>Guoliang Li, Xuanhe Zhou, Ji Sun, Xiang Yu, Yue Han, Lianyuan Jin, Wenbo Li, Tianqing Wang, Shifu Li</i>	
Hyperspace: The Indexing Subsystem of Azure Synapse.....	3043
<i>Rahul Potharaju, Terry Kim, Eunjin Song, Wentao Wu, Lev Novik, Apoorve Dave, Pouria Pirzadeh, Andrew Fogarty, Gurleen Dhody, Jiyi Li, Vidip Acharya, Sinduja Ramanujam, Nico Bruno, Cesar Galindo-legaria, Vivek Narasayya, Surajit Chaudhuri, Anil Nori, Tomas Talius, Raghu Ramakrishnan</i>	
SpeakNav: Voice-based Route Description Language Understanding for Template Driven Path Search.....	3056
<i>Bolong Zheng, Lei Bi, Juan Cao, Hua Chai, Jun Fang, Lu Chen, Yunjun Gao, Xiaofang Zhou, Christian S Jensen</i>	

Railgun: managing large streaming windows under MAD requirements.....	3069
<i>Ana Sofia Gomes, João Oliveirinha, Pedro Cardoso, Pedro Bizarro</i>	
Big Metadata: When Metadata is Big Data	3083
<i>Pavan Edara, Mosha Pasumansky</i>	
Tanium Reveal: A Federated Search Engine for Querying Unstructured File Data on Large Enterprise Networks	3096
<i>Joshua F Stoddard, Adam Mustafa, Naveen Goela</i>	
Hazelcast Jet: Low-latency Stream Processing at the 99.99th Percentile.....	3110
<i>Can Gencer, Marko Topolnik, Viliam Ďurina, Emin Demirci, Basri Kahveci, Ali Gürbüz, Jozsef Bartok, Grzegorz Gierlach, František Hartman, Ufuk Yilmaz, Ondřej Lukas, Mehmet Doğan, Mohamed Mandouh, Marios Fragkoulis, Asterios Katsifodimos</i>	
SparkCruise: Workload Optimization in Managed Spark Clusters at Microsoft.....	3122
<i>Abhishek Roy, Alekh Jindal, Priyanka Gomatam, Xiating Ouyang, Ashit Gosalia, Nishkam Ravi, Swinky Mann, Prakhar Jain</i>	
Watermarks in Stream Processing Systems: Semantics and Comparative Analysis of Apache Flink and Google Cloud Dataflow	3135
<i>Edmon Begoli, Tyler Akidau, Slabva Chernyak, Fabian Hueske, Kathryn Knight, Kenneth Knowles, Daniel Mills, Dan Sotolongo</i>	
The Cosmos Big Data Platform at Microsoft: Over a Decade of Progress and a Decade to Look Forward	3148
<i>Conor Power, Hiren Patel, Alekh Jindal, Jyoti Leeka, Bob Jenkins, Michael Rys, Ed Triou, Dexin Zhu, Lucky Katahanas, Chakrapani Bhat Talapady, Josh Rowe, Fan Zhang, Rich Draves, Ivan Santa, Amrish Kumar</i>	
The Evolution of Amazon Redshift.....	3162
<i>Ippokratis Pandis</i>	
Using VDMS to Index and Search 100M Images	3240
<i>Luis Remis, Chaunte W Lacewell</i>	

Tutorials

On the Limits of Machine Knowledge: Completeness, Recall and Negation in Web-scale Knowledge Bases.....	3175
<i>Simon Razniewski, Hiba Arnaout, Shrestha Ghosh, Fabian Suchanek</i>	
Managing ML Pipelines: Feature Stores and the Coming Wave of Embedding Ecosystems	3178
<i>Laurel Orr, Atindriyo Sanyal, Xiao Ling, Karan Goel, Megan Leszczynski</i>	
Data Augmentation for ML-driven Data Preparation and Integration	3182
<i>Yuliang Li, Xiaolan Wang, Zhengjie Miao, Wang-Chiew Tan</i>	
Array DBMS: Past, Present, and (Near) Future	3186
<i>Ramon Antonio Rodrigues Zalipynis</i>	

Machine Learning for Databases	3190
<i>Guoliang Li, Xuanhe Zhou, Lei Cao</i>	
Extending the Lifetime of NVM: Challenges and Opportunities.....	3194
<i>Saeed Kargar, Faisal Nawab</i>	
New Trends in High-D Vector Similarity Search: AI-driven, Progressive, and Distributed.....	3198
<i>Karima Echihabi, Themis Palpanas, Kostas Zoumpatianos</i>	
Machine Learning for Cloud Data Systems: The Promise, the Progress, and the Path Forward	3202
<i>Alekh Jindal, Matteo Interlandi</i>	

Endowment Awards

It's not just Cookies and Tea	3206
<i>Susan Davidson</i>	
Evolution of a Compiling Query Engine	3207
<i>Thomas Neumann</i>	
Make Your Database System Dream of Electric Sheep: Towards Self-Driving Operation.....	3211
<i>Andrew Pavlo, Matthew Butrovich, Lin Ma</i>	
<i>Prashanth Menon, Wan Shen Lim, Dana Van Aken, William Zhang</i>	

Keynotes

Towards instance-optimized data systems	3222
<i>Tim Kraska</i>	
Knowledge Graphs 2021: A Data Odyssey.....	3233
<i>Gerhard Weikum</i>	

Panel

The future of data(base) education: Is the "cow book" dead?	3239
<i>Zachary G. Ives, Rachel Pottinger, Arun Kumar, Johannes Gehrke, Jana Giceva</i>	