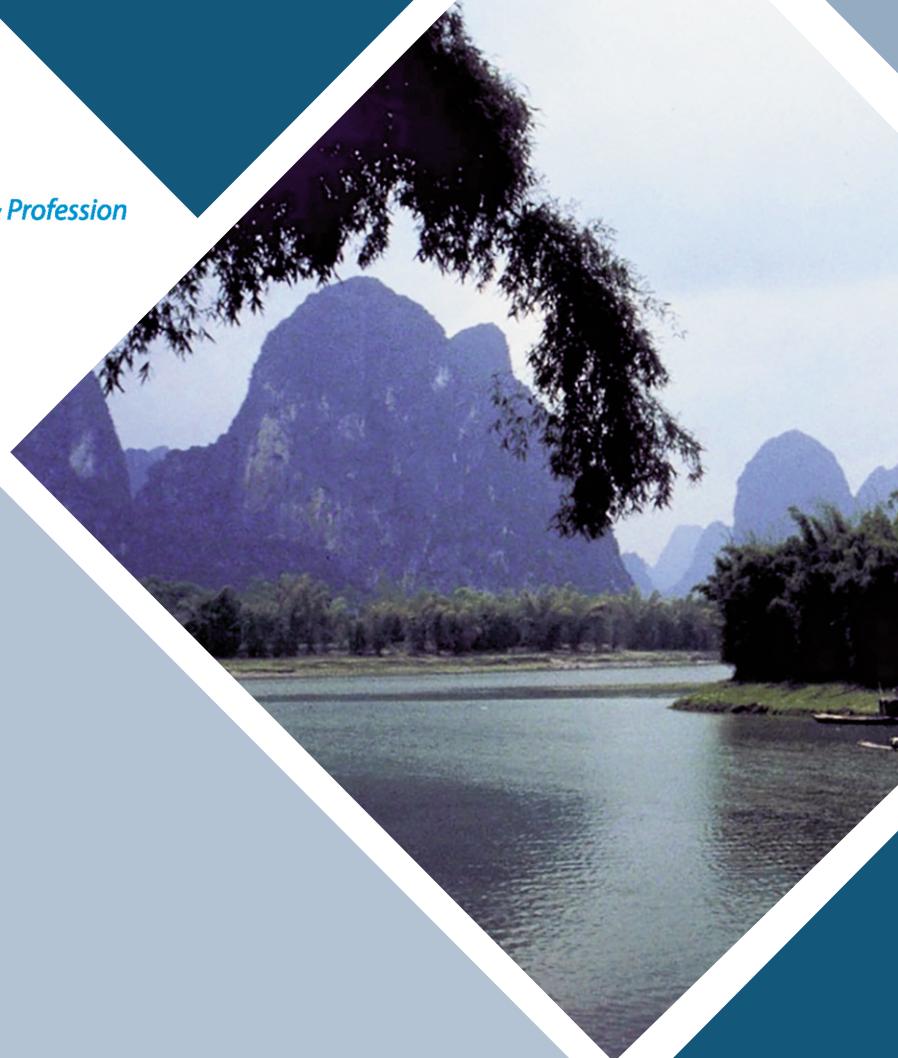


June 20–25, 2021
Virtual Event, China



Association for
Computing Machinery

Advancing Computing as a Science & Profession



SIGMOD '21

Proceedings of the 2021 International Conference on
Management of Data

Sponsored by:

ACM SIGMOD

General Chairs:

Guoliang Li, Tsinghua University, China

Zhanhuai Li, Northwestern Polytechnical University, China

Program Chairs:

Stratos Idreos, Harvard University, USA

Divesh Srivastava, AT&T, USA



**Association for
Computing Machinery**

Advancing Computing as a Science & Profession

The Association for Computing Machinery

1601 Broadway, 10th Floor
New York, NY 10019-7434

Copyright © 2021 by the Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission to republish from: permissions@acm.org or Fax +1 (212) 869-0481.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through www.copyright.com.

ISBN: 978-1-4503-8343-1

Additional copies may be ordered prepaid from:

ACM Order Department

PO Box 30777
New York, NY 10087-0777, USA

Phone: 1-800-342-6626 (USA and Canada)

+1-212-626-0500 (Global)

Fax: +1-212-944-1318

E-mail: acmhelp@acm.org

Hours of Operation: 8:30 am – 4:30 pm ET

Printed in the USA

Welcome to SIGMOD 2021 – The 2021 ACM SIGMOD International Conference on the Management of Data

This year, due to the global uncertainties, travel restrictions, and other potential difficulties associated with Covid-19, SIGMOD is being held entirely online, instead of at its originally planned location of Xi'an, Shaanxi, China. The online SIGMOD conference is complemented with a local physical event at Xi'an, primarily targeting the data management community in China. Despite the challenging times that we find ourselves in, we have an exciting technical program with outstanding research, industrial and demonstration track presentations, keynotes, tutorials, panels, and the awards session.

For the first time, SIGMOD is being held round the clock, with each technical talk presented twice 12 hours apart, to better accommodate online participants from around the world. Also, for the first time, presentations are being grouped into curated sessions to give participants a cohesive, single track experience on a variety of leading edge topics in data management. We are using the latest technologies to keep SIGMOD vibrant, and we will be archiving most SIGMOD presentations, for those who want to review them at a later date.

This year, with the approval of the SIGMOD EC, we introduced two new categories of papers in the Research Track, (a) Data Science & Engineering and (b) Applications, to complement the traditional Data Management category. Data Science & Engineering papers focused on data-intensive components of data science pipelines, solving problems in areas of interest to the community inspired by real applications. Applications papers presented novel applications of data management systems and technologies to inspire future research in the community.

In the Research Track this year, we received 450 research submissions (172 for Round 1 and 278 for Round 2), which were extensively reviewed by 175 program committee members, 23 associate editors, and several external reviewers. We accepted 188 submissions (a 41.8% acceptance rate), most of them after a revision phase that gave authors 10+ weeks to revise and resubmit their papers in response to the reviewer comments. This year we introduced a new set of detailed reviewing instructions focused on reviewing constructively as well as redesigned the review forms to promote constructive reviewing. In addition, we introduced a new step in the reviewing process, the Review Quality week where the associate editors check reviews for certain quality criteria and probe reviewers for constructive rewrites before reviews are released to the authors. In addition, the authors were able to provide structured feedback directly to the associate editors and the program chairs about review quality. Overall, more than 300 reviews were updated for quality during this process leading to a higher number of revision requests.

In addition to the Research Track, the Industrial Track selected 21 papers from 54 submissions; the Demonstration Track selected 27 demonstrations from 75 submissions; the Tutorial Track selected 8 tutorials from 20 submissions and the Student Research Competition selected all 18 submissions for the second round of competition.

This year, we will have two exciting keynote talks, reflecting emerging topics of great interest to the data management community: “Utilizing (and Designing) Modern Hardware for Data-Intensive Computations: The Role of Abstraction” by Kenneth A. Ross (Columbia University) and “Deep Data Integration” by Wang-Chiew Tan (Facebook AI).

In addition, we will have two timely and interesting panels: “Data Management to Social Science and Back in the Future of Work” organized by Sihem Amer-Yahia (CNRS) and Senjuti Basu Roy (New Jersey Institute of Technology), and “Automation of Data Prep, ML, and Data Science: New Cure or Snake Oil?” organized by Arun Kumar (University of California, San Diego).

Thus, SIGMOD 2021 will feature an exciting and rich program, with 188 research papers, 21 industrial papers, 27 demonstrations, 8 tutorials, 2 keynotes, and 2 panels, together with online social and sponsor events (which are being organized as of the writing of this note). Assembling this program and these proceedings requires an immense amount of effort from numerous people, to whom we are very grateful. We thank the members of the SIGMOD organizing committee and the PC members of the various tracks, as well as the staff and volunteers, for doing an outstanding job and going above and beyond what was required. We have been extremely heartened by the level of dedication and professionalism we have seen in the course of organizing SIGMOD 2021, especially during this challenging time.

We are also very grateful to the SIGMOD Executive Committee, as well as former SIGMOD PC chairs, for helping us navigate many issues and for supporting our new initiatives. We thank ACM and Sheridan, especially Lisa Tolles, for helping us put together the proceedings. We are also deeply appreciative of the support team behind Microsoft’s Conference Management Toolkit, who have always been prompt and helpful in answering our questions. Finally, we are extremely grateful to all of our sponsors and supporters. Your continuing backing for our community and for SIGMOD is deeply appreciated.

Welcome to SIGMOD 2021. We hope you will enjoy the conference and see you online!

Guoliang Li
Zhanhuai Li
General Chairs

Stratos Idreos
Divesh Srivastava
Program Chairs

Table of Contents

SIGMOD 2021 Organization xxiv

SIGMOD 2021 Sponsor & Supporters xxxiv

Keynote Talks

- Utilizing (and Designing) Modern Hardware for Data-Intensive Computations:
The Role of Abstraction 1
Kenneth A. Ross (*Columbia University*)
- Deep Data Integration 2
Wang-Chiew Tan (*Facebook AI*)

Award Talks

- A Client-centric Approach to Transactional Databases 3
Natacha Crooks (*University of California, Berkeley*)
- Structure-Aware Machine Learning over Multi-Relational Databases 6
Maximilian Schleich (*University of Washington*)
- Scalable Distributed Transaction Processing on Modern RDMA-enabled Networks 8
Erfan Zamanian (*eBay*)
- Memory-Efficient Search Trees for Database Management Systems 9
Huachen Zhang (*Tsinghua University*)

Research Data Management Track Papers

- MuSE Graphs for Flexible Distribution of Event Stream Processing in Networks 10
Samira Akili (*Humboldt-Universität zu Berlin*), Matthias Weidlich (*Humboldt-Universität zu Berlin*)
- HADAD: A Lightweight Approach for Optimizing Hybrid Complex Analytics Queries 23
Rana Alotaibi (*University of California, San Diego*), Bogdan Cautis (*University of Paris-Saclay*),
Alin Deutsch (*University of California, San Diego*),
Ioana Manolescu (*Inria & Institut Polytechnique de Paris*)
- Fast and Exact Outlier Detection in Metric Spaces: A Proximity Graph-based Approach 36
Daichi Amagata (*Osaka University & PRESTO*), Makoto Onizuka (*Osaka University*),
Takahiro Hara (*Osaka University*)
- Fast Density-Peaks Clustering: Multicore-based Parallelization Approach 49
Daichi Amagata (*Osaka University & PRESTO*), Takahiro Hara (*Osaka University*)
- Exploring Ratings in Subjective Databases 62
Sihem Amer-Yahia (*CNRS, University Grenoble Alpes*), Tova Milo (*Tel Aviv University*),
Brit Youngmann (*Tel Aviv University*)
- SharPer: Sharding Permissioned Blockchains Over Network Clusters 76
Mohammad Javad Amiri (*University of Pennsylvania*),
Divyakant Agrawal (*University of California, Santa Barbara*),
Amr El Abbadi (*University of California, Santa Barbara*)
- FastVer: Making Data Integrity a Commodity 89
Arvind Arasu (*Microsoft Research*), Badrish Chandramouli (*Microsoft Research*),
Johannes Gehrke (*Microsoft Research*), Esha Ghosh (*Microsoft Research*),
Donald Kossmann (*Microsoft Research*), Jonathan Protzenko (*Microsoft Research*),
Ravi Ramamurthy (*Microsoft Research*), Tahina Ramananandro (*Microsoft Research*),
Aseem Rastogi (*Microsoft Research*), Srinath Setty (*Microsoft Research*),
Nikhil Swamy (*Microsoft Research*), Alexander van Renen (*Technical University of Munich*),
Min Xu (*University of Chicago*)

• Worst-Case Optimal Graph Joins in Almost No Space	102
Diego Arroyuelo (<i>Universidad Técnica Federico Santa María & Millennium Institute for Foundational Research on Data</i>), Aidan Hogan (<i>Universidad de Chile & Millennium Institute for Foundational Research on Data</i>), Gonzalo Navarro (<i>Universidad de Chile & Millennium Institute for Foundational Research on Data</i>), Juan L. Reutter (<i>Pontificia Universidad Católica de Chile & Millennium Institute for Foundational Research on Data</i>), Javier Rojas-Ledesma (<i>Universidad de Chile & Millennium Institute for Foundational Research on Data</i>), Adrián Soto (<i>Universidad Adolfo Ibáñez & Millennium Institute for Foundational Research on Data</i>)	
• REDS: Rule Extraction for Discovering Scenarios	115
Vadim Arzamasov (<i>Karlsruhe Institute of Technology</i>), Klemens Böhm (<i>Karlsruhe Institute of Technology</i>)	
• Identifying Insufficient Data Coverage for Ordinal Continuous-Valued Attributes	129
Abolfazl Asudeh (<i>University of Illinois at Chicago</i>), Nima Shahbazi (<i>University of Illinois at Chicago</i>), Zhongjun Jin (<i>University of Michigan</i>), H. V. Jagadish (<i>University of Michigan</i>)	
• A Generalized Approach for Reducing Expensive Distance Calls for A Broad Class of Proximity Problems.....	142
Jees Augustine (<i>The University of Texas at Arlington</i>), Suraj Shetiya (<i>The University of Texas at Arlington</i>), Mohammadreza Esfandiari (<i>New Jersey Institute of Technology</i>), Senjuti Basu Roy (<i>New Jersey Institute of Technology</i>), Gautam Das (<i>University of Texas Arlington</i>)	
• TREETOASTER: Towards an IVM-Optimized Compiler	155
Darshana Balakrishnan (<i>State University of New York at Buffalo</i>), Carl Nuessle (<i>State University of New York at Buffalo</i>), Oliver A. Kennedy (<i>State University of New York at Buffalo</i>), Lukasz Ziarek (<i>State University of New York at Buffalo</i>)	
• To Partition, or Not to Partition, That is the Join Question in a Real System.....	168
Maximilian Bandle (<i>Technische Universität München</i>), Jana Giceva (<i>Technische Universität München</i>), Thomas Neumann (<i>Technische Universität München</i>)	
• Compliant Geo-distributed Query Processing.....	181
Kaustubh Beedkar (<i>Technische Universität Berlin</i>), Jorge-Arnulfo Quiané-Ruiz (<i>Technische Universität Berlin & DFKI</i>), Volker Markl (<i>Technische Universität Berlin & DFKI</i>)	
• BullFrog: Online Schema Evolution via Lazy Evaluation	194
Souvik Bhattacherjee (<i>ServiceNow & University of Maryland, College Park</i>), Gang Liao (<i>University of Maryland, College Park</i>), Michael Hicks (<i>University of Maryland, College Park</i>), Daniel J. Abadi (<i>University of Maryland, College Park</i>)	
• ARM-Net: Adaptive Relation Modeling Network for Structured Data.....	207
Shaofeng Cai (<i>National University of Singapore</i>), Kaiping Zheng (<i>National University of Singapore</i>), Gang Chen (<i>Zhejiang University</i>), H. V. Jagadish (<i>University of Michigan</i>), Beng Chin Ooi (<i>National University of Singapore</i>), Meihui Zhang (<i>Beijing Institute of Technology</i>)	
• Why Do My Blockchain Transactions Fail? A Study of Hyperledger Fabric	221
Jeeta Ann Chacko (<i>Technical University of Munich</i>), Ruben Mayer (<i>Technical University of Munich</i>), Hans-Arno Jacobsen (<i>University of Toronto</i>)	
• PigPaxos: Devouring the Communication Bottlenecks in Distributed Consensus	235
Aleksey Charapko (<i>University of New Hampshire</i>), Ailidani Ailiqiang (<i>Microsoft</i>), Murat Demirbas (<i>University at Buffalo, SUNY</i>)	
• Efficient Exact Algorithms for Maximum Balanced Bipartite Graphs	248
Lu Chen (<i>Swinburne University of Technology</i>), Chengfei Liu (<i>Swinburne University of Technology</i>), Rui Zhou (<i>Swinburne University of Technology</i>), Jiajie Xu (<i>Soochow University</i>), Jianxin Li (<i>Deakin University</i>)	
• Out of Many We are One: Measuring Item Batch with Clock-Sketch	261
Peiqing Chen (<i>Peking University</i>), Dong Chen (<i>Peking University</i>), Lingxiao Zheng (<i>Peking University</i>), Jizhou Li (<i>Peking University</i>), Tong Yang (<i>Peking University & Pengcheng Laboratory</i>)	
• Efficient Approximate Algorithms for Empirical Entropy and Mutual Information	274
Xingguang Chen (<i>The Chinese University of Hong Kong</i>), Sibo Wang (<i>The Chinese University of Hong Kong</i>)	

• Evaluating Temporal Queries Over Video Feeds	287
Yueling Chen (<i>York University</i>), Xiaohui Yu (<i>York University</i>), Nick Koudas (<i>University of Toronto</i>), Ziqiang Yu (<i>Yantai University</i>)	
• Hybrid Evaluation for Distributed Iterative Matrix Computation	300
Zihao Chen (<i>East China Normal University</i>), Chen Xu (<i>East China Normal University</i>), Juan Soto (<i>Technische Universität Berlin</i>), Volker Markl (<i>Technische Universität Berlin</i>), Weining Qian (<i>East China Normal University</i>), Aoying Zhou (<i>East China Normal University</i>)	
• P2H: Efficient Distance Querying on Road Networks by Projected Vertex Separators	313
Zitong Chen (<i>The Chinese University of Hong Kong</i>), Ada Wai-Chee Fu (<i>The Chinese University of Hong Kong</i>), Minhao Jiang (<i>TuSimple</i>), Eric Lo (<i>The Chinese University of Hong Kong</i>), Pengfei Zhang (<i>The Chinese University of Hong Kong</i>)	
• Structural Generalizability: The Case of Similarity Search	326
Yodsawalai Chodpathumwan (<i>King Mongkut's University of Technology North Bangkok</i>), Arash Termehchy (<i>Oregon State University</i>), Stephen A. Ramsey (<i>Oregon State University</i>), Aayam Shrestha (<i>Oregon State University</i>), Amy Glen (<i>Oregon State University</i>), Zheng Liu (<i>Oregon State University</i>)	
• Maximizing Persistent Memory Bandwidth Utilization for OLAP Workloads	339
Björn Daase (<i>Hasso Plattner Institute, University of Potsdam</i>), Lars Jonas Bollmeier (<i>Hasso Plattner Institute, University of Potsdam</i>), Lawrence Benson (<i>Hasso Plattner Institute, University of Potsdam</i>), Tilmann Rabl (<i>Hasso Plattner Institute, University of Potsdam</i>)	
• Active Sampling Count Sketch (ASCS) for Online Sparse Estimation of a Trillion Scale Covariance Matrix	352
Zhenwei Dai (<i>Rice University</i>), Aditya Desai (<i>Rice University</i>), Reinhard Heckel (<i>Technical University of Munich</i>), Anshumali Shrivastava (<i>Rice University</i>)	
• Chucky: A Succinct Cuckoo Filter for LSM-Tree.....	365
Niv Dayan (<i>Pliops</i>), Moshe Twitto (<i>Pliops</i>)	
• On Optimizing the Trade-off between Privacy and Utility in Data Provenance	379
Daniel Deutch (<i>Tel Aviv University</i>), Ariel Frankenthal (<i>Tel Aviv University</i>), Amir Gilad (<i>Duke University</i>), Yuval Moskovitch (<i>University of Michigan</i>)	
• Efficient Exploration of Interesting Aggregates in RDF Graphs	392
Yanlei Diao (<i>Ecole Polytechnique, Institut Polytechnique de Paris & Inria</i>), Pawel Guzewicz (<i>Ecole Polytechnique, Institut Polytechnique de Paris & Inria</i>), Ioana Manolescu (<i>Inria & Ecole Polytechnique, Institut Polytechnique de Paris</i>), Mirjana Mazuran (<i>Inria & Ecole Polytechnique, Institut Polytechnique de Paris</i>)	
• To Not Miss the Forest for the Trees - A Holistic Approach for Explaining Missing Answers over Nested Data	405
Ralf Diestelkämper (<i>Universität Stuttgart - IPVS</i>), Seokki Lee (<i>University of Cincinnati</i>), Melanie Herschel (<i>Universität Stuttgart - IPVS</i>), Boris Glavic (<i>Illinois Institute of Technology</i>)	
• Instance-Optimized Data Layouts for Cloud Analytics Workloads	418
Jialin Ding (<i>Massachusetts Institute of Technology</i>), Umar Farooq Minhas (<i>Microsoft Research</i>), Badrish Chandramouli (<i>Microsoft Research</i>), Chi Wang (<i>Microsoft Research</i>), Yinan Li (<i>Microsoft Research</i>), Ying Li (<i>Microsoft</i>), Donald Kossmann (<i>Microsoft Research</i>), Johannes Gehrke (<i>Microsoft</i>), Tim Kraska (<i>Massachusetts Institute of Technology</i>)	
• Residual Sensitivity for Differentially Private Multi-Way Joins	432
Wei Dong (<i>Hong Kong University of Science and Technology</i>), Ke Yi (<i>Hong Kong University of Science and Technology</i>)	
• JSON Tiles: Fast Analytics on Semi-Structured Data.....	445
Dominik Durner (<i>Technische Universität München</i>), Viktor Leis (<i>Friedrich-Schiller-Universität Jena</i>), Thomas Neumann (<i>Technische Universität München</i>)	
• Incrementalizing Graph Algorithms.....	459
Wenfei Fan (<i>University of Edinburgh & Shenzhen Institute of Computing Sciences</i>), Chao Tian (<i>Alibaba Group</i>), Ruiqi Xu (<i>University of Edinburgh</i>), Qiang Yin (<i>Alibaba Group</i>), Wenyuan Yu (<i>Alibaba Group</i>), Jingren Zhou (<i>Alibaba Group</i>)	

• Making Graphs Compact by Lossless Contraction	472
Wenfei Fan (<i>University of Edinburgh, Shenzhen Institute of Computing Sciences, & Beihang University</i>), Yuanhao Li (<i>University of Edinburgh</i>), Muyang Liu (<i>University of Edinburgh</i>), Can Lu (<i>Shenzhen Institute of Computing Sciences</i>)	
• Klink: Progress-Aware Scheduling for Streaming Data Systems	485
Omar Farhat (<i>University of Waterloo</i>), Khuzaima Daudjee (<i>University of Waterloo</i>), Leonardo Querzoni (<i>Sapienza University of Rome</i>)	
• Conformance Constraint Discovery: Measuring Trust in Data-Driven Systems	499
Anna Fariha (<i>Microsoft & University of Massachusetts Amherst</i>), Ashish Tiwari (<i>Microsoft</i>), Arjun Radhakrishna (<i>Microsoft</i>), Sumit Gulwani (<i>Microsoft</i>), Alexandra Meliou (<i>University of Massachusetts, Amherst</i>)	
• RisGraph: A Real-Time Streaming System for Evolving Graphs to Support Sub-millisecond Per-update Analysis at Millions Ops/s	513
Guanyu Feng (<i>Tsinghua University</i>), Zixuan Ma (<i>Tsinghua University</i>), Daixuan Li (<i>Tsinghua University</i>), Shengqi Chen (<i>Tsinghua University</i>), Xiaowei Zhu (<i>Tsinghua University</i>), Wentao Han (<i>Tsinghua University</i>), Wenguang Chen (<i>Tsinghua University</i>)	
• Efficient Uncertainty Tracking for Complex Queries with Attribute-level Bounds	528
Su Feng (<i>Illinois Institute of Technology</i>), Boris Glavic (<i>Illinois Institute of Technology</i>), Aaron Huber (<i>University at Buffalo, SUNY</i>), Oliver A. Kennedy (<i>University at Buffalo, SUNY</i>)	
• ALLIGN: Aligning All-Pair Near-Duplicate Passages in Long Texts	541
Weiwei Feng (<i>Shanghai Jiao Tong University</i>), Dong Deng (<i>Rutgers University</i>)	
• Adaptive Compression for Fast Scans on String Columns	554
Yannis Foufoulas (<i>University of Athens & Athena Research Center</i>), Lefteris Sidirooulos (<i>University of Athens</i>), Eleftherios Stamatogiannakis (<i>University of Athens</i>), Yannis Ioannidis (<i>Athena Research Center & University of Athens</i>),	
• VF²Boost: Very Fast Vertical Federated Gradient Boosting for Cross-Enterprise Learning	563
Fangcheng Fu (<i>Peking University</i>), Yingxia Shao (<i>BUPT</i>), Lele Yu (<i>Peking University</i>), Jiawei Jiang (<i>ETH Zurich</i>), Huanran Xue (<i>Tencent Inc.</i>), Yangyu Tao (<i>Tencent Inc.</i>), Bin Cui (<i>Peking University</i>)	
• Explaining Black-Box Algorithms Using Probabilistic Contrastive Counterfactuals	577
Sainyam Galhotra (<i>University of Massachusetts, Amherst</i>), Romila Pradhan (<i>University of California, San Diego</i>), Babak Salimi (<i>Unievristy of California, San Diego</i>)	
• Efficiently Answering Durability Prediction Queries	591
Junyang Gao (<i>Google</i>), Yifan Xu (<i>Amazon.com</i>), Pankaj K. Agarwal (<i>Duke University</i>), Jun Yang (<i>Duke University</i>)	
• The Power of Nested Parallelism in Big Data Processing – Hitting Three Flies with One Slap –	605
Gábor E. Gévay (<i>Technische Universität Berlin</i>), Jorge-Arnulfo Quiané-Ruiz (<i>Technische Universität Berlin</i>), Volker Markl (<i>Technische Universität Berlin</i>)	
• Synthesizing Linked Data Under Cardinality and Integrity Constraints	619
Amir Gilad (<i>Duke University</i>), Shweta Patwa (<i>Duke University</i>), Ashwin Machanavajjhala (<i>Duke University</i>)	
• An In-Depth Benchmarking of Text-to-SQL Systems	632
Orest Gkini (<i>Athena Research Center</i>), Theofilos Belmpas (<i>Athena Research Center</i>), Georgia Koutrika (<i>Athena Research Center</i>), Yannis Ioannidis (<i>University of Athens & Athena Research Center</i>)	
• Sliding Window-based Approximate Triangle Counting over Streaming Graphs with Duplicate Edges	645
Xiangyang Gou (<i>Peking University</i>), Lei Zou (<i>Peking University</i>)	
• Releasing Locks As Early As You Can: Reducing Contention of Hotspots by Violating Two-Phase Locking	658
Zhihan Guo (<i>University of Wisconsin-Madison</i>), Kan Wu (<i>University of Wisconsin-Madison</i>), Cong Yan (<i>Microsoft Research</i>), Xiangyao Yu (<i>University of Wisconsin-Madison</i>)	

• Efficient and Effective Algorithms for Revenue Maximization in Social Advertising	671
Kai Han (<i>University of Science and Technology of China</i>),	
Benwei Wu (<i>University of Science and Technology of China</i>), Jing Tang (<i>National University of Singapore</i>),	
Shuang Cui (<i>University of Science and Technology of China</i>), Cigdem Aslay (<i>Aarhus University</i>),	
Laks V.S. Lakshmanan (<i>The University of British Columbia</i>)	
• VSS: A Storage System for Video Analytics	685
Brandon Haynes (<i>Gray Systems Lab & Microsoft</i>), Maureen Daum (<i>University of Washington</i>),	
Dong He (<i>University of Washington</i>), Amrita Mazumdar (<i>University of Washington</i>),	
Magdalena Balazinska (<i>University of Washington</i>), Alvin Cheung (<i>University of California, Berkeley</i>),	
Luis Ceze (<i>University of Washington</i>)	
• Small Selectivities Matter: Lifting the Burden of Empty Samples	697
Axel Hertzschuch (<i>Technische Universität Dresden</i>), Guido Moerkotte (<i>University of Mannheim</i>),	
Wolfgang Lehner (<i>Technische Universität Dresden</i>), Norman May (<i>SAP SE</i>),	
Florian Wolf (<i>SAP SE</i>), Lars Fricke (<i>SAP SE</i>)	
• ReStore - Neural Data Completion for Relational Databases	710
Benjamin Hilprecht (<i>Technical University of Darmstadt</i>), Carsten Binnig (<i>Technical University of Darmstadt</i>)	
• One WITH RECURSIVE is Worth Many GOTOs	723
Denis Hirn (<i>Universität Tübingen</i>), Torsten Grust (<i>Universität Tübingen</i>)	
• Accelerating Triangle Counting on GPU	736
Lin Hu (<i>Peking University</i>), Lei Zou (<i>Peking University</i>), Yu Liu (<i>Peking University</i>)	
• Nova-LSM: A Distributed, Component-based LSM-tree Key-value Store.....	749
Haoyu Huang (<i>University of Southern California</i>),	
Shahram Ghandeharizadeh (<i>University of Southern California</i>)	
• MIDAS: Towards Efficient and Effective Maintenance of Canned Patterns in Visual Graph Query Interfaces	764
Kai Huang (<i>Fudan University & Nanyang Technological University</i>),	
Huey Eng Chua (<i>Nanyang Technological University</i>),	
Sourav S. Bhowmick (<i>Nanyang Technological University</i>),	
Byron Choi (<i>Hong Kong Baptist University</i>), Shuigeng Zhou (<i>Fudan University</i>)	
• Point-to-Hyperplane Nearest Neighbor Search Beyond the Unit Hypersphere	777
Qiang Huang (<i>National University of Singapore</i>), Yifan Lei (<i>National University of Singapore</i>),	
Anthony K. H. Tung (<i>National University of Singapore</i>)	
• Do the Rich Get Richer? Fairness Analysis for Blockchain Incentives	790
Yuming Huang (<i>National University of Singapore</i>), Jing Tang (<i>National University of Singapore</i>),	
Qianhao Cong (<i>National University of Singapore</i>), Andrew Lim (<i>National University of Singapore</i>),	
Jianliang Xu (<i>Hong Kong Baptist University</i>),	
• COMPASS: Online Sketch-based Query Optimization for In-Memory Databases	804
Yesdaulet Izenov (<i>University of California Merced</i>), Asoke Datta (<i>University of California Merced</i>),	
Florin Rusu (<i>University of California Merced</i>), Jun Hyung Shin (<i>University of California Merced</i>)	
• A-Tree: A Dynamic Data Structure for Efficiently Indexing Arbitrary Boolean Expressions	817
Shuping Ji (<i>Institute of Software, Chinese Academy of Sciences</i>),	
Hans-Arno Jacobsen (<i>University of Toronto</i>)	
• Bidirectionally Densifying LSH Sketches with Empty Bins	830
Peng Jia (<i>Xi'an Jiaotong University</i>), Pinghui Wang (<i>Xi'an Jiaotong University</i>),	
Junzhou Zhao (<i>Xi'an Jiaotong University</i>), Shuo Zhang (<i>Xi'an Jiaotong University</i>),	
Yiyan Qi (<i>Xi'an Jiaotong University</i>), Min Hu (<i>China Mobile Research Institute</i>),	
Chao Deng (<i>China Mobile Research Institute</i>),	
Xiaohong Guan (<i>Xi'an Jiaotong University & Tsinghua University</i>)	
• Good to the Last Bit: Data-Driven Encoding with CodecDB.....	843
Hao Jiang (<i>University of Chicago</i>), Chunwei Liu (<i>University of Chicago</i>),	
John Paparrizos (<i>University of Chicago</i>), Andrew A. Chien (<i>University of Chicago</i>),	
Jihong Ma (<i>Alibaba Group</i>), Aaron J. Elmore (<i>University of Chicago</i>)	

• Towards Demystifying Serverless Machine Learning Training	857
Jiawei Jiang (<i>ETH Zurich</i>), Shaoduo Gan (<i>ETH Zurich</i>), Yue Liu (<i>ETH Zurich</i>), Fanlin Wang (<i>ETH Zurich</i>), Gustavo Alonso (<i>ETH Zurich</i>), Ana Klimovic (<i>ETH Zurich</i>), Ankit Singla (<i>ETH Zurich</i>), Wentao Wu (<i>Microsoft Research</i>), Ce Zhang (<i>ETH Zurich</i>)	
• Efficient String Sort with Multi-Character Encoding and Adaptive Sampling	872
Wen Jin (<i>Independent Researcher</i>), Weining Qian (<i>East China Normal University</i>), Aoying Zhou (<i>East China Normal University</i>)	
• Proportionality in Spatial Keyword Search	885
Georgios Kalamatianos (<i>Uppsala University</i>), Georgios J. Fakas (<i>Uppsala University</i>), Nikos Mamoulis (<i>University of Ioannina</i>)	
• Jigsaw: A Data Storage and Query Processing Engine for Irregular Table Partitioning	898
Donghe Kang (<i>The Ohio State University</i>), Ruochen Jiang (<i>The Ohio State University</i>), Spyros Blanas (<i>The Ohio State University</i>)	
• Shedding Light on Opaque Application Queries	912
Kapil Khurana (<i>Indian Institute of Science</i>), Jayant R. Haritsa (<i>Indian Institute of Science</i>)	
• Versatile Equivalences: Speeding up Subgraph Query Processing and Subgraph Matching ..	925
Hyunjoon Kim (<i>Seoul National University & SAP Labs Korea</i>), Yunyoung Choi (<i>Seoul National University</i>), Kunsoo Park (<i>Seoul National University</i>), Xuemin Lin (<i>University of New South Wales</i>), Seok-Hee Hong (<i>The University of Sydney</i>), Wook-Shin Han (<i>Pohang University of Science and Technology (POSTECH)</i>)	
• Rethink the Scan in MVCC Databases.....	938
Jongbin Kim (<i>Hanyang University</i>), Kihwang Kim (<i>Hanyang University</i>), Hyunsoo Cho (<i>Hanyang University</i>), Jaeseon Yu (<i>Hanyang University</i>), Sooyong Kang (<i>Hanyang University</i>), Hyungsoo Jung (<i>Hanyang University</i>)	
• Boosting Graph Similarity Search through Pre-Computation.....	951
Jongik Kim (<i>Jeonbuk National University</i>)	
• Combining Sampling and Synopses with Worst-Case Optimal Runtime and Quality Guarantees for Graph Pattern Cardinality Estimation	964
Kyoungmin Kim (<i>Pohang University of Science and Technology</i>), Hyeonji Kim (<i>Pohang University of Science and Technology</i>), George Fletcher (<i>Eindhoven University of Technology</i>), Wook-Shin Han (<i>Pohang University of Science and Technology</i>)	
• iTurboGraph: Scaling and Automating Incremental Graph Analytics.....	977
Seongyun Ko (<i>POSTECH</i>), Taesung Lee (<i>POSTECH</i>), Kijae Hong (<i>POSTECH</i>), Wonseok Lee (<i>POSTECH</i>), In Seo (<i>POSTECH</i>), Jiwon Seo (<i>Hanyang University</i>), Wook-Shin Han (<i>POSTECH</i>)	
• Building Advanced SQL Analytics From Low-Level Plan Operators	1001
André Kohn (<i>Technische Universität München</i>), Viktor Leis (<i>Friedrich-Schiller-Universität Jena</i>), Thomas Neumann (<i>Technische Universität München</i>)	
• Efficient Deep Learning Pipelines for Accurate Cost Estimations Over Large Scale Query Workload	1014
Johan Zhi Kang Kok (<i>National University of Singapore</i>), Gaurav (<i>GrabTaxi Holdings</i>), Sien Yi Tan (<i>GrabTaxi Holdings</i>), Feng Cheng (<i>GrabTaxi Holdings</i>), Shixuan Sun (<i>National University of Singapore</i>), Bingsheng He (<i>National University of Singapore</i>)	
• Index-Accelerated Pattern Matching in Event Stores	1023
Michael Körber (<i>University of Marburg</i>), Nikolaus Glombiewski (<i>University of Marburg</i>), Bernhard Seeger (<i>University of Marburg</i>)	
• Top-K Deep Video Analytics: A Probabilistic Approach	1037
Ziliang Lai (<i>The Chinese University of Hong Kong</i>), Chenxia Han (<i>The Chinese University of Hong Kong</i>), Chris Liu (<i>The Chinese University of Hong Kong</i>), Pengfei Zhang (<i>The Chinese University of Hong Kong</i>), Eric Lo (<i>The Chinese University of Hong Kong</i>), Ben Kao (<i>University of Hong Kong</i>)	
• Putting Things into Context: Rich Explanations for Query Answers using Join Graphs.....	1051
Chenjie Li (<i>Illinois Institute of Technology</i>), Zhengjie Miao (<i>Duke University</i>), Qitian Zeng (<i>Illinois Institute of Technology</i>), Boris Glavic (<i>Illinois Institute of Technology</i>), Sudeepa Roy (<i>Duke University</i>)	

• Auto-FuzzyJoin: Auto-Program Fuzzy Similarity Joins Without Labeled Examples	1064
Peng Li (<i>Georgia Institute of Technology</i>), Xiang Cheng (<i>Georgia Institute of Technology</i>), Xu Chu (<i>Georgia Institute of Technology</i>), Yeye He (<i>Microsoft Research</i>), Surajit Chaudhuri (<i>Microsoft Research</i>)	
• Building Fast and Compact Sketches for Approximately Multi-Set Multi-Membership Querying	1077
Rundong Li (<i>Xi'an Jiaotong University</i>), Pinghui Wang (<i>Xi'an Jiaotong University</i>), Jiongli Zhu (<i>Xi'an Jiaotong University</i>), Junzhou Zhao (<i>Xi'an Jiaotong University</i>), Jia Di (<i>Xi'an Jiaotong University</i>), Xiaofei Yang (<i>Xi'an Jiaotong University</i>), Kai Ye (<i>Xi'an Jiaotong University</i>)	
• Asynchronous Prefix Recoverability for Fast Distributed Stores.....	1090
Tianyu Li (<i>Massachusetts Institute of Technology</i>), Badrish Chandramouli (<i>Microsoft Research</i>), Jose M. Faleiro (<i>Microsoft Research</i>), Samuel Madden (<i>Massachusetts Institute of Technology</i>), Donald Kossmann (<i>Microsoft Research</i>)	
• Imminence Monitoring of Critical Events: A Representation Learning Approach	1103
Yan Li (<i>University of Massachusetts, Lowell</i>), Tingjian Ge (<i>University of Massachusetts, Lowell</i>)	
• PRISM: Private Verifiable Set Computation over Multi-Owner Outsourced Databases	1116
Yin Li (<i>Dongguan University of Technology</i>), Dhrubajyoti Ghosh (<i>University of California, Irvine</i>), Peeyush Gupta (<i>University of California, Irvine</i>), Sharad Mehrotra (<i>University of California, Irvine</i>), Nisha Panwar (<i>Augusta University</i>), Shantanu Sharma (<i>University of California, Irvine</i>)	
• Combining Aggregation and Sampling (Nearly) Optimally for Approximate Query Processing	1129
Xi Liang (<i>University of Chicago</i>), Stavros Sintos (<i>University of Chicago</i>), Zechao Shang (<i>Snowflake Computing</i>), Sanjay Krishnan (<i>University of Chicago</i>)	
• TENET: Joint Entity and Relation Linking with Coherence Relaxation.....	1142
Xueling Lin (<i>The Hong Kong University of Science and Technology</i>), Lei Chen (<i>The Hong Kong University of Science and Technology</i>), Chaorui Zhang (<i>Huawei Technologies</i>)	
• Don't Look Back, Look into the Future: Prescient Data Partitioning and Migration for Deterministic Database Systems	1156
Yu-Shan Lin (<i>National Tsing Hua University</i>), Ching Tsai (<i>National Tsing Hua University</i>), Tz-Yu Lin (<i>National Tsing Hua University</i>), Yun-Sheng Chang (<i>National Tsing Hua University</i>), Shan-Hung Wu (<i>National Tsing Hua University</i>)	
• Logical Schema Design that Quantifies Update Inefficiency and Join Efficiency	1169
Sebastian Link (<i>University of Auckland</i>), Ziheng Wei (<i>University of Auckland</i>)	
• Properties of Inconsistency Measures for Databases.....	1182
Ester Livshits (<i>Technion</i>), Rina Kochirgan (<i>Technion</i>), Segev Tsur (<i>Technion</i>), Ihab F. Ilyas (<i>University of Waterloo</i>), Benny Kimelfeld (<i>Technion</i>), Sudeepa Roy (<i>Duke University</i>)	
• Graph Iso/Auto-morphism: A Divide-&-Conquer Approach	1195
Can Lu (<i>The Chinese University of Hong Kong</i>), Jeffrey Xu Yu (<i>The Chinese University of Hong Kong</i>), Zhiwei Zhang (<i>Beijing Institute of Technology</i>), Hong Cheng (<i>Chinese University of Hong Kong</i>)	
• Cache-Efficient Fork-Processing Patterns on Large Graphs	1208
Shengliang Lu (<i>National University of Singapore</i>), Shixuan Sun (<i>National University of Singapore</i>), Johns Paul (<i>National University of Singapore</i>), Yuchen Li (<i>Singapore Management University</i>), Bingsheng He (<i>National University of Singapore</i>)	
• Automatic Optimization of Matrix Implementations for Distributed Machine Learning and Linear Algebra.....	1222
Shangyu Luo (<i>Rice University</i>), Dimitrije Jankov (<i>Rice University</i>), Binhang Yuan (<i>Rice University</i>), Chris Jermaine (<i>Rice University</i>)	
• Synthesizing Natural Language to Visualization (NL2VIS) Benchmarks from NL2SQL Benchmarks	1235
Yuyu Luo (<i>Tsinghua University</i>), Nan Tang (<i>Qatar Computing Research Institute, Hamad Bin Khalifa University</i>), Guoliang Li (<i>Tsinghua University</i>), Chengliang Chai (<i>Tsinghua University</i>), Wenbo Li (<i>Tsinghua University</i>), Xuedi Qin (<i>Tsinghua University</i>)	

• MB2: Decomposed Behavior Modeling for Self-Driving Database Management Systems	1248
Lin Ma (<i>Carnegie Mellon University</i>), William Zhang (<i>Carnegie Mellon University</i>), Jie Jiao (<i>Carnegie Mellon University</i>), Wuwen Wang (<i>Carnegie Mellon University</i>), Matthew Butrovich (<i>Carnegie Mellon University</i>), Wan Shen Lim (<i>Carnegie Mellon University</i>), Prashanth Menon (<i>Carnegie Mellon University</i>), Andrew Pavlo (<i>Carnegie Mellon University</i>)	
• MetaInsight: Automatic Discovery of Structured Knowledge for Exploratory Data Analysis	1262
Pingchuan Ma (<i>Hong Kong University of Science and Technology</i>), Rui Ding (<i>Microsoft Research</i>), Shi Han (<i>Microsoft Research</i>), Dongmei Zhang (<i>Microsoft Research</i>)	
• Bao: Making Learned Query Optimization Practical	1275
Ryan Marcus (<i>Massachusetts Institute of Technology & Intel Labs</i>), Parimarjan Negi (<i>Massachusetts Institute of Technology</i>), Hongzi Mao (<i>Massachusetts Institute of Technology</i>), Nesime Tatbul (<i>Massachusetts Institute of Technology & Intel Labs</i>), Mohammad Alizadeh (<i>Massachusetts Institute of Technology</i>), Tim Kraska (<i>Massachusetts Institute of Technology</i>)	
• Hybrid Edge Partitioner: Partitioning Large Power-Law Graphs under Memory Constraints	1289
Ruben Mayer (<i>Technical University of Munich</i>), Hans-Arno Jacobsen (<i>University of Toronto</i>)	
• Rotom: A Meta-Learned Data Augmentation Framework for Entity Matching, Data Cleaning, Text Classification, and Beyond	1303
Zhengjie Miao (<i>Duke University</i>), Yuliang Li (<i>Megagon Labs</i>), Xiaolan Wang (<i>Megagon Labs</i>)	
• Marrying Top-k with Skyline Queries: Relaxing the Preference Input while Producing Output of Controllable Size	1317
Kyriakos Mouratidis (<i>Singapore Management University</i>), Keming Li (<i>Southern University of Science and Technology</i>), Bo Tang (<i>Southern University of Science and Technology</i>)	
• MxTasks: How to Make Efficient Synchronization and Prefetching Easy	1331
Jan Mühlig (<i>TU Dortmund University</i>), Jens Teubner (<i>TU Dortmund University</i>)	
• Enforcing Constraints for Machine Learning Systems via Declarative Feature Selection: An Experimental Study	1345
Felix Neutatz (<i>TU Berlin</i>), Felix Biessmann (<i>Einstein Center Digital Future Berlin & Beuth University Berlin</i>), Ziawasch Abedjan (<i>Leibniz Universität Hannover & L3S Research Center</i>)	
• When the Recursive Diversity Anonymity Meets the Ring Signature	1359
Wangze Ni (<i>Hong Kong University of Science and Technology</i>), Peng Cheng (<i>East China Normal University</i>), Lei Chen (<i>Hong Kong University of Science and Technology</i>), Xuemin Lin (<i>University of New South Wales</i>)	
• Terrace: A Hierarchical Graph Container for Skewed Dynamic Graphs	1372
Prashant Pandey (<i>Lawrence Berkeley National Lab & University of California, Berkeley</i>), Brian Wheatman (<i>Johns Hopkins University</i>), Helen Xu (<i>Massachusetts Institute of Technology</i>), Aydin Buluc (<i>Lawrence Berkeley National Lab & University of California, Berkeley</i>)	
• Vector Quotient Filters: Overcoming the Time/Space Trade-Off in Filter Design	1386
Prashant Pandey (<i>Lawrence Berkeley National Lab & University of California, Berkeley</i>), Alex Conway (<i>VMware Research</i>), Joe Durie (<i>Rutgers University</i>), Michael A. Bender (<i>Stony Brook University</i>), Martin Farach-Colton (<i>Rutgers University</i>), Rob Johnson (<i>VMware Research</i>)	
• Looking for Trouble: Analyzing Classifier Behavior via Pattern Divergence	1400
Eliana Pastor (<i>Politechnico di Torino</i>), Luca de Alfaro (<i>University of California, Santa Cruz</i>), Elena Baralis (<i>Politechnico di Torino</i>)	
• MG-Join: A Scalable Join for Massively Parallel Multi-GPU Architectures	1413
Johns Paul (<i>Nanyang Technological University</i>), Shengliang Lu (<i>National University of Singapore</i>), Bingsheng He (<i>National University of Singapore</i>), Chiew Tong Lau (<i>Nanyang Technological University</i>)	
• LIMA: Fine-grained Lineage Tracing and Reuse in Machine Learning Systems	1426
Arnab Phani (<i>Graz University of Technology</i>), Benjamin Rath (<i>Graz University of Technology</i>), Matthias Boehm (<i>Graz University of Technology</i>)	
• Scalable and Usable Relational Learning With Automatic Language Bias	1440
Jose Picado (<i>Oregon State University</i>), Arash Termehchy (<i>Oregon State University</i>), Alan Fern (<i>Oregon State University</i>), Sudhanshu Pathak (<i>Oregon State University</i>), Praveen Ilango (<i>Oregon State University</i>), John Davis (<i>Oregon State University</i>)	

- **To Share, or not to Share Online Event Trend Aggregation Over Bursty Event Streams** 1452
 Olga Poppe (*Microsoft Gray Systems Lab*), Chuan Lei (*IBM Research - Almaden*),
 Lei Ma (*Worcester Polytechnic Institute*), Allison M. Rozet (*MathWorks*),
 Elke A. Rundensteiner (*Worcester Polytechnic Institute*)
- **Weighted Distinct Sampling: Cardinality Estimation for SPJ Queries** 1465
 Yuan Qiu (*Hong Kong University of Science and Technology*),
 Yilei Wang (*Hong Kong University of Science and Technology*),
 Ke Yi (*Hong Kong University of Science and Technology & SICS, Shenzhen University*),
 Feifei Li (*Alibaba Group*), Bin Wu (*Alibaba Group*), Chaoqun Zhan (*Alibaba Group*)
- **Online Topic-Aware Entity Resolution Over Incomplete Data Streams** 1478
 Weilong Ren (*Kent State University*), Xiang Lian (*Kent State University*),
 Kambiz Ghazinour (*State University of New York*)
- **Dynamic Structural Clustering on Graphs** 1491
 Boyu Ruan (*The University of Queensland*), Junhao Gan (*The University of Melbourne*),
 Hao Wu (*The University of Melbourne*), Anthony Wirth (*The University of Melbourne*)
- **Blockchains vs. Distributed Databases: Dichotomy and Fusion** 1504
 Pingcheng Ruan (*National University of Singapore*),
 Tien Tuan Anh Dinh (*Singapore University of Technology and Design*),
 Dumitrel Loghin (*National University of Singapore*), Meihui Zhang (*Beijing Institute of Technology*),
 Gang Chen (*Zhejiang University*), Qian Lin (*National University of Singapore*),
 Beng Chin Ooi (*National University of Singapore*)
- **Graphsurge: Graph Analytics on View Collections Using Differential Computation** 1518
 Siddhartha Sahu (*University of Waterloo*), Semih Salihoglu (*University of Waterloo*)
- **Correlation Sketches for Approximate Join-Correlation Queries** 1531
 Aécio Santos (*New York University*), Aline Bessa (*New York University*),
 Fernando Chirigati (*Springer Nature*), Christopher Musco (*New York University*),
 Juliana Freire (*New York University*)
- **HedgeCut: Maintaining Randomised Trees for Low-Latency Machine Unlearning** 1545
 Sebastian Schelter (*University of Amsterdam*), Stefan Grafberger (*University of Amsterdam*),
 Ted Dunning (*Hewlett Packard Enterprise*)
- **Self-adaptive Graph Traversal on GPUs** 1558
 Mo Sha (*National University of Singapore*), Yuchen Li (*Singapore Management University*),
 Kian-Lee Tan (*National University of Singapore*)
- **PCOR: Private Contextual Outlier Release via Differentially Private Search** 1571
 Masoumeh Shafieinejad (*University of Waterloo*), Florian Kerschbaum (*University of Waterloo*),
 Ihab F. Ilyas (*University of Waterloo*)
- **Towards Benchmarking Feature Type Inference for AutoML Platforms** 1584
 Vraj Shah (*University of California, San Diego*), Jonathan Lacanlale (*California State University, Northridge*),
 Premanand Kumar (*University of California, San Diego*), Kevin Yang (*University of California, San Diego*),
 Arun Kumar (*University of California, San Diego*)
- **Distributed Stream KNN Join** 1597
 Amirhesam Shahvarani (*Technische Universität München*), Hans-Arno Jacobsen (*University of Toronto*)
- **PGMJoins: Random Join Sampling with Graphical Models** 1610
 Ali Mohammadi Shanghooshabad (*University of Warwick*), Meghdad Kurmanji (*University of Warwick*),
 Qingzhi Ma (*University of Warwick*), Michael Shekelyan (*University of Warwick*),
 Mehrdad Almasi (*University of Warwick*), Peter Triantafillou (*University of Warwick*)
- **At-the-time and Back-in-time Persistent Sketches** 1623
 Benwei Shi (*University of Utah*), Zhuoyue Zhao (*University of Utah*),
 Yanqing Peng (*University of Utah*), Feifei Li (*University of Utah*), Jeff M. Phillips (*University of Utah*)
- **Clonos: Consistent Causal Recovery for Highly-Available Streaming Dataflows** 1637
 Pedro F. Silvestre (*Delft University of Technology*), Marios Frakoulis (*Delft University of Technology*),
 Diomidis Spinellis (*Delft University of Technology*), Asterios Katsifodimos (*Delft University of Technology*)
- **Scalable Multi-Query Execution using Reinforcement Learning** 1651
 Panagiotis Sioulas (*EPFL*), Anastasia Ailamaki (*EPFL & RAW Labs SA*)

• Vertex-centric Parallel Computation of SQL Queries	1664
Ainur Smagulova (<i>University of California, San Diego</i>), Alin Deutsch (<i>University of California, San Diego</i>)	
• Auto-Validate: Unsupervised Data Validation Using Data-Domain Patterns Inferred from Data Lakes	1678
Jie Song (<i>University of Michigan</i>), Yeye He (<i>Microsoft Research</i>)	
• On Saving Outliers for Better Clustering over Noisy Data	1692
Shaoxu Song (<i>Tsinghua University</i>), Fei Gao (<i>Tsinghua University</i>), Ruihong Huang (<i>Tsinghua University</i>), Yihan Wang (<i>Tsinghua University</i>)	
• Why Not Match: On Explanations of Event Pattern Queries	1705
Shaoxu Song (<i>Tsinghua University</i>), Ruihong Huang (<i>Tsinghua University</i>), Yu Gao (<i>Tsinghua University</i>), Jianmin Wang (<i>Tsinghua University</i>)	
• Tuplex: Data Science in Python at Native Code Speed	1718
Leonhard Spiegelberg (<i>Brown University</i>), Rahul Yesantharao (<i>Massachusetts Institute of Technology</i>), Malte Schwarzkopf (<i>Brown University</i>), Tim Kraska (<i>Massachusetts Institute of Technology</i>)	
• Reducing Ambiguity in Json Schema Discovery	1732
William Spoth (<i>University at Buffalo, SUNY</i>), Oliver Kennedy (<i>University at Buffalo, SUNY</i>), Ying Lu (<i>Oracle</i>), Beda Hammerschmidt (<i>Oracle</i>), Zhen Hua Liu (<i>Oracle</i>)	
• Learned Cardinality Estimation for Similarity Queries	1745
Ji Sun (<i>Tsinghua University</i>), Guoliang Li (<i>Tsinghua University</i>), Nan Tang (<i>Hamad Bin Khalifa University</i>)	
• PathEnum: Towards Real-Time Hop-Constrained s-t Path Enumeration	1758
Shixuan Sun (<i>National University of Singapore</i>), Yuhang Chen (<i>National University of Singapore</i>), Bingsheng He (<i>National University of Singapore</i>), Bryan Hooi (<i>National University of Singapore</i>)	
• Slice Tuner: A Selective Data Acquisition Framework for Accurate and Fair Machine Learning Models	1771
Ki Hyun Tae (<i>Korea Advanced Institute of Science and Technology</i>), Steven Euijong Whang (<i>Korea Advanced Institute of Science and Technology</i>)	
• On m-Impact Regions and Standing Top-k Influence Problems	1784
Bo Tang (<i>Southern University of Science and Technology</i>), Kyriakos Mouratidis (<i>Singapore Management University</i>), Mingji Han (<i>Southern University of Science and Technology</i>)	
• Resource-efficient Shared Query Execution via Exploiting Time Slackness	1797
Dixin Tang (<i>University of California, Berkeley</i>), Zechao Shang (<i>Snowflake Computing</i>), William W. Ma (<i>University of Chicago</i>), Aaron J. Elmore (<i>University of Chicago</i>), Sanjay Krishnan (<i>University of Chicago</i>)	
• CoRM: Compactable Remote Memory over RDMA	1811
Konstantin Taranov (<i>ETH Zurich</i>), Salvatore Di Girolamo (<i>ETH Zurich</i>), Torsten Hoefer (<i>ETH Zurich</i>)	
• DFI: The Data Flow Interface for High-Speed Networks	1825
Lasse Thostrup (<i>TU Darmstadt</i>), Jan Skrzypczak (<i>Zuse Institut</i>), Matthias Jasny (<i>TU Darmstadt</i>), Tobias Ziegler (<i>TU Darmstadt</i>), Carsten Binnig (<i>TU Darmstadt</i>)	
• Conditional Cuckoo Filters	1838
Daniel Ting (<i>Tableau Software</i>), Rick Cole (<i>Tableau Software</i>)	
• Parallel Index-Based Structural Graph Clustering and Its Approximation	1851
Tom Tseng (<i>Massachusetts Institute of Technology</i>), Laxman Dhulipala (<i>Massachusetts Institute of Technology</i>), Julian Shun (<i>Massachusetts Institute of Technology</i>)	
• Expand your Training Limits! Generating Training Data for ML-based Data Management	1865
Francesco Ventura (<i>PoliTecnicco di Torino</i>), Zoi Kaoudi (<i>TU Berlin & DFKI GmbH</i>), Jorge Arnulfo Quiané-Ruiz (<i>TU Berlin & DFKI GmbH</i>), Volker Markl (<i>TU Berlin & DFKI GmbH</i>)	
• Self-Tuning Query Scheduling for Analytical Workloads	1879
Benjamin Wagner (<i>Technische Universität München</i>), André Kohn (<i>Technische Universität München</i>), Thomas Neumann (<i>Technische Universität München</i>)	

- **DP-Sync: Hiding Update Patterns in Secure Outsourced Databases with Differential Privacy** 1892
Chenghong Wang (*Duke University*), Johnes Bater (*Duke University*),
Kartik Nayak (*Duke University*), Ashwin Machanavajjhala (*Duke University*)
- **Public Transport Planning: When Transit Network Connectivity Meets Commuting Demand** 1906
Sheng Wang (*New York University*), Yuan Sun (*Monash University*),
Christopher Musco (*New York University*), Zhifeng Bao (*RMIT University*)
- **Interactive Search for One of the Top-k** 1920
Weicheng Wang (*Hong Kong University of Science and Technology*),
Raymond Chi-Wing Wong (*Hong Kong University of Science and Technology*),
Min Xie (*Shenzhen Institute of Computing Sciences*)
- **Towards Enhancing Database Education: Natural Language Generation Meets Query Execution Plans** 1933
Weiguo Wang (*Nanyang Technological University & Xidian University*),
Sourav S. Bhowmick (*Nanyang Technological University*), Hui Li (*Xidian University*),
Shafiq Joty (*Nanyang Technological University*), Siyuan Liu (*Nanyang Technological University*),
Peng Chen (*Xidian University*)
- **Query-by-Sketch: Scaling Shortest Path Graph Queries on Very Large Networks** 1946
Ye Wang (*Australian National University*), Qing Wang (*Australian National University*),
Henning Koehler (*Massey University*), Yu Lin (*Australian National University*)
- **Secure Yannakakis: Join-Aggregate Queries over Private Data** 1969
Yilei Wang (*Hong Kong University of Science and Technology*),
Ke Yi (*Hong Kong University of Science and Technology*)
- **Fast Parallel Algorithms for Euclidean Minimum Spanning Tree and Hierarchical Spatial Clustering** 1982
Yiqiu Wang (*Massachusetts Institute of Technology*), Shangdi Yu (*Massachusetts Institute of Technology*),
Yan Gu (*University of California, Riverside*), Julian Shun (*Massachusetts Institute of Technology*)
- **Unifying the Global and Local Approaches: An Efficient Power Iteration with Forward Push** 1996
Hao Wu (*The University of Melbourne*), Junhao Gan (*The University of Melbourne*),
Zhewei Wei (*Renmin University of China*), Rui Zhang (*Tsinghua University*)
- **A Unified Deep Model of Learning from both Data and Queries for Cardinality Estimation** 2009
Peizhi Wu (*Singtel Cognitive and Artificial Intelligence Lab for Enterprises@NTU*),
Gao Cong (*Nanyang Technological University*)
- **Spatial Independent Range Sampling** 2023
Dong Xie (*The Pennsylvania State University & University of Utah*), Jeff M. Phillips (*University of Utah*),
Michael Matheny (*Amazon & University of Utah*), Feifei Li (*Alibaba Group & University of Utah*)
- **DIV: Resolving the Dynamic Issues of Zero-knowledge Set Membership Proof in the Blockchain** 2036
Zihuan XU (*The Hong Kong University of Science and Technology*),
Lei Chen (*The Hong Kong University of Science and Technology*)
- **HUGE: An Efficient and Scalable Subgraph Enumeration System** 2049
Zhengyi Yang (*The University of New South Wales*), Longbin Lai (*Alibaba Group*),
Xuemin Lin (*The University of New South Wales*), Kongzhang Hao (*The University of New South Wales*),
Wenjie Zhang (*The University of New South Wales*),
- **Multiple Dynamic Outlier-Detection from a Data Stream by Exploiting Duality of Data and Queries** 2063
Susik Yoon (*Korea Advanced Institute of Science and Technology*),
Yooju Shin (*Korea Advanced Institute of Science and Technology*),
Jae-Gil Lee (*Korea Advanced Institute of Science and Technology*), Byung Suk Lee (*University of Vermont*)
- **OmniFair: A Declarative System for Model-Agnostic Group Fairness in Machine Learning** 2076
Hantian Zhang (*Georgia Institute of Technology*), Xu Chu (*Georgia Institute of Technology*),
Abolfazl Asudeh (*University of Illinois at Chicago*), Shamkant B. Navathe (*Georgia Institute of Technology*)

- **Parallelizing Intra-Window Join on Multicores: An Experimental Study** 2089
Shuhao Zhang (*Singapore University of Technology and Design*), Yancan Mao (*National University of Singapore*), Jiong He (*Bytedance*), Philipp M. Grulich (*Technische Universität Berlin*), Steffen Zeuch (*German Research Centre for Artificial Intelligence*), Bingsheng He (*National University of Singapore*), Richard T. B. Ma (*National University of Singapore*), Volker Markl (*Technische Universität Berlin & German Research Centre for Artificial Intelligence*)
- **ResTune: Resource Oriented Tuning Boosted by Meta-Learning for Cloud Databases** 2102
Xinyi Zhang (*Peking University & Alibaba Group*), Hong Wu (*Alibaba Group*), Zhuo Chang (*Alibaba Group & Peking University*), Shuowei Jin (*Alibaba Group*), Jian Tan (*Alibaba Group*), Feifei Li (*Alibaba Group*), Tieying Zhang (*Alibaba Group*), Bin Cui (*Peking University*)
- **Minimizing the Regret of an Influence Provider** 2115
Yipeng Zhang (*RMIT University*), Yuchen Li (*Singapore Management University*), Zhifeng Bao (*RMIT University*), Baihua Zheng (*Singapore Management University*), H. V. Jagadish (*University of Michigan*)
- **EIREs: Efficient Integration of Remote Data in Event Stream Processing** 2128
Bo Zhao (*Humboldt-Universität zu Berlin*), Han van der Aa (*Universität Mannheim*), Thanh Tam Nguyen (*Leibniz Universität Hannover*), Quoc Viet Hung Nguyen (*Griffith University*), Matthias Weidlich (*Humboldt-Universität zu Berlin*)
- **A Learned Sketch for Subgraph Counting** 2142
Kangfei Zhao (*The Chinese University of Hong Kong*), Jeffrey Xu Yu (*Chinese University of Hong Kong*), Hao Zhang (*Chinese University of Hong Kong*), Qiyan Li (*Wuhan University*), Yu Rong (*Tencent AI Lab*)
- **PACE: Learning Effective Task Decomposition for Human-in-the-loop Healthcare Delivery** 2156
Kaiping Zheng (*National University of Singapore*), Gang Chen (*Zhejiang University*), Melanie Herschel (*Universität Stuttgart*), Kee Yuan Ngiam (*National University Health System*), Beng Chin Ooi (*National University of Singapore*), Jinyang Gao (*Alibaba Group*)
- **SIA: Optimizing Queries using Learned Predicates** 2169
Qi Zhou (*Georgia Institute of Technology*), Joy Arulraj (*Georgia Institute of Technology*), Shamkant Navathe (*Georgia Institute of Technology*), William Harris (*Galois Inc*), Jinpeng Wu (*Alibaba Group*)
- **VeriDB: An SGX-based Verifiable Database** 2182
Wenchao Zhou (*Georgetown University & Alibaba Group*), Yifan Cai (*University of Pennsylvania & Alibaba Group*), Yanqing Peng (*University of Utah & Alibaba Group*), Sheng Wang (*Alibaba Group*), Ke Ma (*Shanghai Jiao Tong University*), Feifei Li (*Alibaba Group*)
- **Spitfire: A Three-Tier Buffer Manager for Volatile and Non-Volatile Memory** 2195
Xinjing Zhou (*Tencent Inc.*), Joy Arulraj (*Georgia Institute of Technology*), Andrew Pavlo (*Carnegie Mellon University*), David Cohen (*Intel*)

Research Data Science & Engineering Track Papers

- **AlphaEvolve: A Learning Framework to Discover Novel Alphas in Quantitative Investment** 2208
Can Cui (*National University of Singapore*), Wei Wang (*National University of Singapore*), Meihui Zhang (*Beijing Institute of Technology*), Gang Chen (*Zhejiang University*), Zhaojing Luo (*National University of Singapore*), Beng Chin Ooi (*National University of Singapore*)
- **Adaptive Rule Discovery for Labeling Text Data** 2217
Sainyam Galhotra (*University of Massachusetts, Amherst*), Behzad Golshan (*Megagon Labs*), Wang-Chiew Tan (*Facebook AI*)
- **Fast Processing and Querying of 170TB of Genomics Data via a Repeated And Merged BloOm Filter (RAMBO)** 2226
Gaurav Gupta (*Rice University*), Minghao Yan (*Rice University*), Benjamin Coleman (*Rice University*), Bryce Kille (*Rice University*), R. A. Leo Elworth (*Rice University*), Tharun Medini (*Rice University*), Todd Treangen (*Rice University*), Anshumali Shrivastava (*Rice University*)

- **Shahin: Faster Algorithms for Generating Explanations for Multiple Predictions** 2235
Sona Hasani (*Google*), Saravanan Thirumuruganathan (*QCRI, Hamad Bin Khalifa University*),
Nick Koudas (*University of Toronto*), Gautam Das (*University of Texas Arlington*)
- **Pool of Experts: Realtime Querying Specialized Knowledge in Massive Neural Networks** 2244
Hakbin Kim (*Inha University*), Dong-Wan Choi (*Inha University*)
- **Joint Open Knowledge Base Canonicalization and Linking** 2253
Yinan Liu (*Nankai University*), Wei Shen (*Nankai University*), Yuanfei Wang (*Nankai University*),
Jianyong Wang (*Tsinghua University & Jiangsu Normal University*), Zhenglu Yang (*Nankai University*),
Xiaojie Yuan (*Nankai University*)
- **Heterogeneity-Aware Distributed Machine Learning Training via Partial Reduce** 2262
Xupeng Miao (*Peking University & Tencent Inc.*), Xiaonan Nie (*Peking University*),
Yingxia Shao (*Beijing University of Posts and Telecommunications*), Zhi Yang (*Peking University*),
Jiawei Jiang (*ETH Zürich*), Lingxiao Ma (*Peking University*), Bin Cui (*Peking University*)
- **DataPrep.EDA: Task-Centric Exploratory Data Analysis for Statistical Modeling in Python** 2271
Jinglin Peng (*Simon Fraser University*), Weiyuan Wu (*Simon Fraser University*),
Brandon Lockhart (*Simon Fraser University*), Song Bian (*The Chinese University of Hong Kong*),
Jing Nathan Yan (*Cornell University*), Linghao Xu (*Simon Fraser University*),
Zhixuan Chi (*Simon Fraser University*), Jeffrey M. Rzeszotarski (*Cornell University*),
Jiannan Wang (*Simon Fraser University*)
- **LightNE: A Lightweight Graph Processing System for Network Embedding** 2281
Jiezong Qiu (*Tsinghua University*), Laxman Dhulipala (*Massachusetts Institute of Technology*),
Jie Tang (*Tsinghua University & Microsoft Research*),
Richard Peng (*Georgia Institute of Technology & Microsoft Research*), Chi Wang (*Microsoft Research*)
- **SliceLine: Fast, Linear-Algebra-based Slice Finding for ML Model Debugging** 2290
Svetlana Sagadeeva (*Graz University of Technology*), Matthias Boehm (*Graz University of Technology*)
- **To Intervene or Not To Intervene: Cost based Intervention for Combating Fake News** 2300
Saravanan Thirumuruganathan (*Hamad Bin Khalifa University*),
Michael Simpson (*University of British Columbia*), Laks V.S. Lakshmanan (*University of British Columbia*)
- **Medical Entity Disambiguation Using Graph Neural Networks** 2310
Alina Vretinaris (*IBM Germany & IBM Research*), Chuan Lei (*IBM Research - Almaden*),
Vasilis Eftymiou (*FORTH-ICS & IBM Research*), Xiao Qin (*IBM Research - Almaden*),
Fatma Özcan (*Google & IBM Research*)
- **Consistent and Flexible Selectivity Estimation for High-Dimensional Data** 2319
Yaoshu Wang (*Shenzhen University*), Chuan Xiao (*Osaka University & Nagoya University*),
Jianbin Qin (*Shenzhen University*), Rui Mao (*Shenzhen University*),
Makoto Onizuka (*Osaka University*), Wei Wang (*Dongguan Univ. of Technology & University of New South Wales*),
Rui Zhang (*www.ruihang.info*), Yoshiharu Ishikawa (*Nagoya University*)
- **RobustPeriod: Robust Time-Frequency Mining for Multiple Periodicity Detection** 2328
Qingsong Wen (*DAMO Academy & Alibaba Group*), Kai He (*DAMO Academy & Alibaba Group*),
Liang Sun (*DAMO Academy & Alibaba Group*), Yingying Zhang (*Alibaba Group*),
Min Ke (*Alibaba Group*), Huan Xu (*Alibaba Group*)
- **EquiTensors: Learning Fair Integrations of Heterogeneous Urban Data** 2338
An Yan (*University of Washington*), Bill Howe (*University of Washington*)
- **GPU-Accelerated Graph Label Propagation for Real-Time Fraud Detection** 2348
Chang Ye (*Singapore Management University*), Yuchen Li (*Singapore Management University*),
Bingsheng He (*National University of Singapore*), Zhao Li (*Alibaba Group*), Jianling Sun (*Zhejiang University*)
- **Efficient Graph Summarization using Weighted LSH at Billion-Scale** 2357
Quinton Yong (*University of Victoria*), Mahdi Hajibabi (*University of Victoria*),
Venkatesh Srinivasan (*University of Victoria*), Alex Thomo (*University of Victoria*)
- **ALG: Fast and Accurate Active Learning Framework for Graph Convolutional Networks** ... 2366
Wentao Zhang (*Peking University & Tencent Inc.*), Yu Shen (*Peking University*),
Yang Li (*Peking University*), Lei Chen (*Hong Kong University of Science and Technology*),
Zhi Yang (*Peking University*), Bin Cui (*Peking University*)

- **BurstSketch: Finding Bursts in Data Streams** 2375
 Zheng Zhong (*Peking University*), Shen Yan (*Peking University*), Zikun Li (*Peking University*),
 Decheng Tan (*Peking University*), Tong Yang (*Peking University & Pengcheng Laboratory*),
 Bin Cui (*Peking University*)

Research Applications Track Papers

- **An Ecosystem of Applications for Modeling Political Violence** 2384
 Aline Bessa (*New York University*), Sonia Castelo (*New York University*),
 Rémi Rampin (*New York University*), Aélio Santos (*New York University*),
 Mike Shoemate (*Harvard University*), Vito D’Orazio (*University of Texas at Dallas*),
 Juliana Freire (*New York University*)
- **P²B-Trace: Privacy-Preserving Blockchain-based Contact Tracing to Combat Pandemics** 2389
 Zhe Peng (*Hong Kong Baptist University*), Cheng Xu (*Hong Kong Baptist University*),
 Haixin Wang (*Hong Kong Baptist University*), Jinbin Huang (*Hong Kong Baptist University*),
 Jianliang Xu (*Hong Kong Baptist University*), Xiaowen Chu (*Hong Kong Baptist University*)
- **De-anonymization Attacks on Neuroimaging Datasets** 2394
 Vikram Ravindra (*Purdue University*), Ananth Grama (*Purdue University*)
- **Convergence of Array DBMS and Cellular Automata: A Road Traffic Simulation Case** 2399
 Ramon Antonio Rodriges Zalipynis (*HSE University*)
- **Agile and Accurate CTR Prediction Model Training for Massive-Scale Online Advertising Systems** 2404
 Zhiqiang Xu (*Baidu Research*), Dong Li (*Baidu Search Ads (Phoenix Nest)*), Weijie Zhao (*Baidu Research*),
 Xing Shen (*Baidu Search Ads (Phoenix Nest)*), Tianbo Huang (*Baidu Search Ads (Phoenix Nest)*),
 Xiaoyun Li (*Baidu Research*), Ping Li (*Baidu Research*)

Industrial Track Papers

- **Bringing Cloud-Native Storage to SAP IQ** 2410
 Mohammed Abouzour (*SAP*), Güneş Aluç (*SAP*), Ivan T. Bowman (*SAP*), Xi Deng (*SAP*),
 Nandan Marathe (*SAP*), Sagar Ranadive (*SAP*), Muhammed Sharique (*SAP*), John C. Smirnios (*SAP*)
- **PG-KEYS: Keys for Property Graphs** 2423
 Renzo Angles (*Universidad de Talca, IMFD Chile*), Angela Bonifati (*Lyon 1 University, Liris CNRS & INRIA*),
 Stefania Dumbrava (*ENSIIE & Institute Polytechnique de Paris*),
 George Fletcher (*Eindhoven University of Technology*), Keith W. Hare (*JCC Consulting Inc. & Neo4j*),
 Jan Hidders (*Birkbeck, University of London*), Victor E. Lee (*TigerGraph*), Bei Li (*Google LLC*),
 Leonid Libkin (*University of Edinburgh, ENS-Paris/PSL, & Neo4j*),
 Wim Martens (*University of Bayreuth*), Filip Murlak (*University of Warsaw*),
 Josh Perryman (*Interos Inc.*), Ognjen Savković (*Free University of Bozen-Bolzano*),
 Michael Schmidt (*Amazon Web Services*), Juan Sequeda (*data.world*),
 Sławek Staworko (*U. Lille, INRIA LINKS, CRISTAL CNRS*), Dominik Tomaszuk (*University of Białystok*)
- **SQL Ledger: Cryptographically Verifiable Data in Azure SQL Database** 2437
 Panagiotis Antonopoulos (*Microsoft*), Raghad Kaushik (*Microsoft*), Hanuma Kodavalla (*Microsoft*),
 Sergio Rosales Aceves (*Microsoft*), Reilly Wong (*Microsoft*), Jason Anderson (*Microsoft*),
 Jakub Szymaszek (*Microsoft*)
- **ExDRa: Exploratory Data Science on Federated Raw Data** 2450
 Sebastian Baunsgaard (*Graz University of Technology*), Matthias Boehm (*Graz University of Technology*),
 Ankit Chaudhary (*Technische Universität Berlin*), Behrouz Derakhshan (*DFKI GmbH*),
 Stefan Geißelsöder (*Siemens AG*), Philipp M. Grulich (*Technische Universität Berlin*),
 Michael Hildebrand (*Siemens AG*), Kevin Innerebner (*Graz University of Technology*),
 Volker Markl (*DFKI GmbH & Technische Universität Berlin*), Claus Neubauer (*Siemens AG*),
 Sarah Osterburg (*Siemens AG*), Olga Ovcharenko (*Graz University of Technology*),
 Sergey Redyuk (*Technische Universität Berlin*), Tobias Rieger (*Graz University of Technology*),
 Alireza Rezaei Mahdiraji (*DFKI GmbH*), Sebastian Benjamin Wrede (*Graz University of Technology*),
 Steffen Zeuch (*DFKI GmbH*)

• LogStore: A Cloud-Native and Multi-Tenant Log Database	2464
Wei Cao (<i>Zhejiang University & Alibaba Group</i>), Xiaojie Feng (<i>Alibaba Group</i>), Boyuan Liang (<i>Alibaba Group</i>), Tianyu Zhang (<i>Alibaba Group</i>), Yusong Gao (<i>Alibaba Group</i>), Yunyang Zhang (<i>Alibaba Group</i>), Feifei Li (<i>Alibaba Group</i>)	
• PolarDB Serverless: A Cloud Native Database for Disaggregated Data Centers	2477
Wei Cao (<i>Zhejiang University & Alibaba Group</i>), Yingqiang Zhang (<i>Alibaba Group</i>), Xinjun Yang (<i>Alibaba Group</i>), Feifei Li (<i>Alibaba Group</i>), Sheng Wang (<i>Alibaba Group</i>), Qingda Hu (<i>Alibaba Group</i>), Xuntao Cheng (<i>Alibaba Group</i>), Zongzhi Chen (<i>Alibaba Group</i>), Zhenjun Liu (<i>Alibaba Group</i>), Jing Fang (<i>Alibaba Group</i>), Bo Wang (<i>Alibaba Group</i>), Yuhui Wang (<i>Alibaba Group</i>), Haiqing Sun (<i>Alibaba Group</i>), Ze Yang (<i>Alibaba Group</i>), Zhushu Cheng (<i>Alibaba Group</i>), Sen Chen (<i>Alibaba Group</i>), Jian Wu (<i>Alibaba Group</i>), Wei Hu (<i>Alibaba Group</i>), Jianwei Zhao (<i>Alibaba Group</i>), Yusong Gao (<i>Alibaba Cloud</i>), Songlu Cai (<i>Alibaba Group</i>), Yunyang Zhang (<i>Alibaba Group</i>), Jiawang Tong (<i>Alibaba Group</i>)	
• Citus: Distributed PostgreSQL for Data-Intensive Applications	2490
Umur Cubukcu (<i>Microsoft Corporation</i>), Ozgun Erdogan (<i>Microsoft Corporation</i>), Sumedh Pathak (<i>Microsoft Corporation</i>), Sudhakar Sannakkayala (<i>Microsoft Corporation</i>), Marco Slot (<i>Microsoft Corporation</i>)	
• Real-time Data Infrastructure at Uber	2503
Yupeng Fu (<i>Uber, Inc.</i>), Chinmay Soman (<i>Uber, Inc.</i>)	
• QuiCK: A Queuing System in CloudKit	2517
Kfir Lev-Ari (<i>Apple Inc</i>), Yizuo Tian (<i>Apple Inc</i>), Alexander Shraer (<i>Apple Inc</i>), Chris Douglas (<i>Apple Inc</i>), Hao Fu (<i>Apple Inc</i>), Andrey Andreev (<i>Apple Inc</i>), Kevin Beranek (<i>Apple Inc</i>), Scott Dugas (<i>Apple Inc</i>), Alec Grieser (<i>Apple Inc</i>), Jeremy Hemmo (<i>Apple Inc</i>)	
• Greenplum: A Hybrid Database for Transactional and Analytical Workloads	2530
Zhenghua Lyu (<i>VMware</i>), Huan Hubert Zhang (<i>VMware</i>), Gang Xiong (<i>VMware</i>), Gang Guo (<i>VMware</i>), Haozhou Wang (<i>VMware</i>), Jinbao Chen (<i>VMware</i>), Asim Praveen (<i>VMware</i>), Yu Yang (<i>VMware</i>), Xiaoming Gao (<i>VMware</i>), Alexandra Wang (<i>VMware</i>), Wen Lin (<i>VMware</i>), Ashwin Agrawal (<i>VMware</i>), Junfeng Yang (<i>VMware</i>), Hao Wu (<i>VMware</i>), Xiaoliang Li (<i>VMware</i>), Feng Guo (<i>VMware</i>), Jiang Wu (<i>VMware</i>), Jesse Zhang (<i>VMware</i>), Venkatesh Raghavan (<i>VMware</i>)	
• Toto – Benchmarking the Efficiency of a Cloud Service	2543
Justin Moeller (<i>Microsoft</i>), Zi Ye (<i>Microsoft</i>), Katherine Lin (<i>Microsoft</i>), Willis Lang (<i>Microsoft</i>)	
• Steering Query Optimizers: A Practical Take on Big Data Workloads	2557
Parimarjan Negi (<i>Massachusetts Institute of Technology</i>), Matteo Interlandi (<i>Microsoft</i>), Ryan Marcus (<i>Massachusetts Institute of Technology & Intel Labs</i>), Tim Kraska (<i>Massachusetts Institute of Technology</i>), Mohammad Alizadeh (<i>Massachusetts Institute of Technology</i>), Marc Friedman (<i>Microsoft</i>), Alekh Jindal (<i>Microsoft</i>)	
• ArkDB: A Key-Value Engine for Scalable Cloud Storage Services	2570
Zhu Pang (<i>Alibaba Group</i>), Qingda Lu (<i>Alibaba Group</i>), Shuo Chen (<i>Alibaba Group</i>), Rui Wang (<i>Alibaba Group</i>), Yikang Xu (<i>Alibaba Group</i>), Jiesheng Wu (<i>Alibaba Group</i>)	
• AutoAI-TS: AutoAI for Time Series Forecasting	2584
Syed Yousaf Shah (<i>IBM Thomas J. Watson Research Center</i>), Dhaval Patel (<i>IBM Thomas J. Watson Research Center</i>), Long Vu (<i>IBM Thomas J. Watson Research Center</i>), Xuan-Hong Dang (<i>IBM Thomas J. Watson Research Center</i>), Bei Chen (<i>IBM Research</i>), Peter Kirchner (<i>IBM Thomas J. Watson Research Center</i>), Horst Samulowitz (<i>IBM Thomas J. Watson Research Center</i>), David Wood (<i>IBM Thomas J. Watson Research Center</i>), Gregory Bramble (<i>IBM Thomas J. Watson Research Center</i>), Wesley M. Gifford (<i>IBM Thomas J. Watson Research Center</i>), Giridhar Ganapavarapu (<i>IBM Thomas J. Watson Research Center</i>), Roman Vaculin (<i>IBM Thomas J. Watson Research Center</i>), Petros Zerfos (<i>IBM Thomas J. Watson Research Center</i>)	
• Learning-Aided Heuristics Design for Storage System	2597
Yingtian Tang (<i>Huawei Noah's Ark Lab & University of Pennsylvania</i>), Han Lu (<i>Huawei Noah's Ark Lab & Shanghai Jiao Tong University</i>), Xijun Li (<i>Huawei Noah's Ark Lab & University of Science and Technology of China</i>), Lei Chen (<i>Huawei Noah's Ark Lab</i>), Mingxuan Yuan (<i>Huawei Noah's Ark Lab</i>), Jia Zeng (<i>Huawei Noah's Ark Lab</i>)	

• Consistency and Completeness: Rethinking Distributed Stream Processing in Apache Kafka	2602
Guozhang Wang (<i>Confluent Inc.</i>), Lei Chen (<i>Bloomberg L.P.</i>), Ayushman Dikshit (<i>Expedia Group</i>), Jason Gustafson (<i>Confluent Inc.</i>), Boyang Chen (<i>Confluent Inc.</i>), Matthias J. Sax (<i>Confluent Inc.</i>), John Roesler (<i>Confluent Inc.</i>), Sophie Blee-Goldman (<i>Confluent Inc.</i>), Bruno Cadonna (<i>Confluent Inc.</i>), Apurva Mehta (<i>Confluent Inc.</i>), Varun Madan (<i>Confluent Inc.</i>), Jun Rao (<i>Confluent Inc.</i>)	
• Milvus: A Purpose-Built Vector Data Management System	2614
Jianguo Wang (<i>Zilliz & Purdue University</i>), Xiaomeng Yi (<i>Zilliz</i>), Rentong Guo (<i>Zilliz</i>), Hai Jin (<i>Zilliz</i>), Peng Xu (<i>Zilliz</i>), Shengjun Li (<i>Zilliz</i>), Xiangyu Wang (<i>Zilliz</i>), Xiangzhou Guo (<i>Zilliz</i>), Chengming Li (<i>Zilliz</i>), Xiaohai Xu (<i>Zilliz</i>), Kun Yu (<i>Zilliz</i>), Yuxing Yuan (<i>Zilliz</i>), Yinghao Zou (<i>Zilliz</i>), Jiquan Long (<i>Zilliz</i>), Yudong Cai (<i>Zilliz</i>), Zhenxiang Li (<i>Zilliz</i>), Zhifeng Zhang (<i>Zilliz</i>), Yihua Mo (<i>Zilliz</i>), Jun Gu (<i>Zilliz</i>), Ruiyi Jiang (<i>Zilliz</i>), Yi Wei (<i>Zilliz</i>), Charles Xie (<i>Zilliz</i>)	
• APAN: Asynchronous Propagation Attention Network for Real-time Temporal Graph Embedding	2628
Xuhong Wang (<i>Shanghai Jiao Tong University & Ant Group</i>), Ding Lyu (<i>Shanghai Jiao Tong University</i>), Mengjian Li (<i>Ant Group</i>), Yang Xia (<i>Ant Group</i>), Qi Yang (<i>Ant Group</i>), Xinwen Wang (<i>Ant Group</i>), Xinguang Wang (<i>Ant Group</i>), Ping Cui (<i>Shanghai Jiao Tong University</i>), Yupu Yang (<i>Shanghai Jiao Tong University</i>), Bowen Sun (<i>Ant Group</i>), Zhenyu Guo (<i>Ant Group</i>)	
• Production Machine Learning Pipelines: Empirical Analysis and Optimization Opportunities	2639
Doris Xin (<i>University of California, Berkeley</i>), Hui Miao (<i>Google</i>), Aditya Parameswaran (<i>University of California, Berkeley</i>), Neoklis Polyzotis (<i>Google</i>)	
• FoundationDB: A Distributed Unbundled Transactional Key Value Store	2653
Jingyu Zhou (<i>Apple Inc.</i>), Meng Xu (<i>Apple Inc.</i>), Alexander Shraer (<i>Apple Inc.</i>), Bala Namasivayam (<i>Apple Inc.</i>), Alex Miller (<i>Snowflake Inc.</i>), Evan Tschannen (<i>Snowflake Inc.</i>), Steve Atherton (<i>Snowflake Inc.</i>), Andrew J. Beamon (<i>Snowflake Inc.</i>), Rusty Sears (<i>Apple Inc.</i>), John Leach (<i>Apple Inc.</i>), Dave Rosenthal (<i>Apple Inc.</i>), Xin Dong (<i>Apple Inc.</i>), Will Wilson (<i>antithesis.com</i>), Ben Collins (<i>antithesis.com</i>), David Scherer (<i>antithesis.com</i>), Alec Grieser (<i>Apple Inc.</i>), Young Liu (<i>Apple Inc.</i>), Alvin Moore (<i>Apple Inc.</i>), Bhaskar Muppana (<i>Apple Inc.</i>), Xiaoge Su (<i>Apple Inc.</i>), Vishesh Yadav (<i>Apple Inc.</i>)	
• KEA: Tuning an Exabyte-Scale Data Infrastructure	2667
Yiwen Zhu (<i>Microsoft</i>), Subru Krishnan (<i>Microsoft</i>), Konstantinos Karanasos (<i>Microsoft</i>), Isha Tarte (<i>Microsoft</i>), Conor Power (<i>Microsoft</i>), Abhishek Modi (<i>Microsoft</i>), Manoj Kumar (<i>Microsoft</i>), Deli Zhang (<i>Microsoft</i>), Kartheek Muthyalu (<i>Microsoft</i>), Nick Jurgens (<i>Microsoft</i>), Sarvesh Sakalanaga (<i>Salesforce</i>), Sudhir Darbha (<i>Microsoft</i>), Minu Iyer (<i>Microsoft</i>), Ankita Agarwal (<i>Microsoft</i>), Carlo Curino (<i>Microsoft</i>)	

Demo Track Papers

• DataMangler: A Novel Approach to Data Virtualization	2681
Damianos Chatziantoniou (<i>Athens University of Economics and Business</i>), Verena Kantere (<i>National Technical University of Athens</i>)	
• TSExplain: Surfacing Evolving Explanations for Time Series	2686
Yiru Chen (<i>Columbia University</i>), Silu Huang (<i>Microsoft Research</i>)	
• SRA: Smart Recovery Advisor for Cyber Attacks	2691
Ka-Ho Chow (<i>Georgia Institute of Technology</i>), Umesh Deshpande (<i>IBM Research - Almaden</i>), Sangeetha Seshadri (<i>IBM Research - Almaden</i>), Ling Liu (<i>Georgia Institute of Technology</i>)	
• Transforming ML Predictive Pipelines into SQL with MASQ.....	2696
Francesco Del Buono (<i>Università di Modena e Reggio Emilia</i>), Matteo Paganelli (<i>Università di Modena e Reggio Emilia</i>), Paolo Sottovia (<i>Huawei</i>), Matteo Interlandi (<i>Microsoft</i>), Francesco Guerra (<i>Università di Modena e Reggio Emilia</i>)	
• CAvSAT: Answering Aggregation Queries over Inconsistent Databases via SAT Solving.....	2701
Akhil A. Dixit (<i>University of California, Santa Cruz</i>), Phokion G. Kolaitis (<i>University of California, Santa Cruz & IBM Research</i>)	

• CoCo: Interactive Exploration of Conformance Constraints for Data Understanding and Data Cleaning	2706
	Anna Fariha (<i>University of Massachusetts, Amherst</i>), Ashish Tiwari (<i>Microsoft</i>), Alexandra Meliou (<i>University of Massachusetts, Amherst</i>), Arjun Radhakrishna (<i>Microsoft</i>), Sumit Gulwani (<i>Microsoft</i>)
• BEER: Blocking for Effective Entity Resolution	2711
	Sainyam Galhotra (<i>University of Massachusetts, Amherst</i>), Donatella Firmani (<i>Roma Tre University</i>), Barna Saha (<i>University of California, Berkeley</i>), Divesh Srivastava (<i>AT&T Chief Data Office</i>)
• A System for Automated Open-Source Threat Intelligence Gathering and Management	2716
	Peng Gao (<i>University of California, Berkeley</i>), Xiaoyuan Liu (<i>University of California, Berkeley</i>), Edward Choi (<i>University of California, Berkeley</i>), Bhavna Soman (<i>Microsoft Corporation</i>), Chinmaya Mishra (<i>Microsoft Corporation</i>), Kate Farris (<i>Microsoft Corporation</i>), Dawn Song (<i>University of California, Berkeley</i>)
• Attaining Workload Scalability and Strong Consistency for Replicated Databases with Hihooi	2721
	Michael A. Georgiou (<i>Cyprus University of Technology</i>), Michael Panayiotou (<i>Cyprus University of Technology</i>), Lambros Odysseos (<i>Cyprus University of Technology</i>), Aristodemos Paphitis (<i>Cyprus University of Technology</i>), Michael Sirivianos (<i>Cyprus University of Technology</i>), Herodotos Herodotou (<i>Cyprus University of Technology</i>)
• Dendrite: Bolt-on Adaptivity for Data Systems	2726
	Brad Glasbergen (<i>University of Waterloo</i>), Fangyu Wu (<i>University of Waterloo</i>), Khuzaima Daudjee (<i>University of Waterloo</i>)
• PyExplore: Query Recommendations for Data Exploration without Query Logs	2731
	Apostolos Glenis (<i>Athena Research Center</i>), Georgia Koutrika (<i>Athena Research Center</i>)
• MLINSPECT: A Data Distribution Debugger for Machine Learning Pipelines	2736
	Stefan Grafberger (<i>University of Amsterdam</i>), Shubha Guha (<i>University of Amsterdam</i>), Julia Stoyanovich (<i>New York University</i>), Sebastian Schelter (<i>University of Amsterdam</i>)
• QuTE: Answering Quantity Queries from Web Tables	2740
	Vinh Thinh Ho (<i>Max Planck Institute for Informatics</i>), Konnika Pal (<i>Max Planck Institute for Informatics</i>), Gerhard Weikum (<i>Max Planck Institute for Informatics</i>)
• INCA: Inconsistency-Aware Data Profiling and Querying	2745
	Ousmane Issa (<i>University Clermont Auvergne</i>), Angela Bonifati (<i>Université Lyon 1</i>), Farouk Toumani (<i>University Clermont Auvergne</i>)
• Boomerang: Proactive Insight-Based Recommendations for Guiding Conversational Data Analysis	2750
	Doris Jung-Lin Lee (<i>University of California, Berkeley & IBM Research</i>), Abdul Quamar (<i>IBM Research AI</i>), Eser Kandogan (<i>Megagon Labs & IBM Research</i>), Fatma Özcan (<i>Google & IBM Research</i>)
• IndoorViz: A Demonstration System for Indoor Spatial Data Management	2755
	Yue Li (<i>East China Normal University</i>), Shiyu Yang (<i>Guangzhou University</i>), Muhammad Aamir Cheema (<i>Monash University</i>), Zhou Shao (<i>Monash University</i>), Xuemin Lin (<i>The University of New South Wales</i>)
• RawVis: A System for Efficient In-situ Visual Analytics	2760
	Stavros Maroulis (<i>National Technical University of Athens & ATHENA Research Center</i>), Nikos Bikakis (<i>ATHENA Research Center</i>), George Papastefanatos (<i>ATHENA Research Center</i>), Panos Vassiliadis (<i>University of Ioannina</i>), Yannis Vassiliou (<i>National Technical University of Athens</i>)
• Crosstown Foundry: A Scalable Data-driven Journalism Platform for Hyper-local News	2765
	Luciano Nocera (<i>University of Southern California</i>), George Constantinou (<i>University of Southern California</i>), Luan V. Tran (<i>University of Southern California</i>), Seon Ho Kim (<i>University of Southern California</i>), Gabriel Kahn (<i>University of Southern California</i>), Cyrus Shahabi (<i>University of Southern California</i>)
• A Byzantine Fault Tolerant Storage for Permissioned Blockchain	2770
	Xiaodong Qi (<i>East China Normal University</i>), Zhihao Chen (<i>East China Normal University</i>), Zhao Zhang (<i>East China Normal University</i>), Cheqing Jin (<i>East China Normal University</i>), Aoying Zhou (<i>East China Normal University</i>), Haizhen Zhuo (<i>Ant Group</i>), Quangqing Xu (<i>Ant Group</i>)

• TardisDB: Extending SQL to Support Versioning	2775
Maximilian E. Schüle (<i>Technical University of Munich</i>), Josef Schmeißer (<i>Technical University of Munich</i>), Thomas Blum (<i>Technical University of Munich</i>), Alfons Kemper (<i>Technical University of Munich</i>), Thomas Neumann (<i>Technical University of Munich</i>)	
• GRIP: Constraint-based Explanation of Missing Answers for Graph Queries	2779
Qi Song (<i>Amazon.com</i>), Hancho Ma (<i>Case Western Reserve University</i>), Peng Lin (<i>Washington State University</i>), Yinghui Wu (<i>Case Western Reserve University, Pacific Northwest National Laboratory</i>)	
• FeatTS: Feature-based Time Series Clustering	2784
Donato Tiano (<i>Lyon 1 University</i>), Angela Bonifati (<i>Lyon 1 University</i>), Raymond Ng (<i>University of British Columbia</i>)	
• SOFOS: Demonstrating the Challenges of Materialized View Selection on Knowledge Graphs	2789
Georgia Troullinou (<i>FORTH-ICS</i>), Haridimos Kondylakis (<i>FORTH-ICS</i>), Matteo Lissandrini (<i>Aalborg University</i>), Davide Mottin (<i>Aarhus University</i>)	
• Demonstrating UDO: A Unified Approach for Optimizing Transaction Code, Physical Design, and System Parameters via Reinforcement Learning	2794
Junxiong Wang (<i>Cornell University</i>), Immanuel Trummer (<i>Cornell University</i>), Debabrota Basu (<i>Scool, Inria Lille-Nord Europe</i>)	
• Demonstrating Robust Voice Querying with MUVE: Optimally Visualizing Results of Phonetically Similar Queries	2798
Ziyun Wei (<i>Cornell University</i>), Immanuel Trummer (<i>Cornell University</i>), Connor Anderson (<i>Cornell University</i>)	
• Vertex-Centric Visual Programming for Graph Neural Networks	2803
Yidi Wu (<i>The Chinese University of Hong Kong</i>), Yuntao Gui (<i>The Chinese University of Hong Kong</i>), Tatiana Jin (<i>The Chinese University of Hong Kong</i>), James Cheng (<i>The Chinese University of Hong Kong</i>), Xiao Yan (<i>Southern University of Science and Technology</i>), Peiqi Yin (<i>Southern University of Science and Technology</i>), Yufei Cai (<i>Southern University of Science and Technology</i>), Bo Tang (<i>Southern University of Science and Technology</i>), Fan Yu (<i>Huawei Technologies Co. Ltd</i>)	
• DPGraph: A Benchmark Platform for Differentially Private Graph Analysis	2808
Siyuan Xia (<i>University of Waterloo</i>), Beizhen Chang (<i>University of Waterloo</i>), Karl Knopf (<i>University of Waterloo</i>), Yihan He (<i>New York University</i>), Yuchao Tao (<i>Duke University</i>), Xi He (<i>University of Waterloo</i>)	

Tutorial Track Papers

• Permissioned Blockchains: Properties, Techniques and Applications	2813
Mohammad Javad Amiri (<i>University of Pennsylvania</i>), Divyakant Agrawal (<i>University of California, Santa Barbara</i>), Amr El Abbadi (<i>University of California, Santa Barbara</i>)	
• Querying in the Age of Graph Databases and Knowledge Graphs	2821
Marcelo Arenas (<i>Universidad Católica & IMFD</i>), Claudio Gutierrez (<i>DCC, Universidad de Chile & IMFD</i>), Juan F. Sequeda (<i>data.world</i>)	
• Cohesive Subgraph Search over Big Heterogeneous Information Networks: Applications, Challenges, and Solutions	2829
Yixiang Fang (<i>The Chinese University of Hong Kong, Shenzhen</i>), Kai Wang (<i>University of New South Wales</i>), Xuemin Lin (<i>University of New South Wales</i>), Wenjie Zhang (<i>University of New South Wales</i>)	
• Practical Security and Privacy for Database Systems	2839
Xi He (<i>University of Waterloo</i>), Jennie Rogers (<i>Northwestern University</i>), John Bater (<i>Duke University</i>), Ashwin Machanavajjhala (<i>Duke University</i>), Chenghong Wang (<i>Duke University</i>), Xiao Wang (<i>Northwestern University</i>)	
• A Deep Dive into Deep Learning Approaches for Text-to-SQL Systems	2846
George Katsogiannis-Meimarakis (<i>Athena Research Center</i>), Georgia Koutrika (<i>Athena Research Center</i>)	
• Not your Grandpa's SSD: The Era of Co-Designed Storage Devices	2852
Alberto Lerner (<i>University of Fribourg</i>), Philippe Bonnet (<i>IT University of Copenhagen</i>)	

- **AI Meets Database: AI4DB and DB4AI** 2859
Guoliang Li (*Tsinghua University*), Xuanhe Zhou (*Tsinghua University*),
Lei Cao (*Massachusetts Institute of Technology*)
- **Deep Learning: Systems and Responsibility** 2867
Abdul Wasay (*Harvard University*), Subarna Chatterjee (*Harvard University*),
Stratos Idreos (*Harvard University*)

Panel Track

- **Data Management to Social Science and Back in the Future of Work** 2876
Sihem Amer-Yahia (*CNRS, University Grenoble Alpes*), Senjuti Basu Roy (*New Jersey Institute of Technology*)
- **Automation of Data Prep, ML, and Data Science: New Cure or Snake Oil?** 2878
Arun Kumar (*University of California, San Diego*)

Graduate Student Research Competition Track Abstracts

- **Temporal Dependencies for Graphs** 2881
Morteza Alipourlangouri (*McMaster University*)
- **Data Lakes Empowered by Knowledge Graph Technologies** 2884
Ahmed Helal (*Concordia University*)
- **Wisconsin Benchmark Data Generator: To JSON and Beyond** 2887
Shiva Jahangiri (*Univrsity of California, Irvine*)
- **Framework for Differentially Private Data Analysis with Multiple Accuracy Requirements** 2890
Karl Knopf (*University of Waterloo*)
- **Flow Provenance in Temporal Interaction Networks** 2893
Chrysanthi Kotsyfaki (*University of Ioannina*)
- **Efficient Deterministic Concurrency Control Under Practical Isolation Levels** 2896
Ziliang Lai (*The Chinese University of Hong Kong*)
- **Grouped Learning: Group-By Model Selection Workloads** 2899
Side Li (*University of California, San Diego*)
- **XLJoins** 2902
Ali Mohammadi Shanghooshabad (*University of Warwick*)
- **Raptor: Large Scale Processing of Big Raster + Vector Data** 2905
Samriddhi Singla (*University of California, Riverside*)
- **Efficiently Supporting Adaptive Multi-Level Serializability Models in Distributed Database Systems** 2908
Zhanhao Zhao (*Renmin University of China*)
- **Contextual Data Cleaning with Ontology FDs** 2911
Zheng Zheng (*McMaster University*)
- **Data Summarization with Hierarchical Taxonomy** 2914
Xuliang Zhu (*Hong Kong Baptist University*)

Undergraduate Student Research Competition Track Abstracts

- **Accelerating Product Quantization Query Execution Runtime** 2917
Ikraduya Edian (*Bandung Institute of Technology*)
- **GraphGem: Optimized Scalable System for Graph Convolutional Networks** 2920
Aditya Gemawat (*University of California, San Diego*)
- **Learning Algorithms for Automatic Data Structure Design** 2923
Demi Guo (*Harvard University*)

• Subteam Replacement: Problem Definition and Fast Solution.....	2926
Zhaoheng Li (<i>University of Illinois at Urbana-Champaign</i>),	
Xinyu Pi (<i>University of Illinois at Urbana-Champaign</i>),	
Mingyuan Wu (<i>University of Illinois at Urbana-Champaign</i>)	
• Model-Parallel Model Selection for Deep Learning Systems	2929
Kabir Nagrecha (<i>University of California, San Diego</i>)	
• Index-Based Join Size Estimation Using Adaptive Sampling	2932
Sergiu Pocol (<i>University of Waterloo</i>)	
Author Index	2934

SIGMOD 2021 Organization

General Chairs: Guoliang Li, Tsinghua University (China)
Zhanhuai Li, Northwestern Polytechnical University (China)

Program Chairs: Stratos Idreos, Harvard University (USA)
Divesh Srivastava, AT&T (USA)

Associate Editors: Azza Abouzied, NYU Abu Dhabi (UAE)
Sihem Amer-Yahia, CNRS, Univ. Grenoble Alpes (France)
Phil Bernstein, Microsoft Research (USA)
Sourav Bhowmick, Nanyang Technological University (Singapore)
Rada Chirkova, NC State University (USA)
Graham Cormode, University of Warwick (UK)
Alan Fekete, The University of Sydney (Australia)
Wolfgang Gatterbauer, Northeastern University (USA)
H. V. Jagadish, University of Michigan (USA)
Georgia Koutrika, Athena Research Center (Greece)
Wolfgang Lehner, Technische Universität Dresden (Germany)
Alexandra Meliou, University of Massachusetts (USA)
Renée Miller, Northeastern University (USA)
Jeff Naughton, Google (USA)
Beng Chin Ooi, National University of Singapore (Singapore)
Tamer Ozsu, University of Waterloo (Canada)
Themis Palpanas, University of Paris (France)
Alkis Polyzotis, Google (USA)
Ken Ross, Columbia University (USA)
Timos Sellis, Swinburne University of Technology (Australia)
S. Sudarshan, IIT Bombay (India)
Pinar Tozun, IT University of Copenhagen (Denmark)
Meihui Zhang, Beijing Institute of Technology (China)

Industry Chairs: Anastasia Ailamaki, EPFL (Switzerland)
Fatma Ozcan, Google (USA)

Demonstrations Chairs: Spyros Blanas, Ohio State University (USA)
Katja Hose, Aalborg University (Denmark)

Tutorials Chairs: Julia Stoyanovich, New York University (USA)
Xiaofang Zhou, University of Queensland (Australia)

Panels Chairs: Tiziana Catarci, Sapienza Università di Roma (Italy)
Sam Madden, Massachusetts Institute of Technology (USA)

Students Competition Chairs: Xi He, University of Waterloo (Canada)
Yongjoo Park, University of Illinois Urbana Champaign (USA)

New Researcher Symposium Azza Abouzzied, NYU Abu Dhabi (UAE)
Chairs: Leilani Battle, University of Maryland (USA)

Programming Competition Chairs Donatella Firmani, Roma Tre University (Italy)
Giovanni Simonini, University of Modena and Reggio Emilia (Italy)

Diversity and Inclusion Chairs Avrilia Floratou, Microsoft (USA)
Arun Kumar, University of California San Diego (USA)

Reproducibility Chair Manos Athanassoulis, Boston University (USA)

Reproducibility Advisors Juliana Freire, New York University (USA)
Dennis Shasha, New York University (USA)

Local Organization Chair Qun Chen, Northwestern Polytechnical University (China)

Local Arrangements Vice-Chairs Jiangtao Cui, Xidian University (China)
Pinghui Wang, Xi'an Jiaotong University (China)
Guixin Ye, Northwest University (China)

Sponsorship Chairs Bill Howe, University of Washington (USA)
Yongxin Tong, Beihang University (China)

Mentorship Chair Dong Deng, Rutgers University (USA)
Tianzheng Wang, Simon Fraser University (Canada)

Finance Chair Xuequn Shang, Northwestern Polytechnical University (China)

Registration Chair Ju Fan, Renmin University of China (China)

SIGMOD Publicity Chair Jiannan Wang, Simon Fraser University (Canada)

Exhibit Chair Nan Tang, HBKU (Qatar)

Proceedings Chairs Lei Cao, Massachusetts Institute of Technology (USA)
Asterios Katsifodimos, Delft University of Technology (Holland)

Program Committee: Aaron Elmore, University of Chicago (USA)
Abolfazl Asudeh, University of Illinois at Chicago (USA)
Adriane Chapman, University of Southampton (UK)
Agma Traina, ICMC-USP (Brazil)
Alekh Jindal, Microsoft (USA)
Alexander Boehm, SAP SE (Germany)
Alkis Simitsis, Hewlett Packard Enterprise (USA)
Alon Halevy, Facebook (USA)
Alvin Cheung, University of California, Berkeley (USA)
Andreas Kipf, Massachusetts Institute of Technology (USA)
Anja Gruenheid, Google Inc. (USA)
Anna Fariha, University of Massachusetts Amherst (USA)
Antonis Deligiannakis, Technical University of Crete (Greece)
Arnab Nandi, OSU (USA)
Avigdor Gal, Technion - Israel Institute of Technology (Israel)
Azadeh Nazi, Google Brain (USA)
Baihua Zheng, Singapore Management University (Singapore)
Bailu Ding, Microsoft Research (USA)
Barna Saha, University of California, Berkeley (USA)
Birgitta König-Ries, Friedrich-Schiller-Universität Jena (Germany)
Boris Glavic, Illinois Institute of Technology (USA)
Byron Choi, Hong Kong Baptist University (Hong Kong SAR)
Carsten Binnig, TU Darmstadt (Germany)
Ce Zhang, ETH (Switzerland)
Chenggang Wu, University of California, Berkeley (USA)
Christian Jensen, Aalborg University (Denmark)
Christopher De Sa, Cornell University (USA)
Chunbin Lin, Amazon AWS (USA)
Curtis Dyreson, Utah State University (USA)
Da Yan, University of Alabama at Birmingham (USA)
Danica Porobic, Oracle (USA)
Davide Mottin, Aarhus University (Denmark)
Dumitrel Loghin, National University of Singapore (Singapore)
Eddie Kohler, Harvard University (USA)
Elena Ferrari, University of Insubria, Varese (Italy)
Elisa Bertino, Purdue University (USA)
Eric Lo, Chinese University of Hong Kong (Hong Kong SAR)
Erkang Zhu, Microsoft (USA)
Erietta Liarou, Vertica (USA)
Essam Mansour, Concordia University (Canada)
Evangelia Sitaridi, Amazon Web Services (USA)
Fatemeh Nargesian, University of Rochester (USA)
Fei Chiang, McMaster University (Canada)
Feifei Li, Alibaba Group (USA)

- Program Committee Continued**
- Foto Afrati, NTUA (Greece)
Francesco Bonchi, ISI Foundation, Turin (Italy)
Gabriela Jacques-Silva, Facebook (USA)
George Fakas, Uppsala University (Sweden)
George Fletcher, Eindhoven University of Technology the Netherlands
(Netherlands)
George Kollios, Boston University (USA)
George Papadakis, University of Athens (Greece)
Giovanni Simonini, University of Modena and Reggio Emilia (Italy)
Hakan Ferhatosmanoglu, University of Warwick (UK)
Hank Korth, Lehigh University (USA)
Holger Pirk, Imperial College (UK)
Huachen Zhang, Carnegie Mellon University (USA)
Hyungsoo Jung, Hanyang University (Korea)
Ilaria Bartolini, University of Bologna (Italy)
Immanuel Trummer, Cornell University (USA)
Ioana Manolescu, INRIA and Institut Polytechnique de Paris (France)
Irini Fundulaki, ICS FORTH, Greece (Greece)
Ismail Oukid, Snowflake Computing (Germany)
Jeffrey Xu Yu, Chinese University of Hong Kong (China)
Jennie Rogers, Northwestern University (USA)
Jens Dittrich, Saarland University (Germany)
Jens Teubner, TU Dortmund University (Germany)
Jianguo Wang, Purdue University (USA)
Johann-Christoph Freytag, Humboldt-Universität zu Berlin (Germany)
Johannes Gehrke, Microsoft (USA)
John Paparrizos, University of Chicago (USA)
Jonathan Goldstein, Microsoft (USA)
Jose Faleiro, Microsoft (USA)
Joy Arulraj, Georgia Tech (USA)
Ju Fan, Renmin University of China (China)
Julia Stoyanovich, New York University (USA)
K. Selçuk Candan, Arizona State University (USA)
Karima Echihabi, Mohammed VI Polytechnic University (Morocco)
Khuzaima Daudjee, University of Waterloo (Canada)
Konstantinos Karanasos, Microsoft (USA)
Kostas Zoumpatianos, Harvard University (USA)
Kristen Lefevre, Google (Switzerland)
Kurt Stockinger, ZHAW Zurich University of Applied Sciences (Switzerland)
Laure Berti-Equille, AMU (France)
Laurel Orr, Stanford University (USA)
Lei Chen, HKUST (USA)
Letizia Tanca, Politecnico di Milano (Italy)
Lyublena Antova, Datometry (USA)
Manos Athanassoulis, Boston University (USA)
Mark Callaghan, MongoDB (USA)

Program Committee Continued

Matthias Boehm, Graz University of Technology (Austria)
Maya Ramanath, IIT Delhi (India)
Melanie Herschel, Universität Stuttgart (Germany)
Michael Gubanov, Florida State University (USA)
Michael Mior, Rochester Institute of Technology (USA)
Michaela Hardt, Amazon (USA)
Minos Garofalakis, Technical University of Crete (Greece)
Mirella Moro, Universidade Federal de Minas Gerais (Brazil)
Mohamed Mokbel, University of Minnesota - Twin Cities (USA)
Mohammad Sadoghi, UC Davis (USA)
Mohammed Eunus Ali, BUET (Bangladesh)
Mónica Scannapieco, ISTAT (Italy)
Nesime Tatbul, Intel Labs and Massachusetts Institute of Technology (USA)
Nga Tran, Tableau (USA)
Nikos Mamoulis, University of Ioannina (Greece)
Nikos Ntarmos, University of Glasgow (UK)
Niv Dayan, Harvard (USA)
Oana Balmau, McGill (Canada)
Olga Papaemmanouil, Brandeis University (USA)
Olga Poppe, Microsoft (USA)
Oliver Kennedy, University at Buffalo, SUNY (USA)
Orestis Polychroniou, Amazon (USA)
Paolo Papotti, Eurecom (France)
Paul Suganthan, Google (USA)
Peter Alvaro, University of California, Santa Cruz (USA)
Peter Haas, University of Massachusetts Amherst (USA)
Peter Pietzuch, Imperial College London (UK)
Peter Triantafillou, University of Warwick (UK)
Philippe Bonnet, IT Univ Copenhagen (Denmark)
Pierangela Samarati, Universita delgi Studi di Milano (Italy)
Prithviraj Sen, IBM Almaden Research Center (USA)
Qian Lin, National University of Singapore (Singapore)
Qin Zhang, Indiana University Bloomington (USA)
Qiong Luo, Hong Kong University of Science and Technology (Hong Kong SAR)
Raghav Kaushik, Microsoft (USA)
Rebecca Taft, Cockroach Labs (USA)
Renata Borovica-Gajic, University of Melbourne (Australia)
Reynold Cheng, "The University of Hong Kong, China" (Hong Kong SAR)
Ryan Johnson, Amazon Web Services (USA)
Ryan Marcus, Massachusetts Institute of Technology (USA)
Saravanan Thirumuruganathan, QCRI (Qatar)
Sajjadur Rahman, Megagon Labs (USA)
Sebastian Link, University of Auckland (New Zealand)
Sebastian Michel, TU Kaiserslautern (Germany)

Program Committee Continued

Sebastian Schelter, University of Amsterdam (Netherlands)
Semih Salihoglu, University of Waterloo (Canada)
Senjuti Basu Roy, NJIT (USA)
Sergey Melnik, Google (USA)
Shrainik Jain, University of Washington (USA)
Slava Novgorodov, eBay Research (Israel)
Sonia, Bergamaschi, Università di Modena e Reggio Emilia (Italy)
Subarna Chatterjee, Harvard (USA)
Sudeepa Roy, Duke University (USA)
Sudipto Das, Amazon Web Services (USA)
Tarique Siddiqui, UIUC (USA)
Theodore Johnson, AT&T Labs - Research (USA)
Thomas Neumann, TU Munich (Germany)
Tianzheng Wang, Simon Fraser University (Canada)
Tien Tuan Anh Dinh, Singapore University of Technology and Design
(Singapore)
Tilmann Rabl, HPI, University of Potsdam (Germany)
Umar Farooq Minhas, Microsoft Research (USA)
Vagelis Hristidis, University of California, Riverside (USA)
Vana Kalogeraki, Athens University of Economics and Business (Greece)
Vanessa Braganholo, Fluminense Federal University (Brazil)
Verena Kantere, National Technical University of Athens (Greece)
Volker Markl, Technische Universität Berlin (Germany)
Walid Aref, Purdue University (USA)
Wei Lu, Renmin University of China (China)
Wei Wang, NUS (Singapore)
Wenjie Zhang, University of New South Wales (Australia)
Wolfram Wingerath, Baqend (Germany)
Wook-Shin Han, POSTECH (Korea)
Xi He, University of Waterloo (Canada)
Xiangyao Yu, University of Wisconsin-Madison (USA)
Xiaofei Zhang, University of Memphis (USA)
Xiaokui Xiao, National University of Singapore (Singapore)
Xuemin Lin, University of New South Wales (Australia)
Yael Amsterdamer, Bar-Ilan University (Israel)
Yanyan Shen, Shanghai Jiao Tong University (China)
Yeye He, Microsoft Research (USA)
Yinghui Wu, Case Western Reserve University (USA)
Yuval Moskovich, University of Michigan (USA)
Zack Ives, University of Pennsylvania (USA)
Zhifeng Bao, RMIT University (Australia)
Zi Huang, University of Queensland (Australia)
Zoi Kaoudi, TU Berlin (Germany)
Zsolt István, IMDEA Software (Spain)

Industrial Track PC Members Alexander Shraer, Apple (USA)
Avrilia Floratou, Microsoft (USA)
Bjorn Jonsson, IT University of Copenhagen (Denmark)
C Mohan, Tsinghua University (China)
Cagatay Demiralp, Sigma Computing (USA)
Donald Kossmann, Microsoft Research (USA)
Eleni Tzirita Zacharatou, TU Berlin (Germany)
Eliezer Levy, Huawei (China)
Eser Kandogan, Megagon Labs (USA)
Haixun Wang, Facebook (USA)
Hamid Pirahesh, IBM Research (USA)
Ihab Ilyas, U. of Waterloo (Canada)
Ippokratis Pandis, Amazon Web Services (USA)
Iraklis Psaroudakis, Oracle (USA)
Jingren Zhou, Alibaba Group (China)
Justin Levandoski, Google (USA)
Jyoti Leeka, Microsoft (USA)
Kai Zeng, Alibaba (China)
Matthias Boehm, Graz University of Technology (Austria)
Norman May SAP SE (Germany)
Sudip Roy, Google (USA)
Xiao Qin, IBM Research (USA)
Yongzhu Li, Singlestore.com (USA)
Yuanyuan Tian, IBM Almaden (USA)

Demo Track PC Members Ahmed Eldawy, University of California, Riverside (USA)

Angela Bonifati, Univ. of Lyon (France)

Asterios Katsifodimos, TU Delft (Holland)

Babak Salimi, Unievristy of California, San Diego (USA)

Daniele Dell'Aglio, Aalborg University (Denmark)

Essam Mansour, Concordia University (Canada)

Evaggelia Pitoura, Univ. of Ioannina (Greece)

Florin Rusu, University of California, Merced (USA)

George Fletcher, Eindhoven University of Technology (Holland)

Georgia Koutrika, Athena Research Center (Greece)

Jens Teubner, TU Dortmund University (Germany)

Jiankai Sun, ByteDance (China)

Jiongqian LIANG, Google (USA)

Jyoti Leeka, Microsoft (USA)

Kai-Uwe Sattler, TU Ilmenau (Germany)

Kavitha Srinivas, IBM Research (USA)

Lukasz Golab, University of Waterloo (Canada)

Martin Theobald, University of Luxemburg (Luxembourg)

Matteo Lissandrini, Aalborg University (Denmark)

Michael Gertz, Heidelberg University (Germany)

Michael Grossniklaus, University of Konstanz (Germany)

Michael Gubanov, Florida State University (USA)

Mohammad Sadoghi, University of California, Davis (USA)

Nikhita Vedula, Amazon (USA)

Rainer Gemulla, Universität Mannheim (Germany)

Ralf Schenkel, University of Trier (Germany)

Raul Castro, Fernandez UChicago (USA)

Sebastian Michel, TU Kaiserslautern (Germany)

Shady Elbassuoni, AUB

Srikanta Bedathur, IIT Delhi (India)

Stefania Dumbrava, ENSIIE (France)

Stefanie Scherzinger, University of Passau (Germany)

Subhadeep Sarkar, Boston University (USA)

Tomer Sagi, University of Haifa (Israel)

Ulf Leser, Humboldt-Universität zu Berlin (Germany)

Xiangyao Yu, University of Wisconsin-Madison (USA)

Xiao Hu, Duke University (USA)

Xiaoyi Lu, University of California, Merced (USA)

Xu Chu, GATECH (USA)

Yannis Theodoridis, University of Piraeus (Greece)

External Reviewers Caetano Traina Junior, University of Sao Paulo (Brazil)
Chenhao Ma, University of Hong Kong (Hong Kong SAR)
Chris Conlan, University of Warwick (UK)
Davide Azzalini, Politecnico di Milano (Italy)
Dmytro Bogatov, Boston University (USA)
Fabio Azzalini, Politecnico di Milano (Italy)
Farhana Murtaza Choudhury, University of Melbourne (Australia)
Francesco Pugliese, Italian National Institute of Statistics (Italy)
Guimu Guo, The University of Alabama at Birmingham (USA)
Hao Jiang, University of Chicago (USA)
Jalal Khalil, The University of Alabama at Birmingham (USA)
Jelle Hellings, University of California, Davis (USA)
Kaustubh Beedkar, TU Berlin (Germany)
Luca Gagliardelli, University of Modena & Reggio Emilia (Italy)
Martin Kiefer, TU Berlin (Germany)
Michael Shekelyan, University of Warwick (UK)
Mirela Cazzolato, University of São Paulo (Brazil)
Philipp Grulich, TU Berlin (Germany)
Riu Liu, University of Chicago (USA)
Sabrina De Capitani di Vimercati, Universita degli Studi di Milano (Italy)
Sara Foresti, Universita' degli Studi di Milano (Italy)
Shufan Zhang, University of Waterloo (Canada)
Subhadeep Sarkar, Boston University (USA)
Tanzima Hashem, BUET (Bangladesh)
Teddy Cunningham, University of Warwick (UK)
Wenya Sun, University of Hong Kong (Hong Kong SAR)
William Ma, University of Chicago (USA)
Yikai Zhang, The Chinese University of Hong Kong (Hong Kong SAR)
Yiyuan Xiong, National University of Singapore (Singapore)

SIGMOD 2021 Sponsor & Supporters

Sponsor



Diamond Supporters



Platinum Supporters



FACEBOOK



Megagon Labs

ORACLE

ZTE

Tencent 腾讯

Gold Supporters

