

June 14 –19, 2020  
Portland, OR, USA



Association for  
Computing Machinery

*Advancing Computing as a Science & Profession*



# SIGMOD'20

Proceedings of the 2020 ACM SIGMOD International  
Conference on  
**Management of Data**

*Sponsored by:*

**ACM SIGMOD**

*General Chairs:*

**David Maier (Portland State University, USA)**

**Rachel Pottinger (University of British Columbia, Canada)**

*Program Chairs:*

**AnHai Doan (University of Wisconsin, USA)**

**Wang-Chiew Tan (Megagon Labs, USA)**

*Proceedings Chairs:*

**Abdussalam Alawini (University of Illinois at Urbana-Champaign, USA)**

**Hung Q. Ngo (RelationalAI, USA)**



**Association for  
Computing Machinery**

*Advancing Computing as a Science & Profession*

**The Association for Computing Machinery**  
1601 Broadway, 10<sup>th</sup> Floor  
New York, NY 10019-7434

Copyright © 2020 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](mailto: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](http://www.copyright.com).

**ISBN:** 978-1-4503-6735-6

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](mailto:acmhelp@acm.org)  
Hours of Operation: 8:30 am – 4:30 pm ET

Printed in the USA

# **Welcome to SIGMOD 2020 — The 2020 ACM SIGMOD International Conference on the Management of Data!**

This year, SIGMOD is heading for a new frontier and is being held entirely online, instead of at its originally planned location of Portland, Oregon, USA. The influence of that location is still a bit evident, in the logo and in the schedule, which uses the Pacific time zone. Despite the challenging times that we find ourselves in, we have an exciting program with many more papers than previous SIGMODs. We have tried to maintain as many program elements as possible: technical and industrial talks, keynotes, tutorials, panels, demonstrations, and the awards session.

We are using the latest technologies to keep SIGMOD as vibrant as previous SIGMODs, and we will be archiving more of SIGMOD than for past conferences, for those who want to review the talks and other sessions at a later date.

In the Research Track this year, we received 458 research submissions (159 for Round 1 and 299 for Round 2), which were extensively reviewed by 168 program committee members, 17 area chairs, and the help of several external reviewers. We accepted 123 submissions (a 26.9% acceptance rate).

In addition, with the approval of SIGMOD EC, we introduced a new category called “short paper”. This new category showcases interesting ongoing work, new discoveries, insights or experience, summaries of significant projects, and interdisciplinary work. Short papers might not possess the journal-like quality that reviewers often expect of papers in the regular research category, but they cover interesting work that we (and the reviewers and area chairs) believe should be brought to the wider community's attention now, to rapidly advance the field. After a rigorous reviewing process, we ended up accepting 21 more papers as short papers.

In addition to the Research Track, the Industrial Track selected 21 papers from 48 submissions; the Demonstration Track selected 36 demonstrations out of 100 submissions; the Tutorial Track selected 8 tutorials from 19 submissions; the Student Research Competition selected 19 from 36 submissions; and the Programming Contest selected 5 finalists from 41 submissions.

This year, we will have three exciting keynote talks, reflecting emerging topics of great interest to the data management community: “Systems and ML: When the Sum is Greater than Its Parts” by Ion Stoica (UC Berkeley), “When the Web is your Data Lake: Creating a Search Engine for Datasets on the Web” by Natasha Noy (Google), and “The Challenge of Building Effective, Enterprise-scale Data Lakes” by Awez Syed (Databricks).

In addition, we will have a timely and interesting plenary panel, organized by Surajit Chaudhuri (Microsoft) and Magdalena Balazinska (Univ. of Washington), on “The Next 5 Years: What Opportunities Should the Database Community Seize to Maximize its Impact?”. There is also an industrial panel on startups founded by database researchers, organized by C. Mohan (IBM Almaden Research Center), probably the first such panel for SIGMOD conferences. The panelists will discuss the trials and tribulations of their entrepreneurial efforts, what worked and

what did not, and cover other topics on startups, spanning early-stage to successfully exited ones, both in the USA and Europe.

Thus, SIGMOD 2020 will feature an exciting and rich program, with 144 research papers, 21 industrial papers, 36 demonstrations, 8 tutorials, 3 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 programs, 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 2020, 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 has 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 2020. We hope you will enjoy the conference, and see you online!

**David Maier**  
**Rachel Pottinger**  
*General Chairs*

**AnHai Doan**  
**Wang-Chiew Tan**  
*Program Chairs*

**Abdussalam Alawini**  
**Hung Q. Ngo**  
*Proceedings Chairs*

# Table of Contents

<b>SIGMOD 2020 Organization .....</b>	<b>xxix</b>
---------------------------------------	-------------

<b>SIGMOD 2020 Sponsor &amp; Supporters .....</b>	<b>xxxiii</b>
---	---------------

## **SIGMOD Keynote 1**

• <b>Systems and ML: When the Sum is Greater than Its Parts.....</b>	<b>1</b>
Ion Stoica ( <i>University of California, Berkeley</i> )	

## **Research 1: Crowdsourcing and Visualization**

• <b>Recommending Deployment Strategies for Collaborative Tasks .....</b>	<b>3</b>
Dong Wei, Senjuti Basu Roy ( <i>New Jersey Institute of Technology</i> ), Sihe Amer-Yahia ( <i>CNRS, University Grenoble Alpes</i> )	
• <b>Human-in-the-loop Outlier Detection.....</b>	<b>19</b>
Chengliang Chai ( <i>Tsinghua University</i> ), Lei Cao ( <i>CSAIL, MIT</i> ), Guoliang Li, Jian Li, Yuyu Luo ( <i>Tsinghua University</i> ), Samuel Madden ( <i>CSAIL, MIT</i> )	
• <b>QUAD: Quadratic-Bound-based Kernel Density Visualization .....</b>	<b>35</b>
Tsz Nam Chan, Reynold Cheng ( <i>The University of Hong Kong</i> ), Man Lung Yiu ( <i>Hong Kong Polytechnic University</i> )	
• <b>ShapeSearch: A Flexible and Efficient System for Shape-based Exploration of Trendlines.....</b>	<b>51</b>
Tarique Siddiqui, Paul Luh, Zesheng Wang, Karrie Karahalios ( <i>University of Illinois at Urbana-Champaign</i> ), Aditya Parameswaran ( <i>University of California, Berkeley</i> )	
• <b>Marviq: Quality-Aware Geospatial Visualization of Range-Selection Queries Using Materialization .....</b>	<b>67</b>
Liming Dong ( <i>Tsinghua University</i> ), Qiushi Bai, Taewoo Kim, Taiji Chen ( <i>University of California, Irvine</i> ), Weidong Liu ( <i>Tsinghua University</i> ), Chen Li ( <i>University of California, Irvine</i> )	

## **Research 2: Serverless and Cloud Data Management**

• <b>Transactional Causal Consistency for Serverless Computing.....</b>	<b>83</b>
Chenggang Wu, Vikram Sreekanti, Joseph M. Hellerstein ( <i>University of California, Berkeley</i> )	
• <b>Cost Models for Big Data Query Processing: Learning, Retrofitting, and Our Findings .....</b>	<b>99</b>
Tarique Siddiqui ( <i>Microsoft &amp; University of Illinois at Urbana-Champaign</i> ), Alekh Jindal, Shi Qiao, Hiren Patel, Wangchao Le ( <i>Microsoft</i> )	
• <b>Lambda: Interactive Data Analytics on Cold Data Using Serverless Cloud Infrastructure ....</b>	<b>115</b>
Ingo Müller, Renato Marroquín, Gustavo Alonso ( <i>ETH Zürich</i> )	
• <b>Starling: A Scalable Query Engine on Cloud Functions.....</b>	<b>131</b>
Matthew Perron ( <i>Massachusetts Institute of Technology</i> ), Raul Castro Fernandez ( <i>University of Chicago</i> ), David DeWitt, Samuel Madden ( <i>Massachusetts Institute of Technology</i> )	
• <b>Learning a Partitioning Advisor for Cloud Databases .....</b>	<b>143</b>
Benjamin Hilprecht, Carsten Binnig ( <i>TU Darmstadt</i> ), Uwe Röhm ( <i>The University of Sydney</i> )	

## **Research 3: Machine Learning for Databases I**

• <b>DB4ML - An In-Memory Database Kernel with Machine Learning Support .....</b>	<b>159</b>
Matthias Jasny, Tobias Ziegler ( <i>TU Darmstadt</i> ), Tim Kraska ( <i>Massachusetts Institute of Technology</i> ), Uwe Roehm ( <i>The University of Sydney</i> ), Carsten Binnig ( <i>TU Darmstadt</i> )	
• <b>Active Learning for ML Enhanced Database Systems .....</b>	<b>175</b>
Lin Ma ( <i>Carnegie Mellon University</i> ), Bailu Ding ( <i>Microsoft Research</i> ), Sudipto Das ( <i>Amazon Web Services</i> ), Adith Swaminathan ( <i>Microsoft Research</i> )	

• <b>Qd-tree: Learning Data Layouts for Big Data Analytics</b> .....	193
Zongheng Yang ( <i>University of California, Berkeley</i> ), Badrish Chandramouli, Chi Wang ( <i>Microsoft Research</i> ), Johannes Gehrke ( <i>Microsoft</i> ), Yinan Li, Umar Farooq Minhas, Per-Åke Larson, Donald Kossmann ( <i>Microsoft Research</i> ), Rajeev Acharya ( <i>Microsoft</i> )	
• <b>Facilitating SQL Query Composition and Analysis</b> .....	209
Zainab Zolaktaf, Mostafa Milani, Rachel Pottinger ( <i>University of British Columbia</i> )	
• <b>MONSOON: Multi-Step Optimization and Execution of Queries with Partially Obscured Predicates</b> .....	225
Sourav Sikdar, Chris Jermaine ( <i>Rice University</i> )	

## Research 4: Uncertain, Probabilistic, and Approximate Data

• <b>Causal Relational Learning</b> .....	241
Babak Salimi ( <i>University of Washington</i> ), Harsh Parikh ( <i>Duke University</i> ), Moe Kayali ( <i>University of Washington</i> ), Lise Getoor ( <i>University of California, Santa Cruz</i> ), Sudeepa Roy ( <i>Duke University</i> ), Dan Suciu ( <i>University of Washington</i> )	
• <b>Sample Debiasing in the Themis Open World Database System</b> .....	257
Laurel Orr, Magdalena Balazinska, Dan Suciu ( <i>University of Washington</i> )	
• <b>Stochastic Package Queries in Probabilistic Databases</b> .....	269
Matteo Brucato, Nishant Yadav ( <i>University of Massachusetts, Amherst</i> ), Azza Abouzied ( <i>New York University Abu Dhabi</i> ), Peter J. Haas, Alexandra Meliou ( <i>University of Massachusetts, Amherst</i> )	
• <b>Fast and Reliable Missing Data Contingency Analysis with Predicate-Constraints</b> .....	285
Xi Liang, Zechao Shang, Sanjay Krishnan, Aaron J. Elmore, Michael J. Franklin ( <i>University of Chicago</i> )	
• <b>Mining Approximate Acyclic Schemes from Relations</b> .....	297
Batya Kenig, Pranay Mundra, Guna Prasaad, Babak Salimi, Dan Suciu ( <i>University of Washington</i> )	

## Industry 1: Graph Databases and Knowledge Bases

• <b>AliCoCo: Alibaba E-commerce Cognitive Concept Net</b> .....	313
Xusheng Luo, Luxin Liu, Yonghua Yang, Le Bo, Yuanpeng Cao, Jinghang Wu, Qiang Li, Keping Yang ( <i>Alibaba Group</i> ), Kenny Q. Zhu ( <i>Shanghai Jiao Tong University</i> )	
• <b>A1: A Distributed In-Memory Graph Database</b> .....	329
Chiranjeev Buragohain ( <i>Oracle &amp; Microsoft</i> ), Knut Magne Risvik, Paul Brett, Miguel Castro, Wonhee Cho, Joshua Cowhig, Nikolas Gloy, Karthik Kalyanaraman ( <i>Microsoft</i> ), Richendra Khanna ( <i>Oracle &amp; Microsoft</i> ), John Pao, Matthew Renzelmann, Alex Shamis ( <i>Microsoft</i> ), Timothy Tan ( <i>Amazon</i> ), Shuheng Zheng ( <i>Microsoft &amp; Amazon</i> )	
• <b>IBM Db2 Graph: Supporting Synergistic and Retrofittable Graph Queries Inside IBM Db2</b> .....	345
Yuan Yuan Tian, En Liang Xu, Wei Zhao, Mir Hamid Pirahesh, Sui Jun Tong, Wen Sun ( <i>IBM Research</i> ), Thomas Kolanko, Md. Shahidul Haque Apu, Huijuan Peng ( <i>IBM Cloud and Cognitive Software</i> )	
• <b>An Ontology-Based Conversation System for Knowledge Bases</b> .....	361
Abdul Quamar, Chuan Lei ( <i>IBM Research - Almaden</i> ), Dorian Miller ( <i>IBM Watson Health</i> ), Fatma Özcan ( <i>IBM Research - Almaden</i> ), Jeffrey Kreulen ( <i>IBM Watson Health</i> ), Robert J. Moore, Vasilis Efthymiou ( <i>IBM Research - Almaden</i> )	
• <b>Aggregation Support for Modern Graph Analytics in TigerGraph</b> .....	377
Alin Deutsch ( <i>University of California, San Diego &amp; TigerGraph</i> ), Yu Xu, Mingxi Wu, Victor E. Lee ( <i>TigerGraph</i> )	
• <b>GIANT: Scalable Creation of a Web-scale Ontology</b> .....	393
Bang Liu ( <i>University of Alberta</i> ), Weidong Guo ( <i>Tencent</i> ), Di Niu ( <i>University of Alberta</i> ), Jinwen Luo, Chaoyue Wang, Zhen Wen, Yu Xu ( <i>Tencent</i> )	

## SIGMOD Panel

- **The Next 5 Years: What Opportunities Should the Database Community Seize to Maximize its Impact?** ..... 411  
Magda Balazinska (*University of Washington*), Surajit Chaudhuri (*Microsoft Research*),  
Anastasia Ailamaki (*EPFL*), Juliana Freire (*New York University*), Silesh Krishnamurthy (*Google*),  
Michael Stonebraker (*Massachusetts Institute of Technology*)

## Research 5: Data Provenance

- **Equivalence-Invariant Algebraic Provenance for Hyperplane Update Queries** ..... 415  
Pierre Bourhis (*CNRS, UMR 9189 - CRISTAL*), Daniel Deutch, Yuval Moskovitch (*Tel Aviv University*)
- **Causality-Guided Adaptive Interventional Debugging** ..... 431  
Anna Fariha (*University of Massachusetts, Amherst*), Suman Nath (*Microsoft Research*),  
Alexandra Meliou (*University of Massachusetts, Amherst*)
- **PrIU: A Provenance-Based Approach for Incrementally Updating Regression Models** ..... 447  
Yinjun Wu, Val Tannen, Susan B. Davidson (*University of Pennsylvania*)
- **BugDoc: Algorithms to Debug Computational Processes** ..... 463  
Raoni Lourenço, Juliana Freire, Dennis Shasha (*New York University*)
- **Computing Local Sensitivities of Counting Queries with Joins** ..... 479  
Yuchao Tao (*Duke University*), Xi He (*University of Waterloo*),  
Ashwin Machanavajjhala, Sudeepa Roy (*Duke University*)

## Research 6: Transaction Processing and Query Optimization

- **Long-lived Transactions Made Less Harmful** ..... 495  
Jongbin Kim, Hyunsoo Cho, Kihwang Kim, Jaeseon Yu, Sooyong Kang, Hyungsoo Jung (*Hanyang University*)
- **Chiller: Contention-centric Transaction Execution and Data Partitioning for Modern Networks** ..... 511  
Erfan Zamanian (*Brown University*), Julian Shun (*Massachusetts Institute of Technology*),  
Carsten Binnig (*TU Darmstadt*), Tim Kraska (*Massachusetts Institute of Technology*)
- **Handling Highly Contended OLTP Workloads Using Fast Dynamic Partitioning** ..... 527  
Guna Prasaad (*University of Washington*), Alvin Cheung (*University of California, Berkeley*),  
Dan Suciu (*University of Washington*)
- **A Transactional Perspective on Execute-order-validate Blockchains** ..... 543  
Pingcheng Ruan, Dumitrel Loghin, Quang-Trung Ta (*National University of Singapore*),  
Meihui Zhang (*Beijing Institute of Technology*), Gang Chen (*Zhejiang University*),  
Beng Chin Ooi (*National University of Singapore*)
- **Aggify: Lifting the Curse of Cursor Loops using Custom Aggregates** ..... 559  
Surabhi Gupta (*Microsoft Research India*), Sanket Purandare (*Harvard University*),  
Karthik Ramachandra (*Microsoft Research India*)

## Research 7: Security, Privacy, and Blockchain

- **Querying Shared Data with Security Heterogeneity** ..... 575  
Yang Cao (*University of Edinburgh*),  
Wenfei Fan (*University of Edinburgh, Shenzhen University, & Beihang University*),  
Yanghao Wang (*University of Edinburgh*), Ke Yi (*Hong Kong University of Science*)
- **SAGMA: Secure Aggregation Grouped by Multiple Attributes** ..... 587  
Timon Hackenjos (*FZI Research Center for Information Technology*), Florian Hahn (*University of Twente*),  
Florian Kerschbaum (*University of Waterloo*)
- **Crypt? Crypto-Assisted Differential Privacy on Untrusted Servers** ..... 603  
Amrita Roy Chowdhury (*University of Wisconsin-Madison*), Chenghong Wang (*Duke University*),  
Xi He (*University of Waterloo*), Ashwin Machanavajjhala (*Duke University*),  
Somesh Jha (*University of Wisconsin-Madison*)

- **Estimating Numerical Distributions under Local Differential Privacy** ..... 621  
Zitao Li, Tianhao Wang (*Purdue University*), Milan Lopuhaä-Zwakenberg (*Eindhoven University of Technology*),  
Ninghui Li (*Purdue University*), Boris Škoric (*Eindhoven University of Technology*)
- **FalconDB: Blockchain-based Collaborative Database** ..... 637  
Yanqing Peng (*University of Utah*), Min Du (*University of California, Berkeley*), Feifei Li (*University of Utah*),  
Raymond Cheng, Dawn Song (*University of California, Berkeley*)

## Research 8: Graph Query Processing

- **Exact Single-Source SimRank Computation on Large Graphs** ..... 653  
Hanzhi Wang, Zhewei Wei (*Renmin University of China*), Ye Yuan (*Beijing Institute of Technology*),  
Xiaoyong Du, Ji-Rong Wen (*Renmin University of China*)
- **Distributed Processing of  $k$  Shortest Path Queries over Dynamic Road Networks** ..... 665  
Ziqiang Yu (*Yantai University*), Xiaohui Yu (*York University*),  
Nick Koudas (*University of Toronto*), Yang Liu (*Wilfrid Laurier University*),  
Yifan Li (*York University & Key Laboratory of Urban Land Resources Monitoring and Simulation, MNR*),  
Yueting Chen (*York University*), Dingyu Yang (*Alibaba Group*)
- **On the Optimization of Recursive Relational Queries: Application to Graph Queries** ..... 681  
Louis Jachiet (*LTCI, Télécom Paris*),  
Pierre Genevès, Nils Gesbert, Nabil Layaïda (*University Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG*)
- **Pensieve: Skewness-Aware Version Switching for Efficient Graph Processing** ..... 699  
Tangwei Ying, Hanhua Chen, Hai Jin (*Huazhong University of Science and Technology*)
- **Extending Graph Patterns with Conditions** ..... 715  
Grace Fan (*Brown University*), Wenfei Fan (*University of Edinburgh, Beihang University & Shenzhen University*),  
Yuanhao Li (*University of Edinburgh & Shenzhen University*), Ping Lu (*Beihang University*),  
Chao Tian, Jingren Zhou (*Alibaba Group*)

## Industry 2: Machine Learning and Analytics

- **Elastic Machine Learning Algorithms in Amazon SageMaker** ..... 731  
Edo Liberty, Zohar Karnin, Bing Xiang, Laurence Rouesnel, Baris Coskun, Ramesh Nallapati, Julio Delgado,  
Amir Sadoughi, Yury Astashonok, Piali Das, Can Balioglu, Saswata Chakravarty, Madhav Jha, Philip Gautier,  
David Arpin, Tim Januschowski, Valentin Flunkert, Yuyang Wang, Jan Gasthaus, Lorenzo Stella,  
Syama Rangapuram, David Salinas, Sebastian Schelter, Alex Smola (*Amazon AI*)
- **Timon: A Timestamped Event Database for Efficient Telemetry**  
**Data Processing and Analytics** ..... 739  
Wei Cao (*Zhejiang University & Alibaba Group*), Yusong Gao, Feifei Li, Sheng Wang, Bingchen Lin, Ke Xu,  
Xiaojie Feng, Yucong Wang, Zhenjun Liu, Gejin Zhang (*Alibaba Group*)
- **Vertica-ML: Distributed Machine Learning in Vertica Database** ..... 755  
Arash Fard, Anh Le, George Larionov, Waqas Dhillon, Chuck Bear (*Vertica*)
- **Database Workload Capacity Planning using Time Series Analysis**  
**and Machine Learning** ..... 769  
Antony S. Higginson, Mihaela Dediu, Octavian Arsene (*Oracle Advanced Customer Services*),  
Norman W. Paton, Suzanne M. Embury (*University of Manchester*)
- **The Machine Learning Bazaar: Harnessing the ML Ecosystem**  
**for Effective System Development** ..... 785  
Micah J. Smith, Carles Sala (*Massachusetts Institute of Technology*), James Max Kanter (*Feature Labs*),  
Kalyan Veeramachaneni (*Massachusetts Institute of Technology*)

## SIGMOD Keynote 2

- **When the Web is your Data Lake: Creating a Search Engine for Datasets on the Web** ..... 801  
Natasha Noy (*Google Research*)
- **The Challenge of Building Effective, Enterprise-scale Data Lakes** ..... 803  
Awez Syed (*Databricks*)



## Research 9: Data Cleaning

- **Cleaning Denial Constraint Violations through Relaxation** ..... 805  
Stella Giannakopoulou (*EPFL*), Manos Karpathiotakis (*Facebook*), Anastasia Ailamaki (*EPFL*)
- **On Multiple Semantics for Declarative Database Repairs** ..... 817  
Amir Gilad, Daniel Deutch (*Tel Aviv University*), Sudeepa Roy (*Duke University*)
- **Discovery Algorithms for Embedded Functional Dependencies** ..... 833  
Ziheng Wei (*The University of Auckland*), Sven Hartmann (*Clausthal University of Technology*),  
Sebastian Link (*The University of Auckland*)
- **SCODED: Statistical Constraint Oriented Data Error Detection** ..... 845  
Jing Nathan Yan (*Cornell University*), Oliver Schulte, MoHan Zhang, Jiannan Wang (*Simon Fraser University*),  
Reynold Cheng (*The University of Hong Kong*)
- **A Statistical Perspective on Discovering Functional Dependencies in Noisy Data** ..... 861  
Yunjia Zhang, Zhihan Guo, Theodoros Rekatsinas (*University of Wisconsin-Madison*)

## Research 10: Storage and Indexing

- **Rethinking Logging, Checkpoints, and Recovery for High-Performance Storage Engines** .... 877  
Michael Haubenschild, Caetano Sauer (*Tableau Software*), Thomas Neumann (*Technische Universität München*),  
Viktor Leis (*Friedrich-Schiller-Universität Jena*)
- **Lethe: A Tunable Delete-Aware LSM Engine** ..... 893  
Subhadeep Sarkar, Tarikul Islam Papon, Dimitris Staratzis, Manos Athanassoulis (*Boston University*)
- **BinDex: A Two-Layered Index for Fast and Robust Scans** ..... 909  
Linwei Li, Kai Zhang, Jiading Guo, Wen He, Zhenying He, Yinan Jing, Weili Han, X. Sean Wang (*Fudan University*)
- **Analysis of Indexing Structures for Immutable Data** ..... 925  
Cong Yue, Zhongle Xie (*National University of Singapore*), Meihui Zhang (*Beijing Institute of Technology*),  
Gang Chen (*Zhejiang University*), Beng Chin Ooi (*National University of Singapore*),  
Sheng Wang (*Alibaba Group*), Xiaokui Xiao (*National University of Singapore*)
- **Tree-Encoded Bitmaps** ..... 937  
Harald Lang, Alexander Beischl (*Technical University of Munich*),  
Viktor Leis (*Friedrich Schiller University Jena*), Peter Boncz (*Centrum Wiskunde & Informatica*),  
Thomas Neumann, Alfons Kemper (*Technical University of Munich*)

## Research 11: Machine Learning for Databases II

- **ALEX: An Updatable Adaptive Learned Index** ..... 969  
Jialin Ding (*Massachusetts Institute of Technology*), Umar Farooq Minhas (*Microsoft Research*),  
Jia Yu (*Arizona State University & Microsoft Research*), Chi Wang, Jaeyoung Do, Yinan Li (*Microsoft Research*),  
Hantian Zhang (*Georgia Institute of Technology & Microsoft Research*),  
Badrish Chandramouli (*Microsoft Research*), Johannes Gehrke (*Microsoft*),  
Donald Kossmann, David Lomet (*Microsoft Research*), Tim Kraska (*Massachusetts Institute of Technology*)
- **Learning Multi-Dimensional Indexes** ..... 985  
Vikram Nathan, Jialin Ding, Mohammad Alizadeh, Tim Kraska (*Massachusetts Institute of Technology*)
- **The Case for a Learned Sorting Algorithm** ..... 1001  
Ani Kristo (*Brown University*), Kapil Vaidya (*Massachusetts Institute of Technology*),  
Ugur Çetintemel (*Brown University*), Sanchit Misra (*Intel Labs*),  
Tim Kraska (*Massachusetts Institute of Technology*)
- **QuickSel: Quick Selectivity Learning with Mixture Models** ..... 1017  
Yongjoo Park (*University of Illinois at Urbana-Champaign*),  
Shucheng Zhong, Barzan Mozafari (*University of Michigan – Ann Arbor*)
- **Deep Learning Models for Selectivity Estimation of Multi-Attribute Queries** ..... 1035  
Shohedul Hasan (*University of Texas at Arlington*), Saravanan Thirumuruganathan (*QCRI, HBKU*),  
Jees Augustine (*University of Texas at Arlington*), Nick Koudas (*University of Toronto*),  
Gautam Das (*University of Texas at Arlington*)

## Research 12: Graph Matching and Discovery

- **Efficient Algorithms for Densest Subgraph Discovery on Large Directed Graphs** ..... 1051  
Chenhao Ma (*The University of Hong Kong*), Yixiang Fang (*University of New South Wales*),  
Reynold Cheng (*The University of Hong Kong*), Laks V.S. Lakshmanan (*The University of British Columbia*),  
Wenjie Zhang, Xuemin Lin (*University of New South Wales*)
- **GPU-Accelerated Subgraph Enumeration on Partitioned Graphs** ..... 1067  
Wentian Guo (*National University of Singapore*), Yuchen Li (*Singapore Management University*),  
Mo Sha, Bingsheng He, Xiaokui Xiao, Kian-Lee Tan (*National University of Singapore*)
- **In-Memory Subgraph Matching: An In-depth Study** ..... 1083  
Shixuan Sun, Qiong Luo (*Hong Kong University of Science and Technology*)
- **G-CARE: A Framework for Performance Benchmarking of Cardinality Estimation Techniques for Subgraph Matching** ..... 1099  
Yeonsu Park, Seongyun Ko, Sourav S. Bhowmick, Kyoungmin Kim, Kijae Hong,  
Wook-Shin Han (*POSTECH*)
- **Approximate Pattern Matching in Massive Graphs with Precision and Recall Guarantees** ..... 1115  
Tashin Reza, Matei Ripeanu (*University of British Columbia*),  
Geoffrey Sanders, Roger Pearce (*Lawrence Livermore National Laboratory*)

## Research 13: Data Matching

- **A Comprehensive Benchmark Framework for Active Learning Methods in Entity Matching** ..... 1133  
Venkata Vamsikrishna Meduri (*Arizona State University*),  
Lucian Popa, Prithviraj Sen (*IBM Research, Almaden*), Mohamed Sarwat (*Arizona State University*)
- **ZeroER: Entity Resolution using Zero Labeled Examples** ..... 1149  
Renzhi Wu, Sanya Chaba, Saurabh Sawlani, Xu Chu (*Georgia Institute of Technology*),  
Saravanan Thirumuruganathan (*QCRI, HBKU*)
- **Towards Interpretable and Learnable Risk Analysis for Entity Resolution** ..... 1165  
Zhaoqiang Chen, Qun Chen, Boyi Hou, Zhanhuai Li (*Northwestern Polytechnical University*),  
Guoliang Li (*Tsinghua University*)
- **SLIM: Scalable Linkage of Mobility Data** ..... 1181  
Fuat Basik (*Amazon Web Services*), Hakan Ferhatosmanoğlu (*University of Warwick*),  
Búcrá Gedik (*Bilkent University*)
- **Monotonic Cardinality Estimation of Similarity Selection: A Deep Learning Approach** ..... 1197  
Yaoshu Wang (*Shenzhen University*), Chuan Xiao (*Osaka University & Nagoya University*),  
Jianbin Qin (*Shenzhen University*), Xin Cao, Yifang Sun, Wei Wang (*The University of New South Wales*),  
Makoto Onizuka (*Osaka University*)

## Research 14: Query Optimization and Execution

- **Fast Join Project Query Evaluation using Matrix Multiplication** ..... 1213  
Shaleen Deep (*University of Wisconsin-Madison*), Xiao Hu (*Duke University*),  
Paraschos Koutris (*University of Wisconsin-Madison*)
- **Maintaining Acyclic Foreign-Key Joins under Updates** ..... 1225  
Qichen Wang, Ke Yi (*Hong Kong University of Science and Technology*)
- **Thrifty Query Execution via Incrementability** ..... 1241  
Dixin Tang, Zechao Shang, Aaron J. Elmore, Sanjay Krishnan, Michael J. Franklin (*University of Chicago*)
- **A Method for Optimizing Opaque Filter Queries** ..... 1257  
Wenjia He, Michael R. Anderson, Maxwell Strome, Michael Cafarella (*University of Michigan – Ann Arbor*)
- **Functional-Style SQL UDFs With a Capital ‘F’** ..... 1273  
Christian Duta, Torsten Grust (*University of Tübingen*)

## Research 15: Machine Learning for Cleaning, Integration, and Search

- **Learning to Validate the Predictions of Black Box Classifiers on Unseen Data**..... 1289  
Sebastian Schelter (*New York University*), Tammo Rukat (*Amazon Research*),  
Felix Biessmann (*Beuth University Berlin*)
- **Learning Over Dirty Data Without Cleaning** ..... 1301  
Jose Picado, John Davis, Arash Termehchy, Ga Young Lee (*Oregon State University*)
- **Complaint-driven Training Data Debugging for Query 2.0** ..... 1317  
Weiyuan Wu (*Simon Fraser University*), Lampros Flokas, Eugene Wu (*Columbia University*),  
Jiannan Wang (*Simon Fraser University*)
- **Creating Embeddings of Heterogeneous Relational Datasets for Data Integration Tasks**..... 1335  
Riccardo Cappuzzo, Paolo Papotti (*EURECOM*), Saravanan Thirumuruganathan (*QCRI, HBKU*)
- **Minimization of Classifier Construction Cost for Search Queries**..... 1351  
Shay Gershtein, Tova Milo, Gefen Morami (*Tel Aviv University*), Slava Novgorodov (*eBay Research*)

## Research 16: Graph and Stream Processing

- **Scaling Up Distance Labeling on Graphs with Core-Periphery Properties**..... 1367  
Wentao Li (*CAI, FEIT, University of Technology Sydney*), Miao Qiao (*University of Auckland*),  
Lu Qin, Ying Zhang (*CAI, FEIT, University of Technology Sydney*), Lijun Chang (*University of Sydney*),  
Xuemin Lin (*University of New South Wales*)
- **Factorized Graph Representations for Semi-Supervised Learning from Sparse Data**..... 1383  
Krishna Kumar P. (*IIT Madras*), Paul Langton, Wolfgang Gatterbauer (*Northeastern University*)
- **Reliable Data Distillation on Graph Convolutional Network**..... 1399  
Wentao Zhang (*Peking University & National Engineering Laboratory for Big Data Analysis and Applications*),  
Xupeng Miao (*Peking University*), Yingxia Shao (*Beijing University of Posts and Telecommunications, BUPT*),  
Jiawei Jiang (*ETH Zurich*), Lei Chen (*Hong Kong University of Science and Technology*),  
Olivier Ruas (*Peking University*),  
Bin Cui (*Peking University & National Engineering Laboratory for Big Data Analysis and Applications*)
- **Regular Path Query Evaluation on Streaming Graphs**..... 1415  
Anil Pacaci (*University of Waterloo*), Angela Bonifati (*Lyon 1 University*),  
M. Tamer Özsu (*University of Waterloo*)
- **Timely Reporting of Heavy Hitters using External Memory** ..... 1431  
Prashant Pandey (*Carnegie Mellon University*), Shikha Singh (*Wellesley College*),  
Michael A. Bender (*Stony Brook University*), Jonathan W. Berry (*Sandia National Laboratories*),  
Martín Farach-Colton (*Rutgers University*), Rob Johnson (*VMware Research*),  
Thomas M. Kroege, Cynthia A. Phillips (*Sandia National Laboratories*)

## Industry 3: Cloud and Distributed Databases

- **A Framework for Emulating Database Operations in Cloud Data Warehouses**..... 1447  
Mohamed A. Soliman, Lyublena Antova, Marc Sugiyama, Michael Duller, Amirhossein Aleyasen,  
Gourab Mitra, Ehab Abdelhamid, Mark Morcos, Michele Gage, Dmitri Korablev,  
Florian M. Waas (*Datometry, Inc.*)
- **Taurus Database: How to be Fast, Available, and Frugal in the Cloud**..... 1463  
Alex Depoutovitch, Chong Chen, Jin Chen, Paul Larson, Shu Lin, Jack Ng, Wenlin Cui, Qiang Liu,  
Wei Huang, Yong Xiao, Yongjun He (*Huawei Research*)
- **Reliability Analytics for Cloud Based Distributed Databases**..... 1479  
Mathieu B. Demarne, Jim Gramling, Tomer Verona, Miso Cilimdžic (*Microsoft Corporation*)
- **CockroachDB: The Resilient Geo-Distributed SQL Database**..... 1493  
Rebecca Taft, Irfan Sharif, Andrei Matei, Nathan VanBenschoten, Jordan Lewis, Tobias Grieger,  
Kai Niemi, Andy Woods, Anne Birzin, Raphael Poss, Paul Bardea, Amruta Ranade, Ben Darnell,  
Bram Gruneir, Justin Jaffray, Lucy Zhang, Peter Mattis (*Cockroach Labs*)
- **Azure SQL Database Always Encrypted**..... 1511  
Panagiotis Antonopoulos, Arvind Arasu, Kunal D. Singh, Ken Eguro, Nitish Gupta, Rajat Jain, Raghav Kaushik,  
Hanuma Kodavalla, Donald Kossmann, Nikolas Ogg, Ravi Ramamurthy, Jakub Szymaszek, Jeffrey Trimmer,  
Kapil Vaswani, Ramarathnam Venkatesan, Mike Zwilling (*Microsoft Azure and Microsoft Research*)

## Research 17: Data Exploration and Preparation

- **Automatically Generating Data Exploration Sessions Using Deep Reinforcement Learning**..... 1527  
Ori Bar El, Tova Milo, Amit Somech (*Tel Aviv University*)
- **Auto-Suggest: Learning-to-Recommend Data Preparation Steps Using Data Science Notebooks** ..... 1539  
Cong Yan (*University of Washington*), Yeye He (*Microsoft Research*)
- **IDEBench: A Benchmark for Interactive Data Exploration**..... 1555  
Philipp Eichmann (*Brown University*), Emanuel Zraggen (*Massachusetts Institute of Technology*), Carsten Binnig (*TU Darmstadt*), Tim Kraska (*Massachusetts Institute of Technology*)
- **Database Benchmarking for Supporting Real-Time Interactive Querying of Large Data** ..... 1571  
Leilani Battle (*University of Maryland*), Philipp Eichmann (*Brown University*), Marco Angelini, Tiziana Catarci, Giuseppe Santucci (*University of Rome “La Sapienza”*), Yukun Zheng (*University of Maryland*), Carsten Binnig (*Technical University of Darmstadt*), Jean-Daniel Fekete (*Inria, University Paris-Saclay, CNRS*), Dominik Moritz (*University of Washington*)
- **Benchmarking Spreadsheet Systems** ..... 1589  
Sajjadur Rahman (*University of Illinois at Urbana-Champaign*), Kelly Mack (*University of Washington*), Mangesh Bendre (*VISA Research*), Ruilin Zhang (*University of Southern California*), Karrie Karahalios (*University of Illinois at Urbana-Champaign*), Aditya Parameswaran (*University of California, Berkeley*)

## Research 18: Main Memory Databases and Modern Hardware

- **Order-Preserving Key Compression for In-Memory Search Trees**..... 1601  
Huan Chen Zhang, Xiaoxuan Liu, David G. Andersen (*Carnegie Mellon University*), Michael Kaminsky (*BrdgAI*), Kimberly Keeton (*Hewlett Packard Labs*), Andrew Pavlo (*Carnegie Mellon University*)
- **A Study of the Fundamental Performance Characteristics of GPUs and CPUs for Database Analytics** ..... 1617  
Anil Shanbhag, Samuel Madden (*Massachusetts Institute of Technology*), Xiangyao Yu (*University of Wisconsin-Madison*)
- **Pump Up the Volume: Processing Large Data on GPUs with Fast Interconnects** ..... 1633  
Clemens Lutz (*DFKI GmbH*), Sebastian Breß (*TU Berlin*), Steffen Zeuch (*DFKI GmbH*), Tilmann Rabl (*HPI, University of Potsdam*), Volker Markl (*DFKI GmbH, TU Berlin*)
- **Robust Performance of Main Memory Data Structures by Configuration**..... 1651  
Tiemo Bang (*Technical University of Darmstadt & SAP SE*), Ismail Oukid (*Snowflake Inc.*), Norman May (*SAP SE*), Ilia Petrov (*Reutlingen University*), Carsten Binnig (*Technical University of Darmstadt*)
- **Black or White? How to Develop an AutoTuner for Memory-based Analytics** ..... 1667  
Mayuresh Kunjir (*Duke University*), Shivnath Babu (*Unravel Data Systems*)

## Research 19: Machine Learning Systems and Applications

- **Vista: Optimized System for Declarative Feature Transfer from Deep CNNs at Scale** ..... 1685  
Supun Nakandala, Arun Kumar (*University of California, San Diego*)
- **Optimizing Machine Learning Workloads in Collaborative Environments** ..... 1701  
Behrouz Derakhshan, Alireza Rezaei Mahdiraji (*DFKI GmbH*), Ziawasch Abedjan (*TU Berlin*), Tilmann Rabl (*Hasso Plattner Institute & University of Potsdam*), Volker Markl (*DFKI GmbH & TU Berlin*)
- **GOGGLES: Automatic Image Labeling with Affinity Coding**..... 1717  
Nilaksh Das, Sanya Chaba, Renzhi Wu, Sakshi Gandhi, Duen Horng Chau, Xu Chu (*Georgia Institute of Technology*)

- **DeepSqueeze: Deep Semantic Compression for Tabular Data** ..... 1733  
Amir Ilkhechi, Andrew Crotty, Alex Galakatos, Yicong Mao, Grace Fan, Xiran Shi,  
Ugur Cetintemel (*Brown University*)
- **TRACER: A Framework for Facilitating Accurate and Interpretable Analytics for High Stakes Applications** ..... 1747  
Kaiping Zheng, Shaofeng Cai (*National University of Singapore*),  
Horng Ruey Chua (*National University Health System*), Wei Wang (*National University of Singapore*),  
Kee Yuan Ngiam (*National University Health System*), Beng Chin Ooi (*National University of Singapore*)

## Research 20: Graph Data Management and Analysis

- **Application Driven Graph Partitioning** ..... 1765  
Wenfei Fan (*University of Edinburgh, Beihang University & Shenzhen University*), Ruochun Jin (*University of Edinburgh*), Ping Lu (*BDBC, Beihang University*), Xiaojian Luo (*Alibaba Group*), Ruiqi Xu (*University of Edinburgh*),  
Qiang Yin, Wenyan Yu, Jingren Zhou (*Alibaba Group*)
- **Progressive Top-K Nearest Neighbors Search in Large Road Networks** ..... 1781  
Dian Ouyang (*The University of Sydney*), Dong Wen, Lu Qin (*University of Technology Sydney*),  
Lijun Chang (*The University of Sydney*), Ying Zhang (*University of Technology Sydney*),  
Xuemin Lin (*The University of New South Wales*)
- **Memory-Aware Framework for Efficient Second-Order Random Walk on Large Graphs** ... 1797  
Yingxia Shao (*Beijing University of Posts and Telecommunications*),  
Shiyue Huang, Xupeng Miao, Bin Cui (*Peking University*),  
Lei Chen (*Hong Kong University of Science and Technology*)
- **Hub Labeling for Shortest Path Counting** ..... 1813  
Yikai Zhang, Jeffrey Xu Yu (*Chinese University of Hong Kong*)
- **CHASSIS: Conformity Meets Online Information Diffusion** ..... 1829  
Hui Li (*Nanyang Technological University*), Hui Li (*Xidian University*),  
Sourav S. Bhowmick (*Nanyang Technological University*)

## Research 21: Spatial, Temporal, and Multimedia Data I

- **Architecture-Intact Oracle for Fastest Path and Time Queries on Dynamic Spatial Networks** ..... 1841  
Victor Junqiu Wei (*Noah's Ark Lab, Huawei Technologies*),  
Raymond Chi-Wing Wong (*The Hong Kong University of Science and Technology*),  
Cheng Long (*Nanyang Technological University*)
- **Data Series Progressive Similarity Search with Probabilistic Quality Guarantees** ..... 1857  
Anna Gogolou (*Université Paris-Saclay, Inria, CNRS, LRI & LIPADE, University of Paris*),  
Theophanis Tsandilas (*Université Paris-Saclay, Inria, CNRS, LRI*),  
Karima Echihabi (*IRDA, Rabat IT Center, ENSIAS, Mohammed V University*),  
Anastasia Bezerianos (*Université Paris-Saclay, CNRS, Inria, LRI*),  
Themis Palpanas (*LIPADE, University of Paris & French University Institute (IUF)*)
- **A GPU-friendly Geometric Data Model and Algebra for Spatial Queries** ..... 1875  
Harish Doraiswamy, Juliana Freire (*New York University*)
- **Debunking Four Long-Standing Misconceptions of Time-Series Distance Measures** ..... 1887  
John Paparrizos, Chunwei Liu, Aaron J. Elmore, Michael J. Franklin (*University of Chicago*)
- **MIRIS: Fast Object Track Queries in Video** ..... 1907  
Fayyen Bastani, Songtao He, Arjun Balasingam, Karthik Gopalakrishnan, Mohammad Alizadeh, Hari  
Balakrishnan, Michael Cafarella, Tim Kraska, Sam Madden (*Massachusetts Institute of Technology*)

## Award Talks

- **ACM SIGMOD Jim Gray Dissertation Award W Talk** ..... 1923  
Jose M. Faleiro (*Microsoft Research*)
- **Effective Data Versioning for Collaborative Data Analytics** ..... 1925  
Silu Huang (*Microsoft Research*)

## Research 22: Data Lakes, Web, and Knowledge Graph

- **Organizing Data Lakes for Navigation** ..... 1939  
Fatemeh Nargesian (*University of Rochester*), Ken Q. Pu (*University of Ontario Institute of Technology*),  
Erkang Zhu (*Microsoft Research*), Bahar Ghadiri Bashardoost (*University of Toronto*),  
Renée J. Miller (*Northeastern University*)
- **Finding Related Tables in Data Lakes for Interactive Data Science** ..... 1951  
Yi Zhang, Zachary G. Ives (*University of Pennsylvania*)
- **Web Data Extraction using Hybrid Program Synthesis: A Combination of Top-down and Bottom-up Inference** ..... 1967  
Mohammad Raza, Sumit Gulwani (*Microsoft Corporation*)
- **SPARQL Rewriting: Towards Desired Results** ..... 1979  
Xun Jian (*The Hong Kong University of Science and Technology*),  
Yue Wang (*Shenzhen Institute of Computing Sciences, Shenzhen University*),  
Xiayu Lei, Libin Zheng, Lei Chen (*The Hong Kong University of Science and Technology*)
- **Realistic Re-evaluation of Knowledge Graph Completion Methods: An Experimental Study** ..... 1995  
Farahnaz Akrami, Mohammed Samiul Saeef (*University of Texas at Arlington*),  
Qingheng Zhang, Wei Hu (*Nanjing University*), Chengkai Li (*University of Texas at Arlington*)

## Research 23: OLAP, Data Warehouses, and Key-Value Stores

- **Bitvector-aware Query Optimization for Decision Support Queries** ..... 2011  
Bailu Ding, Surajit Chaudhuri, Vivek Narasayya (*Microsoft Research*)
- **Efficient Join Synopsis Maintenance for Data Warehouse** ..... 2027  
Zhuoyue Zhao, Feifei Li, Yuxi Liu (*University of Utah*)
- **Adaptive HTAP through Elastic Resource Scheduling** ..... 2043  
Aunn Raza, Periklis Chrysogelos (*Ecole Polytechnique Fédérale de Lausanne*),  
Angelos Christos Anadiotis (*Ecole Polytechnique*),  
Anastasia Ailamaki (*Ecole Polytechnique Fédérale de Lausanne*)
- **SPRINTER: A Fast  $n$ -ary Join Query Processing Method for Complex OLAP Queries** ..... 2055  
Yoon-Min Nam Nam, Donghyoung Han Han (*Daegu Gyeongbuk Institute of Science and Technology*),  
Min-Soo Kim Kim (*Korea Advanced Institute of Science and Technology*)
- **Rosetta: A Robust Space-Time Optimized Range Filter for Key-Value Stores** ..... 2071  
Siqiang Luo, Subarna Chatterjee, Rafael Ketsetsidis, Niv Dayan, Wilson Qin,  
Stratos Idreos (*Harvard University*)

## Research 24: Spatial, Temporal, and Multimedia Data II

- **RID: Deduplicating Snapshot Computations** ..... 2087  
Nikos Tsikoudis, Liuba Shrira (*Brandeis University*)
- **Architecting a Query Compiler for Spatial Workloads** ..... 2103  
Ruby Y. Tahboub, Tiark Rompf (*Purdue University*)
- **LISA: A Learned Index Structure for Spatial Data** ..... 2119  
Pengfei Li (*Zhejiang University*), Hua Lu (*Roskilde University*), Qian Zheng (*Nanyang Technological University*),  
Long Yang, Gang Pan (*Zhejiang University*)
- **Effective Travel Time Estimation: When Historical Trajectories over Road Networks Matter** ..... 2135  
Haitao Yuan, Guoliang Li (*Tsinghua University*), Zhifeng Bao (*RMIT University*),  
Ling Feng (*Tsinghua University*)

## Research 25: Social Network Analysis

- **The Solution Distribution of Influence Maximization: A High-level Experimental Study on Three Algorithmic Approaches** ..... 2151  
Naoto Ohsaka (*NEC Corporation*)

- **Influence Maximization Revisited: Efficient Reverse Reachable Set Generation with Bound Tightened** ..... 2167  
Qintian Guo, Sibao Wang (*The Chinese University of Hong Kong*),  
Zhewei Wei, Ming Chen (*Renmin University of China*)
- **Truss-based Community Search over Large Directed Graphs** ..... 2183  
Qing Liu (*Hong Kong Baptist University*), Minjun Zhao (*Zhejiang University*),  
Xin Huang, Jianliang Xu (*Hong Kong Baptist University*), Yunjun Gao (*Zhejiang University*)
- **Densely Connected User Community and Location Cluster Search in Location-Based Social Networks** ..... 2199  
Junghoon Kim (*Nanyang Technological University*), Tao Guo (*Google*),  
Kaiyu Feng, Gao Cong, Arijit Khan (*Nanyang Technological University*),  
Farhana M. Choudhury (*University of Melbourne*)
- **Global Reinforcement of Social Networks: The Anchored Coreness Problem** ..... 2211  
Qingyuan Linghu, Fan Zhang (*Guangzhou University*),  
Xuemin Lin, Wenjie Zhang (*University of New South Wales*), Ying Zhang (*University of Technology Sydney*)

## Industry 4: Advanced Functionality

- **Confidentiality Support over Financial Grade Consortium Blockchain** ..... 2227  
Ying Yan, Changzheng Wei, Xuepeng Guo, Xuming Lu, Xiaofu Zheng, Qi Liu, Chenhui Zhou, Xuyang Song,  
Boran Zhao, Hui Zhang, Guofei Jiang (*Ant Financial Services Group*)
- **PASE: PostgreSQL Ultra-High-Dimensional Approximate Nearest Neighbor Search Extension** ..... 2241  
Wen Yang, Tao Li, Gai Fang, Hong Wei (*Ant Financial Services Group*)
- **Making Search Engines Faster by Lowering the Cost of Querying Business Rules Through FPGAs** ..... 2255  
Fabio Maschi, Muhsen Owaida, Gustavo Alonso (*ETH Zurich*),  
Matteo Casalino, Anthony Hock-Koon (*Amadeus*)
- **Spur: Mitigating Slow Instances in Large-Scale Streaming Pipelines** ..... 2271  
Ke Wang (*Carnegie Mellon University*), Avriella Floratou, Ashvin Agrawal (*Microsoft*),  
Daniel Musgrave (*Netflix*)
- **Entity Matching in the Wild: A Consistent and Versatile Framework to Unify Data in Industrial Applications** ..... 2287  
Yan Yan, Stephen Meyles, Aria Haghighi (*Amperity, Inc.*), Dan Suciu (*University of Washington*)

## Research 26: Usability and Natural Language User Interfaces

- **QueryVis: Logic-based Diagrams help Users Understand Complicated SQL Queries Faster** ..... 2303  
Aristotelis Leventidis, Jiahui Zhang, Cody Dunne, Wolfgang Gatterbauer (*Northeastern University*),  
H.V. Jagadish (*University of Michigan*), Mirek Riedewald (*Northeastern University*)
- **Duoquest: A Dual-Specification System for Expressive SQL Queries** ..... 2319  
Christopher Baik, Zhongjun Jin, Michael Cafarella, H. V. Jagadish (*University of Michigan – Ann Arbor*)
- **SQLCheck: Automated Detection and Diagnosis of SQL Anti-Patterns** ..... 2331  
Prashanth Dintyala, Arpit Narechania, Joy Arulraj (*Georgia Institute of Technology*)
- **DBPal: A Fully Pluggable NL2SQL Training Pipeline** ..... 2347  
Nathaniel Weir (*Johns Hopkins University*), Prasetya Utama (*Technische Universität Darmstadt*),  
Alex Galakatos, Andrew Crotty, Amir Ilkhechi, Shekar Ramaswamy, Rohin Bhushan (*Brown University*),  
Nadja Geisler, Benjamin Härtel, Steffen Eger (*Technische Universität Darmstadt*),  
Ugur Cetintemel (*Brown University*), Carsten Binnig (*Technische Universität Darmstadt*)
- **SpeakQL: Towards Speech-driven Multimodal Querying of Structured Data** ..... 2363  
Vraj Shah, Side Li, Arun Kumar, Lawrence Saul (*University of California, San Diego*)

## Research 27: Distributed and Parallel Processing

- **Near-Optimal Distributed Band-Joins through Recursive Partitioning** ..... 2375  
Rundong Li (*Google*), Wolfgang Gatterbauer, Mirek Riedewald (*Northeastern University*)
- **ChronoCache: Predictive and Adaptive Mid-Tier Query Result Caching** ..... 2391  
Brad Glasbergen, Kyle Langendoen, Michael Abebe, Khuzaima Daudjee (*University of Waterloo*)
- **Cheetah: Accelerating Database Queries with Switch Pruning** ..... 2407  
Muhammad Tirmazi, Ran Ben Basat, Jiaqi Gao, Minlan Yu (*Harvard University*)
- **External Merge Sort for Top-K Queries: Eager input filtering guided by histograms** ..... 2423  
Yannis Chronis (*University of Wisconsin-Madison*), Thanh Do, Goetz Graefe, Keith Peters (*Google Inc*)
- **Automating Incremental and Asynchronous Evaluation for Recursive Aggregate Data Processing** ..... 2439  
Qiang Wang, Yanfeng Zhang (*Northeastern University*), Hao Wang (*Ohio State University*),  
Liang Geng (*Northeastern University*), Rubao Lee, Xiaodong Zhang (*Ohio State University*),  
Ge Yu (*Northeastern University*)

## Research 28: Stream Processing

- **Prompt: Dynamic Data-Partitioning for Distributed Micro-batch Stream Processing Systems** ..... 2455  
Ahmed S. Abdelhamid, Ahmed R. Mahmood, Anas Daghistani, Walid G. Aref (*Purdue University*)
- **Rhino: Efficient Management of Very Large Distributed State for Stream Processing Engines** ..... 2471  
Bonaventura Del Monte, Steffen Zeuch (*Technische Universität Berlin & DFKI GmbH*),  
Tilman Rabl (*Hasso Plattner Institute, University of Potsdam*),  
Volker Markl (*Technische Universität Berlin & DFKI GmbH*)
- **Grizzly: Efficient Stream Processing Through Adaptive Query Compilation** ..... 2487  
Philipp M. Grulich, Breß Sebastian (*Technische Universität Berlin*),  
Steffen Zeuch (*Technische Universität Berlin & DFKI GmbH*),  
Jonas Traub, Janis von Bleichert (*Technische Universität Berlin*), Zongxiong Chen (*DFKI GmbH*),  
Tilman Rabl (*HPI, University of Potsdam*), Volker Markl (*Technische Universität Berlin & DFKI GmbH*)
- **LightSaber: Efficient Window Aggregation on Multi-core Processors** ..... 2505  
Georgios Theodorakis (*Imperial College London*), Alexandros Koliosis (*Graphcore Research*),  
Peter Pietzuch, Holger Pirk (*Imperial College London*)
- **Parallel Index-based Stream Join on a Multicore CPU** ..... 2523  
Amirhesam Shahvarani, Hans-Arno Jacobsen (*Technische Universität München*)

## Research 29: Data Mining and Similarity Search

- **Improving Approximate Nearest Neighbor Search through Learned Adaptive Early Termination** ..... 2539  
Conglong Li (*Carnegie Mellon University*), Minjia Zhang (*Microsoft AI and Research*),  
David G. Andersen (*Carnegie Mellon University*), Yuxiong He (*Microsoft AI and Research*)
- **Theoretically-Efficient and Practical Parallel DBSCAN** ..... 2555  
Yiqiu Wang (*Massachusetts Institute of Technology*), Yan Gu (*University of California, Riverside*),  
Julian Shun (*Massachusetts Institute of Technology*)
- **A Relational Matrix Algebra and its Implementation in a Column Store** ..... 2573  
Oksana Dolmatova (*University of Zürich*), Nikolaus Augsten (*University of Salzburg*),  
Michael H. Böhlen (*University of Zürich*)
- **Locality-Sensitive Hashing Scheme based on Longest Circular Co-Substring** ..... 2589  
Yifan Lei, Qiang Huang, Mohan Kankanhalli, Anthony K. H. Tung (*National University of Singapore*)
- **Continuously Adaptive Similarity Search** ..... 2601  
Huayi Zhang (*Worcester Polytechnic Institute*), Lei Cao (*Massachusetts Institute of Technology*),  
Yizhou Yan (*Worcester Polytechnic Institute*),  
Samuel Madden (*Massachusetts Institute of Technology*), Elke A. Rundensteiner (*Worcester Polytechnic Institute*)



## Tutorials

- **Automating Exploratory Data Analysis via Machine Learning: An Overview** ..... 2617  
Tova Milo, Amit Somech (*Tel Aviv University*)
- **Crowdsourcing Practice for Efficient Data Labeling: Aggregation, Incremental Relabeling, and Pricing** ..... 2623  
Alexey Drutsa, Dmitry Ustalov, Evfrosiniya Zermirnova, Valentina Fedorova, Olga Megorskaya, Daria Baidakova (*Yandex*)
- **State of the Art and Open Challenges in Natural Language Interfaces to Data** ..... 2629  
Fatma Özcan, Abdul Quamar, Jaydeep Sen, Chuan Lei, Vasilis Efthymiou (*IBM Research - Almaden*)
- **SIGMOD 2020 Tutorial on Fairness and Bias in Peer Review and Other Sociotechnical Intelligent Systems** ..... 2637  
Nihar B. Shah, Zachary Lipton (*Carnegie Mellon University*)
- **Le Taureau: Deconstructing the Serverless Landscape & A Look Forward** ..... 2641  
Anurag Khandelwal (*Yale University*), Arun Kejariwal (*Facebook Inc.*), Karthikeyan Ramasamy (*Splunk Inc.*)
- **Beyond Analytics: The Evolution of Stream Processing Systems** ..... 2651  
Paris Carbone (*RISE - Research Institutes of Sweden*), Marios Fragkoulis (*Delft University of Technology*), Vasiliki Kalavri (*Boston University*), Asterios Katsifodimos (*Delft University of Technology*)
- **Optimal Join Algorithms Meet Top- $k$**  ..... 2659  
Nikolaos Tziavelis, Wolfgang Gatterbauer, Mirek Riedewald (*Northeastern University*)
- **Key-Value Storage Engines** ..... 2667  
Stratos Idreos (*Harvard University*), Mark Callaghan (*MongoDB*)

## Demonstrations

- **RASQL: A Powerful Language and its System for Big Data Applications** ..... 2673  
Jin Wang, Guorui Xiao, Jiaqi Gu (*University of California, Los Angeles*), Jiacheng Wu (*Tsinghua University*), Carlo Zaniolo (*University of California, Los Angeles*)
- **PL/SQL Without the PL** ..... 2677  
Denis Hirn, Torsten Grust (*University of Tübingen*)
- **Analysis of Database Search Systems with THOR** ..... 2681  
Theofilos Belpas, Orest Gkini, Georgia Koutrika (*Athena Research Center*)
- **BOOMER: A Tool for Blending Visual P-Homomorphic Queries on Large Networks** ..... 2685  
Yinglong Song (*Nanyang Technological University & Fudan University*), Huey Eng Chua, Sourav S. Bhowmick (*Nanyang Technological University*), Byron Choi (*Hong Kong Baptist University*), Shuigeng Zhou (*Fudan University*)
- **AURORA: Data-driven Construction of Visual Graph Query Interfaces for Graph Databases** ..... 2689  
Sourav S. Bhowmick (*Nanyang Technological University*), Kai Huang (*Nanyang Technological University & Fudan University*), Huey Eng Chua (*Nanyang Technological University*), Zifeng Yuan (*Nanyang Technological University & Fudan University*), Byron Choi (*Hong Kong Baptist University*), Shuigeng Zhou (*Fudan University*)
- **vChain: A Blockchain System Ensuring Query Integrity** ..... 2693  
Haixin Wang, Cheng Xu, Ce Zhang, Jianliang Xu (*Hong Kong Baptist University*)
- **AUDITOR: A System Designed for Automatic Discovery of Complex Integrity Constraints in Relational Databases** ..... 2697  
Wentao Hu, Dongxiang Zhang, Dawei Jiang, Sai Wu, Ke Chen (*Zhejiang University*), Kian-Lee Tan (*School of Computing National University of Singapore*), Gang Chen (*Zhejiang University*)
- **SHARQL: Shape Analysis of Recursive SPARQL Queries** ..... 2701  
Angela Bonifati (*Lyon 1 University*), Wim Martens, Thomas Timm (*University of Bayreuth*)
- **High Performance Distributed OLAP on Property Graphs with Grasper** ..... 2705  
Hongzhi Chen, Bowen Wu, Shiyuan Deng, Chenghuan Huang, Changji Li, Yichao Li, James Cheng (*The Chinese University of Hong Kong*)

- **ProcAnalyzer: Effective Code Analyzer for Tuning Imperative Programs in SAP HANA**..... 2709  
Kisung Park, Taeyoung Jeong, Chanho Jeong, Jaeha Lee, Dong-Hun Lee (*SAP Labs Korea*),  
Young-Koo Lee (*Kyung Hee University*)
- **LATTE: Visual Construction of Smart Contracts**..... 2713  
Sean Tan, Sourav S. Bhowmick, Huey Eng Chua (*Nanyang Technological University*),  
Xiaokui Xiao (*National University of Singapore*)
- **PROUD: PaRallel OUTlier Detection for Streams**..... 2717  
Theodoros Toliopoulos, Christos Bellas, Anastasios Gounaris,  
Apostolos Papadopoulos (*Aristotle University of Thessaloniki*)
- **MithraCoverage: A System for Investigating Population Bias for Intersectional Fairness** .... 2721  
Zhongjun Jin, Mengjing Xu, Chenkai Sun (*University of Michigan – Ann Arbor*),  
Abolfazl Asudeh (*University of Illinois at Chicago*), H. V. Jagadish (*University of Michigan – Ann Arbor*)
- **MC3: A System for Minimization of Classifier Construction Cost**..... 2725  
Shay Gershtein, Tova Milo, Gefen Morami (*Tel Aviv University*), Slava Novgorodov (*eBay Research*)
- **Sentinel: Understanding Data Systems** ..... 2729  
Brad Glasbergen, Michael Abebe, Khuzaima Daudjee, Daniel Vogel, Jian Zhao (*University of Waterloo*)
- **BugDoc: A System for Debugging Computational Pipelines** ..... 2733  
Raoni Lourenço, Juliana Freire, Dennis Shasha (*New York University*)
- **TQVS: Temporal Queries over Video Streams in Action** ..... 2737  
Yueting Chen, Xiaohui Yu (*York University*), Nick Koudas (*University of Toronto*)
- **ExTuNe: Explaining Tuple Non-conformance** ..... 2741  
Anna Fariha (*University of Massachusetts, Amherst*), Ashish Tiwari, Arjun Radhakrishna, Sumit Gulwani  
(*Microsoft*)
- **Interactively Discovering and Ranking Desired Tuples without Writing SQL Queries** ..... 2745  
Xuedi Qin, Chengliang Chai, Yuyu Luo (*Tsinghua University*), Nan Tang (*QCRI, HBKU*),  
Guoliang Li (*Tsinghua University*)
- **Synner: Generating Realistic Synthetic Data** ..... 2749  
Miro Mannino, Azza Abouzied (*New York University Abu Dhabi*)
- **InCognitoMatch: Cognitive-aware Matching via Crowdsourcing** ..... 2753  
Roei Shraga, Coral Scharf, Rakefet Ackerman, Avigdor Gal (*Technion – Israel Institute of Technology*)
- **CoClean: Collaborative Data Cleaning**..... 2757  
Mashaal Musleh (*University of Minnesota*), Mourad Ouzzani, Nan Tang (*QCRI, HBKU*),  
AnHai Doan (*University of Wisconsin*)
- **STAR: A Distributed Stream Warehouse System for Spatial Data**..... 2761  
Zhida Chen, Gao Cong (*Nanyang Technological University*), Walid G. Aref (*Purdue University*)
- **T-REx: Table Repair Explanations** ..... 2765  
Daniel Deutch, Nave Frost, Amir Gilad, Oren Sheffer (*Tel Aviv University*)
- **SVQ++: Querying for Object Interactions in Video Streams** ..... 2769  
Daren Chao, Nick Koudas, Ioannis Xarchakos (*University of Toronto*)
- **F-IVM: Learning over Fast-Evolving Relational Data** ..... 2773  
Milos Nikolic (*University of Edinburgh*), Haozhe Zhang, Ahmet Kara, Dan Olteanu (*University of Oxford*)
- **CoMing: A Real-time Co-Movement Mining System for Streaming Trajectories**..... 2777  
Ziquan Fang, Yunjun Gao, Lu Pan (*Zhejiang University*), Lu Chen (*Aalborg University*),  
Xiaoye Miao (*Zhejiang University*), Christian S. Jensen (*Aalborg University*)
- **Unified Spatial Analytics from Heterogeneous Sources with Amazon Redshift**..... 2781  
Nemanja Borić, Hinnerk Gildhoff, Menelaos Karavelas, Ippokratis Pandis,  
Ioanna Tsalouchidou (*Amazon Web Services*)
- **Big Data Series Analytics Using TARDIS and its Exploitation in Geospatial Applications** .... 2785  
Liang Zhang, Noura Alghamdi, Mohamed Y. Eltabakh, Elke A. Rundensteiner (*Worcester Polytechnic Institute*)
- **CDFShop: Exploring and Optimizing Learned Index Structures** ..... 2789  
Ryan Marcus, Emily Zhang, Tim Kraska (*Massachusetts Institute of Technology*)

• <b>TensorFlow Data Validation: Data Analysis and Validation in Continuous ML Pipelines</b> .....	2793
Emily Caveness, Paul Suganthan G. C., Zhuo Peng, Neoklis Polyzotis, Sudip Roy, Martin Zinkevich ( <i>Google Inc.</i> )	
• <b>Grosbeak: A Data Warehouse Supporting Resource-Aware Incremental Computing</b> .....	2797
Zuozhi Wang ( <i>University of California, Irvine</i> ), Kai Zeng, Botong Huang, Wei Chen, Xiaozong Cui, Bo Wang, Ji Liu, Liya Fan, Dachuan Qu, Zhenyu Hou, Tao Guan ( <i>Alibaba Group</i> ), Chen Li ( <i>University of California, Irvine</i> ), Jingren Zhou ( <i>Alibaba Group</i> )	
• <b>Demonstration of BitGourmet: Data Analysis via Deterministic Approximation</b> .....	2801
Saehan Jo, Immanuel Trummer ( <i>Cornell University</i> )	
• <b>Bring Your Own Data to X-PLAIN</b> .....	2805
Eliana Pastor, Elena Baralis ( <i>Politecnico di Torino</i> )	
• <b>Physical Visualization Design</b> .....	2809
Lana Ramjit, Zhaoning Kong, Ravi Netravali ( <i>University of California, Los Angeles</i> ), Eugene Wu ( <i>Columbia University</i> )	
• <b>Demonstration of Chestnut: An In-memory Data Layout Designer for Database Applications</b> .....	2813
Mingwei Samuel ( <i>University of California, Berkeley</i> ), Cong Yan ( <i>University of Washington</i> ), Alvin Cheung ( <i>University of California, Berkeley</i> )	

## Student Abstracts

• <b>Breaking Down Memory Walls in LSM-based Storage Systems</b> .....	2817
Chen Luo ( <i>University of California, Irvine</i> )	
• <b>Context-Free Path Querying via Matrix Equations</b> .....	2821
Yuliya Susanina ( <i>Saint Petersburg State University</i> )	
• <b>Simulation-based Approximate Graph Pattern Matching</b> .....	2825
Xiaoshuang Chen ( <i>The University of New South Wales</i> )	
• <b>High-Dimensional Vector Similarity Search: From Time Series to Deep Network Embeddings</b> .....	2829
Karima Echihabi ( <i>Mohammed V University</i> )	
• <b>Rethinking Message Brokers on RDMA and NVM</b> .....	2833
Hendrik Makait ( <i>Technische Universität Berlin</i> )	
• <b>Monte Carlo Tree Search for Generating Interactive Data Analysis Interfaces</b> .....	2837
Yiru Chen ( <i>Columbia University</i> )	
• <b>Continuous Prefetch for Interactive Data Applications</b> .....	2841
Haneen Mohammed ( <i>Columbia University</i> )	
• <b>Re-evaluating the Performance Trade-offs for Hash-Based Multi-Join Queries</b> .....	2845
Shiva Jahangiri ( <i>University of California, Irvine</i> )	
• <b>Interactive View Recommendation</b> .....	2849
Xiaozhong Zhang ( <i>University of Pittsburgh</i> )	
• <b>From Worst-Case to Average-Case Analysis: Accurate Latency Predictions for Key-Value Storage Engines</b> .....	2853
Meena Jagadeesan, Garrett Tanzer ( <i>Harvard University</i> )	
• <b>Towards the Scheduling of Vertex-constrained Multi Subgraph Matching Query</b> .....	2857
Kongzhang Hao ( <i>University of New South Wales</i> ), Longbin Lai ( <i>Alibaba Group</i> )	
• <b>Serverless Query Processing on a Budget</b> .....	2861
William Ma ( <i>University of California, Berkeley</i> )	
• <b>Workload-Aware Column Imprints</b> .....	2865
Noah Slavitch ( <i>University of Waterloo</i> )	
• <b>Towards Scalable UDTFs in Noria</b> .....	2869
Justus Adam ( <i>Technische Universität Dresden</i> )	

• <b>Column Partition and Permutation for Run Length Encoding in Columnar Databases .....</b>	2873
Jia Shi ( <i>University of Waterloo</i> )	
• <b>Supporting Database Constraints in Synthetic Data Generation based on Generative Adversarial Networks .....</b>	2875
Wanxin Li ( <i>University of Waterloo</i> )	
• <b>An Evaluation of Methods of Compressing Doubles .....</b>	2879
Jacob Spiegel ( <i>University of Chicago</i> )	
• <b>MemFlow: Memory-Aware Distributed Deep Learning.....</b>	2883
Neil Band ( <i>Harvard University</i> )	
• <b>JSON Schema Matching: Empirical Observations .....</b>	2887
Kunal Waghay ( <i>University of Wisconsin – Madison</i> )	
<b>Author Index .....</b>	2890

# SIGMOD 2020 Organization

**General Chairs:** David Maier (Portland State University, USA)  
Rachel Pottinger (University of British Columbia, Canada)

**Program Chairs:** AnHai Doan (University of Wisconsin, USA)  
Wang-Chiew Tan (Megagon Labs, USA)

**Proceedings Chairs:** Abdussalam Alawini (University of Illinois at Urbana-Champaign, USA)  
Hung Q. Ngo (RelationalAI, USA)

**Tutorial Chairs:** Bettina Kemme (McGill University, Canada)  
Felix Naumann (Hasso Plattner Institute, Germany)

**Workshop Chairs:** Ihab Ilyas (University of Waterloo, Canada)  
Angela Bonifati (Lyon 1 University, France)  
Aditya Parameswaran (University of California, Berkeley)

**Student Research Competition Chairs:** Xi He (University of Waterloo, Canada)  
Eugene Wu (Columbia University, USA)

**New Researcher Symposium Chairs:** Spyros Blanas (The Ohio State University, USA)  
Azza Abouzzied (NYU Abu Dhabi, UAE)

**SIGMOD Programming Contest** Donatella Firmani (Roma Tre University, Italy)  
**Chairs:** Andrea De Angelis (Roma Tre University, Italy)  
Maurizio Mazzei (Roma Tre University, Italy)  
Federico Piai (Roma Tre University, Italy)

**Community Initiatives Chair:** Eduard Dragut (Temple University, USA)

**Local Organization Chair:** Kristin Tufte (Portland State University, USA)

**SIGMOD Local Arrangement Vice-Chairs:** Mohamed Ali (UW Tacoma, USA)  
Min Chen (UW Bothell, USA)  
Dejing Dou (University of Oregon, USA)

**Sponsorship Chairs:** Tilmann Rabl (HPI, University of Potsdam, Germany)  
Bill Howe (University of Washington, USA)

**Mentorship Chairs:** Qiong Luo (HKUST, China)  
Tiaanzheng Wang (Simon Fraser University, Canada)

**Finance Chair:** Jennie Rogers (Northwestern University, USA)

**Registration Chair:** Shawn Bowers (Gonzaga University, USA)

**Web/Information Chair:** Wendy Hui Wang (Stevens Institute of Technology, USA)

**Publicity Chairs:** Leilani Battle (University of Maryland, USA)  
Lukasz Golab (University of Waterloo, Canada)

**Exhibit Chair:** Rafael Fernandez-Moctezuma (Google Inc., USA)

**Demonstration and Workshop Local**

**Arrangements Chair:** Arash Termehchy (Oregon State University, USA)

**Conference Logistics:** John Lateulere (Integrated Management Solutions,  
Conference Management, USA)

**Communication Chair:** Mostafa Milani (University of British-Columbia, Canada)

**Program Committee Area Chairs:** Aditya G Parameswaran (UC Berkeley, USA)  
Andrew Pavlo (Carnegie Mellon University, USA)  
Anthony Tung (National University of Singapore, Singapore)  
Evimaria Terzi (Boston University, USA)  
Gerome Miklau (University of Massachusetts Amherst, USA)  
Jeff Naughton (Google, USA)  
Jeffrey Xu Yu (Chinese University of Hong Kong, China)  
Jens Dittrich (Saarland University, Germany)  
Johannes Gehrke (Microsoft, USA)  
Kian-Lee Tan (National University of Singapore, Singapore)  
Neoklis Polyzotis (Google, USA)  
Pinar Tozun (ITU, Denmark)  
Prithviraj Sen (IBM Almaden Research Center, USA)  
Renée J. Miller (Northeastern University, USA)  
Tamer Ozsu (University of Waterloo, Canada)  
Tova Milo (Tel Aviv University, Israel)  
Yunyao Li (IBM Almaden Research Center, USA)

**Program Committee:** Aaron J Elmore (University of Chicago, USA)  
Abdussalam Alawini (University of Illinois at Urbana-Champaign, USA)  
Alan Fekete (University of Sydney, Australia)  
Alex Beutel (Google, USA)  
Alexander Boehm (SAP SE, Austria)  
Alexandra Meliou (University of Massachusetts Amherst, USA)  
Alexandros Labrinidis (University of Pittsburgh, USA)  
Alin Deutsch (UCSD, USA)  
Alvin Cheung (UC Berkeley, USA)  
Anja Gruenheid (Google Inc., USA)  
Antonios Deligiannakis (Technical University of Crete, Greece)  
Arijit Khan (Nanyang Technological University, Singapore)  
Aristides Gionis (Aalto University, Finland)  
Arun Kumar (University of California, San Diego, USA)

**Program Committee (continued):** Arvind Arasu (Microsoft, USA)  
Ashraf Aboulnaga (Qatar Computing Research Institute, Qatar)  
Assaf Schuster (Technion, Israel)  
Avigdor Gal (Technion, Israel)  
Avrilia Floratou (Microsoft, USA)  
Babak Salimi (University of Washington, USA)  
Badrish Chandramouli (Microsoft, USA)  
Bailu Ding (Microsoft Research, USA)  
Beng Chin Ooi (NUS, Singapore)  
Bill Howe (University of Washington, USA)  
Bin Cui (Peking University, China)  
Bingsheng He (National University of Singapore, Singapore)  
Bolin Ding (Data Analytics and Intelligence Lab, Alibaba Group, USA)  
Boon Thau Loo (University of Pennsylvania, USA)  
Boris Glavic (Illinois Institute of Technology, USA)  
Byron Choi (Hong Kong Baptist University, SAR China)  
Çağatay Demiralp (Megagon Labs, USA)  
Carsten Binnig (TU Darmstadt, Germany)  
Ce Zhang (ETH, Switzerland)  
Chen Chen (Megagon Labs, USA)  
Chen Li (UC Irvine, USA)  
Chengkai Li (The University of Texas at Arlington, USA)  
Chris De Sa (Cornell, USA)  
Chris Jermaine (Rice University, USA)  
Christoph Koch (EPFL, Switzerland)  
Chuan Lei (IBM Research - Almaden, USA)  
Cong Yu (Google, USA)  
Da Yan (University of Alabama at Birmingham, USA)  
Daisy Zhe Wang (University of Florida, USA)  
Daniel Kifer (Pennsylvania State University, USA)  
Danica Porobic (Oracle, USA)  
Daniel Deutch (Tel Aviv University, Israel)  
Dimitris Papadias (HKUST, China)  
Doug Burdick (IBM Research, USA)  
Eduard Dragut (Temple Univ., USA)  
Erhard Rahm (University of Leipzig, Germany)  
Faisal Nawab (UC Santa Cruz, USA)  
Feifei Li (University of Utah, USA)  
Felix Schuhknecht (Saarland University, Germany)  
Flip Korn (Google, USA)  
Florin Rusu (UC Merced, USA)  
Gabriela Jacques-Silva (Facebook, USA)  
Gao Cong (Nanyang Technological University, Singapore)

**Program Committee (continued):** Gautam Das (University of Texas at Arlington, USA)  
Georgios Fakas (Uppsala University, Sweden)  
George Kollios (Boston University, USA)  
Gerhard Weikum (Max-Planck-Institut für Informatik, Germany)  
Glenn M Fung (American Family Insurance, USA)  
Goetz Graefe (Google, USA)  
Goncalo Simoes (Google, USA)  
Graham Cormode (University of Warwick, UK)  
Guoliang Li (Tsinghua University, China )  
Haixun Wang (WeWork Research, USA)  
Han Li (Amazon, USA)  
Holger Pirk (Imperial College, UK)  
Huanchen Zhang (Carnegie Mellon University, USA)  
Immanuel Trummer (Cornell, USA)  
Ippokratis Pandis (Amazon, USA)  
James Cheng (CUHK, Hong Kong)  
Jana Giceva (Imperial College London, UK)  
Jennie Rogers (Northwestern University, UK)  
Jens Teubner (TU Dortmund University, Germany)  
Jian Pei (Simon Fraser University, Canada)  
Jiannan Wang (Simon Fraser University, Canada)  
Jignesh Patel (UW - Madison, USA)  
Jinfeng Li (Megagon Labs, USA)  
Jose M Faleiro (UC Berkeley, USA)  
Joy Arulraj (Georgia Tech, USA)  
Ju Fan (Renmin University of China, China)  
Jun Yang (Duke University, USA)  
Justin Levandoski (Amazon Web Services, USA)  
Karthik Ramachandra (Microsoft Gray Systems Lab, USA)  
Kaushik Chakrabarti (Microsoft Research, USA)  
Ke Yi (Hong Kong University of Science and Technology, Hong Kong)  
Khuzaima Daudjee (University of Waterloo, Canada)  
Kyuseok Shim (Seoul National University, South Korea)  
Laks V.S. Lakshmanan (The University of British Columbia, Canada)  
Leilani Battle (University of Maryland, USA)  
Li Xiong (Emory University, USA)  
Makoto Onizuka (Osaka University, Japan)  
Manos Athanassoulis (Boston University, USA)  
Marco Serafini (University of Massachusetts Amherst, USA)  
Matthias Boehm (Graz University of Technology, Austria)  
Michael Hay (Colgate University, USA)  
Mirek Riedewald (Northeastern University, USA)  
Mohamed Mokbel (University of Minnesota - Twin Cities, USA)



**Program Committee (continued):** Mohammad Sadoghi (University of California-Davis, USA)  
Mourad Ouzzani (Qatar Computing Research Institute, Qatar)  
Nan Tang (Qatar Computing Research Institute, Qatar)  
Nesime Tatbul (Intel Labs and MIT, USA)  
Nick Koudas (University of Toronto, Canada)  
Nikita Bhutani (Megagon Labs, USA)  
Nikos Mamoulis (University of Ioannina, Greece)  
Oliver A Kennedy (University at Buffalo, SUNY, USA)  
Orestis Polychroniou (Amazon, USA)  
Paolo Papotti (Eurecom, France)  
Paraschos Koutris (University of Wisconsin-Madison, USA)  
Paul Suganthan (Google, USA)  
Peter Alvaro (UC Santa Cruz, USA)  
Peter Boncz (CWI, Netherlands)  
Peter Pietzuch (Imperial College London, UK)  
Pradap Konda (Facebook, USA)  
Qiong Luo (Hong Kong University of Science & Technology, China)  
Raghav Kaushik (Microsoft, USA)  
Raymond Chi-Wing Wong (Hong Kong University of Science & Technology, China)  
Rebecca Taft (Cockroach Labs, USA)  
Renata Borovica-Gajic (University of Melbourne, Australia)  
S. Sudarshan (IIT Bombay, India)  
Sanjay Krishnan (University of Chicago, USA)  
Sara Cohen (The Hebrew University of Jerusalem, Israel)  
Saravanan Thirumuruganathan (QCRI, Qatar)  
Sebastian Schelter (New York University, USA)  
Semih Salihoglu (University of Waterloo, Canada)  
Sergey Melnik (Google, USA)  
Sharad Mehrotra (UC Irvine, USA)  
Shivaram Venkataraman (University of Wisconsin-Madison, USA)  
Sidharth Mudgal (Google, USA)  
Slava Novgorodov (eBay Research, Israel)  
Sourav S Bhowmick (Nanyang Technological University, Singapore)  
Sriram Rao (Microsoft, USA)  
Steven Whang (KAIST, South Korea)  
Stratos Idreos (Harvard, USA)  
Sudeepa Roy (Duke University, USA)  
Sudip Roy (Google, USA)  
Sudipto Das (Amazon Web Services, USA)  
Themis Palpanas (Paris Descartes University, France)  
Theodoros Rekatsinas (University of Wisconsin-Madison, USA)  
Thomas Neumann (TUM, Germany)

**Program Committee (continued):** Tim Kraska (MIT, USA)  
Vagelis Hristidis (UC Riverside, USA)  
Viktor Leis (Friedrich Schiller University Jena, Germany)  
Vivek Narasayya (Microsoft Research, USA)  
Walid Aref (Purdue Univ., USA)  
Wei Wang (University of New South Wales, Australia)  
Wei-Shinn Ku (Auburn University, USA)  
Wenfei Fan (University of Edinburgh, UK & Beihang Univ., China)  
Wentao Wu (Microsoft Research, USA)  
Wolfgang Gatterbauer (Northeastern University, USA)  
Xi He (University of Waterloo, Canada)  
Xi Wu (Google, USA)  
Xiangyao Yu (MIT, USA)  
Xiaokui Xiao (National University of Singapore, Singapore)  
Xiaolan Wang (University of Massachusetts Amherst, USA)  
Xu Chu (GATECH, USA)  
Xuemin Lin (University of New South Wales, Australia)  
Yael Amsterdamer (Bar-Ilan University, Israel)  
Yannis Katsis (IBM Research Almaden, USA)  
Yannis Velegrakis (Utrecht University, Netherlands)  
Yeye He (Microsoft Research, USA)  
Yongjoo Park (University of Michigan, USA)  
Yash Govind (University of Wisconsin - Madison, USA)  
Yuanyuan Tian (IBM Research - Almaden, USA)  
Yuliang Li (UC San Diego, USA)  
Ziawasch Abedjan (TU Berlin, Germany)

**External Program Committee:** Michael Abebe (University of Waterloo, Canada)  
Muhammad Abu Bakar Siddique (UC Riverside, USA)  
Darshana Balakrishnan (University at Buffalo, SUNY, USA)  
Dmytro Bogatov (Boston University, USA)  
Tianwen Chen (The Hong Kong University of Science & Technology, China)  
Yannis Chronis (University of Wisconsin, USA)  
Anthony Colas (University of Florida, USA)  
Shaleen Deep (University of Wisconsin, USA)  
Jialin Ding (MIT, USA)  
Karima Echihabi (IRDA - Rabat IT Center & ENSIAS - Mohammed V University, Morocco)  
Yixiang Fang (The University of New South Wales, Australia)  
Anna Fariha (University of Massachusetts Amherst, USA)  
Dhrubjyothi Ghosh (UC Irvine, USA)  
Brad Glasbergen (University of Waterloo, Canada)

**External Program Committee** Paul Grubbs (Cornell University, USA)  
**(continued):** Guimu Guo (The University of Alabama at Birmingham, USA)  
Peeyush Gupta (UC Irvine, USA)  
Surabhi Gupta (Microsoft Research, India)  
Zsolt Istvan (IMDEA Software Institute, Spain)  
Fuad Jamour (UC Riverside, USA)  
Nikos R. Katsipoulakis (Amazon.com, USA)  
Jongik Kim (Jeonbuk National University, Korea)  
Taewoo Kim (Microsoft, USA)  
Longbin Lai (Alibaba Group, USA)  
Qingcan Li (Simon Fraser University, Canada)  
Hao Liu (The Hong Kong University of Science and Technology, China)  
Brandon Lockhart (Simon Fraser University, Canada)  
Merlin Mao (UC Riverside, USA)  
Ryan Marcus (MIT, USA)  
Aisha Mohamed (Qatar Computing Research Institute, Qatar)  
JuHyoung Mun (Boston University, USA)  
Matthaios Olma (Microsoft, USA)  
Jianglin Peng (Simon Fraser University, Canada)  
Changbo Qu (Simon Fraser University, Canada)  
Ryan Rivas (UC Riverside, USA)  
Miguel Rodriguez (University of Florida, USA)  
Ali Sadeghian (University of Florida, USA)  
Subhadeep Sarkar (Boston University, USA)  
Zeyuan Shang (MIT, USA)  
Shantanu Sharma (UC Irvine, USA)  
Mohamed Sharaf (United Arab Emirates University, UAE)  
Anatoli Shein (Vertica, USA)  
Michael Shekelyan (University of Warwick, UK)  
Panos Simatis (HKUST, China)  
William Spoth (University at Buffalo, SUNY, USA)  
Dimitris Tsaras (HKUST, China)  
Kai Wang (The University of New South Wales, Australia)  
Pei Wang (Simon Fraser University, Canada)  
Weicheng Wang (The Hong Kong University of Science and Technology, China)  
Xiaoying Wang (Simon Fraser University, Canada)  
Yifan Wang (University of Florida, USA)  
Xiaoyang Wang (Zhejiang Gongshang University, China)  
Victor Junqiu Wei (Noah's Ark Lab, Huawei Technologies, China)  
Dong Wen (The Sydney University of Technology, Australia)  
Min Xie (Shenzhen Institute of Computing Sciences, China)  
Zhengyi Yang (The University of New South Wales, Australia)

**External Program Committee (continued):** Xiaotian You (The Hong Kong University of Science & Technology, China)  
Kai Zeng (Alibaba, USA)  
Fan Zhang (Guanzhong University, China)  
Zichen Zhu (Boston University, USA)  
Kostas Zoumpatianos (Harvard University, USA & University of Paris, France)

**Industrial Track PC Chairs:** C. Mohan (IBM Research - Almaden, USA)  
Divesh Srivastava (AT&T Labs, USA)

**Industrial Track PC Members:** Alan Halverson (Microsoft, USA)  
Alexander Boehm (SAP SE, Germany)  
Alon Halevy (Facebook, USA)  
Anastasia Ailamaki (EPFL and RAW Labs, Switzerland)  
Calisto Zuzarte (IBM Canada)  
Eric Simon (SAP, France)  
Eva Sitaridi (Amazon, USA)  
Georgia Koutrika (Athena Research Center, Greece)  
Georgios Giannikis (Oracle Labs, Switzerland)  
Hamid Pirahesh (IBM Research, USA)  
Jun Rao (Confluent, USA)  
Kristen Lefevre (Google, Switzerland)  
Latha Colby (Sigma Computing, USA)  
Lyublena Antova (Datometry, USA)  
Mahashweta Das (Visa Research, USA)  
Martin Kersten (CWI, Netherlands)  
Matei Zaharia (Stanford and Databricks, USA)  
Michaela Hardt (Amazon, USA)  
Nicolas Bruno (Microsoft, USA)  
Rajeev Rastogi (Amazon, India)  
Rick Cole (Tableau, USA)  
Sameep Mehta (IBM Research, India)  
Tamraparni Dasu (AT&T Labs - Research, USA)  
Vijayshankar Raman (Google, USA)  
Vladislav Shkapenyuk (AT&T Labs - Research, USA)

**Demonstration Track PC Chairs:** Sihem Amer-Yahia (Laboratoire d'Informatique de Grenoble, France)  
Tilman Rabl (HPI, University of Potsdam, Germany)

**Demonstration Track PC Members:** Alessandro Margara (Politecnico di Milano, Italy)  
Andras Benczur (Hungarian Academy of Sciences, Hungary)  
Asterios Katsifodimos (TU Delft, Netherlands)  
Behrooz Omidvar-Tehrani (NAVER LABS Europe, France)  
Beng Chin Ooi (NUS, Singapore)

**Demonstration Track PC Members** Chen Xu (East China Normal University, China)  
**(continued):** Chengkai Li (UT Arlington, USA)  
Christoph Koch (EPFL, Switzerland)  
Chunyang Ye (Chinese Academy of Sciences, China)  
Cong Yu (Google, USA)  
Emanuel Zgraggen (MIT, USA)  
Evaggelia Pitoura (University of Ioannina, Greece)  
Fei Chiang (McMaster University, Canada)  
George Fletcher (TU Eindhoven, Netherlands)  
Georgia Koutrika (Athena Research Center, Greece)  
Ioana Manolescu (INRIA Saclay, France)  
Jiang Lilong (Twitter, USA)  
Jonas Traub (Technische Universität Berlin, Germany)  
Kaiwen Zhang (ETS Montreal, Canada)  
Khuzaima Daudjee (University of Waterloo, Canada)  
Marco Serafini (University of Massachusetts Amherst, USA)  
Matthias Uflacker (Hasso Plattner Institute, Germany)  
Maya Ramanath (IIT Delhi, India)  
Meikel Poess (Oracle, USA)  
Michael Carey (UC Irvine, USA)  
Michael Grossniklaus (University of Konstanz, Germany)  
Mohammad Sadoghi (University of California, Davis, USA)  
Panagiotis Bouros (Johannes Gutenberg University Mainz, Germany)  
Panagiotis Karras (Aarhus University, Denmark)  
Saravanan Thirumuruganathan (QCRI, Qatar)  
Seif Haridi (KTH, Sweden)  
Senjuti Basu Roy (New Jersey Institute of Technology, USA)  
Shady Elbassuoni (AUB, Lebanon)  
Sourav S Bhowmick (Nanyang Technological University, Singapore)  
Stefanie Scherzinger (OTH Regensburg, Germany)  
Sudeepa Roy (Duke University, USA)  
Themis Palpanas (Paris Descartes University, France)  
Xiang Ni (IBM Research, USA)  
Yannis Velegrakis (Utrecht University, Netherlands)  
Yongluan Zhou (University of Copenhagen, Denmark)

**Best Paper Award Chair:** Alon Halevy (Facebook AI, USA)  
**Best Paper Award Committee:** Anastasia Ailamaki (EPFL, Switzerland)  
Peter Haas (University of Massachusetts, USA)  
Christian S. Jensen (Aalborg University, Denmark)