

June 30–July 5, 2019  
Amsterdam, Netherlands



Association for  
Computing Machinery

*Advancing Computing as a Science & Profession*

**SIGMOD /**  
**A M S P X D S**  
**T E R**  
**D A M 2 X 1 9**



# SIGMOD '19

Proceedings of the 2019 International Conference on  
Management of Data

*General Chairs:*

***Peter Boncz & Stefan Manegold (CWI, Netherlands)***

*Program Chair:*

***Anastasia Ailamaki (EPFL, Switzerland)***

*Program Vice-Chairs:*

***Amol Deshpande (University of Maryland, USA)***

***Tim Kraska (MIT, 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 © 2019 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:** 978-1-4503-5643-5

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 2019 - The 2019 ACM SIGMOD International Conference on the Management of Data!

This year, the conference is held in the city center of Amsterdam, capital of The Netherlands. Amsterdam is an internationally oriented city, home to people with origins from all over the world. This used to be already the case even back in the 16th and 17th century, when Amsterdam was the world's biggest trading and financial center; establishing the world's first stock exchange in 1602.

SIGMOD/PODS 2019 is held in the original Amsterdam Stock and Commodities Exchange, constructed by Dutch architect Berlage between 1896 and 1903, which now serves as the well-equipped Amsterdam Conference Center. This architect and his apprentices (the school of Berlage) left an important mark on the city, being responsible for a major expansion of the city in the early 20th century. The sculptures and drawings in the Exchange building refer to the people behind the commodities traded in the various rooms ("Effecten" - stock; "Graan" - grains), e.g., depicting farmers in the grain exchange room; as a reminder that trading affects society. Amsterdam is a city that offers many cultural activities, including the world-famous classical Concertgebouw Orchestra, as well as many museums (Amsterdam Museum, Rijksmuseum, Rembrandthuis, Anne Frank Huis). In a slight break with SIGMOD tradition, the SIGMOD opening reception will be held one day later, on Tuesday night, when the SIGMOD/PODS attendees will have exclusive access to the Van Gogh museum. The Wednesday conference dinner is organized across the water in Amsterdam North, in Noorderlicht Cafe in a festival-like environment. This used to be harbour area and was less-populated and industrial, but in the recent decade has become a hotspot for nightlife activities.

Amsterdam is also increasingly a hub for data science companies and services, with multiple universities and CWI in the vicinity; which all participate in the organization of SIGMOD/PODS 2019. On Thursday night, after the SIGMOD program finishes, there will be a meetup of Amsterdam Data Science, where the local data science community will be able to mingle with our data management research community.

The SIGMOD 2019 Research Program Committee consists of the Program Chair, two Program Vice Chairs, a core committee with 37 members, and a regular committee with 98 members. During the reviewing period, we solicited additional reviews from 16 external reviewers and occasional input from 10 assistant reviewers. The committee received 430 submissions, out of which 12 were desk-rejected (i.e., without review). There was no bidding; instead, reviewer assignments were made using input from Microsoft's Conference Management System, the Toronto Paper Matching System, and the reviewers' background (the detailed assignment procedure is described in a paper which has been submitted for publication to SIGMOD Record). The core committee members had (roughly) double the reviewing load of the regular committee members, and in addition acted as discussion leaders and meta-reviewers for their assigned papers. There were two rounds of submissions, with deadlines in July and November, respectively. Initially, each paper received three reviews. At this point authors could read the reviews and provide feedback about potential factual errors (disclosed to the reviewers) or sensitive issues about potential mishandling (confidentially to the chair). Two additional reviews were solicited for a paper if (a) the reviewers' expertise level was suboptimal, or (b) if there was significant score discrepancy in the first three reviews, or (c) if it was heading for rejection but

had received a weak accept (or higher) by at least one reviewer. Papers were discussed extensively online; 10 were accepted based on the first round of reviews, while 311 were rejected. The authors of the remaining 97 papers were asked to revise their papers to address reviewers' criticisms; 78 revisions were ultimately accepted for a total of 88 papers which are presented in the research track. Finally, 12 papers were shepherded after acceptance to guarantee that the camera-ready version addresses all of the reviewers' comments.

**Peter Boncz, Stefan Manegold**

*SIGMOD'19 General Chairs*

*CWI, Netherlands*

**Anastasia Ailamaki**

*SIGMOD'19 Program Chair*

*EPFL, Switzerland*

# Table of Contents

**SIGMOD 2019 Organization** ..... xvii

**SIGMOD 2019 Sponsors & Supporters** ..... xi

## **SIGMOD Keynote 1**

Session Chair: Peter Boncz

- **Responsible Data Science** ..... 1  
Lise Getoor (*University of California, Santa Cruz*)

## **Research 1: Query Processing & Optimization 1 -- sponsored by Tableau**

Session Chair: Wolfgang Lehner

- **Exact Cardinality Query Optimization with Bounded Execution Cost**..... 2  
Immanuel Trummer (*Cornell University*)
- **Pessimistic Cardinality Estimation: Tighter Upper Bounds for Intermediate Join Cardinalities** ..... 18  
Walter Cai, Magdalena Balazinska, Dan Suciu (*University of Washington*)
- **Efficiently Searching In-Memory Sorted Arrays: Revenge of the Interpolation Search?** ..... 36  
Peter Van Sandt, Yannis Chronis, Jignesh M. Patel (*University of Wisconsin-Madison*)
- **Iterative Query Processing based on Unified Optimization Techniques** ..... 54  
Kisung Park, Hojin Seo, Mostofa Kamal Rasel, Young-Koo Lee (*Kyung Hee University*), Chanhoo Jeong, Sung Yeol Lee, Chungmin Lee, Dong-Hun Lee (*SAP Labs Korea*)
- **Approximate Distinct Counts for Billions of Datasets**..... 69  
Daniel Ting (*Tableau Software*)
- **Cache-oblivious High-performance Similarity Join**..... 87  
Martin Perdacher, Claudia Plant (*University of Vienna*), Christian Böhm (*Ludwig-Maximilians-Universität*)

## **Research 2: Privacy/Blockchain**

Session Chair: Raghav Kaushik

- **Blurring the Lines between Blockchains and Database Systems: the Case of Hyperledger Fabric**..... 105  
Ankur Sharma, Felix Martin Schuhknecht, Divya Agrawal, Jens Dittrich (*Saarland University*)
- **Towards Scaling Blockchain Systems via Sharding** ..... 123  
Hung Dang, Tien Tuan Anh Dinh, Dumitrel Loghin, Ee-Chien Chang, Qian Lin, Beng Chin Ooi (*National University of Singapore*)
- **vChain: Enabling Verifiable Boolean Range Queries over Blockchain Databases** ..... 141  
Cheng Xu, Ce Zhang, Jianliang Xu (*Hong Kong Baptist University*)
- **Answering Multi-Dimensional Analytical Queries under Local Differential Privacy**..... 159  
Tianhao Wang (*Purdue University*), Bolin Ding, Jingren Zhou, Cheng Hong, Zhicong Huang (*Alibaba Group*), Ninghui Li (*Purdue University*), Somesh Jha (*University of Wisconsin*)
- **APEX: Accuracy-Aware Differentially Private Data Exploration**..... 177  
Chang Ge, Xi He, Ihab F. Ilyas (*University of Waterloo*), Ashwin Machanavajjhala (*Duke University*)
- **Active Sparse Mobile Crowd Sensing Based on Matrix Completion** ..... 195  
Kun Xie, Xiaocan Li (*Hunan University*), Xin Wang (*Stony Brook University*), Gaogang Xie, Jigang Wen (*Institute of Computing Technology & Chinese Academy of Sciences*), Dafang Zhang (*Hunan University*)

## Research 3: Information Extraction

Session Chair: Guoliang Li

- **Autocompletion for Prefix-Abbreviated Input**..... 211  
Sheng Hu, Chuan Xiao (*Nagoya University & Osaka University*), Jianbin Qin (*Shenzhen University*),  
Yoshiharu Ishikawa (*Nagoya University*), Qiang Ma (*Kyoto University*)
- **Progressive Deep Web Crawling Through Keyword Queries For Data Enrichment** ..... 229  
Pei Wang, Ryan Shea, Jiannan Wang (*Simon Fraser University*), Eugene Wu (*Columbia University*)
- **Visual Segmentation for Information Extraction from Heterogeneous Visually Rich Documents** ..... 247  
Ritesh Sarkhel, Arnab Nandi (*Ohio State University*)
- **RRR: Rank-Regret Representative** ..... 263  
Abolfazl Asudeh (*University of Michigan*), Azade Nazi (*Google AI*), Nan Zhang (*Pennsylvania State University*),  
Gautam Das (*University of Texas at Arlington*), H. V. Jagadish (*University of Michigan*)
- **Strongly Truthful Interactive Regret Minimization**..... 281  
Min Xie, Raymond Chi-Wing Wong (*Hong Kong University of Science and Technology*),  
Ashwin Lall (*Denison University*)
- **Verifying Text Summaries of Relational Data Sets**..... 299  
Saehan Jo, Immanuel Trummer, Weicheng Yu (*Cornell University*), Xuezhi Wang, Cong Yu (*Google Research*),  
Daniel Liu, Niyati Mehta (*Cornell University*)

## Industry 1: Data Applications

Session Chair: Marco Serafini

- **QuickInsights: Quick and Automatic Discovery of Insights from Multi-Dimensional Data**... 317  
Rui Ding, Shi Han, Yong Xu, Haidong Zhang, Dongmei Zhang (*Microsoft Research*)
- **ExplainIt! – A Declarative Root-cause Analysis Engine for Time Series Data**..... 333  
Vimalkumar Jeyakumar, Omid Madani, Ali Parandeh, Ashutosh Kulshreshtha, Weifei Zeng, Navindra Yadav  
(*Cisco Tetration Analytics*)
- **Automatically Generating Interesting Facts from Wikipedia Tables**..... 349  
Flip Korn, Xuezhi Wang, You Wu, Cong Yu (*Google Research*)
- **Snorkel DryBell: A Case Study in Deploying Weak Supervision at Industrial Scale**..... 362  
Stephen H. Bach (*Brown University*),  
Daniel Rodriguez, Yintao Liu, Chong Luo, Haidong Shao, Cassandra Xia, Souvik Sen (*Google*),  
Alex Ratner, Braden Hancock (*Stanford University*), Houman Alborzi, Rahul Kuchhal (*Google*),  
Chris Ré (*Stanford University*), Rob Malkin (*Google*)
- **PS2: Parameter Server on Spark** ..... 376  
Zhipeng Zhang (*Peking University & Tencent Inc.*), Bin Cui (*Peking University*),  
Yingxia Shao (*Beijing University of Posts and Telecommunications*), Lele Yu, Jiawei Jiang (*Tencent Inc.*),  
Xupeng Miao (*Peking University & Tencent Inc.*)
- **Entity Matching Meets Data Science: A Progress Report from the Magellan Project**..... 389  
Yash Govind (*University of Wisconsin - Madison*), Pradap Konda (*Facebook*), Paul Suganthan G.C. (*Google*),  
Philip Martinkus (*University of Wisconsin - Madison*), Palaniappan Nagarajan (*Amazon*),  
Han Li, Aravind Soundararajan (*University of Wisconsin - Madison*), Sidharth Mudgal (*Amazon*),  
Jeff R. Ballard, Haojun Zhang (*University of Wisconsin - Madison*), Adel Ardalan (*Columbia University*),  
Sanjib Das (*Google*), Derek Paulsen, Amanpreet Singh Saini (*University of Wisconsin - Madison*),  
Erik Paulson, Youngchoon Park (*Johnson Control*),  
Marshall Carter, Mingju Sun, Glenn M. Fung (*American Family Insurance*),  
AnHai Doan (*University of Wisconsin - Madison*)

## SIGMOD Keynote 2

Session Chair: Stefan Manegold

- **State of Public and Private Blockchains: Myths and Reality**..... 404  
C. Mohan (*IBM Almaden Research Center*)

## Panel

- **The Responsibility Challenge for Data** ..... 412  
H. V. Jagadish (*University of Michigan, Ann Arbor*), Francesco Bonchi (*ISI Foundation*),  
Tina Eliassi-Rad (*Northeastern University*), Lise Getoor (*University of California, Santa Cruz*),  
Krishna Gummadi (*Max Planck Institute for Software Systems*), Julia Stoyanovich (*New York University*)

## Research 4: Distributed Data Management

Session Chair: Holger Pirk

- **An End-to-End Automatic Cloud Database Tuning System Using Deep Reinforcement Learning**..... 415  
Ji Zhang, Yu Liu, Ke Zhou (*Huazhong University of Science and Technology*), Guoliang Li (*Tsinghua University*),  
Zhili Xiao, Bin Cheng, Jiashu Xing (*Tencent Inc.*),  
Yangtao Wang, Tianheng Cheng, Li Liu, Minwei Ran, Zekang Li (*Huazhong University of Science and Technology*)
- **Fast General Distributed Transactions with Opacity** ..... 433  
Alex Shamis (*Microsoft Research*), Matthew Renzelmann (*Microsoft*), Stanko Novakovic (*VMware*),  
Georgios Chatzopoulos (*École Polytechnique Fédérale de Lausanne*),  
Aleksandar Dragojević, Dushyanth Narayanan, Miguel Castro (*Microsoft Research*)
- **The Log-Structured Merge-Bush & the Wacky Continuum** ..... 449  
Niv Dayan, Stratos Idreos (*Harvard University*)
- **RaSQL: Greater Power and Performance for Big Data Analytics with Recursive-aggregate-SQL on Spark**..... 467  
Jiaqi Gu, Yugo H. Watanabe (*University of California, Los Angeles*), William A. Mazza (*University of Naples Federico II*),  
Alexander Shkapsky (*Workday, Inc.*), Mohan Yang (*Google LLC*),  
Ling Ding, Carlo Zaniolo (*University of California, Los Angeles*)

## Research 5: Provenance

Session Chair: Alexandra Meliou

- **Going Beyond Provenance: Explaining Query Answers with Pattern-based Counterbalances**..... 485  
Zhengjie Miao (*Duke University*), Qitian Zeng, Boris Glavic (*Illinois Institute of Technology*),  
Sudeepa Roy (*Duke University*)
- **Explaining Wrong Queries Using Small Examples**..... 503  
Zhengjie Miao, Sudeepa Roy, Jun Yang (*Duke University*)
- **Ariadne: Online Provenance for Big Graph Analytics**..... 521  
Vicky Papavasileiou (*University of California-San Diego*),  
Ken Yocum (*Intuit, Inc. & University of California-San Diego*), Alin Deutsch (*University of California-San Diego*)
- **Hypothetical Reasoning via Provenance Abstraction**..... 537  
Daniel Deutch, Yuval Moskovitch, Noam Rinetzky (*Tel Aviv University*)

## Research 6: Streams

Session Chair: Jonathan Goldstein

- **Event Trend Aggregation Under Rich Event Matching Semantics** ..... 555  
Olga Poppe (*Microsoft Gray Systems Lab*), Chuan Lei (*IBM Research - Almaden*),  
Elke A. Rundensteiner (*Worcester Polytechnic Institute*), David Maier (*Portland State University*)
- **Elasticutor: Rapid Elasticity for Realtime Stateful Stream Processing**..... 573  
LI Wang (*Yitu Technology*), Tom Z. J. Fu (*Advanced Digital Sciences Center*),  
Richard T. B. Ma (*National University of Singapore*), Marianne Winslett (*University of Illinois Urbana-Champaign*),  
Zhenjie Zhang (*Yitu Technology*)
- **Real-Time Multi-Pattern Detection over Event Streams**..... 589  
Ilya Kolchinsky, Assaf Schuster (*Technion, Israel Institute of Technology*)
- **AStream: Ad-hoc Shared Stream Processing**..... 607  
Jeyhun Karimov (*DFKI GmbH*), Tilmann Rabl, Volker Markl (*DFKI GmbH & TU Berlin*)

## Industry 2: Storage & Indexing

Session Chair: Alexander Shraer

- **Nanosecond Indexing of Graph Data With Hash Maps and VLists** ..... 623  
Andrew Carter, Andrew Rodriguez, Yiming Yang, Scott Meyer (*LinkedIn Corporation*)
- **Implementation of Cluster-wide Logical Clock and Causal Consistency in MongoDB**..... 636  
Misha Tyulenev, Andy Schwerin, Asya Kamsky, Randolph Tan, Alyson Cabral, Jack Mulrow (*MongoDB, Inc*)
- **X-Engine: An Optimized Storage Engine for Large-scale E-commerce Transaction Processing** ..... 651  
Gui Huang, Xuntao Cheng, Jianying Wang, Yujie Wang, Dengcheng He, Tieying Zhang, Feifei Li, Sheng Wang, Wei Cao, Qiang Li (*Alibaba Group*)
- **Automatically Indexing Millions of Databases in Microsoft Azure SQL Database**..... 666  
Sudipto Das, Miroslav Grbic, Igor Ilic, Isidora Jovandic, Andrija Jovanovic, Vivek R. Narasayya, Miodrag Radulovic, Maja Stikic, Gaoxiang Xu, Surajit Chaudhuri (*Microsoft Corporation*)

## Research 7: Modern Hardware

Session Chair: Justin Lewandowski

- **Concurrent Prefix Recovery: Performing CPR on a Database**..... 687  
Guna Prasaad (*University of Washington*), Badrish Chandramouli, Donald Kossmann (*Microsoft Research*)
- **BriskStream: Scaling Data Stream Processing on Shared-Memory Multicore Architectures..** 705  
Shuhao Zhang (*National University of Singapore*), Jiong He (*Advanced Digital Sciences Center*),  
Amelie Chi Zhou (*Shenzhen University*), Bingsheng He (*National University of Singapore*)
- **Border-Collie: A Wait-free, Read-optimal Algorithm for Database Logging on Multicore Hardware**..... 723  
Jongbin Kim, Hyeongwon Jang, Seohui Son (*Hanyang University*), Hyuck Han (*Dongduk Women's University*),  
Sooyong Kang, Hyungsoo Jung (*Hanyang University*)
- **Designing Distributed Tree-based Index Structures for Fast RDMA-capable Networks**..... 741  
Tobias Ziegler (*TU Darmstadt*), Sumukha Tumkur Vani (*Brown University*), Carsten Binnig (*TU Darmstadt*),  
Rodrigo Fonseca (*Brown University*), Tim Kraska (*Massachusetts Institute of Technology*)
- **DistME: A Fast and Elastic Distributed Matrix Computation Engine using GPUs** ..... 759  
Donghyoung Han, Yoon-Min Nam, Jihye Lee (*Daegu Gyeongbuk Institute of Science & Technology (DGIST)*),  
Kyongseok Park, Hyunwoo Kim (*Korea Institute of Science and Technology Information (KISTI)*),  
Min-Soo Kim (*Daegu Gyeongbuk Institute of Science & Technology (DGIST)*)
- **GPU-based Graph Traversal on Compressed Graphs** ..... 775  
Mo Sha (*National University of Singapore*), Yuchen Li (*Singapore Management University*),  
Kian-Lee Tan (*National University of Singapore*)

## Research 8: Data Integration/Cleaning

Session Chair: Paolo Papotti

- **Interventional Fairness : Causal Database Repair for Algorithmic Fairness** ..... 793  
Babak Salimi, Luke Rodriguez, Bill Howe, Dan Suciu (*University of Washington*)
- **Uni-Detect: A Unified Approach to Automated Error Detection in Tables**..... 811  
Pei Wang (*Simon Fraser University*), Yeye He (*Microsoft Research*)
- **HoloDetect: Few-Shot Learning for Error Detection** ..... 829  
Alireza Heidari (*University of Waterloo*), Joshua McGrath (*University of Wisconsin-Madison*),  
Ihab F. Ilyas (*University of Waterloo*), Theodoros Rekatsinas (*University of Wisconsin-Madison*)
- **JOSIE: Overlap Set Similarity Search for Finding Joinable Tables in Data Lakes** ..... 847  
Erkang Zhu (*University of Toronto*), Dong Deng (*Rutgers University & Inception Institute of Artificial Intelligence*),  
Fatemeh Nargesian (*University of Toronto*), Renée J. Miller (*Northeastern University*)



- **Raha: A Configuration-Free Error Detection System**..... 865  
 Mohammad Mahdavi, Ziawasch Abedjan (*TU Berlin*),  
 Raul Castro Fernandez, Samuel Madden (*Massachusetts Institute of Technology*), Mourad Ouzzani (*QCRI, HBKU*),  
 Michael Stonebraker (*Massachusetts Institute of Technology*), Nan Tang (*QCRI, HBKU*)
- **Speculative Distributed CSV Data Parsing for Big Data Analytics** ..... 883  
 Chang Ge (*University of Waterloo*),  
 Yinan Li, Eric Eilebrecht, Badrish Chandramouli, Donald Kossmann (*Microsoft Research*)

## Research 9: Query Processing & Optimization 2

Session Chair: Jun Yang

- **CATAPULT: Data-driven Selection of Canned Patterns for Efficient Visual Graph Query Formulation** ..... 900  
 Kai Huang (*Fudan University*), Huey Eng Chua, Sourav S. Bhowmick (*Nanyang Technological University*),  
 Byron Choi (*Hong Kong Baptist University*), Shuigeng Zhou (*Fudan University*)
- **iQCAR: inter-Query Contention Analyzer for Data Analytics Frameworks** ..... 918  
 Prajakta Kalmegh (*Duke University*), Shivnath Babu (*Unravel Data Systems*), Sudeepa Roy (*Duke University*)
- **A Holistic Approach for Query Evaluation and Result Vocalization in Voice-Based OLAP** ..... 936  
 Immanuel Trummer Yicheng Wang, Saketh Mahankali (*Cornell University*)
- **Top- $k$  Queries over Digital Traces**..... 954  
 Yifan Li, Xiaohui Yu (*York University*), Nick Koudas (*University of Toronto*)
- **Visual Road: A Video Data Management Benchmark** ..... 972  
 Brandon Haynes, Amrita Mazumdar, Magdalena Balazinska, Luis Ceze, Alvin Cheung (*University of Washington*)
- **Mining Precision Interfaces From Query Logs**..... 988  
 Qianrui Zhang (*Tsinghua University*), Haoci Zhang, Thibault Sellam, Eugene Wu (*Columbia University*)

## Research 10: Graphs 1

Session Chair: Angela Bonifati

- **Distance-generalized Core Decomposition**..... 1006  
 Francesco Bonchi (*ISI Foundation & Eurecat*), Arijit Khan (*Nanyang Technological University*), Lorenzo Severini (*ISI Foundation*)
- **Unboundedness and Efficiency of Truss Maintenance in Evolving Graphs** ..... 1024  
 Yikai Zhang, Jeffrey Xu Yu (*Chinese University of Hong Kong*)
- **PRSim: Sublinear Time SimRank Computation on Large Power-Law Graphs**..... 1042  
 Zhewei Wei (*Renmin University of China*), Xiaodong He (*4Paradigm Inc.*),  
 Xiaokui Xiao (*National University of Singapore*), Sibow Wang (*Chinese University of Hong Kong*),  
 Yu Liu (*Peking University*), Xiaoyong Du, Ji-Rong Wen (*Renmin University of China*)
- **Scaling Distance Labeling on Small-World Networks**..... 1060  
 Wentao Li (*University of Technology Sydney*), Miao Qiao (*University of Auckland*),  
 Lu Qin, Ying Zhang (*University of Technology Sydney*), Lijun Chang (*University of Sydney*),  
 Xuemin Lin (*University of New South Wales*)
- **Maximizing Welfare in Social Networks under A Utility Driven Influence Diffusion model**...1078  
 Prithu Banerjee (*University of British Columbia*), Wei Chen (*Microsoft Research*),  
 Laks V.S. Lakshmanan (*University of British Columbia*)
- **Efficient Approximation Algorithms for Adaptive Seed Minimization**..... 1096  
 Jing Tang (*National University of Singapore*), Keke Huang (*Nanyang Technological University*),  
 Xiaokui Xiao (*National University of Singapore*), Laks V.S. Lakshmanan (*University of British Columbia*),  
 Xueyan Tang (*Nanyang Technological University*), Aixin Sun, Andrew Lim (*National University of Singapore*)

## Award Talks

- **Data Management on Non-Volatile Memory** ..... 1114  
Joy Arulraj (*Georgia Institute of Technology*)
- **Formal Approaches to Querying Big Data in Shared-Nothing Systems**..... 1115  
Bas Ketsman (*École Polytechnique Fédérale de Lausanne*)

## Research 11: Systems & Machine Learning

Session Chair: Matthias Boehm

- **DeepBase: Deep Inspection of Neural Networks** ..... 1117  
Thibault Sellam, Kevin Lin, Ian Huang (*Columbia University*), Michelle Yang (*University of California, Berkeley*), Carl Vondrick, Eugene Wu (*Columbia University*)
- **BlinkML: Efficient Maximum Likelihood Estimation with Probabilistic Guarantees**..... 1135  
Yongjoo Park, Jingyi Qing Xiaoyang Shen, Barzan Mozafari (*University of Michigan*)
- **SkinnerDB: Regret-Bounded Query Evaluation via Reinforcement Learning** ..... 1153  
Immanuel Trummer, Junxiong Wang, Deepak Maram, Samuel Moseley, Saehan Jo, Joseph Antonakakis (*Cornell University*)
- **Democratizing Data Science through Interactive Curation of ML Pipelines** ..... 1171  
Zeyuan Shang, Emanuel Zraggen (*Massachusetts Institute of Technology*), Benedetto Buratti (*Brown University*), Ferdinand Kossmann (*Massachusetts Institute of Technology*), Philipp Eichmann, Yeounoh Chung (*Brown University*), Carsten Binnig (*Brown University & TU Darmstadt*), Eli Upfal (*Brown University*), Tim Kraska (*Massachusetts Institute of Technology*)

## Research 12: Indexing

Session Chair: Stratos Idreos

- **FITing-Tree: A Data-aware Index Structure** ..... 1189  
Alex Galakatos, Michael Markovitch (*Brown University*), Carsten Binnig (*TU Darmstadt*), Rodrigo Fonseca (*Brown University*), Tim Kraska (*Massachusetts Institute of Technology*)
- **Hyperion: Building the Largest In-memory Search Tree** ..... 1207  
Markus Mäsker (*Johannes Gutenberg University Mainz*), Tim Süß (*University of Applied Science Fulda*), Lars Nagel (*Loughborough University*), Lingfang Zeng (*Huazhong University of Science and Technology*), André Brinkmann (*Johannes Gutenberg University Mainz*)
- **Designing Succinct Secondary Indexing Mechanism by Exploiting Column Correlations ...** 1223  
Yingjun Wu (*IBM Research - Almaden*), Jia Yu (*Arizona State University*), Yuanyuan Tian (*IBM Research - Almaden*), Richard Sidle (*IBM*), Ronald Barber (*IBM Research - Almaden*)
- **AI Meets AI: Leveraging Query Executions to Improve Index Recommendations** ..... 1241  
Bailu Ding, Sudipto Das (*Microsoft Research*), Ryan Marcus (*Brandeis University*), Wentao Wu, Surajit Chaudhuri, Vivek R. Narasayya (*Microsoft Research*)

## Research 13: Fairness, Uncertainty

Session Chair: Ke Yi

- **Designing Fair Ranking Schemes** ..... 1259  
Abolfazl Asudeh, H. V. Jagadish (*University of Michigan*), Julia Stoyanovich (*New York University*), Gautam Das (*University of Texas at Arlington*)
- **Anti-Freeze for Large and Complex Spreadsheets: Asynchronous Formula Computation....** 1277  
Mangesh Bendre, Tana Wattanawaroon, Kelly Mack, Kevin Chang, Aditya Parameswaran (*University of Illinois (UIUC)*)

- **Anytime Approximation in Probabilistic Databases via Scaled Dissociations**..... 1295  
Maarten Van den Heuvel (*University of Antwerp*), Peter Ivanov (*Northeastern University*),  
Wolfgang Gatterbauer (*Northeastern University*), Floris Geerts (*University of Antwerp*),  
Martin Theobald (*University of Luxembourg*)
- **Uncertainty Annotated Databases - A Lightweight Approach for Approximating  
Certain Answers** ..... 1313  
Su Feng (*Illinois Institute of Technology*), Aaron Huber (*University at Buffalo*),  
Boris Glavic (*Illinois Institute of Technology*), Oliver Kennedy (*University at Buffalo*)

## Research 14: Graphs 2

Session Chair: Sourav S Bhowmick

- **Efficient Estimation of Heat Kernel PageRank for Local Clustering** ..... 1339  
Renchi Yang, Xiaokui Xiao (*National University of Singapore*), Zhewei Wei (*Renmin University of China*),  
Sourav S. Bhowmick, Jun Zhao (*Nanyang Technological University*), Rong-Hua Li (*Beijing Institute of Technology*)
- **Fractal: A General-Purpose Graph Pattern Mining System** ..... 1357  
Vinicius Dias, Carlos H. C. Teixeira, Dorgival Guedes, Wagner Meira (*Universidade Federal de Minas Gerais*),  
Srinivasan Parthasarathy (*Ohio State University*)
- **Experimental Analysis of Streaming Algorithms for Graph Partitioning**..... 1375  
Anil Pacaci, M. Tamer Özsu (*University of Waterloo*)
- **Interactive Graph Search**..... 1393  
Yufei Tao (*Chinese University of Hong Kong*), Yuanbing Li, Guoliang Li (*Tsinghua University*)

## Research 15: Graphs 3

Session Chair: Xuemin Lin

- **Optimizing Declarative Graph Queries at Large Scale**..... 1411  
Qizhen Zhang, Akash Acharya (*University of Pennsylvania*), Hongzhi Chen (*Chinese University of Hong Kong*),  
Simran Arora (*University of Pennsylvania*), Ang Chen (*Rice University*),  
Vincent Liu, Boon Thau Loo (*University of Pennsylvania*)
- **Efficient Subgraph Matching: Harmonizing Dynamic Programming,  
Adaptive Matching Order, and Failing Set Together** ..... 1429  
Myoungji Han, Hyunjoon Kim, Geonmo Gu, Kunsoo Park (*Seoul National University*),  
Wook-Shin Han (*Pohang University of Science and Technology (POSTECH)*)
- **CECI: Compact Embedding Cluster Index for Scalable Subgraph Matching** ..... 1447  
Bibek Bhattacharai (*George Washington University*), Hang Liu (*University of Massachusetts, Lowell*),  
H. Howie Huang (*George Washington University*)
- **Efficiently Answering Regular Simple Path Queries on Large Labeled Networks** ..... 1463  
Sarisht Wadhwa, Anagh Prasad, Sayan Ranu, Amitabha Bagchi, Srikanta Bedathur (*IIT Delhi*)
- **Answering Why-questions by Exemplars in Attributed Graphs** ..... 1481  
Mohammad Hossein Namaki, Qi Song, Yinghui Wu (*Washington State University*),  
Shengqi Yang (*WeWork Technology*)
- **An Efficient Index for RDF Query Containment** ..... 1499  
Theofilos Mailis, Yannis Kotidis (*Athens University of Economics and Business*),  
Vaggelis Nikolopoulos (*University of Athens*), Evgeny Kharlamov (*University of Oslo & Bosch Center for AI*),  
Ian Horrocks (*University of Oxford*), Yannis Ioannidis (*Athena Research Centre & University of Athens*)

## Research 16: Machine Learning

Session Chair: Theodoros Rekatsinas

- **Tuple-oriented Compression for Large-scale Mini-batch Stochastic Gradient Descent** ..... 1517  
Fengan Li, Lingjiao Chen, Yijing Zeng (*University of Wisconsin-Madison*),  
Arun Kumar (*University of California, San Diego*),  
Xi Wu, Jeffrey F. Naughton, Jignesh M. Patel (*University of Wisconsin-Madison*)
- **Towards Model-based Pricing for Machine Learning in a Data Marketplace**..... 1535  
Lingjiao Chen, Paraschos Koutris (*University of Wisconsin, Madison*),  
Arun Kumar (*University of California, San Diego*)

- **DBEst: Revisiting Approximate Query Processing Engines with Machine Learning Models...** 1553  
Qingzhi Ma, Peter Triantafyllou (*University of Warwick*)
- **Enabling and Optimizing Non-linear Feature Interactions in Factorized Linear Algebra .....** 1571  
Side Li (*University of California, San Diego*), Lingjiao Chen (*University of Wisconsin-Madison*),  
Arun Kumar (*University of California, San Diego*)
- **Incremental and Approximate Inference for Faster Occlusion-based Deep CNN Explanations.....** 1589  
Supun Nakandala, Arun Kumar, Yannis Papakonstantinou (*University of California, San Diego*)
- **MNC: Structure-Exploiting Sparsity Estimation for Matrix Expressions .....** 1607  
Johanna Sommer (*IBM Germany*), Matthias Boehm (*Graz University of Technology*),  
Alexandre V. Evfimievski, Berthold Reinwald (*IBM Research -- Almaden*),  
Peter J. Haas (*University of Massachusetts, Amherst*)

## Research 17: Scalability

Session Chair: Norman May

- **A Scalable Index for Top-k Subtree Similarity Queries.....** 1624  
Daniel Kocher, Nikolaus Augsten (*University of Salzburg*)
- **A Layered Aggregate Engine for Analytics Workloads.....** 1642  
Maximilian Schleich, Dan Olteanu (*University of Oxford*), Mahmoud Abo Khamis, Hung Q. Ngo (*relationalAI*),  
XuanLong Nguyen (*University of Michigan*)
- **Towards Scalable Hybrid Stores: Constraint-Based Rewriting to the Rescue.....** 1660  
Rana Alotaibi (*University of California, San Diego*), Damian Bursztyn (*Thales*),  
Alin Deutsch (*University of California, San Diego*), Ioana Manolescu (*Inria & Ecole polytechnique*),  
Stamatis Zampetakis (*Orchestra Networks*)
- **MIFO: A Query-Semantic Aware Resource Allocation Policy.....** 1678  
Prajakta Kalmegh (*Duke University*), Shivnath Babu (*Unravel Data Systems*)
- **Dissecting the Performance of Strongly-Consistent Replication Protocols .....** 1696  
Ailidani Ailijiang (*Microsoft*), Aleksey Charapko, Murat Demirbas (*University at Buffalo, SUNY*)
- **FishStore: Faster Ingestion with Subset Hashing.....** 1711  
Dong Xie (*University of Utah*), Badrish Chandramouli, Yanan Li, Donald Kossmann (*Microsoft Research*)

## Industry 3: Data Platforms

Session Chair: Ying Zhang

- **CFS: A Distributed File System for Large Scale Container Platforms.....** 1729  
Haifeng Liu (*University of Science and Technology of China*), Wei Ding, Yuan Chen, Weilong Guo, Shuoran Liu,  
Tianpeng Li, Mofei Zhang, Jianxing Zhao, Hongyin Zhu, Zhengyi Zhu (*JD.com*)
- **Socrates: The New SQL Server in the Cloud .....** 1743  
Panagiotis Antonopoulos, Alex Budovski, Cristian Diaconu, Alejandro Hernandez Saenz, Jack Hu,  
Hanuma Kodavalla (*Microsoft*), Donald Kossmann (*Microsoft Research*), Sandeep Lingam (*Microsoft*),  
Umar Farooq Minhas (*Microsoft Research*), Naveen Prakash, Vijendra Purohit, Hugh Qu,  
Chaitanya Sreenivas Ravella, Krystyna Reisteter, Sheetal Shrotri (*Microsoft*), Dixin Tang (*University of Chicago*),  
Vikram Wakade (*Microsoft*)
- **One SQL to Rule Them All - an Efficient and Syntactically Idiomatic Approach to Management of Streams and Tables.....** 1757  
Edmon Begoli (*Oak Ridge National Laboratory*), Tyler Akidau (*Google Inc.*), Fabian Hueske (*Veriverica*),  
Julian Hyde (*Looker Inc.*), Kathryn Knight (*Oak Ridge National Laboratory*), Kenneth Knowles (*Google Inc.*)
- **Apache Hive: From MapReduce to Enterprise-grade Big Data Warehousing.....** 1773  
Jesús Camacho-Rodríguez, Ashutosh Chauhan, Alan Gates, Eugene Koifman, Owen O'Malley, Vineet Garg,  
Zoltan Haindrich, Sergey Shelukhin, Prasanth Jayachandran, Siddharth Seth, Deepak Jaiswal, Slim Bouguerra,  
Nishant Bangarwa, Sankar Hariappan, Anishek Agarwal, Jason Dere, Daniel Dai, Thejas Nair, Nita Dembla,  
Gopal Vijayaraghavan, Günther Hagleitner (*Hortonworks*)

- **FoundationDB Record Layer: A Multi-Tenant Structured Datastore** ..... 1787  
Christos Chrysafis, Ben Collins, Scott Dugas, Jay Dunkelberger, Moussa Ehsan, Scott Gray, Alec Grieser, Ori Herrnstadt, Kfir Lev-Ari, Tao Lin, Mike McMahon, Nicholas Schiefer, Alexander Shraer (*Apple, Inc.*)
- **Data Platform for Machine Learning** ..... 1803  
Pulkit Agrawal, Rajat Arya, Aanchal Bindal, Sandeep Bhatia, Anupriya Gagneja, Joseph Godlewski, Yucheng Low Timothy Muss, Mudit Manu Paliwal Sethu Raman, Vishrut Shah Bochao Shen, Laura Sugden, Kaiyu Zhao, Ming-Chuan Wu (*Apple Inc.*)

## Student Abstracts

- **Scalable Reservoir Sampling on Many-Core CPUs**..... 1817  
Altan Birler (*Technical University of Munich*)
- **Helios: An Adaptive and Query Workload-driven Partitioning Framework for Distributed Graph Stores** ..... 1820  
Ali Davoudian (*Carleton University*)
- **CAvSAT: A System for Query Answering over Inconsistent Databases** ..... 1823  
Akhil A. Dixit (*University of California, Santa Cruz*)
- **Interactive Visualization For Big Spatial Data**..... 1826  
Saheli Ghosh (*University of California, Riverside*)
- **LSM-Trees and B-Trees: The Best of Both Worlds**..... 1829  
Varun Jain, James Lennon, Harshita Gupta (*Harvard University*)
- **Answering Range Queries Under Local Differential Privacy**..... 1832  
Tejas Kulkarni (*The University Of Warwick*)
- **Fingerprints for Compressed Columnar Data Search**..... 1835  
Carmen Kwan (*University of Waterloo*)
- **Learning to Generate Questions with Adaptive Copying Neural Networks** ..... 1838  
Xinyuan Lu (*Carleton University*)
- **Towards Understanding Data Analysis Workflows using a Large Notebook Corpus**..... 1841  
Mohammed Suhail Rehman (*University of Chicago*)
- **Query-Driven Learning for Next Generation Predictive Modeling & Analytics** ..... 1844  
Fotis Savva (*University of Glasgow*)
- **SpeakQL: Towards Speech-driven Multimodal Querying** ..... 1847  
Vraj Shah (*University of California, San Diego*)
- **Arachnid: Generalized Visual Data Cleaning**..... 1850  
Conder L. Shou, Amita Shukla (*Columbia University*)
- **Generating Selective Filters for Access Method and Physical Design Evaluation**..... 1853  
Pranav Subramaniam (*University of Chicago*)
- **Deep Query Optimization**..... 1856  
Tin Vu (*University of California, Riverside*)
- **Recommending Deployment Strategies in Crowdsourcing Platforms**..... 1859  
Dong Wei (*New Jersey Institute of Technology*)
- **Bootstrapping an End-to-End Natural Language Interface for Databases**..... 1862  
Nathaniel Weir (*Brown University*), Prasetya Utama (*TU Darmstadt*)

## Demonstrations

- **GraphWrangler: An Interactive Graph View on Relational Data** ..... 1865  
Nafisa Anzum, Semih Salihoglu, Daniel Vogel (*University of Waterloo*)
- **Apollo: A Dataset Profiling and Operator Modeling System** ..... 1869  
Tasos Bakogiannis, Ioannis Giannakopoulos (*National Technical University of Athens*), Dimitrios Tsoumakos (*Ionian University*), Nectarios Koziris (*National Technical University of Athens*)

- **MapRepair: Mapping and Repairing under Policy Views** ..... 1873  
Angela Bonifati, Ugo Comignani (*Lyon 1 University & Liris CNRS*), Efthymia Tsamoura (*University of Oxford*)
- **Data Debugging and Exploration with Vizier** ..... 1877  
Mike Brachmann (*University at Buffalo*), Carlos Bautista, Sonia Castelo (*New York University*),  
Su Feng (*Illinois Institute of Technology*), Juliana Freire (*New York University*),  
Boris Glavic (*Illinois Institute of Technology*), Oliver Kennedy (*University of Buffalo*),  
Heiko Müller, Rémi Rampin (*New York University*), William Spoth (*University at Buffalo*), Ying Yang (*Oracle*)
- **Large Scale Graph Mining with G-Miner** ..... 1881  
Hongzhi Chen, Xiaoxi Wang, Chenghuan Huang, Juncheng Fang Yifan Hou, Changji Li, James Cheng  
(*Chinese University of Hong Kong*)
- **Demonstration of Nimbus: Model-based Pricing for Machine Learning in a Data Marketplace** ..... 1885  
Lingjiao Chen, Hongyi Wang (*University of Wisconsin-Madison*), Leshang Chen (*University of Pennsylvania*),  
Paraschos Koutris (*University of Wisconsin-Madison*), Arun Kumar (*University of California, San Diego*)
- **Peering through the Dark: An Owl’s View of Inter-job Dependencies and Jobs’ Impact in Shared Clusters** ..... 1889  
Andrew Chung (*Carnegie Mellon University*), Carlo Curino, Subru Krishnan, Konstantinos Karanasos (*Microsoft*),  
Panagiotis Garefalakis (*Imperial College London*), Gregory R. Ganger (*Carnegie Mellon University*)
- **Capturing and Querying Structural Provenance in Spark with Pebble**..... 1893  
Ralf Diestelkämper, Melanie Herschel (*Universität Stuttgart*)
- **CLASH: A High-Level Abstraction for Optimized, Multi-Way Stream Joins over Apache Storm**..... 1897  
Manuel Dossinger, Sebastian Michel, Constantin Roudsarabi (*TU Kaiserslautern*)
- **Visual Exploration of Time Series Anomalies with Metro-Viz**..... 1901  
Philipp Eichmann, Franco Solleza (*Brown University*),  
Nesime Tatbul (*Intel Labs and Massachusetts Institute of Technology*), Stan Zdonik (*Brown University*)
- **BlockchainDB - Towards a Shared Database on Blockchains** ..... 1905  
Muhammad El-Hindi, Martin Heyden (*TU Darmstadt*), Arvind Arasu (*Microsoft Research*),  
Carsten Binnig (*TU Darmstadt*), Donald Kossmann, Ravi Ramamurthy (*Microsoft Research*)
- **Cost-Effective, Workload-Adaptive Migration of Big Data Applications to the Cloud** ..... 1909  
Victor Giannakouris, Alejandro Fernandez, Alkis Simitsis, Shivnath Babu (*Unravel Data Systems*)
- **MithraRanking: A System for Responsible Ranking Design**..... 1913  
Yifan Guan, Abolfazl Asudeh, Pranav Mayuram, H. V. Jagadish (*University of Michigan*),  
Julia Stoyanovich (*New York University*), Jerome Miklau (*University of Massachusetts, Amherst*),  
Gautam Das (*University of Texas at Arlington*)
- **MorphStore - In-Memory Query Processing based on Morphing Compressed Intermediates LIVE** ..... 1917  
Dirk Habich, Patrick Damme, Annett Ungethüm, Johannes Pietrzyk, Alexander Krause, Juliana Hildebrandt,  
Wolfgang Lehner (*Technische Universität Dresden*)
- **Fluid: A Blockchain based Framework for Crowdsourcing**..... 1921  
Siyuan Han, Zihuan Xu, Yuxiang Zeng, Lei Chen (*Hong Kong University of Science and Technology*)
- **MigCast: Putting a Price Tag on Data Model Evolution in NoSQL Data Stores**..... 1925  
Andrea Hillenbrand, Maksym Levchenko, Uta Störl (*Darmstadt University of Applied Sciences*),  
Stefanie Scherzinger (*OTH Regensburg*), Meike Klettke (*University of Rostock*)
- **PgCuckoo: Laying Plan Eggs in PostgreSQL’s Nest**..... 1929  
Denis Hirn, Torsten Grust (*Universität Tübingen*)
- **Demonstration of ModelarDB: Model-Based Management of Dimensional Time Series**..... 1933  
Søren Kejser Jensen, Torben Bach Pedersen, Christian Thomsen (*Aalborg University*)
- **Estimating Cardinalities with Deep Sketches**..... 1937  
Andreas Kipf, Dimitri Vorona, Jonas Müller (*Technical University of Munich*),  
Thomas Kipf (*University of Amsterdam*), Bernhard Radke, Viktor Leis (*Technical University of Munich*),  
Peter Boncz (*Centrum Wiskunde & Informatica*), Thomas Neumann, Alfons Kemper (*Technical University of Munich*)

- **Coconut Palm: Static and Streaming Data Series Exploration Now in your Palm** ..... 1941  
Haridimos Kondylakis (*FORTH-ICS*), Niv Dayan, Kostas Zoumpatianos (*Harvard University*),  
Themis Palpanas (*Paris Descartes University*)
- **NeMeSys - A Showcase of Data Oriented Near Memory Graph Processing**..... 1945  
Alexander Krause, Thomas Kissinger, Dirk Habich, Wolfgang Lehner (*Technische Universität Dresden*)
- **Ratel: Interactive Analytics for Large Scale Trajectories** ..... 1949  
Haoda Li, Guoliang Li, Jiayang Liu, Haitao Yuan, Haiquan Wang (*Tsinghua University*)
- **NEURON: Query Execution Plan Meets Natural Language Processing  
For Augmenting DB Education** ..... 1953  
Siyuan Liu, Sourav S. Bhowmick, Wanlu Zhang Shu Wang, Wanyi Huang, Shafiq Joty  
(*Nanyang Technological University*)
- **CrowdGame: A Game-Based Crowdsourcing System for Cost-Effective Data Labeling** ..... 1957  
Tongyu Liu, Jingru Yang, Ju Fan, Zhewei Wei (*Renmin University of China*), Guoliang Li (*Tsinghua University*),  
Xiaoyong Du (*Renmin University of China*)
- **RATest: Explaining Wrong Relational Queries Using Small Examples** ..... 1961  
Zhengjie Miao, Sudeepa Roy, Jun Yang (*Duke University*)
- **NAVIGATE: Explainable Visual Graph Exploration by Examples**..... 1965  
Mohammad Hossein Namaki, Qi Song, Yinghui Wu (*Washington State University*)
- **Pivotal Greenplum® for Kubernetes: Demonstration of Managing Greenplum  
Database on Kubernetes: Massive Parallel Processing Relational Database in the Cloud** .... 1969  
Jemish Patel, Goutam Tadi, Oz Basarir, Lawrence Hamel, David Sharp Fei Yang, Xin Zhang (*Pivotal Software Inc*)
- **NEWS: News Event Walker and Summarizer**..... 1973  
Radityo Eko Prasajo, Mouna Kacimi, Werner Nutt (*Free University of Bozen-Bolzano*)
- **ANMAT: Automatic Knowledge Discovery and Error Detection through Pattern  
Functional Dependencies** ..... 1977  
Abdulhakim Qahtan, Nan Tang, Mourad Ouzzani (*Qatar Computing Research Institute, HBKU*),  
Yang Cao (*University of Edinburgh*), Michael Stonebraker (*Massachusetts Institute of Technology*)
- **DuckDB: an Embeddable Analytical Database**..... 1981  
Mark Raasveldt, Hannes Mühleisen (*CWI*)
- **ChronosDB in Action: Manage, Process, and Visualize Big Geospatial Arrays in the Cloud** .... 1985  
Ramon Antonio Rodrigues Zalipynis (*National Research University Higher School of Economics*)
- **Ursprung: Provenance for Large-Scale Analytics Environments** ..... 1989  
Lukas Rupperecht (*IBM Research - Almaden*), James C. Davis (*Virginia Tech & IBM Systems*),  
Constantine Arnold (*IBM Research - Almaden*), Alexander Lubbock, Darren Tyson (*Vanderbilt University*),  
Deepavali Bhagwat (*IBM Research - Almaden*)
- **Unit Testing Data with Deequ**..... 1993  
Sebastian Schelter, Felix Biessmann, Dustin Lange, Tammo Rukat, Phillip Schmidt, Stephan Seufert,  
Pierre Brunelle, Andrey Taptunov (*Amazon Research*)
- **Natural Language Querying of Complex Business Intelligence Queries** ..... 1997  
Jaydeep Sen, Fatma Ozcan, Abdul Quamar (*IBM Research AI*), Greg Stager (*IBM Canada*),  
Ashish Mittal, Manasa Jammie, Chuan Lei, Diptikalyan Saha, Karthik Sankaranarayanan (*IBM Research AI*)
- **Demonstration of SpeakQL: Speech-driven Multimodal Querying of Structured Data**..... 2001  
Vraj Shah, Side Li, Kevin Yang, Arun Kumar, Lawrence Saul (*University of California, San Diego*)
- **C<sup>2</sup>Metadata: Automating the Capture of Data Transformations from Statistical  
Scripts in Data Documentation** ..... 2005  
Jie Song, George Alter, H. V. Jagadish (*University of Michigan*)
- **[Demo] Low-latency Spark Queries on Updatable Data**..... 2009  
Alexandru Uta (*Vrije Universiteit Amsterdam*), Bogdan Ghit (*Databricks*),  
Ankur Dave (*University of California, Berkeley*), Peter Boncz (*Centrum Wiskunde & Informatica*)
- **SVQ: Streaming Video Queries** ..... 2013  
Ioannis Xarchakos, Nick Koudas (*University of Toronto*)

- **FindYourFavorite: An Interactive System for Finding the User’s Favorite Tuple in the Database** ..... 2017  
Min Xie, Tianwen Chen, Raymond Chi-Wing Wong (*Hong Kong University of Science and Technology*)
- **PIClean: A Probabilistic and Interactive Data Cleaning System** ..... 2021  
Zhuoran Yu, Xu Chu (*Georgia Institute of Technology*)

## Tutorials

- **Towards Democratizing Relational Data Visualization** ..... 2025  
Nan Tang (*Qatar Foundation*), Eugene Wu (*Columbia University*), Guoliang Li (*Tsinghua University*)
- **Exploring the Data Wilderness through Examples** ..... 2031  
Davide Mottin (*Aarhus University*), Matteo Lissandrini (*Aalborg University*), Yannis Velegarakis (*Utrecht University*), Themis Palpanas (*Paris Descartes University*)
- **Database and Distributed Computing Foundations of Blockchains** ..... 2036  
Sujaya Maiyya, Victor Zakhary, Mohammad Javad Amiri, Divyakant Agrawal, Amr El Abbadi (*University of California, Santa Barbara*)
- **Classical and Contemporary Approaches to Big Time Series Forecasting** ..... 2042  
Christos Faloutsos (*Carnegie Mellon University & Amazon*), Jan Gasthaus, Tim Januschowski, Yuyang Wang (*AWS AI Labs*)
- **Data Pipelines for User Group Analytics** ..... 2048  
Behrooz Omidvar-Tehrani, Sihem Amer-Yahia (*University of Grenoble Alpes and CNRS*)
- **From Auto-tuning One Size Fits All to Self-designed and Learned Data-intensive Systems** ..... 2054  
Stratos Idreos (*Harvard University*), Tim Kraska (*Massachusetts Institute of Technology*)
- **Schemas and Types for JSON Data: From Theory to Practice** ..... 2060  
Mohamed-Amine Baazizi (*Sorbonne Université, LIP6 UMR 7606*), Dario Colazzo (*Université Paris-Dauphine, PSL Research University*), Giorgio Ghelli (*Università di Pisa*), Carlo Sartiani (*Università della Basilicata*)

## Workshop Summaries

- **GRADES-NDA 2019: Joint International Workshop on Graph Data Management Experiences & Systems and Network Data Analytics** ..... 2064  
Akhil Arora (*École Polytechnique Fédérale de Lausanne*), Arnab Bhattacharya (*IIT Kanpur*), George Fletcher (*TU Eindhoven*)
- **DEEM 2019: Workshop on Data Management for End-to-End Machine Learning** ..... 2066  
Sebastian Schelter (*New York University*), Neoklis Polyzotis (*Google*), Manasi Vartak (*Massachusetts Institute of Technology*), Stephan Seufert (*Amazon Research*)
- **DSMM’19: The 5<sup>th</sup> Workshop on Data Science for Macro-modeling with Financial and Economic Datasets** ..... 2068  
Douglas Burdick (*IBM Research - Almaden*), Rajasekar Krishnamurthy (*IBM Watson*), Louiqa Raschid (*University of Maryland*)
- **DaMoN 19: The 15th International Workshop on Data Management on New Hardware** ..... 2070  
Thomas Neumann (*Technische Universität München*), Ken Salem (*University of Waterloo*)
- **International Workshop on Human-In-the-Loop Data Analytics (HILDA)** ..... 2072  
Leilani Battle (*University of Maryland, College Park*), Surajit Chaudhuri, Arnab Nandi (*Microsoft*)
- **Overview of the 2<sup>nd</sup> International Workshop on Exploiting Artificial Intelligence Techniques for Data Management (aiDM’19)** ..... 2073  
Rajesh Bordawekar (*IBM T. J. Watson Research Center*), Oded Shmueli (*Technion*)
- **SBD’19: Fourth Edition of the International Workshop on Semantic Big Data** ..... 2075  
Sven Groppe (*University of Lübeck*), Le Gruenwald (*University of Oklahoma*)

**Author Index** ..... 2077



# SIGMOD 2019 Organization

## **SIGMOD Program Chair:**

Anastasia Ailamaki (EPFL, Switzerland)

## **SIGMOD Program Vice-Chairs:**

Amol Deshpande (University of Maryland, USA)

Tim Kraska (MIT, USA)

## **SIGMOD General Chairs:**

Peter Boncz (CWI & Vrije Universiteit Amsterdam, The Netherlands)

Stefan Manegold (CWI & Universiteit Leiden, The Netherlands)

## **SIGMOD Honorary Chair:**

Martin Kersten (CWI & Universiteit van Amsterdam, The Netherlands)

## **SIGMOD Local Arrangements/Organization Chairs:**

George Fletcher (Eindhoven University of Technology, The Netherlands)

Asterios Katsifodimos (Delft University of Technology, The Netherlands)

## **SIGMOD Sponsorship Chairs:**

Semih Salihoglu (University of Waterloo, Canada)

Tilman Rabl (HPI Potsdam, Germany)

## **SIGMOD Finance Chair:**

Hannes Mühleisen (CWI, Amsterdam & Vrije Universiteit Amsterdam, The Netherlands)

## **SIGMOD Registration Chair:**

Maurice van Keulen (University of Twente, Enschede, The Netherlands)

## **SIGMOD Core Program Committee:**

Ashraf Aboulnaga (Qatar Computing Research Institute, Qatar)

Azza Abouzied (New York University Abu Dhabi, United Arab Emirates)

Gustavo Alonso (ETH Zürich, Switzerland)

Peter Alvaro (University of California, Santa Cruz, USA)

## **SIGMOD Publicity/Social Media Chairs:**

Jan Hidders (Vrije Universiteit Brussel, Belgium)

Torsten Grust (Universität Tübingen, Germany)

## **SIGMOD Web/Information Chair:**

Holger Pirk (Imperial College London, UK)

## **SIGMOD Mentorship Chairs:**

Parth Nagarkar (New Mexico State University, USA),

Qiong Luo (Hong Kong University of Science & Technology, Hong Kong, China)

## **SIGMOD Proceedings Chair:**

Ziawasch Abedjan (TU Berlin, Germany)

## **SIGMOD Tutorials Chair:**

Ioana Manolescu (INRIA)

Hakan Hacigumus (Google, USA)

## **SIGMOD New Researcher Symposium Chairs:**

Katja Hose (Aalborg University, Denmark)

Spyros Blanas (Ohio State University, USA)

## **SIGMOD Workshops Chairs:**

Ihab Ilyas (University of Waterloo, CA)

Angela Bonifati (Lyon 1 University, FR)

Benny Kimelfeld (Technion, IL)

## **SIGMOD Undergrad Research Contest:**

Jana Giceva (Imperial College London, UK)

Eugene Wu (Columbia University, USA)

## **SIGMOD Programming Contest:**

Ravi Rajwar (Intel, USA)

Pinar Tözün (IT University of Copenhagen, Denmark)

Sourav Bhowmick (Nanyang Technological University, Singapore)

Carsten Binnig (Brown University, USA)

Matthias Boehm (IBM, USA)

Angela Bonifati (University of Lyon, France)

K. Selcuk Candan (Arizona State University, USA)

Gautam Das (University of Texas, Arlington, USA)

Daniel Deutch (Tel Aviv University, Israel)  
Johannes Gehrke (Microsoft, USA)  
Jonathan Goldstein (Microsoft, USA)  
Bill Howe (University of Washington, USA)  
Stratos Idreos (Harvard University, USA)  
Ihab Ilyas (University of Waterloo, Canada)  
Zachary Ives (University of Pennsylvania, USA)  
Chris Jermaine (Rice University, USA)  
Raghav Kaushik (Microsoft, USA)  
Martin Kersten (CWI, Netherlands)  
Wolfgang Lehner (Technical University of Dresden,  
Germany)  
Justin Levandoski (Microsoft, USA)  
Feifei Li (University of Utah, USA)  
Xuemin Lin (University of New South Wales,  
Australia)  
Alexandra Meliou (University of Massachusetts  
Amherst, USA)

**SIGMOD Regular Program Committee:**

Ioannis Alagiannis (Microsoft, USA)  
Foteini Alvanaki (EPFL, Switzerland)  
Nicolas Ancaux (French Institute for Research in  
Computer Science and Automation, France)  
Raja Appuswamy (Eurecom, France)  
Manos Athanassoulis (Harvard University, USA)  
Peter D Bailis (Stanford University, USA)  
Alex Beutel (Google, USA)  
Spyros Blanas (Ohio State University, USA)  
Alexander Boehm (SAP SE, Germany)  
Renata Borovica (University of Melbourne,  
Australia)  
Lei Cao (MIT, USA)  
Raul Castro Fernandez (MIT, USA)  
Badrish Chandramouli (Microsoft, USA)  
Lei Chen (University of Santo Tomas, Philippines)  
Shimin Chen (University of Chinese Academy of  
Sciences, China)  
Reynold Cheng (Univ. of Hong Kong, Hong Kong)  
James Cheng (Chinese Univ. of Hong Kong, Hong  
Kong)  
Xu Chu (Georgia Institute of Technology, USA)  
Carlo Curino (Microsoft, USA)  
Mahashweta Das (Hewlett-Packard, USA)  
Sudipto Das (Microsoft, USA)  
Chris De Sa (Cornell University, USA)

Paolo Papotti (Eurecom, France)  
Srinivasan Parthasarathy (Ohio State University,  
USA)  
Holger Pirk (Imperial College London, UK)  
Evaggelia Pitoura (University of Ioannina, Greece)  
Theodoros Rekatsinas (Stanford University, USA)  
Dan Suciu (University of Washington, USA)  
Jens Teubner (Technical University of Dortmund,  
Germany)  
Marcos Antonio Vaz Salles (University of  
Copenhagen, Denmark)  
Jiannan Wang (Simon Fraser University, Canada)  
Jun Yang (Duke University, USA)  
Ke Yi (Hong Kong University of Science &  
Technology, Hong Kong)  
Nan Zhang (Pennsylvania State University, USA)

Dong Deng (Massachusetts Institute of Technology,  
USA)  
AnHai Doan (University of Washington, USA)  
Eduard Dragut (Temple University College of  
Education, USA)  
Georgios Fakas (Uppsala University, Sweden)  
Wenfei Fan (University of Edinburgh, UK)  
Alan Fekete (University of Sydney, Australia)  
George Fletcher (Eindhoven University of  
Technology, Netherlands)  
Wolfgang Gatterbauer (Northeastern Univ., USA)  
Jana Giceva (Imperial College London, UK)  
Boris Glavic (Illinois Institute of Technology, USA)  
Torsten Grust (University of Tübingen, Germany)  
Alon Halevy (Megagon Labs, USA)  
Michael Hay (Colgate University, USA)  
Melanie Herschel (University of Stuttgart,  
Germany)  
Vagelis Hristidis (UC Riverside, USA)  
H. V. Jagadish (University of Michigan, USA)  
Alekh Jindal (Microsoft, USA)  
Ryan Johnson (Amazon Web Services)  
Eser Kandogan (IBM, USA)  
Manos Karpathiotakis (EPFL, Switzerland)  
Alfons Kemper (Technical University of Munich,  
Germany)

Oliver Kennedy (University at Buffalo, USA)  
Hideaki Kimura (Oracle, USA)  
Donald Kossmann (Microsoft, USA)  
Paraschos Koutris (University of Washington, USA)  
Rajasekar Krishnamurthy (IBM, USA)  
Arun Kumar (UC San Diego, USA)  
Viktor Leis (Technical University of Munich, Germany)  
Guoliang Li (Tsinghua University, China)  
Chengkai Li (University of Texas, Arlington, USA)  
Ashwin Machanavajjhala (Duke University, USA)  
Sam Madden (MIT, USA)  
Nikos Mamoulis (University of Ioannina, Greece)  
Norman May (SAP SE, Germany)  
Gerome Miklau (University of Massachusetts Amherst, USA)  
Ingo Müller (ETH Zürich, Switzerland)  
Thomas Neumann (Technical University of Munich, Germany)  
Milos Nikolis (University of Oxford, UK)  
Ismail Oukid (SAP SE, Germany)  
Tamer Ozsu (University of Waterloo, Canada)  
Themis Palpanas (Paris Descartes, France)  
Ippokratis Pandis (Amazon, USA)  
Charalampos Papamanthou (University of Maryland, USA)  
Andy Pavlo (Carnegie Mellon University, USA)  
Danica Porobic (Oracle, USA)  
Dan Ports (University of Washington, USA)  
Iraklis Psaroudakis (Oracle, Switzerland)  
Ravi Ramamurthy (Microsoft, USA)  
Mary Roth (IBM, USA)

**SIGMOD External Program Committee:**

Panagiotis Antonopoulos (Microsoft, USA)  
Akhil Arora (EPFL, Switzerland)  
Magdalena Balazinska (University of Washington, USA)  
Stella Giannakopoulou (EPFL, Switzerland)  
Georgia Koutrika (Athena Research Center, Greece)  
Barzan Mozafari (University of Michigan, USA)  
Rachel Pottinger (University of British Columbia, Canada)  
Babak Salimi (University of Washington, USA)  
Hanghang Tong (Arizona State University, USA)  
Charalampos Tsourakakis (Boston University, USA)

Sudeepa Roy (Duke University, USA)  
Sudip Roy (Google, USA)  
Florin Rusu (UC Merced, USA)  
Mohammad Sadoghi (UC Davis, USA)  
Yasushi Sakurai (Kumamoto University, Japan)  
Ken Salem (University of Waterloo, Canada)  
Semih Salihoglu (University of Waterloo, Canada)  
Mohamed Sarwat (Arizona State University, USA)  
**SIGMOD Regular Program Committee (continued):**  
Sebastian Schelter (New York University, USA)  
Prithviraj Sen (IBM, USA)  
Divesh Srivastava (AT&T, USA)  
Radu Stoica (IBM, Switzerland)  
Julia Stoyanovich (Drexel University, USA)  
Wang-Chiew Tan (Megagon Labs, USA)  
Nan Tang (Qatar Computing Research Institute, Qatar)  
Yufei Tao (Chinese University of Hong Kong, Hong Kong)  
Nesime Tatbul (MIT, USA)  
Farhan Tauheed (Oracle, USA)  
Arash Termehchy (Oregon State University, USA)  
Martin Theobald (University of Luxembourg, Luxembourg)  
Immanuel Trummer (Cornell University, USA)  
Eugene Wu (Columbia University, USA)  
Da Yan (University of Alabama, Birmingham, USA)  
Cong Yu (Google, USA)  
Ce Zhang (ETH Zürich, Switzerland)  
Yongluan Zhou (University of Copenhagen, Denmark)  
Wenchao Zhou (Georgetown University, USA)

Nalini Venkatasubramanian (University of California, Irvine, USA)  
Yinghui Wu (Washington State University, USA)  
Yaoliang Yu (University of Waterloo, Canada)  
Matei Zaharia (Stanford University, USA)  
Emanuel Zraggen (Massachusetts Institute of Technology, USA)  
Kostas Zoumpatianos (Harvard University, USA)

**SIGMOD Assisting Program Committee:**

Peng Cheng (Hong Kong University of Science & Technology, Hong Kong)  
Stefania Dumbrava (INRIA, France)  
Cibele Freire (Wellesley College, USA)

Sainyam Galhotra (University of Massachusetts, Amherst, USA)  
Zeynep Korkmaz (University of Waterloo, USA)  
Anil Pacaci (University of Waterloo, USA)  
Aida Sheshbolouki (University of Waterloo, USA)  
Yue Wang (Hong Kong University of Science & Technology, Hong Kong)  
Xiaofei Zhang (University of Memphis, USA)  
Libin Zheng (Hong Kong University of Science & Technology, Hong Kong)

**SIGMOD Industrial Track PC Chairs:**

Lyublena Antova (Datometry)  
Jignesh Patel (UW Madison, USA)

**SIGMOD Industrial Track PC Members:**

Alexander Shraer (Apple, USA)  
Allison Holloway (Oracle, USA)

**SIGMOD Demo Track PC Members:**

Alexandros Koliouisis (Imperial College London, UK)  
Anisoara Nica (SAP SE, Waterloo)  
Bertram Ludaescher (University of Illinois, USA)  
Boris Glavic (Illinois Institute of Technology, USA)  
Danica Porobic (Oracle)  
Dong-Wan Choi (Inha University)  
Eric Lo (Chinese University of Hong Kong, China)  
Fabian Suchanek (Telecom Paris Tech Univ., France)  
Fabio Porto (LNCC)  
Florin Rusu (UC Merced, USA)  
Georgia Koutrika (Athena Research Center)  
Guoliang Li (Tsinghua University)  
Hannes M<sup>o</sup>hleisen (CWI)  
Huiping Cao (New Mexico State University, USA)  
Iman Elghandour (University of Copenhagen)  
Jaydeep Sen (IBM Research AI, USA)  
Johann Gamper (Free Univ. of Bozen-Bolzano, Italy)  
Kai Zheng (University of Electronic Science & Technology of China)  
Katja Hose (Aalborg University)  
Kyriakos Mouratidis (Singapore Management University)  
Lei Chen (Hong Kong University of Science & Technology)  
Manos Karpathiotakis (Facebook)  
Maria Luisa Sapino (U. Torino)

Avrilia Floratou (Microsoft, USA)  
Danica Porobic (Oracle, USA)  
Entong Shen (Amazon Web Services, USA)  
Jianjun Chen (Huawei US Research Center, USA)  
Josiane Xavier Parreira (Siemens AG Austria)  
Marco Serafini (University of Massachusetts Amherst, USA)  
Mark Callaghan (Facebook, USA)  
Markus Weimer (Microsoft)  
Mohamed Soliman (Datometry, USA)  
Per-Ake Larson (University of Waterloo, Canada)  
Vaishnavi Sashikanth (Google, USA)  
Ying Zhang (MonetDB Solutions, Netherlands)

**SIGMOD Demo Track PC Chairs:**

Thomas Heinis (Imperial College London, UK)  
Fatma Ozcan (IBM Almaden, USA)

Maya Ramanath (IIT Delhi)  
Mitesh Vasa (IBM Watson - Almaden, USA)  
Mohamed Eltabakh (Teradata)  
Mohammad Sadoghi (UC Davis, USA)  
Mustafa Canim (IBM Research)  
Nan Tang (Qatar Computing Research Institute, HBKU)  
Niketan Pansare (IBM Almaden, USA)  
Odysseas Papapetrou (EPFL)  
Olga Papaemmanouil (Brandeis University)  
Parth Nagarkar (NMSU)  
Pelin Angin (METU, Turkey)  
Peter Fischer (University of Augsburg)  
Renata Borovica-Gajic (University of Melbourne)  
Ricardo Torres (IC-Unicamp)  
Semih Salihoglu (University of Waterloo, Canada)  
Senjuti Basu Roy (NJ Institute of Technology, USA)  
Stefanie Scherzinger (OTH Regensburg)  
Sudip Roy (Google)  
Till Westmann (Couchbase)  
Vasilis Eftymiou (IBM Research - Almaden, USA)  
Verena Kantere (University of Ottawa, Canada)  
Xiaochun Yang (Northeastern University, USA)  
Xu Chu (GATECH, USA)  
Yongxin Tong (Beihang University)

# SIGMOD 2019 Sponsors & Supporters

Sponsors:



Diamond Supporter:



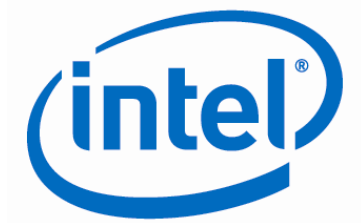
Platinum Supporters:



Gold Supporters:



IBM Research AI



Megagon Labs



Silver Supporters:



Startup Supporters:



Platinum Publisher:



Gold Publisher:



Silver Publishers:



Student Support:



Institutional Support:



Centrum Wiskunde & Informatica