



Association for  
Computing Machinery

*Advancing Computing as a Science & Profession*

## NEWS RELEASE

**Contact:** Jim Ormond  
212-626-0505  
[ormond@hq.acm.org](mailto:ormond@hq.acm.org)

### **ACM SIGMOD/PODS CONFERENCE EXPLORES FUTURE APPLICATIONS OF ADVANCED DATA MANAGEMENT**

*Virtual Conference to Feature Keynotes, Tutorials and Programming Contests  
Involving the World's Foremost Data Scientists*

**New York, NY, June 10, 2020** – The Association for Computing Machinery's Special Interest Group on Management of Data (ACM SIGMOD), together with PODS, the premier international conference on the theoretical aspects of database systems, will host the [2020 ACM SIGMOD/PODS](#) conference from June 14-June 19, 2020. The conference will be held virtually, with all the sessions available online.

The conference is a leading international forum for database researchers, practitioners, developers, and users to explore cutting-edge ideas and results, and to exchange techniques, tools, and experiences in all aspects of data management.

Data management plays a vital role in the development and efficacy of machine learning, biometric analysis and remote sensors, among other technologies. And as data continues to be collected at an unprecedented scale, a growing number of industries rely significantly on data management to inform their future operations and prosperity.

“We are at an unprecedented moment in the history of technology, but more specifically in the history of data management,” said SIGMOD Program Co-chairs AnHai Doan, University of Wisconsin, Madison, and Wang-Chiew Tan, Megagon Labs. “Autonomous vehicles, chatbots, streaming television recommendations—these are just a few examples of the technologies that have been enabled by creative data management over the past few years. This gathering provides us with an opportunity to share in our creativity and think about where the next groundbreaking application of data management might lie.”

“This year's program includes a wide selection of research papers reflecting the conference's tradition of being the premier forum on the theoretical foundations of data management,” added PODS Program Chair Yufei Tao, Chinese University of Hong Kong. “At the same time, we continue to broaden our scope. 2020 PODS attendees will learn about the latest research in data management and machine learning, data privacy and security, as well as data ethics.”

## **2020 ACM SIGMOD/PODS HIGHLIGHTS**

### **KEYNOTE TALKS**

#### **MODS KEYNOTES**

##### **“Grand Challenges in Exploring, Understanding, and Searching a Billion Datasets”**

###### **Part 1: “When the Web is Your Data Lake: Creating a Search Engine for Datasets on the Web”**

*Natasha Noy, Google*

In this keynote, Noy will discuss her work on Dataset Search at Google. Dataset Search provides search capabilities over potentially all dataset repositories on the Web. She will talk about the open ecosystem for describing and citing datasets that we hope to encourage and the technical details on how the Google team went about building Dataset Search.

###### **Part 2: “The Challenge of Building Effective Data Lakes”**

*Awez Syed, Databricks*

There has been a rapid rise in the popularity of data lakes as the data infrastructure for modern analytics and data science. In this talk, Syed will describe the real-world implementation patterns of data lakes and give an overview of the many open challenges in deploying successful, enterprise-scale data lakes.

##### **“Systems and ML: When the Sum Is Greater than Its Parts”**

*Ion Stoica, University of California, Berkeley*

The research at the intersection between machine learning (ML) and systems has the potential to fuel the innovation over the next decade. When used together, ML and systems can lead to a virtuous cycle, where systems accelerate ML algorithms, and ML algorithms lead to faster systems. In this talk, Stoica will discuss the research his team has done at RISELab (UC Berkeley) to enable this virtuous cycle.

#### **PODS KEYNOTE**

##### **"word2vec, node2vec, graph2vec, X2vec: Towards a Theory of Vector Embeddings of Structured Data"**

*Martin Grohe, RWTH Aachen University*

Vector representations of graphs and relational data, whether hand-crafted feature vectors or learned representations, enable us to apply standard data analysis and machine learning techniques to these forms of structured data. Starting with a brief overview of embedding techniques commonly used in practice, in this talk Grohe will discuss theoretical ideas that have proved useful for analyzing and designing practical vector embeddings and that may provide a foundation for future work in the area.

#### **MODS AWARDS**

##### **Edgar F. Codd Innovations Award**

###### **Beng Chin Ooi, National University of Singapore**

Ooi's research interests include database systems, distributed and decentralized systems, machine learning and large-scale analytics, in the aspects of system architectures, performance issues, security, accuracy and correctness. He has been building large scale data systems for supporting advanced analytics and has released a number of them as open sources, which include Apache SINGA, a

distributed deep learning platform that has been designed based on the principles of scalability, efficiency, elasticity and usability.

### **SIGMOD Best Paper Award**

#### **“Pump Up the Volume: Processing Large Data on GPUs with Fast Interconnects” (Co-Winner)**

*Clemens Lutz, TU Berlin; Sebastian Breß, Snowflake Computing, TU Berlin; Steffen Zeuch, TU Berlin, IAM Group; Tilmann Rabl, Hasso Plattner Institute, University of Potsdam; Volker Markl, TU Berlin*

GPUs have long been discussed as accelerators for database query processing because of their high processing power and memory bandwidth. However, two main challenges limit the utility of GPUs for large-scale data processing: (1) the on-board memory capacity is too small to store large data sets, yet (2) the interconnect bandwidth to CPU main memory is insufficient for ad hoc data transfers. In this paper, the authors investigate how a fast interconnect can resolve these scalability limitations using the example of NVLink 2.0. NVLink 2.0 is a new interconnect technology that links dedicated GPUs to a CPU.

#### **“ShapeSearch: A Flexible and Efficient System for Shape-based Exploration of Trendlines” (Co-Winner)**

*Tarique Siddiqui, University of Illinois, Urbana-Champaign; Paul Luh, University of Wisconsin-Madison; Zesheng Wang, Roblox Corporation; Karrie Karahalios, University of Illinois; Aditya Parameswaran, University of California, Berkeley.*

Identifying trendline visualizations with desired patterns is a common and fundamental data exploration task. The authors propose ShapeSearch, an efficient and flexible pattern-searching tool, that enables the search for desired patterns via multiple mechanisms: sketch, natural-language, and visual regular expressions.

### **SIGMOD Jim Gray PhD Dissertation Award**

**This award is given to the best database systems dissertation from the previous year.**

#### **“High Performance Multi-core Transaction Processing via Deterministic Execution”**

*Jose Faleiro, Researcher, Microsoft*

This dissertation proposes and explores the use of deterministic execution to address challenges in database systems arising from new hardware, changes in application and deployment environments and scalability issues. Deterministic execution ensures that the database’s final state is always determined by its input list of transactions; in other words, the input list of transactions is the same as the total order of transactions that determines the database’s state.

### **Jim Gray PhD Dissertation Honorable Mention**

#### **“Effective Data Versioning for Collaborative Data Analytics”**

*Silu Huang, Microsoft Research Redmond Lab*

With the massive proliferation of datasets in a variety of sectors, data science teams in these sectors spend vast amounts of time collaboratively constructing, curating, and analyzing these datasets. However, no existing systems enable us to effectively store, track, and query these versioned datasets, leading to massive redundancy in versioned data storage and making true collaboration and sharing impossible. In her thesis, Huang proposes solutions for versioned data management for collaborative data analytics.

## **PODS AWARDS**

### **PODS Best Paper Award**

#### **“A Framework for Adversarially Robust Streaming Algorithms”**

*Omri Ben-Eliezer, Tel Aviv University; Rajesh Jayaram, David P. Woodruff, Carnegie Mellon University; Eylon Yogev, Boston University and Tel Aviv University*

In this paper, the authors investigate the adversarial robustness of streaming algorithms. They also develop several generic tools allowing one to efficiently transform a non-robust streaming algorithm into a robust one in various scenarios.

### **PODS Alberto Mendelzon Test of Time Award**

#### **“Optimizing Linear Counting Queries under Differential Privacy”**

*Chaol Li, Google; Michael Hay, Colgate University, Tumult Labs; Vibhor Rastogi, Facebook; Jerome Miklau, University of Massachusetts, Amherst; Andrew McGregor, University of Massachusetts, Amherst*

Differential privacy is a robust privacy standard that has been successfully applied to a range of data analysis tasks. This paper, published in PODS 2010, proposes a matrix-based two-phase algorithm that has been fundamental to further developments in the field. Most notably is the adoption of the approach as the data-privacy mechanism in the United States 2020 Census. The paper received about 300 citations, both in database theory and data systems papers, underscoring its additional impact.

Additional papers, tutorials, research sessions and in-depth expositions will be presented throughout the multi-day conference. For a complete list of papers and a full schedule of activities, please visit: [2020 ACM SIGMOD/PODS](#).

### **About SIGMOD**

[The ACM Special Interest Group on Management of Data \(SIGMOD\)](#) is concerned with the principles, techniques and applications of database management systems and data management technology. Our members include software developers, academic and industrial researchers, practitioners, users, and students. SIGMOD sponsors the annual SIGMOD/PODS conference, one of the most important and selective in the field.

### **About ACM**

[ACM, the Association for Computing Machinery](#), is the world’s largest educational and scientific computing society, uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field’s challenges. ACM strengthens the computing profession’s collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

###