

Editor's Notes

Welcome to the September 2020 issue of the ACM SIGMOD Record!

This issue starts with the Database Principles column featuring an article by Hu and Yi. The authors survey recent algorithms for join evaluation under the Massively Parallel Computation (MPC) model exemplified by the MapReduce and Spark frameworks. The article provides a formal perspective on processing join-aggregate queries over annotated relations; this formalization covers many query types, including standard aggregate queries and conjunctive queries. The authors study the data complexity of algorithms for multi-round processing of joins under the MPC model, conveying crisp and practically applicable characterizations of the algorithms by cases. The article also provides lower-bound results, whose importance is in ruling out the possibility of certain types of join algorithms for certain query classes. The formal results discussed in the article are summarized in two detailed tables, and the exposition contains illustrations featuring specific join examples. The article also presents open problems and provides a discussion of related and further work.

The Surveys column features an article by Hameed and Naumann that studies the state of the art in data-preparation tools, including the accompanying research and development opportunities. Here, data preparation is understood as a collection of processes for preparing raw data for ingestion into downstream applications, such as data-analytics tools and data-management systems. The article provides a thorough classification of data-preparation features and commercial tools, a wealth of examples, and extensive documentation and evaluation results of the features on three data sets obtained from public data repositories. The recommendations include an articulation of the features and abilities that are still lacking in the state-of-the-art tools, as well as a discussion of challenging research problems in the emerging area of data preparation.

The Distinguished Profiles column features Goetz Graefe, recipient of the SIGMOD Innovations Award, the SIGMOD Test of Time Award, the ICDE Distinguished Paper Award, and the ACM Software System Award, all for his work on query processing. Goetz's Ph.D. is from the University of Wisconsin-Madison. He was an HP Fellow, and currently works for Google. In this interview, Goetz talks about the research topics to which he has made contributions over the years, and about how traditional results can be extended to apply in new contexts. He shares his thoughts about open problems in query optimization, outlining ideas for sources of potential solutions for lack of robustness in query performance. Goetz talks about the success of the Cascades framework, as well as about how he would build a query optimizer today from scratch. He provides insights on a range of other topics, including thoughts on ways in which academia, industry, and industry research labs do research, as well as on his teaching at Dagstuhl. Goetz shares ideas on promising long-term hard systems problems, discusses what he would change about himself as a computer-science researcher, and provides (surprising) advice for fledgling or midcareer database researchers.

The Industry Perspectives column features an article by Jindal about lessons that he has learned from the CloudViews project at the Gray Systems Lab at Microsoft. While using his experience from the CloudViews project as a detailed running example, the author provides excellent advice that can be readily used by any new researchers to set themselves up for success in a product group. The article walks the reader through all the steps of the process, from initial selection of a promising product and of the right problem to work on, all the way to publishing and to then moving on. The author discusses common challenges and pitfalls, and shares excellent advice, including tips on building collaborations, on success metrics, and on publishing. Much of the advice in this engaging

and enjoyable article would also be of interest to beginner researchers in general, both in graduate school and in the workplace, including in academia.

This issue features two reports. First, it is our pleasure to include a detailed report on running the SIGMOD/PODS conference in June 2020 in the online-only mode due to the COVID-19 situation. Preparing and running the conference was challenging, as it had originally been planned to take place as a “live” event in Portland, Oregon, with many months of work already invested into realizing that plan. The organizers of the conference detail in the article the sequence of the events and decisions in the process of rearranging the conference into the fully remote-access mode. (At some point, there was even discussion of canceling the conference altogether.) The article presents experiences of the conference organizers, information gathered in the process, quotes from participants, and recommendations on running analogous events. The authors also outline and discuss the most-requested features that ended up not being supported by SIGMOD/PODS 2020, and provide statistics from the post-conference survey and from logs of Zoom sessions during the conference. Lessons from the conference would be of interest to those who organize online-only conferences or conferences with an online component. The second article in the Reports column, by Stonebraker, Mattson, Kraska, and Gadepally, reports on the outcomes of the fourth annual workshop on Polystores (POLY’19), which took place in Los Angeles, CA USA in August 2019 in conjunction with the VLDB conference. The article focuses on the theme of the workshop that explored the implications of data-privacy regulations such as General Data Privacy Regulations (GDPR) in the context of heterogeneous data and data-management systems. The authors outline some major approaches and directions presented during the data-privacy portion of the POLY’19 workshop, specifically in the scope of the GDPR right to be forgotten and of support for restrictions on data usage known as “purposes.” The authors also outline the unaddressed issues and emerging research directions. The workshop materials are available online; in addition, the authors point the readers to the POLY’20 workshop, which also featured data privacy in its program.

On behalf of the SIGMOD Record Editorial board, I hope that you enjoy reading the September 2020 issue of the SIGMOD Record!

Your submissions to the SIGMOD Record are welcome via the submission site:

<https://mc.manuscriptcentral.com/sigmodrecord>

Prior to submission, please read the Editorial Policy on the SIGMOD Record’s website:

<https://sigmodrecord.org/sigmod-record-editorial-policy/>

Rada Chirkova

September 2020

Past SIGMOD Record Editors:

Yanlei Diao (2014-2019)	Ioana Manolescu (2009-2013)	Alexandros Labrinidis (2007-2009)
Mario Nascimento (2005-2007)	Ling Liu (2000-2004)	Michael Franklin (1996-2000)
Jennifer Widom (1995-1996)	Arie Segev (1989-1995)	Margaret H. Dunham (1986-1988)
Jon D. Clark (1984-1985)	Thomas J. Cook (1981-1983)	Douglas S. Kerr (1976-1978)
Randall Rustin (1974-1975)	Daniel O’Connell (1971-1973)	Harrison R. Morse (1969)