

The Next 5 Years: What Opportunities Should the Database Community Seize to Maximize its Impact?

Magda Balazinska
University of Washington, Seattle

Surajit Chaudhuri
Microsoft Research

Anastasia Ailamaki
EPFL

Juliana Freire
NYU

Sailesh Krishnamurthy
Google

Michael Stonebraker
MIT

Panel Overview

The database research community has been spectacularly successful in impacting the industry and academia since the invention of the relational model. Examples of innovation in the last decade include columnar storage for data analytic platforms, cloud data services, HTAP systems, and a new generation of data wrangling systems. Nonetheless, critical self-assessment by the community and identifying key opportunities for the future is essential if we are to continue the tradition of impactful research.

In the Fall of 2018, following a long tradition that dates back to 1988 [1], and five years after the last such meeting [2], a group of approximately thirty database researchers gathered at the University of Washington, Seattle for two days to discuss the opportunities we have as a community for impactful research. A report from that meeting is now available [3]. The discussions in the Seattle meeting focused not only on technical challenges and opportunities but also on topics related to how we organize ourselves as a community.

This SIGMOD panel will provide a forum for the broader database community to review and debate the findings from the Seattle Report on Database Research [3] as well as to identify other considerations that need to be taken into account.

Permission to make digital or hard copies of part or all 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. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

SIGMOD '20, June 14–19, 2020, Portland, OR, USA.

© 2020 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6735-6/20/06.

DOI: <https://doi.org/10.1145/3318464.3394837>

Some of the research challenges from the report that we will discuss in the panel include:

- *Data to Insight*: How can we leverage our expertise in data cleaning and data integration for accelerating discovery of insights? How can we make it easy for data scientists to use both query systems and machine learning inferences in a seamless way?
- *Data Governance*: How do we support tracking of provenance across data stores efficiently and at scale? What breakthroughs are needed to support data privacy in data platforms? What are the unique challenges in ethical data science?
- *Cloud Database Services*: How does the architecture of database services adapt to and take full advantage of disaggregation of resources in the cloud? How do we architect hybrid clouds that span on-premise data systems and cloud data services? What should be the next generation “serverless” cloud data service architecture?
- *Novel Scenarios for Database Systems*: What are the unique challenges of data lakes that go beyond traditional SQL data warehouse and today’s big data architectures? What are the implications of edge computing?
- *Implications of Hardware Trends*: How will the era of heterogeneous computation and modern hardware impact the architecture of our engines?

On the community front, we will discuss the changes that may be needed in our conferences and the reviewing process to encourage early stage sharing of potentially impactful ideas. In the era of cloud services, another important question is understanding how the academia and the industry can collaborate as database services from the public clouds continue to gain popularity. Last but not the least, we will discuss our role in data science education.

Role of Panel Members

The moderators will open the panel with a summary of the Seattle Report on Database Research and will frame some of the open questions (e.g., as discussed earlier in this document) for the panel members.

Anastasia Ailamaki will wear two hats in the panel. In addition to her perspectives on challenges in systems research, as the 2019 SIGMOD Program Committee Chair, she can provide her view from the ground on “what is working” and “what is broken” in our peer review process.

Juliana Freire has wide ranging experience in data science and is deeply involved on many innovative data-driven applications. Thus, her thoughts on how best to influence data science will be valuable. As the chair of the SIGMOD executive, she can comment on the evolution of database conferences.

Sailesh Krishnamurthy brings a wealth of expertise in cloud systems and will address the key challenges in evolution of cloud data services. He will be the voice of the “industry” and will also reflect on how best the academia can partner with cloud providers to have impact.

Mike Stonebraker will reflect on how to accelerate innovation and systems research. He will also share his view of our role in Data to Insights pipeline. His wide breadth of experience across research as well as the start-up ecosystem enables him to provide a unique perspective.

Expected outcome

We expect the panel to serve as a forum to update the broader community on the Seattle Report on Database Research [3] and seek their criticism and feedback. The technical part of discussions will also result in a frank assessment of topics where we are “polishing the round ball” and where there are big opportunities.

We also hope to identify concrete action items that the SIGMOD and VLDB Executive can take to improve the state of the community. Last but not the least, we want to provide the academia with a better understanding of

different avenues to have impact, including collaboration with the industry.

In summary, we hope that the Seattle Report on Database Research [3] and the discussion in the panel with the broader community will help database researchers and industry participants identify key opportunities that we should seize as the community for an impactful future.

Moderators:

Magdalena Balazinska is Professor and Director of the Paul G. Allen School of Computer Science & Engineering at the University of Washington. Magdalena's research interests are in the field of database management systems. Her current research focuses on data management for data science, big data systems, cloud computing, and image and video analytics. She also served as Co-Editor-in-Chief for Volume 13 of the Proceedings of the Very Large Data Bases Endowment (PVLDB) journal and as PC co-chair for the corresponding, prestigious VLDB'20 conference. Magdalena holds a Ph.D. from the Massachusetts Institute of Technology (2006). Magdalena received the inaugural VLDB Women in Database Research Award (2016) for her work on scalable distributed data systems. She also received an ACM SIGMOD Test-of-Time Award (2017) for her work on fault-tolerant distributed stream processing and a 10-year most influential paper award (2010) from her earlier work on reengineering software clones. Magdalena received an NSF CAREER Award (2009), the UW CSE ACM Teaching Award (2013), the Jean Loup Baer Career Development Professorship in Computer Science and Engineering (2014-2017), two Google Research Awards (2011 and 2018), an HP Labs Research Innovation Award (2009 and 2010), a Rogel Faculty Support Award (2006), a Microsoft Research Graduate Fellowship (2003-2005), and multiple best-paper (and “best of”) awards.

Surajit Chaudhuri is a Distinguished Scientist at Microsoft Research and leads the Data Management, Exploration and Mining group. His current areas of interest are data analytics for big data platforms, self-manageability, and cloud database services. Working with his colleagues in Microsoft Research, he helped incorporate the Index Tuning Wizard (and subsequently the Database Engine Tuning Advisor) and Data Cleaning technology in Microsoft SQL Server. Surajit is an ACM Fellow, a recipient of the ACM SIGMOD Edgar F. Codd Innovations Award, ACM SIGMOD Contributions Award, a VLDB 10-year Best Paper Award, and an IEEE

Data Engineering Influential Paper Award. Surajit received his Ph.D. from Stanford University.

Panelists:

Anastasia Ailamaki is a Professor of Computer and Communication Sciences at EPFL as well as the CEO and co-founder of RAW Labs SA, a Swiss company that develops enables digital transformation for enterprises through real-time analysis of heterogeneous big data. Previously, she was on the faculty of the Computer Science Department at CMU, where she held the Finmeccanica endowed chair. She has received the 2019 ACM SIGMOD Edgar F. Codd Innovations Award, the 2019 EDBT Test of Time award, the 2018 Nemitsas Prize in Computer Science, an ERC Consolidator Award (2013), the European Young Investigator Award from the European Science Foundation (2007), an Alfred P. Sloan Research Fellowship (2005), and ten best-paper awards in database, storage, and computer architecture conferences. She is an ACM fellow, an IEEE fellow, and an elected member of the Swiss, the Belgian, and the Cypriot National Research Councils.

Juliana Freire is a Professor of Computer Science and Data Science at New York University. She is the elected chair of the ACM Special Interest Group on Management of Data (SIGMOD) and a council member of the Computing Research Association's Computing Community Consortium (CCC). She was the lead investigator and executive director of the NYU Moore-Sloan Data Science Environment. Her research interests are in large-scale data analysis, curation and integration, visualization, provenance management, and web information discovery. She has made fundamental contributions to data management methods and tools that address problems introduced by emerging applications including urban analytics and computational reproducibility. Freire has published over 200 technical papers (including 8 award-winning papers), several open-source systems, and is an inventor of 12 U.S. patents. She is an ACM Fellow and a recipient of an NSF CAREER, two IBM Faculty awards, and a Google Faculty Research award. Her research has been funded by the National Science Foundation, DARPA, Department of Energy, National Institutes of Health, Sloan Foundation, Gordon and Betty Moore Foundation, W. M. Keck Foundation, Google, Amazon, AT&T Research, Microsoft Research, Yahoo! and IBM. She received M.Sc. and Ph.D. degrees in computer science from the State University of New York at Stony Brook.

Sailesh Krishnamurthy is a Senior Director of Engineering in Google, overseeing cloud-native database services in Google. Previously, he was an Engineering Manager at Amazon Web Services (AWS) where he led engineering for the Amazon Aurora database engine

compatible with MySQL. Prior to AWS, Sailesh was at Cisco Systems via the acquisition of Truviso, a real-time streaming data analytics software company that he co-founded to commercialize his prior academic research.

Michael Stonebraker has been a pioneer of data base research and technology for more than a quarter of a century. He was the main architect of the INGRES relational DBMS, and the object-relational DBMS, POSTGRES. These prototypes were developed at the University of California at Berkeley where Stonebraker was a Professor of Computer Science for twenty-five years. More recently at M.I.T. he was a co-architect of the Aurora/Borealis stream processing engine, the C-Store column-oriented DBMS, the H-Store transaction processing engine, the SciDB array DBMS, and the Data Tamer data curation system. Presently he serves as Chief Technology Officer of Paradigm4 and Tamr, Inc. Professor Stonebraker was awarded the ACM System Software Award in 1992 for his work on INGRES. Additionally, he was awarded the first annual SIGMOD Innovation award in 1994 and was elected to the National Academy of Engineering in 1997. He was awarded the IEEE John Von Neumann award in 2005 and the ACM Turing award in 2014. Stonebraker is an Adjunct Professor of Computer Science at M.I.T. and is continuing to work on big data,

References

- [1] Philip A. Bernstein, Umeshwar Dayal, David J. DeWitt, Dieter Gawlick, Jim Gray, Matthias Jarke, Bruce G. Lindsay, Peter C. Lockemann, David Maier, Erich J. Neuhold, Andreas Reuter, Lawrence A. Rowe, Hans-Jörg Schek, Joachim W. Schmidt, Michael Schrefl, Michael Stonebraker: Future Directions in DBMS Research - The Laguna Beach Participants. SIGMOD Rec. 18(1): 17-26 (1989)
- [2] Daniel J. Abadi, Rakesh Agrawal, Anastasia Ailamaki, Magdalena Balazinska, Philip A. Bernstein, Michael J. Carey, Surajit Chaudhuri, Jeffrey Dean, AnHai Doan, Michael J. Franklin, Johannes Gehrke, Laura M. Haas, Alon Y. Halevy, Joseph M. Hellerstein, Yannis E. Ioannidis, H. V. Jagadish, Donald Kossmann, Samuel Madden, Sharad Mehrotra, Tova Milo, Jeffrey F. Naughton, Raghu Ramakrishnan, Volker Markl, Christopher Olston, Beng Chin Ooi, Christopher Ré, Dan Suciu, Michael Stonebraker, Todd Walter, Jennifer Widom: The Beckman Report on Database Research. SIGMOD Rec. 43(3): 61-70 (2014)
- [3] Daniel Abadi, Anastasia Ailamaki, David Andersen, Peter Bailis, Magdalena Balazinska, Philip A. Bernstein, Peter A. Boncz, Surajit Chaudhuri, Alvin Cheung, AnHai Doan, Luna Dong, Michael J. Franklin, Juliana Freire, Alon Y. Halevy, Joseph M. Hellerstein, Stratos Idreos, Donald Kossmann, Tim Kraska, Sailesh Krishnamurthy, Volker Markl, Sergey Melnik, Tova Milo, C. Mohan, Thomas Neumann, Beng Chin Ooi, Fatma Ozcan, Jignesh Patel, Andrew Pavlo, Raluca A. Popa, Raghu Ramakrishnan, Christopher Ré, Michael Stonebraker, Dan Suciu: The Seattle Report on Database Research. SIGMOD Rec. 48(4): 44-53 (2019)