

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
School of Engineering Faculty Personnel Record

Date: August 4, 2020

Name: Tim Kraska
Department: Electrical Engineering and
Computer Science

1. Date of Birth: September 1980
2. Citizenship: Germany (U.S. permanent resident)

3. Education:

<u>School</u>	<u>Degree</u>	<u>Date</u>
Westfälische Wilhelms-Universität Münster, Germany	B.S.	2004
University of Sydney, Australia	M.S.	2006
Westfälische Wilhelms-Universität Münster, Germany	M.S.	2006
ETH Zurich, Switzerland	Ph.D.	2010

4. Title of Thesis for Most Advanced Degree:
Building Database Applications in the Cloud

5. Principal Fields of Interest:

Data Management, Machine Learning for Systems, Systems for Data Science

6. Name and Rank of Other Department Faculty in the Same Field:

Samuel Madden, Professor
David DeWitt, Adjunct Professor
Michael Stonebraker, Adjunct Professor

7. Name and Rank of Faculty in Other Departments in the Same Field:

None

8. Non-MIT Experience (including military service):

<u>Employer</u>	<u>Position</u>	<u>Beginning</u>	<u>Ending</u>
ICN GmbH & Co. KG, Germany	Consultant	March 1997	Feb. 2003
McKinsey & Company, Germany	Associate Project Manager and Software Developer	Aug. 2002	June 2004
Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Germany	Software Developer	April 2006	Aug. 2006
ETZ Zurich, Switzerland	Research Assistant	Dec. 2006	March 2010
University of California, Berkeley	Postdoctoral Scholar	March 2010	Dec. 2012

Brown University	Assistant Professor	Jan. 2013	Dec. 2017
Google, Mountain View	Visiting Researcher	Jan. 2017	Aug. 2017
9. History of MIT Appointments:			
<u>Rank</u>		<u>Beginning</u>	<u>Ending</u>
Associate Professor without Tenure, EECS		Jan. 2018	present
10. Consulting Record:			
<u>Firm</u>		<u>Beginning</u>	<u>Ending</u>
Marine Polymer Technologies, Burlington		Jan. 2016	Dec. 2017
TravelClick, New York		March 2016	Dec. 2017
Google Brain, Mountain View		Sept. 2017	present
Samsung Next Q Fund		June 2018	present
Paul, Weiss, Rifkind, Wharton & Garrison LLP		Dec. 2018	present
einblick analytics, inc.		Jan. 2020	present
11. Department and Institute Committees, Other Assigned Duties:			
<u>Activity</u>		<u>Beginning</u>	<u>Ending</u>
Ph.D. Admissions Committee (Computer Science, Brown University), member		2013	2017
Presidential Data Science Initiative (Brown University), member		2014	2018
Master of Data Science Curriculum Committee (Brown University), member		2015	2016
Initiatives and Vision (Computer Science, Brown University), member		2015	2015
Computing committee (Computer Science, Brown University), member		2016	2016
Co-founder and co-director of DSAIL@CSAIL http://dsail.csail.mit.edu/		2018	present
Ph.D. Admissions Committee		2019	2019
12. Professional service:			
<u>Activity</u>		<u>Beginning</u>	<u>Ending</u>
W3C XQuery working group, member		2007	2010
CloudApp 2010, Program Committee		2010	2010
SIGMOD 2011, Program Committee		2011	2011
SIGMOD Repeatability 2011, Program Committee		2011	2011
EDBT 2011		2011	2011
International Workshop on Data Analytics in the Cloud (DanaC) in conjunction with EDBT/ICDT 2011, Program Committee Chair		2011	2011
SOCC 2012, Program Committee		2012	2012
ICDE 2012, Program Committee		2012	2012

SOCC 2013, Program Committee	2013	2013
EuroSys 2013, Program Committee	2013	2013
VLDB 2013 Ph.D. Workshop, Program Committee	2013	2013
NEDB Summit 2013, Program Committee	2013	2013
DMC 2013, Program Committee	2013	2013
IEEE Internet Computing, Editor	2013	2014
CIKM 2014, Program Committee	2014	2014
SIGMOD 2014, Program Committee	2014	2014
VLDB 2014, Program Committee	2014	2014
ICDE 2014, Program Committee	2014	2014
Information Systems Journal, Editor	2014	2014
New England Database Summit (NEDB) 2014, Program Committee	2014	2014
Chair		
CIKM 2015, Program Committee	2015	2015
SIGMOD 2015, Program Committee	2015	2015
VLDB 2015, Program Committee	2015	2015
Northeast Big Data Innovation Hub, Steering committee member and main representative of Rhode Island	2015	2018
Usenix HotCloud 2015, Program Committee Chair	2015	2015
OSDI 2016, Program Committee	2016	2016
VLDB 2016, Program Committee	2016	2016
ICDE 2016, Program Committee	2016	2016
Workshop on Human-in-the-Loop Data Analytics, co-located with SIGMOD 2016	2016	2016
Usenix HotCloud, Advisory Board member	2016	2019
IEEE Data Engineering Bulletin, Editor	2016	2018
SIGMOD 2016 System Track, Program Committee Chair	2016	2016
NEDB Data Sharing Workshop 2016, Program Committee Chair	2016	2016
Northeast Big Data Hub Workshop, Organizer	2017	2017
KDD 2017, Program Committee	2017	2017
SIGMOD 2017, Program Committee	2017	2017
ICDE 2017, Program Committee	2017	2017
CIDR 2017, Program Committee	2017	2017
WWW 2017, Program Committee	2017	2017
VLDB 2018, Program Committee	2018	2018
SIGMOD 2018, Program Committee	2018	2018
Northeast Big Data Innovation Hub, advisory board	2018	present
CIDR 2019, Program Committee	2019	2019
NEDB 2019, Organizer and Program Chair	2019	2019
SYSML 2019, Program Committee	2019	2019
SIGMOD 2019, Program Co-Chair	2019	2019
Poly19@VLDB, Program Committee	2019	2019
CIDR 2020, Program Committee	2020	2020
SML 2020, Program Committee	2020	2020
SIGMOD 2020, Program Committee	2020	2020
Poly20@VLDB, Program Committee	2020	2020

13. Awards Received:

<u>Award</u>	<u>Date</u>
ACM TODS – Best of SIGMOD invitation	2008
Prospective Researcher Fellowship, Swiss National Science Foundation	2010
VLDB Best Demo Award	2011
ICDE Best Paper Award	2013
Robotics Science and Systems (AAAI-RSS) Blue Sky Award	2015
VLDB Best Demo Award	2015
Honorable Mention for the TCDE Early Career Award	2015
AFOSR Young Investigator Research Award	2015
NSF CAREER Award	2015
Google Research Award	2015
ACM TODS - Best of SIGMOD invitation	2016
Google PhD Fellowship (for Ph.D. student Andrew Crotty)	2016
Early Career Research Achievement Award, Brown University	2017
Appointed a name chair: Manning Assistant Professor of Computer Science, Brown University	2017
VMware Systems Research Award	2017
Alfred P. Sloan Research Fellow in Computer Science	2017
Appointed a name chair: ITT Career Development Associate Professor in Computer Technology, MIT	2018
VLDB Early Career Research Contribution Award	2018

14. Current Organization Membership:

<u>Organization</u>	<u>Offices Held</u>
Gesellschaft für Informatik e.V.	Member, 2006 - present
Association for Computing Machinery	Member, 2012 - present
ACM Special Interest Group on Management of Data	Member, 2012 - present
Institute of Electrical and Electronics Engineers (IEEE)	Member, 2013 - present
DHV – German Association of University Professors and Lecturers	Member, 2013 - present

15. Patents and Patent Applications Pending:

1. Mathias Brantner, David Graf, Donald Kossmann, Tim Kraska, “Fine-grained and concurrent access to a virtualized disk in a distributed system,” US application no. 12/350,197, July 2009.

16. Professional Registration:

None

17. Major New Products, Processes, Designs, or Systems:

Zorba, a compliant W3C XQuery implementation, open-source project: <http://www.zorba.io/>

MLI, an API for Distributed Machine Learning, <https://github.com/amplab/MLI> has been incorporated into Apache Spark MLlib <http://spark.apache.org/mlib/>

RMI, open-source implementations of the original learned index using the recursive model index <https://github.com/learnedsystems/RMI>

SOSD, Search on Sorted Data Benchmark
<https://github.com/learnedsystems/SOSD>

Alex, a ML-enhanced range index (in collaboration with Microsoft)
<https://github.com/microsoft/ALEX>

Park, an open-source platform for learning augmented computer systems
<https://github.com/park-project/park>

BAO, bandit optimizer
<https://github.com/learnedsystems/baoforpostgresql>

Northstar, an Interactive Data Science Systems, now commercialized by einblick analytics, inc.

Educational Contributions Other Than Classroom Teaching of Tim Kraska

1. Teaching materials developed that illustrate teaching effectiveness or innovativeness:
 - i. FT13: Design of Brown University's first Data Science course "CS1951-A: Data Science."
 - ii. ST15: Design of a 1-week Ph.D. course on "Integrated Analytics on Big Data" at the University of Aalborg, Denmark.
 - iii. ST15: Creation of the Data Science Teaching Initiative (www.teachingdatascience.org) (2015-2019).
 - iv. FT15: Development of a first Google Glass teaching tool (it shows latex equations/lecture notes from the PowerPoint slides while presenting).
 - v. ST16-FT16: Design of Brown's 2-credit Master of Data Science "DATA1030: An introduction to Topics in Data and Computational Science."
 - vi. FT19: Design of MIT's first Systems for Data Science class "6.S080 Software Systems for Data Science."
2. Education contributions, apart from classroom performance and supervision, such as new educational programs and curricula developed by the candidate:
 - i. Design of Brown's Master of Data Science Curriculum, offered for the first time in Fall 2017. <https://www.brown.edu/initiatives/data-science/masters-degree>

3. Contributions to the educational commons:

I advised the following students as part of the Undergraduate Research Opportunities Program (UROP) at MIT:

ST18	Wooders, Sarah	"Improving Sorting Performance by Learning the Dataset's Distribution"
IAP-ST19	Mustafi, Urmi	"Visualization of Data with Learned Indexes"
ST19-FT19	Lam, Jason	"Time Series Data Visualization" (now advised as a master's student)
ST19	Peng, Anthony B.	"Gene Sequence Search"
ST19-ST20	Yesantharao, Rahul V.	"Tuplex" (led to paper submission to OSDI'20 [under review])
ST19	Barnett, Daniel C.	"Apache Arrow for Northstar"
ST19	Gayle, Ricardo M.	"Data Wrangling for Northstar"
ST19-ST20	Ho, Darryl	"Learned DNA Sequence Search" (led to paper at the MLSys workshop [4.21])
ST19	Montas, Enrique B.	"Learned Data Structures: Incremental Adaptivity of an Updatable Learned Multi-Dimensional Index"

Educational Contributions Other Than Classroom Teaching of Tim Kraska

FT19-ST20	Richardson, Yaateh H.	SuperUROP: Applications of Learned Data Structures (received the “2020 SuperUROP award”)
FT19-ST20	Cen, Lujing	“Use Machine Learning to Take Out the Trash: A Learned Garbage Collector” (led to paper at MAPL@PLDI’20 [4.26])
FT19	Chen, Shiyu	“Debugging in Northstar”
FT19-ST20	Hernandez, Adriano	“Let programmers be lazy! Btrees edition”
FT19	Yang, Adela Y.	“Debugging in Northstar”
FT19	Zhang, Emily Y.	“Tuning Learned Index Structures” (led to demo paper at SIGMOD’20 [3.62])
ST20	Lu, Mindren D.	“Enhanced Video Conferencing”

Publications of Tim Kraska

1. Books

None.

2. Papers in Refereed Journals

1. Tim Kraska, Martin Hentschel, Gustavo Alonso, Donald Kossmann, “Consistency Rationing in the Cloud: Pay only when it matters,” Proceedings of the VLDB Endowment (PVLDB), 2(1), pp. 253-264, 2009.
2. Roger Bamford, Vinayak Borkar, Matthias Brantner, Peter M. Fischer, Daniela Florescu, David Graf, Donald Kossmann, Tim Kraska, Dan Muresan, Sorin Nasoi, Markos Zacharioudakis, “XQuery Reloaded” (Industrial Paper), Proceedings of the VLDB Endowment (PVLDB), 2(2), pp. 1342-1353, 2009.
3. Donald Kossmann, Tim Kraska, Simon Loesing, Stephan Merkli, Raman Mittal, Flavio Pfaffhauser, “Cloudy: A Modular Cloud Storage System” (Demo Paper), Proceedings of the VLDB Endowment (PVLDB), 3(2), pp. 1533-1536, 2010.
4. AnHai Doan, Michael J. Franklin, Donald Kossmann, Tim Kraska, “Crowdsourcing Applications and Platforms: A Data Management Perspective” (Tutorial), Proceedings of the VLDB Endowment (PVLDB), 4(12), pp. 1508-1509, 2011.
5. Michael Armbrust, Kristal Curtis, Tim Kraska, Armando Fox, Michael J. Franklin, David A. Patterson, “Success-Tolerant Query Processing in the Cloud,” Proceedings of the VLDB Endowment (PVLDB), 5(3), pp.181-192, 2011.
6. Amber Feng, Michael Franklin, Donald Kossmann, Tim Kraska, Samuel Madden, Sukriti Ramesh, Reynold Xin, “CrowdDB: Sourcing the VLDB Crowd” (Demo Paper), Proceedings of the VLDB Endowment (PVLDB), 4(12), pp. 1387-1390, 2011. (**Best Demo Award**)
7. Jiannan Wang, Tim Kraska, Michael J. Franklin, Jianhua Feng, “CrowdER: Crowdsourcing Entity Resolution,” Proceedings of the VLDB Endowment (PVLDB), 5(11), pp.1483-1494, 2012.
8. Andrew Crotty, Alex Galakatos, Tim Kraska, “Tuplware: Distributed Machine Learning on Small Clusters,” IEEE Data Engineering Bulletin 37(3), pp. 63-76, 2014. **
9. Ugur Çetintemel, Jiang Du, Tim Kraska, Samuel Madden, David Maier, John Meehan, Andrew Pavlo, Michael Stonebraker, Erik Sutherland, Nesime Tatbul, Kristin Tufte, Hao Wang, Stanley B. Zdonik, “S-Store: A Streaming NewSQL System for Big Velocity Applications” (Demo Paper), Proceedings of the VLDB Endowment (PVLDB), 7(13), pp. 1633-1636, 2014. **
10. Andrew Crotty, Alex Galakatos, Kayhan Dursun, Tim Kraska, Carsten Binnig, Ugur Cetintemel, Stan Zdonik, “An Architecture for Compiling UDF-centric Workflows,” Proceedings of the VLDB Endowment (PVLDB), 8(12), pp. 1466-1477, 2015. **

** Outgrowth of supervised student research

Publications of Tim Kraska

11. John Meehan, Nesime Tatbul, Stan Zdonik, Cansu Aslantas, Ugur Cetintemel, Jiang Du, Tim Kraska, Samuel Madden, David Maier, Andrew Pavlo, Michael Stonebraker, Kristin Tufte, Hao Wang, “S-Store: Streaming Meets Transaction Processing,” Proceedings of the VLDB Endowment (PVLDB), 8(13), pp. 2134-2145, 2015. **
12. Sanjay Krishnan, Jiannan Wang, Michael J. Franklin, Ken Goldberg, Tim Kraska, “Stale View Cleaning: Getting Fresh Answers from Stale Materialized Views,” Proceedings of the VLDB Endowment (PVLDB), 8(12), pp.1370-1381, 2015.
13. Beth Trushkowsky, Tim Kraska, Michael J. Franklin, Purnamrita Sarkar, Venketaram Ramachandran, “Crowdsourcing Enumeration Queries: Estimators and Interfaces,” IEEE Transactions on Knowledge and Data Engineering 27(7), pp. 1796-1809, 2015.
14. Andrew Crotty Alex Galakatos Emanuel Zraggen Carsten Binnig Tim Kraska, “Vizdom: Interactive Analytics through Pen and Touch” (Demo Paper), Proceedings of the VLDB Endowment (PVLDB), 8(12), pp. 2024-2035, 2015. (**Best Demo Award**) **
15. Aaron Elmore, Jennie Duggan, Michael Stonebraker, Magdalena Balazinska, Ugur Cetintemel, Vijay Gadepally, Jeff Heer, Bill Howe, Jeremy Kepner, Tim Kraska, Samuel Madden, David Meier, Tim Mattson, Stavros Papadopoulos, Jeff Parkhurst, Nesime Tatbul, Manasi Vartek, Stan Zdonik, “A Demonstration of the BigDAWG Polystore System” (Demo Paper), Proceedings of the VLDB Endowment (PVLDB), 2015.
16. Sanjay Krishnan, Jiannan Wang, Michael J. Franklin, Ken Goldberg, Tim Kraska, Tova Milo, Eugene Wu, “SampleClean: Fast and Reliable Analytics on Dirty Data,” IEEE Data Eng. Bull. 38(3), pp. 59-75, 2015.
17. Emanuel Zraggen, Alexander Galakatos, Andrew Crotty, Jean-Daniel Fekete, Tim Kraska, “How Progressive Visualizations Affect Exploratory Analysis,” IEEE Transactions on Visualization and Computer Graphics (TVCG), 99(PP), 2016. **
18. Philipp Eichmann, Emanuel Zraggen, Zheguang Zhao, Carsten Binnig, Tim Kraska, “Towards a Benchmark for Interactive Data Exploration,” IEEE Data Engineering Bulletin, 39(4), pp. 50-61, 2016. **
19. Carsten Binnig, Andrew Crotty, Alex Galakatos, Tim Kraska, Erfan Zamanian, “The End of Slow Networks: It's Time for a Redesign,” Proceedings of the VLDB Endowment (PVLDB), 9(7), pp. 528-539, 2016. **
20. Beth Trushkowsky, Tim Kraska, Michael J. Franklin, Purnamrita Sarkar, “Answering enumeration queries with the crowd,” Communications of the ACM, 59(1), pp. 118–127, 2016. (**Listed in ACM Computing Reviews “21st Annual Best of Computing: Notable Books and Articles,” 2016**)
21. Abdallah Salama, Carsten Binnig, Tim Kraska, Ansgar Scherp, Tobias Ziegler, “Rethinking Distributed Query Execution on High-Speed Networks,” IEEE Data Engineering Bulletin, pp. 27-37, 2017. **

** Outgrowth of supervised student research

Publications of Tim Kraska

22. Erfan Zamanian, Carsten Binnig, Tim Kraska, Tim Harris, “The End of a Myth: Distributed Transactions Can Scale,” Proceedings of the VLDB Endowment (PVLDB), 2017. *(Nominated by Mike Stonebraker for ACM Computing Reviews “2016’s Best Books and Articles in Computing”)* **
23. Yeounoh Chung, Sanjay Krishnan, Tim Kraska, “A Data Quality Metric (DQM): How to Estimate the Number of Undetected Errors in Data Sets,” Proceedings of the VLDB Endowment (PVLDB), 2017. **
24. Alexander Galakatos, Andrew Crotty, Emanuel Zraggen, Carsten Binnig, Tim Kraska, “Revisiting Reuse for Approximate Query Processing,” Proceedings of the VLDB Endowment (PVLDB), 2017. **
25. Michael Mortensen, Gaelen Adam, Thomas Trikalinos, Tim Kraska, Byron Wallace, “An Exploration of Crowdsourcing Citation Screening for Systematic Reviews,” Research Synthesis Methods, 8(3), pp. 366-386, 2017. **
26. Yeounoh Chung, Michael Lind Mortensen, Carsten Binnig, Tim Kraska, “Estimating the Impact of Unknown Unknowns on Aggregate Query Results,” ACM Trans. Database Syst. 43(1): 3:1-3:37, 2018. **
27. Tim Kraska, “Northstar: An Interactive Data Science System,” Proceedings of the VLDB Endowment (PVLDB), 11(12), pp. 2150-2164, 2018.
28. Yeounoh Chung, Sacha Servan-Schreiber, Emanuel Zraggen, Tim Kraska, “Towards Quantifying Uncertainty in Data Analysis & Exploration,” IEEE Data Eng. Bull. 41(3), pp. 15-27, 2018. **
29. Anastasia Ailamaki, Periklis Chrysogelos, Amol Deshpande, Tim Kraska, “The SIGMOD 2019 Research Track Reviewing System,” SIGMOD Record, 48(2), pp. 47-54, 2019.
30. Junjay Tan, Thanaa Ghanem, Matthew Perron, Xiangyao Yu, Michael Stonebraker, David J. DeWitt, Marco Serafini, Ashraf Aboulnaga, Tim Kraska, “Choosing A Cloud DBMS: Architectures and Tradeoffs,” Proceedings of the VLDB Endowment (PVLDB), 12(12), pp. 2170-2182, 2019. **
31. Leonhard F. Spiegelberg, Tim Kraska. “Tuplex: Robust, Efficient Analytics When Python Rules” (Demo), Proceedings of the VLDB Endowment (PVLDB), 12(12), pp. 1958-1961, 2019. **
32. Ryan C. Marcus, Parimarjan Negi, Hongzi Mao, Chi Zhang, Mohammad Alizadeh, Tim Kraska, Olga Papaemmanouil, Nesime Tatbul, “Neo: A Learned Query Optimizer,” Proceedings of the VLDB Endowment (PVLDB), 12(11), pp. 1705-1718, 2019. **
33. Erfan Zamanian, Xiangyao Yu, Michael Stonebraker, Tim Kraska, “Rethinking Database High Availability with RDMA Networks,” Proceedings of the VLDB Endowment (PVLDB), 12(11), pp. 1637-1650, 2019. **
34. Nadiia Chepurko, Ryan Marcus, Emanuel Zraggen, Raul Castro Fernandez, Tim Kraska, David Karger, “ARDA: Automatic Relational Data Augmentation for Machine

** Outgrowth of supervised student research

Learning,” Proceedings of the VLDB Endowment (PVLDB), VLDB Endowment (PVLDB), 13(9), pp. 1373-1387, 2020.

35. Ryan Marcus, Andreas Kipf, Alexander van Renen, Mihail Stoian, Sanchit Misra, Alfons Kemper, Thomas Neumann, Tim Kraska, “Benchmarking Learned Indexes” Experiments & Analyses), Proceedings of the VLDB Endowment (PVLDB), to appear.
 36. Jialin Ding, Vikram Nathan, Mohammad Alizadeh, Tim Kraska, “Tsunami: A Learned Multi-dimensional Index for Correlated Data and Skewed Workloads,” Proceedings of the VLDB Endowment (PVLDB), to appear. **
3. Proceedings of Refereed Conferences
1. Tim Kraska, Uwe Röhm, “Genea: Schema-Aware Mapping of Ontologies into Relational Databases,” Proceedings of the International Conference on Management of Data (COMAD), pp. 92-103, 2006.
 2. Joshua Wing Kei Ho, Tristan Manwaring, Seok-Hee Hong, Uwe Röhm, David Cho Yau Fung, Kai Xu, Tim Kraska, David Hart, “PathBank: Web-Based Querying and Visualziation of an Integrated Biological Pathway Database,” Proceedings of the International Conference on Computer Graphics, Imaging and Visualization (CGIV), pp. 84-89, 2006.
 3. Irina Botan, Peter M. Fischer, Daniela Florescu, Donald Kossmann, Tim Kraska, Rokas Tamosevicius, “Extending XQuery with Window Functions,” Proceedings of the International Conference on Very Large Data Bases (VLDB), pp. 75-86, 2007.
 4. Matthias Brantner, David Graf, Daniela Florescu, Donald Kossmann, Tim Kraska, “Building a Database on S3,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 251-264, 2008. (***Best of SIGMOD*** invite by the *ACM Transactions on Database Systems (TODS) journal*)
 5. Ghislain Fourny, Donald Kossmann, Tim Kraska, Markus Pilman, Daniela Florescu, “XQuery in the Browser” (Demo Paper), Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1337-1340, 2008.
 6. Ghislain Fourny, Markus Pilman, Daniela Florescu, Donald Kossmann, Tim Kraska, Darin McBeath, “XQuery in the Browser,” Proceedings of the International Conference on World Wide Web (WWW), pp. 1011-1020, 2009.
 7. Donald Kossmann, Tim Kraska, Simon Loesing, “An Evaluation of Alternative Architectures for Transaction Processing in the Cloud,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 579-590, 2010.
 8. Michael Franklin, Donald Kossmann, Tim Kraska, Sukriti Ramesh, Reynold Xin, “Crowddb: Answering Queries with Crowdsourcing,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 61-72, 2011.

** Outgrowth of supervised student research

Publications of Tim Kraska

9. Jan Schaffner, Dean Jacobs, Tim Kraska, Hasso Plattner, “The Multi-Tenant Data Placement Problem,” Proceedings of the International Conference on Advances in Databases, Knowledge, and Data Applications, 2012.
10. Sean Goldberg, Jeff Depree, Daisy Zhe Wang, Tim Kraska, “CASTLE: Crowd-Assisted System for Text Labeling and Extraction,” Proceedings of the Conference on Human Computation & Crowdsourcing (HCOMP), 2013.
11. Beth Trushkowsky, Tim Kraska, Michael J. Franklin, “A Framework for Adaptive Crowd Query Processing,” Proceedings of the Conference on Human Computation & Crowdsourcing (HCOMP), 2013.
12. Evan Sparks, Ameet Talwalkar, Virginia Smith, Xinghao Pan, Joseph Gonzales, Tim Kraska, Michael Jordan, Michael Franklin, “MLI: An API for Distributed Machine Learning,” Proceedings of the IEEE International Conference on Data Mining (ICDM), 2013.
13. Jiannan Wang, Guoliang Li, Tim Kraska, Michael J. Franklin, Jianhua Feng, “Leveraging Transitive Relations for Crowdsourced Joins,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 229-240, 2013.
14. Jan Schaffner, Tim Januschowski, Megan Kercher, Tim Kraska, Hasso Plattner, Michael J. Franklin, Dean Jacobs, “RTP: Robust Tenant Placement for Elastic In-Memory Database Clusters,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 773-784, 2013.
15. Michael Armbrust, Eric Liang, Tim Kraska, Armando Fox, Michael J. Franklin, David A. Patterson, “Generalized Scale Independence Through Incremental Precomputation,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 625-636, 2013.
16. Tim Kraska, Gene Pang, Michael Franklin, Samuel Madden, Alan Fekete, “MDCC: Multi-Data Center Consistency,” Proceedings of the Eurosys Conference, pp. 113-126, 2013.
17. Beth Trushkowsky, Tim Kraska, Michael J. Franklin, Purnamrita Sarkar, “Crowdsourced Enumeration Queries,” Proceedings of the IEEE International Conference on Data Engineering (ICDE), 2013. (*Best Paper Award*)
18. Tim Kraska, Ameet Talwalkar, John Duchi, Rean Griffith, Michael Franklin, Michael Jordan, “MLbase: A Distributed Machine-learning System,” Proceedings of the Conference on Innovative Data Systems Research (CIDR), 2013.
19. Gianluca Demartini, Beth Trushkowsky, Tim Kraska, Michael Franklin, “CrowdQ: Crowdsourced Query Understanding,” Proceedings of the Conference on Innovative Data Systems Research (CIDR), 2013.
20. Gene Pang, Tim Kraska, Michael Franklin, Samuel Madden, Alan Fekete, “PLANET: Making Progress with Commit Processing in Unpredictable Environments,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 3-14, 2014.

Publications of Tim Kraska

21. Jiannan Wang, Sanjay Krishnan, Michael J. Franklin, Ken Goldberg, Tim Kraska, Tova Milo, “A sample-and-clean framework for fast and accurate query processing on dirty data,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 469-480, 2014.
22. Evan R. Sparks, Ameet Talwalkar, Daniel Haas, Michael J. Franklin, Michael I. Jordan, Tim Kraska, “Automating Model Search for Large Scale Machine Learning,” Proceedings of the ACM Symposium on Cloud Computing (SOCC), pp. 368-380, 2015.
23. Abdallah Salama, Carsten Binnig, Tim Kraska, Erfan Zamanian, “Cost-based Fault-tolerance for Parallel Data Processing,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 285-297, 2015. **
24. Andrew Crotty, Alex Galakatos, Kayhan Dursun, Tim Kraska, Ugur Cetintemel, Stan Zdonik, “Tuplware: ‘Big’ Data, Big Analytics, Small Clusters,” Proceedings of the Conference on Innovative Data Systems Research (CIDR), 2015. **
25. Yeounoh Chung, Michael Lind Mortensen, Carsten Binnig, Tim Kraska, “Estimating the Impact of Unknown Unknowns on Aggregate Query Results,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 861-876, 2016. (*“Best of SIGMOD” invite by the ACM Transactions on Database Systems (TODS) journal*) **
26. Sanjay Krishnan, Jiannan Wang, Michael J. Franklin, Ken Goldberg, Tim Kraska, “PrivateClean: Data Cleaning and Differential Privacy,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 937-951, 2016.
27. Gabriel Lyons, Vinh Tran, Carsten Binnig, Ugur Çetintemel, Tim Kraska, “Making the Case for Query-by-Voice with EchoQuery” (Demo Paper), Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 2129-2132, 2016. **
28. Carsten Binnig, Lorenzo De Stefani, Tim Kraska, Eli Upfal, Emanuel Zgraggen, Zheguang Zhao, “Towards Sustainable Insights, or why polygamy is bad for you,” Proceedings of the Conference on Innovative Data Systems Research (CIDR), 2017. **
29. Tim Kraska, Elkhan Dadashov, Carsten Binnig, “Spotlytics: How to Use Cloud Market Places for Analytics?,” BTW, pp. 361-380, 2017. **
30. Christoph Pinkel, Carsten Binnig, Ernesto Jimenez-Ruiz, Evgeny Kharlamov, Andriy Nikolov, Andreas Schwarte, Christian Heupel, Tim Kraska, “IncMap: A Journey towards Ontology-based Data Integration,” BTW, pp. 145-164, 2017.
31. Zheguang Zhao, Lorenzo De Stefani, Emanuel Zgraggen, Carsten Binnig, Eli Upfal, Tim Kraska, “Controlling False Discoveries During Interactive Data Exploration,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 527-540, 2017. **

** Outgrowth of supervised student research

Publications of Tim Kraska

32. Kayhan Dursun, Carsten Binnig, Ugur Cetintemel, Tim Kraska, “Revisiting Reuse in Main Memory Database Systems,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1275-1289, 2017.
33. Zheguang Zhao, Emanuel Zgraggen, Lorenzo De Stefani, Carsten Binnig, Eli Upfal and Tim Kraska, “Safe Visual Data Exploration” (Demo Paper), Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1671-1674, 2017. **
34. Bill Howe, Michael J. Franklin, Laura M. Haas, Tim Kraska, Jeffrey D. Ullman, “Data Science Education: We’re missing the Boat, Again,” Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 1473-1474, 2017.
35. Tim Kraska, “Approximate Query Processing For Interactive Data Science,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 525, 2017.
36. Tim Kraska, “Letter from the Special Issue Editor,” IEEE Data Eng. Bull. 40(1,4): 2, 2017.
37. Tim Kraska, Alex Beutel, Ed H. Chi, Jeffrey Dean, Neoklis Polyzotis, “The Case for Learned Index Structures,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 489-504, 2018.
38. Linnan Wang, Jinmlan Ye, Yiyang Zhao, Wei Wu, Ang Li, Shuaiwen Leon Song, Zenglin Xu, Tim Kraska, “Superneurons: dynamic GPU memory management for training deep neural networks,” Proceedings of the Symposium on Principles and Practice of Parallel Programming (PPoPP), pp. 41-53, 2018. **
39. Emanuel Zgraggen, Zheguang Zhao, Robert C. Zeleznik, Tim Kraska, “Investigating the effect of the Multiple Comparisons Problem in Visual Analysis,” Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI), pp. 479, 2018. **
40. Kevin Zeng Hu, Michiel A. Bakker, Stephen Li, Tim Kraska, Cesar A. Hidalgo, “VizML: A Machine Learning Approach to Visualization Recommendation,” Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI), pp. 128, 2019. **
41. Matthew Perron, Zeyuan Shang, Tim Kraska, Michael Stonebraker, “How I Learned to Stop Worrying and Love Re-optimization,” Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 1758-1761, 2019. **
42. Yeounoh Chung, Tim Kraska, Neoklis Polyzotis, Ki Hyun Tae, Steven Euijong Whang, “Slice Finder: Automated Data Slicing for Model Validation,” Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 1550-1553, 2019. **
43. Stratos Idreos, Tim Kraska, “From Auto-tuning One Size Fits All to Self-designated and Learned Data-intensive Systems” (Tutorial), Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 2054-2059, 2019.

** Outgrowth of supervised student research

Publications of Tim Kraska

44. Alex Galakatos, Michael Markovitch, Carsten Binnig, Rodrigo Fonseca, Tim Kraska, “FITing-Tree: A DATA-aware Index Structure,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1189-1206, 2019. **
45. Zeyuan Shang, Emanuel Zgraggen, Benedetto Buratti, Ferdinand Kossmann, Philipp Eichmann, Yeounoh Chung, Carsten Binnig, Eli Upfal, Tim Kraska, “Democratizing Data Science through Interactive Curation of ML Pipelines,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1171-1188, 2019. **
46. Tobias Ziegler, Sumukha Tumkur Vani, Carsten Binnig, Rodrigo Fonseca, Tim Kraska, “Designing Distributed Tree-based Index Structures for Fast RDMA-capable Networks,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 741-758, 2019. **
47. Hongzi Mao, Parimarjan Nagi, Akshay Narayan, Hanrui Wang, Jiacheng Yang, Haonan Wang, Ryan Marcus, Ravichandra Addanki, Mehrdad Khani Shirkoohi, Songtao He, Vikram Nathan, Frank Cangialosi, Shaileshh Bojja Venkatakrishnan, Wei-Hung Weng, Song Han, Tim Kraska, Mohammad Alizadeh, “Park: An Open Platform for Learning-Augmented Computer Systems,” NeurIPS, pp. 2490-2502, 2019. **
48. Madelon Hulsebos, Kevin Zeng Hu, Michael A. Bakker, Emanuel Zgraggen, Arvind Satyanarayan, Tim Kraska, Cagatay Demiralp, Cesar A. Hidalgo, “Sherlock: A Deep Learning Approach to Semantic Data Type Detection,” KDD, pp. 1500-1508, 2019. **
49. Lorenzo De Stefani, Leonhard F. Spiegelberg, Eli Upfal, Tim Kraska, “VizCertify: A Framework for Secure Visual Data Exploration,” Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 241-251, 2019. **
50. Tim Kraska, Mohammad Alizadeh, Alex Beutel, Ed H. Chi, Ani Kristo, Guillaume Lecierc, Samuel Madden, Hongzi Mao, Vikram Nathan, “SageDB: A Learned Database System,” Proceedings of the Conference on Innovative Data Systems Research (CIDR), 2019. **
51. Kevin Zeng Hu, Snehal Kumar (Neil) S. Gaikwad, Madelon Hulsebos, Michiel A. Bakker, Emanuel Zgraggen, Cesar A. Hidalgo, Tim Kraska, Guoliang Li, Arvind Satyanarayan, Cagatay Demiralp, “VizNet: Towards A Large-scale Visualization Learning and Benchmarking Repository,” Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI), pp. 662, 2019. **
52. Andrew Crotty, Alex Galakatos, Tim Kraska, “Getting Swole: Generating Access-aware Code with Predicate Pullups,” Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 1273-1284, 2020. **
53. Songtao He, Favyen Bastani, Arjun Balasingam, Karthik Gopalakrishnan, Mohammad Alizadeh, Hari Balakrishnan, Michael Cafarella, Tim Kraska, Sam Madden, “BeeCluster: Enable Application-Aware Predictive Optimization in Drone Orchestration System,”

** Outgrowth of supervised student research

Publications of Tim Kraska

- Proceedings of the ACM International Conference on Mobile Systems, Applications, and Services (MobiSys 2020), pp. 299-311, 2020. **
54. Favyen Bastani, Songtao He, Arjun Balasingam, Karthik Gopalakrishnan, Mohammad Alizadeh, Hari Balakrishnan, Michael Cafarella, Tim Kraska, Samuel Madden, "MIRIS: Fast Object Track Queries in Video," Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1907-1921, 2020. **
 55. Philipp Eichmann, Emanuel Zraggen, Carsten Binnig, Tim Kraska, "IDEBench: A Benchmark for Interactive Data Exploration," Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1555-1569, 2020. **
 56. Matthias Jasný, Tobias Ziegler, Tim Kraska, Uwe Roehm, Carsten Binnig, "DB4ML: An In-Memory Database Kernel with Machine Learning Support," Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 159-173, 2020. **
 57. Vikram Nathan, Jialin Ding, Mohammad Alizadeh, Tim Kraska, "Learned Multi-Dimensional Indexing," Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 985-1000, 2020. **
 58. Jialin Ding, Umar Farooq Minhas, Jia Yu, Chi Wang, Hantian Zhang, Yinan Li, Badrish Chandramouli, Jaeyoung Do, Johannes Gehrke, Tim Kraska, Donald Kossmann, David Lomet, "ALEX: An Updatable Adaptive Learned Index for Range Queries," Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 969-984, 2020. **
 59. Erfan Zamanian, Julian Shun, Carsten Binnig, Tim Kraska, "Chiller: Contention-centric Transaction Execution and Data Partitioning for Modern Networks," Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp 511-526, 2020. **
 60. Ani Kristo, Kapil Vaidya, Ugur Cetintemel, Sanchit Misra, Tim Kraska, "The Case for a Learned Sorting Algorithm," Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1001-1016, 2020. **
 61. Ryan Marcus, Emily Zhang, Tim Kraska, "CDFShop: Exploring and Optimizing Learned Index Structures" (Demo), Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 2789-2792, 2020. **
 62. Jeremy Kepner, Andreas Kipf, Darren Engwirda, Navin Vembar, Michael S Jones, Lauren Milechin, Vijay Gadepally, Chris Hill, Tim Kraska, William Arcand, David Bestor, William Bergeron, Chansup Byun, Matthew Hubbell, Michael Houle, Andrew Kirby, Anna Klein, Julie Mullen, Andrew Prout, Albert Reuther, Antonio Rosa, Sid Samsi, Charles Yee, Peter Michaleas, "Fast Mapping onto Census Blocks", Proceedings of the IEEE High Performance Extreme Computing Conference (HPEC), to appear.

4. Other Major Publications

** Outgrowth of supervised student research

Publications of Tim Kraska

1. Carsten Binnig, Donald Kossmann, Tim Kraska, Simon Loesing, “How is the Weather tomorrow? Towards a Benchmark for the Cloud,” DBTest Workshop in conjunction with SIGMOD 2009, Providence, RI, 2009. **
2. Donald Kossmann, Tim Kraska, “Data Management in the Cloud: Promises, State-of-the-art, and Open Questions,” *Datenbank Spektrum*, 10(3), pp. 121-129, 2010.
3. Tim Kraska (ed), “XML Query (XQuery) 1.1 Use Cases,” W3C Working Draft, 2010.
4. Philippe Bonnet, Stefan Manegold, Matias Bjørling, Wei Cao, Javier Gonzalez, Joel Granados, Nancy Hall, Stratos Idreos, Milena Ivanova, Ryan Johnson, David Koop, Tim Kraska, René Müller, Dan Olteanu, Paolo Papotti, Christine Reilly, Dimitris Tsirogiannis, Cong Yu, Juliana Freire, Dennis Shasha, “Repeatability and workability evaluation of SIGMOD 2011,” *SIGMOD Record* 40(2), pp. 45-48, 2011.
5. Tim Kraska, “Big Data Column,” *IEEE Internet Computing*, since 2012.
6. Simon Loesing Martin Hentschel, Tim Kraska, Donald Kossmann, “Stormy: An Elastic and Highly Available Streaming Service in the Cloud,” Workshop on Data analytics in the Cloud (DanaC) in conjunction with EDBT/ICDT, Berlin, Germany, 2012.
7. Tim Kraska, Beth Trushkowsky, “The New Database Architectures,” *IEEE Internet Computing*, 17(3), pp. 72-75, 2013.
8. Tim Kraska, “Finding the Needle in the Big Data Systems Haystack,” *IEEE Internet Computing* 17(1), pp. 84-86, 2013.
9. Elkhan Dadashov, Ugur Çetintemel, Tim Kraska, “Putting Analytics on the Spot: Or How to Lower the Cost for Analytics,” *IEEE Internet Computing*, 18(5), pp. 70-73, 2014. **
10. John Oberlin, Maria Meier, Tim Kraska, Stefanie Tellex, “Acquiring Object Experiences at Scale,” Proceedings of the AAAI-RSS Special Workshop on the 50th Anniversary of Shakey: The Role of AI to Harmonize Robots and Humans, co-located with the Robotics Science and Systems Conference (RSS), 2015. (*Won the Blue Sky Award*) **
11. Dalia Kaulakiene, Christian Thomsen and, Torben Bach Pedersen, Ugur Cetintemel, Tim Kraska, “SpotADAPT: Spot-Aware (re-)Deployment of Analytical Processing Tasks on Amazon EC2,” Proceedings of the ACM Eighteenth International Workshop on Data Warehousing and OLAP (DOLAP), pp. 19-23, 2015. **
12. Muhammad El-Hindi, Zheguang Zhao, Carsten Binnig, Tim Kraska, “VisTrees: fast indexes for interactive data exploration,” Workshop on Human-In-the-Loop Data Analytics (HILDA) in conjunction with SIGMOD, 2016. **
13. Andrew Crotty, Alex Galakatos, Emanuel Zraggen, Carsten Binnig, Tim Kraska, “The case for interactive data exploration accelerators (IDEAs),” Workshop on Human-In-the-Loop Data Analytics (HILDA) in conjunction with SIGMOD, 2016. **

** Outgrowth of supervised student research

Publications of Tim Kraska

14. Yue Guo, Carsten Binnig, Tim Kraska “What you see is not what you get! Detecting Simpson's Paradoxes during Data Exploration,” Workshop on Human-In-the-Loop Data Analytics (HILDA) in conjunction with SIGMOD, 2017. **
15. Xiangyao Yu, Vijay Gadepally, Stan Zdonik, Tim Kraska, Michael Stonebraker, “FastDAWG: Improving Data Migration in the BigDAWG Polystore System,” Workshop on Heterogeneous Data Management, Polystores, and Analytics for Healthcare in conjunction with VLDB, pp. 3-15, 2018. **
16. Carsten Binnig, Benedetto Buratti, Yeounoh Chung, Cyrus Cousins, Tim Kraska, Zeyuan Shang, Eli Upfal, Robert C. Zeleznik, Emanuel Zraggen, “Towards Interactive Curation & Automatic Tuning of ML Pipelines,” Proceedings of the International Workshop on Data Management for End-to-End Machine Learning in conjunction with SIGMOD, pp. 1-14, 2018. **
17. Peter A. Boncz, Stefan Manegold, Anastasia Ailamaki, Amol Deshpande, Tim Kraska, “Proceeding of the 2019 International Conference on Management of Data, SIGMOD conference 2019, Amsterdam, The Netherlands, June 30-July 5, 2019,” ACM, ISBN 978-1-4503-5643-5, 2019.
18. Tim Kraska, Michael Stonebraker, Michael L. Brodie, Sacha Servan-Schreiber, Daniel J. Weitzner, “SchengenDB: A Data Protection Database Proposal,” Workshop on Heterogeneous Data Management, Polystores, and Analytics for Healthcare in conjunction with VLDB, pp.24-38, 2019. **
19. Zeyuan Shang, Emanuel Zraggen, Philipp Eichmann, Tim Kraska, “Niseko: a Large-Scale Meta-Learning Dataset,” Workshop on Meta-Learning in conjunction with NeurIPS, 2019. **
20. Vikram Nathan, Jialin Ding, Mohammad Alizadeh, Tim Kraska, “Learned Multi-dimensional Indexing,” Workshop on Machine Learning Systems in conjunction with NeurIPS, 2019. **
21. Andreas Kipf, Ryan Marcus, Alexander van Renen, Mihail Stoian, Alfons Kemper, Tim Kraska, Thomas Neumann, “SOSD: A Benchmark for Learned Indexes,” Workshop on Machine Learning for Systems in conjunction with NeurIPS, 2019. **
22. Zeyuan Shang, Emanuel Zraggen, Tim Kraska, “Alpine Meadow: A System for Interactive AutoML,” MLSys: Workshop on Systems for ML in conjunction with NeurIPS, 2019. **
23. Darryl Ho, Jialin Ding, Sanchit Misra, Nesime Tatbul, Vikram Nathan, Vasimuddin Md, Tim Kraska, “LISA: Towards Learned DNA Sequence Search,” MLSys: Workshop on Systems for ML in conjunction with NeurIPS, 2019. **
24. Ani Kristo, Kapil Vaidya, Ugur Cetintemel, Tim Kraska, “The Case for a Learned Sorting Algorithm,” AI Systems Workshop in conjunction with SOSP, 2019. **
25. Gennady L. Andrienko, Natalia V. Andrienko, Steven Mark Drucker, Jean-Daniel Fekete, Danyel Fisher, Stratos Idreos, Tim Kraska, Guoliang Li, Kwan-Liu Ma, Jock D.

** Outgrowth of supervised student research

Publications of Tim Kraska

- Mackinlay, Antti Oulasvirta, Tobias Schreck, Heidrun Schumann, Michael Stonebraker, David Auber, Nikos Bikakis, Panos K. Chrysanthis, George Papastefanatos, Mohamed A. Sharaf, "Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications," EDBT/ICDT Workshops, 2020.
26. Lujing Cen, Ryan Marcus, Hongzi Mao, Justin Gottschlich, Mohammad Alizadeh, Tim Kraska, "Learned garbage collection," MAPL@PLDI, pp. 38-44, 2020.
27. Andreas Kipf, Ryan Marcus, Alexander van Renen, Mihail Stoian, Alfons Kemper, Tim Kraska, Thomas Neumann, "RadixSpline: a single-pass learned index," aiDM@SIGMOD, pp 5:1 – 5:5, 2020.
28. Parimarjan Negi, Ryan Marcus, Hongzi Mao, Nesime Tatbul, Tim Kraska, Mohammad Alizadeh, "Cost-Guided Cardinality Estimation: Focus Where it Matters," Workshop on Self-Managing Database Systems (SMDB@ICDE). 2019.**
5. Internal Memoranda and Progress Reports
None.
6. Invited Lectures
January 2013, "MLbase: A Distributed Machine-learning System," New England Database Day (NEDBday), Boston, MA.
May 2013, "Tomorrow's Data Management Systems," Oracle, Redmond, CA.
May 2013, "Tomorrow's Data Management Systems," TU Kaiserslautern, Germany.
May 2013, "Tomorrow's Data Management Systems," University of Mannheim, Germany.
June 2013, "CrowdER: Crowdsourcing Entity Resolution," DIMAC Big Data Integration Workshop, Rutgers, NJ.
June 2013, "MLBase: A User-friendly System for Distributed Machine Learning," AT&T, Bedminster, NJ.
June 2013, "IBM Workshop on Big Data Analytics," IBM, Yorktown Heights, NY.
October 2013, "MLbase," Simons Institute for the Theory of Computing, Berkeley, CA.
October 2013, "MLbase," Google, Seattle, WA.
October 2013, "MLbase," University of Washington, Seattle, WA.
October 2013, "The 1-Phase Commit Protocol," University of Waterloo, Waterloo, Canada.
December 2013, "MDCC: Multi Data Center Consistency," Google, Mountain View, CA.
April 2014, "Tupeware," Intel, San Jose, CA.
June 2014, Panel: "Should we all be teaching 'Intro to Data Science' instead of 'Intro to Databases'?" (**panelist**), SIGMOD, Snowbird, UT.

** Outgrowth of supervised student research

Publications of Tim Kraska

April 2014, “An Integrated Approach to Data Science,” University of Pennsylvania, Philadelphia, PA.

August 2014, “Tupleware,” Oracle, Redmond, CA.

August 2014, “Tupleware,” SAP, Palo Alto, CA.

August 2014, “Tupleware,” VMware, Palo Alto, CA.

October 2014, “Tupleware: Redefining Modern Analytics” Hardcore Data Science, Strata + Hadoop World NYC, New York, NY.

June 2015, Panel: “Machine Learning and Databases: The Sound of Things to Come or a Cacophony of Hype?” (**panelist**), SIGMOD, Melbourne, Australia.

June 2015, Panel: “New Researcher Symposium: How To Fail” (**panelist**), SIGMOD, Melbourne, Australia.

March 2015, “Tupleware,” NAVSEA, Newport, RI.

May 2015, “Tupleware and MLBase,” New England Machine Learning Day (NEML), Boston, MA.

May 2015, “The End of Slow Networks: It’s Time for a Redesign,” ACM IEEE Computer Society Joint Seminar, Boston, MA.

June 2015, “The End of Slow Networks,” Oracle, Redmond, CA.

September 2015, “Tupleware,” IEEE High Performance Extreme Computing Conference, Waltham, MA.

November 2015, “Towards Human-In-The-Loop Data Exploration,” Worcester Polytechnic Institute, Worcester, MA.

November 2015, “Democratizing Data Science,” Technical University Munich, Germany.

January 2016, “Making Distributed Transactions Scale,” New England Database Day (NEDBday), Boston, MA.

January 2016, “Interactive Data Science,” Oracle, Redmond, CA.

May 2016, Panel: “Dark Data: Are we solving the right problems?” (**moderator**), IEEE International Conference on Data Engineering, Helsinki, Finland.

June 2016, “The End of Slow Networks: It's Time for a Redesign,” Airforce Young Investigator meeting, Arlington, VA.

July 2016, “The End of Slow Networks,” Mellanox, Sunnyvale, CA.

August 2016, “Interactive Data Science,” Aalborg University, Aalborg, Denmark.

August 2016, “How Not to Fail” (keynote), VLDB Ph.D. workshop, Delhi, India.

August 2016, “Interactive Data Science,” IIT Delhi, India.

September 2016, “Interactive Data Science,” Google, Mountain View, CA.

October 2016, “Interactive Data Science,” University of Massachusetts, Amherst, MA.

Publications of Tim Kraska

November 2016, “Interactive Data Science,” RWTH Aachen, Germany.

January 2017, Panel: “Machine learning and storage: Why is our community not at the table?” (**panelist**), Conference on Innovative Data Systems Research (CIDR), Santa Cruz, CA.

January 2017, “Quantifying the Uncertainty in Data Exploration,” University of Washington, Seattle, WA.

January 2017, Panel: “Data Sharing and Integration: Infrastructure and Best Practices” (**panelist**, invited on behalf of Governor O’Malley, Mayor Murray of Seattle, and the University of Washington), Big Data + Human Services Workshop, Seattle, WA.

April 2017, Panel: “Data science education: We’re missing the boat, again” (**panelist**), IEEE International Conference on Data Engineering (ICDE), San Diego, CA.

May 2017, “Approximate Query Processing for Interactive Data Science” (**keynote**), SIGMOD, Chicago, IL.

June 2017, “Interactive Data Science,” IBM Research, San Jose, CA.

June 2017, “Interactive Data Science,” at University of California, Berkeley, CA.

September 2017, “Interactive Data Science,” Amazon, Berlin, Germany.

October 2017, “Interactive Data Science” (**keynote**), Workshop on Data Systems for Interactive Analysis co-located with VIS 2017, Phoenix, AZ.

November, 2018, “Interactive Data Science,” Dagstuhl Seminar on Connecting Visualization and Data Management Research, Dagstuhl, Germany.

January 2018, “Learned Index Structures,” Facebook, New York, NY.

January 2018, “Learned Index Structures,” IBM, San Jose, CA.

January 2018, “Learned Index Structures,” New England Database Day (NEDB Day), Boston, MA.

January 2018, “Learned Index Structures,” Microsoft Research, Redmond, CA.

January 2018, “Learned Index Structures,” Cloudera, San Francisco, CA.

April 2018, “Learned Index Structures,” Extreme Large Database Conference (XLDB), Palo Alto, CA.

April 2018, “The End of a Myth: Distributed Transactions Can Scale,” Future Cloud Workshop, Boston, MA.

May 2018, “Machine learning just ate algorithms in one large bite” (**keynote**), O’Reilly Artificial Intelligence Conference, New York, NY.

May 2018, “Learned index structures,” O’Reilly Artificial Intelligence Conference, New York, NY.

May 2018, “Learned index structures,” Data Science Summit, Tel Aviv, Israel.

May 2018, “Interactive Data Science,” Tel Aviv University, Tel Aviv, Israel.

Publications of Tim Kraska

- June 2018, “Interactive Data Science,” SAP, Waldorf, Germany.
- June 2018, “Learning Data Systems Components” (**keynote**), Workshop on Exploiting Artificial Intelligence Techniques for Data Management (aiDM) in conjunction with SIGMOD, Houston, TX.
- June 2018, Panel: “Data-Driven and Machine Learning Empowered Autonomous DB Systems” (**panelist**), Alibaba workshop at SIGMOD, Houston, TX.
- July 2018, “Learned index structures,” Microsoft Research Faculty Summit, Redmond, WA.
- August 2018, “Interactive Data Science” (**award talk**), VLDB, Rio de Janeiro, Brazil.
- November 2018, “ML for Systems and Systems for ML,” Oracle, Redwood City, CA.
- November 2018, “Computing Systems Disruptions,” Disruptive Technology Review at Microsoft in front of Microsoft’s leadership team including **Satya Nadella (CEO)** and **Kevin Scott (CTO)**, Redmond, WA.
- December 2018, “Computing Systems Disruptions,” TTI Vanguard, San Francisco, CA.
- January 2019, “SageDB – Towards an Instance Customized DB,” Conference on Innovative Data Systems Research (CIDR), Asilomar, CA.
- March 2019, “SageDB – Towards an Instance Optimized DB,” Google, Cambridge, MA.
- March 2019, “SageDB – Towards an Instance Optimized DB,” University of California, Berkeley, CA.
- April 2019, “The Case for Learned Algorithms, Data Structures, and Systems” (**keynote**), New England Networking and Systems Day, Cambridge, MA.
- May 2019, “The Case for Learned Algorithms, Data Structures, and Systems,” University of Wisconsin-Madison, Madison, WI.
- May 2019, “The Case for Learned Algorithms, Data Structures, and Systems,” ETH Zurich, Switzerland.
- May 2019, “The Case for Learned Algorithms, Data Structures, and Systems,” Technical University of Munich, Munich, Germany.
- June 2019, “Interactive Data Science,” Tsinghua University, Beijing, China.
- June 2019, “Interactive Data Science,” BOE, Beijing, China.
- June 2019, “Interactive Data Science,” Talking Data, Beijing, China.
- June 2019, “The Case for Learned Algorithms, Data Structures, and Systems” (**keynote**), O’Reilly AI Conference, Beijing, China.
- June 2019, “The Case for Learned Algorithms, Data Structures, and Systems,” SAP, Waldorf, Germany.
- July 2019, Tutorial: “From Auto Tuning One Size Fits All to Self-designed and Learned Data Intensive Systems,” SIGMOD, Amsterdam, Netherlands.
- August 2019, “The Case for Learned Algorithms, Data Structures, and Systems,” TTI-Chicago Summer Workshop on Learning Based Algorithm Workshop, Chicago, IL.

Publications of Tim Kraska

August 2019, “The Case for Learned Algorithms, Data Structures, and Systems,” Google, Kirkland, WA.

August 2019, “SchengenDB: A Data Protection Database Proposal.” Poly/DMAH@VLDB 2019, Los Angeles, CA.

September 2019, “Interactive Data Science,” IBM AI Systems Day, Cambridge, MA.

November 2019, “Fast Networks and the Next Generation of Transactional Database Systems,” International Workshop on High Performance Transaction Systems (HPTS), Asilomar, CA.

December 2019, “The Case for Learned Algorithms, Data Structures, and Systems,” Google, Cambridge, MA.

February 2020, “Towards Learned Algorithms, Data Structures, and Systems” (**keynote**), AAAI Cloud Intelligence Workshop, New York, NY.

April 2020, “Towards Learned Algorithms, Data Structures, and Systems” (**online keynote with over 10k people watching the video stream**), Database forum of the Chinese database society.

August 2020, “Towards Learned Algorithms, Data Structures, and Systems” (**online keynote**), Applied AI for Database Systems and Applications (AIDB) @VLDB.

Theses Supervised by Tim Kraska

	Total	Completed	In Progress
Bachelor's	2	2	0
Master's	18	18	0
MEng	0	0	0
Engineer's	0	0	0
Doctoral			
As Supervisor	11	5	6
As Reader	9	7	2

Bachelor's Theses

Herlihy, Anna, "Compilation Techniques for Distributed Analytics," Brown University, 2014.

Correa Orozco, David, "TeachWithGlass: Improving the Teaching Experience through Google Glass," Brown University, 2015.

Master's Theses

Tarnutzer, Christian, "Streaming XQueryP," ETHZ, 2008.

Rutishauser, Jonas, "Lazy evaluation of semantic mapping rules," ETHZ, 2008.

Unternährer, Thomas, "Cloud bursting for cloudy," ETHZ, 2009.

Egger, Daniel, "SQL in the Cloud," ETHZ, 2009.

Zhang, Haoning, "Indexed Semantic Mapping Rules," ETHZ, 2009.

Moreno, João, "A testing framework for cloud storage systems," ETHZ, 2010.

Merkli, Stephan, "Streaming in the cloud," ETHZ, 2010.

Kirkwood, Rachael, "Analyzing the Nutrition of Children's Menus: Obtaining the Unstructured Dataset," Brown University, 2014.*

Askari, Mani, "Postgres on Infiniband," Brown University, 2015.*

Lillie, Giselle L., "Crowdsourcing Smart Home Use Cases," Brown University, 2015.*

Dadashov, Elkhan, "Spotlytics," Brown University, 2015.*

Shao, Qiming, "Spark, BlinkDB and Sampling," Brown University, 2016.*

Jin, Yan, "ML Transactions," Brown University, 2017.*

Ly, Andy, "AutoML in Vizdom," Brown University, 2017.*

Zhang, Zeyuan, "Democratizing Data Science through Interactive Curation of ML Pipelines," MIT, 2018.

* Independent study. (Brown University does not require CS Master's students to write a thesis; instead, students are encouraged to complete an independent study.)

Theses Supervised by Tim Kraska

Sedlar, Katie, “Exploring Learned Indexes for Approximate Query Processing and Visual Interfaces,” MIT, 2019.

Lam, Jason, “Predicate Push-Down for Interactive Data Exploration,” MIT, 2020.

Runnels, Wesley, “Automatic Feature Engineering,” MIT, 2020.

MEng Theses

None.

Engineer’s Theses

None.

Doctoral Theses, Supervisor

Chung, Yeounoh, “Quantifying Uncertainty in Data Exploration,” Brown University, 2018.

Galakatos, Alex, “Accelerating Interactive Data Exploration,” Brown University, 2018.

Crotty, Andrew, “Code Generation for In-Memory Data Analytics,” Brown University, 2018.

Zraggen, Emanuel, “Towards Accessible Data Analysis,” Brown University, 2018. (Co-advised with Andries Van Dam, Brown University)

Zamanian, Erfan, “Distributed Transaction Processing on Modern RDMA-enabled Networks,” Brown University, 2020.

Spiegelberg, Leonhard, Brown University, expected 2021. (Co-advised with Malte Schwarzkopf, Brown University)

Kristo, Ani, Brown University, expected 2021.

Vaidya, Kapil, MIT, expected 2022.

Ding, Jialin, MIT, expected 2022.

Zeng, Anna, MIT, expected 2023.

Sacha Servan-Schreiber, MIT, expected 2023. (Co-supervised with Srinu Devadas)

Doctoral Theses, Reader

Gong, Weiwei, “Improving Main-Memory Database (MMDB) Concurrency,” University of Massachusetts Boston, 2014.

Kalinin, Alexander, “Integrated Search and Exploration Over Large Multidimensional Data,” Brown University, September 2016.

Cao, Lei, “Outlier Detection in Big Data,” Worcester Polytechnic Institute, March 2016.

Dursun, Kayhan, “Query Processing for Data Analytics on Modern Multicore Systems,” Brown University, 2019.

Hu, Kevin, “Automating Data Visualization through Recommendation,” MIT, 2019.

De Stefani, Lorenzo, “Probabilistic approaches for rigorous and efficient analysis of statistical properties of large datasets,” Brown University, 2020.

Eichmann, Philipp, “Visual Methods for Exploring and Forecasting Time Series,” Brown University, 2020.

Theses Supervised by Tim Kraska

Mao, Hongzi, "Building Self-Improving Network Computer Systems," MIT, expected 2020.

Lu, Yi, "Fast Transactions in Distributed and Highly Available Databases," MIT, expected 2020.

Postdoctoral Associates and Fellows Supervised by Tim Kraska

Current Postdocs

Name	Dates of Appointment	PhD Granting Institution
Marcus, Ryan	April 2019 - present	Brandeis University
Sabek, Ibrahim	Feb. 2020 - present	University of Minnesota
Kipf, Andreas	April 2020 - present	TU Munich

Previous Postdocs

Name	Current Title	Current Employer
Zraggen, Emanuel	CTO	einblick analytics, inc.

Teaching Evaluations of Tim Kraska

<i>Term</i>	<i>Course Number</i>	<i>Course Title</i>	<i>Role</i>	<i>Course Type</i>	<i># Students Registered</i>	<i># Survey Responses</i>	<i>Instructor's Evaluation</i>	<i>Course Evaluation</i>	<i>Scale</i>
ST13	CSCI1580	Information Retrieval and Web Search (Brown University)	Instructor	Lecture	23	18	1.53	2.53	5*
FT13	CSCI1950W	Topics in Data Science (Brown University)	Instructor	Seminar	20	15	2.4	2.67	5*
ST14	CSCI1951A	Data Science (Brown University)	Instructor	Lecture	44	34	1.66	1.81	5*
ST15	CSCI1951A	Data Science (Brown University)	Instructor	Lecture	114	89	1.55	1.85	5*
ST16	CSCI1951A	Data Science (Brown University)	Instructor	Lecture	188	120	2.22	2.54	5*
FT16	CSCI2951V	Systems for Interactive Data Exploration (Brown University)	Instructor	Seminar	11	7	1.17	1.29	5*
ST18	6.033	Computer System Engineering	Recitation Instructor	Lecture	393	161	6.6	5.3	7
FT18	6.814	Database Systems	Instructor	Lecture	48	17	5.6	5.6	7
ST19	6.033	Computer System Engineering	Recitation Instructor	Lecture	358	159	6.4	5.1	7
FT19	6.S080	Software Systems for Data Science	Instructor	Lecture	74	20	6.6	6.1	7

* Very effective (1) – very ineffective (5)

Photograph of Tim Kraska

