Jialin Ding

jialind@amazon.com jialinding.github.io Revised 11/2023 **EDUCATION** Massachusetts Institute of Technology 2018-2022 PhD, Computer Science Advisor: Tim Kraska **Stanford University** 2014 - 2018Bachelor of Science with Distinction, Electrical Engineering Minor in Economics PUBLICATIONS 1. Automated Multidimensional Data Layouts in Amazon Redshift. Jialin Ding, Matt Abrams, Sanghita Bandyopadhyay, Luciano Di Palma, Yanzhu Ji, Davide Pagano, Gopal Paliwal, Panos Parchas, Pascal Pfeil, Orestis Polychroniou, Gaurav Saxena, Aamer Shah, Amina Voloder, Sherry Xiao, Davis Zhang and Tim Kraska. Under submission to SIGMOD 2024 Industrial Track. 2. Learning Bit Allocations for Z-Order Layouts in Analytic Data Systems. Jenny Gao, Jialin Ding, Sivaprasad Sudhir and Samuel Madden. Systems for ML Workshop @ NeurIPS 2023. 3. SageDB: An Instance-Optimized Data Analytics System. Jialin Ding, Ryan Marcus, Andreas Kipf, Vikram Nathan, Aniruddha Nrusimha, Kapil Vaidya, Alexander van Renen and Tim Kraska. VLDB 2023. 4. APEX: A High-Performance Learned Index on Persistent Memory. Baotong Lu, Jialin Ding, Eric Lo, Umar Farooq Minhas and Tianzheng Wang. VLDB 2022. 5. Self-Organizing Data Containers. Samuel Madden, Jialin Ding, Tim Kraska, Sivaprasad Sudhir, David Cohen, Timothy Mattson and Nesime Tatbul. CIDR 2022. 6. Tsunami: A Learned Multi-dimensional Index for Correlated Data and Skewed Workloads. Jialin Ding, Vikram Nathan, Mohammad Alizadeh and Tim Kraska. VLDB 2021. 7. Instance-Optimized Data Layouts for Cloud Analytics Workloads. Jialin Ding, Umar Farooq Minhas, Badrish Chandramouli, Chi Wang, Yinan Li, Ying Li, Donald Kossmann, Johannes Gehrke and Tim Kraska. SIGMOD 2021. 8. Cortex: Harnessing Correlations to Boost Query Performance. Vikram Nathan, Jialin Ding, Tim Kraska and Mohammad Alizadeh. CoRR 2020. 9. The Case for Learned Spatial Indexes. Varun Pandey, Alexander van Renen, Andreas Kipf, Ibrahim Sabek, Jialin Ding and Alfons Kemper. AIDB Workshop @ VLDB 2020. 10. ALEX: An Updatable Adaptive Learned Index. Jialin Ding, Umar Farooq Minhas, Jia Yu, Chi Wang, Jaeyoung Do, Hantian Zhang, Yinan Li, Badrish Chandramouli, Johannes Gehrke, Donald Kossmann, David Lomet and Tim Kraska. SIGMOD 2020.

- 11. Learning Multi-dimensional Indexes. Vikram Nathan^{*}, Jialin Ding^{*}, Mohammad Alizadeh and Tim Kraska. *SIGMOD 2020.*
- 12. LISA: Towards Learned DNA Sequence Search. Darryl Ho, Jialin Ding, Sanchit Misra, Nesime Tatbul, Vikram Nathan, Vasimuddin Md and Tim Kraska. Systems for ML Workshop @ NeurIPS 2019. Oral Presentation.
- 13. Learning Multi-dimensional Indexes. Vikram Nathan^{*}, Jialin Ding^{*}, Mohammad Alizadeh and Tim Kraska. *ML for Systems Workshop @ NeurIPS 2019. Oral Presentation.*
- 14. SageDB: A Learned Database System. Tim Kraska, Mohammad Alizadeh, Alex Beutel, Ed Chi, Jialin Ding, Ani Kristo, Guillaume Leclerc, Samuel Madden, Hongzi Mao and Vikram Nathan. *CIDR 2019.*
- 15. A Machine-compiled Database of Genome-wide Association Studies. Volodymyr Kuleshov, Jialin Ding, Christopher Vo, Braden Hancock, Alexander Ratner, Yang Li, Christopher R, Serafim Batzoglou and Michael Snyder *Nature Communications 2019.*
- 16. Moment-Based Quantile Sketches for Efficient High Cardinality Aggregation Queries. Edward Gan, Jialin Ding, Kai Sheng Tai, Vatsal Sharan and Peter Bailis. *VLDB 2018*.
- 17. Efficient Mergeable Quantile Sketches using Moments. Edward Gan, Jialin Ding and Peter Bailis. SysML 2018. Extended Abstract.
- 18. MacroBase: Prioritizing Attention in Fast Data. Firas Abuzaid, Peter Bailis, Jialin Ding, Edward Gan, Samuel Madden, Deepak Narayanan, Kexin Rong and Sahaana Suri. *TODS 2018.*
- A Machine-Compiled Database of Genome-Wide Association Studies. Volodymyr Kuleshov, Jialin Ding, Braden Hancock, Alexander Ratner, Christopher Re, Serafim Batzoglou and Michael Snyder. *ISMB 2017. Short Paper.*

TEACHING AND SERVICE	• Reviewer, VLDB Demo Track, 2022-2023			
	 Reviewer, VLDB Journal, 2023 Teaching Assistant, 6.887: Machine Learning for Systems, Fall 2021 Student Volunteer, VLDB 2021 			
			• Reviewer, TKDE 2020	
			INVITED	Towards Practical Instance-Optimized Systems
	TALKS	Facebook/Meta	February 2022	
	Learned Index Structures for Dynamic and Multi-Dimensional Data			
	University of Washington (NWDS Seminar)	February 2021		
	Instance-optimized Indexing and Storage			
	Cornell University (DB Seminar)	October 2020		
	LADSIOS Workshop @ VLDB	August 2021		
	Stanford Systems Seminar	March 2022		
	Learning Multi-dimensional Indexes			
	Boston University (MiDAS Seminar)	April 2020		
	New England Database Day	January 2020		

• Facebook Fellowship, 2021–2023

- NSF Graduate Research Fellowship Program, Honorable Mention, 2018
- MIT Jacobs Presidential Fellowship, 2018

INDUSTRY EXPERIENCE

2022–Present Applied Scientist II, Amazon • Conduct research on instance-optimized database systems as part of a research team embedded within AWS Redshift.

Research Intern, Microsoft Research, Redmond Summer 2020

• Led research on a data layout framework for cloud analytics services, with applications to Azure Synapse, resulting in a SIGMOD 2021 publication.

Research Intern, Microsoft Research, Redmond Summer 2018

• Led research on an updatable learned index structure, resulting in a SIGMOD 2020 publication.

Software Engineer Intern, Google

• As part of Google Safe Browsing, implemented a MapReduce pipeline to integrate Chrome browser incident data into the evaluation of user downloads.

Software Engineer Intern, Thumbtack

• Worked on SEO, automatic text generation, and recommendation systems.

FELLOWSHIPS

AND AWARDS

Summer 2016

Summer 2015