Ryan Marcus

Assistant Professor of Computer Science
University of Pennsylvania
Philadelphia, PA
`rcmarcus@seas.upenn.edu`

RESEARCH INTERESTS

Machine learning for systems, especially databases and clouds. Machine programming, parallel and distributed systems, high performance computing, programming languages & compilers.

EDUCATION

Postdoc, Computer Science, MIT   July 2023
Supervisor: Tim Kraska   Cambridge, MA. USA

Ph.D. Computer Science, Brandeis University   April 2019
Advisor: Olga Papaemmanouil   Waltham, MA. USA

M.A. Computer Science, Brandeis University   January 2016
Bachelors, University of Arizona   May 2014

PUBLICATIONS

1. Lyric Doshi, Vincent Zhuang, Gaurav Jain, <u>Ryan Marcus</u>, Haoyu Huang, Deniz Altinbuken, Eugene Brevdo, Campbell Fraser. "Kepler: Robust Learning for Faster Parametric Query Optimization." ACM Special Interest Group in Data Management (**SIGMOD 23)**, 2023.

2. Gaurav Saxena, Mohammad Rahman, Naresh Chainani, Chunbin Lin, George Caragea, Fahim Chowdhury, <u>Ryan Marcus</u>, Tim Kraska, Ippokratis Pandis, Balakrishnan (Murali) Narayanaswamy. "Auto-WLM: Machine Learning Enhanced Workload Management in Amazon Redshift." ACM Special Interest Group in Data Management (**SIGMOD 23)**, 2023.

3. Christoph Anneser, Nesime Tatbul, David Cohen, Zhenggang Xu, Prithvi Pandian, Nikolay Leptev, <u>Ryan Marcus</u>. "AutoSteer: Learned Query Optimization for Any SQL Database." Proceedings of the VLDB Endowment (**VLDB 23**), Volume 16, Issue 12, 2023

4. Chenyuan Wu, Bhavana Mehta, Mohammad Javad Amiri, <u>Ryan Marcus</u>, Boon Thau Loo. "AdaChain: A Learned Adaptive Blockchain." Proceedings of the VLDB Endowment (**VLDB 23**), Volume 16, Issue 8, 2023.

5. Parimarjan Negi, Ziniu Wu, Andreas Kipf, Nesime Tatbul, <u>Ryan Marcus</u>, Sam Madden, Tim Kraska, Mohammad Alizadeh. "Robust Query Driven Cardinality Estimation under Changing Workloads." Proceedings of the VLDB Endowment (**VLDB 23**), Volume 16, Issue 6, 2023.

6. Jialin Ding, <u>Ryan Marcus</u>, Andreas Kipf, Vikram Nathan, Aniruddha Nrusimha, Kapil Vaidya, Alexander van Renen, Tim Kraska. "SageDB: An Instance-Optimized Data Analytics System." Proceedings of the VLDB Endowment (**VLDB 22**), Volume 15, Issue 13, 2022.

7. <u>Ryan Marcus</u>, Parimajan Negi, Hongzi Mao, Nesime Tatbul, Mohammad Alizadeh, Tim Kraska. "Bao: Making Learned Query Optimization Practical." ACM Special Interest Group in Data Management (**SIGMOD 21**), 2021. ★ *best paper award* ★ *SIGMOD research highlights*

8. Parimarjan Negi, Matteo Interlandi, <u>Ryan Marcus</u>, Mohammad Alizadeh, Tim Krask, Marc Friedman Alekh Jindal. "Steering Query Optimizers: A Practical Take on Big Data Workloads." ACM Special Interest Group in Data Management, Industry (**SIGMOD 21**), 2021. ★ *best paper honorable mention*

9. <u>Ryan Marcus</u>, Andreas Kipf, Alex van Renen, Mihail Stoian, Sanchit Misra, Alfons Kemper, Thomas Neumann, Tim Kraska. "Benchmarking Learned Indexes." Proceedings of the VLDB Endowment (**VLDB 21**), Volume 14, Issue 1, 2021.

10. Parimarjan Negi, <u>Ryan Marcus</u>, Andreas Kipf, Hongzi Mao, Nesime Tatbul, Tim Kraska, Mohammad Alizadeh. "Flow-Loss: Learning Cardinality Estimates That Matter." Proceedings of the VLDB Endowment (**VLDB 21**), Volume 14, Issue 11, 2021.

11. Nadiia Chepurko, <u>Ryan Marcus</u>, Emanuel Zgraggen, Raul Castro Fernandez, Tim Kraska, David Karger. "ARDA: Automatic Relational Data Augmentation for Machine Learning." Proceedings of the VLDB Endowment (**VLDB 20**), Volume 12, Issue 11, 2020.

12. Solomon Garber, <u>Ryan Marcus</u>, Antonella DiLillio, James Storer. "Low Bitrate Compression of Video with Dynamic Background." Data Compression Conference (**DCC 20**), 2020.

13. Hongzi Mao, Parimarjan Negi, Akshay Narayan, Hanrui Wang, Jiacheng Yang, Haonan Wang, <u>Ryan Marcus</u>, Ravichandra Addanki, Mehrdad Khani, Songtao He, Vikram Nathan, Frank Cangialosi, Shaileshh Venkatakrishnan, Wei-Hung Weng, Song Han, Tim Kraska, Mohammad Alizadeh. "Park: An Open Platform for Learning Augmented Computer Systems." 33rd Conference on Neural Information Processing Systems (**NeurIPS 19**), 2019.

14. <u>Ryan Marcus</u>, Parimarjan Negi, Hongzi Mao, Chi Zhang, Mohammad Alizadeh, Tim Kraska, Olga Papaemmanouil, Nesime Tatbul. "Neo: A Learned Query Optimizer." Proceedings of the VLDB Endowment (**VLDB 19**), Volume 12, Issue 11, 2019.

15. <u>Ryan Marcus</u>, Olga Papaemmanouil. "Plan-Structured Deep Neural Network Models for Query Performance Prediction." Proceedings of the VLDB Endowment (**VLDB 19**), Volume 12, Issue 11, 2019.

16. Bailu Ding, Sudipto Das, <u>Ryan Marcus</u>, Wentao Wu, Surajit Chaudhuri, Vivek R. Narasayya. "AI Meets AI: Leveraging Query Executions to Improve Index Recommendations." ACM Special Interest Group in Data Management (**SIGMOD 19**), 2019.

17. Solomon Garber, Aaditya Prakash, <u>Ryan Marcus</u>, Antonella DiLillio, James Storer. "Compact Representations of Dynamic Video Background Using Motion Sprites." Data Compression Conference (**DCC 19**), 2019.

18. <u>Ryan Marcus</u>, Olga Papaemmanouil. "Towards a Hands-Free Query Optimizer through Deep Learning." 9th Biennial Conference on Innovative Data Systems Research (**CIDR 19**), 2019.

19. <u>Ryan Marcus</u>, Olga Papaemmanouil, Sofiya Semenova, Solomon Garber. "NashDB: An Economic Approach to Fragmentation, Replication and Provisioning for Elastic Databases." 37th ACM Special Interest Group in Data Management (**SIGMOD 18**), 2018.

20. <u>Ryan Marcus</u>, Olga Papaemmanouil. "Releasing Cloud Databases from the Chains of Performance Prediction Models." 8th Biennial Conference on Innovative Data Systems Research (**CIDR 17**), 2017.

21. <u>Ryan Marcus</u>, Olga Papaemmanouil. "WiSeDB: A Learning-Based Workload Management Advisor for Cloud Databases." Proceedings of the VLDB Endowment (**VLDB 16**), Volume 9, Issue 10, 2016.

DEMOS

1. Christoph Anneser, Mario Petruccelli, Nesime Tatbul, David Cohen, Zhenggang Xu, Prithviraj Pandian, Nikolay Laptev, <u>Ryan Marcus</u>, Alfons Kemper. "QO-Insight: Inspecting Steered Query Optimizers (Demo)." VLDB Endowment (**VLDB 23)**, Volume 16, Issue 12, 2023.

2. <u>Ryan Marcus</u>, Emily Zhang, Tim Kraska. "CDFShop: Exploring and Optimizing Learned Index Structures (Demo)." ACM Special Interest Group in Data Management (**SIGMOD 20**), 2020.

3. <u>Ryan Marcus</u>, Chi Zhang, Shuai Yu, Geoffrey Kao, Olga Papaemmanouil. "NashDB: Fragmentation, Replication, and Provisioning using Economic Methods (Demo)." VLDB Endowment (**VLDB 19**), Volume 12, Issue 10, August 2019.

4. <u>Ryan Marcus</u>, Sofiya Semenova, Olga Papaemmanouil. "A Learning-based Service for Cost and Performance Management of Cloud Databases (Demo)." IEEE International Conference on Data Engineering (**ICDE 17**), 2017.

WORKSHOP PAPERS

1. <u>Ryan Marcus</u>. "Learned Query Superoptimization." Workshop on Applied AI for Database Systems and Applications (**AIDB @ VLDB 23**), 2023.

2. Bhavana Mehta, Neelesh Chinnakonda Ashok Kumar, Prashanth S Iyer, Mohammad Javad Amiri, Boon Thau Loo, <u>Ryan Marcus</u>. "Towards Adaptive Fault-Tolerant Sharded Databases." Workshop on Applied AI for Database Systems and Applications (**AIDB @ VLDB 23**), 2023.

3. Andreas Kipf, Dominik Horn, Pascal Pfeil, <u>Ryan Marcus</u>, Tim Kraska. "LSI: A Learned Secondary Index Structure." Fifth International Workshop on Exploiting Artificial Intelligence Techniques for Data Management (**aiDM @ SIGMOD 22**), 2022.

4. Mihail Stoian, Andreas Kipf, <u>Ryan Marcus</u>, Tim Kraska. "PLEX: Towards Practical Learned Indexing." The 3rd International Workshop on Applied AI for Database Systems and Applications (**AIDB @ VLDB 21**) 2021.

5. Lujing Cen, Andreas Kipf, <u>Ryan Marcus</u>, Tim Kraska. "LEA: A Learned Encoding Advisor for Column Stores." Fourth International Workshop on Exploiting Artificial Intelligence Techniques for Data Management (**aiDM @ SIGMOD 21**), 2021.

6. Chi Zhang, <u>Ryan Marcus</u>, Anat Kleiman, Olga Papaemmanouil. "Buffer Pool Aware Query Scheduling via Deep Reinforcement Learning." The 2nd International Workshop on Applied AI for Database Systems and Applications (**AIDB @ VLDB 20**), 2020.

7. Lujing Cen, <u>Ryan Marcus</u>, Hongzi Mao, Justin Gottschlich, Mohammad Alizadeh, Tim Kraska. "Learned Garbage Collection." The 4th Annual ACM SIGPLAN Machine Learning and Programming Languages Workshop (**MAPL @ PLDI 20**), 2020.

8. Andreas Kipf*, <u>Ryan Marcus</u>*, Alexander van Renen*, Mihail Stoian, Alfons Kemper, Tim Kraska, Thomas Neumann. "RadixSpline: A Single-Pass Learned Index." Third International Workshop on Exploiting Artificial Intelligence Techniques for Data Management (**aiDM @ SIGMOD 20**), 2020. (* equal contribution)

9. Parimarjan Negi, <u>Ryan Marcus</u>, Hongzi Mao, Nesime Tatbul, Tim Kraska, Mohammad Alizadeh. "Cost-Guided Cardinality Estimation: Focus Where it Matters." Workshop on Self-managing Databases at the IEEE International Conference on Data Engineering (**SMDB @ ICDE 20**), 2020.

10. Andreas Kipf, <u>Ryan Marcus</u>, Alexander van Renen, Mihail Stoian, Alfons Kemper, Tim Kraska, Thomas Neumann. "Learning to Search." Machine Learning for Systems Workshop @ 33rd Conference on Neural Information Processing Systems (**ML4Sys @ NeurIPS 19**), 2019.

11. Hongzi Mao, Akshay Narayan, Parimarjan Negi, Hanrui Wang, Jiacheng Yang, Haonan Wang, Mehrdad Khani, Songtao He, Ravichandra Addanki, <u>Ryan Marcus</u>, Frank Cangialosi, Wei-Hung Weng, Song Han, Tim Kraska, Mohammad Alizadeh. "Park: An Open Platform for Learning Augmented Computer Systems." International Conference on Machine Learning Workshop on Reinforcement Learning for Real Life (**RL4RL @ ICML 19**), 2019. ★ *best paper award*

12. <u>Ryan Marcus</u>, Olga Papaemmanouil. "Deep Reinforcement Learning for Join Order Enumeration." First International Workshop on Exploiting Artificial Intelligence Techniques for Data Management (**aiDM @ SIGMOD 18**), 2018.

13. <u>Ryan Marcus</u>, Olga Papaemmanouil. "Workload Management for Cloud Databases via Machine Learning." Workshop on Cloud Data Management at the IEEE International Conference on Data Engineering (**CloudDM @ ICDE 16**), 2016.

HONORS & AWARDS

SIGMOD Record Research Highlights	ACM SIGMOD, 2022
SIGMOD Best Paper	ACM SIGMOD, 2021
RL4RL @ ICML Best Paper Award	ACM ICML, 2019
Computer Science Outstanding Teaching Fellow Award	Brandeis University, 2015
Michtom Fellowship	Brandeis University, 2014
CEDA Summa Cum Laude National Debate Scholar	Idaho State University, 2013

EMPLOYMENT

- **University of Pennsylvania, Assistant Professor [2023-]**: Supervised graduate students, lead a research lab, prepared papers for publications, taught undergraduate and graduate courses.

- **AWS, Applied Scientist [2022-2023]**: Worked on query prediction models for Redshift AutoWLM. Designed and implemented an ML model to predict query scalability. Prepared papers for publication.

- **MIT, Postdoc [2019-2022]**: Applying machine learning to database systems, including query optimization, index structures, cardinality estimation. Making database systems learn from their mistakes and customize themselves to the user's workload.

- **Intel Labs Scientist [2020-2022]**: Researching machine programming, the automation of software construction. Code similarity, automatic garbage collection, autonomous systems.

- **Brandeis University, Ph.D. Student and Research Assistant [2014-2019]**: WiSeDB, a learning-based cost and performance management service for cloud databases. Supervised and reinforcement learning techniques to automate resource provisioning and workload scheduling by generating models that minimize monetary cost, while supporting performance SLAs.

- **Google, Ph.D. Intern [2018]**: Improving ad spend and revenue modeling for the Attribution 360 team. Developed convergence detection method to speed up model training 33%. Developed aggregation methods to model performance in low-variance scenarios.

- **Microsoft Research, Ph.D. Intern [2017]**: ML techniques to predict query performance regressions due to physical design changes. Online transfer learning via learned kernels to adapt models to new datasets.

- **HP Vertica, Software Development Intern [2016]**: Design and implementation of improved query plan execution algorithms, achieving a 4x performance boost for rollup analytic queries.

- **Los Alamos National Lab, HPC Engineer [2009-2015]**: Developed a machine-learning framework for automatic performance analysis, currently used to analyze high-performance hydrodynamics codes. Parallelized serial algorithms on GPUs. Designed novel median filter approximation algorithm, improved 3D image reconstruction performance by 3x. Implemented exascale co-design neutron transport code.

TEACHING EXPERIENCE

University of Pennsylvania

Co-instructor, Big Data Analytics (ongoing)	Spring 2024
Instructor, Topics in Database Management Systems (ongoing)	Fall 2023

Brandeis University

Instructor, Database Management Systems (4.81 / 5, n = 108)	Spring 2019
T.A. for Networked Information Systems (4.6 / 5, n = 6)	Spring 2018
T.A. for Advanced Programming Techniques (4.5-4.7 / 5, n = 32-51)	Spring 2016 – Fall 2017
T.A. for Databases Management Systems (4.4 / 5, n = 45)	Fall 2015
T.A. for Theory of Computation (4.5 / 5, n = 70)	Spring 2015
T.A. for Operating Systems (4.8 / 5, n = 56)	Fall 2014

SERVICE

- Brandeis Ph.D. Seminar Organizer 2016 - 2019 http://rm.cab/seminars

VLDB 2025	Program committee
SIGMOD 2025	Program committee
SoCC 2023	Program committee
VLDB 2023	aiDM
VLDB 2022	Program committee
OOPSLA 2021	Program committee
ICDE 2021	Program committee
SIGMOD 2021	Program committee
ICDE 2020	SMDB
SIGMOD 2020	Demos, aiDB, DEEM
VLDB 2020	Program committee, aiDM
PLDI 2020	MAPL
ICDE 2019	Demos
SIGMOD 2019	aiDB
VLDB 2019	aiDM

INVITED TALKS AND SEMINARS

- "Machine Learning Powered Query Optimization." University of Washington, Allen School Colloquium. December 15th, 2022. (slides) (video)
- "Machine Learning Meets Query Optimization." Cornell University, database group seminars. March 27th, 2023. (slides)
- "Machine Learning for Query Optimization." Boston University, database guest lecture series. April 18th, 2023. (slides)
- "Kepler: Robust Learning for Faster Parametric Query Optimization." Invited speaker for BTW '23. March 7th, 2023. (slides)

- "Open Source Learned Query Optimization." MIT / Intel / Meta Optimizer Symposium. May 22nd, 2023. ([slides](#))
- "Machine Learning for Query Optimization… and beyond!" Google YouTube Seminar, May 24th, 2023. ([slides](#))
- "Learned Query Optimization in PrestoDB." PrestoCon, Presto track sessions, December 6th, 2023. ([slides](#)) ([video](#))
- "Learned Query Superoptimization: PQO and Beyond." TUDa-HPI Database Seminar. December 13th, 2023. ([slides](#))