

Research Interests Systems, Cloud-Native Databases, Instance-Optimized Data Systems

Education **Massachusetts Institute of Technology** 2020 – Present
Doctor of Philosophy Student in Computer Science (PhD) Cambridge, MA

- *Advisor:* [Tim Kraska](#), Data Systems Group

University of Toronto 2018 – 2020
Master of Science in Computer Science (MSc) Toronto, ON

- *Advisor:* [Gennady Pekhimenko](#), EcoSystem Group (Computer Systems and Networks)
- *Thesis:* Habitat: Prediction-guided Hardware Selection for Deep Neural Network Training

University of Waterloo 2013 – 2018
Bachelor of Software Engineering (BSE) Waterloo, ON

- 94.65% Cumulative Average, Graduated with Distinction and on the Dean's Honors List

Publications **Check Out the Big Brain on BRAD: Simplifying Cloud Data Processing with Learned Automated Data Meshes**

Tim Kraska*, Tianyu Li*, Samuel Madden*, Markos Markakis*, Amadou Ngom*, Ziniu Wu*, [Geoffrey X. Yu](#)*
Proceedings of the VLDB Endowment (VLDB), Vol. 16, No. 11, 2023. Vision Paper

TreeLine: An Update-In-Place Key-Value Store for Modern Storage

[Geoffrey X. Yu](#)*, Markos Markakis*, Andreas Kipf*, Per-Åke Larson, Umar Farooq Minhas, Tim Kraska
Proceedings of the VLDB Endowment (VLDB), Vol. 16, No. 1, 2022

Habitat: A Runtime-Based Computational Performance Predictor for Deep Neural Network Training

[Geoffrey X. Yu](#), Yubo Gao, Pavel Golikov, Gennady Pekhimenko
USENIX Annual Technical Conference (USENIX ATC), 2021

Skyline: Interactive In-Editor Computational Performance Profiling for Deep Neural Network Training

[Geoffrey X. Yu](#), Tovi Grossman, Gennady Pekhimenko
ACM Symposium on User Interface Software and Technology (UIST), 2020

* Denotes equal contribution.

Demonstrations **Skyline: Interactive In-Editor Performance Visualizations and Debugging for DNN Training**

[Geoffrey X. Yu](#), Tovi Grossman, Gennady Pekhimenko
Conference on Machine Learning and Systems (MLSys), 2020

TBD Suite: Benchmarking and Profiling Tools for DNNs

[Geoffrey X. Yu](#), Hongyu Zhu, Anand Jayarajan, Bojian Zheng, Abhishek Tiwari, Gennady Pekhimenko
Conference on Machine Learning and Systems (MLSys), 2019

Teaching Experience **Teaching Assistant** Massachusetts Institute of Technology
6.5830 / 6.5831 – Database Systems Fall 2022

- *Instructors:* [Samuel Madden](#) and [Michael Cafarella](#)

Teaching Assistant University of Toronto
CSC 209 – Software Tools and Systems Programming Fall 2018, Winter 2019

**Professional
Experience**

Intel Corporation
North America – Remote

Research Graduate Intern
May – August 2021

- *Mentors:* Nesime Tatbul and David Cohen
- Led research on developing new designs for concurrent learned indexes that support database transactions.

Facebook
Menlo Park, CA

Software Engineering Intern
September – December 2017

- Utilized C++ to develop a new queuing service for the Async Tier—Facebook’s asynchronous job execution platform, responsible for handling tens of millions of jobs per minute.
- Designed and implemented a mechanism to reschedule failed jobs, requiring development across new and legacy infrastructure services on the Async Tier.
- Improved operational efficiency by implementing scripts to facilitate database shard management.

Facebook
Menlo Park, CA

Software Engineering Intern
January – April 2017

- Implemented robust Java code diagnostics for [Nuclide](#), enabling Java editing support for massive projects with over 60,000 source files.
- Engineered and developed Nuclide’s Java type resolution mechanism, which was capable of handling source code files with more than 10,000 transitive dependencies.
- Added partial rendering to the Nuclide console, increasing the number of entries that could be retained in the user interface by 20 times.

Apple
Cupertino, CA

Software Engineering Intern
May – August 2016

- [Contributed to Apache Cassandra](#) to improve disaster recovery across data centers.
- Prototyped a Cassandra partition blocking mechanism in Java to support nodes in production.
- Improved Cassandra timeout enforcement and introduced custom per-query timeouts
- Enhanced Cassandra error handling to enable error propagation from replica nodes to clients.

Coursera
Mountain View, CA

Software Engineering Intern
September – December 2015

- Developed a near real-time course tracking tool by leveraging Scala, Kafka, and JavaScript with React—providing an aggregate view of over 1700 courses on Coursera.
- Wrote a scheduled job in Scala to validate user role invariants across multiple data stores.
- Utilized Scala and JavaScript with React to build a suite of administrative course editing tools end-to-end, replacing legacy tools that lacked comprehensive validations.

Coursera
Mountain View, CA

Software Engineering Intern
January – April 2015

- Engineered robust course authoring tools using React and Flux that were made available to over 100 Coursera partner institutions.
- Built the front end for an instructor to learner mass email communication tool. This tool has been used to send out over 800,000 emails.

Milq (now Layer 6, acquired by TD Bank Group)
Toronto, ON

Full Stack Software Engineering Intern
May – August 2014

- Developed the tag page—an aggregation of content about a specific hashtag.
- Engineered scalable data API endpoints in Node.js using Redis and MongoDB to deliver matched content to the tag page.
- Developed functionality that fetched, resized, and stored video thumbnails into an Amazon S3 bucket for new and existing content added to Milq.

Awards and Honors

- **NSERC Postgraduate Scholarship – Doctoral (PGS D)** 2020 – 2023
Merit-based doctoral research scholarship awarded by the Government of Canada (63,000 CAD).
- **NSERC Alexander Graham Bell Canada Graduate Scholarship (CGS D)** 2020 – 2023 (declined)
Merit-based doctoral research scholarship awarded by the Government of Canada (105,000 CAD).
- **NSERC Canada Graduate Scholarship – Master’s (CGS M)** 2019 – 2020
Merit-based master’s research scholarship awarded by the Government of Canada (17,500 CAD).
- **Snap Research Scholarship** January 2019
Awarded by Snap Inc. for research quality, technical skills, and communication skills (10,000 USD).
- **Vector Institute Scholarship in Artificial Intelligence** 2018 – 2019
Awarded to top students pursuing AI-related master’s studies in Ontario (17,500 CAD).
- **Queen Elizabeth II Graduate Scholarship in Science and Technology** 2018 – 2019
Merit-based scholarship awarded to students pursuing graduate studies in Ontario (15,000 CAD).
- **Wolfond Fellowship, University of Toronto** 2018 – 2019
Awarded at the University of Toronto for academic achievements (10,000 CAD).
- **Sandford Fleming Foundation Award for Academic Excellence** June 2018
Best graduating academic record in the Software Engineering program at the University of Waterloo.
- **University of Waterloo Faculty of Engineering First in Class Award** May 2017
Highest term average in the Software Engineering program at the University of Waterloo (Fall 2016).
- **Wish Scholarship, University of Waterloo** January 2016
Awarded by the University of Waterloo based on academic merit.
- **University of Waterloo Faculty of Engineering Upper Year Scholarship** May 2015
Awarded by the Faculty of Engineering at the University of Waterloo based on academic merit.
- **Software Engineering Entrance Scholarship, University of Waterloo** September 2013
Awarded to first year Software Engineering students based on academic and extracurricular achievements.
- **University of Waterloo President’s Scholarship of Distinction** September 2013
Awarded to undergraduate students with an admissions average of 95% or greater.
- **Toronto District School Board Top Scholar** June 2013
Highest high school graduating average in the Toronto District School Board (Toronto, ON).

Talks

- **Check Out the Big Brain on BRAD: Simplifying Cloud Data Processing with Learned Automated Data Meshes** September 15, 2023
Cockroach Labs Virtual
- **TreeLine: An Update-In-Place Key-Value Store for Modern Storage** August 31, 2023
VLDB 2023 Vancouver, BC
- **Check Out the Big Brain on BRAD: Simplifying Cloud Data Processing with Learned Automated Data Meshes** August 29, 2023
VLDB 2023 Vancouver, BC
- **Learning-Based Creation of Data Mesh Architectures** March 10, 2023
North East Database Day 2023 Boston, MA
- **Habitat: A Runtime-Based Computational Performance Predictor for Deep Neural Network Training** July 15, 2021
USENIX ATC 2021 Virtual
- **Skyline: Interactive In-Editor Computational Performance Profiling for Deep Neural Network Training** October 21, 2020
UIST 2020 Virtual