

Michael Abebe

mtabebe@uwaterloo.ca

<https://cs.uwaterloo.ca/~mtabebe/>

Waterloo, Ontario, Canada

(226)-791-1039

RESEARCH EXPERIENCE

University of Waterloo – Graduate Research Assistant

Waterloo, Ontario

May 2016 - Present

My research interests lie in the area of distributed data management. I design and build large-scale data management systems that adapt based on workload trends to execute more efficiently.

- I built a distributed database system that adaptively adjusts its physical design via changes to data partitioning, replication and master location to efficiently execute transactions (MorphoSys, VLDB 2020)
- I designed and implemented a distributed database system that adaptively changes data mastership to eliminate the overheads of distributed commits (DynaMast, ICDE 2020)
- I developed an erasure coded storage system that reduces data access latency using dynamic data placement and access strategies (EC-Store, ICDCS 2018)
- I worked on caching systems that predictively execute and cache database query results (ChronoCache, SIGMOD 2020, Apollo, EDBT 2018)

EDUCATION

University of Waterloo

Ph.D. Student, Computer Science

May 2016 - Present

- Advisor: Khuzaima Daudjee
- Research topic: Adaptive distributed data management systems

University of Waterloo

Honours Bachelor of Computer Science, with Distinction, Dean's Honours List
Bioinformatics Option, Co-operative Program

May 2016

SCHOLARLY CONTRIBUTIONS

- **Abebe, M.**, Glasbergen, B., Daudjee, K. (2020) MorphoSys: Automatic Physical Design Metamorphosis for Distributed Database Systems (*PVLDB*)
- Glasbergen, B., **Abebe, M.**, Daudjee, K., Levi, A. (2020) Sentinel: Universal Analysis and Insight for Data Systems (*PVLDB*)
- **Abebe, M.**, Glasbergen, B., Daudjee, K. (2020) DynaMast: Adaptive Dynamic Mastering for Replicated Systems (*ICDE*)
- Glasbergen, B., Langendoen, K., **Abebe, M.**, Daudjee, K. (2020) ChronoCache: Predictive and Adaptive Mid-Tier Query Result Caching (*SIGMOD*)
- Glasbergen, B., **Abebe, M.**, Daudjee, K., Vogel, D., Zhao, J. (2020) Demo: Sentinel: Understanding Data Systems (*SIGMOD Demonstration Track*)
- Glasbergen, B., **Abebe, M.**, Daudjee, K., Foggo, S., Pacaci, A. (2018) Apollo: Learning Query Correlations for Predictive Caching in Geo-Distributed Systems (*SIGMOD*)
- **Abebe, M.**, Glasbergen, B., Daudjee, K. (2019) WatDFS: A Project for Understanding Distributed Systems in the Undergraduate Curriculum (*SIGCSE*)
- **Abebe, M.**, Daudjee, K., Glasbergen, B., Tian, F. (2018) EC-Store: Bridging the Gap Between Storage and Latency in Distributed Erasure Coded Systems (*ICDCS*)
- Glasbergen, B., **Abebe, M.**, Daudjee, K. (2018) Tutorial: Adaptive Replication and Partitioning in Data Systems (*Middleware*)
- Glasbergen, B., **Abebe, M.**, Daudjee, K., Foggo, S., Pacaci, A. (2018) Apollo: Learning Query Correlations for Predictive Caching in Geo-Distributed Systems (*EDBT*)

- Wu, L., Gingery, M., **Abebe, M.**, Arambula, D., ..., & Zimmerly, S. (2017) Diversity-Generating Retroelements: Natural Variation, Classification and Evolution Inferred from a Large-Scale Genomic Survey (*Nucleic Acids Research*)
- **Abebe, M.**, Candales, M. Duong, A., ... & Semper, C. (2013). A pipeline of programs for collecting and analyzing group II intron retroelement sequences from GenBank. (*Mobile DNA*)

WORK EXPERIENCE

Facebook - Software Engineer Intern

Seattle, Washington

August - December 2015

Engineer on the next-generation file system for the data warehouse.

- I designed, built, and deployed a distributed log search system for the data warehouse fleet. This system provides sub-second responses to queries across large clusters. To improve debugging, I customized a logging framework within the storage layer to allow developers to trace request flow between services and servers.
- Within the storage system, I decreased the memory consumed during inter-process communication by 2/3 by developing a new RPC layer for large data buffers.
- I improved the operational workflow of the team by building new dashboards to track SLA violations and irregularities in service performance.

Facebook - Software Engineer Intern

Seattle, Washington

January - April 2015

Engineer on Core Systems, which provides the infrastructure to manage services across the entire fleet of servers.

- I prototyped a new approach to service discovery at Facebook: Service Discovery as a Service, which decreased resource use compared to the original approach by an order of magnitude, and increased reliability.
- To provide ultra-low latency responses, I made changes to the Thrift compiler (RPC framework) to transparently cache incoming requests and outgoing responses.
- I improved the operations of services that were owned by the team by building emergency recovery tools and adding new metrics for monitoring services.

Palantir - Software Engineer Intern

Palo Alto, California

May - August 2014

Engineer on Palantir Gotham, which provides data integration and transformation to enable expert analysis

- I prototyped a web application that allowed collaborative investigation within Palantir Gotham, including sharing the state of the investigation as a primitive data type in a messaging application

University of Calgary - Bioinformatics Research Assistant

Calgary, Alberta

May - August 2012, January-April 2013

Automated the discovery and analysis of genetic elements using bioinformatics tools.

- I built data collection pipelines to discover, annotate and analyze new classes of mobile DNA elements.

SERVICE

- External Reviewer: VLDB (2017-2021), ICDE (2017-2020), SIGMOD (2018-2021), Middleware (2017)
- Graduate Representative: Undergraduate Academic Plans Committee – David R. Cheriton School of Computer Science (2017-2020)
 - This committee oversees and maintains the undergraduate curriculum for computer science at the University of Waterloo
- Vice President Finance and Member-at-Large: University of Waterloo Bioinformatics Club (2011-2016)

- In this role, I helped secure the funds to purchase a new server and desktop computer for the club. These computing resources helped members learn more about the tools and software used in bioinformatics.

TEACHING

- Teaching Assistant: CS 454 (Distributed Systems)
 - Guest lectures on erasure coded storage systems, cloud storage systems and MapReduce.
 - Developed new distributed file system assignments (WatDFS, published in SIGCSE 2019).
- Supervision of Undergraduate Research Assistants:
 - Dmytro Shynkevych (2021) – Query arrival prediction in database systems
 - Brad Huang (2020) – Machine learning models for cost estimates in database systems
 - Alya Berciu (2019) – Evaluation of data placement on recovery in erasure-coded storage systems.
 - Zheng Tao (2018) – Holistic data placement upon recovery in erasure-coded storage systems.
 - Andrew Lau (2018) – Evaluation of remastering strategies in distributed databases.
 - Felix Tian (2017) – Fault detection in erasure-coded storage systems.
 - Scott Foggo (2017) – Scalable distributed cache management.
 - Youngbin Kim (2016) – Session management in distributed databases.

AWARDS

- NSERC Alexander Graham Bell Canada Graduate Scholarship (CGS-D) (2018-2021, \$35,000 / year)
- ACM SIGMOD Best Demonstration Award (*Sentinel: Understanding Data Systems*) (2020)
- University of Waterloo Computer Science Teaching Assistant Award (2019)
- Facebook Fellowship - Emerging Scholar (CGS-D) (2018-2020, \$42,000 / year)
- NSERC Alexander Graham Bell Canada Graduate Scholarship (2017, \$17,500)
- Ontario Graduate Scholarship (2016, \$15,000)
- David R. Cheriton Graduate Scholarship (2016-2017, \$10,000 / year)
- University of Waterloo President's Graduate Scholarship (2016-2017, \$5000 / year)
- University of Waterloo Graduate Entrance Scholarship (2016, \$5000)
- University of Waterloo Term Dean's Honours List (2011 - 2016)
- University of Waterloo President's International Experience Award (2013, \$1500)
- Duke of Edinburgh Gold Award (2010)