Barzan Mozafari

Computer Science & Engineering
2260 Hayward St., Room 4769
Ann Arbor, MI 48109, USA

Phone: (734) 763-3669
E-mail: mozafari@umich.edu
Web: `http://web.eecs.umich.edu/~mozafari/`
LinkedIn: linkedin.com/in/barzan-mozafari

Affiliation

University of Michigan, Ann Arbor, MI, USA

Associate Professor, Computer Science and Engineering

Keebo, Ann Arbor, MI, USA

Founder and CEO, 2020–2026

Education

Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Postdoctoral Associate, CSAIL, 2011–2013

University of California Los Angeles (UCLA), Los Angeles, CA, USA

Ph.D., Computer Science, 2006–2011
M.S., Computer Science, 2006–2008

Shahid Beheshti University, Tehran, Iran

B.S., Computer Engineering, 2002–2006

Research Interests

Database Systems, Machine Learning, Artificial Intelligence

Awards and Honors

- Morris Wellman Faculty Development Award, 2017.
- National Science Foundation CAREER Award, 2016.
- ACM SIGMOD Best Demo Award, 2014.
- EuroSys Best Paper Award, 2013.
- ACM SIGMOD Best Paper Award, 2012.
- ACM SIGMOD Honorable Mention Award, 2010.
- UC San Diego Data Mining Competition for Grads/Postdocs, 3^{rd} place, 2008.
- Graduate Division Fellowship, 2007–2008.
- Graduate Division Fellowship, 2006–2007.
- National Matriculation Examination for graduate studies in computer science, 2^{nd} place in 2005, and 9^{th} place in 2004.
- National Matriculation Examination for graduate level in computer engineering, 10^{th} place in Iran, 2005.
- ACM/ICPC Contest, Asia/Tehran, 9^{th} place in 2004, 11^{th} place in 2003.
- World Robocup Rescue Simulation League, Lisbon, Portugal, 5^{th} place, 2003.
- Silver medal in National Informatics Olympiad, 2001.

Research Impact

Cloud-warehouse optimization commercialized through Keebo, *2023*. Keebo developed the first autonomous cloud data-warehouse optimization platform and was used by Fortune 100 companies.

BlinkDB ideas commercialized through SnappyData and TIBCO, *2019*. BlinkDB was commercialized through SnappyData, which was later acquired by TIBCO.

BlinkDB and VerdictDB brought interactive approximate analytics to open source, *2018*. BlinkDB and VerdictDB made research on bounded, interactive approximate analytics available through public software, project artifacts, and reproducible system descriptions.

Transaction schedulers adopted by MySQL and MariaDB, *2018*. Our CATS algorithm has become the default transaction scheduler in MySQL and Percona Server for MySQL, estimated to be running on hundreds of millions of deployments worldwide. Our earlier VATS algorithm was also adopted as the default transaction scheduler in older versions of MariaDB.

Grants

Total funding raised as either **PI** or **Co-PI: $6.6M** ($6M research; $600K educational)

Verdict: Cross-Platform Analytics at a Fraction of the Cost. *Sponsor:* Alibaba, *Role:* PI, *Period:* 2018.

Approximate Cluster Computing. *Sponsor:* Google, *Role:* PI, *Period:* 2017.

Commercializing YowP: Your Own Planner. *Sponsor:* Michigan Translational Research and Commercialization (MTRAC), *Role:* PI, *Period:* 2017–2018.

Bridging the Gap between Academia and Industry: Workshop on Approximate Computing. *Sponsor:* NSF, *Role:* PI, *Period:* 2017.

YowP: Your Own Planner. *Sponsor:* Michigan Translational Research and Commercialization (MTRAC), *Role:* PI, *Period:* 2016–2017.

XPS: FULL: A Cross-Layer Approach Toward Low-Latency Data-Parallel Applications in Rack-Scale Computing. *Sponsor:* NSF, *Role:* Co-PI, *Period:* 2016–2019.

CAREER: Designing A Predictable Database: An Overlooked Virtue. *Sponsor:* NSF, *Role:* PI, *Period:* 2016–2021.

I-Corps: Pain-free Database Administration via Workload Intelligence. *Sponsor:* NSF, *Role:* PI, *Period:* 2016.

MRI: Acquisition of Conflux, A Novel Platform for Data-Driven Computational Physics. *Sponsor:* NSF, *Role:* Co-PI, *Period:* 2015–2018.

CPS: GOALI: Synergy: Maneuver and Data Optimization for High Confidence Testing of Future Automotive Cyberphysical Systems. *Sponsor:* NSF, *Role:* Co-PI, *Period:* 2015–2018.

Transforming Analytical Learning in the Era of Big Data. *Sponsor:* NIH, *Role:* Co-PI, *Period:* 2015–2018.

Educational Grant. *Sponsor:* Microsoft, *Role:* PI, *Period:* 2015–2016.

An Open-Platform Lab Sandbox for Big Data Education. *Sponsor:* Intel, *Role:* PI, *Period:* 2015.

Achieving Robustness in Cloud-based Distributed Databases. *Sponsor:* Microsoft Azure, *Role:* PI, *Period:* 2014–2015.

Robust Physical Designers. *Sponsor:* Amazon AWS, *Role:* PI, *Period:* 2014.

Educational Grant. *Sponsor:* Amazon, *Role:* PI, *Period:* 2013–2014.

Selected Publications

Jie Liu and **Barzan Mozafari**, "GenRewrite: Query Rewriting via Large Language Models", In Proceedings of **SIGMOD**, 2026.

Rui Dong, Jie Liu, Yuxuan Zhu, Cong Yan, **Barzan Mozafari**, and Xinyu Wang, "SlabCity: Whole-Query Optimization using Program Synthesis", In Proceedings of **VLDB**, 2023.

Barzan Mozafari, Jie Liu, Yongjoo Park, Morteza Tolee, Wenbo Wu, and Chi Zhang, "Making Data Clouds Smarter at Keebo: Automated Warehouse Optimization using Data Learning", In Proceedings of **SIGMOD**, 2023.

Junghwan Kim and **Barzan Mozafari**, "Provable Memorization Capacity of Transformers", In Proceedings of **ICLR**, 2023.

Rui Liu and **Barzan Mozafari**, "Communication-efficient Distributed Learning for Large Batch Optimization", In Proceedings of **ICML**, 2022.

Rui Liu and **Barzan Mozafari**, "Transformer with Memory Replay", In Proceedings of **AAAI**, 2022.

Rui Liu, Tianyi Wu, and **Barzan Mozafari**, "Adam with Bandit Sampling for Deep Learning", In Proceedings of **NeurIPS**, 2020.

Yongjoo Park, Jingyi Qing, Xiaoyang Shen, and **Barzan Mozafari**, "BlinkML: Efficient Maximum Likelihood Estimation with Probabilistic Guarantees", In Proceedings of **SIGMOD**, 2019.

Tanvir Ahmed Khan, Yifan Zhao, Gilles Pokam, **Barzan Mozafari**, and Baris Kasikci, "Huron: Hybrid False Sharing Detection and Repair", In Proceedings of **PLDI**, 2019.

Jarrid Rector-Brooks, Jun-Kun Wang, and **Barzan Mozafari**, "Revisiting Projection-Free Optimization for Strongly Convex Constraint Sets", In Proceedings of **AAAI**, 2019.

Rui Liu, Tianyi Wu, and **Barzan Mozafari**, "A Bandit Approach to Maximum Inner Product Search", In Proceedings of **AAAI**, 2019.

Yongjoo Park, **Barzan Mozafari**, Joseph Sorenson, and Junhao Wang, "VerdictDB: Universalizing Approximate Query Processing", In Proceedings of **SIGMOD**, 2018.

Dong Young Yoon, Mosharaf Chowdhury, and **Barzan Mozafari**, "Distributed Lock Management with RDMA: Decentralization without Starvation", In Proceedings of **SIGMOD**, 2018.

Boyu Tian, Jiamin Huang, **Barzan Mozafari**, Grant Schoenebeck, and Thomas F. Wenisch, "Contention-aware lock scheduling for transactional databases", In Proceedings of **VLDB**, 2018.

Kevin Eykholt, Atul Prakash, and **Barzan Mozafari**, "Ensuring Authorized Updates

in Multi-user Database-Backed Applications", In Proceedings of the **USENIX Security Symposium**, 2017.

Yongjoo Park, Ahmad Shahab Tajik, Michael Cafarella, and **Barzan Mozafari**, "Database Learning: Toward a Database that Becomes Smarter Every Time", In Proceedings of **SIGMOD**, 2017.

Jiamin Huang, **Barzan Mozafari**, Grant Schoenebeck, and Thomas F. Wenisch, "A Top-Down Approach to Achieving Performance Predictability in Database Systems", In Proceedings of **SIGMOD**, 2017.

Jiamin Huang, **Barzan Mozafari**, and Thomas F. Wenisch, "Statistical Analysis of Latency Through Semantic Profiling", In Proceedings of **EuroSys**, 2017.

Dong Young Yoon, Ning Niu, and **Barzan Mozafari**, "DBSherlock: A Performance Diagnostic Tool for Transactional Databases", In Proceedings of **SIGMOD**, 2016.

Yongjoo Park, Michael Cafarella, and **Barzan Mozafari**, "Neighbor-Sensitive Hashing", In Proceedings of **VLDB**, 2015.

Barzan Mozafari, Eugene Zhen Ye Goh, and Dong Young Yoon, "CliffGuard: A Principled Framework for Finding Robust Database Designs", In Proceedings of **SIGMOD**, 2015.

Barzan Mozafari, Purnamrita Sarkar, Michael Franklin, Michael Jordan, and Samuel Madden, "Scaling Up Crowd-Sourcing to Very Large Datasets: A Case for Active Learning", In Proceedings of **VLDB**, 2014.

Kai Zeng, Shi Gao, **Barzan Mozafari**, and Carlo Zaniolo, "The Analytical Bootstrap: A New Method for Fast Error Estimation in Approximate Query Processing", In Proceedings of **SIGMOD**, 2014.

Sameer Agarwal, Henry Milner, Ariel Kleiner, Ameet Talwalkar, Michael Jordan, Samuel Madden, **Barzan Mozafari**, and Ion Stoica, "Knowing When You're Wrong: Building Fast and Reliable Approximate Query Processing Systems", In Proceedings of **SIGMOD**, 2014.

Barzan Mozafari, Carlo Curino, Alekh Jindal, and Samuel Madden, "Performance and Resource Modeling in Highly-Concurrent OLTP Workloads", In Proceedings of **SIGMOD**, 2013.

Sameer Agarwal, **Barzan Mozafari**, Aurojit Panda, Henry Milner, Samuel Madden, and Ion Stoica, "BlinkDB: Queries with Bounded Errors and Bounded Response Times on Very Large Data", In Proceedings of **EuroSys**, 2013. [**Best Paper Award**]

Barzan Mozafari, Carlo Curino, and Samuel Madden, "DBSeer: Resource and Performance Prediction for Building a Next Generation Database Cloud", In Proceedings of **CIDR**, 2013.

Barzan Mozafari, Kai Zeng, Loris D'Antoni, and Carlo Zaniolo, "High-performance Complex Event Processing Over Hierarchical Data", In **ACM TODS**, 2013.

Barzan Mozafari, Kai Zeng, and Carlo Zaniolo, "High-Performance Complex Event Processing over XML Streams", In Proceedings of **SIGMOD**, 2012. [**Best Paper Award**]

Hetal Thakkar, Nikolay Laptev, Hamid Mousavi, **Barzan Mozafari**, Vincenzo Russo, and Carlo Zaniolo, "SMM: A Data Stream Management System for Knowledge Discovery", In

Proceedings of **ICDE**, 2011.

Barzan Mozafari, Kai Zeng, and Carlo Zaniolo, "From Regular Expressions to Nested Words: Unifying Languages and Query Execution for Relational and XML Sequences", In Proceedings of **VLDB**, 2010.

Barzan Mozafari and Carlo Zaniolo, "Optimal Load Shedding with Aggregates and Mining Queries", In Proceedings of **ICDE**, 2010.

Barzan Mozafari and Carlo Zaniolo, "Publishing Naive Bayesian Classifiers: Privacy without Accuracy Loss", In Proceedings of **VLDB**, 2009.

Barzan Mozafari, Hetal Thakkar, and Carlo Zaniolo, "Verifying and Mining Frequent Patterns from Large Windows over Data Streams", In Proceedings of **ICDE**, 2008.

Rodrigo B. Almeida, **Barzan Mozafari**, and Junghoo Cho, "On the Evolution of Wikipedia", In Proceedings of **ICWSM**, 2007.

Other Publications

Wen He, Yongjoo Park, Idris Hanafi, Jacob Yatvitskiy, **Barzan Mozafari**, "Demonstration of VerdictDB, the Platform-Independent AQP System", In proceedings of **SIGMOD**, 2018. [Demo]

Dong Young Yoon, **Barzan Mozafari**, Douglas P. Brown, "DBSeer: Pain-free Database Administration through Workload Intelligence", In proceedings of **PVLDB**, 2015. [Demo]

Barzan Mozafari, "Verdict: A System for Stochastic Query Planning", In proceedings of **CIDR**, 2015.

Kai Zeng, Shi Gao, Jiaqi Gu, **Barzan Mozafari**, and Carlo Zaniolo, "ABS: a System for Scalable Approximate Queries with Accuracy Guarantees", In proceedings of **SIGMOD**, 2014. [**Best Demo Award**]

Kai Zeng, Mohan Yang, **Barzan Mozafari**, Carlo Zaniolo, "Complex Pattern Matching in Complex Structures: the XSeq Approach", In proceedings of **ICDE**, 2013. [Demo]

Sameer Agarwal, Aurojit Panda, **Barzan Mozafari**, Anand P. Iyer, Samuel Madden, Ion Stoica, "Blink and It's Done: Interactive Queries on Very Large Data", In proceedings of **PVLDB**, 5(12): 1902-1905, 2012. [Demo]

Barzan Mozafari, Kai Zeng, Carlo Zaniolo, "K*SQL: a Unifying Engine for Sequence Patterns and XML", In proceedings of **SIGMOD**, 2010. [**Honorable Mention** Demo]

Hetal Thakkar, **Barzan Mozafari** and Carlo Zaniolo, "Continuous Post-Mining of Association Rules in a Data Stream Management System", In Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, edited by Yanchang Zhao, Chengqi Zhang, and Longbing Cao, Information Science Reference, May, 2009.

Hetal Thakkar, **Barzan Mozafari** and Carlo Zaniolo "A Data Stream Mining System", In proceedings of the International Conference on Data Mining (**ICDM**), 2008.

Hetal Thakkar, **Barzan Mozafari** and Carlo Zaniolo, "Designing an Inductive DSMS: the Stream Mill Experience", In proceedings of the International Workshop on Scalable Stream Processing Systems (**SSPS**), 2008.

Teaching Experience

Introduction to Database Management Systems, University of Michigan Ann Arbor

Advanced Database Management Systems, University of Michigan Ann Arbor

Current Topics in Databases, University of Michigan Ann Arbor

Entrepreneurial Experience

Keebo, Founder and CEO, 2020–2026, Ann Arbor, MI, USA
Database intelligence company focused on automated cloud data warehouse optimization.

SnappyData Inc., Architect and Strategic Advisor, Pune, India
The first commercial approximate query processing engine.

Professional Activities

Co-chair, Distributed Infra., Systems, Programming, & AI (DISPA) Workshop, 2020.
Chair, Approximate Computing for Affordable and Interactive Analytics (ACAIA), 2017.
Co-chair, SIGMOD Workshops, 2016.
Co-chair, CloudDM, 2016.
PC, VLDB, 2015, 2016, 2017, 2018, 2019.
PC, SIGMOD, 2014, 2016, 2017, 2018.
PC, ICDE, 2013.
PC, SoCC, 2015.
PC, Scientific & Statistical Database Management Conf., 2013
PC, CIKM, 2013.