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/`

Affiliation

University of Michigan, Ann Arbor, MI, USA

Assistant Professor (July 2013-now), Computer Science and Engineering Division

Education

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

Postdoctoral Associate (July 2011-June 2013), CSAIL, Mentor: Samuel Madden

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

Ph.D. (2006-2011), Computer Science, Advisor: Carlo Zaniolo

- Thesis: "High-Performance Pattern Discovery and Detection for Databases and Data Streams"

M.Sc. (2006-2008), Computer Science, GPA: 4.0, Advisor: Carlo Zaniolo

- Thesis: "Mining Frequent Patterns from Large Windows over Data Streams"

Shahid Beheshti University, Tehran, Iran

B.S., Computer Engineering (2002-2006)

National Organization for Development of Exceptional Talents, Sanandaj, Iran

Diploma, Mathematics and Physics (1998-2002)

Research Interests

Databases, Large-Scale Data-Intensive Systems, Database-as-a-Service, Crowdsourcing, Data Mining and Applied Machine Learning

Awards and Honors

- ACM SIGMOD's Best Demo Award for "ABS: a System for Scalable Approximate Queries with Accuracy Guarantees", 2014.
- Eurosys 2013's "Best Paper Award" for "BlinkDB: Queries with Bounded Errors and Bounded Response Times on Very Large Data", 2013.
- ACM SIGMOD 2012's "Best Paper Award" for "High-Performance Complex Event Processing over XML Streams", 2012.
- ACM SIGMOD's Honorable Mention for demo paper "K*SQL: A Unifying Engine for Sequence Patterns and XML", 2010.
- UC San Diego Data Mining Competition for Grads/Postdocs, 3rd place, 2008.
- Graduate Division Fellowship, 2007-2008.
- Graduate Division Fellowship, 2006-2007.
- National Matriculation Examination for graduate level in computer science, 2nd place in 2005, and 9th place in 2004 (among more than 4000 examinees in Iran).
- National Matriculation Examination for graduate level in computer engineering, 10th place in Iran, 2005.
- ACM/ICPC Programming Contest, Asia/Tehran site, 9th place in 2004, $11th$ place in 2003.
- World Robocup Rescue Simulation League, Lisbon, Portugal, 5th in 2003.
- Silver medal in National Informatics Olympiad (a nationwide competition in programming and computer science theory for selecting the national team for International Olympiad in Informatics), 2001.

Research Impact

Creator, Variance-Aware Transaction Scheduling, 2014-2016. In a project aimed at improving the predictability of database systems, we created the first profiler capable of analyzing a complex DBMS codebase and identifying its main source of latency variance. Through our findings, we designed a new scheduling algorithm, called *Variance-Aware Transaction Scheduling (VATS)*, which was quickly adopted by the MySQL community (as of MySQL 5.7.17) and even chosen as the default algorithm in MariaDB (see `https://github.com/MariaDB/server/pull/248`), a major distributor of MySQL with over 2M+ installations around the world.

Creator and Developer, DBSeer Open Source Project, 2011-2015. I started the DBSeer project in 2011, as a software suite that uses the state-of-the-art machine learning and causality techniques to help database administrators monitor, diagnose, tune, analyze and predict the performance of their database system in an easier, more accurate and more effective manner. DBSeer is now an open-source project (see `http://dbseer.org/video`) with a growing user-base, currently under consideration by a major distributor of MySQL for Linux. DBSeer has been very well received in the community (see `http://web.eecs.umich.edu/~mozafari/press.html#dbseer`).

Co-Designer, BlinkDB Open Source Project, 2011-2015. I was one of the main designers of BlinkDB since its inception. BlinkDB is a large-scale distributed query processing system that uses approximation techniques to query petabytes of data within seconds. As an open-source project (`http://blinkdb.org`), BlinkDB received much attention in the Hadoop community, was adopted by Databricks (the company commercializing Apache Spark), influenced similar approximation features in many other engines (e.g., Facebook's Presto and Oracle's 12C release of approximate aggregates), and successfully commercialized by Pivotal and General Electric (called SnappyData).

Designer and Co-Developer, K*SQL and XSeq Systems, 2009-2012. I started the X*SQL and XSeq projects, as the main designer and developer. K*SQL and XSeq are the first complex event processing systems to support hierarchical data, e.g., genomic sequences, software traces, and JSON/XML event streams. These theoretical and practical contributions enabled new applications, leading to an honorable award in SIGMOD 2010 and a best paper award in SIGMOD 2012.

Grants

CAREER: Designing A Predictable Database: An Overlooked Virtue, $500K. *Sponsor:* NSF, *PI:* Mozafari, *Period:* 2016–2021.

XPS: FULL: A Cross-Layer Approach Toward Low-Latency Data-Parallel Applications in Rack-Scale Computing, $825K. *Sponsor:* NSF, *PI:* Chowdhury, *co-PI:* Mozafari, *Period:* 2016–2019.

I-Corps: Pain-free Database Administration via Workload Intelligence, $50K. *Sponsor:* NSF, *PI:* Mozafari, *Period:* 2016.

MRI: Acquisition of Conflux, A Novel Platform for Data-Driven Computational Physics, $3.46M. *Sponsor:* NSF, *PI:* Duraisamy, *co-PI:* Mozafari, Evrard, Alberto, and Garikipati, *Period:* 2015–2018.

CPS: GOALI: Synergy: Maneuver and Data Optimization for High Confidence Testing of Future Automotive Cyberphysical Systems, $775K. *Sponsor:* NSF, *PI:* Kolmanovsky, *co-PI:* Mozafari, Atkins, and Oliver. *Period:* 2015–2018.

Transforming Analytical Learning in the Era of Big Data, $481K. *Sponsor:* NIH,

PI: Bhramar, *co-PI:* Johnson, Mozafari, and Nguyen. *Period:* 2015–2018.

Educational Grant, $23.4K. *Sponsor:* Microsoft, *PI:* Mozafari, *Period:* 2015–2016.

An Open-Platform Lab Sandbox for Big Data Education, $39K. *Sponsor:* Intel, *PI:* Mozafari, 2015.

Achieving Robustness in Cloud-based Distributed Databases, $20K. *Sponsor:* Microsoft Azure, *PI:* Mozafari, 2014–2015.

Robust Physical Designers, $6K. *Sponsor:* Amazon AWS, *PI:* Mozafari, 2014.

Educational Grant, $3.5K. *Sponsor:* Amazon, *PI:* Mozafari, *Period:* 2013–2014.

Selected Publications

Barzan Mozafari, Jags Ramnarayan, Sudhir Menon, Yogesh Mahajan, Soubhik Chakraborty, Hemant Bhanawat, Kishor Bachhav, "SnappyData: A Unified Cluster for Streaming, Transactions and Interactice Analytics", In proceedings of **CIDR**, 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, and Ning Niu, "A Handbook for Building an Approximate Query Engine", A Book Chapter in Approximate Query Processing and Applications, IEEE Special Issue, 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, Purna Sarkar, Michael Franklin, Michael Jordan, Samuel Madden, "Scaling a Crowd-Sourced Database", In proceedings of **VLDB**, 8(2), 2015.

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, 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, Ion Stoica, "BlinkDB: Queries with Bounded Errors and Bounded Response Times on Very Large Data", In **EuroSys**, 2013. [**Best Paper Award**] `http://blinkdb.org`

Barzan Mozafari, Carlo Curino, Samuel Madden, "DBSeer: Resource and Performance Prediction for Building a Next Generation Database Cloud", In proceedings of **CIDR**, 2013. `http://dbseer.org`

Barzan Mozafari, Kai Zeng, Carlo Zaniolo, "Complex Event Processing over Big Hierarchical Data", Invited to Special Issue of ACM Transactions on Database Systems (**TODS**) for SIGMOD 2012 Best Papers, 2013.

Barzan Mozafari, Kai Zeng, Carlo Zaniolo, "High-Performance Complex Event Processing over XML Streams", In proceedings of **SIGMOD**, 253-264, 2012. [**Best Paper Award**].

Hetal Thakkar, Nikolay Laptev, Hamid Mousavi, **Barzan Mozafari**, Vincenzo Russo, Carlo Zaniolo, "SMM: A Data Stream Management System for Knowledge Discovery", In proceedings of **ICDE**, 757-768, 2011.

Barzan Mozafari, Kai Zeng, Carlo Zaniolo, "From Regular Expressions to Nested Words: Unifying Languages and Query Execution for Relational and XML Sequences", In proceedings of **VLDB**, 3(1): 150-161, 2010.

Barzan Mozafari, Carlo Zaniolo, "Optimal Load Shedding with Aggregates and Mining Queries", In proceedings of **ICDE**, 76-88, 2010.

Barzan Mozafari, Carlo Zaniolo, "Publishing Naive Bayesian Classifiers: Privacy without Accuracy Loss", In proceedings of **VLDB**, 2(1): 1173-1185, 2009.

Barzan Mozafari, Hetal Thakkar, Carlo Zaniolo, "Verifying and Mining Frequent Patterns from Large Windows over Data Streams", In proceedings of **ICDE**, 179-188, 2008.

Rodrigo B. Almeida, **Barzan Mozafari**, Junghoo Cho, "On the Evolution of Wikipedia". In proceedings of the International Conf. on Weblogs and Social Media (**ICWSM**), 2007.

Other Publications

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

Industrial Experience

Adconion Media Group (largest private ad network worldwide), Santa Monica, CA, USA
Research Intern **Summer of 2009**
Developed an offline testbed to allow fast and easy evaluation of different machine learning models and parameters, in terms of their impact on conversion rate of the displayed ads, without having to deploy those models on live traffic.

- **Accomplishments:** A projected uplift of 11% for CPA (Cost Per Action) ads.

Microsoft Research, Redmond , Washington, USA
Research Intern **Summer of 2008**
Worked on optimal query plans for distributed aggregates in a Map-Reduce cluster.

- **Accomplishments:** A set of empirical heuristics for choosing the number of partial aggregation layers based on different parameters of the query and the cluster.

Yahoo Headquarters, Sunnyvale , California, USA
Data Mining Research Intern **3 months, 2007**
Developed real-time models to maximize the revenue through lifting conversion rate of Pay-Per-Click ads while satisfying Pay-Per-Impression commitments.

- **Accomplishments:** (joint with my mentor Shu-Yao Chien and 3 other employees in the Audience Realtime Targeting team) Lifting conversion rate of ads across Yahoo property by more than 150% while serving personalized headlines to +175 million Yahoo users, with a projected increase of the annual revenue by $5.5 M.

Membership

Board of Advisors, SnappyData Inc.

Professional Activities

Co-chair, SIGMOD Workshops, 2016.
Co-chair, CloudDM, 2016.
PC, VLDB, 2015, 2016, 2017.
PC, SIGMOD, 2014, 2016.
PC, ICDE, 2013.
PC, SoCC, 2015.
PC, Scientific & Statistical Database Management Conf., 2013
PC, CIKM, 2013.