

EDUCATION

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

Postdoctoral Associate (July 2011-now), Big Data at 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

- 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, 3<sup>rd</sup> place, 2008.
- Graduate Division Fellowship, 2007-2008.
- Graduate Division Fellowship, 2006-2007.
- National Matriculation Examination for graduate level in computer science, 2<sup>nd</sup> place in 2005, and 9<sup>th</sup> place in 2004 (among more than 4000 examinees in Iran).
- National Matriculation Examination for graduate level in computer engineering, 10<sup>th</sup> place in Iran, 2005.
- ACM/ICPC Programming Contest, Asia/Tehran site, 9<sup>th</sup> place in 2004, 11<sup>th</sup> place in 2003.
- World Robocup Rescue Simulation League, Lisbon, Portugal, 5<sup>th</sup> in 2003.
- Silver medal in National Informatics Olympiad (Remark: This is a nationwide contest in programming and computer science theory for selecting the national team for International Olympiad in Informatics), 2001.

RESEARCH  
GRANTS

- Veolia Environment’s Research Funding, awarded on 6/29/2012, for \$40K:  
– “Revisiting Big-Data Analytics: A General Framework for Efficient Decision Support over Uncertain Data”, PI: Barzan Mozafari.

## 1. Large-Scale Data-Intensive Systems, Cloud Computing:

- Complex Event Processing Systems (SIGMOD'10, '12, PVLDB'10, TODS'13, ICDE'13)
  - *Designing query languages and systems for high-performance complex event processing; Extending pattern matching algorithms to Nested-Word Automata*
- Database Clouds (Database-as-a-Service) (CIDR'13, SIGMOD'13)
  - *Performance modeling and prediction for building the next generation of database clouds; Providing elasticity, performance isolation and resource attribution*
- Large-Scale Distributed Systems (PVLDB'12, EuroSys'13)
  - *Designing a massively parallel, approximate query engine to support interactive analytics with bounded error on terabytes of data*
- Mining Data Streams (SSPS'08, ICDM'08, ICDE'08, ICDE'10, ICDE'11)
  - *Designing a data stream management system with mining functionalities*
  - *Single-pass algorithms for frequent pattern mining; Efficient data structures for containment counting*
  - *Optimal sampling by modeling uncertainty propagation in complex query networks*

## 2. Crowdsourcing, Social Web:

- Crowd-enabled Database Systems (under submission)
  - *Designing active learning algorithms for crowdsourcing very large datasets*
- User Modeling (ICWSM'07, PVLDB'09)
  - *Modeling collaborative content generation on Wikipedia*
  - *Publishing privacy-preserving Naïve Bayesian Classifiers with zero accuracy loss*

[P18] **Barzan Mozafari**, Carlo Curino, Alekh Jindal, Samuel Madden, “Performance and Resource Modeling in Highly-Concurrent OLTP Workloads”, To appear in **SIGMOD**, 2013.

[P17] 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>

[P16] **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>

[P15] **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, To appear in Dec 2013.

[P14] **Barzan Mozafari**, Kai Zeng, Carlo Zaniolo, “High-Performance Complex Event Processing over XML Streams”, In proceedings of **SIGMOD**, 253-264, 2012. [**Best Paper Award**]. Acceptance rate: 16.6%.

[P13] 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. Acceptance rate: 19.8%.

[P12] **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 **PVLDB**, 3(1): 150-161, 2010. Acceptance rate: 15.8%.

[P11] **Barzan Mozafari**, Carlo Zaniolo, “Optimal Load Shedding with Aggregates and Mining Queries”, In proceedings of **ICDE**, 76-88, 2010. Acceptance rate: 12.5%.

[P10] **Barzan Mozafari**, Carlo Zaniolo, “Publishing Naive Bayesian Classifiers: Privacy without Accuracy Loss”, In proceedings of **PVLDB**, 2(1): 1173-1185, 2009. Acceptance rate: 16.7%.

[P9] **Barzan Mozafari**, Hetal Thakkar, Carlo Zaniolo, “Verifying and Mining Frequent Patterns from Large Windows over Data Streams”, In proceedings of **ICDE**, 179-188, 2008. Acceptance rate: 19%.

[P8] 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

[P7] Kai Zeng, Mohan Yang, **Barzan Mozafari**, Carlo Zaniolo, “Complex Pattern Matching in Complex Structures: the XSeq Approach”, In proceedings of **ICDE**, 2013. [Demo]

[P6] 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]

[P5] Kai Zeng, **Barzan Mozafari**, Shi Gao, and Carlo Zaniolo, “Uncertainty Propagation in Complex Query Networks on Data Streams: A New Paradigm for Load Shedding”, Technical Report 120016, UCLA, 2011.

[P4] **Barzan Mozafari**, Kai Zeng, Carlo Zaniolo, “K\*SQL: a Unifying Engine for Sequence Patterns and XML”, In proceedings of **SIGMOD**, 2010. [**Honorable Mention Demo**]

[P3] 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.

[P2] Hetal Thakkar, **Barzan Mozafari** and Carlo Zaniolo “A Data Stream Mining System”, In proceedings of the International Conference on Data Mining (**ICDM**), 2008.

[P1] 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.

UNDER  
SUBMISSION

[U2] **Barzan Mozafari**, Purnamrita Sarkar, Michael Franklin, Michael Jordan, Samuel Madden, “Active Learning for Crowd-Sourced Databases”, Nov 2012.  
<http://arxiv.org/abs/1209.3686>

[U1] **Barzan Mozafari**, Stephen Tu, Andrew Lamb, Samuel Madden, Michael Stonebraker, “A Black-box Approach to Database Design through Robust Optimization”, Dec 2012.

TEACHING  
EXPERIENCE

- TA for Introduction to Machine Language Programming, UCLA, Spring 2008.
- TA for Introduction to Computer Programming, UCLA, Fall and Winter 2007.
- TA (as an undergrad) for Algorithms, Shahid Beheshti University, 2004-2005.
- TA (as an undergrad) for Computer Networks, Shahid Beheshti University, 2005.
- TA (as an undergrad) for Data Structures, Amirkabir University of Technology, 2005.

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.

INVITED TALKS

- EMC head-quarters, Hopkinton, Massachusetts, November, 2012.
- University of Pennsylvania, November, 2012.
- Database Workload Management Workshop, Dagstuhl, Germany, July, 2012.
- SIGMOD, Scottsdale, Arizona, May, 2012.
- UC Berkeley AMPLab's Summer Retreat, May, 2012.
- UC Santa Cruz, October, 2012.
- VLDB, Singapore, September, 2010.
- SIGMOD, Indianapolis, Indiana, June, 2010.
- ICDE, Long Beach, California, March, 2010.
- ICDM, Pisa, Italy, December, 2008.
- University of Calabria, Rende, Italy, December 2008.
- ICWSM, Boulder, Colorado, March, 2007.

PROFESSIONAL  
ACTIVITIES

PC (Program Committee), Intl. Conf. on Data Engineering (ICDE), 2013.  
PC, Data Management in the Cloud (DMC) Workshop, 2013.  
PC, Intl. Conf. on Scientific & Statistical Database Management (SSDBM), 2013  
PC, DOA-Trusted Cloud Intl. Conf. on Secure Virtual Infrastructures, 2013  
PC, Intl. Conf. on Information and Knowledge Management (CIKM) 2013.

REFERENCES

**Prof. Samuel Madden**

Director of BigData@CSAIL  
MIT EECS Department, CSAIL  
Room 32-G938, 32 Vassar St.  
Cambridge, MA 02139  
Phone: (617) 258-6643  
Email: madden@csail.mit.edu

**Prof. Ion Stoica**

UC Berkeley Computer Science Division  
465 Soda Hall, Berkeley, CA 94720  
Phone: (510) 643-4007  
Email: istoica@cs.berkeley.edu

**Prof. Carlo Zaniolo**

UCLA Computer Science Department  
3532D Boelter Hall  
420 Westwood Plaza  
Los Angeles, CA 90095, USA  
Phone: (310) 825-8137  
Email: zaniolo@cs.ucla.edu

**Prof. Michael J. Franklin**

UC Berkeley Computer Science Division  
465 Soda Hall, Berkeley, CA 94720  
Phone: (510) 642-1662  
Email: franklin@cs.berkeley.edu