Shivaram Venkataraman

901 Drake Street, 401, Madison, WI 53715, 217-766-2227
shivaram@cs.wisc.edu, http://shivaram.org

RESEARCH INTERESTS

Computer Systems, Databases, Machine Learning

EDUCATION

PhD, Computer Science
University of California, Berkeley,
December 2017
Advisors: Michael J. Franklin, Ion Stoica

Master of Science, Computer Science
University of Illinois, Urbana-Champaign
May 2011
Advisor: Roy H. Campbell

Bachelor of Engineering (Honors), Computer Science
Birla Institute of Technology and Science, Pilani, India
June 2006

EMPLOYMENT

Assisstant Professor 8/18-Present
University of Wisconsin - Madison, WI

Post-Doctoral Researcher 9/17-8/18
Microsoft Research - Redmond, WA

- Worked with Systems Research Group on developing scalable systems support for large scale machine learning.

Research Intern 5/11-8/11
HP Research Labs - Palo Alto, CA

- Worked with Indrajit Roy and Robert S. Schreiber on the design and implementation of a distributed data processing engine on R. This project was later released by HP Labs as DistributedR and is now a part of Vertica's advanced analytics toolkit.

Research Intern 5/10-8/10
HP Research Labs - Palo Alto, CA

- Worked with Niraj Tolia and Parthasarathy Ranganathan on the design of storage systems and data structures structures for non-volatile memory.

Software Engineer 8/06-8/09
Google Inc, Bangalore, India

- *Mobile Search, 2006-2007*: Developed, deployed and maintained the infrastructure for google search over short messaging service (SMS).
- *Orkut, 2007-2009*: Worked on a project to enable sharing of photographs between users on Orkut, a social networking website. I designed and implemented the storage layer for this application. I also enhanced the security features of the system by building controls to prevent distributed denial of service attacks.

AWARDS

Best Vision Paper, SoCC 2017
Best Paper runner-up, Applied Data Science Track, KDD 2016
Communications of the ACM Research Highlight, 2014
Best of VLDB 2012, Invitation to the VLDB Journal
UC Berkeley Outstanding Graduate Student Instructor Award, 2012
Siebel Scholar, Class of 2011
Outstanding Teaching Assistant, University of Illinois, Urbana-Champaign, 2009

PUBLICATIONS

Journal Articles

Xiangrui Meng, Joseph Bradley, Burak Yavuz, Evan Sparks, Shivaram Venkataraman, Davies Liu, Jeremy Freeman, DB Tsai, Manish Amde, Sean Owen, Doris Xin, Reynold Xin, Michael J Franklin, Reza Zadeh, Matei Zaharia, Ameet Talwalkar *MLlib: Machine Learning in Apache Spark* JMLR 17(34):17, 2016

Peter Bailis, Shivaram Venkataraman, Michael J Franklin, Joseph M. Hellerstein, Ion Stoica *Quantifying Eventual Consistency with PBS* The VLDB Journal, April 2014, 23(2):279302. BEST OF VLDB 2012 SPECIAL ISSUE

Conference Publications

Anand Padmanabha Iyer, Zaoxing Liu and Xin Jin, Shivaram Venkataraman, Vladimir Braverman, Ion Stoica *ASAP: Fast, Approximate Pattern Mining at Scale* OSDI 2018

Kevin Hsieh, Ganesh Ananthanarayanan, Peter Bodik, Shivaram Venkataraman, Paramvir Bahl, and Matthai Philipose, Phillip B. Gibbons, Onur Mutlu *Focus: Querying Large Video Datasets with Low Latency and Low Cost* OSDI 2018

Luo Mai, Kai Zeng, Rahul Potharaju, Le Xu, Steve Suh, Shivaram Venkataraman, Paolo Costa, Terry Kim, Saravanam Muthukrishnan, Vamsi Kuppa, Sudheer Dhulipalla, Sriram Rao *Chi: A Scalable and Programmable Control Plane for Distributed Stream Processing Systems* VLDB 2018

Shivaram Venkataraman, Aurojit Panda, Kay Ousterhout, Michael Armbrust, Ali Ghodsi, Michael J. Franklin, Benjamin Recht, Ion Stoica *Drizzle: Fast and Adaptable Stream Processing at Scale* SOSP 2017

Eric Jonas, Qifan Pu, Shivaram Venkataraman, Ion Stoica, Benjamin Recht *Occupy the Cloud: Distributed Computing for the 99%* SoCC 2017

Stephen Tu, Shivaram Venkataraman, Ashia C. Wilson, Alex Gittens, Michael I. Jordan, Benjamin Recht *Breaking Locality Accelerates Block Gauss-Seidel* ICML 2017

Evan R. Sparks, Shivaram Venkataraman, Tomer Kaftan, Michael J. Franklin, Benjamin Recht *KeystoneML: Optimizing Pipelines for Large-Scale Advanced Analytics* ICDE 2017

Omid Alipourfard, Jianshu Chen, Hongqiang Liu, Shivaram Venkataraman, Minlan Yu, Ming Zhang *CherryPick: Adaptively Unearthing the Best Cloud Configurations* USENIX NSDI 2017

Shivaram Venkataraman, Zongheng Yang, Michael J Franklin, Ben Recht, Ion Stoica *Ernest: Efficient Performance Prediction for Large Scale Advanced Analytics* USENIX NSDI 2016

Shivaram Venkataraman, Zongheng Yang, Davies Liu, Eric Liang, Hossein Falaki, Xiangrui Meng, Reynold Xin, Ali Ghodsi, Michael Franklin, Ion Stoica, Matei Zaharia *SparkR: Scaling R Programs with Spark* ACM SIGMOD 2016

Reza Zadeh, Xiangrui Meng, Alexander Ulanov, Burak Yavuz, Li Pu, Shivaram Venkataraman, Evan Sparks, Aaron Staple, Matei Zaharia *Matrix Computations and Optimization in Apache Spark* ACM KDD 2016

Shivaram Venkataraman, Aurojit Panda, Ganesh Ananthanarayanan, Michael Franklin, Ion Stoica *The Power of Choice in Data-Aware Cluster Scheduling* USENIX OSDI 2014

Shivaram Venkataraman, Erik Bodzsar, Indrajit Roy, Alvin AuYoung, and Robert S. Schreiber *Presto: Distributed Machine Learning and Graph Processing with Sparse Matrices* ACM Eurosys 2013

Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, Ion Stoica *Probabilistically Bounded Staleness for Practical Partial Quorums* VLDB 2012. Invited to Best of VLDB 2012 VLDBJ Issue. Selected as CACM research highlight.

Andrew Wang, Shivaram Venkataraman, Sara Alspaugh, Ion Stoica, and Randy Katz *Cake: Enabling High-level SLOs on Shared Storage Systems* ACM SoCC 2012

Reza Farivar, Harshit Kharbanda, Shivaram Venkataraman, Roy Campbell *An Algorithm for Fast Edit Distance Computation on GPUs* IEEE Innovative Parallel Computing (InPar) 2012

Shivaram Venkataraman, Niraj Tolia, Parthasarathy Ranganathan, Roy Campbell *Consistent and Durable Data Structures for Non-Volatile Byte-Addressable Memory* USENIX FAST 2011

Elllick Chan, Shivaram Venkataraman, Francis David, Amey Chaugule, Roy Campbell *Forenscope: A Framework for Live Forensics* ACSAC 2010

Abhishek Verma, Xavier Llora, Shivaram Venkataraman, David Goldberg and Roy Campbell *Scaling eCGA Model Building via Data Intensive Computing* IEEE CEC 2010

Workshop Publications
Xinghao Pan, Shivaram Venkataraman, Zizheng Tai, Joseph Gonzalez *Hemingway: Modeling Distributed Optimization Algorithms* Machine Learning Systems Workshop 2016

Kay Ousterhout, Aurojit Panda, Joshua Rosen, Shivaram Venkataraman, Reynold Xin, Sylvia Ratnasamy, Scott Shenker, Ion Stoica *The Case for Tiny Tasks in Compute Clusters* HotOS 2013

Andrew Wang, Shivaram Venkataraman, Sara Alspaugh, Ion Stoica, and Randy Katz *Sweet Storage SLOs with Frosting* HotCloud 2012

Shivaram Venkataraman, Indrajit Roy, Alvin AuYoung, and Robert S. Schreiber *Using R for Iterative and Incremental Processing* HotCloud 2012

Shivaram Venkataraman, Niraj Tolia, Parthasarathy Ranganathan, Roy Campbell *Redesigning Data Structures for Non-Volatile Byte-Addressable Memory* Non-Volatile Memories Workshop 2011

Ellick Chan, Shivaram Venkataraman, Nadia Tkach, Kevin Larson, Alejandro Gutierrez and Roy H. Campbell *Characterizing Data Structures for Volatile Forensics* Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE), 2011

Contributed Articles
Matei Zaharia, Reynold S. Xin, Patrick Wendell, Tathagata Das, Michael Armbrust, Ankur Dave, Xiangrui Meng, Josh Rosen, Shivaram Venkataraman, Michael J. Franklin, Ali Ghodsi, Joseph Gonzalez, Scott Shenker, Ion Stoica *Apache Spark: A Unified Engine for Big Data Processing* Communications of the ACM, November 2016

Demonstrations
Peter Bailis, Shivaram Venkataraman, Michael Franklin, Joseph M. Hellerstein, and Ion Stoica *PBS at Work: Advancing Data Management with Consistency Metrics* Demo at SIGMOD 2013

Technical Reports
Evan R. Sparks, Shivaram Venkataraman, Tomer Kaftan, Michael J. Franklin, Benjamin Recht *KeystoneML: Optimizing Pipelines for Large-Scale Advanced Analytics* arXiv preprint arXiv:1610.09451

Stephen Tu, Rebecca Roelofs, Shivaram Venkataraman, Ben Recht *Large Scale Kernel Learning using Block Coordinate Descent* arXiv preprint arXiv:1602.05310

Theses
Storage system design for non-volatile byte-addressable memory using consistent and durable data structures *Masters Thesis* University of Illinois, Urbana-Champaign 2011

TALKS

Drizzle: Fast and Adaptable Stream Processing at Scale
- Google, Ads Data Infrastructure, Sep 2016
- Spark Summit, June 2016

Ernest: Efficient Performance Prediction for Large Scale Advanced Analytics
- NSDI, March 2016

SparkR: Scaling R Programs with Spark
- ACM SIGMOD, June 2016
- useR!, June 2016
- Spark Summit, July 2015
- Hadoop Summit, June 2014

The Power of Choice in Data-Aware Cluster Scheduling
- OSDI, September 2014

Presto: Distributed Machine Learning and Graph Processing with Sparse Matrices
- Eurosys, April 2013
- HotCloud, June 2012

Probabilistically Bounded Staleness for Practical Partial Quorums (joint talk with Peter Bailis)
- VLDB, August 2012

Consistent and Durable Data Structures for Non-Volatile Byte-Addressable Memory
- FAST, February 2011
- Non-Volatile Memory Workshop, March 2011

TEACHING & MENTORING

Undergraduate Research Mentor 11/14-present
UC Berkeley - Berkeley, CA
- Guided undergraduate students during various research projects:
  - *Zongheng Yang, EECS 2015*, now at Google, starting PhD at UC Berkeley in 2017
  - *Tomer Kaftan, EECS 2013*, PhD student at University of Washington
  - *Zizheng Tai, EECS 2018 (Expected)*, Junior at UC Berkeley

Graduate Student Instructor 1/14-5/14
UC Berkeley - Berkeley, CA

- *GSI, Introduction to Data Science*: Designed a new data science course for senior undergraduate students. Created lab assignments that were part of the class every week and homeworks that incorporated web programming, statistics, and the ability to manipulate data sets with code.

Graduate Student Instructor 8/12-12/12
UC Berkeley - Berkeley, CA

- *GSI, Introduction to Database Systems*: Instructed two weekly sections of around 35 junior and senior undergraduate students on core database topics. Designed and organized homework assignment using PostgreSQL and helped instructor in course organization, grading.

Teaching Assistant 8/09-5/10
University of Illinois - Urbana-Champaign

- *Teaching Assistant, Data Structures*: Led weekly discussion and lab sections for freshman and sophomore undergraduate students. Also helped the instructor in grading and organizing material for honors students.

SERVICE

- Program committee member for USENIX ATC 2018, USENIX HotCloud 2018
- AMPCamp organizer 2012, 2013, 2014.
- Organized UC Berkeley Visit Day 2016 for incoming students in Systems, Networking, Databases.
- AMPLab Cloud Seminar Organizer in Fall 2015.
- Paper reviews for SIGMOD 2016, Eurosys 2013, SOCC 2012, OSDI 2012
- Guest lecturer for CS 267 at UC Berkeley in Spring 2014, Spring 2015.

REFERENCES

Michael J. Franklin (co-advisor)
Liew Family Chair of Computer Science
Computer Science
University of Chicago
mjfranklin@cs.uchicago.edu

Ion Stoica (co-advisor)
Professor
Electrical Engineering & Computer Science
University of California, Berkeley
istoica@cs.berkeley.edu

Benjamin Recht
Associate Professor
Electrical Engineering & Computer Science, Statistics
University of California, Berkeley
brecht@berkeley.edu

Matei Zaharia
Assistant Professor
Computer Science
Stanford University
matei@cs.stanford.edu