Yanlei Diao

Office: School of Computer Science
University of Massachusetts Amherst
140 Governors Drive
Amherst, MA 01003-9264

Phone: (413) 545-1135
Fax: (413) 545-1249
Email: yanlei@cs.umass.edu
URL: http://www.cs.umass.edu/~yanlei

RESEARCH INTERESTS

Information infrastructures and data management systems,
big data systems, data stream systems, interactive data exploration, uncertain data management,
high-performance genomic data processing, RFID/sensor data management,
flash memory databases, publish/subscribe, XML message brokering

EDUCATION

University of California, Berkeley   Berkeley, CA
Ph.D., Electrical Engineering and Computer Science   August 2005
Advisor: Professor Michael J. Franklin
Thesis: "Query Processing for Large-Scale XML Message Brokering"

Hong Kong University of Science and Technology   Hong Kong, China
M.S., Computer Science   June 2000
Advisor: Professor Hongjun Lu
Thesis: "Learning-based Web Query Processing"

Fudan University   Shanghai, China
B.S., Computer Science   July 1998

ACADEMIC POSITIONS

Associate Professor   09/2011 – present
School of Computer Science
University of Massachusetts Amherst

Assistant Professor   09/2005 – 08/2011
Department of Computer Science
University of Massachusetts Amherst

Research Intern   02/2003 – 08/2003
BEA Systems, San Jose, CA

Summer Intern   06/2001 – 08/2001
IBM Almaden Research Center, San Jose, CA

HONORS AND AWARDS

- **CRA-W Borg Early Career Award**, 2013 (*one female scientist selected each year across all fields of Computer Science and Engineering for significant contributions in research and outreach*)

- **Distinguished Alumni**, School of Computer Science, Fudan University, Shanghai China, 2013

- **IBM Innovation Award on Scalable Data Analytics**, 2010

- **National Science Foundation CAREER Award**, 2008

- **Finalist for Microsoft Research New Faculty Fellowship**, 2008

- **Nomination for Distinguished Teaching Award**, University of Massachusetts Amherst, Fall 2007 and Spring 2008

- **Distinguished Faculty Lecture Series**, University of Texas at Austin, 2005

- **ACM-SIGMOD Dissertation Award Honorable Mention**, 2005

- First Class Honor, Fudan University, 1998

- Intel Fellowship, Fudan University, 1996-1998

- GE Fellowship, Fudan University, 1995-1996

PROFESSIONAL SERVICE

- ### ADVISORY BOARDS AND AWARD COMMITTEES

  ACM SIGMOD Executive Committee, April 2014 – present
  ACM SIGMOD Software Systems Award Committee, September 2014 – present
  Big Data Advisory Board, Mass Technology Leadership Council, January 2014 - present

- ### JOURNALS

  Editor-in-Chief, ACM SIGMOD Record, April 2014 – present
  Associate Editor, ACM Transactions on Database Systems (TODS), 2014 – 2016
  Associate Editor, Proceedings of VLDB (PVLDB), 2012 – 2013, 2015 – 2016

  Reviewer, ACM Transactions on Database Systems (TODS)
  Reviewer, International Journal on Very Large Data Bases (VLDB Journal)
  Reviewer, ACM Transactions on Information Systems (TOIS)
  Reviewer, ACM Transactions on Internet Technology (TOIT)
  Reviewer, IEEE Transactions on Knowledge & Data Engineering (TKDE)
  Reviewer, Journal of Computer Science and Technology (JCST)

- ### CONFERENCES AND WORKSHOPS

  Area Chair, ACM SIGMOD Conference, 2014 – 2015
  Area Chair and Organizing Committee Member, the Abstract track and Gong Show at the Conference on Innovative Data Systems Research (CIDR), 2011 – 2015
  Chair, New England Database (NEDB) Summit, 2011

Co-Chair, New Researcher Symposium, ACM SIGMOD Conference, 2010 – 2011
Co-Chair, International Workshop on Data Management for Sensor Networks, 2008 – 2009
Steering Committee, International Workshop on Data Management for Sensor Networks, 2010
Publicity Chair, ACM SIGMOD Conference on Management of Data (SIGMOD), 2009
Co-organizer, New England Database Society Seminar Series, 2009-present

Committee Member, ACM International Conference on Management of Data (SIGMOD), 2011, 2014
Committee Member, International Conference on Very Large Data Base (VLDB), 2007, 2008, 2009, 2010, 2013, 2014
Committee Member, IEEE International Conference on Data Engineering (ICDE), 2007, 2008, 2009, 2010
Committee Member, ACM International Conference on Information and Knowledge Management (CIKM), 2010
Committee Member, Conference on Innovative Data Systems Research (CIDR), 2009, 2011, 2013
Committee Member, International Workshop on Networking Meets Databases (NetDB), 2008
Committee Member, International Workshop on RFID Data Management (RFDM), 2008
Committee Member, International Workshop on Data Management for Sensor Networks, 2007-2010
Committee Member, International Workshop on Scalable Stream Processing, 2007
Committee Member, International XML Database Symposium (XSym), 2006
Committee Member, International Workshop on Peer-to-Peer Information Management, 2006
Committee Member, ACM International Conference on Management of Data (SIGMOD), Demonstrations, 2005

- **Panels**
  Panelist, International Workshop on Data Management for Sensor Networks (DMSN), 2010
  Panelist, International Workshop on RFID Data Management (RFDM), 2008
  Panelist, National Science Foundation, 2007

- **Diversity Efforts**
  Mentor and Speaker, CRA-W Graduate Cohort, 2011, 2013, 2014
  Mentor and Participant, CRA-W Distributed Mentor Project (DMP), 2010, 2012, 2013
  Massachusetts Aspirations in Computing Affiliate Award (MACAA), Cambridge MA, 2010 – 2014
  Dot Diva Launch Event, Cambridge MA, September 2010
  DataBase Mentoring (DB Me) Workshop, ACM SIGMOD 2010
  Speaker, Women in Science Club at UMass Boston, April 2010

UNIVERSITY SERVICE

- **Chair**
  Recruitment and Outreach Committee of School of Computer Science, 2012 – 2015

- **Member**
  Faculty Recruiting Committee, 2008-2009, 2010-2011, 2013-2015
  Dean Search committee, 2014-2015
  Undergraduate Committee, 2009-2010
  Graduate Committee, 2007-2008, 2009-2010
  Diversity Committee, 2005-2008
  Personnel Committee, 2006-2007, 2010-2011
  Graduate Admissions Committee, 2005-2006

- **Co-Organizer**
  First UMass Big Data Workshop, September 2013

- **Speaker and Panelist**
  Lab Description Seminar, Fall 2005, Fall 2006, Fall 2007, Fall 2009
  Professionalism Seminar on Job Hunting, Fall 2005
  CS Women Luncheon Meeting, Fall 2005
  Professionalism Seminar on Job Hunting, Fall 2005

PUBLICATIONS

I. DATA MANAGEMENT: DEVICES AND PLATFORMS

[1]   Buduo Li, Ed Mazur, Yanlei Diao, Andrew McGregor, and Prashant Shenoy. "Scalla: A Platform for Scalable One-pass Analytics using MapReduce". *ACM Transactions on Database Systems (**TODS**)*, 37(4):27-64, December 2012. ***Special Issue on Best Papers of SIGMOD 2011***.

[2]   Buduo Li, Ed Mazur, Yanlei Diao, Andrew McGregor, and Prashant Shenoy. "A Platform for Scalable On-pass Analytics using MapReduce". In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (***SIGMOD***), 985-996, June 2011. (*Selected as one of the best four papers*)

[3]   Devesh Agrawal, Deepak Ganesan, Ramesh Sitaraman, Yanlei Diao, and Shashi Singh. "Lazy-Adaptive Tree: An Optimized Index Structure for Flash Devices." In *Proceedings of the 35th International Conference on Very Large Data Bases* (***VLDB***), 361-372, August 2009. (*16.7% acceptance rate*)

[4]   Ed Mazur, Boduo Li, Yanlei Diao, and Prashant Shenoy. "Towards Scalable One-Pass Analytics using MapReduce." In Proceedings of the *1st International Workshop on Data Intensive Computing in the Clouds, May 2011.* 10 pages.

II. DATA MANAGEMENT: ANALYTICS AND QUERY PROCESSING

Complex Event Processing over Streams

[5]   Haopeng Zhang, Yanlei Diao, and Neil Immerman. "On Complexity and Optimization of Expensive Queries in Complex Event Processing." In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (***SIGMOD***), 2014. (*25.6% acceptance rate*)

[6]   Haopeng Zhang, Yanlei Diao, and Neil Immerman. "Recognizing Patterns in Streams with Imprecise Timestamps." ***Information Systems***, *Elsevier*, 38(3), 1187-1211, November 2013.

[7]   Haopeng Zhang, Yanlei Diao, and Neil Immerman. "Recognizing Patterns in Streams with Imprecise Timestamps." *Journal "Proceedings of the VLDB Endowment" (**PVLDB**)*, 3(1): 244-255, 2010.

[8]   Jagrati Agrawal, Yanlei Diao, Daniel Gyllstrom, and Neil Immerman. "Efficient Pattern Matching over Event Streams". In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (***SIGMOD***), pages 147-160, June 2008. (*18% acceptance rate*)

[9] Eugene Wu, Yanlei Diao, and Shariq Rizvi. "High-Performance Complex Event Processing over Streams." In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (***SIGMOD***), pages 407-418, June 2006. (*13% acceptance rate*; ***received over 600 citations*** *according to Google Scholar*)

[10] Daniel Gyllstrom, Jagrati Agrawal, Yanlei Diao, and Neil Immerman. "On Supporting Kleene Closure over Event Streams." In *Proceedings of the 24th International Conference on Data Engineering* (*ICDE*), pages 1391-1393, April 2008. (*31% acceptance rate*)

[11] Daniel Gyllstrom, Eugene Wu, Hee-Jin Chae, Yanlei Diao, Patrick Stahlberg, and Gordon Anderson. "SASE: Complex Event Processing over Streams." System Demo. In *Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research* (*CIDR*), pages 407-411, January 2007.

Interactive Data Exploration

[12] Kyriaki Dimitriadou, Olga Papaemmanouil, and Yanlei Diao. "Explore-by-Example: An Automatic Query Steering Framework for Interactive Data Exploration." In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (***SIGMOD***), 2014. (*25.6% acceptance rate*)

[13] Ugur Cetintemel, Mitch Cherniack, Justin DeBrabant, Yanlei Diao, Kyriaki Dimitriadou, Alex Kalinin, Olga Papaemmanouil, Stan Zdonik. "Query Steering for Interactive Data Exploration." In the 6th Biennial *Conference on Innovative Data Systems Research (CIDR)*, 2013. (Vision paper).

[14] Kyriaki Dimitriadou, Olga Papaemmanouil and Yanlei Diao. "Interactive data exploration based on user relevance feedback." In *Proceedings of the 30th International Conference on Data Engineering Workshops*, ICDE, Chicago, IL, USA, March 31 - April 4, 2014.

Uncertain Data Management

[15] Thanh T.L. Tran, Yanlei Diao, Charles Sutton, and Anna Liu. "Supporting User-Defined Functions on Uncertain Data". *Journal "Proceedings of the VLDB Endowment" (**PVLDB**)*, 6(6), 469-480, August 2013.

[16] Thanh T. L. Tran, Liping Peng, Yanlei Diao, Andrew McGregor, and Anna Liu. "CLARO: Modeling and Processing Uncertain Data Streams." ***VLDB Journal***, 21(5):651-676, November 2012.

[17] Liping Peng, Yanlei Diao, and Anna Liu. "Optimizing Probabilistic Query Processing on Continuous Uncertain Data". *Journal "Proceedings of the VLDB Endowment" (**PVLDB**)*, 4(11): 1169-1180, 2011.

[18] Thanh Tran, Andrew McGregor, Yanlei Diao, Liping Peng, and Anna Liu. "Conditioning and Aggregating Uncertain Data Streams: Going Beyond Expectations." *Journal "Proceedings of the VLDB Endowment" (**PVLDB**)*, 3(1): 1302-1313, 2010.

[19] Thanh Tran, Liping Peng, Boduo Li, Yanlei Diao, and Anna Liu. "PODS: Modeling and Processing of High-Volume Uncertain Data Streams." In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (***SIGMOD***), 159-170, June 2010. (*20.8% acceptance rate*)

[20] Yanlei Diao, Boduo Li, Anna Liu, Liping Peng, Charles Sutton, Thanh Tran, and Michael Zink. "Capturing Uncertainty in High-Volume Stream Processing." In *Proceedings of the 4th Biennial*

Conference on Innovative Data Systems Research (**CIDR**), January 2009. 11 pages. (*27.1% acceptance rate*)

XML Query Processing

[21] Yanlei Diao and Michael J. Franklin. "High-Performance XML Message Brokering." *Data Stream Management: Processing High-Speed Data Streams* (Garofalakis, Gehrke, and Rastogi, eds.). Springer Data-Centric Systems and Applications Series, November 2010.

[22] Yanlei Diao and Michael J. Franklin. "Publish/Subscribe over Streams." *Encyclopaedia of Database Systems*. Springer, 2009. 6 pages.

[23] Yanlei Diao and Michael J. Franklin. "XML Publish/Subscribe." *Encyclopaedia of Database Systems*. Springer, 2009. 6 pages.

[24] Yanlei Diao. "Query Processing for Large-Scale XML Message Brokering." PhD Dissertation, University of California, Berkeley, August 2005. ***ACM-SIGMOD Dissertation Award Honorable Mention***. http://www.cs.umass.edu/~yanlei/publications/Yanlei_thesis.pdf.

[25] Yanlei Diao, Mehmet Altinel, Michael J. Franklin, Hao Zhang, and Peter Fischer. "Path Sharing and Predicate Evaluation for High-Performance XML Filtering." *ACM Transactions on Database Systems* (**TODS**) 28(4): 467-516, December 2003. ACM Press, New York, NY.

[26] Yanlei Diao, Daniela Florescu, Donald Kossmann, Michael J. Carey, and Michael J. Franklin. "Implementing Memoization in a Streaming XQuery Processor." In *Proceedings of the 2nd International XML Database Symposium* (**XSym**), pages 35-50, August 2004. (*25% acceptance rate*)

[27] Yanlei Diao and Michael J. Franklin. "Query Processing for High-Volume XML Message Brokering." In *Proceedings of the 29th International Conference on Very Large Data Bases* (**VLDB**), pages 261-272, September 2003. (*14% acceptance rate*)

[28] Yanlei Diao, Peter Fischer, Michael J. Franklin, and Raymond To. "YFilter: Efficient and Scalable Filtering of XML Documents." System Demo. In *Proceedings of the 18th International Conference on Data Engineering* (**ICDE**), pages 341-342, February 2002.

[29] Yanlei Diao and Michael J. Franklin. "High-Performance XML Filtering: An Overview of YFilter." *IEEE Data Engineering Bulletin* 26(1): 41-48, March, 2003.

III. DATA MANAGEMENT: SPECIFIC APPLICATION DOMAINS

RFID/Sensor Networks

[30] Yanming Nie, Richard Cocci, Zhao Cao, Yanlei Diao, and Prashant Shenoy. "SPIRE: Efficient Inference and Compression over RFID Streams." *IEEE Transactions on Knowledge and Data Engineering* (**TKDE**), 24(1): 141-155, January 2012.

[31] Zhao Cao, Charles Sutton, Yanlei Diao, and Prashant Shenoy. "Distributed Inference and Query Processing for RFID Tracking and Monitoring." *Journal "Proceedings of the VLDB Endowment" (PVLDB)*, 4(5): 326-337, 2011.

[32] Exploiting the Interplay Between Memory and Flash Storage In Embedded Sensor Devices. Devesh Agrawal, Boduo Li, Zhao Cao, Deepak Ganesan, Yanlei Diao, Prashant Shenoy. In *Proceedings of the IEEE International Conference on Embedded and Real-Time Computing Systems and Applications* (**RTCSA**), August 2010. 10 pages. (*23% acceptance rate*)

[33] Thanh Tran, Charles Sutton, Richard Cocci, Yanming Nie, Yanlei Diao, and Prashant Shenoy. "Probabilistic Inference over RFID Streams in Mobile Environments." In *Proceedings of the 25th International Conference on Data Engineering* (**ICDE**), pages 1096-1107, March 2009. (*17% acceptance rate*)

[34] Yanlei Diao, Deepak Ganesan, Gaurav Mathur, and Prashant Shenoy. "Re-thinking Data Management for Storage-centric Sensor Networks." In *Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research* (**CIDR**), pages 22-31, January 2007. (*44% acceptance rate*)

[35] Zhao Cao, Yanlei Diao, and Prashant Shenoy. "Architectural Considerations for Distributed RFID Tracking and Monitoring." In *Proceedings of the 5th International Workshop on Networking Meets Databases* (*NetDB*), October 2009. 6 pages.

[36] Richard Cocci, Thanh Tran, Yanlei Diao, and Prashant Shenoy. "Efficient Data Interpretation and Compression over RFID Streams." In *Proceedings of the 24th International Conference on Data Engineering* (*ICDE*), pages 1445-1447, April 2008. (*31% acceptance rate*)

[37] Richard Cocci, Yanlei Diao, and Prashant Shenoy. "SPIRE: Scalable Processing of RFID Event Streams." In *Proceedings of the 5th RFID Academic Convocation*, April 2007. 6 pages.

High-Performance Genomic Data Processing

[38] Abhishek Roy, Yanlei Diao, and Toby Bloom. "Building Highly-Optimized, Low-Latency Pipelines for Genomic Data Analysis." In *Proceedings of the 7th Biennial Conference on Innovative Data Systems Research* (**CIDR**), January 2015. 12 pages.

[39] Abhishek Roy, Yanlei Diao, Evan Mauceli, Yiping Shen, Bai-Lin Wu: Massive Genomic Data Processing and Deep Analysis. *Proceedings of the VLDB Endowment* (*PVLDB*), 5(12): 1906-1909 (2012).

IV. NETWORKING AND DATA DISSEMINATION

[40] Fang Yu, Yanlei Diao, Randy Katz, and T. V. Lakshman. "Fast Packet Pattern Matching Algorithms." *Algorithms for Next Generation Network Architecture* (Graham Cormode and Marina Thottan, eds.). Springer Computer Communications and Networks Series, October 2009.

[41] Fang Yu, Zhifeng Chen, Yanlei Diao, T.V. Lakshman, and Randy H. Katz. "Fast and Memory-Efficient Regular Expression Matching." *ACM/IEEE Symposium on Architectures for Networking and Communications Systems* (**ANCS**), pages 93-102, December 2006. (*19% acceptance rate*)

[42] Yanlei Diao, Shariq Rizvi, and Michael J. Franklin. "Towards an Internet-Scale XML Dissemination Service." In *Proceedings of the 30th International Conference on Very Large Data Bases* (**VLDB**), pages 612-623, August 2004. (*16% acceptance rate*)

V. INFORMATION RETRIEVAL AND DATA MINING

[43] Michael Bendersky, W. Bruce Croft, and Yanlei Diao. "Quality-Biased Ranking of Web Documents". In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining* (**WSDM**), 10 pages, 2011. (*8.3% for oral presentation*)

[44] Desislava Petkova, W. Bruce Croft, and Yanlei Diao. "Refining Keyword Queries for XML Retrieval by Combining Content and Structure." In *Proceedings of 31st European Conference on Information Retrieval* (**ECIR**), pages 662-669, April 2009. (*31.9% acceptance rate*)

[45] Yanlei Diao, Hongjun Lu, Songting Chen, and Zengping Tian. "Toward Learning Based Web Query Processing." In *Proceedings of the 26th International Conference on Very Large Data Bases* (**VLDB**), pages 317-328, September 2000. (*15% acceptance rate*)

[46] Yanlei Diao, Hongjun Lu, and Dekai Wu. "A Comparative Study of Classification Based Personal E-mail Filtering." In *Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining* (**PAKDD**), pages 408-419, April 2000. (*29% acceptance rate*)

[47] Songting Chen, Yanlei Diao, Hongjun Lu and Zengping Tian. "FACT: A Learning Based Web Query Processing System." System Demo. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data* (*SIGMOD*), page 587, May 2000.

VI. SOFTWARE ENGINEERING

[48] Kaituo Li, Christoph Reichenbach, Yannis Smaragdakis, Yanlei Diao, and Christoph Csallner. "SEDGE: Symbolic Example Data Generation for Dataflow Programs." In the *Proceedings of the 28th IEEE/ACM International Conference on Automated Software Engineering* (**ASE**), 2013. (*13.6% acceptance rate*)

RESEARCH SUPPORT

STATE AND FEDERAL FUNDING

- Yanlei Diao (PI) and Toby Bloom. "Collaborative Research: ABI Development: A New Platform for Highly-Optimized, Low-Latency Pipelines for Genomic Data Analysis." **National Science Foundation**, DBI-1356486, $865,382.00, 09/15/2014-08/31/2017.

- Christopher Hill (PI), Prashant Shenoy (UMass PI), Yanlei Diao (co-PI) with other Co-PIs. Commonwealth Computing Cloud for Data Driven Biology (C3DDB). **Mass Life Sciences Center**, **$4.5 million**, for purchasing clusters at MGHPCC for data-intensive computing in life sciences.

- Yanlei Diao (PI) and Anna Liu. "III: Small: High-Performance Complex Processing of Continuous Uncertain Data." **National Science Foundation**, IIS-1218524, $495,961, 09/01/2012-08/31/2015.

    o REU Supplement, $16,000, 06/2013-08/2015.

- Kevin Fu (PI), Yanlei Diao (co-PI), Charles Ross, Deepak Ganesan, Wayne Burleson. "MRI: Acquisition of an RFID Testbed Using Renewable Energy for Object Identification and Habitat Monitoring." **National Science Foundation** CNS-0923313, $450,010, 10/2009-09/2012.

- Yanlei Diao (PI) and Anna Liu. "III-COR-small: Capturing Data Uncertainty in High-Volume Stream Processing." **National Science Foundation** IIS-0812347, $441,594, 09/2008-08/2011.

    o REU Supplement, $16,000, 06/2010-05/2011.

- Yanlei Diao (PI). "CAREER: Efficient, Robust RFID Stream Processing for Tracking and Monitoring." **National Science Foundation** IIS-0746939, $588,879, 06/2008-05/2013.

    o REU Supplement, $16,000, 07/02/2009-06/30/2010.

- Deepak Ganesan (PI), Yanlei Diao (co-PI), Prashant Shenoy. "NeTS-NOSS: STONES: Storage-centric Networked Embedded Systems." **National Science Foundation** CNS-0626873, $500,000, 09/2006-08/2010.

INDUSTRY, UNIVERSITY, AND OTHERS

- Yanlei Diao (PI) and Prashant Shenoy (PI). "Big Data Informatics Initiative: Platforms, Algorithms, and Applications for Real-time Big Data Analytics." **UMass President's Science & Technology Initiatives Fund Award**, $167,500.

- Gerome Miklau (PI), Yanlei Diao, and Evimaria Terzi. "A Platform for Data-Intensive Cybersecurity Monitoring." **Advanced Cybersecurity Center**, $75,000 in total and $50,000 for UMass Amherst.

- Yanlei Diao (PI), Li-Jun Ma, Samuel Madden, and Yiping Shen. "A Cloud Service for Massive Scale Genomic Data Processing and Deep Analysis." **Massachusetts Green High Performance Computing Center (MGHPCC) Seed Fund**, $144,000 and $94,000 for UMass Amherst.

- Yanlei Diao (PI). "Streaming Analytics in Large-Scale Monitoring Applications." **IBM Scalable Data Analytics for A Smarter Planet Innovation Award**. $20,000.

- Yanlei Diao (PI) and Andrew McGregor. "Data Analytics in the Cloud: Exact Answers Fast, Approximate Answers Faster." **NEC Labs Research Award**. $40,000.

- Yanlei Diao (PI) and Prashant Shenoy. "Streamlining Large-Scale Data Analysis: Exact Answers Fast, Approximate Answers Faster**." Google Research Award**, $50,000.

- Yanlei Diao (PI). "In-Network Event Processing over Distributed Streams." **Cisco Research Gift**, $89,916.

SOFTWARE

MASS-SCALLA 0.1 (http://scalla.cs.umass.edu/): A scalable, low-latency analytics platform that fundamentally transforms the batch-oriented MapReduce cluster computing paradigm into an incremental parallel processing paradigm, and further extends to near real-time analytics.

SASE 1.0 (http://sase.cs.umass.edu/): A stream processing system developed at the University of Massachusetts Amherst that provides fast pattern matching over event streams. The release of SASE 1.0 contains the core stream processing engine of the SASE system.

YFilter 1.0 (http://yfilter.cs.umass.edu/): A high-performance filtering system that allows users or applications to submit queries to be continuously executed against streaming XML messages. This release supports queries written in a subset of XPath 1.0.

INVITED TALKS

- "Scalable, Low-Latency Data Analytics and its Applications"

  Google, April 11 2014
  China East Normal University, China, June 7 2013
  Beihuang University, January 11, 2013
  INRIA Research Center, December 20, 2012
  Hong Kong University of Science and Technology, January 18, 2012
  Hong Kong Baptist University, January 18, 2012
  Fudan University, October 22, 2011

- "Career Paths in Databases"

  DataBase Mentoring (DB Me) Workshop in conjunction with SIGMOD 2010, June 11 2010

- "Addressing New Challenges in Data Stream Processing"

  University of Wisconsin Madison, September 1, 2010
  AT&T Research Center, August 3 2010
  IBM Watson Research Center, August 2 2010
  Microsoft Research Silicon Valley, May 11 2010
  Yahoo! Research, May 10 2010
  Portland State University, May 7 2010
  Massachusetts Institute of Technology, April 29 2010
  University of Pittsburgh and Carnegie Mellon University, April 19 2010
  University of California San Diego, April 14 2010
  University of California Los Angeles, November 17 2009
  University of California Irvine, November 16 2009
  University of California Berkeley, November 13 2009
  IBM Almaden Research Center, July 28 2009
  HP Labs, July 27 2009

- "An Optimized Index Structure for Flash Devices"

  Non-volatile Memories Workshop, April 13 2010

- "SASE+: Expressing and Processing Complex Event Patterns over Streams"

  Worcester Polytechnic Institute, Department Colloquium, April 25, 2008
  New England Database Day, Cambridge, MA, February 4, 2008
  Renming University and Beihuang University, Beijing, China, January 4, 2008
  Microsoft Research Center, Redmond, WA, October 12, 2007
  StreamBase Systems, Lexington, Massachusetts, February 16, 2007

- "SPIRE: Scalable Processing of RFID Event Streams"

  International Workshop on RFID Data Management, April 7, 2008

- "High-Performance Complex Event Processing over Streams"

Hong Kong Baptist University, Hong Kong, January 16, 2007
Fudan University, China, September 7, 2006

- "Query Processing for Large-Scale XML Message Brokering"

  Tsinghua University and Peking University, September 5-6, 2006
  Distinguished Faculty Lecture Series, University of Texas at Austin, December 12-14, 2005
  AT&T Research Lab, December 15, 2005

- "XML Filtering, Transformation, and Routing with YFilter"

  Cisco Systems, January 5, 2007
  Database Seminar, University of Wisconsin, Madison, December 17, 2004
  IBM Almaden Research Center, September 23, 2004
  Oracle Corporation, December 9, 2003
  BEA Systems, May 22, 2003

TEACHING

Instructor   **University of Massachusetts Amherst**

Information Systems (CMPSCI 445), Fall 2007, S2008, F2009, Fall 2012, Spring 2013, Spring 2015
Database Design and Implementation (CMPSCI 645), S2006 (with G. Miklau), S2007, S2009, S2011
Advanced Topics in Database Systems (CMPSCI 745), F2008, F2010, F2013
Hot Topics in Databases (CMPSCI 691TD), Spring 2009, S2010
Advanced Database Systems (CMPSCI 691LL), F2006
Independent Study Courses: CMPSCI 701, F2007, S2008; CMPSCI496, S2008, S2010

Graduate Student Instructor   **University of California, Berkeley**

Introduction to Database Systems (CS 186), Fall 2000

Teaching Assistant   **Hong Kong University of Science and Technology**

Database Management Systems (COMP231), Spring 2000
Principles of Database Design (COMP332), Fall 1999
Computer and Programming Fundamentals II (COMP103), Spring 1999
Computer and Programming Fundamentals I (COMP102), Fall 1998

ADVISING

- **Current PhD Students**
  Liping Peng
  Boduo Li (*finalist of the 2011 Facebook Graduate Fellowship*)
  Haopeng Zhao
  Abhishek Roy
  Yunmeng Ban (*recipient of 2014 Microsoft Research Graduate Women's Scholarship and 2014 Jim Gray Graduate Scholarship at UMass Amherst*)
  Wenzhao Liu

- **Past Students**
  Thanh Tran (PhD, 2013), Twitter.
      Thesis: *High-Performance Processing of Continuous Uncertain Data*.
  Edward Mazur (MS, 2011), Google.
      Thesis: *Towards Scalable One-Pass Analytics using MapReduce*.
  Richard Cocci (MS, 2008), Harvard Law School.
      Thesis: *Efficient Data Interpretation and Compression over RFID Streams*.
  Ravishankar Rajamony (MS, 2008), Goldman Sachs.
      Thesis: *Improved Memory Management in XML Data Stream Processing*.
  Daniel Gyllstrom (MS, 2007), UMass Amherst.
      Thesis: *On Supporting Kleene Closure over Event Streams*.

- **Graduate Student Researchers**
  Junghee Jo (Fall 2008-Spring 2009), Jagrati Agrawal (Spring 2007-Spring 2008),
  Hee-Jin Chae (Summer 2006),

- **Undergraduate Student Researchers**
  Rebecca Bryan (2013), Said Mastawi (2013), Yang Tang (2013), Sofya Vorotnikova (2012),
  DiHuynh (2012), Andrew Huang (Spring 2008), Jacob Mitchell (Fall 2009), Alex Jackson (Spring
  2010), Eugene Wu (Fall 2006)

- **Dissertation Committees**
  AlperOkcan (2014), Northeastern University
  Asterios Katsifodimos (2013), INRIA, France
  Desislava Petkova (2008), University of Massachusetts Amherst
  Xiaoyong Liu (2006), Jiwoon Jeon (2006), University of Massachusetts Amherst
  Mo Liu (2009-2010), Worcester Polytechnic Institute

- **Visiting Students**
  Zhao Cao (2008-2010), IBM Research, China
  Yanming Nie (2007-2008), Northwestern Polytechnical University, China