

Yuanyuan Tian

yuanyuan.t@gmail.com

<http://researcher.watson.ibm.com/researcher/view.php?person=us-ytian>

650 Harry Rd
San Jose, CA 95120
Phone: (734) 709 8497

Education

- Sep 2003 – Aug 2008 **University of Michigan** Ann Arbor, MI
Ph.D. in Computer Science & Engineering Advisor: Jignesh M. Patel
Thesis: Querying Graph Databases
Designed and developed a graph querying system to support various sophisticated analytics on graphs including graph matching and graph summarizations.
- Sep 2003 – Dec 2005 **University of Michigan** Ann Arbor, MI
M.S. in Computer Science GPA: 8.77 / 9.00
- Sep 1999 – Jul 2003 **Peking University** Beijing, China
B.S. in Computer Science & Technology GPA: 91.52 / 100

Professional Experience

- Oct 2008 – Present **IBM Research - Almaden** San Jose, CA
Technical Lead, Research Staff Member
Current Work:
 - Co-founder and a lead developer of the Wildfire system for hybrid transactional and analytical processing (HTAP), which has been released as the IBM Db2 Event Store product.
 - Leading a project on online machine learning model management.
 - Leading a project on supporting graph analytics inside relational databases.Past Work:
 - Led multiple SQL-for-Big-Data projects with some techniques incorporated into the IBM Db2 BIG SQL product.
 - Co-founder and a lead developer of a distributed machine learning system, which was open sourced as Apache SystemML.
 - Led multiple projects on graph processing and analytics with new technical results, and initiated customer and product interactions.
- May 2004 – Aug 2008 **University of Michigan** Ann Arbor, MI
Graduate Student Research Assistant, Advisor: Jignesh M. Patel
 - Invented novel and efficient methods for querying and mining graphs and sequences.
 - Designed and developed an efficient graph querying system that was applied to solve the core driving biological problems in the National Center for Integrative Biomedical Informatics (NCIBI).

Selected Awards

- 2018 **EDBT 2018 Best Paper Award** EDBT 2018
B. Hentschel, P. J. Haas, Y. Tian: Temporally-Biased Sampling for Online Model Management
- 2016 **Outstanding Technical Achievement Award** IBM Research - Almaden

For the technical contributions of her research on join algorithms for Big Data.

- | | | |
|------|---|--|
| 2016 | Eminence & Excellence Award
For her research contribution on Hybrid Transactional and Analytics Processing (HTAP). | IBM Almaden Research Center |
| 2015 | IBM A-Level Accomplishment
For the impact of her research on join algorithms for Big Data. | IBM Research - Almaden |
| 2015 | IBM A-Level Accomplishment
For her contributions to the Apache open source SystemML project. | IBM Research - Almaden |
| 2013 | High Value Patent Application Award
For her US and International patents on systems and methods for processing machine learning algorithms in a MapReduce environment | IBM Research - Almaden |
| 2012 | Eminence & Excellence Award
For the impact of her research on the IBM InfoSphere BigInsights product and her leadership in graph analytics research. | IBM Research - Almaden |
| 2011 | Eminence & Excellence Award
For the impact of her research on the IBM InfoSphere BigInsights product. | IBM Research - Almaden |
| 2010 | Winner of CS Workshop Innovation Proposals
For proposals with innovation and potential impact on business.
(One of the only two winners in CS department.) | IBM Research - Almaden |
| 2008 | Distinguished Achievement Award
Awarded to selected doctoral students with research excellence.
(The only recipient in CSE division.) | College of Engineering, University of Michigan |
| 2007 | 2nd Place, CSE Honor Competition
For the best research and presentations among CSE graduate students. | CSE Division, University of Michigan |
| 2007 | Rackham Predoctoral Fellowship
Awarded to selected doctoral students with outstanding dissertation research.
(The only recipient in CSE division.) | University of Michigan |
| 2003 | Rackham Graduate Fellowship
To selected incoming graduate students with outstanding academic records. | University of Michigan |
| 2002 | Yang-Wang Scholarship
To top CS students in senior year with outstanding academic achievements.
(The 1 st class scholarship.) | Peking University |
| 2001 | Xiyue Scholarship
To selected undergraduate students with outstanding academic achievements. | Peking University |
| 2000 | Legend Scholarship
To selected undergraduate students with outstanding academic achievements. | Peking University |

Professional Service

Program Committee Chair

- VLDB 2019 Workshop Track
- ICDE 2017 Demo Track
- CIKM 2013 Poster Track
- 3rd Workshop on Large Scale Network Analysis (LSNA 2014)
- 5th Workshop on Graph Data Management (GDM 2014)
- 2nd Workshop on Large Scale Network Analysis (LSNA 2013)
- 4th Workshop on Graph Data Management (GDM 2013)
- 1st Workshop on Large Scale Network Analysis (LSNA 2012)

Editorship:

- Section Editor, Encyclopedia on Big Data Technologies
- Associate Editor, PVLDB, 2018

Panelist:

- NSF Advisory Panel, 2016
- NSF Advisory Panel, 2013
- NSF Career Mentoring Panel, ICDE 2012

Program Committee Member: VLDB 2019, SIGMOD 2018, VLDB 2017, VLDB 2016 Industrial Track, TKDE 2016 Poster Track, VLDB 2015, ICDE 2014, WISE 2013, SIGMOD 2012, GDM 2012, VLDB 2011 Industrial Track, DBSocial 2011, GDM 2011, ICDE 2011, GDM 2010, VLDB 2009.

Journal Reviewer: VLDB Journal (2014, 2017), TODS (2013, 2015), Statistical Analysis and Data Mining (2009), Information System (2010, 2011, 2013), ACM Transactions on Intelligent Systems and Technology (2010), Distributed and Parallel Databases (2012).

Reviewer for Books: Data Processing Techniques in the Era of Big Data.

Reviewer for Research Grants: Research Grants Council (RGC) of Hong Kong (2010, 2011).

Reviewer for Awards: The NCWIT Award for Aspirations in Computing.

Publications

- **Book and Book Chapter**

Systems for Big Graph Analytics

D. Yan, Y. Tian, J. Cheng. *SpringerBriefs in Computer Science, Springer, 2017.*

Big Graph Analytics Platforms

D. Yan, Y. Bu, Y. Tian, A. Deshpande. *Foundations and Trends in Databases, Vol. 7: No. 1-2, pp 1-195, 2017.*

Interactive Graph Summarization

Y. Tian, J. M. Patel. *Book Chapter, Link Mining: Models, Algorithms and Applications, Springer, 2010.*

- **Encyclopedia Article**

Hybrid Systems Based on Traditional Database Extensions

Y. Tian. *Encyclopedia of Big Data Technologies*, 2018.

Wildfire: HTAP for Big Data

R. Barber, V. Raman, R. Sidle, Y. Tian, P. Tozun. *Encyclopedia of Big Data Technologies*, 2018.

- **Journal Article**

Building a Hybrid Warehouse: Efficient Joins Between Data Stored in HDFS and Enterprise Warehouse

Y. Tian, F. Ozcan, T. Zou, R. Goncalves, H. Pirahesh. *ACM Transactions on Database Systems (TODS)*, 2016. **(Invited as Best of EDBT 2015)**

SystemML's Optimizer: Plan Generation for Large-Scale Machine Learning Programs

M. Boehm, D. Burdick, A. Evfimievski, B. Reinwald, F. R. Reiss, P. Sen, S. Tatikonda, Y. Tian. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 37(3), 2014.

Hybrid Parallelization Strategies for Large Scale Machine Learning in SystemML

M. Boehm, S. Tatikonda, B. Reinwald, P. Sen, Y. Tian, D. R. Burdick, S. Vaithyanathan. *The Proceedings of the VLDB Endowment (PVLDB)*, 7(7), 2014.

From "Think Like a Vertex" to "Think Like a Graph"

Y. Tian, A. Balmin, S. A. Corsten, S. Tatikonda, J. McPherson. *The Proceedings of the VLDB Endowment (PVLDB)*, 7(3), 2013. **(Best of VLDB 2014)**

A Platform for eXtreme Analytics

A. Balmin, K. Beyer, V. Ercegovic, J. McPherson, F. Ozcan, H. Pirahesh, E. Shekita, Y. Sismanis, S. Tata, Y. Tian. *IBM Journal of Research and Development*, 57(3/4), 2013.

CoHadoop: Flexible Data Placement and Its Exploitation in Hadoop

M. Eltabakh, Y. Tian, F. Ozcan, R. Gemulla, A. Krettek, J. McPherson. *The Proceedings of the VLDB Endowment (PVLDB)*, 2011.

Michigan Molecular Interactions R2: From Interacting Proteins to Pathways

G. Tarcea, T. Weymouth, A. Ade, A. Bookvich, J. Gao, V. Mahavisno, Z. Wright, A. Chapman, M. Jayapandian, A. Ozgur, Y. Tian, J. Cavalcoli, B. Mirel, J. Patel, D. Radev, B. Athey, D. States and H. V. Jagadish. *Nucleic Acids Research* 37 (Database issue):D642-6, 2009.

Periscope/GQ: A Graph Querying Toolkit

Y. Tian, J. M. Patel, V. Nair, S. Martini, M. Kretzler. *The Proceedings of the VLDB Endowment (PVLDB)*, 1(2), 2008.

SAGA: A Subgraph Matching Tool for Biological Graphs

Y. Tian, R. C. McEachin, C. Santos, D. J. States, J. M. Patel. *Bioinformatics Journal*, 23(2): 232-239, 2007.

Practical Methods for Constructing Suffix Trees

Y. Tian, S. Tata, R. A. Hankins, J. M. Patel. *Very Large Data Base Journal (VLDBJ)*, 14(3): 281-299, 2005.

- **Conference Paper**

Temporally-Biased Sampling for Online Model Management

B. Hentschel, P. J. Haas, Y. Tian. *The 21st International Conference on Extending Database Technology (EDBT)*, 2018. **(EDBT 2018 Best Paper Award)**

Hybrid Transactional/Analytical Processing: A Survey

F. Özcan, Y. Tian, P. Tözün. *The 37th ACM SIGMOD International Conference on Management of Data (SIGMOD), Tutorial*, 2017.

Evolving Databases for New-Gen Big Data Applications

C. Garcia-Arellano, R. Barber, M. Huras, R. Grosman, C. Mohan, R. Mueller, F. Özcan, H. Pirahesh, V. Raman, R. Sidle, A. Storm, Y. Tian, P. Tözün, D. Zilio, G. Lohman. *The 2017 biennial Conference on Innovative Data Systems Research (CIDR)*, 2017.

Wildfire: Concurrent Blazing Data Ingest and Analytics

R. Barber, M. Huras, G. Lohman, C. Mohan, R. Mueller, F. Özcan, H. Pirahesh, V. Raman, R. Sidle, O. Sidorkin, A. Storm, Y. Tian, P. Tözün. *The 36th ACM SIGMOD International Conference on Management of Data (SIGMOD), Demo*, 2016.

Big Graph Analytics Platforms

D. Yan, Y. Bu, Y. Tian, A. Deshpande, J. Cheng. *The 36th ACM SIGMOD International Conference on Management of Data (SIGMOD), Tutorial*, 2016.

Resource Elasticity for Large-Scale Machine Learning

B. Huang, M. Boehm, Y. Tian, B. Reinwald, S. Tatikonda, F. R. Reiss. *The 35th ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2015.

Joins for Hybrid Warehouses: Exploiting Massive Parallelism in Hadoop and Enterprise Data Warehouses

Y. Tian, T. Zou, F. Ozcan, R. Goncalves, H. Pirahesh. *The 18th International Conference on Extending Database Technology (EDBT)*, 2015. **(Best of EDBT 2015)**

A Generic Solution to Integrate SQL and Analytics for Big Data

N. R. Katsipoulakis, Y. Tian, F. Ozcan, B. Reinwald, H. Pirahesh. *The 18th International Conference on Extending Database Technology (EDBT), Vision Paper*, 2015.

Dynamic Interaction Graphs with Probabilistic Edge Decay

W. Xie, Y. Tian, Y. Sismanis, A. Balmin, P. J. Haas. *The 31st International Conference on Data Engineering (ICDE)*, 2015.

Distributed Graph Summarization

X. Liu, Y. Tian, Q. He, W. Lee, J. McPherson. *The 23rd ACM International Conference on Information and Knowledge Management (CIKM)*, 2014.

Scalable Topic-Specific Influence Analysis on Microblogs

B. Bi, Y. Tian, Y. Sismanis, A. Balmin, J. Cho. *The 7th ACM Conference on Web Search and Data Mining (WSDM)*, 2014.

Compiling Machine Learning Algorithms with SystemML

M. Boehm, D. Burdick, A. Evfimievski, B. Reinwald, P. Sen, S. Tatikonda, **Y. Tian**. *ACM Symposium on Cloud Computing (SoCC), Poster, 2013.*

Event-based Social Networks: Linking the Online and Offline Social Worlds

X. Liu, Q. He, **Y. Tian**, W. Lee, J. McPherson, J. Han. *The 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2012.*

Scalable and Numerically Stable Descriptive Statistics In SystemML

Y. Tian, S. Tatikonda, B. Reinwald. *The 28th International Conference on Data Engineering (ICDE), 2012.*

SystemML: Declarative Machine Learning on MapReduce

A. Ghoting, R. Krishnamurthy, E. Pednault, B. Reinwald, V. Sindhwani, S. Tatikonda, **Y. Tian**, S. Vaithyanathan. *The 27th International Conference on Data Engineering (ICDE), 2011.*

A Comparison of Join Algorithms for Log Processing in MapReduce

S. Blanas, J. M. Patel, V. Ercegovic, J. Rao, E. J. Shekita, **Y. Tian**. *The 30th ACM International Conference on Management of Data (SIGMOD), 2010.*

Discovery-Driven Graph Summarization

N. Zhang, **Y. Tian**, J. M. Patel. *The 26th International Conference on Data Engineering (ICDE), 2010.*

Efficient Aggregation for Graph Summarization

Y. Tian, R. A. Hankins, J. M. Patel. *The 28th ACM International Conference on Management of Data (SIGMOD), 2008.*

TALE: A Tool for Approximate Large Graph Matching

Y. Tian, J. M. Patel. *The 24th International Conference on Data Engineering (ICDE), 2008.*

Patents

Sparsity-Driven Matrix Representation to Optimize Operational and Storage Efficiency

B. Reinwald, S. Tatikonda, **Y. Tian**. *US Patent 9,454,472.*

Identifying Influencers for Topics in Social Media

A. Balmin, B. Bi, Y. Sismanis, **Y. Tian**. *US Patent 9,449,096.*

Subgraph-Based Distributed Graph Processing

A. Balmin, S. A. Corsten, J. McPherson, S. Tatikonda, **Y. Tian**. *US Patent 9,400,767.*

Hybrid Parallelization Strategies for Machine Learning Programs on Top of MapReduce

M. Boehm, D. Burdick, B. Reinwald, P. Sen, S. Tatikonda, **Y. Tian**, S. Vaithyanathan. *US Patent 9,286,044.*

Systems and Methods for Processing Machine Learning Algorithms in a MapReduce Environment

D. R. Burdick, A. Ghoting, R. Krishnamurthy, E. P. Pednault, B. Reinwald, V. Sindhvani, S. Tatikonda, **Y. Tian**, S. Vaithyanathan. *US Patent 8,612,368*.

Invited Talks

- Aug 2017 **Big Data Analytics: From SQL to Machine Learning and Graph Analysis (Keynote)**
BigDas Workshop, SIGKDD'2017, August 2017
- May 2017 **Hybrid Transactional/Analytical Processing (Tutorial)**
SIGMOD 2017
- Jun 2016 **Big Graph Analytics Platforms (Tutorial)**
SIGMOD 2016
- Nov 2013 **Giraph++: From "Think Like a Vertex" to "Think Like a Graph"**
Facebook, Menlo Park, CA
- May 2013 **Large Scale Topic-specific Influence Analysis on Microblogs**
CS Department, UC Santa Barbara, CA
- May 2013 **Large Scale Topic-specific Influence Analysis on Microblogs**
Database Group, UC Santa Cruz, CA
- Aug 2012 **SystemML: Declarative Machine Learning on MapReduce**
Database Group, Department of Computer Science and Technology, Peking University, Beijing, China
- Aug 2012 **SystemML: Declarative Machine Learning on MapReduce**
IBM China Research Lab, Beijing, China
- Apr 2012 **SystemML: Declarative Machine Learning on MapReduce**
Database Group, Department of Computer Science, University of Maryland, College Park, MD
- Oct 2007 **Querying Graph Databases**
CSE Honor Competition, CSE Division, University of Michigan, Ann Arbor, MI
- Aug 2007 **Towards Summarizing and Understanding Large Graphs**
Nokia Research Center, Palo Alto, CA
- Apr 2007 **SAGA+TALE: Fast and Flexible Graph Matching Tools**
National Center for Integrative Biomedical Informatics (NCIBI), Ann Arbor, MI
- Jan 2006 **SAGA: A Fast and Flexible Graph Matching Tool**
National Center for Integrative Biomedical Informatics (NCIBI), Ann Arbor, MI

Other Professional Experience

Jun 2007 - Aug 2007 **Nokia Research Center** Palo Alto, CA

Research Intern

- Applied graph mining methods to analyze call logs.
- Researched on methods for summarizing large social networks.

Jan 2005 - Apr 2005

University of Michigan

Ann Arbor, MI

Graduate Student Instructor

- Led the weekly 1-hour discussion section of EECS 485: Web Databases.
- Held office hours to help students with lecture material and projects.
- Designed and graded course projects.

May 2002 - Jun 2003

Peking University

Beijing, China

Undergraduate Student Research Assistant

- Worked on the web database design for a departmental online information system.

Jul 2002 - Mar 2003

Bright Ocean Corporation

Beijing, China

Part-time Intern

- Designed and developed data visualization tools used in China Mobile Corporation's Decision Support System for telecommunication network operation, administration and maintenance.