

Stanford



Matei Zaharia

Associate Professor of Computer Science and, by courtesy, of Electrical Engineering

Bio

BIO

Homepage: <https://cs.stanford.edu/~matei/>

ACADEMIC APPOINTMENTS

- Associate Professor, Computer Science
- Associate Professor (By courtesy), Electrical Engineering

LINKS

- Matei Zaharia's Homepage: <https://cs.stanford.edu/~matei/>

Teaching

COURSES

2021-22

- Machine Learning Systems Seminar: CS 528 (Aut, Win, Spr)
- Principles of Data-Intensive Systems: CS 245 (Win)
- Value of Data and AI: CS 320 (Win)

2020-21

- Principles of Data-Intensive Systems: CS 245 (Win)
- Value of Data and AI: CS 320 (Win)

2019-20

- Principles of Data-Intensive Systems: CS 245 (Win)
- Value of Data and AI: CS 320 (Win)

STANFORD ADVISEES

Doctoral Dissertation Reader (AC)

Qian Li

Doctoral Dissertation Co-Advisor (AC)

Lingjiao Chen, Omar Khattab

Master's Program Advisor

Victoria DiMelis, Rishab Gargeya, Kai Kato, Max Sobol Mark, Megan Worrel, Emily Yang, Eric Zhang

Doctoral (Program)

Jared Davis, Trevor Gale, Peter Kraft, Liana Patel, Deepti Raghavan, Keshav Santhanam, Gina Yuan

Publications

PUBLICATIONS

- **Machine Learned Cellular Phenotypes Predict Outcome in Ischemic Cardiomyopathy.** *Circulation research*
Rogers, A. J., Selvalingam, A., Alhusseini, M. I., Krummen, D. E., Corrado, C., Abuzaid, F., Baykaner, T., Meyer, C., Clopton, P., Giles, W. R., Bailis, P., Niederer, S. A., Wang, et al
2020
- **DIFF: a relational interface for large-scale data explanation** *VLDB JOURNAL*
Abuzaid, F., Kraft, P., Suri, S., Gan, E., Xu, E., Shenoy, A., Ananthanarayan, A., Sheu, J., Meijer, E., Wu, X., Naughton, J., Bailis, P., Zaharia, et al
2020
- **Machine Learning to Classify Intracardiac Electrical Patterns during Atrial Fibrillation.** *Circulation. Arrhythmia and electrophysiology*
Alhusseini, M. I., Abuzaid, F., Rogers, A. J., Zaman, J. A., Baykaner, T., Clopton, P., Bailis, P., Zaharia, M., Wang, P. J., Rappel, W., Narayan, S. M.
2020
- **Approximate Selection with Guarantees using Proxies** *PROCEEDINGS OF THE VLDB ENDOWMENT*
Kang, D., Gan, E., Bailis, P., Hashimoto, T., Zaharia, M.
2020; 13 (11): 1990–2003
- **PREDICTING SUDDEN CARDIAC DEATH BY MACHINE LEARNING OF VENTRICULAR ACTION POTENTIALS**
Selvalingam, A., Alhusseini, M., Rogers, A. J., Krummen, D., Abuzaid, F. M., Baykaner, T., Clopton, P., Bailis, P., Zaharia, M., Wang, P., Narayan, S.
ELSEVIER SCIENCE INC.2020: 427
- **Fleet: A Framework for Massively Parallel Streaming on FPGAs**
Thomas, J., Hanrahan, P., Zaharia, M., ACM
ASSOC COMPUTING MACHINERY.2020: 639–51
- **Blazelt: Optimizing Declarative Aggregation and Limit Queries for Neural Network-Based Video Analytics** *PROCEEDINGS OF THE VLDB ENDOWMENT*
Kang, D., Bailis, P., Zaharia, M.
2019; 13 (4): 533–46
- **To Index or Not to Index: Optimizing Exact Maximum Inner Product Search**
Abuzaid, F., Sethi, G., Bailis, P., Zaharia, M., IEEE
IEEE.2019: 1250–61
- **Optimizing Data-Intensive Computations in Existing Libraries with Split Annotations**
Palkar, S., Zaharia, M., ACM
ASSOC COMPUTING MACHINERY.2019: 291–305
- **PipeDream: Generalized Pipeline Parallelism for DNN Training**
Narayanan, D., Harlap, A., Phanishayee, A., Seshadri, V., Devanur, N. R., Ganger, G. R., Gibbons, P. B., Zaharia, M., ACM
ASSOC COMPUTING MACHINERY.2019: 1–15
- **TASO: Optimizing Deep Learning Computation with Automatic Generation of Graph Substitutions**
Jia, Z., Padon, O., Thomas, J., Warszawski, T., Zaharia, M., Aiken, A., ACM
ASSOC COMPUTING MACHINERY.2019: 47–62
- **From Laptop to Lambda: Outsourcing Everyday Jobs to Thousands of Transient Functional Containers**
Fouladi, S., Romero, F., Iyer, D., Li, Q., Chatterjee, S., Kozyrakis, C., Zaharia, M., Winstein, K., USENIX Assoc
USENIX ASSOC.2019: 475–88
- **DIFF: A Relational Interface for Large-Scale Data Explanation** *PROCEEDINGS OF THE VLDB ENDOWMENT*
Abuzaid, F., Kraft, P., Suri, S., Gan, E., Xu, E., Shenoy, A., Ananthanarayan, A., Sheu, J., Meijer, E., Wu, X., Naughton, J., Bailis, P., Zaharia, et al

2018; 12 (4): 419–32

- **Structured Streaming: A Declarative API for Real-Time Applications in Apache Spark**
Armbrust, M., Das, T., Torres, J., Yavuz, B., Zhu, S., Xin, R., Ghodsi, A., Stoica, I., Zaharia, M., Das, G., Jermaine, C., Bernstein, P., Eldawy, et al
ASSOC COMPUTING MACHINERY.2018: 601–13
- **MISTIQUE: A System to Store and Query Model Intermediates for Model Diagnosis**
Vartak, M., da Trindade, J. F., Madden, S., Zaharia, M., Das, G., Jermaine, C., Bernstein, P., Eldawy, A.
ASSOC COMPUTING MACHINERY.2018: 1285–1300
- **NoScope: Optimizing Neural Network Queries over Video at Scale** *PROCEEDINGS OF THE VLDB ENDOWMENT*
Kang, D., Emmons, J., Abuzaid, F., Bailis, P., Zaharia, M.
2017; 10 (11): 1586–97
- **Splinter: Practical Private Queries on Public Data**
Wang, F., Yun, C., Goldwasser, S., Vaikuntanathan, V., Zaharia, M., USENIX Assoc
USENIX ASSOC.2017: 299–313
- **DIY Hosting for Online Privacy**
Palkar, S., Zaharia, M., Assoc Comp Machinery
ASSOC COMPUTING MACHINERY.2017: 1–7
- **Making Caches Work for Graph Analytics**
Zhang, Y., Kiriansky, V., Mendis, C., Amarasinghe, S., Zaharia, M., Nie, J. Y., Obradovic, Z., Suzumura, T., Ghosh, R., Nambiar, R., Wang, C., Zang, H., BaezaYates, et al
IEEE.2017: 293–302
- **Apache Spark: A Unified Engine for Big Data Processing** *COMMUNICATIONS OF THE ACM*
Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, et al
2016; 59 (11): 56–65
- **Voodoo - A Vector Algebra for Portable Database Performance on Modern Hardware** *PROCEEDINGS OF THE VLDB ENDOWMENT*
Pirk, H., Moll, O., Zaharia, M., Madden, S.
2016; 9 (14): 1707–18
- **MLlib: Machine Learning in Apache Spark** *JOURNAL OF MACHINE LEARNING RESEARCH*
Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D. B., Amde, M., Owen, S., Xin, D., Xin, R., Franklin, et al
2016; 17
- **GraphFrames: An Integrated API for Mixing Graph and Relational Queries**
Dave, A., Jindal, A., Li, L., Xin, R., Gonzalez, J., Zaharia, M., ACM
ASSOC COMPUTING MACHINERY.2016
- **FairRide: Near-Optimal, Fair Cache Sharing**
Pu, Q., Li, H., Zaharia, M., Ghodsi, A., Stoica, I., USENIX Assoc
USENIX ASSOC.2016: 393–406
- **SparkR: Scaling R Programs with Spark**
Venkataraman, S., Yang, Z., Liu, D., Liang, E., Falaki, H., Meng, X., Xin, R., Ghodsi, A., Franklin, M., Stoica, I., Zaharia, M., ACM SIGMOD
ASSOC COMPUTING MACHINERY.2016: 1099–1104
- **Introduction to Spark 2.0 for Database Researchers**
Armbrust, M., Bateman, D., Xin, R., Zaharia, M., ACM SIGMOD
ASSOC COMPUTING MACHINERY.2016: 2193–94
- **Yggdrasil: An Optimized System for Training Deep Decision Trees at Scale**
Abuzaid, F., Bradley, J., Liang, F., Feng, A., Yang, L., Zaharia, M., Talwalkar, A., Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, Garnett, R.
NEURAL INFORMATION PROCESSING SYSTEMS (NIPS).2016
- **Matrix Computations and Optimization in Apache Spark**

Zadeh, R., Meng, X., Ulanov, A., Yavuz, B., Pu, L., Venkataraman, S., Sparks, E., Staple, A., Zaharia, M., Assoc Comp Machinery
ASSOC COMPUTING MACHINERY.2016: 31–38

- **Scaling Spark in the Real World: Performance and Usability** *PROCEEDINGS OF THE VLDB ENDOWMENT*
Armbrust, M., Das, T., Davidson, A., Ghodsi, A., Or, A., Rosen, J., Stoica, I., Wendell, P., Xin, R., Zaharia, M.
2015; 8 (12): 1840–43
- **Vuvuzela: Scalable Private Messaging Resistant to Traffic Analysis**
van den Hooff, J., Lazar, D., Zaharia, M., Zeldovich, N., Assoc Comp Machinery
ASSOC COMPUTING MACHINERY.2015: 137–52
- **Spark SQL: Relational Data Processing in Spark**
Armbrust, M., Xin, R. S., Lian, C., Huai, Y., Liu, D., Bradley, J. K., Meng, X., Kaftan, T., Franklinton, M. J., Ghodsi, A., Zaharia, M., ACM SIGMOD
ASSOC COMPUTING MACHINERY.2015: 1383–94
- **Optimally designing games for behavioural research** *PROCEEDINGS OF THE ROYAL SOCIETY A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES*
Rafferty, A. N., Zaharia, M., Griffiths, T. L.
2014; 470 (2167): 20130828
- **A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples** *GENOME RESEARCH*
Naccache, S. N., Federman, S., Veeraraghavan, N., Zaharia, M., Lee, D., Samayoa, E., Bouquet, J., Greninger, A. L., Luk, K., Enge, B., Wadford, D. A., Messenger, S. L., Genrich, et al
2014; 24 (7): 1180–92
- **Multi-Resource Fair Queueing for Packet Processing** *ACM SIGCOMM COMPUTER COMMUNICATION REVIEW*
Ghodsi, A., Sekar, V., Zaharia, M., Stoica, I.
2012; 42 (4): 1–12
- **Managing Data Transfers in Computer Clusters with Orchestra** *ACM SIGCOMM COMPUTER COMMUNICATION REVIEW*
Chowdhury, M., Zaharia, M., Ma, J., Jordan, M. I., Stoica, I.
2011; 41 (4): 98–109