

Matthias Boehm

Graz University of Technology
Computer Science and Biomedical Engineering
Institute of Interactive Systems and Data Science
BMK endowed chair for Data Management

Office: 8010 Graz, Inffeldgasse 13/V, PZ 205 014
m.boehm@tugraz.at



Matthias Boehm is a BMK-endowed professor for data management at Graz University of Technology (<https://www.tugraz.at/home/>), Austria, and a research area manager for data management at the colocated Know-Center GmbH (<https://www.know-center.tugraz.at/en/>), Austria. Prior to joining TU Graz in 2018, he was a research staff member at IBM Research - Almaden (<https://www.research.ibm.com/labs/almaden/>), CA, USA, with a major focus on compilation and runtime techniques for declarative, large-scale machine learning in Apache SystemML (<https://systemml.apache.org/>). Matthias received his Ph.D. from Dresden University of Technology (<https://wwwdb.inf.tu-dresden.de/>), Germany in 2011 with a dissertation on cost-based optimization of integration flows. His previous research also includes systems support for time series forecasting as well as in-memory indexing and query processing. Matthias is a recipient of the 2016 VLDB Best Paper Award, a 2016 SIGMOD Research Highlight Award, and a 2016 IBM Pat Goldberg Memorial Best Paper Award.

Current Projects: Apache SystemDS (<https://github.com/apache/systemml>) (An open source ML system for the end-to-end data science lifecycle), ExDRa (<https://www.exdra.de/>) (exploratory data science and federated ML over raw data, w/ Siemens, DFKI, and TU Berlin)

Team

The DAMSLab (<https://damslab.github.io/>) (data management for data science laboratory) is a cross-organizational research group uniting the data management group of TU Graz and the research area data management of the colocated Know-Center.

We're looking for motivated PhD, master, and bachelor students to join our team. Our research focuses on building ML systems and tools for simplifying the data science lifecycle – from data integration over model training to deployment and scoring – via high-level language abstractions and specialized compiler and runtime techniques. If you're interested, please contact me directly via email.

Open Bachelor/Master Thesis Topics ([./thesis_topics.htm](#))

Publications

This publication list covers the last six years. For a full list see DBLP (https://dblp.uni-trier.de/pers/hd/b/Boehm_0001:Matthias) and Google Scholar (<https://scholar.google.com/citations?user=jrPG2AUAAAAJ&hl=en>).

2020

- Matthias Boehm, Iulian Antonov, Sebastian Baunsgaard, Mark Dokter, Robert Ginthör, Kevin Innerebner, Florijan Klezin, Stefanie Lindstaedt, Arnab Phani, Benjamin Rath, Berthold Reinwald, Shafaq Siddiqi, Sebastian Benjamin Wrede: SystemDS: A Declarative Machine Learning System for the End-to-End Data Science Lifecycle **CIDR 2020**. [paper, (resources/cidr2020.pdf)slides (resources/cidr2020_talk.pptx)]

2019

- Johanna Sommer, Matthias Boehm, Alexandre V. Evfimievski, Berthold Reinwald, Peter J. Haas: MNC: Structure-Exploiting Sparsity Estimation for Matrix Expressions. **SIGMOD 2019**. [paper (resources/sigmod2019.pdf), slides (resources/sigmod2019_talk.pptx), poster (resources/sigmod2019_poster.pdf)]
- Ahmed Elgohary, Matthias Boehm, Peter J. Haas, Frederick R. Reiss, Berthold Reinwald: Compressed Linear Algebra for Large-Scale Machine Learning. **Commun. ACM 2019** 62(5). [paper (resources/cacm2019.pdf), Link (<https://dl.acm.org/citation.cfm?doid=3328504.3318221>)]
- Matthias Boehm, Arun Kumar, Jun Yang: Data Management in Machine Learning Systems. Synthesis Lectures on Data Management 11 (1), **Morgan & Claypool Publishers 2019**. [book (<https://www.morganclaypool.com/doi/10.2200/S00895ED1V01Y201901DTM057>)]
- Matthias Boehm, Alexandre V. Evfimievski, Berthold Reinwald: Efficient Data-Parallel Cumulative Aggregates for Large-Scale Machine Learning. **BTW 2019**. [paper (resources/btw2019.pdf), slides (resources/btw2019_talk.pptx)]

2018

- Matthias Boehm, Berthold Reinwald, Dylan Hutchison, Prithviraj Sen, Alexandre V. Evfimievski, Niketan Pansare: On Optimizing Operator Fusion Plans for Large-Scale Machine Learning in SystemML. **PVLDB 2018** 11(12). [paper (resources/vldb2018.pdf), slides (resources/vldb2018_talk.pptx), poster (resources/vldb2018_poster.pdf)]
- Ahmed Elgohary, Matthias Boehm, Peter J. Haas, Frederick R. Reiss, Berthold Reinwald: Compressed Linear Algebra for Large-Scale Machine Learning. **VLDB Journal 2018** 27(5). [paper (resources/vldb2018.pdf), link (<https://link.springer.com/article/10.1007%2Fs00778-017-0478-1>)]
- Matthias Boehm: Apache SystemML – Declarative Large-Scale Machine Learning. **Encyclopedia of Big Data Technologies 2018**. [paper (resources/encyclopedia2018.pdf)]
- Niketan Pansare, Michael Dusenberry, Nakul Jindal, Matthias Boehm, Berthold Reinwald, Prithviraj Sen: Deep Learning with Apache SystemML. **SysML 2018**. [paper (resources/sysml2018.pdf)]

2017

- Arun Kumar, Matthias Boehm, Jun Yang: Data Management in Machine Learning: Challenges, Techniques, and Systems. **SIGMOD 2017**. [paper (resources/sigmod2017.pdf), slides (resources/sigmod2017_tutorial.pdf), video (https://www.youtube.com/watch?v=U8J0Dd_Z5wo)]
- Ahmed Elgohary, Matthias Boehm, Peter J. Haas, Frederick R. Reiss, Berthold Reinwald: Scaling Machine Learning via Compressed Linear Algebra. **SIGMOD Record 2017** 46(1). [paper (resources/sigmodrec2017.pdf)]
- Tarek Elgamal, Shangyu Luo, Matthias Boehm, Alexandre V. Evfimievski, Shirish Tatikonda, Berthold Reinwald, Prithviraj Sen: SPOOF: Sum-Product Optimization and Operator Fusion for Large-Scale Machine Learning. **CIDR 2017**. [paper (resources/cidr2017.pdf), slides (resources/cidr2017_talk.pptx)]

2016

- Ahmed Elgohary, Matthias Boehm, Peter J. Haas, Frederick R. Reiss, Berthold Reinwald: Compressed Linear Algebra for Large-Scale Machine Learning. **PVLDB 2016** 9(12). [paper (resources/vldb2016a.pdf), slides (resources/vldb2016a_talk.pptx), poster (resources/vldb2016a_poster.pdf)]
- Matthias Boehm, Michael Dusenberry, Deron Eriksson, Alexandre V. Evfimievski, Faraz Makari Manshadi, Niketan Pansare, Berthold Reinwald, Frederick Reiss, Prithviraj Sen, Arvind Surve, Shirish Tatikonda: SystemML: Declarative Machine Learning on Spark. **PVLDB 2016** 9(13). [paper (resources/vldb2016b.pdf), slides (resources/vldb2016b_talk.pptx)]
- Matthias Boehm, Alexandre V. Evfimievski, Niketan Pansare, Berthold Reinwald: Declarative Machine Learning - A Classification of Basic Properties and Types. **CoRR 2016** abs/1605.05826. [paper (resources/corr2016.pdf)]

2015

- Arash Ashari, Shirish Tatikonda, Matthias Boehm, Berthold Reinwald, Keith Campbell, John Keenleyside, P. Sadayappan: On Optimizing Machine Learning Workloads via Kernel Fusion. **PPOPP 2015**. [paper (resources/ppopp2015.pdf)]
- Botong Huang, Matthias Boehm, Yuanyuan Tian, Berthold Reinwald, Shirish Tatikonda, Frederick R. Reiss: Resource Elasticity for Large-Scale Machine Learning. **SIGMOD 2015**. [paper (resources/sigmod2015.pdf), slides (resources/sigmod2015_talk.pptx), poster (resources/sigmod2015_poster.pdf)]
- Matthias Boehm: Costing Generated Runtime Execution Plans for Large-Scale Machine Learning Programs. **CoRR 2015** abs/1503.06384. [paper (resources/corr2015.pdf)]

2014

- Matthias Boehm, Douglas R. Burdick, Alexandre V. Evfimievski, Berthold Reinwald, Frederick R. Reiss, Prithviraj Sen, Shirish Tatikonda, Yuanyuan Tian: SystemML's Optimizer: Plan Generation for Large-Scale Machine Learning Programs. **IEEE Data Eng. Bull. 2014** 37(3). [paper (resources/debull2014.pdf)]
- Matthias Boehm, Dirk Habich, Wolfgang Lehner: On-Demand Re-Optimization of Integration Flows. **Inf. Syst. 2014** 45. [paper (resources/jinfsys2014.pdf)]
- Peter D. Kirchner, Matthias Boehm, Berthold Reinwald, Daby M. Sow, J. Michael Schmidt, Deepak S. Turaga, Alain Biem: Large Scale Discriminative Metric Learning. IPDPS Workshop **ParLearning 2014**. [paper (resources/parlearning2014.pdf), slides (resources/parlearning2014_talk.pptx)]
- Matthias Boehm, Shirish Tatikonda, Berthold Reinwald, Prithviraj Sen, Yuanyuan Tian, Douglas Burdick, Shivakumar Vaithyanathan: Hybrid Parallelization Strategies for Large-Scale Machine Learning in SystemML. **PVLDB 2014** 7(7). [paper (resources/vldb2014.pdf), slides (resources/vldb2014_talk.pptx), poster (resources/vldb2014_poster.pdf)]

Teaching

Summer 2020

- Data Management (INF.01017UF (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=228955&pSpracheNr=2&pMUISuche=FALSE), INF.02018UF (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=228953&pSpracheNr=2&pMUISuche=FALSE), 706.010 (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=229611&pSpracheNr=2&pMUISuche=FALSE)). [course website (teaching/ss20_dbs/index.htm)]
- Architecture of Machine Learning Systems (706.550 (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=225081&pSpracheNr=2&pMUISuche=FALSE)). [course website (teaching/ss20_amls/index.htm)]

Winter 2019/20

- Data Management (INF.01017UF (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=229612&pSpracheNr=2&pMUISuche=FALSE), INF.02018UF (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=229613&pSpracheNr=2&pMUISuche=FALSE), 706.010 (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=229551&pSpracheNr=2&pMUISuche=FALSE)). [course website (teaching/ws1920_dbs/index.htm)]
- Data Integration and Large-Scale Analysis (706.520 (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=227337&pSpracheNr=2&pMUISuche=FALSE)). [course website (teaching/ws1920_dia/index.htm)]
- Architecture of Database Systems (706.543 (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=221921&pSpracheNr=2&pMUISuche=FALSE)). [course website (teaching/ws1920_adbs/index.htm)]

Summer 2019

- Database Systems (INF.01014UF (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=215511&pSpracheNr=2&pMUISuche=FALSE), 706.004 (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=211614&pSpracheNr=2&pMUISuche=FALSE)). [course website (teaching/ss19_dbs/index.htm)]
- Architecture of Machine Learning Systems (706.550 (https://online.tugraz.at/tug_online/wbLv.wbShowLVDetail?pStpSpNr=219917&pSpracheNr=2&pMUISuche=FALSE)). [course website (teaching/ss19_aml/index.htm)]

Service

This list summarizes PC memberships and review activities, again of the last six years.

- Track Chair (Data Science) BTW 2021, GI working group Data Science (since 03/2020)
- Program Committee SIGMOD 2021, SIGMOD 2021 Industry, CIDR 2021
- Program Committee SIGMOD 2020, PVLDB 2020, ICDE 2020, CIDR 2020, DEEM 2020, PVLDB PhD 2020; Journal Reviewer VLDBJ 2020
- Program Committee SIGMOD 2019, PVLDB 2019, ICDE 2019, EDBT 2019, DEEM 2019, AIDB 2019; Journal Reviewer TKDE 2019
- Program Committee PVLDB 2018, EDBT 2018 Industry, DEEM 2018, WebDB 2018, EBDVF 2018
- Program Committee ICDE 2017 Demo, DEEM 2017; Journal Reviewer SIGMOD Record 2017/18
- Journal Reviewer TKDE 2016/17, ACM Computing Surveys 2016, IBM Journal R&D 2016; External Reviewer CIKM 2016
- Program Committee SSDBM 2015; Journal Reviewer Information Systems 2015; External Reviewer SIGMOD Record 2015

Acknowledgements

Our research group is grateful for funding support from BMVIT, the BMVIT/FFG program "ICT of the Future", TU Graz, AVL LIST, Infineon Technologies Austria, Magna Steyr Fahrzeugtechnik, voestalpine Stahl Donawitz, and Know-Center.