Sebastian Baunsgaard Email: seb@bgaard.dk - baunsgaard@tu-berlin.de

Homepage: **baunsgaard.github.io** | Github: **github.com/baunsgaard** | Publications: **scholar.google.com**

PROFILE

I am a PhD student at TU Berlin, supervised by Matthias Boehm, working in the cross-section of **data management** and **machine learning systems**. I have finished my thesis and will defend August 2025. My research focuses on exploiting data structures in programs via workload-aware compression that can execute linear algebra operations directly on compressed formats without decompression.

I frequently contribute to and mentor others in **open source** projects and am a PMC member in Apache SystemDS with over 800 personal open-source commits and 50 mentored projects in primarily Java, Python, and an R-like declarative language DML. My contributions range from high-level compilation techniques to low-level tight-loop optimizations of kernels. Of particular note is my setup of automated testing infrastructure in SystemDS that executes the > 500'000 SystemDS tests in the cloud in less than 30 minutes.

SKILLS

Ordered by proficiency

- Java, Python
- Torch, TensorFlow, Spark, Hadoop
- Linux
- Scala, Jekyll, JS
- GoLang

ACADEMIC EXPERIENCE

BERLIN UNIVERSITY OF TECHNOLOGY — Ph.D. Student — August 2022 - Present

- Followed my Ph.D. advisor, Matthias Boehm, from Graz to TU Berlin.
- Research focus: Compression in the End-to-end Data Sience Lifecycle.
- M.Sc. and B.Sc. thesis supervisor on topics like quantization exploitation in neural networks, general sparsity exploitation, data augmentation of satellite images, and physical artifact/noise detection in MRI Scans.
- Lecturer - Seminar on ML and Data Management (2023, 2023/2024, 2024, 2024/2025, 2025)
  Managed, supervised, and presented a list of proposed research papers students could select.
- Project Mentor - Large-scale Data Engineering Project (2023, 2023/2024, 2024, 2024/2025, 2025)
  Mentored numerous custom projects leading to novel open-source contributions.
- TA - Architecture of ML Systems (2023, 2024, 2025)
  Independently designed and taught exercises on ECG classification, satellite data processing for Star and Galaxy classification from unaligned multi spectral images, and Earth building segmentation using Sentinel-2 and OpenStreetMap data.

GRAZ UNIVERSITY OF TECHNOLOGY — Ph.D. Student — Jan 2020 - July 2022

- Research focus: Workload-aware Compressed Linear Algebra
- TA - Architecture of ML Systems & Data Integration and Large-Scale Analysis (2020, 2021a, 2021b, 2022)
  Supervised a bachelor thesis on homomorphic encryption, contributed to building a federated backend in SystemDS, and mentored many custom projects for internal features in SystemDS.

IT UNIVERSITY COPENHAGEN — Research/Teaching Assistant — August 2018 - December 2019

- Research focus: Speech recognition on limited hardware resources, supervised by Pınar Tözün.
- Worked closely with the IT depatment on setup and testing of hardware in a SLURM cluster.
- TA in Big Data Technical, Large Scale Data Analysis and Process Innovation (2018 - 2019)
  Assisting in challenging cross-field courses with computer science- and business informatics students.

WORK EXPERIENCE

PHASE ONE — Test Engineer — June 2018 – February 2019

- Created and executed installing scripts for custom Linux distros and verified hardware components behavior.
- Taught and outsourced installing and verification procedures to a hardware factory in Thailand.
- Constructed automated testing using Jenkins of software-hardware interaction on USB2-3, ethernet, Wi-Fi, focus, and exposure via a scripting language called Robot Framework, Bash, and Python.

ZEN DATA APS — Full Stack Developer — October 2016 – August 2017

- DevOps, and deployment of web-content using a LAMP stack with Java integration.

REFERENCES

Matthias Boehm — **matthias.boehm@tu-berlin.de**
Tenured Professor — Technische Universität Berlin
Big Data Engineering (DAMS) Lab

Pınar Tözün — **pito@itu.dk**
Associate Professor — IT University of Copenhagen
Intensive Systems and Applications (DASYA) Lab

SELECTED PUBLICATIONS

- BWARE: Morphing-based Compression for Data-centric ML Pipelines, Under Submission
  ***Sebastian Baunsgaard**, Matthias Boehm*
- Adversarial Inputs for Linear Algebra Backends, **ICML** 2025
  *Jonas Möller, Lukas Pirch, Felix Weissberg, **Sebastian Baunsgaard**, Thorsten Eisenhofer, Konrad Rieck*
- AWARE: Workload-aware, Redundancy exploiting Linear Algebra, **SIGMOD** 2023
  ***Sebastian Baunsgaard**, Matthias Boehm*
- Federated Data Preparation, Learning, and Debugging in Apache SystemDS, **CIKM** 2022
  ***Sebastian Baunsgaard**, Matthias Boehm, Kevin Innerebner, Mito Kehayov, Florian Lackner, Olga Ovcharenko, Arnab Phani, Tobias Rieger, David Weissteiner, Sebastian Benjamin Wrede*
- ExDRa: Exploratory Data Science on Federated Raw Data. **SIGMOD** 2021
  ***Sebastian Baunsgaard**, Matthias Boehm, Ankit Chaudhary, Behrouz Derakhshan, Stefan Geißelsöder, Philipp M. Grulich, Michael Hildebrand, Kevin Innerebner, Volker Markl, Claus Neubauer, Sarah Osterburg, Olga Ovcharenko, Sergey Redyuk, Tobias Rieger, Alireza Rezaei Mahdiraji, Sebastian Benjamin Wrede, Steffen Zeuch*
- Training for Speech Recognition on Coprocessors. **ADMS@VLDB** 2020
  ***Sebastian Baunsgaard**, Sebastian Benjamin Wrede, Pınar Tözün*
- SystemDS: A Declarative Machine Learning System for the End-to-End Data Science Lifecycle. **CIDR** 2020
  *Matthias Boehm, Iulian Antonov, **Sebastian Baunsgaard**, Mark Dokter, Robert Ginthör, Kevin Innerebner, Florijan Klezin, Stefanie N. Lindstaedt, Arnab Phani, Benjamin Rath, Berthold Reinwald, Shafaq Siddiqui, Sebastian Benjamin Wrede*

EDUCATION

BERLIN UNIVERSITY OF TECHNOLOGY

Defence Planned August 2025

- Doktor der Ingenieurwissenschaften, Dr.-Ing
  Thesis: Workload-aware Compressed Linear Algebra for Data-centric Machine Learning Pipelines

IT-UNIVERSITY COPENHAGEN

- Master of Science, Software Development, Specializing in Advanced Computing
  Thesis: Scalable Speech Recognition — Completed June 2019 - GPA 9.73
- Bachelor of Science, Software Development
  Thesis: Multi-objective Genetic Algorithms — Completed June 2017 - GPA 7.33

LANGUAGES

• Danish(Native) • English • German(B2)

COMPETITIONS & AWARDS

SENTINEL-2 INNOVATION COMPETITION

2019

- 3. place, competing in innovative solutions to use Sentinel-2 images at ITU.

BEST MASTER THESIS TALKS AT ITU

2019

- ITU invited presentation of excellent master theses on speech recognition.

VOLUNTEER EXPERIENCE

CONFERENCE REVIEWER

- Sigmod PC 2026, EuroSYS Shadow PC Member 2024, PVLDB Emergency Reviewer 2023.

REPRODUCIBILITY REVIEWER

- Sigmod 2024, (Sigmod 2025)

VOLUNTEERING SERVICE

- VLDB Copenhagen 2022

PMC IN APACHE SYSTEMDS

2020 - present

- In the program management committee indicating a high level of autonomy in the Apache ecosystem.