

ARNAB PHANI

Graz, Austria
Cell: +436766568224
Email: phaniarnab@gmail.com
<https://github.com/phaniarnab>



EDUCATION

PhD in Computer Science **April 2019 - Present**

Graz University of Technology, Austria

Dissertation title: “Reproducibility, Reuse and Resource Elasticity for ML Workloads”

Supervisor: [Matthias Boehm](#)

Master of Technology in Software Systems. **2014 – 2016**

Birla Institute of Technology and Science (BITS), Pilani.

CGPA: 9.02

Dissertation title: “Commit Time Materialized View Maintenance for Bulk Load Operations in Teradata”

RESEARCH INTERESTS

My research interest lies broadly in the intersection of Data Management, Machine Learning and Systems, an emerging area referred to as Systems for ML or ML Systems.

RESEARCH & INDUSTRY EXPERIENCE

University Assistant **April 2019 - Present**

Institute of Interactive Systems and Data Science, TU Graz, Austria

- Fine-grained lineage tracing and reuse of intermediates in end-to-end ML pipelines.
- Unified memory management for multi-backend ML systems.
- Parallelization strategies for feature transformations.
- Research and develop internals of [Apache SystemDS](#).
- DSL/Compiler for large scale data analysis.

Sr. Software Engineer **July 2010 – March 2019**

Teradata Labs, India

- Developed features inside the query execution engine of **Teradata database**.
- Design and implementation of [Read Committed isolation level](#) for Teradata.
- Design and development of [Fast-mode Column Add feature](#), which allows adding columns in fast mode by storing the default column values in table header, without touching any rows.
- Worked on [global space accounting](#), integrating OpenMP into table operator and other features.

PUBLICATIONS

University Assistant (PhD Student)

Institute of Interactive Systems and Data Science, TU Graz, Austria

- Sebastian Baunsgaard et al. 2022. Federated Data Preparation, Learning, and Debugging in Apache SystemDS (Demo). In CIKM 2022.
- **Arnab Phani** et al. 2022. UPLIFT: Parallelization Strategies for Feature Transformations in Machine Learning Workloads. In PVLDB 2022.
- **Arnab Phani** et al. 2021. LIMA: Fine-grained Lineage Tracing and Reuse in Machine Learning Systems. In SIGMOD.
- Matthias Boehm et al. 2020. SystemDS: A Declarative Machine Learning System for the End-to-End Data Science Lifecycle. In CIDR.

Sr. Software Engineer

Teradata Labs, India

- **Arnab Phani**, Chandrasekhar Tekur, RKN Sai Krishna. 2019. Commit Time Materialized View Maintenance for Bulk Load Operations in Teradata. In ICECCT.
- Chandrasekhar Tekur, **Arnab Phani**, RKN Sai Krishna. 2019. Improving performance by avoiding transaction logging on Load Isolated tables in Teradata. In ICECCT.
- RKN Sai Krishna, Chandrasekhar Tekur, **Arnab Phani**. 2019. RepliSmart: A Smart Replication framework for optimal query throughput in read-heavy environments. In COMAD/CODS.

TEACHING

University Assistant (PhD Student)

Institute of Interactive Systems and Data Science, TU Graz, Austria

- **Lectures:** Transaction Processing and Concurrency, Distributed Storage and Data Analysis in Data Management. Modern Concurrency Control in Architecture of DB Systems. Model Serving Systems and Techniques in Architecture of ML Systems.
- **Teaching Assistant:** Data Management 2022, 2021, 2020, 2019. Architecture of DB Systems 2021, 2020.
- **Talks:** LIMA: Fine-grained Lineage Tracing and Reuse in Machine Learning Systems in SIGMOD 2021, UPLIFT: Parallelization Strategies for Feature Transformations in Machine Learning Workloads in VLDB 2022.

COMMUNITY CONTRIBUTIONS

- **Apache SystemDS:** I am a PMC member of Apache SystemDS, and the release manager of SystemDS 2.0 and 2.1 releases. I regularly contribute to SystemDS for my research.
- **Experiments:** All the experiments presented in papers are publicly available.
- **Benchmarks:** I published a benchmark and reference implementations for feature transformation workloads.

DATE: 25.09.2022

PLACE: Graz, Austria