ARNAB PHANI

Berlin, Germany
Email: phaniarnab@gmail.com, arnab.phani@tu-berlin.de
Website: https://phaniarnab.github.io/

EDUCATION

PhD in Computer Science **2019 - 2014**
TU Berlin, Germany
Grade: Summa cum laude
Dissertation title: "Fine-grained Reuse and Feature Transformations in Machine Learning Systems"
Supervisor: Matthias Boehm

M.Tech 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. I explore different aspects of the ML system internals to address high computational redundancy. In particular, I am implementing a novel framework for efficient, fine-grained lineage tracing and multi-backend reuse of intermediates inside ML systems.

SELECTED PUBLICATIONS

- **Arnab Phani** et al. 2025. MEMPHIS: Holistic Lineage-based Reuse and Memory Management for Multi-backend ML Systems. In EDBT.
- **Arnab Phani** et al. 2022. UPLIFT: Parallelization Strategies for Feature Transformations in Machine Learning Workloads. In PVLDB.
- **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.
- **Arnab Phani**, Chandrasekhar Tekur, RKN Sai Krishna. 2019. Commit Time Materialized View Maintenance for Bulk Load Operations in Teradata. In ICECCT.

RESEARCH & INDUSTRY EXPERIENCE

Research Assistant **April 2019 - Present**
TU Berlin, Germany, TU Graz, Austria
- Primary contributor to Apache SystemDS, an open source end-to-end ML system.
- Contributing to ML system internals from compiler to multi-backend runtime.

Sr. Software Engineer **July 2010 – March 2019**
Teradata Labs, India
- Contributed to query execution engine of **Teradata database**.
- Design and implementation of Read Committed isolation level, Fast Column Add, Global Space Accounting, and many other features.

TEACHING & OPEN SOURCE CONTRIBUTIONS

- **Teaching Assistant:** Architecture of DB Systems, and Data Integration and Large-scale Analysis courses.
- **Invited Talks:** A Tutorial Workshop on ML Systems @ BTW 2023, @ AWS Berlin, 2024.
- **Apache SystemDS**: PMC member and Release Manager (2.0, 2.1) of Apache SystemDS.
- **Reproducibility:** Availability and reproducibility of all paper experiments.
- **Benchmarks:** FTBench benchmark for feature transformation workloads with reference implementations.
- **Services:** SIGMOD 2026 (PC member)

DATE: 22.12.2024 PLACE: Berlin, Germany