ARNAB PHANI

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

EDUCATION

PhD in Computer Science **2019 - 2024**
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"

SUMMARY

I specialize in Data Management. During my PhD, I explored different aspects of the data **system internals** to address high computational redundancy within ML tasks. I am a regular contributor to **Apache SystemDS**, a leading open-source system for end-to-end data science. In addition, I have a strong background in relational database systems having worked extensively on the **query engine** of **Teradata** prior to my PhD.

SELECTED PROJECTS

- Holistic Lineage-based **Reuse and Memory Management** for Multi-backend ML Systems (EDBT 2025).
- **Parallelization Strategies** for Feature Transformations in Machine Learning Workloads (PVLDB 2022).
- Fine-grained **Lineage Tracing** and Reuse in Machine Learning Systems (SIGMOD 2021).
- SystemDS: A **Machine Learning System** for the End-to-End Data Science Lifecycle (CIDR 2020).
- Commit Time **Materialized View Maintenance** for Bulk Load Operations in Teradata (ICECCT 2019).
- Fast-path Column Add in Teradata Database

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.
- System internals from compiler to multi-backend runtime (CPU, Spark, GPU).

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.

OPEN-SOURCE CONTRIBUTIONS

- **Apache SystemDS**: Regular contributor to Apache SystemDS.
- **Reproducibility:** Availability and reproducibility of all paper experiments.
- **Benchmarks:** FTBench benchmark for feature transformation workloads with reference implementations.
- **Invited Talks:** A Tutorial Workshop on ML4Sys and Sys4ML, BTW 2023, AWS Berlin, 2024.

DATE: 18.12.2024 PLACE: Berlin, Germany