ARNAB PHANI

# Berlin, Germany

Email: [phaniarnab@gmail.com](mailto: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](https://mboehm7.github.io/)

### M.Tech in Software Systems. 2014 – 2016

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

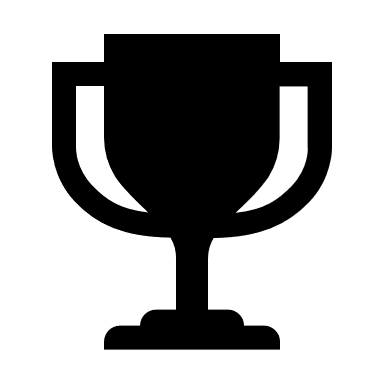
*CGPA*: 9.02

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

## SUMMARY

I specialize in **large scale data systems**, with a strong focus on **optimizing performance and efficiency**. During my PhD, I explored various aspects of AI/ML **system internals** to mitigate computational redundancy in machine learning workflows. Prior to my PhD, I worked extensively on the massively parallel **query processing engine** of **Teradata Database**.

## SELECTED PUBLICATIONS

* **Arnab Phani** et al. 2025. MEMPHIS: Holistic Lineage-based Reuse and Memory Management for Multi-backend ML Systems. In EDBT (**Best Paper Award** ).
* **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](https://github.com/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](https://docs.teradata.com/r/SQL-Request-and-Transaction-Processing/June-2020/Transaction-Processing/Load-Isolation), [Fast Column Add](https://support.teradata.com/knowledge?id=kb_article_view&sys_kb_id=efaf66dc47c2655886f3405c346d4346), [Global Space Accounting](https://docs.teradata.com/r/Enterprise_IntelliFlex_VMware/Database-Administration/Managing-Space-Operational-DBAs/Global-Space-Accounting), and many other features.

## OPEN SOURCE CONTRIBUTIONS

* **Apache SystemDS**: PMC member and Release Manager (2.0, 2.1) of Apache SystemDS.
* **Reproducibility:** Availability and reproducibility of code and [all paper experiments](https://github.com/damslab/reproducibility/tree/master).
* **Benchmarks:** FTBench [benchmark](https://www.vldb.org/pvldb/vol15/p2929-phani.pdf) for feature transformation workloads with [reference implementations](https://github.com/damslab/reproducibility/tree/master/vldb2022-UPLIFT-p2528/FTBench).
* **Invited Talks:** A Tutorial Workshop on ML4Sys and Sys4ML, BTW 2023, AWS Berlin, 2024
* **Services:** SIGMOD 2026 (PC member)

DATE: 05.03.2025 PLACE: Berlin, Germany