

# **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.
- 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**.
- Engaged in design and implementation of [Read Committed isolation level](#) for Teradata.
- Engaged in design and development of Fast Column Add feature, that allows users to add 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*

- **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.

## **COMMUNITY CONTRIBUTIONS**

I am a PMC member of [Apache SystemDS](#), and the release manager of SystemDS 2.0 release.

DATE: 16.01.2021

PLACE: Graz, Austria