

AARATI KAKARAPARTHY

Email: aaratik@cs.wisc.edu

Website: <http://pages.cs.wisc.edu/~aaratik>

Github: <https://github.com/aarati-K>



EDUCATION

Program	Institution	GPA	Year
PhD in Computer Science	University of Wisconsin, Madison	3.92/4	2019 – present
Master's in Computer Science	University of Wisconsin, Madison	3.88/4	2017 – 2019
Bachelor's in Computer Science and Engineering	IIT Madras	9.13/10	2011 – 2015

RESEARCH INTERESTS

Improving the performance of RDBMS by adapting to the workload properties and the underlying hardware.
Developing synergy between systems for relational query processing, data science, and machine learning.

PROFESSIONAL EXPERIENCE

Research Intern with DMX Systems Group at Microsoft Research, Redmond *(June – August 2020)*
advised by Vivek Narasayya and Christian Konig

- Predicted resource usage metrics (CPU, memory, disk) of databases in the cloud.
- Developed database scheduling policies using predicted metrics to reduce resource violations in the cloud.

Research Assistant at Microsoft Gray Systems Lab *(January 2018 – August 2021)*
advised by Prof. Jignesh Patel, UW Madison

- Lightweight online learning for low-latency data structures such as hash tables, with the ability to re-learn on-the-fly.
- Learning unique parameters of commercial SSDs using novel benchmarks; improving the performance of database systems by adapting them to the specific parameters of an SSD.
- Studied the implementation of common database operations on FPGAs using Microsoft's Catapult architecture.

Full Stack Developer at Zenefits in Vancouver, Canada & Bangalore, India *(October 2015 – July 2017)*

- Developed RESTful API services to integrate third-party applications with the Zenefits platform, as part of the platform engineering team.
- Managed end-to-end lifecycle of the services involving development, testing, deployment and monitoring.
- Redesigned the Zenefits dashboard codebase, while collaborating across teams to integrate multiple sub-products.
- Won second prize in an internal hackathon with around 50 participating teams.

Summer Intern at Samsung Electronics in Suwon, South Korea *(May – July 2014)*

- Developed an SMS spam detector as a combination of a generalized global model and a personalized local model.
- Designed a personalized lightweight entropy-based classification model for local classification.

Summer Intern at Google in Bangalore, India *(May – July 2013)*

- Designed a topic recommendation system for Google Baraza (a question-answer website similar to Quora), depending on the users' community and past activity.
- Implemented a prototype recommendation system using Flume and BigTable.

INVITED TALKS

- “VIP Hashing – Adapting to Skew in Popularity of Data on-the-fly” *2023*
to the **Microsoft Azure Databases Team, India** *(invited by Karthik Ramachandra)*
- “VIP Hashing – Adapting to Skew in Popularity of Data on-the-fly” *2022*
to the **Microsoft Azure Databases Team, Redmond** *(invited by Hanuma Kodavalla)*
- “FPGA for Aggregate Processing – The Good, The Bad, and The Ugly” *2021*
as part of **Microsoft GSL Talk Series** *(invited by Rathijit Sen)*

Splitting Dataframes for Memory-Efficient Data Analysis, (submitted) VLDB 2023*Aarati Kakaraparthi, Jignesh M. Patel*

- Developed a technique called splitting that can be automatically applied to tabular data to reduce redundancy.
- Implemented split dataframes in Ibis which reduced the memory consumption of Python notebooks by 22-57% and improved the running time by 13-70%.

Fine-Grained Hardware Profiling – Are You Using the Right Tools?, (under revision) SIGMOD Record 2023*Aarati Kakaraparthi, Jignesh M. Patel*

- Evaluated the correctness of tools for fine-grained hardware profiling on Linux platforms.
- Developed a library *PMU-metrics* which enables μ s-scale profiling on Intel CPUs.

VIP Hashing - Adapting to Skew in Popularity of Data on the Fly, VLDB 2022*Aarati Kakaraparthi, Jignesh M. Patel, Brian P. Kroth, Kwanghyun Park*

- Online learning using lightweight computation for adapting hash tables to skew in the workload.
- Transparently reconfigures the hash tables on-the-fly in the presence of insert and delete operations.
- First feasible attempt at learning with low-latency hash tables, with the ability to re-learn on-the-fly.

Tenant Placement in Oversubscribed Database-as-a-Service Clusters, VLDB 2022*Arnd Christian König, Yi Shan, Tobias Ziegler, Aarati Kakaraparthi, Willis Lang, Justin Moeller, Ajay Kalhan, Vivek Narasayya*

- Leveraging historical database tenant traces to predict resource usage of new tenants in the cloud.
- Using resource usage predictions to inform placement of tenants on physical nodes to minimize resource violations.
- Evaluation using production resource usage traces from Azure SQL DB show significant reduction over state-of-the-art tenant placement techniques.

FPGA for Aggregate Processing: The Good, The Bad, and The Ugly, ICDE 2021*Zubeyr F. Eryilmaz, Aarati Kakaraparthi, Jignesh M. Patel, Rathijit Sen, Kwanghyun Park*

- Compared *remote-main-memory* and *bump-in-the-wire* CPU-FPGA architectures for aggregation query processing; hardware trends indicate that the latter is likely to remain superior in the future.
- Discuss concerns of resource ceiling and bandwidth ceiling for aggregate query processing; high switching times pose challenges to the practical usability of CPU-FPGA architectures.

Optimizing Databases by Learning Hidden Parameters of Solid State Drives, VLDB 2020*Aarati Kakaraparthi, Jignesh M. Patel, Kwanghyun Park, Brian P. Kroth*

- Benchmarking commercial SSDs to uncover their internal operations and physical properties.
- Developed methods to deduce important parameters of SSDs, and proposed applications to database systems.
- Developed new techniques and implemented them in SQLite3 and MariaDB, which led to a significant improvement of 29% in the read throughput.

The Case for Unifying Data Loading in Machine Learning Clusters, HotCloud 2019*Aarati Kakaraparthi, Abhay Venkatesh, Amar Phanishayee, Shivaram Venkataraman*

- Building synergy between ML frameworks and the storage layer for better performance.
- Developed OneAccess, which improves data loading by performing sequential accesses through reservoir sampling.
- > 3.5x improvement in loading time over PyTorch's in-built data loader, for the MS-COCO Detection dataset.

SCHOLASTIC ACHIEVEMENTS

- | | |
|--|------|
| • DeWitt Graduate Fellowship in the Databases Group, UW Madison | 2022 |
| • All India Rank 108 in IIT-JEE (<i>top 0.02% nationwide</i>) | 2011 |
| • All India Rank 947 in AIEEE (<i>top 0.1% nationwide</i>) | 2011 |
| • Top 1% nationwide in Indian National Physics, Chemistry and Astronomy Olympiads | 2009 |

COURSES

- | | | |
|------------------------------------|----------------------------------|-------------------------------------|
| • Topics in DBMS | • Advanced Operating Systems | • Advanced Computer Networks |
| • Big Data Systems | • Mathematical Foundations of ML | • Cryptography and Network Security |
| • Advanced Computer Architecture I | • Data Mining | • Advanced Theory of Computation |