# Sriram Rao

sriram.rao04@gmail.com | +1 949-560-3250

github | linkedin


## Experience

### University of California, Irvine — Sep 2020 - present

*Graduate Student Researcher* — *Irvine, CA*

- *Janus*: Designed a database plugin to minimize latency via forecast-based data pre-compute. Resource reallocation between processing and pre-compute could cut latency 100%, per estimates. (Ongoing.)
- *GenIE*: Developed a framework to integrate data generators into databases to call implicitly during analysis. Employed to integrate HYSPLIT (simulator) with EnrichDB to enable 0 setup for follow-up queries.
- *Rush*: Created a pipeline execution system for workflows expressible as a directed acyclic graph of tasks. Code is available on GitHub as a sample.

*Teaching Assistant* — *Irvine, CA*

- Facilitated instruction and engagement in advanced courses like Introduction to and Principles of Data Management, Beyond SQL, and Projects in Databases and Web Apps, across 200+ students per course.
- Each week, led 3 practice-oriented sessions (30 students each) and 2 office hours to enhance comprehension and application of concepts, with guidance on programming assignments and debugging.
- Collaborated on creating comprehensive lecture slides, 30 quiz questions, 7 programming assignments, and 20 exam questions per course, with solutions; enhanced student understanding systematically.

### Dremio Corporation — Jun 2022 - Sep 2022

*Software Engineer - PhD Intern* — *Remote, CA*

- Devised proof-of-concept to refine row count estimation in data lakes by logging accurate statistics observed during query execution. These statistics could, then, be used for subsequent query planning. Familiarized with Apache Calcite, Iceberg. Idea based on LEO, DB2's learning optimizer.

### Microsoft IDC — Jun 2016 - Sep 2020

*Software Engineer 2* — *Bengaluru, India*

Engineered and deployed efficient systems to process big data and generate reports for Microsoft Ads.

- Piloted a Spark Streaming proof-of-concept pipeline to compute the statistical significance of A/B testing results 3 times faster than existing batched methods.
- Redesigned a workflow management system used for distributed Extract-Transform-Load (ETL) in 100+ workflows. Reduced the time to deploy primary use cases from 1 hour to < 2 min, simplifying releases.
- Refactored the cache configuration component of a middle-tier system through Aspect-Oriented Programming. Decreased the code needed to configure API caches 5x and the codebase size by 300 lines.
- Contributed to teammates' success with 30+ design reviews, 70+ code reviews, and guidance with code, design, debugging, and timely communication and mitigation of issues while on call.

### Microsoft IDC — May 2015 - Jun 2015

*Summer Intern* — *Bengaluru, India*

- Analyzed ingestion and query response times of 3 data storage mechanisms under normal and stress loads. Mechanisms tested: Kusto (now, Azure Data Explorer), MongoDB, and column-store SQL tables.
- Concluded that Kusto suited the use-case of log analysis (response $< 5$s) while tables in the column-store format were efficient for aggregation-based queries (response $< 1$s).

## Publications

- S. Dinesh, **S. Rao**, and K. Chandrasekaran, "Traceback: A Forensic Tool for Distributed Systems," Proc. 3rd Int. Conf. Adv. Comput. Netw. Informatics, pp. 17–27, 2016.
- GenIE: Generator-Driven Iterative Data Exploration. Integrating data generators, like simulators or benchmark data producers, into databases. Publication is in progress. Poster presented at alumni meet.
- Janus: Autonomous resource allocation and result pre-computation for future workloads. (In progress.)

## Skills

**Programming Languages** — Python | C# | Java | SQL | C++

**Technologies** — Big Data | Databases | Storage Formats | NoSQL | Query Processing, Operators | Spark
Docker | Cloud Services (Azure) | Distributed Systems | Flask (Python) | .NET | Spring (Java)

## Education

**MS and PhD in Computer Science** **Sep 2020 - Jul 2026 (anticipated)**

*University of California, Irvine (GPA: 3.96/4)* *Irvine, CA*

- Collaborating with Prof. Sharad Mehrotra on workload-aware pre-computation of results from expensive functions (e.g., ML inference, simulation).
- Relevant coursework: Principles of Data Management, Transaction Processing, Introduction to Machine Learning, Analysis of Programming Languages

**Bachelor of Technology in Computer Engineering** **Jun 2012 - May 2016**

*National Institute of Technology, Karnataka (GPA: 8.42/10)* *Surathkal, India*