RAJ SHAH

rajshah@gatech.edu | (470) 452-9783 | linkedin.com/in/qraj | rajshah-1.github.io/ | Atlanta, GA

EDUCATION

Georgia Institute of Technology | GPA: 4/4

Aug 2024 – May 2026

Master of Science in Computer Science (Computing Systems)

Indian Institute of Technology Bhubaneswar | GPA: 9.69/10 (Rank: 1/850)

Jul 2018 - May 2022

Bachelor of Technology in Computer Science (Highest GPA amongst all undergraduates)

Skills

- Programming: C++, Java, Python, SQL, JavaScript.
- Systems & Infrastructure: Distributed systems (Kafka, Solace), Databases, Operating Systems, Cloud (AWS, Azure).
- Performance Engineering: Multithreading, Java Native Interface (JNI), benchmarking (JMH).
- Machine Learning Systems: Model compression, retrieval-augmented generation (RAG), PyTorch.
- Software Engineering Practices: System design, CI/CD, testing (JUnit, Mockito), Unix/Linux environments.

EXPERIENCE

Barclays Investment Bank (Statistical Modelling and Development) | New York, US

Quant Dev Intern

Jun 2025 – Aug 2025

- Improved p99 latency by 23% by profiling the Java quoting pipeline with JMH, diagnosing cross-socket NUMA traffic in a JNI-backed C++ library, and optimizing synchronization.
- Increased throughput by 2.6× and cut message size by 50% by migrating inter-process communication from stringencoded messages to Protocol Buffers.
- Designed a **real-time event stream** to convert yield-space quotes to price-space valuations to have a unified fair price.

The D. E. Shaw Group (Middleware team) | Hyderabad, India

Senior Member Technical

Jan 2024 – Jul 2024

- **Decreased ops workload by 30%** by enhancing recovery of partial transaction log files and adding integrity checks against distributed replica warehouses.
- Boosted trading middleware performance by 40% by optimizing cryptography implementation with OpenSSL and Java Native Interface, eliminating garbage generation.
- Led the UI design of the intern evaluation portal and a UI overhaul of the monitoring tool based on user feedback.
- Guided and mentored **two summer interns** through the completion of the internship.

Member Technical

Jun 2022 – Jan 2024

- Reduced start-up time of stateful Kafka applications by 93% by engineering a Kafka Application Recovery Protocol.
- Co-presented Kafka Application Recovery Protocol at the Great International Developer Summit in Bengaluru.
- Created a Slack bot that allows users to run custom scripts securely with user credentials through Slack commands, **earning Runner-up at Infinity Hacks 2022**, a company-wide 24-hour hackathon, among 350 participants.
- Set up system infra such as load balancers, CNAMEs, Kerberos, CI/CD, and CRONs for 15+ services.

Software Developer Intern

May 2021 – Jul 2021

• Developed an **all-in-one monitoring and debugging tool for Kafka** users in the firm, providing a convenient way of analyzing topics, metrics, brokers, and consumer groups of Kafka clusters.

PROJECTS

Classmate AI | React Native, Flask, LLM, RAG, Python Celery, Azure, Full-stack, Figma

- Built a cross-platform app for real-time transcription, AI summaries, and Q&A, improving lecture review efficiency by 30% in
 pilot tests (50+ students).
- Architected a fault-tolerant Whisper + LLM pipeline and ran A/B tests on note layouts and Q&A interface, **showing 41% higher** engagement and 2× faster response times.

RAG Pipeline for Textbook QA | Python, FAISS, Sentence Transformers, Chunking, RAG

- Designed a retrieval-augmented generation (RAG) system that indexes 1,000+ textbook pages into FAISS with dense + BM25.
- Implemented **DocETL style query planning, ensemble re-ranking, and structured logging**, visualization, with support for llama.cpp inference and document chunking.

Publications

- Presented Tree Gradient Coding Considering Communication Delays and Partial Stragglers at IEEE ICC Denver, CO, 2024.
- Presented Improved Tree Gradient Coding with Non-uniform Computation Load at IEEE ICC, Montreal, QC, 2021.