

# ANURAG CHAKRABORTY

+1 226 8997477 ✉ a8chakra@uwaterloo.ca <https://www.linkedin.com/in/anu7ag/>  
<https://anuchak.github.io/>

## EDUCATION

---

### University of Waterloo

MMath (Thesis), Computer Science (GPA: 91.25/100)

2022 – Present

Waterloo, Canada

### Jadavpur University

BE, Computer Science and Engineering (GPA: 8.48/10)

2016 – 2020

Kolkata, India

## RESEARCH STATEMENT

---

I am interested in building data management systems that can support efficient processing and execution at scale and computer systems in general. I have been working on an embedded graph database management system (**Kùzu**) that aims to be scalable and provide good query speeds on a single node. I enjoy low level programming in concurrent settings and performance tuning of systems.

## RESEARCH EXPERIENCE

---

### Graduate Researcher | Advisor: Dr. Semih Salihođlu

September 2022 - Present

- Currently working on Kùzu, a new graph database management system. Kùzu is an in-process columnar GDBMS built for scalable processing of graph workloads (complex pattern matching, many-to-many join based analytical queries). It incorporates techniques such as factorization, vectorized query execution and worst-case optimal join (WCOJ) for cyclic join queries.
- My research focus in Kùzu has been on how to efficiently implement query processing for recursive joins. I wrote a specialized recursive join physical operator that introduces both inter and intra-operator parallelism.
- For my masters thesis work, I've made the following contributions -
  - (i) Modifying Kùzu's morsel driven parallelism architecture to efficiently execute recursive join queries (weighted / unweighted shortest path, all shortest path & variable length).
  - (ii) Scheduling policies with work stealing at runtime to distribute work among active threads. Designing lock free data structures for concurrent progress and work sharing between threads.
  - (iii) Assessing different optimizations at query planning stage and runtime such as MS-BFS, direction optimizing BFS to determine which query cases (reachability, path length or returning full path) would benefit from them.

### Undergraduate Researcher | Advisor: Dr. Nandini Mukherjee

July 2019 - June 2020

- Our objective was performance enhancement of query execution based on MapReduce paradigm, for which we developed a custom filesystem application called Node Guided Map-Reduce (NGMR) using Node.js.
- NGMR filesystem maintains an additional two level index structure at the Master & Client Node, to keep track of particular keys from large datasets. During query execution this metadata is used to reduce shuffling cost of intermediate results.
- This project was part of my Undergraduate Thesis and resulted in a **paper**.

## WORK EXPERIENCE

---

### Goldman Sachs Group Inc.

Engineering Analyst

July 2020 – July 2022

Bangalore, India

- Primarily worked on a Distributed Calculation Engine on top of Apache Spark
- Worked on migrating On-Prem Spark Job Server to AWS EMR with support for long-running Spark Contexts
- Migrated On-Prem Calculation Engine to Cloud using Serverless Compute (Fargate + ECS) and integration with API Gateway

## TECHNICAL SKILLS

---

**Languages:** C/C++, Java, Bash, OpenCypher, SQL

**Technologies/Frameworks:** Linux Perf, Intel VTune, GDB, HDFS, Apache Spark, Docker, AWS (S3, EC2, EMR, Lambda, ECS, Fargate, API Gateway), Git