Raihaan Sandhu

438-340-2559 | raihaan.sandhu@mail.utoronto.ca | linkedin.com/in/raihaansandhu | GaminRick7

Education

University of Toronto **Expected Graduation: May 2028**

Bachelor of Computer Science *4.0/4.0 GPA*

Relevant Coursework: Systems Programming, Object-Oriented Programming, Computer Organization, Data Structures

Experience

Software Engineer Intern Summer 2026

IBM *Toronto, ON*

- Incoming SWE Intern at IBM on the Z/OS Compilers team

Software Engineer (Distributed Compute Platform) Sep 2025 – Present

UTMIST *Toronto, ON*

- Designed a distributed job scheduler in Go and Redis Streams for **500+ developers**, enabling automatic rebalancing across GPU/CPU worker nodes, processing upto **200 jobs/day**
- Implemented GPU-based routing using Redis consumer groups and heartbeat-based health tracking to match jobs to compatible hardware (AMD/NVIDIA/CPU), **reducing failed dispatches by 90%**
- Built sandboxed job execution by spinning up isolated Docker containers per job with volume-mounted storage and per-node resource caps with automatic container cleanup on completion or failure
- Shipped a 7-endpoint REST control plane for job submission, status polling, and cluster health, serving as the single integration contract for the dashboard, CLI, and ops teams

Projects

GFS-Inspired Distributed File System | *C, TCP, Make* GitHub

- Built a distributed file system in C over TCP sockets, **modeled after the 2003 Google File System paper**
- Created a remote storage of large files with **0% data loss** by splitting files into multiple chunks across 3 servers
- Achieved **125MB/s throughput** across 400MB transfers by routing all data directly between clients and chunk servers
- Automated file chunk recovery by detecting dead servers and re-replicating their chunks to maintain 3x redundancy

ScholarSearch Engine | *Python, FastAPI, PostgreSQL, ChromaDB, Redis, Docker* GitHub

- Developed a full-stack research paper search engine to query across a **database of 1M+ research papers**
- Improved search accuracy by **3x** by fusing BM25 scores and BERT semantic similarity into a single ranking pipeline
- **Reduced repeat query latency by 10x** by caching frequent search results in Redis, eliminating redundant lookups
- Enabled researchers to map academic influence by building an interactive D3.js citation graph ranked by incoming citation count, surfacing the most referenced papers in any research area

LeetCLI | *C++, GraphQL, Gemini* GitHub

- Created a CLI tool for LeetCode with AI code insights, enabling problem submission, and performance analytics
- Enabled developers to solve **3,500+ LeetCode problems** locally using a CLI, improving workflow speed
- Implemented submission reliability and security by integrating LeetCode GraphQL API with session management and CSRF protection, ensuring safe and accurate problem execution for all users.
- Containerized Docker image enabling **zero-setup** installation and per-command usage via aliases and bind mounts

Technical Skills

Languages: C/C++, Python, Go, Java, JavaScript/TypeScript, Bash, SQL

Frameworks & Tools: PostgreSQL, Docker, Kubernetes, Redis, ChromaDB, Spring Boot, Nginx, Jenkins, Linux, Git

Infrastructure Tools: AWS S3, AWS Lambda, AWS DynamoDB, Apache HDFS