

Himanshu Rai

+91 7310791343 — raihimanshu22345@gmail.com — [linkedin](#) — [github](#)

Software Engineer skilled in building high-scale, distributed systems and multi-region infrastructure. Strong background in full-stack and AI-driven application development, with experience across DevOps and multi-agent systems. Committed to building reliable, high-performance systems and enhancing digital experiences with emerging technologies.

Experience

Versa Networks

Mar 2025 – Present

Software Engineer

- Maintained and extended a high-scale **URL Reputation Manager (URM)** system in **Go**, serving multi-tier URL/domain reputation lookups (in-memory trie/radix-tree cache, Redis, Cassandra) across multi-region **GKE** clusters.
- Designed a zero-downtime **Cassandra** schema migration tool with token-range parallelism, SHA-256 re-keying, and checkpointing for large-scale keyspace migrations.
- Implemented a lock-free in-memory trie/radix-tree **localcache** package using atomic pointer swaps for concurrent-safe reads, combined with an **LRU** eviction layer; reducing p99 request latency from over 400ms to under 150ms.
- Designed and built a highly available, scalable, and fault-tolerant **ClickHouse** analytics platform with a Go ingestion service, and deployed an **MCP server** that lets users query the analytics DB through any AI agent, including a web-based chatbot built on **GCP Vertex AI** models; replacing the need for static dashboards.
- Worked on the **Assets Management** system and its configuration sync pipeline, and designed a **Tenant Contacts** module enabling the support team to look up tenant contact information for sending notifications.

FRONT.PAGE (YC S21)

Jul 2023 – Mar 2025

Software Development Engineer

- Built and maintained a full-stack web application using **TypeScript, React, Redux, and Node.js**, integrating client and server components with a hybrid rendering strategy; combining Incremental Static Regeneration (with HTML cached in Redis) and server-side rendering to optimize load performance.
- Enhanced application functionality and real-time responsiveness by implementing polling and WebSocket-based communication between client and server.
- Implemented an optimized, fast post search functionality using **Elasticsearch**.

Projects

Shipyard Cortex

- Built a full-stack content creation platform using Go and React with TypeScript, featuring AI-powered semantic search, rich text editing, and real-time collaborative editing features.
[Github Repository](#) | [View Project](#)

Agent Bridge

- Developed a lightweight multi-agent communication bridge enabling AI agents to exchange messages, mentions, and shared context via the Model Context Protocol (MCP), backed by SQLite for persistence.
[Github Repository](#)

Skills and Technologies

Languages Go, Python, Typescript, SQL, Shell Scripting
Cloud & DevOps GCP, GKE/Kubernetes, Docker, Docker Compose, Prometheus
Backend Node.js, Express.js, FastAPI, Kafka, RabbitMQ
Frontend React.js, Next.js, Redux

Databases Cassandra, ClickHouse, PostgreSQL, MySQL, MongoDB, Redis
Data Tools BigQuery, DuckDB, Amplitude, Pandas
Others Linux, LLMs, RAG, Ollama, MCP, React-Query, Socket.IO, Prisma, SSE, Celery

Education

Graphic Era Deemed To be University

Bachelor of Technology in Computer Science and Engineering — CGPA : 9.13

Dehradun, India

July 2019 – June 2023