

VARUN SINGH

varun7singh10@gmail.com

[GitHub](#)

Phone: +91-8779608702

EDUCATION

Indian Institute of Information Technology Jabalpur

Dec 2021 – May 2025

B.Tech, Computer Science and Engineering

SKILLS

Languages: Golang, TypeScript, Python, C++, SQL

Backend & Systems: FastAPI, NestJS, Bull, Redis, PostgreSQL, RabbitMQ, gRPC, Docker, Kubernetes, S3 (DigitalOcean/AWS)

AI/LLM: LangChain, RAG Pipelines, pgvector, FAISS, Cross-Encoder Reranking, Hybrid Search, ReAct Agents, Prompt Engineering

WORK EXPERIENCE

Associate Software Developer (SDE-1)

Jul 2025 — Present

Oracle

Hybrid

- Co-developed an LLM-driven test configuration generation system for 5G Network Functions, implementing a FAISS-backed retrieval layer over versioned 150+ NF documents to enable context-aware structured output, improved reproducibility and reduced manual test-configuration authoring time by 50% across 6+ NF teams
- Refactored the NF logging pipeline by replacing centralized log servers with artifact-backed storage, introducing deterministic upload flows and reduced stale artifact storage costs by 35% through retention cleanup GitLab runners.
- Orchestrated deployment of internal utilities throughout Kubernetes environments by enhancing in-house CI pipelines, debugged and streamlined rollout workflows to reduce manual setup steps and improve first-pass deployment success within clusters from 60% to 95%.

Project Intern

Jan 2025 – Jun 2025

Oracle

Hybrid

- Contributed to performance run reporting infrastructure by building utilities to extract and structure worker-node utilization metrics and KPIs into standardized report tables, enhancing resource bottleneck visibility during NF performance runs.
- Identified, resolved production UI defects and refactored state management logic of the test scheduling dashboard, boosting dashboard stability for 75+ engineers cutting UI-related support tickets by 80%.

Consulting Backend Engineer

Feb 2024 – Apr 2024

Trijog

Remote

- Architected and delivered a modular LMS backend in NestJS with role-based access control, and engineered an asynchronous media processing pipeline using Bull and FFmpeg to transcode uploads into adaptive-bitrate HLS streams with thumbnail generation; reduced upload-to-stream time from 30 minutes to 5 minutes (83%) while enabling lower-bandwidth streaming via adaptive quality selection.
- Designed secure media delivery by implementing a Cloudflare Worker-based CDN proxy to authorize HLS segment access to prevent direct object storage exposure, strengthening content protection and improving distribution control.

PROJECTS

Repi: Developed an autonomous log investigation app that ingests distributed service logs into PostgreSQL via an async chunking + HNSW-indexed embedding pipeline, retrieves evidence using hybrid BM25 + pgvector search fused with Reciprocal Rank Fusion and cross encoder reranking, drives root-cause analysis through a multi-step ReAct loop with tool-call orchestration, with a streaming dashboard (SSE).

Collapser: Open-sourced a high-performance gRPC sidecar proxy in Go to mitigate thundering-herd effects via request collapsing, implementing thread-safe inflight coordination and caching to achieve up to 1000:1 collapse ratios under 10k+ concurrent load.

ACHIEVEMENTS

Winner, Smart India Hackathon 2023 — selected among 50,000+ submissions nationwide.

Selected for Code for GovTech (C4GT), contributing to government-grade open-source digital public infrastructure.

LeetCode rating 1728+ (Top 11%), solved 600+ algorithmic problems