

AMIT KUMAR

contacttoamit00@gmail.com | [Oxamit.com](https://oxamit.com) | [Linkedin](#) | [Github](#) | [Docker Hub](#)

EXPERIENCE

Senior Software Engineer, Coupang Dec 2025 – Present

- Contribution to order cancel workflow migration from legacy acitiviti based framework to temporal
- Implemented Aspect based Resilience4j circuit breakers in Spring Boot and Webflux microservices, drastically reducing thread choking and mitigating cascading failures to enhance service availability.
- Developed a Slack Domain Bot leveraging internal AI agents framework and MCP to power domain-level QnA, accelerating issue debugging and root-cause analysis.

Software Development Engineer 3, Network18 Dec 2023 – Dec 2025

- Designed a media ingestion platform for real-time ground reports using multimodal LLMs and vector embeddings, Qdrant and PostGIS for semantic and geo-based clustering, improving newsroom discovery speed
- Built a RAG-based VSCode extension powered by a self-hosted LLM using vLLM, improving developer code assistance efficiency by 25%
- Built a scalable, multi-tenant API platform with Kong Gateway, cutting API latency and load by 50% and saving 1 week of DevOps effort per service via automated quota, access, and monitoring controls.
- Pioneered an automated video generation pipeline using FFmpeg to convert articles into livestream-ready YouTube content, enabling support for 100+ peak concurrent viewers—first of its kind in the organization.
- Replaced static PHP plugin with a React-based, CI/CD-managed system using remote builds, removing cross-team dependencies and speeding up delivery cycles.

Senior Software Engineer, BYJU'S May 2022 – Dec 2023

- Re-architected the Video Packager service in Go using gRPC and Redis-backed multi-master coordination, improving packaging speed by up to 40%
- Cut infrastructure costs by over 90% through ECS tuning and AWS region migration.
- Optimized CDN delivery with unified storage and custom DNS for per-business unit tracking and analytics
- Built a high-concurrency FIFA quiz service handling 2500 RPS with sub-10ms response time using Go, ECS, DynamoDB, and Redis

Software Engineer, Toppr (Acquired by BYJU'S) June 2020 – May 2022

- Implemented semantic search for academic questions using OpenSearch, BERT embeddings, and ANN with p95 latency of 100ms
- Developed a fast type-ahead search service leveraging optimized Elasticsearch queries
- Built image optimization pipeline using S3, CloudFront, Lambda@Edge, and CloudFront Functions to serve PNG/JPG/WebP formats with improved SEO and browser support.

EDUCATION

IIT (ISM) Dhanbad, B.Tech in Electrical Engineering CGPA: 8.50 | 2016-2020

LBSS Sr. Sec School, Sr. High School, RK Puram, New Delhi 83% | 2015

Projects

- **Personal Kubernetes Cluster**: Bare-metal K8s on Oracle Ampere VM [\[Blog\]](#)
- **Type-ahead Service**: Movie search with fuzzy matching using word inverted index (Rust, Next.js) [\[Demo\]](#)
- **IconGenerator**: AI icon creator with vector similarity search (OpenAI API, Hugging Face) [\[Demo\]](#)
- **Serverless Container**: Golang/Knative implementation and Rust/V8 isolate version. 5s and <5ms cold start.

Skills

Languages & Frameworks: Golang, Python (Django, FastAPI, Flask), Rust, Node.js, JavaScript/TypeScript, React.js, Next.js, TailwindCSS, GraphQL, gRPC, REST, WebAssembly

AI & ML: LLM Integration, RAG Systems, BERT, Vector Search, Hugging Face, OpenAI API, Semantic Search

Databases & Storage: PostgreSQL, Redis, DynamoDB, Cassandra, Elasticsearch, OpenSearch, Qdrant DB

Infrastructure: AWS, GCP, Cloudflare, Docker, Kubernetes, CI/CD, NGINX, Unix

Hardware & Systems: Arduino, Raspberry Pi, Computer Networks, System Design