

Majid Ali

Backend Engineer | Cloud Infrastructure & AI/LLM Tooling | 88% Cost Reduction, MCP Servers

majidalee1.github.io | [LinkedIn](#) | [GitHub](#)

aleemii270@gmail.com | +92 348 1837238

Karachi, Pakistan

Summary

Backend engineer with 5+ years building distributed systems, cloud infrastructure, and production APIs, with recurring hands-on AI/LLM tooling experience – including a Model Context Protocol server exposing a database as LLM-callable tools and a 7-agent LangGraph coaching platform. Led a 4-person team architecting event-driven microservices on AWS EKS/Fargate, and independently redesigned cloud infrastructure to cut monthly spend by roughly 88%. Comfortable owning systems end-to-end in fast-paced startup environments.

Technical Skills

Cloud & Infrastructure

- **AWS** (EKS, ECS, Fargate, Lambda, Cognito, Elastic Beanstalk) – production deployment & scaling
- **Azure** (App Services, Functions, Application Gateway) – infrastructure redesign, cost optimization
- **Kubernetes** & Docker – containerized microservices, event-driven architecture (NATS Streaming)
- Multi-cloud migration, distributed systems architecture, high-throughput data pipelines

Backend & APIs

- **Node.js**, Express, Koa2, **TypeScript** – REST API design & production backend services
- **Python** – MCP server development, backend services, database tooling, automation scripting
- WebRTC, Server-Sent Events streaming, real-time systems (Socket.io, PubNub) for production apps

AI / LLM Tooling

- **Model Context Protocol (MCP)** – server design, LLM-callable tool & resource exposure
- **LangChain / LangGraph** – multi-agent orchestration, multi-provider LLM integration (Claude, OpenAI, Gemini)

Data & Databases

- MySQL, PostgreSQL, MongoDB, Redis – schema design, caching, query optimization at scale
- Amazon Redshift, ClickHouse – OLAP data warehousing, cost-optimized migration at scale

Reliability & DevOps

- CI/CD (GitHub Actions, GitLab CI) – deployment automation, multi-environment release pipelines
- Production security & observability – VAPT remediation, monitoring, incident response

Experience

Cloud Infrastructure, AI Tooling & Platform Reliability

Jun 2024 – Present

Technical Operations Lead, Sehat Kahani

- Built an MCP (Model Context Protocol) server exposing Sehat Kahani's production Azure MySQL database as LLM-callable tools, with a Chainlit natural-language chat interface for querying and management
- Led a cross-cloud migration to AWS Elastic Beanstalk for a live telemedicine platform, operating Azure and AWS in parallel across blob storage, Cognito auth triggers, and application monitoring
- Drove third-party penetration-test remediation (VAPT) across three production product lines of a healthcare platform handling protected patient data, alongside Redis caching and multi-tenant schema evolution
- Led development of an in-progress AI feature using prompt engineering against a base LLM to auto-generate SOAP-format clinical notes from patient consultations, aimed at cutting documentation time

Enterprise API Integration & Platform Delivery

Sep 2023 – Jun 2024

Senior Project Manager, Zetsol Technologies

- Coordinated enterprise integrations with major banking, insurance, and telecom partners onto a multi-tenant healthcare platform, each running as a white-labeled deployment with its own environment stack

Distributed Backend Systems & Cloud Infrastructure

Dec 2022 – Sep 2023

Backend & DevOps Team Lead, BitsPro

- Led a 4-person backend team building an event-driven microservices warehouse management system (Node.js, TypeScript, NATS Streaming) for the largest retail-first cannabis logistics operator in the US
- Architected and operated a containerized microservices platform on AWS EKS and ECS running on Fargate behind load balancers, running independently-deployable services communicating over an event bus

- Owned the data-warehousing layer for warehouse inventory search across 3 fulfillment centers, migrating from Amazon Redshift to ClickHouse to cut infrastructure cost while preserving fast columnar queries
- Delivered a multi-warehouse B2B SaaS operations suite – zone management, per-resource RBAC, and compliance-grade cycle-count auditing for multi-site operators in a heavily regulated logistics industry
- Built CI/CD pipelines automating lint, test, and deployment gates for a multi-service AWS microservices fleet, cutting manual release steps across independently-deployable backend services daily

Backend Engineering & Cloud Cost Optimization

Mar 2021 – Dec 2022

Full Stack Developer → Development Team Lead, Zetsol Technologies

- Redesigned a national telemedicine platform's Azure infrastructure around Application Gateway, App Services, and Azure Functions, cutting monthly cloud spend from ~\$12,000 to ~\$1,500 – an ~88% reduction
- Built the real-time video and audio consultation engine for a telemedicine platform using WebRTC (OpenTok) on a Koa2/TypeScript backend with MySQL, plus PubNub-backed realtime chat and notifications
- Built and owned CI/CD for a production healthcare platform, automating lint, test, and deploy stages across five environments spanning three separate product lines on a shared monorepo codebase
- Built the internal Angular operations console used to run a national telemedicine platform – doctor and clinic management, corporate account administration, and operational reporting dashboards

Education

Master in Software Engineering Sindh Agriculture University, Hyderabad, Pakistan	Jul 2022 – Present
Artificial Intelligence Specialization Isra University, Hyderabad, Pakistan	2020 – 2021
Bachelor of Information Technology (BSIT) Sindh Agriculture University, Tando Jam, Hyderabad, Pakistan	2017 – 2021

Selected Freelance & Contract Projects

- **Founder, Codenetic:** Founded and run a software consultancy delivering client products end to end, earning \$200K+ across 300+ applications built, 70+ published to the App Store and Google Play
- **Myaigi:** Architected a multi-agent AI platform using LangChain/LangGraph, orchestrating 7 specialized agents for routing, generation, QA, scheduling, and coaching behind an intent classifier
- **Myaigi – Streaming & Memory:** Designed a real-time agent pipeline streaming AI content over Server-Sent Events, backed by Pinecone vector memory and Cassandra for durable multi-session state
- **Myaigi – AI Safety:** Designed a dedicated QA and safety-review agent gating all AI-generated content before delivery, enforcing output quality inside the agent's orchestration graph itself
- **PSX MCP Server:** Contributed to an open-source Model Context Protocol server for Pakistan Stock Exchange data, fixing production scraper breakage after a data-portal migration, shipping new features
- **Baba Telehealth – AI Triage:** Fine-tuned and deployed an OpenAI model for AI-assisted patient triage and symptom assessment, integrated into a telehealth platform's patient intake workflow
- **Baba Telehealth – Backend:** Designed and deployed a Node.js/Express and MongoDB backend on AWS EC2 under PM2, integrating Twilio Video consultations, Socket.io realtime, and Stripe payments
- **AI Personalization Pipeline:** Delivered an AI personalization layer for a wellness app, building an OpenAI-backed survey-to-persona classification pipeline generating individualized programs
- **Custodial Crypto Wallet Backend:** Built backend services for a multi-blockchain crypto wallet, integrating a blockchain-as-a-service SDK for deposit-address generation and cross-chain transfers