GHULAM MUHAMMAD

PRINCIPAL SOFTWARE ENGINEER | SENIOR FULL STACK DEVELOPER

Lahore, Pakistan • muhammmadghulamm@gmail.com • linkedin.com/in/g-muhammad

PROFESSIONAL SUMMARY

Full Stack Developer and engineering leader with 8 years of experience designing scalable web platforms, distributed services, and data-intensive products for international clients. Leads cross-functional teams of 20+ engineers and owns delivery from discovery and architecture through production operations. Deep expertise in TypeScript, Node.js, React/Next.js, PostgreSQL, GraphQL, cloud infrastructure, CI/CD, and event-driven systems, with practical experience building RAG and LLM-enabled applications.

CORE SKILLS

Full Stack: TypeScript, JavaScript, Node.js, Express.js, NestJS, React, Next.js, GraphQL, REST, WebSockets
Data & Messaging: PostgreSQL, MongoDB, MySQL, ClickHouse, Redis, Kafka, RabbitMQ, GCP Pub/Sub, BullMQ
Cloud & DevOps: AWS, GCP, Azure, Docker, Kubernetes, Terraform, GitHub Actions, Azure DevOps, CircleCI
AI Engineering: Python, FastAPI, RAG, LlamaIndex, LangChain/LangGraph, Qdrant, LLaMA, Qwen, Ollama
Leadership: System design, technical strategy, client discovery, estimation, mentoring, code review, release management

PROFESSIONAL EXPERIENCE

Principal Software Engineer & Tech Team Lead | Dec 2023 – Present

CodingCops | Lahore / Chicago • Hybrid, international clients

- Lead and mentor 20+ frontend, backend, and DevOps engineers through micro-team ownership, sprint planning, architecture reviews, code reviews, and delivery governance; improved delivery velocity by 40%.
- Own client discovery, requirements, feasibility analysis, technical estimation, solution architecture, and production delivery across AI, IoT, analytics, and enterprise platforms.
- Architected Kafka and GCP Pub/Sub microservices processing millions of IoT records daily while sustaining 99.9% availability; delivered GraphQL services handling 10K+ requests per minute.
- Optimized PostgreSQL/Cloud SQL using indexing, query tuning, and connection pooling, improving response time by 45%; introduced Redis caching and performance-monitoring practices across services.
- Built automated GitHub Actions and Azure DevOps release pipelines with blue-green deployment patterns, reducing deployment time from two hours to 15 minutes.
- Led delivery of an AI procurement-negotiation platform, covering RAG architecture, open-source model evaluation, real-time LLM integration, and production stack selection.

Full Stack Engineer | Aug 2018 – Nov 2023

WebEvis Pvt Ltd | Dubai, UAE • Remote, international clients

- Delivered 5+ enterprise products spanning CRM, e-commerce, operations, and customer platforms, contributing across discovery, architecture, backend APIs, responsive web interfaces, deployment, and support.
- Designed modular Node.js/TypeScript services and React/Next.js applications serving 500K+ monthly active users; established reusable components, service boundaries, and API conventions for maintainable delivery.
- Built and documented GraphQL and REST APIs with validation, error handling, Swagger/OpenAPI specifications, background jobs, and real-time WebSocket workflows.
- Partnered directly with UAE-based clients to translate business workflows into technical plans, estimates, milestones, and sprint-ready tasks; communicated delivery risks and tradeoffs to non-technical stakeholders.
- Designed PostgreSQL and MongoDB schemas, indexes, and query patterns that reduced execution time by 50%; implemented Redis caching for high-traffic and data-intensive workflows.
- Integrated Stripe and other payment gateways using secure webhook processing, idempotency, and PCI-aligned practices; implemented authentication and RBAC with JWT, Passport.js, OAuth2, and Auth0.
- Containerized services with Docker and deployed scalable workloads on AWS EC2, RDS, and S3; built CI/CD workflows and supported monitoring, troubleshooting, releases, and production incidents.
- Mentored developers through pull-request reviews, technical walkthroughs, debugging sessions, and delivery standards while coordinating frontend and backend implementation across concurrent projects.

SELECTED PROJECTS

Power AI — Enterprise Knowledge & Operations Platform | *React, Python, PostgreSQL, RAG, AWS/GCP*

- Architected a secure self-hosted AI platform consolidating 52+ tools across five construction departments, with multi-source ingestion, department/project RBAC, Google SSO, audit logging, and approval-controlled automation.
- Designed multi-source ingestion from Procore, Google Workspace, contracts, and PDFs, plus an AI-assisted form library with human approval checkpoints and governed Gmail delivery.

Calculum AI — Procurement Negotiation Platform | *FastAPI, WebSockets, Qdrant, LlamaIndex, LangGraph*

- Led backend architecture for streamed multi-turn supplier negotiations; built RAG over negotiation and supplier knowledge and evaluated LLaMA/Qwen, embeddings, serving frameworks, and vector databases.
- Implemented real-time WebSocket sessions and service boundaries for LLM orchestration, retrieval, supplier context, and negotiation-state management.

AcwaPower Nomac — Plant Monitoring & Analytics | *NestJS, Kafka, ClickHouse, PostgreSQL, ADLS, Azure*

- Built real-time IoT/ETL pipelines and sub-second analytics over billion-row datasets; reduced cloud storage cost by 70% using ADLS and Parquet and enabled zero-downtime releases.
- Engineered an Azure Event Grid-triggered ADLS-to-ClickHouse migration service for near-real-time synchronization across multiple desalination plant sites.

ATLAS-10 | *Express.js, TypeScript, PostgreSQL, Redis, Next.js, WebSockets*

- Developed complex REST APIs and real-time workflows; combined Redis caching with database query optimization to reduce response times by 55% and added structured production logging with Winston.

UBU | *Express.js, TypeScript, PostgreSQL, Next.js, AWS*

- Built scalable EC2/RDS/S3 infrastructure with load balancing, autoscaling, JWT-based RBAC, and AI/ML endpoint integrations that contributed to a 20% increase in user retention.

ADDITIONAL IMPACT

- Reduced a large-scale data synchronization cycle from three months to three days with 99.9% accuracy through a Change Data Capture pipeline.
- Delivered projects for clients across the United States, UAE, and Saudi Arabia while balancing architecture, hands-on development, leadership, and stakeholder communication.
- Improved API response times by up to 60% through schema design, indexing, caching, connection pooling, and service-level performance analysis.
- Managed production monitoring, incident troubleshooting, release coordination, and infrastructure changes across AWS, Azure, and GCP environments.