

LIVIN MATHEW

Solutions Architect

Cochin, India | +91 9744142626 | livinmathew99@gmail.com | livinmathew.com

SUMMARY

Solutions Architect with 11+ years building distributed, serverless, and enterprise GenAI systems on AWS. Specialized in Retrieval-Augmented Generation (RAG), multi-agent orchestration, vector search, LLM evaluation frameworks, and multi-region architectures.

Delivered:

- 35% infrastructure cost reduction
- 30% improvement in answer relevance
- 95% legacy tech debt elimination
- SOC 2, HIPAA & HITRUST-compliant systems
- Platforms scaling to millions of users

Combines deep cloud architecture expertise with strong product execution and operational rigor.

CORE EXPERTISE

GenAI & Search:

RAG Architectures · LangGraph · Multi-Agent Systems · Semantic Caching · Vector Search (OpenSearch, pgvector, Qdrant) · LLM Evaluation Frameworks · Prompt Optimization

Cloud & Systems:

AWS (Lambda, Fargate, ECS, Step Functions, DynamoDB, S3, OpenSearch, Bedrock, EventBridge, SQS) · Event-Driven Architecture · Multi-Region Deployments · DR (RTO/RPO) · CDK · Terraform · SAM

Security & Observability:

OIDC · OAuth2 · RBAC · SOC 2 · HIPAA · HITRUST · Secure SDLC · CodeQL · OWASP ZAP · OpenTelemetry

Languages & Frameworks:

Go · Node.js · Python · TypeScript · React

PROFESSIONAL EXPERIENCE

TEKsystems (Client: Sage UK)

Product Lead | Jan 2024 – Present

Architected a multi-region enterprise GenAI knowledge platform on AWS.

- Designed multi-agent RAG system (FastAPI + LangGraph, 6 concurrent agents) with semantic caching and Claude integration.
 - Built Bedrock-powered semantic search ingesting 500k+ documents nightly into OpenSearch vector index.
 - Delivered multilingual responses across 7+ locales with sub-second cache latency.
 - Engineered distributed ingestion workflows using Step Functions (10 concurrent maps).
 - Built 15-metric LLM evaluation framework (groundedness, relevance, latency, token cost), reducing inference cost by 35%.
 - Reduced proposal preparation time from 2 hours to ~10 minutes via automated RAG workflows.
 - Implemented multi-region CDK stacks (us-east-1, eu-central-1) with GitHub Actions CI/CD and OIDC federation.
 - Integrated OpenTelemetry observability; achieved zero security findings during audit.
-

ChiSquare Labs (Client: SiteRx US)

Solutions Architect | Apr 2023 – Dec 2023

- Led monolith-to-serverless migration (Lambda, DynamoDB, EventBridge), improving scalability and reducing infra cost 35%.
 - Developed reusable CDK templates standardizing logging, monitoring, and security; adopted by 20+ engineers.
 - Directed SOC 2 compliance initiatives and secure SDLC implementation.
-

Metric Tree Labs (Client: SiteRx US)

Solutions Architect | Jun 2020 – Mar 2023

- Architected HIPAA-compliant data pipeline passing HITRUST audit on first attempt.
 - Designed DR strategy with defined RTO/RPO targets and validated failover drills.
 - Eliminated 95% legacy technical debt through modernization and architecture refactoring.
 - Led cross-functional delivery teams, increasing client revenue 40% and developer productivity 30%.
-

Riafy Technologies

Software Architect | Mar 2019 – May 2020

- Scaled Android applications to 10M+ downloads through backend modernization.
 - Improved legacy AngularJS platform performance by 90%.
-

Lmnrtr Tech

CEO and Co-Founder | Nov 2014 – Mar 2019

- Bootstrapped startup to acqui-hire exit; owned architecture, hiring, delivery, and P&L.
 - Reduced operational costs 25% and tripled traffic through scalable cloud architecture and growth optimization.
-

PROJECTS

- [KloudArch](#) - Platform converting architecture diagrams into Terraform (healthcare pilots).
 - [S3 Virus Scanner](#) - Serverless malware scanning & quarantine pipeline.
 - [Go-Fake](#) - Schema-aware CLI mock data generator.
 - create-sequelize-express - NPM scaffolding tool (2k+ downloads).
-

EDUCATION

Bachelor of Technology in Computer Science (coursework completed through 8th semester)
MG University, Kerala (2014 - 2018)