

Kofi E. Ampaabeng, Ph.D.

Staff / Senior Backend & Distributed Systems Engineer — JVM · Data Platforms · Multi-Tenant Systems

Washington, DC–Baltimore Area · Open to Remote · kofi@capewesley.com · www.ampaabeng.com · linkedin.com/in/kofi-ampaabeng *code proprietary (self-hosted GitLab); CW DevOps & RegAuth open-sourcing*

SUMMARY

Hands-on backend & distributed-systems engineer — 20+ years building **data platforms and multi-tenant systems on the JVM and Python**.

- **Architected RegAuth** (Spark/Solr over the full U.S. federal legal corpus) and **Vigil** (live streaming NLP/LLM), and built/mentored the teams that delivered them.
 - **Designed a multi-tenant, event-driven core-banking ledger** — single-writer posting, transactional outbox, immutable audit.
 - Fluent across the data-product stack: SQL & schema design, distributed processing (Spark, Celery), search (Solr), streaming (Kafka), plus the observability and operational discipline high-scale systems demand.
 - Go deep on correctness; care about the data that downstream algorithms and ML models consume.
-

TECHNICAL SKILLS

- **Languages:** Java (11–21), Python, TypeScript/JavaScript, SQL, Go; C++ (earlier).
- **Distributed data:** Apache **Spark** (PySpark / Spark SQL), partitioned **Parquet**, **Celery** distributed workers, **Apache Solr**, **Kafka**, Redis.
- **Backend/frameworks:** Spring Boot 3/4, multi-module Gradle, FastAPI, Hono/Bun, Express; event-driven architecture (transactional outbox, inbox dedup), DDD.
- **Datastores:** PostgreSQL, Oracle 10g–21c EE, MS SQL Server; SQLAlchemy, Liquibase, Drizzle.
- **Multi-tenancy & correctness:** tenant/legal-entity isolation, per-tenant sharding, immutable append-only audit, maker-checker, idempotency keys.
- **ML-adjacent data:** NLP feature pipelines (obligation extraction, summarization, content features); self-hosted LLM inference (vLLM/LiteLLM/Ollama) with PII/injection guardrails; ML text classification.
- **Platform engineering:** built **CW DevOps** — a provider-neutral internal developer platform (38 providers/11 categories, Terraform self-service, Pipeline IR → 6 CI platforms) with governed AI delivery (Claude worker behind a merge-blocking compliance engine) and an MCP server.
- **Cloud & IaC:** **Terraform** (AWS/Azure/GCP), **AWS** (VPC, RDS, EC2, IAM, ECS, S3), **Azure** (Container Apps, DevOps); cost-aware right-sizing — cloud-native vs. self-hosted to budget & business indicators.
- **Ops & platform:** Self-hosted **bare-metal data center** (on-prem private cloud, GPU compute); Docker, OpenShift, Kubernetes; Prometheus/Grafana, structured logging + correlation IDs; Apache/ModSecurity; CI/CD (Azure DevOps), GitFlow, SemVer.

- **Honest deltas:** Scala (JVM-fluent; Spark experience is PySpark) and consumer-hyperscale latency are growth areas, not claims.

SELECTED SYSTEMS

RegAuth — Distributed Regulatory-Data Platform (*visionary, architect & engineering manager*)

One of the few production systems that renders the full U.S. federal legal/regulatory corpus as structured, queryable data (obligation extraction, regulatory quantification) — a **scrape** → **clean** → **NLP** → **index** → **serve** pipeline over CFR, USC, Public Laws, Executive Orders, Statutes, and the Federal Register. Productionizes and extends the RegData/QuantGov approach I helped pioneer at Mercatus. I owned the vision and architecture, stayed hands-on, and **hired and mentored the team of junior engineers who built it with me** — original work predating AI code generation. - **Distributed processing:** PySpark / **Spark SQL** exporting NLP features to **partitioned Parquet** with tuned writer configs (snappy compression, filter pushdown, partition/file sizing); **Celery** distributed task workers with Redis-backed state; pluggable local/S3 storage. - **Data products for downstream ML/algorithms:** obligation-extraction, summarization, and content/language-feature pipelines that produce the structured datasets researchers and models consume. - **Search & serving:** Postgres → **Apache Solr** indexing bridge; React SPA + Express BFF; operator CLI. - **Engineering discipline:** Python monorepo (stable leaf-core / pipeline / Solr-API / studio / webapp) with a **SemVer'd public-API core** (2-minor deprecation window), Azure DevOps CI/CD, GitFlow, typed surfaces, and runbooks for on-call. - *Lineage:* productionizes the RegCensus/RegAuthorities data work I led at the Mercatus Center.

Vigil (Medusa) — Live Social-Media Intelligence Platform (*visionary, architect & engineering manager*)

A **live, multi-source social + news ingestion and AI-analysis platform** — larger infra than RegAuth — correlating real-time public discourse with structured socioeconomic data. Architected by me and delivered by a team of junior engineers I hired and mentored. - **Distributed ingestion & streaming:** Celery worker fleets (**scrape / analyse / ingest / stream**) on a Redis broker with Flower monitoring; concurrent scrapers for Twitter/X (GraphQL), Facebook & Threads, and 11 news sources. - **AI/NLP at scale:** NLU **text-intelligence** — coordination detection, sentiment, summarization — with **local LLM inference (Ollama, qwen2.5)** and Claude/OpenAI fallback: ML/LLMs consuming a live data stream. - **Serving & correlation:** **Solr** index + FastAPI; correlated against structured datasets (elections, census, income, health, World Bank/FRED) with an ML prediction model; Next.js dashboard. - Multi-service, fully containerized (custom Docker Hub images), self-hosted; PostgreSQL 16, Nginx TLS.

CWC Banking — Multi-Tenant, Event-Driven Core-Banking Ledger (*beta · architected & owned*)

IFRS-first core-banking platform; I own the architecture. - **Single-writer ledger** pipeline (PostingRequest → Template → Journal) as the sole GL writer; **three-currency** model (IAS 21) on every journal line. - **Event-driven** delivery via **transactional outbox** (written in the same DB tx as the state change) with consumer-side **inbox dedup**; idempotency keys on all external commands; Kafka transport. - **Multi-tenant isolation** (tenant + legal-entity), per-tenant DB sharding via a pool map; immutable, correlation-chained audit; maker-checker approvals. - Spring Boot 4 / Java 21, 46 modules, strict layered boundaries (circular deps are build errors); Oracle EE + PostgreSQL.

RSS Bank — Regulatory Reporting & Core Banking at Production Scale (*production · flagship*)

Live since 2019 (founded at Excenit, now Cape Wesley); automates statistical reporting/submission to the **Bank of Ghana**; deployed to banks including **air-gapped** sites. - **Real client base in Ghana & South Africa** — Fidelity Bank, Izwe S&L, ASA International, Sinapi Aba, Adehyeman, GRA Co-op Credit Union — proving multi-tenant / multi-jurisdiction in production. - Multi-module JVM application with submodule services (Spring Boot MFA, Excel report-gen API, **Python ML sector-classification** service) integrating core-banking backends (**FlexCube, T24/ASA, Calypso**). - Multi-tenant / multi-jurisdiction by design; OpenShift production; mature, multi-year, ~2,300-commit codebase.

USPS Change-of-Address — Platform Re-Architecture (*at ECS*)

Technical owner of a national-scale federal web system — and a clean demonstration of high-agency delivery in unfamiliar territory: - **Led the WebSphere → WildFly migration with no prior experience in either application server** — learned both runtimes on the job and delivered. - **Re-architected the application across major Spring Boot versions** (1.x → 4.x). - **Built a custom email-queuing system from scratch** when organizational constraints precluded Kafka/ActiveMQ. - **Re-architected authentication to a backend-for-frontend (BFF) model**, removing OAuth2 from the frontend apps; owned security and the Apache tier.

CW DevOps — Internal Developer Platform (*designed & built*)

A provider-neutral developer platform: **38 provider implementations across 11 categories** swappable via Python protocols; **Terraform self-service** for AWS/Azure/GCP (VPC/RDS/ECS/IAM); a **Pipeline IR** rendering one definition to 6 CI platforms (incl. Jenkins Groovy); and **governed AI delivery** — autonomous Claude worker (ticket → tests → auto-retry → MR) behind a **merge-blocking compliance engine**, exposed via an MCP server (~30 tools). **Four developer surfaces** — CLI (34+ commands), a VSCode extension (sidebar + AI assistant), web UI, and MCP — with one source of auth/config. DORA metrics, multi-tenant, self-hostable/air-gappable. ~32,000 LoC, **2,682 tests**.

Self-Hosted Private Cloud — Cape Wesley Data Center (*built & operated solo*)

Designed, built, and operate an **on-premise data center** that hosts the entire Cape Wesley stack — a deliberate cloud-repatriation to eliminate recurring cloud hosting/infrastructure cost. - Own compute, storage, networking, reverse-proxy/TLS, and GPU capacity for the full portfolio. - Orchestrated **on-prem private AI** — Open WebUI + **Ollama** (and vLLM GPU inference with LiteLLM cloud fallback) for internal team use; ordered guardrail pipeline (PII redaction → injection defense → audit), Prometheus/Grafana, multi-service Docker Compose.

EXPERIENCE

Cape Wesley — Founder & Principal Engineer · 2024–Present · Remote (*concurrent with ECS*) Design and build regulated banking, IAM, and data platforms end-to-end (systems above). Hands-on across backend, distributed data, and infrastructure; set engineering standards and run the full build→deploy→verify lifecycle. Hired and built a **globally distributed engineering team (Ghana, Rwanda, US)** serving clients across South Africa, Ghana, Pakistan, and the US.

Excenit — Co-Founder & Executive Director (*fintech startup*) · 2019–2024 · Ghana/US (*subsumed into Cape Wesley*) Built the RSS regulatory/core-banking platform and a banking client base across Ghana & South Africa from zero.

ECS — Senior Backend Engineer · 2022–Present · Fairfax, VA Technical owner of the **USPS Change-of-Address** web application (cleared, SF-85P). WebSphere→WildFly migration (no prior experience in either — learned on the job), Spring Boot 1.x→4.x re-architecture, custom email-queue (Kafka/ActiveMQ precluded), and a BFF auth re-architecture removing OAuth2 from the frontends.

Mercatus Center, George Mason University — Senior Research Fellow & Director, Open Health Policy · 2019–2023 · Arlington, VA (*Progressed from Research Fellow & Data Scientist.*) Built and led the data-science team; built the **HealthRegData** product and the **RegCensus** data platform + API and Python/R client libraries; drove **ML classification** of regulatory text; mentored engineers in SQL, R, and computational methods. (*The RegCensus/RegData work became the production RegAuth platform above.*)

IMPAQ International — Research Associate / Task Lead · 2013–2019 · Columbia, MD Designed technical-solution architectures for federal data systems (SNAP national research database; point-of-sale capture); led large-scale public/administrative data management and analysis for U.S. DOL/USDA.

Fidelity Bank Ghana — Infrastructure Manager · 2006–2007 · Accra Owned the Oracle 10g environment (tuning, backup/recovery, capacity planning) and core-banking rollout infrastructure for a startup bank.

Agricultural Development Bank (Ghana) — Database Administrator · 2005–2006 · Accra Oracle + MS SQL Server in a distributed environment; backup/recovery, capacity planning, banking-software DB upgrades.

Bsystems Ghana — Senior Developer & Database Consultant · 2003–2005 · Accra Led full-SDLC delivery of a financial asset-management system for Ghana's largest insurer and an ERP for a manufacturer; trained client teams on Oracle.

EDUCATION

- **Ph.D., Economics** — Clark University, 2013
- **B.A., Computer Science & Economics (First Class Honors)** — University of Ghana, 2004

Certifications: PgMP — Program Management Professional (PMI), in progress.

Selected publication: “The Long-term Effects of Malnutrition on Cognitive Achievement,” *Journal of Health Economics* 32(6), 2013. Full list available.
