VEDANTA DHOBLEY

Solutions Engineer

vedanta.dhobley@gmail.com • 732-609-6676 • GitHub • Website • LinkedIn

Engineering scalable data solutions for Global Finance and Risk at Citigroup. Building distributed pipelines that process mortgage, loan, and pricing data for risk, compliance, and wealth management. Personal projects leverage local LLM inference, durable workflow orchestration, auto-scaling infrastructure, and real-time data ingestion.

Work Experience

Backend Data Engineer — **2021 – Present**
Citigroup — *Jersey City, NJ*

- **Orchestration Platform:** Replaced legacy monolithic ETL with **Spring Boot** microservices, **Apache Kafka** job queuing, and **Apache Livy** for dynamic **Spark** job submission; enabled parallel execution and automated retries across distributed cluster
- **Retail Pipeline API:** Built **Python** RESTful API connecting **Oracle SQL**, **MongoDB**, and **S3** for real-time pipeline auditing; collaborated with front-end developers and operations teams to ship intuitive dashboard tracking ingestion status and data lineage
- **Database Encryption:** Developed **Java** encryption service using **NIST Format Preserving Encryption** (FF1/FF3-1) for PII redaction on international mortgage and risk datasets; achieved **25% performance gain** via multithreaded processing; passed security audit
- **Automated ETL:** Built scheduled SQL-to-**HDFS** pipeline with **Spring Scheduler** and **CyberArk** credential management; extracted and certified email notification system adopted as shared dependency across multiple teams
- **Developer Tooling:** Documented reusable deployment scripts with **Swagger** API docs, **Bash** automation for **Python** virtual environments, and structured **JSON** logging; reduced onboarding time for new engineers

Personal Projects

*Open source, containerized with **Docker**, deployed to production:* vedanta.systems

"Found Footy" – Autonomous Football Video Pipeline — GitHub | Live

- Real-time goal detection across **100+ teams**; fully autonomous pipeline produces ranked highlight videos within minutes
- Distributed microservices with **Temporal.io** orchestration, **MongoDB**, and custom **auto-scaling service** that monitors queue depth and scales workers (2–8 replicas) dynamically
- **RAG pipeline** using **Wikidata SPARQL** + local **llama.cpp** inference for team alias resolution; **vision model** validation to filter non-soccer content from keyword searches
- Built perceptual hash deduplication with dense 0.25s frame sampling, matching videos across resolutions using Hamming distance on 64-bit dHash frames
- 10-container **Docker Compose** stack: Temporal, PostgreSQL, MongoDB, MinIO S3, workers, **Selenium** browser automation for Twitter/X discovery, noVNC debugging

"Vedanta Systems" – Portfolio & Live Project Dashboard — GitHub | Live

- **React + TypeScript** frontend with **Vite** and **Tailwind CSS**; **Express** backend serving RESTful APIs for project data and live updates
- **Server-Sent Events (SSE)** streaming for real-time "Found Footy" updates; backend triggers `/refresh` via **Docker network** to push new goals to connected clients
- **nginx** reverse proxy routes single-port requests: static files to React build, API calls to Express server
- Deployed to vedanta.systems via **Cloudflare Tunnel** for zero-exposed-port hosting with automatic SSL; backend services communicate over Docker network

"Legal Tender" – Campaign Finance Graph Database — GitHub (In Development)

- **Dagster** asset pipeline ingesting FEC bulk data (~10GB / cycle) into **ArangoDB** graph; variable-depth traversals trace donations through PAC transfer chains to terminal sources
- Enables queries like "trace all donation paths from donor X through N PACs to final candidates"
- Employer resolution via **Wikidata** entity linking + **llama.cpp** inference; processing ~5M contributions across 2020–2024 election cycles

Education & Technical Skills

BS Applied Sciences in Engineering *(2016–2020)*
Rutgers University, New Brunswick

AWS Solutions Architect – Associate *(2024)*
Amazon Web Services

Languages: Python, Java, GoLang, SQL, Bash
Frameworks: FastAPI, Flask, Spring Boot, React
Orchestration: Temporal.io, Dagster, Prefect
Data: MongoDB, PostgreSQL, ArangoDB, Oracle, Kafka, Spark, HDFS
Infrastructure: AWS, Docker, Express, Jenkins, Git, S3, nginx
AI Tools: llama.cpp (local inference), Claude Code (development)