

Ammar Hamdi Hammouch *Data Engineer*

ahlan@ahammouch.me | 055 498 4459 | Riyadh | linkedin.com/in/ammam-hammouch | github.com/AmmarsGeeks

PROFESSIONAL SUMMARY

Building data pipelines and platforms, not just queries. Built an end-to-end LLM-Ops data platform covering all three legs — batch orchestration on Airflow, streaming and anomaly detection on Kafka, and a DuckDB medallion warehouse — wired together with schema contracts that fail in CI. At Entropy, I built a dataset quality pipeline validates CSV, XML, and JSON and catches corrupt files before they reach downstream systems. At Thmanyah, I built a multi-source aggregation engine that cut manual reporting by ~60%.

EDUCATION

Bachelor in Computer Science, Jazan University
GPA: 4.86 / 5 — First Class Honours

2021 – 2025

PROFESSIONAL EXPERIENCE

Software Engineer, Entropy

09/2025 – Present

LLM evaluation platform benchmarking Arabic-language model performance.

- Built a dataset quality pipeline that validates CSV, XML, and JSON uploads — format checks, metadata validation, and corrupt-file detection — stopping bad data before it reaches downstream annotation and evaluation systems, orchestrated on BullMQ + Redis.
- Connected an internal RAG chatbot to Microsoft Teams and Slack via FastAPI, handling connection management and query routing to an indexed internal knowledge base.
- Contributing to dataset management and annotation pipeline infrastructure for the Arabic-language LLM benchmarking platform.

Integration Specialist, Thmanyah

01/2025 – 05/2025

Media and podcast company. Internal systems across HR, TA, and Marketing.

- Designed a multi-source aggregation engine pulling from Tally, Google Forms, Airtable, and Excel into an internal CRM.
- Configurable flows similar to n8n, auto-distributed performance reports, and roughly 60% less manual reporting time.
- Built an internal HR platform covering payroll, notice periods, and 3-month probation evaluations — PDF report generation via Puppeteer, job queuing on BullMQ + Redis, and heavy schema migration work on Drizzle ORM.

React Native Developer, Motmaina Medical Center

04/2024 – 12/2024

Telehealth consulting platform for mental health.

- Built the React Native app end to end — state architecture managing health session state, patient–doctor handoffs, and booking flows across the full care cycle.
- Built the admin dashboard in Next.js with React Query (hydration, interceptors, SSR, query invalidation) for real-time visibility into sessions, providers, and clinical activity.

Technology Integration Intern, Reachware

08/2023 – 04/2024

iPaaS platform serving SMB and enterprise clients across the Middle East.

- Built data integration pipelines between Snowflake and ServeMe for financial reporting and analytics, handling schema mapping and scheduled ETL across the two systems.
- Contributed to Oracle NetSuite and Oracle Fusion ERP integration architecture, focused on finance module sync between the two systems.
- Worked on ZATCA e-Invoice Phase 2 (FATOORA) compliance integration: XML schema validation, cryptographic invoice signing, and real-time clearance against the ZATCA API.

Enterprise Application Developer, COMT

08/2022 – 08/2023

F&B technology.

- Architected a full POS platform from scratch, comparable to Foodics and Rewaa: inventory, product catalog, and categories built into a transaction cycle that runs from entry through order, payment, and receipt in under 30 seconds.
- Built offline-first architecture with real-time sync, so the system keeps running through network interruptions and reconciles when the connection returns.

PROJECTS

Balsam, LLM Evaluation Pipeline

Batch LLM evaluation platform running an 11-task Airflow DAG: dataset validation, Spark ETL for preprocessing, human-in-the-loop annotation routing, and three scoring tiers — classical metrics, semantic similarity, and LLM-as-a-Judge — producing ranked leaderboards. The pipeline ingests datasets through format validation and metadata quality checks before they reach the evaluation stage, with schema contracts enforced in CI so downstream dashboards never see broken data.

Mizan, Streaming Metrics and Anomaly Detection

Real-time observability service on Kafka and Faust, streaming evaluation metrics with EWMA and z-score anomaly detection. Anomalies surface as dashboard alerts when scoring drift or pipeline latency crosses thresholds. Sinks to TimescaleDB for time-series querying and Apache Iceberg for long-term analytical storage, making historical trend analysis possible across months of evaluation runs.

Basira, dbt + DuckDB Medallion Warehouse

Medallion-architecture warehouse (bronze/silver/gold) built on dbt and DuckDB, feeding Metabase dashboards. Joins LLM evaluation history from Balsam with operational telemetry from Mizan across three marts: cost (token usage, compute hours), quality (scoring accuracy, annotator agreement), and ops (pipeline latency, throughput). Schema contracts are enforced in CI — a breaking upstream change fails at the boundary rather than silently corrupting a downstream dashboard. The warehouse layer turns raw evaluation logs and streaming metrics into queryable business views that the platform team uses daily.

Flow Studio *🔗 , DAG Execution Engine on Postgres*

Webhook-driven aggregation and document engine that runs form submissions through a user-defined DAG. Leaderless by design — the advance protocol is a Postgres 'FOR UPDATE' transaction every worker runs, so fan-in fires exactly once and crash recovery is the normal code path. Isolated light and heavy worker classes (forked React-PDF renderers) run on separate BullMQ queues, keeping a runaway render from taking down email delivery.

IterateForge *🔗 , Durable Workflow Engine*

Durable workflow engine in TypeScript that runs inside a Node process on Postgres with no external orchestrator. Journals every step result and suspension point so a workflow killed mid-execution resumes from its last completed step across crashes and redeploys. Memoized step primitives, durable sleeps, external signals, and flow versioning that catches replay-incompatible edits at registration.

SKILLS

Languages

Python, SQL, TypeScript, JavaScript

Databases

PostgreSQL, Snowflake, MySQL, MongoDB

Cloud

GCP (Cloud Run, Cloud Functions, Pub/Sub, Firebase), AWS (S3, Lambda)

Data Engineering

Airflow, PySpark, DuckDB, Kafka, TimescaleDB, ETL, ELT

Infrastructure

Redis, BullMQ, Docker, Webhooks

Frameworks

FastAPI, NestJS, Next.js

CERTIFICATIONS

AWS Cloud Practitioner CLF-C02 — AWS

LANGUAGES

Arabic — Native | **English** — Professional