

ZIAD EL FATIH

+1(609) 540-3000 Los Angeles, CA

ziad.elfatih@gmail.com [linkedin.com/in/ziadelfatih](https://www.linkedin.com/in/ziadelfatih) github.com/ElfatihZiad

SKILLS

- **Programming:** Python, SQL, Bash
- **Data Engineering:** Spark (PySpark), Airflow, Kafka, Web Crawling, Batch & Streaming Processing, CDC
- **Cloud & Infrastructure:** AWS (S3, EMR, Glue, Lambda), GCP (BigQuery, Dataflow), Terraform, Docker, Kubernetes, CI/CD
- **AI & Data Systems:** Audio & Multimodal Data Processing, Data Acquisition, Dataset Preparation & Quality, NLP Pipelines, Data Enrichment, Embeddings, Vector Search, OpenSearch, PostgreSQL

EXPERIENCE

Senior Data Engineer

The Home Depot

Apr 2026 – Present

Remote, USA

- Engineered components of the **Enterprise Knowledge Layer (EKL)**, a knowledge graph connecting metadata, lineage, semantics, and relationships across enterprise analytical systems.
- Designed **property-graph models** across The Home Depot's **petabyte-scale EDW and BigQuery environment**, enabling graph-based discovery and semantic context across datasets, tables, views, and business concepts.
- Developed the **Monday Morning Reporting** analytics use case using Databricks Genie / GenBI, Metric Views, Apache Iceberg, and BigQuery, enabling natural-language exploration of governed enterprise metrics.

Senior Data Engineer

Digital Plastics Inc. (plastics.com)

Sep 2025 – Mar 2026

Boston, MA, USA

- Designed and built a data platform supporting **millions of engineering material records** across PostgreSQL, OpenSearch, and Neo4j, directly powering production search, retrieval, and AI-assisted product experiences.
- Built and optimized **OpenSearch indexing, filtering, and ranking pipelines** supporting fast faceted search across millions of highly structured and semi-structured records.
- Developed data ingestion and transformation workflows to normalize heterogeneous materials data into production-ready representations for search, analytics, and downstream AI applications.
- Designed a backend orchestration layer for query routing, retrieval planning, and context assembly across structured databases, search indexes, and AI services.
- Implemented hybrid retrieval combining structured filtering, vector similarity search, and relationship-aware retrieval to improve result relevance and contextual grounding.

Senior Data Engineer

DeepSearch Labs

Oct 2022 – Jun 2024

London, UK

First data hire at an AI startup building a multimodal intelligence and search platform.

- Designed and operated a cloud-native **AWS data platform** using **Airflow, PySpark, and Amazon EMR**, processing **3TB+ of multimodal data daily** (web, PDFs, audio, video) for production search and machine learning applications.
- Built serverless ingestion pipelines using **AWS Lambda, Glue, and Step Functions** to process high-volume unstructured data from web crawlers, document uploads, and external APIs.
- Developed a modular ingestion framework supporting **APIs, SFTP, relational databases, schema evolution, and CDC**, reducing new data-source onboarding time by **75%**.
- Built distributed enrichment and indexing workflows integrating NLP models, embedding generation, vector search, and entity-aware ranking for production search and document intelligence systems.

- Deployed containerized data services using **Docker and ECS Fargate** and managed cloud infrastructure with **Terraform and GitHub Actions CI/CD**.

Data Engineer

Go & Dev Consulting

Mar 2019 – May 2022

Casablanca, Morocco

- Delivered end-to-end data platform projects across retail, fintech, logistics, and media clients, covering streaming, cloud migration, analytics, and ML data workloads.
- Built **Spark Structured Streaming** pipelines on Databricks processing GPS and IoT telemetry from Kafka, with windowed aggregations and watermarking for real-time operational analytics.
- Migrated legacy SQL Server and Hadoop workloads to **AWS Glue, S3, and Redshift**, improving analytical query performance by up to **60%**.
- Built **GCP Dataflow and BigQuery** pipelines processing transaction and user-event data for reporting and ML feature generation.
- Developed dimensional **dbt** models, SCD Type 2 pipelines, data quality tests, and regulatory reporting datasets.

EDUCATION

Master of Computer Science

Maharishi International University

Relevant Coursework: Machine Learning, Cloud Computing, Algorithms, Database Management Systems

Oct 2024 – Oct 2026

Fairfield, IA

M.Eng in Computer Science

National School for Computer Science and Systems Analysis,

Relevant Coursework: Information Retrieval, Data Warehousing, Big Data Analytics, Machine Learning

Oct 2018 – Jun 2022

Rabat, Morocco

Classes Préparatoires aux Grandes Écoles (CPGE)

Intensive undergraduate program in Mathematics, Physics, and Engineering Sciences

Jul 2018

Casablanca, Morocco