

# ZIAD EL FATIH

Fairfield, IA 52557 • +1-609-540-3000 • [ziad.elf@gmail.com](mailto:ziad.elf@gmail.com)  
<https://www.linkedin.com/in/ziadelfatih/>

## DATA ENGINEER

Experienced Software & Data Engineer with over 5 years of experience building scalable, cloud-native data pipelines across AWS and GCP, Specialized in batch and streaming ETL, data platform architecture, and modern data stacks (Airflow, dbt, Kafka, BigQuery, Redshift). Key contributor at Deep Search Labs, leading the development of an Advanced Graph-Based Search Engine, Proficient in building efficient ETL pipelines, implementing cloud-based architectures, integrating modern data stack technologies, and backend development. Comfortable working across the entire software development lifecycle ensuring high-quality and efficient delivery of projects.

- Data pipelines • DevOps • ETL • Data Warehousing • Data Modeling • Cloud computing • Batch Processing
- Stream Processing • Data Orchestration • Containerization • Infrastructure as Code • CI/CD

**Languages:** Python

**Web:** Django, Flask, FastAPI

**Web Services:** REST, Client-Side Rendering, Server-Side Rendering, Microservices

**Web/App Servers:** Apache, NGINX, uvicorn

**Frameworks:** Django, React, NextJS

**Databases:** PostgreSQL, MongoDB, Elasticsearch, Redis, Neo4j, Redshift, BigQuery

**Design Patterns:** Singleton, Factory, Observer, ETL, Lambda Architecture, Data Lake Pattern, Change Data Capture, Data Sharding, Batch processing, Stream Processing, Data Partitioning

**SDLC:** Agile, Waterfall

**Tools:** Docker, kubernetes, Airflow, DBT, Terraform, CI/CD, Airbyte, Dagster, AWS Glue, Azure Data Factory, AWS lambda, pub/sub, Dask, polars, Superset, Pytorch, Kibana, Power BI

**Platforms:** AWS, GCP, Snowflake, Linux (Debian), NixOS, Windows, MacOS

**Big Data:** Spark, Hadoop, Kafka, Flink

## PROFESSIONAL EXPERIENCE

**CAREER NOTE:** Completed on-campus studies and currently taking distance education courses to complete a **Master's Degree in Computer Science** (Available for full-time, W-2 employment).

### DEEPSEARCH LABS, London, United Kingdom • 10/2022 – 06/2024

Deepsearch Labs is the next-gen research search engine delivering real-time, unique, and actionable insights to the most sophisticated Financial Services firms.

#### Senior Data Engineer

- Built distributed ETL pipelines using Airflow, AWS Glue and Spark on Kubernetes, processing 3TB+ daily with 99.5% SLA adherence and autoscaling for traffic spikes
- Designed cloud-native data platforms on AWS (Redshift) and GCP (BigQuery), automating infrastructure with Terraform reducing client onboarding from weeks to hours
- Designed and led development of Deepsearch Labs' modular ingestion platform, implementing reusable scrapers and API connectors that expanded data collection capacity by 10x and streamlined onboarding.
- Provisioned and optimized Elasticsearch, sharded MongoDB, and Neo4j clusters, achieving 99.9% uptime and cutting query latency from 1s to under 100ms amid 4x traffic growth.
- Built and scaled the company's knowledge graph system using Neo4j and vector databases, enabling semantic search and relationship discovery across 500M+ documents for the flagship AI product.
- Implemented dbt for modular SQL-based transformations, data lineage, and data quality checks on Snowflake and BigQuery, reducing data-related production incidents by 60% in six months.
- Developed Python-based ingestion services for diverse sources (REST APIs, SFTP, SQL Server), applying incremental loads, schema normalization, and partitioning strategies prior to BigQuery integration.
- Built a modern data stack MVP using BigQuery, Airbyte, dbt and and Dagster on GCP, enabling end-to-end analytics with lineage tracking.

**Technologies Used:** Python, AWS, GCP, Snowflake, Airflow, Kubernetes, Docker, Terraform, CI/CD, MongoDB, Elasticsearch, Dagster, Neo4j.

## **GO & DEV, Casablanca, Morocco • 03/2019 – 05/2022**

A global consulting firm focused on data transformation and AI-driven solutions to accelerate digital adoption and business impact.

### **Data Engineer**

- Played a key role in migrating on-prem workloads to Snowflake and BigQuery, including data modeling, access control setup, and pipeline development.
- Optimized multi-terabyte data transformations and aggregations with Apache Spark SQL and custom RDD/DataFrame operations, boosting query performance and resource efficiency.
- Built streaming pipelines with Kafka, Spark Structured Streaming, and DynamoDB for fraud detection, time-sensitive analytics, and real-time inventory tracking.
- Architected federated data platform across EMEA using Kubernetes and GCP with region-specific policies and cross-region synchronization
- Assisted in client-facing delivery, translating business data needs into clear technical requirements, and tailoring data architectures to match security, volume, and latency constraints.
- Designed reusable transformation modules in SQL and PySpark, standardizing pipeline logic across multiple client deployments.

**Technologies Used:** Python, Snowflake, GCP, Spark, AWS, Kubernetes, terraform, Redshift, BigQuery.

---

## **ACADEMIC PROJECTS**

### **Maharishi International University (2025)**

- **Scalable Data Pipeline with Spark, dbt, and Airflow:** Developed an automated data pipeline using Spark, dbt, and Airflow for batch and streaming workloads with Kafka ingestion and Redshift warehousing, ensuring quality, modularity, and full data lineage.
  - **Serverless Cloud Architecture for Real-Time Data Visualization:** Built a real-time data visualization pipeline using AWS Lambda, API Gateway, and DynamoDB, with a React frontend on S3/CloudFront, and automated CI/CD via GitHub Actions and Terraform.
- 

## **PERSONAL PROJECTS**

**Cloud-native Data Platform (2025):** Provisioned AWS infrastructure with Terraform (EC2, S3, IAM, Redshift), orchestrated data pipelines via cron jobs to load data from S3 to Redshift, applied DBT for transformations, and created an interactive dashboard with Apache Superset.

**NLP project (2022):** Built a web scraping pipeline with Scrapy and MongoDB, performing topic modeling (LDA) and sentiment analysis to track theme sentiment trends. Automated scraping, processing, and analysis using Apache Airflow and Docker for daily execution.

---

## **EDUCATION**

### **Master of Science in Computer Science**

*(In progress via distance education; expected completion 06/2027)*

Maharishi International University, Fairfield, Iowa

*Key Courses: Enterprise Architecture, Algorithms, Cloud Computing, Web Application Programming*

### **Master of Engineering in Computer Science**

National Higher School for Computer Science and Systems Analysis, Rabat, Morocco (08/2022)

*Key Courses: Data Structures and Algorithms, Computer Architecture, OOP, Operating Systems, Database Management Systems, Computer Networks, Graph theory, Machine Learning, Information Retrieval, Big data, Data Mining*

### **Undergraduate studies in Mathematics, Physics, and Programming.**

Classes préparatoires aux grandes écoles, Casablanca, Morocco (07/2018)