



Aymane MAGHOUTI

DATA ENGINEER | BIG DATA | AI SOLUTIONS

🏠 Calais (62100)

📅 Date of birth 05/07/2002

✉ aymanemaghouti16@gmail.com

💻 Open to remote work

🇫🇷 French

☎ +33 7 76 43 11 82

🌐 aymane-maghouti.github.io

PORTFOLIO

🌐 @aymane-maghouti

🐙 @aymane-maghouti

LANGUAGES

Arabic
Native

French
Courant

English
Courant

TECHNICAL SKILLS

Programming languages & Frameworks
Python - Flask - FastAPI - Java - Spring - Spring Boot - Spring DATA JPA - ReactJS - SQL

Databases
MySQL - Oracle - PostgreSQL - SQL Server - Mongo DB - Cassandra

Big data tech & BI tools
Hadoop - Spark - Kafka - Hive - HBase - Snowflake - DBT - Airflow - SSIS - SSRS - Power BI

Cloud Services

- **AWS** : S3, Athena, QuickSight, RDS
- **Azure** : Blob Storage, Data Factory, Databricks, Azure ML
- **GCP** : Bigquery, GCS, Looker

Functional Skills
ETL - ELT - Data modeling - Data warehousing - Data visualization - Database Optimisation

ML & DL & Data Science
Scikit Learn - Tensorflow (Keras) - Regression - Classification - Clustering.

INTERESTED

Learn new technologies

Sport (Football- basketball)

I am a Data & AI Engineer with experience in data analysis and data platform migration, real-time analytics, and BI solutions. I have worked on migrating enterprise platforms (Oracle → SQL Server, SSIS, Kafka, Power BI) and developing AI-powered data pipelines. Currently pursuing a Master's (M2) in Complex Systems Engineering at EILCO, France, I focus on delivering scalable insights and intelligent analytics by combining Big Data, Cloud, and AI.

WORK EXPERIENCE

Freelance Data & AI Engineer (Part-time, Remote)

Since February 2025 Shiftbricks AI Hoceima, Morocco

- **Designed and deployed AI-powered pipelines** processing **50K+ Arabic legal documents**, enabling scalable document digitization and retrieval.
 - **Developed ingestion & enrichment workflows** with Azure Document Intelligence, Cosmos DB, and PostgreSQL, improving **data structuring accuracy**.
 - **Implemented embedding pipelines** using Azure OpenAI & Cohere, generating **1024-dim vectors** that boosted semantic search precision to **80%+** with custom chunking strategy for RAG applications.
 - **Automated metadata extraction and hierarchical structuring**, reducing manual annotation time from **2h per document** → **15min**.
 - **Contributed to system architecture** of a knowledge management platform integrating Big Data, AI, and Cloud, ensuring **99.9% uptime** and scalability to **1M+ documents**.
- Tech Stack** : Azure Cloud, PostgreSQL, MongoDB, Python, beautifulsoup

Data Platform Migration & Real-Time Analytics Implementation

From February 2025 to June 2025 SetGet Consulting - Client Inwi Casablanca, Morocco

- **Led migration of INWI's legacy Oracle platform** → **Microsoft SQL Server**, reducing infrastructure and licensing costs by **30%** while modernizing the enterprise data ecosystem.
 - **Developed automated SSIS ETL pipelines** handling **50M+ rows/day**, cutting processing time from **2h** → **4min** and reducing operational effort by **90%**.
 - **Implemented partition-based purge strategies**, decreasing job runtimes from **6h** → **1min** and freeing **350GB+ storage/month**.
 - **Optimized query performance** by implementing **columnstore indexing**, reducing analytics query runtime from **30-60min** → **<1min**.
 - **Enabled real-time analytics** by integrating Kafka & Python, processing **500K+ events/15min** with **<2s latency**, boosting Quality team's anomaly detection speed by **80%**.
 - **Built a monitoring system** (SQL Server views + SSRS dashboards) to track ETL jobs, storage usage, and system health, enabling proactive troubleshooting and ensuring **99.9% platform uptime**.
 - **Delivered reporting solutions** with Power BI & SSRS, reducing manual reporting time by **70%**.
 - **Optimized infrastructure usage**, lowering total cost of ownership (TCO) and ensuring scalability for future data growth.
- Tech Stack**: SQL Server, SSIS, Power BI, SSRS, Kafka, Python.

Data Ingestion Pipeline for AI Application

From June 2024 to September 2024 Shiftbricks AI Hoceima, Morocco

- **Designed and implemented a Medallion-architecture pipeline** transforming **10K+ unstructured documents** into structured datasets, enabling scalable downstream AI applications.
 - **Ensured data quality and consistency** by integrating validation rules and schema checks, reducing ingestion errors by **35%**.
 - **Automated and orchestrated workflows with Apache Airflow**, managing **daily batch pipelines (5+ DAGs)** with built-in error handling and retry policies, ensuring **99% job success rate**.
 - **Developed a FastAPI + React.js validation tool**, enabling business users to review and approve processed records, cutting manual validation time from **2h** → **20min**.
 - **Implemented monitoring and logging** within Airflow to track job failures, execution latency, and data freshness, ensuring transparency across the ingestion pipeline.
- Tech Stack**: Python, FastAPI, React.js, MongoDB, PostgreSQL, Airflow, Prompt Engineering, Git/GitLab, Big Data Architecture, Docker

WORK EXPERIENCE

● Data Collection and Analysis of Train Circulations and Activity Dashboard Design

From July 2024 to September 2024 ONCF Rabat, Morocco

- **Developed a Python ETL pipeline** to collect and process **10M+ train circulation records/day**, loading them into a SQL Server data warehouse for centralized analytics.
- **Designed a star schema** (fact + 8 dimensions) optimized for performance, reducing query latency by **60%** and enabling efficient KPI reporting.
- **Built interactive Power BI dashboards** presenting key operational metrics (train delays, circulation patterns, capacity utilization), improving decision-making for the planning team.
- **Automated daily data refresh workflows** with SQLAlchemy, ensuring up-to-date analytics with minimal manual intervention.
- **Delivered insights** that improved on-time performance monitoring, directly supporting operational efficiency.

Tech Stack: SQL Server, Python, SQLAlchemy, Power BI

COLLABORATION

Sol Plaatje University (South Africa) | Jul 30, 2025

Invited as a Data Engineering Speaker (Visio Conference).

Presented big data fundamentals, data engineering architectures, and AI integration to 20+ students, followed by Q&A.

ENSAH – Al Hoceima, Morocco | Feb 26, 2024

Delivered a hands-on workshop on ETL pipelines for Data Warehousing.

Introduced fundamentals of ETL pipeline design, guided students through developing their first ETL pipeline, and supervised practical exercises with 30+ participants.

UM6P – College of Computing (Morocco, Remote) | Jan 2025

Worked with a professor on *Forecasting Large-Scale Renewable Energy Sources*.

Tools: Python, NumPy, Pandas, APIs, Git/GitHub.

Sol Plaatje University (South Africa - Remote) | Jan – Apr 2025

Collaborated with a Data Science instructor to develop a *Sign Language Recognition Application* – Edge-Cloud Hybrid Architecture for real-time, privacy-preserving recognition.

Tools: HDFS, Kafka, Spark Streaming, Python, React Native, REST APIs, FastAPI.

ENSAH – Al Hoceima, Morocco | Mar 7, 2024

Conducted a training session on Machine Learning (Logistic Regression).

Covered theoretical foundations of logistic regression, then led an end-to-end project on building a classification model and deploying it into a production environment, with 30+ participants.

LATEST PROJECTS

End-to-End AI Solution Migration – From On-Premises to Azure Cloud

- Developed an AI-powered smartphone price prediction system integrating **FastAPI** (model API), **Spring Boot** (backend), **React Native** (mobile app), **PostgreSQL**, and **Power BI** dashboards.
- Implemented an **MLOps** pipeline with **MLFlow** and **Apache Airflow** to automate model training, versioning, and deployment.
- Migrated the entire architecture to **Azure Cloud**, leveraging Azure App Service, Azure Database for PostgreSQL, Azure Machine Learning, and Power BI Service.
- Delivered a scalable, cloud-native solution with automated retraining, real-time prediction APIs, and interactive dashboards for market insights.

Smartphone Price Prediction Using Lambda Architecture in a Big Data Environment

- Implemented a Lambda Architecture-based big data solution supporting real-time and batch processing.
- Integrated **Kafka**, **HBase**, and **PostgreSQL** for scalable ingestion, real-time views, and batch analytics.
- Built real-time prediction pipelines using **XGBoost** for low-latency inference.
- **Developed a Spring Boot service** to serve real-time **price predictions** via REST API.
- **Created Power BI dashboards** for batch analytics and historical trend analysis.
- **Orchestrated batch workflows** with **Apache Airflow** and managed storage using **HDFS**.

EDUCATION

● Master Ingénierie des Systèmes Complexes

Since September 2025 Ecole d'ingénieurs du Littoral Côte d'Opale Calais, France

● Engineering Degree in Data Engineering

From September 2020 to June 2025

National School of Applied Sciences of Al Hoceima Al Hoceima

- Specialized in **Data Engineering & Big Data**, covering advanced topics in **databases**, **distributed systems**, **cloud computing**, and **machine learning**.
- Acquired strong hands-on expertise in **ETL/ELT pipelines**, **data modeling**, **data warehousing**, **Hadoop/Spark ecosystems**, **streaming with Kafka**, and **BI solutions (Power BI, SSIS, SSRS)**.
- Developed several academic and professional projects, including **real-time analytics pipelines**, **cloud-native AI applications**, and **big data architectures**.
- **Graduated 1st in class (Promotion Major)**, demonstrating academic excellence and leadership in collaborative projects.