

GHAITH BENHAJ

☎ (+216)97132477 ✉ ghaith.benhaj@gmail.com [in linkedin.com/in/ghaith-benhaj-29403a18a/](https://www.linkedin.com/in/ghaith-benhaj-29403a18a/) github.com/kAif-gh

Education

National Engineering School of Tunis (ENIT)

Bachelor of Science in Computer Science

Tunis, Tunisia

National Engineering School of Tunis (ENIT), University of Paris City

Double Master's Degree Information Processing and Complexity of Life

Tunis, Tunisia / Paris, France

Certificats

- Azure Fundamentals from **Microsoft**
- Certified Kubernetes Administrator **CKA**
- Machine Learning from Stanford University, **Coursera**
- Python for Data Science, AI Development from **IBM**

Work Experience

Equinor

06/2024 – Present

Consultant Data Engineer

Tunis, Tunisia

- Orchestrated an **end-to-end pipeline** to update data across systems, leveraging **MinIO** for backfilling data (processing over **10 TB** of historical data), **Redis** for real-time updates (reducing data retrieval latency by **35%**), **Kafka** for streaming topic publishing (handling **500,000** messages per day), and **DeltaTables** on **Azure** for data insertion and updates (improving data availability by **20%**). All steps were seamlessly orchestrated and deployed using **ArgoCD** Workflows, reducing deployment time by **40%**.
- Built **Argo** Workflows to automate **Spark** jobs, filtering data based on predefined conditions, which enhanced processing efficiency by **40%** and reduced manual intervention by 10 hours per week.
- Implemented a **Kafka**-based data streaming pipeline that processed **10 GB/day** of JSON data, transforming it into a format optimized for **GraphQL** queries. This solution reduced query response times by **25%** and increased system throughput to support **100,000 requests per hour**.

SambacLab

09/2023 – 05/2024

Software & Data Engineer

Tunis, Tunisia

- Improved system efficiency by replacing **Spark Jobs** with **Rust**-based implementation, cutting resource usage by **23%** and boosting performance and cost-effectiveness.
- Created a robust **Data Conflict Resolution** system for a Collaborative App, employing a multiplayer approach with Rust to uphold data integrity and enhance user experience, reducing conflict instances by **30%**.
- Constructed a streamlined data ingestion process for extracting and loading data into **Deltalake**, optimizing the performance of the collaborative app, resulting in a **40%** increase in data processing speed.

Projects

Profit Boosting via Traffic Prediction In Value Digital Services | *PySpark, Python, Scrapy*

02/2023 – 09/2023

- Spearheaded the development of a **Dashboard** and a **Deep Learning** model using Graph Data, optimizing "Advertising Boards" placement and boosting marketing profits by **10%** with **85%** prediction accuracy.
- Engineered an end-to-end **ETL** Pipeline for Tunisian Traffic Time Series data, scrapping, transforming and storing it in a Graph Database, reducing data processing time by **42%**.

Violence Detection in Tunisian Surveillance Cameras | *Python, Selenium, Machine Learning*

03/2023 – 04/2023

- Designed a sophisticated model leveraging computer vision techniques to identify instances of violence in surveillance camera footage, reaching a precision of **84%**
- Developed a customized web scraping solution to collect video data categorized with instances of violence, serving as valuable training data for machine learning models, resulting in a dataset of over **13,000** labeled instances.

Technical Skills

Languages: Python, Java, Scala, PySpark, SQL, PostgreSQL, Rust, Bash

Tools/Technologies/Frameworks: Git, Docker, Kubernetes, Terraform, ArgoCD, Azure, Oracle, BigQuery, Airflow, DeltaLake, Databricks, Kafka, Spark, Grafana, Pandas, Plotly

Languages

French: Level **B2**

English: Level **C1**