

# Hamza EL Karoui

## Senior Data Engineer & Scala Enthusiast

### EXPERIENCE

#### **DATAFAB** | BIG DATA ENGINEER | CONSULTANT | 5 YEARS

Datafab is a consultancy working in particular for the finance, insurance, energy, and luxury sectors.

Mar 2021 – Oct 2023 | BNP Paribas S.A. | Paris, France

- Taking ownership of building data solutions to provide actionable insights into key business metrics.
- Tuning Spark jobs performance, optimizing the code and solving data skew issues.
- Developed ML inference job using Spark + Deep Java Library + PyTorch.
- Migrating scheduling from legacy systems to Airflow.

KEYWORDS: SCALA/SPARK, KUBERNETES, SKAFFOLD, KUSTOMIZE, AIRFLOW.

Jun 2019 – Feb 2021 | Kering S.A. | Paris, France

- Designing, developing and maintaining a world-class, high-performing data pipelines to enhance the business insights using Spark 2.4 running on AWS EMR.
- Developed Spark Streaming applications integrated with Kafka as broker & Cassandra for data consolidation.
- Designed a multi-channel marketing attribution engine to replace old batch jobs:
  - Reduced compute time and used resources (from **7h30** with 1TB ram to **30 min** with only 20GB).
  - Migrated Spark/SQL jobs to AWS Athena + Glue + Lambda function for data source refresh.

KEYWORDS: SCALA/SPARK, PYSPARK, PYTHON, AWS, EMR, ATHENA, DELTALAKE, NIFI

Nov 2018 – Mai 2019 | Make.org | Paris, France

- Developed an inference API using Akka Http & Actors to serve NLP models saved in PMML format.
- Anomaly detection based on access logs and tracking data using PySpark.

KEYWORDS: AKKA, KAFKA, PYSPARK, PANDAS, DOCKER, GITLAB CI.

#### **ORANGE S.A.** | INTERNSHIP

6 months Apr 2018 – Oct 2018 | Paris, France

Automated Process Discovery based on Email Mining:

- Developed a machine learning approach to classify emails inside a corporation.
- Proposed an unsupervised approach to discover business process from the classified email.

Keywords: PROCESS MINING, PANDAS, SKLEARN, SPACY, NLTK, FLASK, PYSPARK.

### CERTIFICATES

#### **LIGHTBEND ACADEMY: SCALA LANGUAGE - PROFESSIONAL** | 2023

ID Number: d28a6ba0773d47e3ac0c03efdd5e9af9

#### **LIGHTBEND ACADEMY: AKKA CLUSTER - FUNDAMENTALS** | 2023

ID Number: db24a79bcac347e69c675e12b354621e

#### **AWS CERTIFIED CLOUD PRACTITIONER** | 2023

In progress



www.sharek.dev



github.com/helkaroui



helkarou@gmail.com



+33 6 30 71 35 50

### SKILLS

#### PROGRAMMING

Proficient:

Scala • Python • SQL

Familiar:

Java • Shell • Git

#### TOOLS/PLATFORMS

Zeppelin • Jupyter • Airflow • Yarn

#### DATA PROCESSING

Spark (scala & python) • Nifi • Akka Streams • Kafka • ksqldb

#### DATABASES

Cassandra • PostgreSQL • MongoDB • Elasticsearch

#### CLOUD/SOLUTIONS

AWS: EMR, EC2, EKS, Athena, Lambda

#### CONTAINERS & DEVOPS

Docker • k8s • Skaffold • Kustomize • Ansible • Jenkins • GitlabCI

### EDUCATION

#### IMT ATLANTIQUE

MASTER'S IN COMPUTER SCIENCE

Sep 2016 - Sep 2018 | Brest, France

#### SUP'COM

MASTER'S IN TELECOMMUNICATIONS

Sep 2012 - Sep 2016 | Tunis, Tunisia

### LANGUAGES

FRENCH: FLUENT

ENGLISH : PROFICIENT

Arabic: NATIVE