

Hamza KAROUI

Data Engineer - Freelance

+33 06 30 71 35 50

www.sharek.dev

Paris, France

[in hamza-el-karoui](https://www.linkedin.com/in/hamza-el-karoui)

helkarou@gmail.com

[helkaroui](https://github.com/helkaroui)

Éducation

Ingénieur Généraliste

Spécialité: Data Science

GPA: 3.6

Télécom Bretagne - France

Sep 2016 – Oct 2018

Ingénieur Télécommunication

Spécialité: Réseaux de Télécommunications

Distinction: Bourse pour double diplôme à Télécom Bretagne

Sup'Com - Tunisie

Sep 2014 – Août 2016

Classes Préparatoires

Filière: Math Physique

Rang national: 87 / 2500

IPEIN - Tunisie

Mai 2012 – Août 2014

Expérience Professionnelle

DATAFAB

conseil informatique

Nov 2018 – Oct 2023

Paris, France

- Participe au process de recrutement & la préparation des consultants aux entretiens clients
- Participe à la création des formations internes

BNP Paribas – CIB

Senior Data Engineer & Scala Evangelist

Mar 2021 – Oct 2023

- Intégration du framework Quantexa: Framework propriétaire, basé sur Spark, pour la détection de blanchiment d'argent dans le commerce (AML Trading).
 - Implémentation des règles métiers: scores et alertes AML
 - Optimisation des jobs Spark et résolution des problèmes de Data Skew en production
 - Évangélisation du langage Scala au sein de l'équipe et veille sur le respect des bonnes pratiques
 - Création de playbooks Ansible pour le setup des clusters Spark en mode Standalone
 - Mettre en place du monitoring (Spark Metrics)
 - Migration de Spark Standalone vers kubernetes
- Réalisation de différents POC:
 - Étude de comparaison de différent solution de stockage objets: Dell ECS, IBM Cleversafe et Flash-blade.
 - Spark sur k8s (opérateur vs spark submit)
 - Création de service d'inférence ML développé à l'aide de Spark + Deep Java Library + PyTorch, pour extraire et structurer les adresses dans les messages SWIFT
 - Mise en échelle d'un modèle Speech-To-Text pour transcrire les enregistrements des réunions en mode Batch

Stack Technique: Cloud privé, Scala/Spark, k8s, Kustomize, Skaffold, Jenkins CI, Elasticsearch, Kafka, Quantexa 1.6 & 2.4

Kering

Junior Data Engineer

Juin 2019 – Fev 2021

- Participation à la création du nouveau datalake du groupe
 - Concevoir, développer et planifier des jobs d'ingestion de données en Scala/Spark & PySpark
 - Consolidation et enrichissement de données marketing, clients et ventes issues de différents systèmes: 10 TB / Jours.
 - Génération de KPI Marketing: les modèles d'attribution & contribution.

- Conception d'un engin de calcul de l'attribution marketing (ATO-Engine)
 - L'objectif du projet est le calcul du retour sur investissement des campagnes marketing afin de:
 - Avoir un KPI commun à toutes les marques pour suivre les performances CRM
 - Identifier l'impact de chaque canal
 - Améliorer et adapter les canaux CRM et leur mélange
 - Volumétrie: + 1 Milliard de mails.
 - Résultats:
 - Réduction du temps de calcul par rapport aux jobs existants (de 7h30 à 1TB de Ram => 30min à 20GB)
 - Facilité de l'intégration de nouveaux canaux marketing grâce à une DSL customizable
- Reporting BI
 - Production des analyses pour des besoins business ponctuels sur Zeppelin.
 - Industrialisation de dashboards BI pour le suivi des campagnes marketing: Spark / Druid / Tableau

Stack Technique: AWS, EMR: Spark/Spark Streaming, Nifi for Scheduling & ETL, Docker, Jenkins CI, Cassandra, AWS Aurora, Druid, Elasticsearch, Kafka.

Make.org

Jan 2019 – Mar 2019

Junior Data Engineer

- Création de modèles NLP pour la classification de textes courts
- Création d'une API d'inférence en Scala (framework Akka) qui sert les modèles entraînés en Python

Stack Technique: Scala, Python, Akka, SciKit Learn, PySpark, Word2vec, Superset, Flask, Docker, Gitlab CI/CD, Rundeck

Societe Generale - Komeos

Nov 2018 – Dec 2018

Junior Data Scientist

- Indexation de documents .pdf dans Elasticsearch
- Développement de modèles d'extraction de tableau (Deep learning) dans les images
- Création d'un outil web de navigation et recherche dans les documents

Python, Elasticsearch, Flask, AWS API GATEWAY / EC2 / S3

Orange Labs

Mar 2018 – Oct 2018

Stage de fin d'étude

- Catégorisation des mails selon la logique métier de l'entreprise (cycle Data Mining classique)
- Extraction d'un modèle de processus métier à partir des emails classifiés

KERAS, Sci-kit learn, Spacy, NLTK, Flask, PySpark

Compétences techniques

Langages	Scala, Python, SQL, Java, Shell, Go (en cours)
Outils	Zeppelin, Jupyter, Airflow, Yarn
Data processing	Spark (PySpark included), Nifi, Akka Streams, Kafka, ksqldb
Bases de données	Cassandra, PostgreSQL, MongoDB, Elasticsearch
CONTAINERS & DEVOPS	Docker, k8s, Skaffold, Kustomize, Ansible, Jenkins, GitlabCI
Cloud Computing	AWS: EMR, EC2, EKS, Athena, Lambda

Certification

Scala Language - Professional	Lightbend Academy	2023
Akka Cluster - Fundamentals	Lightbend Academy	2023
AWS Certified Cloud Practitioner	AWS (En cours)	2023