





HUSSEIN AWALA


Sr. Data Engineer


 0761654896

 hussein@awala.fr

 Paris, FR

 husseinawala

 hussein-awala

 hussein.awala.fr



EXPERIENCE

Apache Airflow Committer & PMC member

Apache Software Foundation

 Apr 2023 – Ongoing

 remotely

- Active contributor and reviewer, member of the security team, migrating Github Actions runners to K8S, mentoring new contributors.
- Contributor to other Apache projects (Iceberg, Hudi, ...), and open-sourcer (spark-on-k8, async-batcher, ...).

Sr. Data Engineer

Voodoo

 Jan 2024 – Ongoing

 Paris, FR

- Developing a GDPR-compliant Lakehouse using Apache Iceberg to replace the legacy Parquet data lake and Redshift DWH, and reduce models latency 10X and monthly cost from 70k to 6k.
- Developing tools to run Spark SQL queries on K8S via Airflow with automatic Iceberg tables maintenance (optimize and vacuum).
- Developing optimized FastAPI and gRPC async APIs for ML inference which buffer requests and process them in batches.
- Developing a package to easily create online and offline features (transformation and aggregation of raw events) which improved the ML team's throughput by 2X.

Sr. Data Engineer

Leboncoin

 Oct 2021 – Jan 2024

 Paris, FR

- Developed low-latency Kafka stream applications via Spring Cloud Stream and KStreams on K8S for fraud detection (<100ms).
- Developed a scalable in-house feature store (DynamoDB) with Kafka stream apps for online features and Spark jobs for offline features (>1B/day).
- Designed and implemented a Lakehouse architecture (Apache Hudi) which replaced the legacy datalake for GDPR compliance and to reduce the cost and the processing time (compaction, z-order, indexing, ...).
- Configure Spark on Kubernetes cluster and migrate existing YARN (EMR) jobs to the new cluster, reducing cost (4X) and latency.

Data Engineer & Head of Data

Data4Risk

 Feb 2019 – Oct 2021

 Paris, FR

- Developed stream applications and batch ETL using Pyspark and Kafka to collect and process weather data.
- Implemented a low-latency Lakehouse using Delta Lake and Spark streaming on K8S, to create SCD type 1 and type 2 for a financial platform.
- Led the Data team: 2 DEs and 4 DSs working on different projects (computer vision ML models for satellite image processing, financial platform to study EU companies and give recommendations to fund holders, ...)

STRENGTHS

Self-learning

Hard-working

Persuasive

Determination

Motivator & Leader

Decision Making

Product Management

Creative Thinking

Solving complex problems

EDUCATION

M.S. in Data Science

Grenoble INP ENSIMAG

 Sep 2018 – Sep 2019

M.S. in AI & Big Data

Lebanese University Faculty of Science 1

 Sep 2017 – Jul 2018

B.S. in Computer Science

Lebanese University Faculty of Science 5

 Sep 2014 – Jul 2017

TOP SKILLS

Python

JAVA

Spark

Airflow

Kafka

Iceberg

Hudi

SQL

Athena

Docker

Kubernetes

helm

Terraform

AWS

GCP

Avro

DynamoDB

MongoDB


Redis

Argo


FastAPI

OOP

MOST PROUD OF



Competition in all arenas
to become a PMC member on one of the largest open source projects, working with thousands of talented people



Persistence & Hard Working
where despite Lebanon's difficult conditions and long working hours of my part-time job to finance my studies, I was one of the first graders



Voodoo Growth
Since joining Voodoo, I have played a significant role in improving the data stack to reduce latency and processing cost and solving all the challenges that allowed us to manage the BeReal social media application load (x100 of the previous load before) after the acquisition of the application on June 2024.

LANGUAGES

Arabic

French

English

