

AUBIN MAZET

Paris, France · mazetaubin@gmail.com · +33649282124 · <https://www.linkedin.com/in/aubin-mazet/>

SKILLS

Programming Language: Python, Scala, SQL, R
Big Data: Spark, Cloud (Azure, AWS), Kubernetes
Data Science: Time Series (forecast, anomaly detection), Generative models (GAN, VAE, GPT3)
ML Engineering: MLOps, DevOps (CI/CD, infra-as-code), APIs
Project Management: Agile methods, Management consulting, Data strategy

WORK EXPERIENCE

Quantmetry

ML Engineer - Data Engineer

Paris, France
Sep 2019 - Present

- **Sports retailer - Data Engineer**
 - Developed and industrialized daily ETL jobs on billion rows in Spark Scala
 - Developed Airflow pipelines to create transient EMR clusters and orchestrate spark jobs
 - Developed CI/CD pipelines to build, test and deploy spark jobs on different environments
Scala, AWS, Airflow
- **Testing, inspection and certification (TIC) - Data Engineer**
 - Deployed a state-of-the-art big data platform based on a microservices architecture (Kubernetes, Spark, Airflow, Jupyterhub, Gitlab CI, Minio, Helm, Delta)
 - Developed ETL spark jobs for processing data
 - Set-up Airflow on K8S to launch and orchestrate spark jobs
Scala, Kubernetes, Airflow
- **Pharmaceutical industry - Data Engineer**
 - Developed a streaming pipeline to ingest transactions based on an event-driven architecture
 - Deployed and optimized a distributed PostgreSQL database to handle real-time analytical queries on more than 5 billions rows)
Python, Azure, PostgreSQL
- **Water and waste management - Machine Learning Engineer**
 - Industrialized a predictive maintenance model with K8S to serve models prediction to several plants
 - Developed a MLOps Python template to industrialize machine learning models on Azure
 - Trained and up-skilled data scientists on DevOps and clean code.
Python, Azure Machine Learning, AKS
- **Food industry - Machine Learning Engineer**
 - Developed a digital twin offering real-time piloting recommendations based on sensor data.
 - Industrialised the model and the data ingestion pipeline on Azure using serverless resources
 - Developed CI/CD pipelines to deploy the application and the infrastructure
Python, R, Azure

Air France KLM

Data Scientist

Paris, France
Sep 2018 - Sep 2019

- **Fraud detection on credit cards, loyalty programs, flight bookings**
 - Deep learning: CNN, variational autoencoders, entity embeddings in NN
 - Graph database (Neo4J)
- **Revenue forecasting for finance department**
 - Machine learning: boosted trees, time series models
 - Creation of visualization tools with Flask and Spotfire

EDUCATION

Paris Dauphine University

Master's degree Finance

Paris, France
Sep 2016 - Sep 2019

Lund University

Semester Exchange

Lund, Sweden
Sep 2017 - Feb 2018

PROJECTS

Spark-Airflow-on-K8S Helm, Spark, Kubernetes, Airflow

<https://github.com/aubinmazet/spark-airflow-on-K8S>

Starter kit to run spark jobs with Airflow on Kubernetes

Common Readability Kaggle *Python, HuggingFace, Tensorflow*
<https://github.com/aubinmazet/common-readability-kaggle>
Kaggle challenge to evaluate the readability of a text

CERTIFICATIONS

Microsoft Certified - Azure Data Engineer Associate

Credential ID : H414-8260

May 2020

Deep Learning Certification - deeplearning.ai

Credential ID : ZSG3A9HA8AYY

Sep 2019