

YOUSRA ADDALI

Data Engineer (Python, Spark, Big Data, AWS, SQL)

5 ans d'expérience

 yosraaddali@gmail.com  0605646905  <https://github.com/yosraa>  <https://leetcode.com/u/yosraa>

COMPÉTENCES TECHNIQUES

- **Langages** : Python, Scala, SQL
- **Frameworks**: Django, Flask
- **Data**: Spark, Kafka, Hadoop, HDFS
- **DB** : Mongo, Postgres, redis
- **Cloud**: AWS (Lambda, SQS, S3, RDS, CloudWatch, IAM, VPC, EC2, Glue, kinesis, ECS)
- **Outils** : Git, GitLab CI/CD, Docker, Kubernetes, Jira, Confluence
- **Test** : unit test, pytest
- **Méthodologie** : agile Scrum

EXPERIENCES PROFESSIONNELLES

Avril 2022 – Juin 2024

Data Engineer @ Client dans le luxe

Projet : « Selector » est un outil d'aide à la décision pour les artisans qui vise à optimiser l'utilisation des matières première lors la confection des produits :

- Conception et développement des APIs.
- Création des scripts SQL
- Correction de bugs
- Optimisation de code
- Mise en place de PoC
- Développement d'un algorithme de positionnement des gabarits de sacs sur les peaux
- Développement des tests unitaires
- Migration d'un pipeline de donnée vers aws

Environnement : Python 3.11, Pyspark, Poetry, Flask 2.0.3, Marshmallow, Sqlalchemy 2.4.1, boto3 1.21, AWS (lambda, sqs, s3, cloudwatch, IAM, EC2, RDS), gitlabCI/CD, Postgres 13.5, Flyway 10.18, pytest 7.1.1, Docker, Bash

Septembre 2021 – Février 2022

Data Engineer @ Société Générale :

Le projet « Haussmann » vise à mettre à disposition des données raffinées à une équipe de Data Scientist :

- Ingérer des données pour les rapatrier
- Création des jobs de transformation pour les mettre au format attendu
- Orchestration des traitements
- Maintenance de la plateforme Data qui exécute les traitements

Environnement : Scala 2.13, Spark 2.11, environnement Hadoop (hdfs, hive 1.0), Jenkins, cloudera

Février 2020 – Septembre 2021

Data Engineer @ Enedis :

Développement et industrialisation d'une solution de maintenance prédictive destinée à anticiper les incidents et les crises applicatives.

- Mise en place de pipeline de collecte en temps réel et batch
- Mise en place de Dashboard permettant de présenter les résultats du modèle

- Maintenir et améliorer les pipelines de traitement des données.
- Création d'un pipeline CI/CD.
- Correction des bugs.
- Dockeriser et déployer des services sur kubernetes
- Mettre en production des algorithmes de ML sur des instances GPU
- Rédaction des documentations techniques
- Assurer les correctifs en production pour tout problème critique
- Préparer des présentations

Environnement : Python 3.7, Pyspark, ML (FBprophet), Multiprocessing, hdfs 2.5.8, kafka 2.5, Docker, kubernetes, gitlabCI/CD, Redis, Grafana, avro

Juin 2018 – décembre 2019

Développeur FS / Data Engineer @ UBS

Projet : Construction d'une architecture micro-service d'un projet de détection d'actes de blanchiment des capitaux des clients bancaires

- Développement de scripts Spark pour la récupération et le traitement des données
- Construction d'un moteur capable à affecter un score pour chaque client en fonction de règles métier
- Implémentation d'APIs REST pour exposer les résultats d'analyse
- Mise en place de tests unitaires automatisés
- Mise en place de POC
- Correction de bugs

Environnement : Java 8, Drools 6, Spring Boot 2.0.0, Mongo 3.6, Oracle 11g, Maven 3.5.3, Git, Tomcat 8.5.3, Kafka, Apache Spark 2, Scala 2.12

Novembre 2017 – Mai 2018

Développeur FS B / Data Engineer @ Lafayette

Le projet Smartpass est un outil de recherche permettant de s'informer sur les métiers, les entreprises et les formations

- Refonte d'une application java en python
- Développement de scripts pour la collecte de données

Environnement : Python 3.6, Java 8, Spring Boot 2.0.0, Mongo 3.6, Maven 3.5.3, Git, Tomcat 8.5.3

FORMATION & CERTIFICATIONS

- 2017 : Diplôme National d'Ingénieur en Science Appliquées et Technologique – Spécialité Génie Logiciel et Systèmes d'information – Institut Supérieur d'Informatique

LANGUES

- Anglais : technique