

Jason Mfou'ou Oyono

(+33) 644713939
Mobile dans toute la France
jasonmfououoyono@gmail.com

Data Engineer
Disponible de suite

Portfolio: jasonPortfolio.com
github.com/mojkej
linkedin.com/in/jasonoyono11

Data Engineer expérimenté, spécialisé dans la conception et l'optimisation de pipelines de données, le développement d'API et la visualisation avancée. Capable de transformer des données complexes en informations claires et exploitables pour orienter la stratégie des équipes métiers.

COMPETENCES

Programmation :	Python, Scala, SQL
Outils :	Power BI , Git/Github , Kubernetes, Docker , DBT, PySpark
Base de données :	PostgreSQL, MySQL, MongoDB
Cloud :	GCP, AWS, Databricks
Outils ETL/ELT :	Airflow, Talend, Airbyte
Data Quality :	Soda, Pytest
Communication :	Français (Maternelle), Anglais(B2)
Certifications :	IBM Data Engineering (Coursera)
Soft Skills :	Adaptabilité, Travail d'équipe, Autonome
Centre d'intérêts :	Tennis, Football, Lecture, Veille technologique, Jeux vidéos, Échecs

EDUCATION

Mastère spécialisé Big Data Grenoble INP-Ensimag & Grenoble école de Management	Septembre 2023 — Decembre 2024
Ingénieur de Conception en Génie Informatique École Nationale Supérieure de Polytechnique de Yaoundé (ENSPY)	Septembre 2017 — Septembre 2022

EXPERIENCE PROFESSIONNELLE

Data Engineer Spark Cleantech	Avril 2024 — Octobre 2024 Gif-sur-Yvette, France
---	---

Mission : Optimisation des pipelines de données et visualisation des données

- Mise en place d'un pipeline ELT avec Airbyte Cloud pour centraliser les données générées par les instruments de laboratoire(Réduction des délais d'accès aux données de 50 % grâce à une centralisation efficace.)
- Migration des données locales de PostgreSQL vers une base de données sur OVH Cloud, assurant un accès sécurisé et centralisé.
- Transformation des données via DBT pour répondre aux besoins des expérimentateurs.
- Développement d'une API permettant de récupérer et d'analyser les données stockées sur le cloud, facilitant l'exploration des données.
- Conception d'un dashboard interactif avec Dash, permettant de visualiser les données en temps réel et de suivre les événements expérimentaux. (Utilisé par plus de 5 scientifiques pour analyser plus de 1000 points de données par expérience, améliorant la prise de décision expérimentale.)

Environnement technique : Python , Dash , Airbyte cloud (ELT) , SQL, PostgreSQL , Docker , ovhcloud , DBT , CI/CD, Github, VSCode, Jira, Méthode Agile, Pandas

Data Engineer CAA (Caisse Autonome D'Amortissement)	Octobre 2022 — Septembre 2023 Yaoundé, Cameroun
---	--

Mission : Architecture data pour optimiser la gestion et l'analyse des emprunts public

- Organisation de 5+ ateliers avec les équipes métier, permettant d'identifier et de formaliser les besoins en matière de données financières.
- Développement de dashboards interactifs avec Power BI pour visualiser les KPI financiers.
- Utilisation d'Airflow pour orchestrer les workflows de données, y compris l'extraction, le chargement des données dans BigQuery et la transformation via DBT
- Mise en place de Datadog pour surveiller les pipelines de données et assurer leur disponibilité.
- Formation de 10+ collaborateurs à l'utilisation des nouveaux outils et processus, améliorant l'adoption et l'autonomie des équipes

Environnement technique : Python , GCP, SQL, Datadog, Power BI, Airflow, DBT, CI/CD, Github, FastAPI,Terraform, PySpark, VSCode

Data Engineer

Neero

Janvier 2022 — Août 2022

Douala, Cameroun

Mission : Automatisation de la validation des comptes dans une fintech camerounaise : Extraction d'informations et vérification biométrique.

- Extraction des informations des documents officiels camerounais en utilisant Amazon Textract et des expressions régulières.
- Conception et développement d'un microservice Python orchestrant plusieurs modules de traitement (OCR, reconnaissance faciale, web scraping, analyse vidéo).
- parallélisation du processus de comparaison entre le selfie utilisateur et les images stockées dans Amazon S3 pour accélérer la détection de profils déjà créés.
- Développement d'un système de web scraping avec Selenium pour vérifier des informations spécifiques sur des sites web officiels.

Environnement technique : Python , Django Rest , Selenium , PySpark , AWS (Textract , Rekognition , S3) , Tensorflow , Keras , Computer vision, CI/CD, Github, VSCode

Data Engineer

CAA (Caisse Autonome D'Amortissement)

Janvier 2021 — Septembre 2021

Yaoundé, Cameroun

Mission : Participation à la structuration et à l'automatisation des flux de données financières

- Mise en place de scripts Python et de requêtes SQL pour automatiser certaines extractions et chargements de données.
- virtualisation des environnements de test (VMware, vCenter)
- configuration des accès utilisateurs via Active Directory.

Environnement technique : Python , SQL, PostgreSQL, Github, Active Directory, CI/CD, Pytest

PROJETS PERSONNELS

Projet 1 : Pipeline ELT de données YouTube avec Airflow et Docker

 GitHub

- Mise en place d'un pipeline ELT automatisé pour extraire, charger et transformer les données issues de l'API YouTube.
- Intégration avec Apache Airflow pour l'orchestration des tâches et planification des extractions.
- Stockage et gestion des données dans PostgreSQL.
- Vérification de la qualité des données à l'aide de Soda.
- Test unitaire et test d'intégration avec pytest.

Environnement technique : Python, CI/CD, Docker, Bash, Soda, Pytest, Apache Airflow, PostgreSQL, API YouTube, VSCode, Postman, Github

Projet 2 : Développement d'un pipeline ETL et analyse des performances du championnat espagnol

 GitHub

- Conception d'un pipeline ETL complet en Python pour extraire, transformer et charger les données des saisons de la Liga espagnole (2012–2021)
- Extraction automatisée et récursive des fichiers CSV depuis une arborescence de dossiers, avec journalisation des étapes et gestion des erreurs dans un fichier de logs.
- Transformation des données : parsing des dates, séparation des scores, création d'une colonne de résultat (victoire, défaite, nul) et nettoyage des anomalies.
- Analyse avec PySpark pour agréger les performances par équipe et par saison, et calculer des indicateurs statistiques (buts marqués, taux de victoire, etc.)

Environnement technique : Python, Pandas , PySpark , CI/CD, Github, jupyter notebook, VSCode