

# ASSANE DIAKHOUMPA

## Data Engineer

### CONTACT

+221 77 092 07 40

assanediakhoumpa4@gmail.com

Rufisque-Dakar

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

### COMPETENCES

#### Langages & Programmation

- Python (Pandas, PySpark)
- Scala

#### Big Data & Traitement Distribué

- Hadoop (HDFS, YARN)
- Apache Spark (batch & streaming)
- Kafka (producers/consumers, topics, partitions)

#### Bases de Données & Stockage

- SQL : MySQL, PostgreSQL
- NoSQL : MongoDB, Cassandra, Redis
- DynamoDB (LocalStack, AWS)

#### Pipelines & Orchestration

- Airflow
- CI/CD : GitHub Actions, GitLab CI
- Docker & Docker Compose

#### Cloud & Streaming

- AWS (Kinesis, Lambda, S3, DynamoDB)
- LocalStack pour simulation cloud
- Architecture streaming temps réel

### CERTIFICAT

- Introduction to Data Engineer (IBM) Coursera
- Pandas, Matplotlib, Seaborn, Numpy Coursera

### PROFILE

Suite à mon **BUT STID**, je me suis orienté vers le **Data Engineering**. Je maîtrise la collecte, le nettoyage et la transformation des données, avec une solide base en **statistiques** et en **informatique Décisionnel**. Je travaille avec **Python**, **SQL** et **R**, et je suis à l'aise avec la création de **pipelines**, l'**intégration** de données et la mise en place de solutions fiables pour exploiter les données.

### EXPERIENCE

**Étude statistique et modélisation du paludisme** 2023 - 2024  
Mission Data Analyst - Centre de Santé de Rufisque

- Collecte et ingestion de données sanitaires provenant de **4 postes de santé, avec gestion d'historiques multi-années**.
- Nettoyage, transformation et structuration des données via des pipelines automatisés (**Python**, **Power Query**).
- Réalisation d'analyses exploratoires pour valider la qualité des données (**tendances, saisonnalités, répartition par sexe et âge**).
- Visualisation et monitoring de la qualité des données avec Python (**pandas**, **matplotlib**).
- Mise en place d'un pipeline de prévision basé sur des modèles de séries temporelles (**ARIMA**) pour anticiper les cas de paludisme.
- Production d'**insights** opérationnels permettant d'améliorer la **prise de décision** en santé publique grâce à des données fiables et bien structurées.

**Data Engineer - Pipeline Météo & Aérien (Projet DataFlow360)** 2025 - en cours

- Collecte et ingestion des données météo et aériennes provenant d'aéroports d'Afrique de l'Ouest (**APIs**, **scraping Selenium**, **données synthétiques** et **fichiers plats**).
- Nettoyage, transformation et structuration des données via des **pipelines Python** automatisés pour assurer **qualité**, **cohérence** et **réutilisabilité**.
- Mise en place d'une **architecture de streaming cloud simulée**, basée sur **AWS Kinesis** (via **LocalStack**) pour l'ingestion **temps réel**, permettant d'absorber des **flux continus**.
- Traitement des flux en temps réel à l'aide de fonctions **AWS Lambda** simulées, orchestrant la validation, la **normalisation** et le routage des données.
- Stockage des données traitées dans **DynamoDB (LocalStack)**, conçu pour supporter un modèle scalable et orienté événements.
- Conteneurisation** complète du projet avec **Docker** pour uniformiser l'environnement et simplifier les déploiements.

### EDUCATION

**BUT Statistique & Informatique Décisionnelle** 2023 - 2024  
(équivalent Licence Pro Data Science)

université Alioune de bambey / UADB

**Certification en Development Data** 2025

Orange Digital Cenger / ODC