



✉ babacarsarr.tech@gmail.com  
📍 Île-de-France  
📅 23 ans  
☎ 07 75 80 67 99  
🔗 <https://babacarsar.github.io/portofolio>

## Réseaux sociaux

🐦 @Babacarsar  
in @babacar-sarr-a3888a247

## Compétences techniques

### Cloud, ETL et Big Data

- Orchestration : Apache Airflow, AWS Lambda
- Intégration de données : ETL avec KNIME, Talend, Dataiku
- Data Warehouse : Snowflake, PostgreSQL, SQL Server, Oracle
- Bases NoSQL : MongoDB, Cassandra
- Stockage cloud : Amazon S3, MinIO, Databricks

### DevOps et CI/CD

- Git, CI/CD, GitHub, GitLab,
- Docker, Kubernetes
- Terraform Jenkins, Ansible

### Langages et Programmation

- Python avancé
- SQL, R
- Java, VBA
- Shell Script

### Mathématiques et Statistiques

- Statistiques descriptives, régressions, corrélations,
- Modèles supervisés (clustering, classification)

### Outils et Environnements

- Jupyter Notebook, Google Colab
- Power BI, Tableau, Excel avancé, Looker
- QGIS

## Soft Skills

### Storytelling

### Communication

### Collaboration

adaptabilité et gestion de projet agile

## Langues

Anglais  
B2

# Babacar SARR

## Recherche d'une alternance en tant que Data Analyst à partir de septembre 2025, avec un rythme de 2 semaines en entreprise et 1 semaine en formation

Passionné par la gestion et l'analyse de données, je conçois des pipelines de traitement performants et transforme les données en insights stratégiques. J'ai réalisé plusieurs projets en ingestion de données Big Data, automatisation ETL et reporting interactif sous Power BI et Tableau. Je souhaite aujourd'hui mettre à profit mes compétences à la fois en data engineering et en data analysis pour accompagner les entreprises dans la valorisation de leurs données.

## Expériences professionnelles

### Stage en Data Analyst

De novembre 2023 à mai 2024 **Malick Galaxie Groupe** Dakar



- Collecte automatisée et intégration de données multi-sources** via KoboToolbox, avec extraction des jeux de données sous format CSV et ingestion dans des modèles de données sous Excel et Power Query.
- Nettoyage, normalisation et transformation des données** : suppression des doublons, gestion des valeurs manquantes, typage et structuration pour analyse tabulaire.
- Conception et optimisation de **modèles de données relationnels** pour reporting sous Power BI, en intégrant des relations entre tables de faits et dimensions.
- Développement de **tableaux de bord dynamiques Power BI** avec indicateurs clés de performance (KPI) : taux de désenclavement, évolution du revenu moyen, couverture en infrastructures de base, évolution des indicateurs sur plusieurs périodes.
- Production d'**analyses exploratoires et statistiques descriptives** pour identification de patterns et outliers, via Power BI et Excel.
- Rédaction de rapports analytiques et data storytelling** à destination des décideurs, accompagnés de cartographies thématiques et visualisations multicritères.

## Projets

### Exploitation d'une plateforme Big Data à partir d'une situation réelle

- Automatisation** de la récupération de données toutes les 2h via **AWS Lambda**, à partir des API **OpenWeather** (météo) et **AQICN** (qualité de l'air), orchestrée avec **CloudEvents**.
- Stockage des données brutes dans un **Data Lake (Amazon S3)**, traitement ETL avec **AWS Glue**, et chargement dans un **Data Warehouse PostgreSQL (Amazon RDS)**.
- Mise en place d'une infrastructure cloud évolutive via **Terraform (Infrastructure as Code)**.
- Surveillance du pipeline avec **AWS CloudWatch** (logs, métriques, alertes).
- Conception de **tableaux de bord interactifs** avec **Power BI** (KPI météo et pollution, seuils d'alerte).
- Réalisation de visualisations géospatiales avec **QGIS** pour localiser les zones les plus impactées.

### Analyse Big Data des trajets VTC à New York

- Collecte automatisée** des données VTC via **Python**, stockage dans un **Data Lake MinIO** (compatible S3).
- Traitement ETL **visuel** avec **KNIME** : nettoyage, enrichissement et structuration des données.
- Chargement dans un **Data Warehouse PostgreSQL** dockerisé, avec une **modélisation en flocon** pour optimiser les requêtes analytiques.
- Mise en place de l'**orchestration complète** avec **Apache Airflow** pour automatiser l'ensemble du pipeline.
- Pipeline opérationnel traitement de 120 000 lignes sans interruption, qualité vérifiée avec **Soda Core**.
- Conception de **tableaux de bord dynamiques** avec **Power BI** pour analyser les flux.

## Parcours Académiques

### MSc Expert Ingénierie des Données

De 2024 à 2026 **EPSI, Courbevoie** Paris



### Licence Génie Logiciel

De 2020 à 2023 **Unipro** Dakar

