|  |  |
| --- | --- |
| **Babacar GUEYE**  [**Lien Page LinkedIn**](https://www.linkedin.com/in/babacargueye1/)  [**Lien Page GitHub**](https://github.com/humbledDS) | **Data Engineer Cloud & Big Data| Consultant Data**  **Tel : +33 6 79 81 97 72**  **Email :** [**babacar.work2024@gmail.com**](mailto:babacar.work2024@gmail.com)  **Adresse : 12 rue Capitaine Guynemer, Rosny Sous-Bois, 93110** |
|  |  |

Ingénieur Data spécialisé dans le développement de pipelines et d’architectures cloud. Expertise dans la conception de solutions data end-to-end, combinant ingénierie backend, modélisation prédictive, l'automatisation des processus data et l'optimisation des performances à grande échelle.

**EDUCATION**

**Université de Montpellier** | *DU Big Data, Data Science et Analyse des risques (****Mention TB****)* | **10/2023 – 09/2024**

**Université** **de Caen** | *Master 2 Statistiques Appliquées et Analyse Décisionnelle (****Mention Bien****)* | **09/2021 – 09/2023**

**Université de Caen** | *Licence 3 Mathématiques Appliquées et Licence 3 Informatique* | **09/2019 – 09/2021**

**Certifications**: AWS Cloud Practitioner, Microsoft Azure Fundamentals, Databricks Associate Developer

**COMPETENCES**

|  |
| --- |
| **Data Engineering**: Python (Pandas, PySpark), Apache Airflow, Apache Kafka, Spark, Databricks, Snowflake, Dbt **Cloud & DevOps**: AWS (S3, EC2, Lambda, Glue, Redshift), GCP, Docker, Kubernetes, Terraform  **Modelisation & Analytics**: Modélisation dimensionnelle, OLAP, Data Warehousing, lakehouse, SQL,NoSQL, ELT, ETL **CI/CD & Versioning**: Git, GitHub Actions, GitLab CI  **Visualisation & BI**: Streamlit, PowerBI, Tableau, Looker |

**EXPERIENCES**

**Ingénieur Data & IA | Freelance – Projet R&D pour un cabinet de Conseil en Strategie****| 10/2024 – 05/2025**

Conception d’un outil de **pré due-diligence automatisée** pour accélérer les analyses financières des consultants.

* Création d’un pipeline d’ingestion périodique de données via les API de sources officielles (INSEE, DataGouv) et de rapports financiers.
* Agrégation et enrichissement de la data (chiffres clés, dirigeants, M&A, rapports financiers) via Spark & Python.
* Mise en place d’un moteur de recherche et d’un scoring automatisé de la santé financière des entreprises cibles.
* Interface web pour restitution synthétique des insights financiers et business, avec mise à jour automatique.
* Projet ayant réduit le temps moyen d’analyse préliminaire de 60%.

**TrouveTout (Start-up) – *Data Engineer ETL & Cloud*****| 07/2023 – 05/2024**

* Conçu et implémenté des pipelines ETL avec Python et Airflow, réduisant le temps de traitement de 30%.
* Développé une architecture backend scalable en Django pour optimiser la gestion et l'intégration des données.
* Mise en place d’un CI/CD complet (Docker, GitHub Actions) et d’environnements (Dev, Staging, Prod)
* Surveillance proactive de la qualité et fiabilité des données (monitoring, alerting)

**Orange (Telecom) – *Stage Data Science / Data Engineering*****| 11/2022 – 05/2023**

* Gestion de Projet : Amélioration du taux de satisfaction base client B2B au sein du département Marketing
* Implémentation d'un modèle prédictif de churn améliorant la précision de 15%.
* Migration de l'infrastructure sur AWS (S3, Glue, Redshift), optimisant la performance et les coûts.
* Analyse des données clients (200k+ entrées) pour identifier des segments à risque.

**Laboratoire de Mathématiques Nicolas ORESME** | Stage *Data Analyst* **| 03/2021 – 07/2021**

* Optimisé les pipelines de traitement de données, réduisant le temps d'exécution de 20%.
* Effectué des analyses statistiques avancées, incluant la PCA pour la réduction dimensionnelle.
* Conçu des dashboards interactifs avec SAS Visual Analytics pour la visualisation de résultats.

**PROJETS PERSONNELS | LANGUES |CENTRES D’INTERET**

* **Stock Markets Analytics + Dashboard** : Etude et Analyse sur 30 ans de données du S&P 500, Prédictions et Stratégie de portefeuille basé sur un croisement de la data et de conjecture macroéconomique. (Lien)
* **Cloud ML Pipeline :** Déploiement automatisé de modèles sur AWS (MLflow, Docker).
* **LLM RAG :** Système intelligent de question/réponse sur documents (LangChain, Pinecone, OpenAI).

**Langues : Français** (natif), **Anglais** (Courant)

**Centres d’intérêt** : Football, lecture, Gaming, Food, Art, Gym, et voyages