

Abidi Houssem

Joinville-le-pont, Île-de-France | houssemabidi2010@gmail.com | +[33] 0651081363 |

linkedin.com/in/houssemabidi512

EXPÉRIENCE PROFESSIONNELLE

DATABEANS (partenaire databricks) | INGÉNIEUR BIG DATA

11.2021 – 06.2024

- Développer la solution Lighthouse qui sert à:
 - Surveiller la santé des tables du Lakehouse
 - Examiner la disposition des données et les schémas d'accès
 - Fournir des insights essentiels sur la qualité du regroupement des données.
 - Suggérer les optimisations nécessaires pour améliorer l'efficacité opérationnelle du système
 - Améliorer la performance du Lakehouse et réduire son coût total de possession.
 - Travailler dans une équipe Big Data composée d'un tech lead et de quatre ingénieurs de données.
 - Contribuer à la conception de la solution
 - Identifier les métriques permettant la surveillance de l'état de l'ordre des données
 - Contribuer à la création d'un pipeline de données (ETL) avec une architecture multi hop (Bronze, Silver, Gold)
 - Développer des jobs Spark pour faire l'ingestion des données et alimenter les couches Bronze et Silver en Streaming.
 - Développer des jobs Spark pour faire les agrégations nécessaires et alimenter la couche Gold en Batch.
 - Optimiser et faire le tuning des jobs Spark.
 - Élaborer et développer un tableau de bord Power BI visant à mettre en lumière les indicateurs clés de performance (KPI) qui vont assister les utilisateurs dans la sélection des opérations de maintenance sur le lakehouse (optimize, Z-order, compact, repartition, etc.).
 - Déploiement de la solution sur Azure Microsoft Fabric • Effectuer les tests unitaires et les tests d'intégration.
 - Participer aux cérémonies Agile: réunions quotidiennes, planification des sprints, rétrospectives...

Environnement technologique Étant: Databricks, Microsoft Azure, Microsoft Fabric, Delta Lake, Scala, Apache Spark, Git

- Rédiger et réviser les articles et les publications Big Data publiés sur la page Medium et la page LinkedIn de DataBeans (département Big Data d'IMBA AFRICA) :
 - <https://medium.com/@databeans-blogs>
 - <https://www.linkedin.com/company/databeans>

- Développer un benchmark pour comparer la performance Delta Lake vs. Hudi vs. Iceberg. (Résultats publiée dans the DATA +AI summit 2022)
 - Travailler dans une équipe Big Data composée d'un directeur technique et de trois ingénieurs de données.
 - Contribuer à la conception de la solution
 - Déployer et intégrer la solution en utilisant Terraform sur AWS :
 - AWS EMR pour création du cluster
 - AWS RDS comme external hive metastore
 - AWS S3 pour les cluster logs
 - Participer aux cérémonies Agile: réunions quotidiennes, planification des sprints, rétrospectives...

Environnement technologique Étant: Terraform, AWS (EMR, RDS, S3, IAM, VPC, EC2), Git

- Développer une solution pour rendre les multi-statement transactions possible sur Delta Lake.
 - Travailler dans une équipe Big Data composée d'un responsable technique et un ingénieur Big Data
 - Contribuer à la conception de la solution
 - Développer la solution de bout en bout
 - Faire l'optimisation et le tuning des jobs spark
 - Déployer et intégrer la solution sur les plateformes AWS EMR et Databricks / pour rendre le projet could agnostic
 - Effectuer les tests unitaires et les tests d'intégration
 - Participer aux cérémonies agiles : daily meeting, sprint planning, rétros...

Environnement technologique Étant: AWS EMR, DataBricks, Spark Scala, Delta Lake, Git

- Développer une application web qui permet :
 - Aux freelances de trouver des opportunités de travail en ligne.
 - D'utiliser l'extraction de données et le scraping pour collecter des informations sur les opportunités disponibles sur différents sites web.
 - Construire une plateforme conviviale en utilisant Flask et BeautifulSoup pour collecter et présenter les offres d'emploi freelance.
 - Collaborer avec des équipes interfonctionnelles pour recueillir les exigences, assurant ainsi que l'application correspond aux besoins et attentes des utilisateurs.
 - Effectuer des tests unitaires et des tests d'intégration pour garantir la fiabilité et la performance de l'application web.

Environnement technologique Étant: Python, Flask, BeautifulSoup, Angular

COMPÉTENCES TECHNIQUES

Langages de Programmation

Scala, Python, Java , SQL

Technologies Big Data

Apache Spark, Databricks, YARN,
Delta Lake, Azure, Microsoft Fabric

Visualisation De Données

Power BI

LANGUAGES		
Anglais (Courant)	Arabe (Courant)	Français (Courant)

CERTIFICATIONS
Databricks Lakehouse Fundamentals
Databricks GCP Platform Architect
Databricks Platform Architect
Databricks AWS Platform Architect
Databricks Platform Administrator

EDUCATION	
ESPRIT Computer Engineering Degree	Ariana, Tunisia 11.2020 – 06.2023
ISET Higher Diploma in Information Technology	Kasserine, Tunisia 10.2017 – 10.2020
