

EL HAFA Hajar — Data Engineer

Île-de-France

☎ +33 6 65 64 91 43 • ✉ hajarelhafa.pro@gmail.com

Expérience Professionnelle

Shortways

Data Engineer

Paris

Avril 2024 – Présent

Contexte : Orange – Projet Smart Ticketing et assistant Shortways

- Conception de pipelines de données sur AWS pour la gestion du projet Smart Ticketing.
- Création et automatisation de jobs ETL avec AWS Glue (PySpark) pour le traitement des données, comprenant :
 - Extraction des données brutes depuis AWS S3.
 - Transformation des données en nettoyant et enrichissant les ensembles à l'aide de scripts PySpark.
 - Chargement des données préparées vers AWS Athena pour une analyse rapide.
- Gestion et optimisation du Data Lake pour un stockage structuré et conforme aux exigences RGPD, notamment :
 - Mise en place de partitions pour organiser les données par date et type d'événement.
 - Implémentation de politiques IAM pour sécuriser l'accès aux données sensibles.
- Nettoyage, contrôle et validation des données pour garantir leur fiabilité et conformité :
 - Utilisation de requêtes SQL dans AWS Athena pour identifier les anomalies et incohérences.
 - Vérification automatique des schémas des fichiers JSON et CSV ingérés.
- Mise en place de processus CI/CD pour l'automatisation et le déploiement des workflows de traitement de données :
 - Utilisation de **GitHub Actions** pour le contrôle des versions et l'automatisation des tests.
 - Configuration de pipelines CI/CD avec **AWS**.
 - Déploiement automatisé des jobs AWS Glue en environnement de production.
 - Supervision des workflows pour garantir une intégration continue et des déploiements fiables.
- Ingestion des données depuis AWS Data Lake vers les ensembles de données des tableaux de bord, en suivant ces étapes :
 - Récupération des fichiers sources depuis AWS S3 à l'aide de scripts Python.
 - Transformation des données dans un format tabulaire optimisé avec PySpark.
 - Chargement des données nettoyées dans des tables Athena pour l'exploitation directe par les analystes.
- Développement et automatisation de tableaux de bord interactifs sur AWS QuickSight :
 - Création de visualisations dynamiques pour suivre les indicateurs de performance du projet Smart Ticketing.
 - Intégration de fonctions avancées comme les filtres interactifs et les alertes automatiques basées sur les seuils définis.
 - Collaboration avec les parties prenantes pour ajuster les visualisations selon leurs besoins.

Technologies/Outils : Python, PySpark, AWS (S3, Glue, Athena, QuickSight, boto3), Docker, Kubernetes, Git, GitHub, MySQL, TypeScript, Elastic Search, Scala.

Edico

Business Analyst

Gennevilliers

Mars 2023 – Septembre 2023

Contexte : Gestion et optimisation de projets techniques pour garantir des solutions adaptées et des livrables de qualité

- Gestion de projet en cherchant la solution technico-économique pour des clients comme **Orange, TDF, Covage**.
- Gestion financière, incluant la gestion des **attachements clients**.
- Exploitation d'outils et interfaces client (**SHAREPOINT, JIRA, GISNET**).
- Production des calculs de charge des supports techniques (**CAP FT, COMAC**).
- Garantie des délais de livraison et de la qualité du produit final.
- Analyse de la structure d'un réseau existant cuivre afin de créer ou modifier le réseau FTTH.
- Réalisation des récolements, c'est-à-dire correction des incohérences entre les ingénieries effectuées en bureau d'études et la réalité du terrain.
- Mise en place d'une chaîne de traitement automatisée pour les contrôles des plans de déploiement FTTH (scripts Python).

Technologies/Outils : Networks, QGIS, AUTOCAD, Netgeo, ESRI, FME, ArcGIS, Python, PostGIS .

- Travail sur l'intégration approfondie des données SIG pour une meilleure visualisation et analyse des infrastructures de télécommunication.
- Préparation de rapports périodiques sur l'avancement des projets, en se concentrant sur les aspects financiers et opérationnels du déploiement du réseau fibre optique.
- Dimensionnement du réseau fibre optique.
- Traitement de courbes et mesures de fibre optique.

Technologies/Outils : QGIS, Microsoft Excel, FiberGIS.

Formation

Ecole PMN

Nanterre*Master 2: Data Engineer*

Data Engineering, Cloud Computing (Azure), outils DevOps (**Docker, Kubernetes**), et programmation fonctionnelle (**Scala**).

EIDD

Paris*Diplôme d'Ingénieur Systèmes Physiques*

Programmation, Méthodes Numériques, Statistiques, Mathématiques Appliquées, Traitement de Signal.

Compétences Techniques

Langages: Python, Spark, Scala, R, C, HTML, CSS, JS.

Outils: Docker, Kubernetes, Jenkins, Hadoop, Power BI, MySQL, VMWare, Talend, Apache Kafka.

Plateformes: AWS, Azure, GCP.

Méthodologies: **Méthodes agiles** (Scrum, Kanban), gestion des sprints, cérémonies et outils associés (Jira, Confluence).

Certifications

Professional Scrum Master I (PSM I): Certifié

Microsoft Power BI Data Analyst Associate: En cours

AWS Certified Data Engineer - Associate: En cours

Langues

Français: Bilingue

Anglais: Courant