

# Ghania DZIRI

## DATA ENGINEER

Ingénieure en Informatique, avec plus de 4 ans d'expériences dans la data, de l'intégration jusqu'à la restitution métier en passant par l'analyse et le traitement de la donnée brute. Actuellement, je suis à la recherche de nouveaux challenges.



dzirighania@gmail.com



Ile de France



07 51 11 69 99

### Compétences

#### Big Data :

Hadoop (HDFS), Hive (HQL), Oozie, Shell

#### Langages de

##### Programmation :

Python, C, JavaScript, HTML, CSS

#### Frameworks :

Apache Spark (PySpark, Spark SQL), JupyterLab (VsCode, Apache Superset), AnaConda (Jupyter Notebook, Spyder, RStudio )

#### Outils:

JIRA, GitLab, ServiceNow, MobaXterm, DBeaver

#### Bibliothèques :

Pandas, Numpy, Sikit-Learn, Mlib, Keras, TensorFlow, Matplotlib

#### Base de Données :

SQL

#### Systèmes d'exploitation :

Unix, Windows

#### Méthodes :

Agile Scrum, Cycle V

#### DataViz

PowerBI, Dataiku

### Langues

#### Français

#### Anglais

Niveau Professionnel

### Atouts

#### Autonomie

#### Esprit d'équipe

#### Rigueur et efficacité

### Diplômes et Formations



#### Master Big Data et Fouille de Données

Université PARIS 8, Seine-Saint-Denis

De 2017 à 2019



#### Licence Micro Informatique et Machines Embarquées

Université PARIS 8, Seine-Saint-Denis

De 2016 à 2017



#### Licence Science et Technologie Spécialité Automatique des Systèmes

Université Mouloud Mammeri, Tizi Ouzou, Algérie

De 2013 à 2016

### Expériences Professionnelles



#### Data Engineer

Depuis mai 2022 (2 ans+)

SOCIETE GENERALE ASSURANCES, La Défense

#### Projet :

**Réalisation de Pipelines de traitement de données d'Assurances Prévoyance sous une architecture Hadoop**  
**Maintenance corrective et évolutive des différents traitements développés.**

De Juin 2023 à Aujourd'hui :

#### Data Engineer au sein de l'équipe DataHub

- Travail en mode agile en assistant aux différentes cérémonies de la FT Prévoyance
- Conception et mise en œuvre de pipeline de données : nettoyage, uniformisation et transformation de données au sein des Datamarts en partant des données brutes déversées depuis les silos de gestion, stockés sur le Data Lake
- Mise en œuvre de l'ensemble des règles de gestion spécifiques pour répondre au besoin des différentes directions métier
- Réalisation des tests unitaires en utilisant le Query Unit
- Automatisation et Ordonnancement des traitements
- Gestion des mises en production des livrables de différents Datamarts, Clients, Evènements et Sinistres
- Gestion des incidents de production et suivi du RUN
- Revue du code pour les autres MOE de l'équipe
- Mise à jour de la documentation du projet.

De Mai 2022 à Juin 2023 :

#### Data Engineer au sein de l'équipe EDGE (Datalake)

- Participation aux différentes cérémonies agiles de la FT EDGE
- Conception d'un traitement pour chargement des données en mode suppression dans le cadre d'amélioration de gestion des données stockés dans le data lake
- Contribution au développement du traitement générique de conversion des tables internes en tables externes (tables hive)
- Evolution du script de chargement de différentes sources de données brutes reçues en entrée du datalake
- Garantir la maintenabilité et les performances des programmes développés en optimisant les requêtes SQL et les scripts développés en SHELL.
- Réaliser des tests unitaires et d'intégration
- Rédaction et mise à jour de la documentation.

**Environnement Technique :** HADOOP ( Hive, Oozie), Shell, Gitlab, Jira, Nexus, Maven, SONAR, DBeaver, MobaXterm.

## Expériences Professionnelles

### Data Engineer

De mai 2021 à mars 2022 (10 mois)

FRANCAISE DES JEUX, Aix-en-Provence

**Projet : Cadrage et réalisation des évolutions d'une infrastructure de type Datalake sur la base d'une architecture Hadoop (distribution HDFS) et mise en place d'une plateforme Data Science**

- Revue des sprint et planification des User Story en mode Agile
- Création de pipeline de traitement de données batch en utilisant Spark
- Suivi du Run afin de s'assurer du chargement des données en entrée du Data Lake.
- Gestion des configurations et de l'infrastructure en utilisant SaltStack
- Création des pipelines CI/CD en utilisant Gitlab Ci
- Monitoring de la plateforme (Scripting SHELL)
- Partitionnement des données dans Hdfs
- Gestion des Incidents

**Environnement technique :** *Apache Hadoop, Apache Spark, Gitlab, HIVE, HUE, Git, Kafka, Jira, ServiceNow*

**Projet : Responsable Opérationnelle d'Application sur une solution Analytics ayant pour objectif la lutte contre la fraude dans le domaine du jeu**

- Suivi du Monitoring de l'application SaferGame (TDB)
- Industrialisation de nouveaux flux de données (batch et streaming) pour alimenter les différents Use Case Métier
- Investigation sur les problématiques de retard d'ingestion de données et proposition des solutions pour leurs implémentations
- Communication auprès des métiers (remontée des nouveaux cas frauduleux)
- Gestion des Incidents

**Environnement technique :** *Control-M, DataStudio, SQL Developer (ETL), ServiceNow*

### Data Engineer

De décembre 2019 à décembre 2020 (1 an)

CREDIT AGRICOLE - CAPS -, Saint-Quentin-en-Yvelines

**Projet : Fiabilisation et Amélioration du système de reporting**

**Objectif :** La division CAPS souhaite agréger les informations de transactions des caisses régionales afin d'industrialiser la méthode de production des différents tableaux de bord

- Recueil et analyse du besoin
- Import et chargement des données dans HDFS
- Préparation et nettoyage des données de transactions
- Traitement des données d'une manière distribuée avec Spark
- Mise en œuvre de l'ensemble des règles de gestion spécifiques
- Développement du traitement d'agrégation ( PySpark )
- Automatisation du processus du traitement de données
- Rédaction de documentations sur le projet

**Environnement technique :** *Apache HADOOP (PySpark, Hue, HDFS), Distribution MapR, Scripting Shell*

## Expériences Académiques

### Consultante Big Data

De février 2019 à août 2019 (6 mois)

MICROPOLE, Levallois-Perret

**Objectif : Prototypage, étude et implémentation des algorithmes de Data Science dans un contexte Big Data (analytics, machine learning)**

- Recueil et analyse des besoins utilisateurs
- Interrogation des différentes sources de données émises par les capteurs
- Préparation et analyse des données d'exploitation PySpark (DataFrames)
- Étude des algorithmes d'analyse de données géométriques (clustering, calcul d'enveloppes, reconnaissance de forme)
- Évaluation et validation des algorithmes
- Implémentation du modèle de Machine Learning (DBSCAN, ConvexHull)
- Rédaction du Rapport de fin d'études sur le Projet d'entreprise

**Environnement technique :** *Linux, Anaconda (Spyder), Python (Mlib, Sickit-Learn).*

### Stagiaire en Master 1 Big Data et Data Science

D'avril 2018 à août 2018 (4 mois)

Laboratoire LIASD – Université paris 8, Seine-Saint-Denis

**Objectif : Apprentissage profond (deep learning) pour la classification d'images**

- État de l'art sur l'intelligence Artificielle et différents modèles du deep learning
- Extraction du jeu de données d'images sur la plateforme Kaggle
- Validation du modèle d'apprentissage "CNN"
- Prétraitement du Modèle d'apprentissage (Entraînement 80%, Test 20%)
- Développement du modèle pour la Classification multi-classes
- Rédaction du Rapport de stage Master 1

**Environnement technique :** *Linux, PyCharm, Python (TensorFlow, Keras), Scripting Shell*