IBRAHIMA DIATTARA

Email: [ibrahima-diattara1992@hotmail.com](mailto:ibrahima-diattara1992@hotmail.com)

TEL:06 05 72 56 05

**Certifié Microsoft Azure Data Engineer | Google Cloud Professional Data Engineer | Databricks Spark Scala**

# Parcours Professionnel

## EDF depuis Octobre 2023

**Rôle** : Consultant Big Data

**Projet TEGG.** L’objectif de ce projet est de collecter les données des capteurs piézomètres, d'appliquer des règles métiers à ces données, et de générer des rapports d'analyse

### Réalisations :

* Récupérer les données dans Amazon S3 et Kafka
* Contrôler la qualité des données
* Appliquer des règles métiers
* Déposer les données dans Neo4j

### Environnement technique :

Nifi, Java, Neo4j, Spark/Scala, kafla

## SOCIETE GENERALE JANVIER 2021 – Octobre 2023

**Rôle** : Consultant Big Data

**Projet kafka Consumer Lag.** L’objectif de ce projet est de surveiller les retards des consommateurs d’Apache Kafka à travers des tableaux de bord et des systèmes d’alerte afin d’éviter toute perte de données

### Réalisations :

* Installer Jolokia et Kafka lag Expoter
* Faire le Tuning des processors Nifi et des partitions kafka
* Monitorer la consommation de KSQLDB
* Récupérer les metrics de Kafka

### Environnement technique :

Nifi, Kafka, Jolokia, Kafka Lag Exporter, KsqlDB Grafana , Druid , Telegarfe, Prometheus,

**Projet DHR-Usage-Layer.** L’objectif de ce projet est de rendre exploitable la BV (Business View) technique de la DRH

### Réalisations :

* Automatiser la création ou la suppression des tables hive
* Mettre à jour les partitions des tables hive selon une source de données
* Collecter les données de Hive et les inscrire dans une base de données PostgreSQL
* Indexer les logs applicatifs dans Elasticsearch

### Environnement technique :

Spark/Scala, HUE/hive, kafka, Elastcisearch/kibana

**Projet Upgrade HDF.** L’objectif de ce projet et d’étudier les impacts de l’upgrade HDF Réalisations :

* Etudier l’impact de l’upgrade sur les flows nifi
* Etudier le comportement de Kafka après l’upgrade
* Optimiser les flux de données après upgrade

### Environnement technique :

Python, NIFI, Kafka, Ranger, ElasticSearch/Kibana, HDFS, Fluentd, Telagraf , KSQLDB

## UNIVERSITE PARIS SORBONNE DEPUIS JUILLET 2021

Rôle : Professeur Data Engineering (Databricks, Google cloud , Azure, Spark/Scala, Nifi, KAFKA ELK)

## SNCF FÉVRIER 2018 -JUIN 2020

**Rôle** : Consultant Big Data

### Projet HDF :

Dans le cadre du projet de migration Azure de la plateforme Big Data SNCF, il a été décidé de démarrer un chantier spécifique à l’ingestion de données en utilisant HDF

### Réalisations :

* + Mettre en place un système d’ingestion avec Nifi
  + Sécuriser la plateforme d’ingestion avec Keytool
  + Mettre en place un système d’alerte et de monitoring en utilisant grafana et InfluxDB

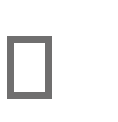
### Environnement technique:

HDF(Nifi, Kafka, zookeeper, grafana, ambari), influxDB, keytool

### Projet DQM :

L objective de ce projet est de développer une application transverse qui permet de transférer les données d'input vers raw

### Réalisations

* + Appliquer certaines vérifications sur les données (pas de colonne manquante…)Faire des transformations de format (CSV to JSON, JSON to parquet, XML to CSV, ...)
  + Orchestrer les Jobs Spark/Scala via Rundeck en utilisant azure service bus comme système d’alerte

### Environnement technique :

Spark, Scala,maven , Jenkins,Nexus, Intellij ,HDFS, Azure service bus et Rundeck

### Projet Sillon :

Estimer les coûts de péage des sillons selon la définition des différentes redevances, suivre et contrôler la facturation des sillons (infrastructure des trains)

### Réalisations :

* + Récolter les données avec NIFI
  + Stocker les données dans HDFS
  + Développer une application Spark/Scala pour nettoyer et agréger les données

### Environnement technique :

Nifi, HDFS/ HADOOP, Spark,Scala ,Maven , Github Jenkis, Nexus

**UNIVERSITE DE THIES (SENAGAL) AVRIL 2020 JANVIER 2022**

Rôle : Professeur Spark Scala , ETL (Nifi) et Machine Learning/MLib

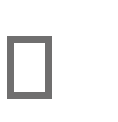
## RENAULT OCTOBRE 2016 - OCTOBRE 2017

### Rôle: APPRENTI ARCHITECTE BIG DATA

**Projet: Rebook**

Optimiser les tâches des garagistes, pour que ces derniers passent moins de temps autour des véhicules lors des contrôles

### Réalisations :

* + Récolte des données(Nifi),
  + Stockage données chaudes(ElasticSearch),
  + Stockage données froides (HDFS/ HDOOP) Utiliser kibana pour la visualisation

### Projet: Maintenance prédictive

Les robots reçoivent pour exécuter leurs tâches. Il arrive que ces robots s’arrêtent ou tombent en panne, bloquant ainsi toute la chaîne de production. Les besoins de l’équipe maîtrise d’œuvre manufacturing BI sont doubles : Quantifier les arrêts sur panne par tranche de durée

Prédire les pannes

### Réalisations :

* + Récolte des données(Nifi)
  + Stockage données (HDFS/ HDOOP)
  + Etablir un modèle statistique pour détecter des séquences de pannes à l’aide de machine learning (Zeppelin/Spark MLib /Scala)
  + Quantifier les arrêts sur panne par tranche de durée avec Hive en utilisant HUE

### Environnement technique :

* + HDF(nifi, kafka, zookepper, Storm, ambari)
  + HDP (HDFS, Spark, Zeppelin,Oozie, Sqoop, Ranger, Hive,HBase)
  + ELK(ElasticSearch, logstash &Kibana)
  + HUE : Hadoop User Experience

# Parcours Professionnel - Stage

## SAFRAN AVRIL 2016-SEPTEMBRE 2016 :

Rôle: STAGE OUTILS SYSTÈME ET BASES DE DONNÉES

### Réalisations:

* + Analyser l'existant et développer un outil qui permet de consulter et d’administrer les bases de données Oracle
  + Mettre en place un système de log, de cryptage et de décryptage, afin de superviser les administrateurs

**Environnement technique :** Java/swing et toad-

## ALSTOM JUIN 2015 -SEPTEMBRE 2015 :

Rôle: Stage développement d'interface/base de données **Réalisations:**

* + Modifier la structure de la base de données et développer des interfaces de saisie
  + Compiler de nombreuses informations venant pour la plupart de fichier Excel, afinde les implanter dans la base et générer des fichiers excel pour la planification des risques

### Environnement technique :

Bootstrap, Php, JavaScript, Jquery, dom et MySQL Workbench

## COMPÉTENCES TECHNIQUES

|  |  |
| --- | --- |
| **ETL** | Nifi |
| **Reporting** | ELK/Grafana/DataDog |
| **Base de données** | Hbase, Cassandra, Redis, MongoDB, InfluxDB, Neo4J, MYSQL |
| **Langages de développement** | Scala/Java/Groovy/Python/Bash |

**COMPÉTENCES MÉTHODOLOGIQUES**

|  |  |
| --- | --- |
| **Méthodes** | Agile |
| **Outil** | Jira/Confluence |

## FORMATION ET CERTIFICATIONS

**2021** ➢ Databricks Certified Associate Developer for Apache Spark Scala

* + - Microsoft Certified Azure Data Engineer
    - Google Cloud Professional Data Engineer

**2016** Master2 Data Scientist (Université Paris Saclay)

**2015 Master**1 Business Intelligence (Université Paris Saclay)

**2014** Licence3 Informatique (Université Paris Sud, Orsay)