



emmanuel.harel@gmail.com

Formation

DEA en physique
- Université de Paris VII

Développeur Java
- IFOCOP

Certifications

Data Engineer with Python
- Datacamp

Cloud Devops AWS
- Blent AI

AWS Certified Solutions Architects
- A Cloud Guru

Langues

Français

Anglais

Emmanuel

Data Engineer

Expérience
3
ans

Mes motivations

Passionné par le traitement de données, j'installe vos briques techniques et branche vos systèmes entre eux

Compétences clés

Techniques :

- **Langages** : Java, Python, Kotlin, Shell
- **Manipulation des données** : SQL, Pandas, Numpy, Spark
- **Visualisation des données** : Matplotlib, Matlab
- **Tests** : JUnit, Pytest
- **API** : FastAPI, REST
- **Cloud** : Amazon Web Services (S3, Lambda, EC2, IAM, ...)
- **Orchestration** : Airflow
- **Gestion du code**: Git
- **Gestion de l'environnement** : Docker, Kubernetes
- **CICD**: Jenkins, Gitlab-ci
- **IAC**: Terraform, ansible

Fonctionnelles :

- **Méthodologie** : Agile Jira,/Scrum
- **Finance**: portfolio management, index, stock
- **Communication** : Vulgarisation d'un sujet technique

Expériences professionnelles

Group M – Data Engineer (1 an 2 mois)

Développement de jobs pour transformer et écrire des analytics du trafic sur les sites web des clients

Enedis – Data Engineer (2 ans)

Mise en place d'un ETL temps réel pour archiver les flux des données Enedis dans un Datalake

CNAM – Data Engineer (2 mois)

Analyse temps fréquence d'un signal de moteur PSA pour la mise en évidence des mouvements de chaque pistons.

Références :

Informations des contacts disponibles sur demande.

- Alexandre L. (Groupe Covéa) : Tech Lead
- Frédéric A. (Reacteev) : Coach agile
- Jaouad B. (Indépendant) : Cloud Data engineer / DevOps

Group M

Data Engineer

1 an 2 mois

Projet :

Développer des jobs de traitement de données en temps réel ou en batch

Contexte :

Les rôles de l'entreprise étaient la collecte, la mesure et la structuration des données pour l'écosystème média de leurs clients.

Missions :

- développer des jobs Spark pour transformer et écrire sur HDFS des analytics du trafic (impression, click) des sites web des clients
- développer une API hautement responsive avec faible latence (<50ms) avec un contract Thrift
- Développement de batchs Java de traitement de nuit
- Migration de données SQL

Environnement technique :

Spring, spring boot2, REST, Spark, Hdfs, Thrift, Postman, Jooq, Mockito, Kafka, Redis, Storm, ssl two ways, GIT, MAVEN, Jenkins, Linux, Java 8

Enedis

Data Engineer

2 ans

Projet :

Mise en place d'un ETL pour archiver, intégrer les flux de données Enedis au sein d'un Datalake basé sur Kafka et HDFS

Contexte :

Enedis a mis en place le plus grand réseau de distribution d'électricité d'europe. Ses data scientists ont besoin de données construites en fonction de leurs analyses.

Missions :

- Scripting bash
- Développement API REST
- Customization de la configuration des brokers et connecteurs source (données xml, csv) Kafka
- Création de jobs Azkaban pour lancer une mise à jour des données entre un NAS et une autre machine
- Mise en place de pipelines Jenkins afin de démarrer les process Kafka connect, les process de mise à jour des schémas avro
- Pipeline Jenkins pour builder et déployer sur Openshift

Environnement technique :

REST APIs · Extract, Transform, Load (ETL), Kafka connect, Spring Boot, git, jenkins, Apache Kafka, Azkaban

CNAM

Data Engineer

2 mois

Projet :

Mise en place de méthodes d'analyse numérique temps fréquence des signaux de montée en régime de moteurs PSA

Contexte :

Le laboratoire d'acoustique du Conservatoire National des Arts et Métiers travaille entre autre sur les méthodes d'analyses de validation des montées en régime des moteurs PSA

Missions :

- Analyse spectrale par Fourier de la montée en régime
- Mise en évidence des limites de cette méthode
- Tests d'un algorithme basé sur Wigner-Ville
- Rédaction technique

Environnement technique :

Matlab, analyse numérique, Fourier, Wigner-Ville, Windows

PORTFOLIO

AWS portfolio pipeline disponible sur demande

Exemple d'architecture :

- Traitement "Realtime"

Kafka producer (python) => Kafka Brokers (EC2) => Kafka Consumer => AWS S3 => AWS Glue Crawler => Amazon Athena (SQL)

- Batch processing :

Airflow DAG (EC2 instance) => Python API consumer => S3

- API GraphQL + Step Function => Lambda Function => Third party API + Parameter Store => DynamoDb