Alaa **Zreik** Senior Data Engineer 8 ans d'expérience

Paris, France

Ingénieur Data avec plus de 8 ans d'expérience, j'accompagne les entreprises dans la conception et l'optimisation de solutions data à forte valeur ajoutée. Avec une solide expérience sur des projets internationaux et des environnements exigeants, je maîtrise l'ensemble de la chaîne data : modélisation, automatisation de pipelines (Python, spark, databricks, AWS), data warehousing, reporting avancé et gouvernance des données. Mon approche privilégie la fiabilité, la performance, et l'autonomie des équipes métiers grâce à des outils modernes et des process éprouvés. J'interviens sur : création de pipelines ETL/ELT, optimisation des coûts cloud, data modeling, intégration de modèles prédictifs et accompagnement stratégique. Je vous aide à transformer vos données en leviers de décisions concrètes, tout en garantissant la qualité et la sécurité de vos actifs data.

COMPÉTENCES

Programming ★ ★ ★ ★	
Python ★ ★ ★ ★	
Spark ★ ★ ★ ★	
AWS ★★★★	
Airflow ★ ★ ★ ★	
Deep/Machine learning ★ ★ ★ ★	×
Terraform ★ ★ ★ ★	

Step Functions ★ ★ ★ ★ ★ Databricks ★ ★ ★ ★ Cloud Watch ★ ★ ★ ★ Grafana ★ ★ ★ ★

Git * * * * * PowerBl ★ ★ ★ ★ ★ Jenkins ★ ★ ★ ★

SQL * * * * * Kafka ★ ★ ★ ★ MongoDB ★ ★ ★ ★ Aurora ★ ★ ★ ★

Jira ★ ★ ★ ★ Agile (SAFe) ★ ★ ★ ★ Gitlab ★ ★ ★ ★



EXPÉRIENCE PROFESSIONNELLE

Présent | Lead Data / Senior Data Engineer, TotalEnergies, PARIS, France Juillet 2024 | Principales fonctions exercées :

- > Responsabilité de la performance des pipelines, la qualité des données et la rigueur du code au sein d'une équipe composée de deux analystes et deux ingénieurs.
- > Migration des pipelines de données depuis Azure Synapse Analytics vers Databricks sur AWS, avec utilisation de Delta Lake sur S3.
- > Implémentation de la gouvernance des données via Unity Catalog, pour centraliser la gestion des accès, des métadonnées et de la sécurité.
- > Conception et développement de pipelines ETL scalables avec Apache Spark sur Databricks, intégrés à Amazon Aurora (RDS) pour alimenter les APIs métiers.
- > Mise en place d'une architecture streaming temps réel avec Delta Live Tables (DLT) et Autoloader, pour l'ingestion et la transformation continue des données d'Aurora à destination de Power BI.
- > Création des modèles sémantiques Power BI adaptés aux besoins métier, facilitant l'analyse en libreservice et la compréhension des données par les utilisateurs finaux.
- > Configuration de la connexions et des autorisations entre Power BI et Databricks, en lien avec les équipes sécurité et gouvernance.
- > Définition de workflows orchestrés assurant fiabilité, traçabilité et performance des traitements de
- > Contribution à la gestion de l'infrastructure avec une approche Infrastructure-as-Code (IaC) pour des déploiements reproductibles.
- > Mise en place de pipelines CI/CD automatisés avec GitHub Actions, garantissant l'intégration, les tests et le déploiement continus.
- > Supervision proactive des traitements et de la qualité des données via Grafana.
- > Rédaction de la documentation technique de l'architecture et des traitements pour assurer la maintenabilité de la plateforme.

AWS Databricks Azure Python Spark Delta lake Delta live table Unity Catalog S3 Aurora Terraform Grafana Github actions CI/CD PowerBI

Juillet 2024 | Senior Aws Data Engineer, BPI, MAISONS-ALFORT, France Octobre 2022 | Principales fonctions exercées :

- > Conception et déploiement d'architectures data évolutives et sécurisées, adaptées aux besoins transversaux complexes du client dans le secteur bancaire.
- > Optimisation de l'architecture cloud globale, entraînant une réduction des coûts opérationnels de plus de 60% grâce à la refonte et à l'automatisation des infrastructures.
- > Développement d'applications data critiques assurant un flux continu, performant et fiable des données à travers les systèmes.
- > Ingestion massive des données brutes dans un Data Lake S3 via des traitements batch et streaming, garantissant une base de données robuste et à jour.
- > Implémentation de pipelines ETL complexes pour la transformation et le chargement des données vers S3 et MongoDB, utilisant AWS Glue Jobs et Lambda Functions.
- > Orchestration avancée des workflows de données à l'aide de Apache Airflow et AWS Step Functions, pour assurer automatisation, résilience et tracabilité des traitements.
- > Supervision opérationnelle des pipelines avec CloudWatch et DataDog, anticipant les incidents et optimisant la maintenance.
- > Automatisation de la gestion des ressources cloud via Terraform, pour des déploiements reproductibles, sécurisés et versionnés.
- > Contribution à la mise en place des pipelines CI/CD avec GitLab et Jenkins, accélérant les cycles de développement et de déploiement.
- > Rédaction de documentation technique complète, facilitant la maintenance et la montée en compétences des équipes.
- > Collaboration dans un cadre Agile à grande échelle, suivant la méthodologie SAFe, pour une coordination optimale et une livraison rapide de valeur.

AWS Python Kafka Spark S3 Glue Step Functions Athena MongoDB Terraform Datadog Git CI/CD Jenkins

Septembre 2022 Data Engineer | Scientist, BnF, Paris, France

Octobre 2018

Titre: Conception et développement d'une plateforme d'ingestion et de traitement des données de la Bibliothèque nationale de France (BNF)

Principales fonctions exercées :

- > Conception et développement d'une plateforme d'ingestion et de traitement de données en temps réel pour détecter automatiquement la dégradation physique des documents de la BnF.
- > Intégration des données multi-sources (lecteurs, documents, conservation) dans un data pipeline big data basé sur des frameworks distribués.
- > Modélisation des données et définition du modèle sémantique à l'aide d'une ontologie métier, pour représenter de manière structurée les entités, événements et relations du domaine documentaire.
- > Conception et mise en œuvre de pipelines de traitement batch et streaming sur données relationnelles et non-structurées, avec des workflows orchestrés via Apache Airflow.
- > Collaboration étroite avec des Data Scientists pour la préparation, la transformation et la livraison des données pertinentes à des fins de recherche ou d'analyse.
- > Participation à l'enrichissement du datawarehouse documentaire et au déploiement de processus d'ingestion automatisés

Apache Spark Airflow Python Pandas PySpark Jupyter Intelligence artificielle

Septembre 2018 | Data Analyst, Musée du Louvre, PARIS, France

Mars 2018

Titre : Étude de la qualité des données du Musée du Louvre

Sujet : Le stage s'est concentré sur le problème de la qualité des données de conservation au Musée du

Principales fonctions exercées :

- > Collecter, intégrer et traiter les données du Louvre
- > Développer une démonstration pour évaluer la qualité des données de conservation

Python Apache Spark OWL Protege

É ÉDUCATION

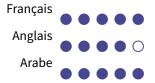
- 2022 Doctorat en informatique, Université de Paris Saclay, Versailles, France
- 2018 Master en Systèmes d'information et Intelligence des données, Université libanaise, Beyrouth, Liban
- 2016 Licence en informatique, Université libanaise, Beyrouth, Liban

CERTIFICATIONS

En cours AWS Solutions Architect Professional

2023 Google Cloud Certified Professional Data Engineer







> Musique : Piano> Sports : Cross-fit et football