

DATA ENGINEER

Le Data Engineer prépare l'architecture adéquate prête à recevoir les projets de data science et configure un datalake ou un datahub.

Il prépare ainsi les flux de données qui y sont injectées qui, après analyse, sont intégrées dans les modèles. Il fait appel aux langages et outils adaptés pour programmer les modèles et les algorithmes associés conçus par les Data Scientists.

Enfin, en utilisant les techniques de parallélisation et les techniques de Big Data, il met en production les modèles mis au point par les Data Scientists.



En suivant le parcours Data Engineer de l'École de la Data & de l'IA de Business & Decision, j'ai pu me former et me certifier dans de nombreux domaines et technologies, ce qui m'a permis d'être plus polyvalent et de gagner en crédibilité en tant que consultant junior, et ce, dès le début de ma carrière.

Antoine
Consultant Data
Nantes



3 mois ≈ 400 heures



Profil recherché : 1 à 2 ans d'expérience dans le domaine de la data (alternance acceptée.)

PARCOURS DATA ENGINEER

> UNE COMBINAISON DE MODULES

▶ Alimentation et préparation des données pour les algorithmes de data science

- Auditer la qualité des données
- Réparer, nettoyer et recoder les données
- Injecter les données dans le datalake / datahub

▶ Préparation et automatisation des traitements de données sur une infrastructure Big Data

- Tester et corriger le fonctionnement des infrastructures Big Data : datalake / datahub
- Spécifier, programmer & optimiser les traitements des données dans le datalake / datahub

▶ Spécification, programmation, industrialisation et recette des algorithmes de data science

- Spécifier, programmer & optimiser les algorithmes, les modèles de data science
- Industrialiser, sécuriser, rendre résilients et utilisables les modèles de data science
- Maintenir les modèles et algorithmes en état de fonctionnement



MÉTHODOLOGIE

- CRISP,
- DevOps
- Agile



LANGAGE & FRAMEWORK

- SQL, Java, Python, Scala
- SAS, Knime, Dataiku
- NoSQL, Kafka, Hive
- Nifi, Oozie



DOMAINES TECHNOLOGIQUES

- Big Data, Spark, Cloud,
- Hadoop, Data Science,
- Machine Learning, Statistiques