

INTRODUCTION À L'INGÉNIERIE DES DONNÉES

Durée : 2 jours (14 heures)

CONNAISSANCES PREALABLES

- Connaissances générales en informatique
- Connaissances de base en bases de données relationnelles
- Notions élémentaires en SQL
- Une première expérience de manipulation de données est recommandée
- Aucune connaissance préalable des outils Big Data n'est nécessaire.

PROFIL DES STAGIAIRES

- Futurs Data Engineers
- Développeurs souhaitant évoluer vers les métiers de la Data
- Data Analysts
- Administrateurs de bases de données
- Architectes SI et Data débutants
- Chefs de projets Data
- Consultants BI et décisionnel
- Toute personne souhaitant comprendre les fondamentaux de l'ingénierie des données.

OBJECTIFS

À l'issue de cette formation, les participants seront capables de :

- Comprendre le rôle et les responsabilités du Data Engineer
- Identifier les composants d'une plateforme Data moderne
- Comprendre les architectures de traitement et de stockage des données
- Différencier les approches ETL, ELT et Streaming
- Comprendre les principes des pipelines de données
- Identifier les principaux outils utilisés dans l'écosystème Data Engineering
- Comprendre les enjeux de qualité, gouvernance et sécurité des données
- Concevoir une architecture Data simple répondant à un besoin métier.

CERTIFICATION PREPAREE

Aucune

METHODES PEDAGOGIQUES

- Présentations illustrées par des cas concrets
- Démonstrations de plateformes Data modernes
- Études de cas inspirées de projets réels
- Ateliers collaboratifs

- Exercices d'analyse d'architectures
- Travaux pratiques de conception de pipelines

FORMATEUR

- Consultant expert Data Engineering et architectures Data modernes, disposant d'une expérience significative dans la conception de plateformes Data, Big Data et Cloud Data.

METHODE D'EVALUATION DES ACQUIS

- Quiz de validation des connaissances
- Exercices pratiques
- Études de cas
- Ateliers de conception
- Évaluation continue par le formateur
- Validation des acquis en fin de formation.

CONTENU DU COURS

Module 1 – Comprendre le rôle du Data Engineer (1h30)

Objectifs

- Comprendre les missions du Data Engineer
- Identifier sa place dans un projet Data

Contenu

- Les métiers de la Data :
 - Data Engineer
 - Data Analyst
 - Data Scientist
 - ML Engineer
 - Data Architect
- Missions et responsabilités du Data Engineer
- Cycle de vie de la donnée
- Positionnement dans les projets Data et IA

Mise en pratique

- Analyse des rôles dans différents projets Data
- Atelier de cartographie des compétences Data

Module 2 – Les fondations des architectures Data modernes (2h)

Objectifs

- Comprendre l'organisation des plateformes Data
- Identifier les composants clés d'une architecture moderne

Contenu

- Architecture Data traditionnelle
- Data Warehouse
- Data Lake
- Data Lakehouse
- Architecture distribuée
- Stockage et calcul

- Cloud Data Platforms
- Introduction aux architectures orientées données

Mise en pratique

- Analyse de plusieurs architectures réelles
- Atelier de comparaison des modèles Data

Module 3 – Collecter et intégrer les données (2h)

Objectifs

- Comprendre les mécanismes d'alimentation des plateformes Data
- Identifier les principales sources de données

Contenu

- Sources de données :
 - Bases relationnelles
 - API
 - Applications métier
 - Logs
 - IoT
- Concepts ETL et ELT
- Ingestion Batch
- Ingestion temps réel
- Qualité des données lors de l'intégration

Mise en pratique

- Étude d'un flux d'alimentation de données
- Identification des étapes d'un processus d'intégration

Module 4 – Concevoir des pipelines de données (2h)

Objectifs

- Comprendre le fonctionnement d'un pipeline Data
- Identifier les bonnes pratiques de conception

Contenu

- Définition d'un pipeline Data
- Extraction
- Transformation
- Chargement
- Orchestration des traitements
- Gestion des dépendances
- Monitoring des traitements

Mise en pratique

- Atelier de conception d'un pipeline de données
- Cartographie d'un flux de transformation

Module 5 – Découverte des outils du Data Engineering (2h)

Objectifs

- Comprendre le rôle des principaux outils du marché
- Identifier les technologies adaptées aux besoins

Contenu

Traitement et transformation

- Apache Spark
- Databricks
- dbt

Orchestration

- Apache Airflow
- Prefect

Streaming

- Apache Kafka
- Apache Flink

Stockage

- PostgreSQL
- MongoDB
- Data Lakes Cloud

Mise en pratique

- Analyse d'une chaîne complète de traitement
- Étude comparative des outils

Module 6 – Gouvernance, qualité et sécurité des données (1h30)**Objectifs**

- Comprendre les enjeux de maîtrise de la donnée
- Identifier les bonnes pratiques de gouvernance

Contenu

- Qualité des données
- Référentiels et métadonnées
- Data Lineage
- Catalogues de données
- Gestion des accès
- Sécurité des données
- RGPD et conformité

Mise en pratique

- Atelier d'identification des risques Data
- Analyse d'un processus de gouvernance

Module 7 – Data Engineering dans le Cloud (1h30)**Objectifs**

- Comprendre les évolutions vers les plateformes Cloud
- Identifier les services Data majeurs

Contenu

- Concepts Cloud Data
- Services AWS :
 - S3
 - Glue
 - Redshift
- Services Azure :
 - Data Factory
 - Synapse Analytics
- Services Google Cloud :
 - BigQuery
 - Dataflow
- Avantages et limites du Cloud

Mise en pratique

- Comparaison des offres Cloud Data
- Choix d'une architecture selon un contexte donné

Module 8 – Atelier de synthèse : concevoir une plateforme Data moderne (1h30)

Objectifs

- Mettre en application l'ensemble des notions étudiées
- Construire une architecture cohérente répondant à un besoin métier

Contenu

- Analyse du besoin
- Identification des sources
- Choix des technologies
- Définition des flux
- Gouvernance et supervision

Mise en pratique

- Cas fil rouge complet
- Conception d'une architecture Data cible
- Présentation des choix techniques
- Débriefing collectif

Notre référent handicap se tient à votre disposition au [01.71.19.70.30](tel:01.71.19.70.30) ou par mail à <mailto:referent.handicap@edugroupe.com> pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.