Data Engineering moderne : pipelines, orchestration et lakehouse (dbt, Airflow, Databricks)

Lien : https://innov-systems.com/formation/data-engineering-moderne-pipelines-orchestration-et-lakehouse-dbt-airflow-databricks

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI206

CATÉGORIE
Data Science et Data Engineering

Innov Systems

OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- Comprendre les principes du Data Engineering moderne
- Concevoir une architecture Data Lakehouse performante
- Développer des pipelines de données fiables et évolutifs
- Mettre en œuvre les transformations ELT avec dbt
- Orchestrer les workflows de données avec Apache Airflow
- Exploiter Databricks et Apache Spark pour le traitement distribué
- Garantir la qualité, la gouvernance et la traçabilité des données
- Industrialiser les pipelines grâce aux pratiques DataOps et CI/CD
- Sécuriser les plateformes de données et assurer leur conformité
- Déployer des solutions Data Engineering adaptées aux besoins de l'entreprise
- Optimiser les performances et la supervision des traitements de données
- Réaliser des projets complets de Data Engineering de niveau professionnel

POUR QUI ?

- Data Engineers
- Data Architects
- Data Scientists
- Analytics Engineers
- Développeurs Big Data
- Ingénieurs ETL/ELT
- Administrateurs de plateformes de données
- Ingénieurs Cloud
- DevOps et DataOps Engineers
- Consultants en Data Engineering
- Responsables BI et plateformes analytiques
- Toute personne souhaitant concevoir, déployer et industrialiser des pipelines de données modernes avec dbt, Airflow et Databricks

Innov Systems

Programme détaillé

1 / Fondamentaux du Data Engineering moderne

- Rôle du Data Engineer dans les architectures de données
- Évolution des entrepôts de données vers le Lakehouse
- Concepts de DataOps et d'architecture orientée données

2 / Architecture Data Lakehouse

- Composants d'une architecture Lakehouse
- Data Lake, Data Warehouse et Lakehouse : comparaison
- Formats de stockage modernes (Delta Lake, Apache Iceberg, Apache Hudi)

3 / Ingestion et intégration des données

- Collecte de données batch et temps réel
- Connecteurs, API et Change Data Capture (CDC)
- Gestion des données structurées, semi-structurées et non structurées

4 / Transformation des données avec dbt

- Modélisation analytique avec dbt
- Tests, documentation et versionnement des modèles
- Bonnes pratiques ELT et gestion des dépendances

5 / Orchestration des pipelines avec Apache Airflow

- Concepts de DAG, tâches et planification
- Gestion des workflows de données
- Monitoring, alertes et reprise sur incident

Innov Systems

6 / Plateforme Databricks et traitement distribué

- Architecture Databricks et Apache Spark
- Traitement distribué avec PySpark
- Optimisation des performances et gestion des clusters

7 / Qualité, gouvernance et catalogage des données

- Contrôle de la qualité des données
- Métadonnées, Data Catalog et Lineage
- Gouvernance et sécurité des données

8 / Industrialisation des pipelines de données

- CI/CD pour les pipelines Data Engineering
- Automatisation des déploiements
- Observabilité, supervision et optimisation des performances

9 / Sécurité et conformité des plateformes de données

- Gestion des accès et des identités
- Chiffrement, masquage et protection des données sensibles
- Conformité réglementaire (RGPD, gouvernance des données)

10 / Cas pratiques et déploiement d'une plateforme Lakehouse

- Construction d'un pipeline de bout en bout
- Déploiement d'une architecture Data Lakehouse avec dbt, Airflow et Databricks
- Bonnes pratiques d'exploitation en production

Innov Systems

Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

Prochaines dates programmées

04 au 08 Jan. 2027	Présentiel - Casablanca
01 au 05 Mars 2027	Présentiel - Casablanca
26 au 30 Avr. 2027	Présentiel - Casablanca
21 au 25 Juin 2027	Présentiel - Casablanca
16 au 20 Août 2027	Présentiel - Casablanca

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Réservation & Renseignements

Téléphone : +212 522 247 210
Email : contact@innov-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés