

Mis à jour le 27/07/2023

S'inscrire

Formation Delta Lake

2 jours (14 heures)

Présentation

Garantissez la fiabilité et la qualité des données de votre data lake. Delta Lake est une couche de stockage ouvert utilisée pour améliorer les performances de votre data lake. Le Delta Lake est une technologie rassemblant tous vos types de données (semi-structurée, structurée et non structurée) à un même endroit. En proposant une source unique pour vos données, Delta Lake garantit leurs véracités. L'outil prend en charge la validation des schémas et les transactions ACID, lui permettant d'offrir la fiabilité manquante dans les lacs de données traditionnels. Vos données seront sécurisées grâce au protocole Delta Sharing et à l'intégration native du Unity Catalog. Vous pourrez ainsi partager vos données en toute confiance avec vos collaborateurs. Cette formation Delta Lake vous enseignera les bénéfices du programme, la manipulation des tables, le Delta Time Travel ainsi que le streaming. Un module complémentaire d'une journée vous est proposé si vous souhaitez intégrer Delta Lake à Amazon Web Service. Comme toujours, nous vous enseignerons la dernière version de l'outil : [Delta Lake 1.0](#).

Objectifs

- Manipuler des tables avec Delta Lake
- Utiliser le Delta Time Travel
- Utiliser Delta Lake pour le streaming

Public visé

- Business analysts
- Data analysts
- Data scientists
- Data engineers
- Data miners

Pré-requis

- Connaissance en SQL et en gestion de base de données
- Connaissance en Big Data

Programme de notre Formation Delta Lake

Introduction à Delta Lake

- Le rôle des data lakes dans l'architecture Big Data
- La véracité des lacs de données
- L'architecture Delta
- Comment Delta Lake peut améliorer la fiabilité des données ?
- Pourquoi utiliser Delta Lake avec Databricks ?
- Installation

Débuter avec l'outil

- Initialiser Spark
- Charger ses données
- Présentation des différents formats de table
- Créer sa première table
- Le transaction log

Manipuler des tables

- Consulter et configurer l'historique des tables
- Conversion des tables parquet en table delta
- Gérer et optimiser la concurrence
- Suppression et mise à jour

Delta Time Travel

- Modifier ses données correctement
- Qu'est-ce-que le time travel ?
- Les répertoires Delta
- Utiliser le time travel
- Vacuum et rétention de données

Delta streaming

- La pipeline du data engineer avec Delta Lake
- Les meilleurs pratiques pour un processus robuste
- Les bonnes pratiques avant l'intégration de Spark Streaming
- Streaming avec Delta
- Utiliser Delta en source de stream

Module complémentaire : Delta Lake avec AWS (1 jour supplémentaire)

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.