



Formation

Delta Lake

DataTipsLearning.com

Tarif :
1 200 €
HT

Durée :
2 Jours
(14h)

Lieu :
PARIS/
Distancielle/
Votre Local

Contact :

contact@datatipslearning.com

+33 0751386021

Réf. DTPSL-DLKE-2022

[S'inscrire](#)

Plus de détails :

Objectifs

Public / Prérequis

Programme détaillé

Méthode pédagogique

Le progrès du moteur Apache Spark et de la plateforme Databricks offre maintenant la possibilité de combiner les avantages d'un data lake et celles d'un Data Warehouse, pour créer finalement un Data Lakehouse qui traite une énorme quantité de données en assurant l'acidité de ces données.

Via un concept de log, Delta Lake permet à la fois d'unifier le traitement en batch et en streaming, et le versioning de données, qui sont la base d'un Lakehouse.

Objectifs

- Comprendre le concept Delta Lake.
- Création d'une table Delta Lake.
- Delete, update, merge, & upsert dans une Delta Lake.
- Requêter une table par version.
- Comprendre le concept des logs.
- Les fondamentaux des pipelines Delta Live.
- Les optimisations de Delta Lake.

Public

- Data Engineer
- Data Scientist
- Data Analyste
- Data Architectes

Prérequis

- Connaissance d'au moins un langage de programmation.
- Connaissance de quelques notions de base de données : Table, Ligne, Colonne, ...

Programme détaillé

1. Les concepts Data Lake et Data Warehouse
2. Introduction Databricks
3. Introduction Delta Lake
4. Les propriétés ACID du stockage
5. L'architecture Data Lakehouse
6. Le concept des logs d'un Delta Lake
7. Créer une table Delta Lake
8. Ingestion de données dans Delta Lake
9. Reads & Writes des tables en batch
10. Reads & Writes des tables en streaming
11. Delete, Update, Merge & Upserts d'une table
12. Modification des données d'une table Delta Lake existante

13. Time travel
14. Protocole de versioning d'une table
15. Les niveaux d'isolation
16. Contrôle de la concurrence
17. Évolution du schéma
18. Sharing data
19. Data catalog
20. Optimisations de Delta Lake
21. Gouvernance et sécurité dans un Data Lakehouse
22. Utility commands d'une table
23. Delta column mapping
24. Delta Lake APIs
25. Fondamentaux des tables Delta Live
26. Pipeline des tables Delta Live
27. Data quality des tables Delta Live
28. Change data capture avec des tables Delta Live

Méthode pédagogique

La formation se compose d'une partie théorique, et également une partie pratique représentant 60% de de la formation.

La partie pratique contient plusieurs exercices sous forme de notebook Databricks avec les corrections, avec aussi un projet à la fin de la formation comme simulation d'une prod.

Chaque jour, une évaluation rapide des connaissances est effectuée avant de commencer les nouvelles parties de la formation.

A la fin, une synthèse globale est délivré aux stagiaires, renforcé par un projet prod.

Finalement, une évaluation QCM est proposée.

Un support de cours sera remis à chaque stagiaire comprenant les slides, les exercices et les corrigés et un git du projet prod.

Une feuille de présence est fournie en fin de formation avec une certificat de complétion de formation pour chaque stagiaire.

Le formateur est un Data Engineer expert, qui intervient sur le sujet depuis plusieurs années en formation mais aussi en conseil.