

Mis à jour le 26/04/2024

S'inscrire

Formation Apache Iceberg

3 jours (21 heures)

Présentation

Vous cherchez un format de table pour l'ensemble de vos données analytiques ? Notre formation Apache Iceberg vous permettra de comprendre comment utiliser pour toutes vos analyses statistiques.

Ce framework open source vous offre une gestion efficace des tables massives avec plusieurs fonctionnalités comme l'intégrité [transactionnelle](#) et la compatibilité multiplateforme.

Durant ce cours, vous découvrirez la traçabilité des données qui permettent de suivre l'historique complet des modifications.

Cette formation devrait vous permettre d'acquérir une expertise dans la gestion des données à grande échelle, et vous devriez être capable de concevoir et gérer des données massives.

Comme pour toutes, la formation Apache Iceberg vous sera présentée avec ses toutes [dernières nouveautés](#) (à la date de rédaction de l'article).

Objectifs

- Comprendre les fondamentaux d'Apache Iceberg
- Apprendre les techniques d'optimisation des performances des tables
- Savoir écrire et exécuter des requêtes de lecture et d'écriture dans Iceberg
- Comprendre la gestion des catalogues et la sécurisation des données
- Maîtriser l'architecture d'Iceberg ainsi que les couches de données et métadonnées

Public visé

- Data scientists
- Data analysts

- Data engineers
- Architectes

Pré-requis

- Compréhension des concepts fondamentaux des data lakes et des bases de données
- Connaissance en programmation (Python) et du SQL

PROGRAMME DE NOTRE FORMATION APACHE ICEBERG

Introduction à Apache Iceberg

- Initiation
- Entrepôt de données
- Lac de données
- Découvrir ce qu'est Hive
- Formats de table

Architecture

- La couche des données
- La couche des métadonnées
 - Fichiers manifestes
 - Listes de manifestes
 - Fichiers de métadonnées
 - Fichiers macareux
- Le catalogue

Cycle de vie des requêtes d'écriture et de lecture

- Écrire des requêtes dans Apache Iceberg
 - Créer le tableau
 - Insérer la requête
 - Fusionner la requête
- Lecture de requêtes dans Apache Iceberg
 - La requête SELECT
 - La requête du voyage dans le temps

Optimisation des performances des tables

- Compactage

- Pratique avec le compactage
 - Stratégies
 - Automatisation
- Tri
- Ordre Z
- Partitionnement
 - Partitionnement caché
 - Évolution des partitions
 - Autres considérations de partitionnement
- Copie sur écriture et fusion sur lecture

Catalogues

- Exigences d'un catalogue Iceberg
- Comparaison du catalogue
- Migration du catalogue
 - Utilisation de la CLI de migration du catalogue Apache Iceberg
 - Utiliser un moteur

Apache Spark

- Configuration
- Opérations du langage de définition de données
- Lecture des données
- Procédures d'entretien des tables Iceberg

Moteur de requête SQL de Dremio

- Entretien des tables icebergs
- Réécrire les fichiers de données
- Réécrire les manifestes

Colle AWS

- Configuration
 - Conception d'une base de données Glue
 - Configuration du travail Glue ETL
- Création d'une table via le catalogue de données Glue
 - Lire le tableau
 - Insérer les données

Gouvernance et sécurité

- Sécuriser les fichiers de données

- Sécuriser et gouverner au niveau de la couche sémantique
 - Meilleures pratiques en matière de couche sémantique
 - Drémio
 - Trino
- Sécuriser et gouverner au niveau du catalogue

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.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

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.