Delta Lake : Architecture Lakehouse, Time Travel et Gestion ACID des Data Lakes

Lien : https://innov-systems.com/formation/delta-lake-architecture-lakehouse-time-travel-et-gestion-acid-des-data-lakes

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI150

CATÉGORIE
Big Data : fondamentaux et architecture

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Comprendre les principes de l’architecture Lakehouse
- Maîtriser les fonctionnalités avancées de Delta Lake
- Implémenter la gestion ACID dans un Data Lake
- Exploiter le Time Travel pour l’audit et la traçabilité
- Concevoir des pipelines de données fiables
- Intégrer Delta Lake avec Apache Spark
- Optimiser les performances des traitements Big Data
- Mettre en œuvre des mécanismes de gouvernance des données
- Gérer l’évolution des schémas de données
- Déployer des architectures Lakehouse sur le cloud
- Comparer Delta Lake aux autres technologies du marché
- Étudier des cas concrets d’implémentation industrielle

POUR QUI ?

- Data Engineers
- Architectes Data
- Data Architects
- Data Scientists
- Ingénieurs Big Data
- Développeurs Spark
- Administrateurs plateformes Data
- Consultants Data & Analytics
- Responsables gouvernance des données
- Chefs de projet Data
- Toute personne impliquée dans la conception et l’exploitation de plateformes de données modernes

Innov Systems

Programme détaillé

1 / Introduction à Delta Lake et au concept Lakehouse

- Limites des Data Lakes traditionnels
- Émergence de l’architecture Lakehouse
- Positionnement de Delta Lake dans l’écosystème Big Data

2 / Fondamentaux de l’architecture Lakehouse

- Convergence Data Warehouse et Data Lake
- Stockage ouvert et formats de données
- Cas d’usage analytiques et opérationnels

3 / Architecture interne de Delta Lake

- Journal des transactions (Transaction Log)
- Fichiers Parquet et métadonnées
- Gestion des versions des données

4 / Gestion ACID dans les Data Lakes

- Atomicité, Cohérence, Isolation et Durabilité
- Contrôle de concurrence optimiste
- Fiabilité des pipelines de données

5 / Time Travel et gestion des versions

- Consultation des données historiques
- Reconstitution des états passés
- Cas d’usage d’audit et de conformité

Innov Systems

6 / Ingestion et traitement des données

- Batch processing et streaming
- Intégration avec Apache Spark
- Gestion des mises à jour et suppressions

7 / Optimisation des performances

- Partitionnement des données
- Compaction et commande OPTIMIZE
- Data Skipping et Z-Ordering

8 / Gouvernance et qualité des données

- Contraintes et validation des données
- Gestion des schémas (Schema Enforcement)
- Évolution des schémas (Schema Evolution)

9 / Intégration dans les plateformes modernes

- Databricks et écosystème Delta Lake
- Interopérabilité avec Apache Iceberg et Hudi
- Déploiement sur AWS, Azure et Google Cloud

10 / Études de cas et projets pratiques

- Construction d’un Lakehouse complet
- Implémentation du Time Travel
- 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

08 au 12 Fév. 2027	Présentiel - Casablanca
05 au 09 Avr. 2027	Présentiel - Casablanca
31 Mai au 04 Juin 2027	Présentiel - Casablanca
26 au 30 Juil. 2027	Présentiel - Casablanca
20 au 24 Sep. 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