Apache Iceberg : Concevoir et Administrer des Data Lakehouses Performants et Évolutifs

Lien : https://innov-systems.com/formation/apache-iceberg-concevoir-et-administrer-des-data-lakehouses-performants-et-evolutifs

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI148

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 des architectures Data Lakehouse
- Maîtriser les concepts fondamentaux d’Apache Iceberg
- Concevoir des plateformes de données évolutives
- Administrer des tables Iceberg en production
- Optimiser les performances des traitements analytiques
- Intégrer Iceberg avec Spark, Trino, Flink et autres moteurs
- Gérer l’évolution des schémas de données
- Mettre en place une gouvernance efficace des données
- Sécuriser les environnements Data Lakehouse
- Réduire les coûts d’exploitation des plateformes Big Data
- Superviser et maintenir des infrastructures de données modernes
- Étudier des cas concrets d’implémentation en entreprise

POUR QUI ?

- Data engineers
- Architectes Big Data
- Administrateurs de plateformes de données
- Développeurs Spark et Hadoop
- Data architects
- Consultants en gestion des données
- Responsables infrastructures data
- DevOps et Cloud engineers
- Data analysts avancés
- Chefs de projets data
- Toute personne impliquée dans la conception ou l’exploitation de Data Lakehouses

Innov Systems

Programme détaillé

1 / Introduction aux Data Lakehouses et à Apache Iceberg

- Évolution des architectures Data Warehouse, Data Lake et Lakehouse
- Présentation d'Apache Iceberg et de ses avantages
- Cas d'usage modernes de gestion des données

2 / Architecture et concepts fondamentaux d'Apache Iceberg

- Tables Iceberg et métadonnées
- Snapshots, manifests et partitionnement
- Gestion des versions de données

3 / Installation et configuration d'Apache Iceberg

- Prérequis techniques et architecture cible
- Intégration avec Hadoop, Spark et le cloud
- Configuration des catalogues de données

4 / Gestion des données dans Iceberg

- Création et administration des tables
- Insertion, mise à jour et suppression de données
- Gestion des schémas évolutifs

5 / Optimisation des performances

- Partitionnement avancé
- Compaction et optimisation des fichiers
- Réduction des coûts de requêtes

Innov Systems

6 / Intégration avec les moteurs analytiques

- Apache Spark et Iceberg
- Trino, Presto et Flink
- Interopérabilité multi-moteurs

7 / Gouvernance et qualité des données

- Gestion des métadonnées
- Contrôle de la qualité des données
- Traçabilité et auditabilité

8 / Sécurité et administration

- Gestion des accès et permissions
- Sécurisation des données dans le cloud
- Surveillance et maintenance des plateformes

9 / Exploitation d’un Data Lakehouse en production

- Stratégies de déploiement
- Gestion de la montée en charge
- Supervision et optimisation continue

10 / Études de cas et projets pratiques

- Mise en œuvre d’un Data Lakehouse avec Iceberg
- Cas d’usage Big Data en entreprise
- Bonnes pratiques d’architecture et d’administration

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

01 au 05 Fév. 2027	Présentiel - Casablanca
29 Mars au 02 Avr. 2027	Présentiel - Casablanca
24 au 28 Mai 2027	Présentiel - Casablanca
19 au 23 Juil. 2027	Présentiel - Casablanca
13 au 17 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