Mettre en Œuvre un Lakehouse avec Microsoft Fabric

Lien : https://innov-systems.com/formation/mettre-en-uvre-un-lakehouse-avec-microsoft-fabric

DURÉE	RÉFÉRENCE	CATÉGORIE
3 jours (21h)	**VSC335**	**Microsoft Azure**

OBJECTIFS DE LA FORMATION

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

- Concevoir et déployer une architecture lakehouse complète dans Microsoft Fabric
- Maîtriser l'ingestion, la transformation et l'orchestration des données
- Implémenter des pipelines de données robustes avec Apache Spark et Delta Lake
- Optimiser les performances et la gouvernance des données dans un lakehouse
- Créer des solutions d'analyse en temps réel et batch avec Microsoft Fabric

POUR QUI ?

- Ingénieurs données (Data Engineers)
- Architectes données
- Développeurs BI
- Responsables techniques data

Innov Systems

Programme détaillé

1 / INTRODUCTION À L'ÉCOSYSTÈME MICROSOFT FABRIC

- Vue d'ensemble de la plateforme Microsoft Fabric et ses composants
- Architecture de référence pour un lakehouse moderne
- Différences entre lakehouse, data warehouse et data lake
- Cas d'usage et scénarios d'implémentation

2 / CONCEPTION ET PROVISIONNEMENT D'UN LAKEHOUSE

- Principes de conception d'un lakehouse performant
- Création et configuration d'un espace de travail Fabric
- Mise en place d'un lakehouse avec structure optimisée
- Gestion des autorisations et sécurité d'accès

3 / INGESTION DE DONNÉES : STRATÉGIES ET BONNES PRATIQUES

- Méthodes d'ingestion : batch vs streaming
- Chargement de fichiers dans le lakehouse (CSV, Parquet, JSON)
- Création de tables managées et externes
- Gestion des schémas et évolution des données

4 / FONDAMENTAUX D'APACHE SPARK DANS FABRIC

- Architecture Spark dans Microsoft Fabric
- Configuration des pools Spark et optimisation des ressources
- Introduction aux notebooks Spark et leur utilisation
- Premiers pas avec les DataFrames Spark

5 / TRAITEMENT AVANCÉ AVEC APACHE SPARK

Innov Systems

- Manipulation avancée des DataFrames (filtrage, agrégation, jointures)
- Optimisation des requêtes Spark et gestion de la mémoire
- Utilisation de Spark SQL pour des transformations complexes
- Création de vues temporaires et catalogues de données

6 / DELTA LAKE : GESTION TRANSACTIONNELLE DES DONNÉES

- Architecture et principes de Delta Lake
- Création et configuration de tables Delta
- Opérations ACID : INSERT, UPDATE, DELETE, MERGE
- Time travel et versioning des données
- Optimisation des tables Delta (OPTIMIZE, Z-ORDER)

7 / STREAMING DE DONNÉES AVEC SPARK STRUCTURED STREAMING

- Concepts du streaming structuré dans Spark
- Ingestion de données en temps réel dans des tables Delta
- Gestion des fenêtres temporelles et agrégations streaming
- Monitoring et gestion des erreurs en streaming

8 / ORCHESTRATION AVEC DATA FACTORY PIPELINES

- Architecture des pipelines dans Microsoft Fabric
- Création de pipelines d'orchestration de bout en bout
- Utilisation de l'activité Copy Data et ses configurations avancées
- Paramétrage et dynamisation des pipelines

9 / TRANSFORMATION DE DONNÉES AVEC DATAFLOWS GEN2

- Architecture et capacités des Dataflows Gen2
- Création de transformations Power Query avancées
- Intégration des Dataflows dans les pipelines
- Comparaison Dataflows vs Notebooks Spark : quand utiliser quoi

Innov Systems

10 / REQUÊTAGE ET ANALYSE AVEC SQL ANALYTICS ENDPOINT

- Utilisation du SQL Analytics Endpoint du lakehouse
- Requêtes T-SQL sur les tables Delta
- Création de vues SQL pour l'analyse
- Connexion avec Power BI et outils externes

11 / PERFORMANCE ET OPTIMISATION

- Stratégies de partitionnement des données
- Optimisation des lectures avec le data skipping
- Gestion du cache et des statistiques
- Monitoring des performances Spark et détection des goulots

12 / GOUVERNANCE ET BONNES PRATIQUES

- Mise en place de la gouvernance des données
- Gestion des métadonnées et documentation
- Stratégies de sauvegarde et disaster recovery
- Sécurité au niveau ligne et colonne

13 / SUPERVISION ET MISE EN PRODUCTION

- Monitoring des pipelines et alerting
- Gestion des dépendances entre pipelines
- Stratégies de déploiement (Dev, Test, Prod)
- Troubleshooting et résolution d'incidents

14 / PROJET PRATIQUE ET CAS D'USAGE RÉEL

- Implémentation d'un lakehouse complet de bout en bout
- Intégration de sources multiples (API, fichiers, bases de données)
- Création d'un pipeline de données automatisé

Inov Systems

- Présentation et validation du projet

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

25 au 27 Nov. 2026	Distanciel
06 au 08 Jan. 2027	Présentiel - Casablanca
10 au 12 Mars 2027	Présentiel - Casablanca
05 au 07 Mai 2027	Présentiel - Casablanca
30 Juin au 02 Juil. 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

Innov Systems — Tous droits réservés