Implémenter une solution engineering données avec Azure Databricks (DP-3027)

Description

Maîtrisez le traitement de données avec Azure Databricks

Azure Databricks devient un pilier essentiel dans l'univers de la data engineering moderne. Grâce à sa puissance, sa flexibilité et ses outils intégrés, cette plateforme vous permet de traiter de grands volumes de données en toute efficacité. Cette formation DP-3027 vous offre une maîtrise complète des pipelines de données dans le cloud, à travers des cas concrets et des modules progressifs.

Une formation pratique pour les ingénieurs data d'aujourd'hui

Ce cours a été conçu pour répondre aux besoins réels des professionnels de la donnée. Vous y apprendrez à mettre en place des solutions d'ingénierie robustes, à automatiser les flux de travail et à garantir la qualité des données en production. Les modules couvrent l'ensemble du cycle de vie des pipelines, du traitement en temps réel avec Spark Structured Streaming jusqu'à l'implémentation de workflows CI/CD, en passant par l'automatisation via Azure Data Factory.

Contenu du cours
Module 1 : Effectuer un traitement incrémentiel avec Spark Structured Streaming

- Configurer des sources de données en temps réel pour le traitement incrémentiel
- Optimiser Delta Lake pour le traitement incrémentiel dans Azure Databricks
- Gérer les données tardives et les événements non ordonnés dans le traitement incrémentiel
- Stratégies de monitoring et de réglage des performances pour le traitement incrémentiel dans Azure Databricks

Module 2 : Implémenter des modèles d'architecture de diffusion en continu avec Delta Live Tables

- Architectures pilotées par les événements avec des tables Delta Live
- Ingérer des données avec le streaming structuré
- Maintenir la cohérence et la fiabilité des données avec la diffusion en continu structurée
- Mettre à l'échelle des charges de travail de streaming avec des tables Delta Live

Module 3 : Optimiser les performances avec Spark et Delta Live Tables

- Optimiser les performances avec Spark et Delta Live Tables
- Effectuer l'optimisation basée sur les coûts et le réglage des requêtes
- Utiliser la capture des changements de données (CDC)
- Utiliser la mise à l'échelle automatique améliorée
- Implémenter des métriques d'observabilité et de qualité des données

Module 4 : Implémenter des workflows d'intégration continue et livraison continue (CI/CD) dans Azure Databricks

- Implémenter la gestion de version et l'intégration Git

© Copyright ITTA – Tous les contenus sont protégés par le droit d'auteur.

- Effectuer des tests d’intégration et des tests unitaires
- Gérer et configurer votre environnement
- Implémenter des stratégies de restauration et de restauration par progression

Module 5 : Automatiser des charges de travail avec des travaux Azure Databricks

- Implémenter la planification et l’automatisation de travaux
- Optimiser les workflows avec des paramètres
- Traiter la gestion des dépendances
- Implémenter la gestion des erreurs et les mécanismes de nouvelle tentative
- Découvrir les règles et meilleures pratiques

Module 6 : Gérer la confidentialité et la gouvernance des données avec Azure Databricks

- Mettre en œuvre des techniques de chiffrement de données dans Azure Databricks
- Gérer les contrôles d’accès dans Azure Databricks
- Mettre en œuvre le masquage des données et l’anonymisation dans Azure Databricks
- Utiliser le partage de données sécurisé et les infrastructures de conformité dans Azure Databricks
- Utiliser la gestion des métadonnées et la traçabilité des données
- Mettre en œuvre l’automatisation de la gouvernance dans Azure Databricks

Module 7 : Utiliser des entrepôts SQL dans Azure Databricks

- Prise en main de SQL Warehouses
- Créer des bases de données et des tables
- Créer des requêtes et des tableaux de bord

Module 8 : Exécuter des notebooks Azure Databricks avec Azure Data Factory

- Comprendre les notebooks et pipelines Azure Databricks
- Créer un service lié pour Azure Databricks
- Utiliser une activité Notebook dans un pipeline
- Utiliser des paramètres dans un carnet

Lab / Exercices

- Ce cours vous donne un accès exclusif au laboratoire officiel Microsoft, vous permettant de mettre en pratique vos compétences dans un environnement professionnel.

Documentation

- Accès à Microsoft Learn, la plateforme d'apprentissage en ligne Microsoft, offrant des ressources interactives et des contenus pédagogiques pour approfondir vos connaissances et développer vos compétences techniques.

Profils des participants

- Ingénieur Data
- Développeur Big Data
- Architecte Cloud
- Consultant Data
- Administrateur Azure
- Responsable BI

Connaissances Préalables

© Copyright ITTA – Tous les contenus sont protégés par le droit d’auteur.

- Savoir utiliser un ordinateur
- Comprendre les bases du cloud

Objectifs

- Créer des flux de données en temps réel
- Construire des architectures avec Delta Live Tables
- Améliorer la performance des traitements
- Déployer des workflows CI/CD dans Databricks
- Automatiser les tâches avec Azure Data Factory
- Sécuriser et gouverner les données
- Utiliser les entrepôts SQL dans Databricks
- Exécuter des notebooks depuis Azure Data Factory

Description
Implémenter une solution engineering données avec Azure Databricks (DP-3027)
Niveau
Intermédiaire
Prix de l'inscription en Présentiel (CHF)
900
Prix de l'inscription en Virtuel (CHF)
850
Durée (Nombre de Jours)
1
Reference
DP-3027

© Copyright ITTA – Tous les contenus sont protégés par le droit d’auteur.