

Implémenter une solution Data Analytics Solution avec Azure Databricks

Durée: 1 Jour Réf de cours: M-DP3011

Résumé:

Apprenez à exploiter la puissance d'Apache Spark et des clusters performants fonctionnant sur la plateforme Azure Databricks pour exécuter d'importants workloads d'ingénierie des données dans le cloud. Azure Databricks est un service cloud qui fournit une plateforme évolutive pour l'analyse de données à l'aide d'Apache Spark. Azure Databricks est construit sur Apache Spark et permet aux ingénieurs de données et aux analystes d'exécuter des tâches Spark pour transformer, analyser et visualiser les données à l'échelle. Delta Lake est une zone de stockage relationnel open source pour Spark que vous pouvez utiliser pour mettre en œuvre une architecture de lac de données dans Azure Databricks. Azure Databricks fournit des entrepôts SQL qui permettent aux analystes de données de travailler avec des données en utilisant des requêtes SQL relationnelles familières. L'utilisation de pipelines dans Azure Data Factory pour exécuter des notebooks dans Azure Databricks vous permet d'automatiser les processus d'ingénierie des données à l'échelle du cloud.

Cette formation vous aide à préparer à l'examen DP-203 : Data Engineering on Microsoft Azure

Mis à jour le 17/02/2025

Public visé:

Ce cours s'adresse aux personnes souhaitant travailler avec Azure Spark et Azure Databricks pour manipuler des données dans le cloud.

Objectifs pédagogiques:

- | | |
|---|---|
| ■ À l'issue de la formation, les participants seront capables : | ■ D'utiliser Delta Lake dans Azure Databricks |
| ■ D'explorer Azure Databricks | ■ D'utiliser SQL Warehouse dans Azure Databricks |
| ■ D'utiliser Apache Spark dans Azure Databricks | ■ D'exécuter Azure Databricks Notebooks avec Azure Data Factory |

Pré-requis:

Les stagiaires doivent avoir le prérequis suivant :

- une connaissance du Cloud et de l'analyse des données.

Contenu:

Module 1 : Explorer Azure Databricks

- Provisionner un espace de travail Azure Databricks.
- Identifier les charges de travail principales et les personas pour Azure Databricks.
- Décrire les concepts clés d'une solution Azure Databricks.

Module 2 : Utiliser Apache Spark dans Azure Databricks

- Décrire les éléments clés de l'architecture Apache Spark.
- Créer et configurer un cluster Spark.
- Décrire les cas d'utilisation de Spark.
- Utiliser Spark pour traiter et analyser des données stockées dans des fichiers.
- Utiliser Spark pour visualiser les données.

Module 3 : Utiliser Delta Lake dans Azure Databricks

- Décrire les principales caractéristiques et capacités de Delta Lake.
- Créer et utiliser des tables Delta Lake dans Azure Databricks.
- Créer des tables de catalogue Spark pour les données Delta Lake.
- Utiliser les tables Delta Lake pour les données en continu.

Module 4 : Utiliser les entrepôts SQL dans Azure Databricks

- Créer et configurer les entrepôts SQL dans Azure Databricks.
- Créer des bases de données et des tables.
- Créer des requêtes et des tableaux de bord.

Module 5 : Exécuter les Notebooks Azure Databricks avec Azure Data Factory

- Décrire comment les notebooks d'Azure Databricks peuvent être exécutés dans un pipeline.
- Créer un service lié Azure Data Factory pour Azure Databricks.
- Utiliser une activité Notebook dans un pipeline.
- Transmettre des paramètres à un notebook.

Méthodes pédagogiques :

Les participants réalisent un test d'évaluation des connaissances en amont et en aval de la formation pour valider les connaissances acquises pendant la formation.

Un support de cours officiel sera remis aux stagiaires

Autres moyens pédagogiques et de suivi:

- Compétence du formateur : Les experts qui animent la formation sont des spécialistes des matières abordées et ont au minimum cinq ans d'expérience d'animation. Nos équipes ont validé à la fois leurs connaissances techniques (certifications le cas échéant) ainsi que leur compétence pédagogique.
- Suivi d'exécution : Une feuille d'émargement par demi-journée de présence est signée par tous les participants et le formateur.
- En fin de formation, le participant est invité à s'auto-évaluer sur l'atteinte des objectifs énoncés, et à répondre à un questionnaire de satisfaction qui sera ensuite étudié par nos équipes pédagogiques en vue de maintenir et d'améliorer la qualité de nos prestations.

Délais d'inscription :

- Vous pouvez vous inscrire sur l'une de nos sessions planifiées en inter-entreprises jusqu'à 5 jours ouvrés avant le début de la formation sous réserve de disponibilité de places et de labs le cas échéant.
- Votre place sera confirmée à la réception d'un devis ou "booking form" signé. Vous recevrez ensuite la convocation et les modalités d'accès en présentiel ou distanciel.
- Attention, si cette formation est éligible au Compte Personnel de Formation, vous devrez respecter un délai minimum et non négociable fixé à 11 jours ouvrés avant le début de la session pour vous inscrire via moncompteformation.gouv.fr.

Accueil des bénéficiaires :

- En cas de handicap : plus d'info sur globalknowledge.fr/handicap
- Le Règlement intérieur est disponible sur globalknowledge.fr/reglement