

Mis à jour le 01/07/2024

S'inscrire

Formation Azure Databricks

2 jours (14 heures)

Présentation

Azure Databricks est la solution pour manipuler un nombre massif de clusters de base de données. Azure Databricks s'appuie sur la puissance du Cloud de Microsoft pour simplifier votre processus Big Data.

Azure Databricks est une version améliorée d'Apache Spark géré entièrement par Azure. Le système est alors facilement scalable et contrôlable afin de réduire vos coûts durant la phase de développement.

Le système dispose de différents connecteurs comme ADSL, Azure SQL DW ou encore PowerBI. C'est également un outil hautement sécurisé, intégrant le cadre de sécurité Azure Active Directory.

Notre formation Azure Databricks vous enseignera comment utiliser cet outil lors de votre processus ETL. Nous vous présenterons comment extraire, transformer et charger des données avec Azure Databricks, nous évoquerons aussi la création de dashboard et de pipelines.

Notre formation Azure Databricks présentera la dernière version de l'outil à savoir, [Azure Databricks](#).

Objectifs

- Bien connaître les spécificités de Azure Databricks
- Extraire les données avec Azure Databricks
- Savoir comment transformer et charger ses données
- Utiliser les dashboards et déployer son processus

Public visé

- Développeurs
- Data Engineer
- Architectes
- Administrateurs système
- Data miners
- Data scientists
- Data analysts
- Business intelligence analysts
- Market intelligence analysts

Pré-requis

- Avoir idéalement suivi nos formations [Spark ML](#) ou [Spark Tuning Avancé](#)
- Connaissance en SQL et Python

Pré-requis logiciel

- Disposer d'un compte Azure
- Disposer d'un compte Databricks

Programme de la formation Azure Databricks

INTRODUCTION

- Présentation de Databricks
- Les bénéfices de l'outil
- Databricks vs Apache Spark
- Présentation de l'interface
- Créer son workspace
- Présentation des notebooks
- Créer les composants d'Azure Databricks (cluster, job, pool)

EXTRACTION DES DONNÉES

- Les différentes tables (schemas, databases, tables)
- Les sources et les destinations
- Naviguer sur les fichiers système
- Le dossier Filestore
- Importer les données depuis son ordinateur
- Importer les données du web
- Importer les données avec SQL et Python
- Connecter ses données Excel, PowerBI et PyCharm

TRANSFORMATION DES DONNÉES

- Trouver et modifier les valeurs null
- Supprimer les données dupliquées
- Supprimer les valeurs extrêmes
- Mise en cache
- Compression des données

CHARGEMENT DES DONNÉES

- Méthodes de sauvegarde
- Managed vs unmanaged tables
- Gérer les partitions
- Connexion avec Azure SQL, Oracle et MongoDB

DASHBOARD ET DÉPLOIEMENT DU PROCESSUS

- Notebooks de tableaux de bord
- Créer un travail planifié pour rafraîchir le dashboard
- Présentation d'un pipeline Azure Databricks CI/CD
- Développer et valider son code
- Définir son pipeline de build
- Définir son pipeline de mise en version

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.