

Formation : Databricks pour data engineers

Formation pratique - 3j - 21h00 - Réf. DKI
Prix : 2380 CHF H.T.

NEW

En trois jours, cette formation offre aux ingénieurs data les clés pour exploiter tout le potentiel d'Azure Databricks dans un contexte cloud moderne. De la création de notebooks à l'orchestration de workflows, en passant par l'utilisation des langages Python et SQL et l'usage du Datastore. Elle met l'accent sur la performance, l'automatisation et la valorisation des données au service de projets data industrialisés.

Objectifs pédagogiques

À l'issue de la formation, le participant sera en mesure de :

- ✓ Comprendre et utiliser les différentes fonctionnalités de Databricks
- ✓ Créer et manipuler des notebooks
- ✓ Mettre en place des jobs
- ✓ Utiliser Python et SQL avec Databricks
- ✓ Comprendre et utiliser le Datastore de Databricks
- ✓ Créer des dashboards avec le SQL Analytics
- ✓ Connecter Databricks et Power BI

Public concerné

Data engineers, collaborateurs impliqués dans la mise en place, l'automatisation et l'industrialisation de pipelines data sur des plateformes cloud modernes.

Prérequis

Bonne connaissance de SQL et de la manipulation de données. Connaissance de base de Python.

Vérifiez que vous avez les prérequis nécessaires pour profiter pleinement de cette formation en faisant [ce test](#).

PARTICIPANTS

Data engineers, collaborateurs impliqués dans la mise en place, l'automatisation et l'industrialisation de pipelines data sur des plateformes cloud modernes.

PRÉREQUIS

Bonne connaissance de SQL et de la manipulation de données.

Connaissance de base de Python.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

Méthodes et moyens pédagogiques

Travaux pratiques

Échanges, exercices et travaux pratiques.

Méthodes pédagogiques

Active

Modalités d'évaluation

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

Programme de la formation

1 Introduction

- Azure : les grands principes du cloud, présentation et intérêts

2 Databricks : vue d'ensemble

- Présentation d'Azure Databricks : les concepts (data lakehouse, ETL, data warehouse)
- Gouvernance des données : concepts du Unity Catalog
- Avantages de Databricks

3 Les espaces de travail

- Les ressources Azure Databricks
- La création d'un espace de travail
- La connexion de l'espace de travail à des sources de données externes

Travaux pratiques

Créer un espace de travail et le connecter à des sources de données externes.

4 Les différents types de ressources : clusters et SQL warehouse

- Présentation
- Les différents types de calculs
- La création d'un cluster
- La gestion du cluster

Travaux pratiques

Identifier les différents types de calculs, créer et gérer un cluster.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les formations pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque formation ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le participant a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Pour toute question ou besoin relatif à l'accessibilité, vous pouvez joindre notre équipe PSH par e-mail à l'adresse psh-accueil@orsys.fr.

5 Unity Catalog

- La création d'un catalogue
- La connexion à des données à partir de Unity Catalog
- L'ajout d'une table dans le schéma du catalogue
- La définition des autorisations sur une table du catalogue
- Focus : connexion à des bases de données SQL
- D'autres méthodes pour ajouter des données au catalogue

Travaux pratiques

Créer un catalogue et y connecter des données. Ajouter une table, définir des autorisations et maîtriser la connexion à des bases de données SQL.

6 Databricks : les notebooks

- Les objectifs et les intérêts des notebooks
- Le principe des cellules
- Les commandes
- Les magic command
- Programmation PySpark, SQL...
- L'utilisation des cellules et la visualisation d'un résultat
- L'export et le partage des notebooks

Travaux pratiques

Connecter Databricks sur des outils de stockage (ADLS, base de données SQL, API), requêter, transformer et stocker des données dans Databricks. Comprendre les différents formats possibles (Parquet, Delta, table externe, etc.).

7 Principe de Spark et Delta

- Présentation et prise en main du Spark SQL
- Concepts de delta lake
- Gestion du cycle de vie des tables : merge delta, delete/insert

8 Orchestration des traitements

- L'appel de notebooks depuis Azure data factory et la configuration
- L'utilisation des workflows Databricks
- Le paramétrage et la configuration des notebooks Databricks depuis un workflow
- Introduction à Databricks Delta Live Tables

Travaux pratiques

Manipuler les workflows Databricks, paramétrer et configurer des notebooks Databricks depuis un workflow.

9 Databricks SQL

- Databricks SQL : administration
- L'éditeur SQL
- La création et l'exécution d'une requête
- Planifier une requête
- Créer des alertes
- Le partage du résultat d'une requête
- Les options disponibles sur les résultats d'une requête : export, visualisation, filtres d'affichage, paramétrage d'un résultat

Travaux pratiques

Créer une requête, en partager le résultat.