

# Programme de formation

## Azure Databricks

(Mis à jour le 09/02/2024)

### DESCRIPTION DE LA FORMATION :

Dans cette formation, vous apprendrez à utiliser le service analytique Azure Databricks afin de vous permettre d'accélérer le traitement de vos données ! Vous allez découvrir les usages et fonctionnalités du service, puis vous serez amené à créer et configurer un cluster, analyser les données avec PySpark sur un notebook, gérer les tables ou encore créer et lire des secrets avec Databricks.

### OBJECTIFS PÉDAGOGIQUES :

A l'issue de cette formation, les participants seront en capacité de :

- Provisionner des clusters Azure Databricks
- Maitriser les notebooks Azure Databricks
- Orchestrer des travaux sur Azure Databricks
- Utiliser Unity Catalog avec Azure Databricks
- Gérer les secrets sur Azure Databricks
- Optimiser son utilisation d'Azure Databricks
- Implémenter du CI/CD avec Azure Databricks

### MÉTHODES PÉDAGOGIQUES :

- Cette formation sera principalement constituée de théorie et d'ateliers techniques qui permettront d'être rapidement opérationnel.
- Support : un support de cours en anglais sera remis aux participants à la fin de la formation.
- Evaluation : Les acquis sont évalués tout au long de la formation par le formateur (Questions régulières, travaux pratiques, QCM ou autres méthodes).
- Formateur : le tout animé par un consultant-formateur expérimenté, nourri d'une expérience terrain.
- Satisfaction : à l'issue de la formation, chaque participant répond à un questionnaire d'évaluation qui est ensuite analysé en vue de maintenir et d'améliorer la qualité de nos formations.
- Suivi : une feuille d'émargement par demi-journée de présence est signée par chacun des participants.
- Cette formation peut être dispensée en format inter-entreprises ou intra-entreprise sur demande et en mode présentiel comme en distanciel.

## Introduction

- Comprendre le service Azure Databricks.
- Découvrir les fonctionnalités et les usages du service.
- Appréhender l'architecture de Azure Databricks.
- Déployer une ressource Azure Databricks.

## PROGRAMME DE FORMATION

### Provisionner des clusters Azure Databricks

- Se rappeler de l'architecture d'Apache Spark.
- Connaître les différents types et modes de clusters.
- Configurer un cluster Azure Databricks.
- Créer et configurer un cluster Azure Databricks.

### Maitriser les Notebooks Azure Databricks

- Comprendre les notebooks et leurs utilités.
- Connaître les commandes magiques.
- Utiliser les utilitaires Databricks et le profilage intégré.
- Comprendre le versioning des notebooks.
- Analyser des données avec PySpark sur un notebook Databricks.

### Orchestrer des travaux sur Azure Databricks

- Comprendre la notion de job et de job multitâches.
- Savoir configurer un job Azure Databricks.
- Maitriser les pipelines Delta Live Tables.
- Créer et configurer un job Azure Databricks.
- Créer et orchestrer un pipeline DLT.

### Utiliser Unity Catalog avec Azure Databricks

- Comprendre l'utilisation de Unity Catalog.
- Configurer Unity Catalog.
- Organiser les catalogues.
- Gérer les permissions avec le niveau account.
- Créer des tables externes ou managées.

### Gérer les secrets sur Azure Databricks

- Comprendre la notion de secret sur Azure Databricks.
- Utiliser le Databricks CLI pour ajouter des secrets.
- Créer et lire des secrets sur Databricks.

### Optimiser son utilisation d'Azure Databricks

- Comprendre le modèle de billing du service.
- Connaître les bonnes pratiques pour réduire les coûts.
- Choisir le bon type de cluster.

### Implémenter du CI/CD avec Databricks

- Comprendre le CI/CD dans le cadre Azure Databricks.
- Implémenter un pipeline CI/CD sur Azure Devops.

Pour suivre cette formation, vous devez avoir suivi la formation « AZ-900 : Azure Fundamentals » où avoir un niveau équivalent. Des notions de Python sont nécessaire et des notions de Spark sont recommandées.

**PRÉ** Un niveau d'anglais B1 est recommandé, retrouvez les niveaux sur ce lien : [Classification des niveaux de langue](#)

Un PC avec un double écran est recommandé pour plus de confort.

: 2 jours (14 heures)

Spécialistes de la donnée, Data Ingénieurs, PO Data

**DUREE** Intermédiaire

**INTERLOCUTEURS :**

**NIVEAU :**