



# Formation Azure Databricks

**DataTipsLearning.com**

**Tarif :**  
**1600 € HT**

**Durée :**  
**3 Jours**  
**(21h)**

**Lieu :**  
**PARIS/**  
**Distancielle/**  
**Votre Local**

## **Contact :**

contact@datatipslearning.com

+33 0751386021

Réf. DTPSL-AZDTB-2022

[S'inscrire](#)

## **Plus de détails :**

Objectifs

Public / Prérequis

Programme détaillé

Méthode pédagogique

Azure Databricks est une plateforme d'analytique données optimisée pour le cloud Microsoft Azure. Databricks propose trois environnements pour développer des applications impliquant beaucoup de données : Databricks SQL, Databricks Data Science & Engineering et Databricks Machine Learning.

Azure Databricks fournit les dernières versions de Spark et permet une intégration transparente aux bibliothèques open source, création rapide des clusters avec un environnement managé grâce à la mise à l'échelle et à la disponibilité d'Azure.

# Objectifs

- Présenter Databricks et Apache Spark.
- Créer un Workspace et un cluster.
- Se connecter à de différentes sources pour lire et écrire des données.
- Comprendre les différents types de cluster et de ressources.
- Mettre en œuvre un traitement en streaming.
- Connecter Databricks à des outils gestion de versions comme Git.
- Comprendre la partie DevOps & Sécurité de Databricks.

# Public

- Data Engineer
- Data Architectes
- Data Scientist
- Data Analyst

# Prérequis

- Connaissance de base d'Apache Spark.
- Quelques connaissances du Cloud.

# Programme détaillé

- I. Introduction à Azure Databricks
- II. Présentation des fondamentaux de Spark
- III. Conception de clusters Spark : Workspace, Runtimes, Clusters, Notebooks, Workflows & Libraires
- IV. Importer des données dans Databricks
- V. Connexion vers et depuis Databricks
- VI. Traitement en batch et en temps réel dans Databricks
- VII. Montage d'Azure Data Lake dans Databricks
- VIII. ML dans Databricks
- IX. Graphe dans Databricks
- X. Système de fichiers Databricks (DBFS)
- XI. Lecture et écriture depuis et vers divers services Azure et formats de fichiers

- XII. Intégration de Git avec Databricks Repos
- XIII. Les pools Databricks
- XIV. Les instances spot de Databricks
- XV. Les managed resource groups de Databricks
- XVI. Présentation de Databricks Delta Lake
- XVII. Databricks avec Azure Key Vault, App Configuration et Log Analytics
- XVIII. DevOps CI/CD pour Azure Databricks
- XIX. Sécurité et Monitoring dans Azure Databricks
- XX. Gouvernance des données
- XXI. Data sharing

## Méthode pédagogique

La formation se compose d'une partie théorique, et également une partie pratique représentant 60% de de la formation.

La partie pratique contient plusieurs exercices sous forme de notebook Databricks avec les corrections, avec aussi un projet à la fin de la formation comme simulation d'une prod.

Chaque jour, une évaluation rapide des connaissances est effectuée avant de commencer les nouvelles parties de la formation.

A la fin, une synthèse globale est délivré aux stagiaires, renforcé par un projet prod.

Finalement, une évaluation QCM est proposée.

Un support de cours sera remis à chaque stagiaire comprenant les slides, les exercices et les corrigés et un git du projet prod.

Une feuille de présence est fournie en fin de formation avec une certificat de complétion de formation pour chaque stagiaire.

Le formateur est un Data Engineer expert, qui intervient sur le sujet depuis plusieurs années en formation mais aussi en conseil.