

# Bonnes pratiques Big Data & Analytics avec Python/PySpark en environnement Databricks

Maîtriser les bonnes pratiques d'une approche Big Data & Analytics en Python/PySpark dans un environnement Databricks

7h

## Objectifs pédagogiques

- Mettre en oeuvre les notions essentielles du data management (cycle de vie des données, qualité des données, architecture et stockage, gestion des données sensibles)
- Mettre en oeuvre les bonnes pratiques de codage et de manipulation de données avec Python/PySpark en environnement Databricks (efficacité, maîtrise de coûts, réduction de l'empreinte environnementale)
- Mettre en oeuvre les bonnes pratiques de sécurité et dans l'utilisation des librairies
- Passer efficacement de code type SAS vers du code Python/PySpark

Modalités d'évaluation : Les objectifs pédagogiques sont évalués à travers la réalisation des parties pratiques (labs dirigés) sous la supervision du formateur délivrant la session de formation.

## Public cible

- Chargé d'études statistiques, Actuaire et Data Scientists
- Consultants, développeurs, chefs de projet, data scientists, data engineers

# Prérequis

- Avoir suivi le cours 'Python pour la science des données en environnement Databricks' ou avoir des connaissances équivalentes
- Avoir suivi le cours 'PySpark pour la science des données en environnement Databricks' ou avoir des connaissances équivalentes

## Programme

### **Module 1: les notions essentielles du data management**

- Cycle de vie des données
- Qualité des données
- Architecture et stockage
- Gestion des données sensibles

Ateliers:

- Data Management avec Python/PySpark en environnement Databricks

### **Module 2: bonnes pratiques de codage et de manipulation de données**

- Performance du code et des requêtes
- Efficience maîtrise de coûts
- Réduction de l'empreinte environnementale
- Pratiques de sécurité et dans l'utilisation des librairies

Ateliers:

- Appliquer les meilleurs pratiques de codage et de manipulation de données avec Python/PySpark en environnement Databricks

### **Module 3: De SAS vers Python / PySpark en environnement Databricks**

- SAS vs Python/PySpark
- De SAS vers Python
- Accélérateurs: bonnes pratiques, IA génératives, ...

Ateliers:

- Cas pratiques pour passer de SAS vers Python/PySpark en environnement Databricks.