



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

Durée : 3 jours

Public :	Data Engineer , Data Architects , Data Scientist , Data Analyst	
Pré-requis :	Connaissance d'apache Spark, cloud, python	
Objectifs :	<p>Cette formation vous fera découvrir l'environnement Databricks, vous apprendra à générer développer des applications impliquant beaucoup de données</p> <p>A l'issue de cette formation, vous aurez acquis les compétences nécessaires pour :</p> <ul style="list-style-type: none"> • 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. 	<p><i>Nombre de stagiaires</i></p> <p>De 3 à 10</p>
Référence :		
Demandeurs d'emploi :	Contactez-nous pour connaître les remises Pôle Emploi	<p><i>Tarifs H.T.</i></p> <p>Standard : 1700 €</p>
Lieu :	Paris, Bordeaux, Lyon, Nantes, Toulouse et Aix-en-Provence	
<p>Pour les personnes en situation de handicap, nous mettrons tout en œuvre pour vous accueillir ou pour vous réorienter. Vous pouvez nous contacter au formation@lecoledeladata.com</p>		

■ Dates et horaires

3 jours à définir - De 9h à 12h30 et de 14h à 17h30

■ Contenu pédagogique – Programme de la formation Databricks

I. Introduction : Découvrir Databricks

- Problématique
- Qu'est-ce que Databricks ?
- Architecture du produit Databricks
- Apprendre à installer/configurer Databricks

Cas Pratique : Configurer 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

Moyens pédagogiques et d'encadrement mis en œuvre :

Nous alternerons régulièrement théorie et pratique : chacun dispose d'un poste de travail pour mettre en pratique au fur et à mesure les notions abordées

Supports pédagogiques

Il est remis aux stagiaires :

- les logiciels (libres et gratuits), en toute légalité
- le plan de déroulement de la formation
- une bibliographie complète
- des schémas explicatifs (sous forme numérique)

Modalités de suivi et dispositif d'appréciation des résultats :

Évaluation formative en continu ; Commentaire personnel en fin de formation si souhaité. Bilan commun le dernier jour.