Databricks pour Utilisateurs

Description

Cette formation de 3 jours vous permettra de découvrir et de maîtriser les principales fonctionnalités de Databricks en passant par les bases de l'environnement, la gestion des notebooks, l'ingestion et la transformation des données, et jusqu'à l'optimisation des performances.

Durée

3.00 jours (14.21 heures)

Profils des stagiaires et prérequis

Profils :

- Data Analysts & Business Analysts souhaitant utiliser Databricks SQL pour interroger et analyser les données
- Data Engineers cherchant à maîtriser PySpark et Delta Lake pour construire des pipelines de données
- Chefs de projet & Responsables Data souhaitant mieux comprendre la plateforme pour la mise en œuvre de projets analytiques
- Data Scientists désirant exploiter Databricks MLflow et les modèles d'apprentissage automatique

Prérequis :

- Connaissances de base en SQL et Python
- Sensibilisation aux concepts Big Data et Cloud Computing
- Expérience en manipulation de données avec des outils analytiques (Excel, Pandas, SQL, etc.)

Objectifs pédagogiques

- Comprendre les fondamentaux de Databricks et son positionnement dans l'écosystème Big Data
- Explorer l'interface Databricks et manipuler des notebooks collaboratifs
- Ingérer, transformer et analyser des jeux de données à l'aide de PySpark et SQL
- Maîtriser l'architecture des clusters Databricks et optimiser les performances
- Construire des pipelines de données et orchestrer des workflows
- Appliquer des cas d'usage concrets autour de la business intelligence, de la data science et du machine learning

Programme détaillé

- Jour 1 : Introduction à Databricks et Premiers Pas
 - Présentation de Databricks
 - Prise en main de l'interface
 - Atelier Pratique : Manipulation des Notebooks
- Jour 2 : Gestion et Transformation des Données avec Databricks
 - Chargement et Exploration des Données
 - Traitement et Transformation avec Spark
 - Optimisation des Performances

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. +33172254082 | E-mail : info@stat4decision.com | www.stat4decision.com

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: Ile de France)

Organisme de formation certifié Qualiopi pour ses actions de formation

- Atelier Pratique : Pipeline de Traitement des Données
- Jour 3 : Automatisation, Orchestration et Machine Learning
 - Automatisation et Orchestration
 - Databricks et Machine Learning
 - Cas Pratique : Mise en Production d'un Pipeline
 - Clôture et Perspectives

Organisation de la formation

Moyens pédagogiques et techniques

- Support PDF / HTML
- Cas d'usages de données
- Outils utilisé : plateforme Databricks

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Quiz et QCM à la fin de chaque module pour valider l'acquisition des connaissances théoriques.
- Mini-projets guidés
- Auto-évaluation sur la compréhension et la mise en application des concepts.
- Feedback individuel : chaque participant se voit proposer un bilan de ses points forts et axes d'amélioration.

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. +33172254082 | E-mail : info@stat4decision.com | www.stat4decision.com

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: Ile de France)

Organisme de formation certifié Qualiopi pour ses actions de formation