

Mis à jour le 20/06/2024

S'inscrire

Formation DataFold

3 jours (21 heures)

Présentation

Automatisez vos tests de données grâce à notre formation DataFold. DataFold est l'outil de référence pour le [data testing](#) en ingénierie des données.

Notre cours vous enseignera l'installation et le paramétrage de l'outil, vous le prendrez en main pour la première fois afin d'exécuter vos premiers tests (test de validation, test de schéma et détection d'anomalies).

Nous vous apprendrons à écrire des tests de données avancées via des paramètres et des assertions personnalisés, vous saurez également valider vos pipelines pour assurer la qualité de vos actifs.

Ce cours comprend un chapitre QA essentiel pour l'analyse de vos modifications lors de vos tests. Ce programme comprend également une partie [CI/CD](#) pour appliquer l'intégration continue au sein de vos pipelines de données.

Par ailleurs, vous connaîtrez les bonnes pratiques pour gérer correctement vos projets data en alliant collaboration, documentation et gestion des versions sur DataFold.

Objectifs

- Maîtriser les fondamentaux du data testing
- Mettre en place des techniques avancées de test de données
- Mettre en place des pipelines de validation des données
- Intégrer DataFold dans ses workflows CI/CD
- Automatiser les tests et les déploiements data

Public visé

- Chefs de projet
- Data analysts
- Data scientists
- Data engineers
- Business analysts

Pré-requis

Expérience en gestion des données.

PROGRAMME DE NOTRE FORMATION DATAFOLD

INTRODUCTION À DATAFOLD

- Présentation de DataFold
- Les fonctionnalités clés
- Installation et configuration initiale
- Tour d'horizon de l'interface utilisateur

LES FONDAMENTAUX DU DATA TESTING

- Introduction aux tests de données
- Types de tests de données
 - Validation des données
 - Test des schémas
 - Détection des anomalies
- Créer des tests basiques
- Définition des cas de test
- Mise en place des sources de données
- Rédaction et exécution des tests
- Interprétation des résultats

Techniques avancées de test de données

- Scénarios de tests complexes
- Écrire des cas de test avancés
- Utilisation d'assertions personnalisées
- Tests paramétrés

VALIDATION DES DONNÉES

- Techniques de validation des données
- Automatisation des tests de validation
- Mise en place de pipelines de validation des données
- Automatiser les notifications et les rapports

- Analyse des résultats de validation
- TP : Validation des données dans un pipeline

QA

- Comprendre l'impact des modifications
- Utilisation de DataFold pour analyser les impacts des changements
- Prévion des impacts sur les pipelines de données
- TP : Test QA

INTÉGRATION ET DÉPLOIEMENT CONTINU (CI/CD)

- Principes CI/CD pour les pipelines de données
- Intégration de DataFold dans les workflows CI/CD
- Automatisation des tests et des déploiements
- Surveillance et gestion des déploiements
- TP : Mise en place d'un workflow CI/CD avec DataFold

PIPELINES DE DONNÉES

- Optimisation des performances
- Résoudre les goulots d'étranglement
- Le suivi des performances
- TP : Optimisation d'un pipeline de données

SÉCURITÉ DES DONNÉES

- Principes de sécurité des données
- Sécuriser ses données
- Surveillance des accès et des utilisations

GESTION DE PROJETS DATA

- Les bonnes pratiques de collaboration
- Gestion de projets de données avec DataFold
- Documentation et communication des résultats
- Gestion des versions et des modifications
- TP : Collaboration sur un projet de données

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.