

DevOps pour la data

Description

Imaginez pouvoir automatiser vos déploiements, intégrer des tests continus, et garantir une infrastructure stable et scalable pour vos applications d'analyse de données.
Cette formation DevOps orientée Data vous offre l'opportunité de maîtriser les principes et outils DevOps appliqués à la data science.

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data Analysts et Business Analysts souhaitant mieux structurer leurs workflows et automatiser leurs analyses
- Data Scientists voulant apprendre à industrialiser et déployer leurs modèles
- Chefs de projet Data désirant mieux comprendre et intégrer les bonnes pratiques DevOps dans leurs équipes
- Ingénieurs Data voulant améliorer la mise en production des pipelines de données

Prérequis :

- Expérience en manipulation de données et analyses statistiques
- Aucune connaissance préalable en DevOps ou Cloud requise

Objectifs pédagogiques

- Comprendre les principes fondamentaux du DevOps appliqués aux projets data
- Maîtriser Git et GitHub/GitLab pour le versioning et la collaboration
- Mettre en place des workflows d'Intégration et Déploiement Continu (CI/CD) pour les scripts analytiques et modèles ML
- Automatiser le traitement et l'orchestration des pipelines de données avec Apache Airflow
- Déployer des modèles et analyses dans un environnement cloud
- Appliquer les bonnes pratiques de surveillance et de monitoring des analyses et modèles en production

Programme détaillé

- Introduction au DevOps appliqué à la Data Science
 - Définition et principes fondamentaux du DevOps
 - Différences entre DevOps classique et MLOps / DataOps
 - Pourquoi le DevOps est crucial pour les Data Analysts et Data Scientists
 - Présentation des outils utilisés : Git, Docker, Airflow, MLflow
- Gestion du Versioning et de la Collaboration avec Git
 - Rappels sur Git et bonnes pratiques du versioning

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

- Utilisation avancée de GitHub/GitLab pour la collaboration
- Gestion des branches, des merge requests et des revues de code
- Automatisation des workflows avec GitHub Actions ou GitLab CI/CD
- Atelier pratique : Mise en place d'un projet analytique versionné avec GitHub/GitLab
- Orchestration et Automatisation des Pipelines de Données
 - Introduction aux pipelines de données
 - Automatisation et planification des traitements avec Apache Airflow
 - Création de DAGs (Directed Acyclic Graphs) pour orchestrer des tâches analytiques
 - Intégration d'Airflow avec Python, SQL et APIs
 - Atelier pratique : Mise en place d'un pipeline Airflow pour l'automatisation d'une analyse
- Industrialisation des Scripts et Déploiement de Modèles en Production
 - Introduction aux concepts d'Intégration Continue (CI) et Déploiement Continu (CD)
 - Utilisation de Docker pour la containerisation des analyses et modèles ML
 - Déploiement des modèles avec FastAPI et Flask
 - Introduction aux solutions cloud : AWS SageMaker, Azure ML, Google Vertex AI
 - Atelier pratique : Déploiement d'une API analytique ou d'un modèle ML dans un conteneur Docker
- Surveillance et Monitoring des Analyses et Modèles en Production
 - Pourquoi et comment monitorer un projet data ?
 - Utilisation de MLflow pour suivre les performances des modèles ML
 - Bonnes pratiques pour la gestion des erreurs et des logs
 - Surveillance et alerte sur la dérive des modèles et qualité des données
 - Atelier pratique : Monitoring et mise à jour d'un modèle en production avec MLflow
- Synthèse et Bonnes Pratiques du DevOps pour la Data
 - Retour sur les principes clés abordés
 - Revue des bonnes pratiques pour l'industrialisation des analyses
 - Plan d'action pour intégrer les outils DevOps dans son quotidien
 - Session de questions-réponses et retour d'expérience

Organisation de la formation

Moyens pédagogiques et techniques

- Documents supports de formation projetés.
- Exposés théoriques
- Etude de cas concrets
- Outils : VSCode, GitHub, GitLab pour pratiquer et illustrer le devops

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

- Cas pratiques validés par le formateur
- Mises en situation.
- Exercices avec auto-corrections à faire pendant la formation
- Évaluation de la compréhension par le biais de quiz interactifs

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