

Mis à jour le 13/12/2023

S'inscrire

Formation Data Version Control : gestion avancée des données

2 jours (14 heures)

Présentation

Notre formation Data Version Control vous permettra de maîtriser DVC pour une gestion optimale de vos données. Vous serez en mesure de gérer facilement les versions de vos ensembles de données, de suivre les modifications apportées au fil du temps et de revenir à des versions antérieures si nécessaire.

Notre formation DVC vous permettra de partager et de synchroniser vos ensembles de données, pour une collaboration efficace sur vos projets. Vous pourrez efficacement gérer des ensembles de données volumineux en optimisant le stockage, en utilisant des fonctionnalités d'échantillonnage et en économisant de l'espace.

Dans cette formation, vous découvrirez comment intégrer DVC avec d'autres outils couramment utilisés dans le domaine de la science des données, tels que TensorFlow, PyTorch et MLflow, pour une intégration harmonieuse dans vos workflows. Vous bénéficierez de conseils pratiques, de bonnes pratiques et d'astuces pour structurer votre projet, optimiser les performances et résoudre les problèmes courants.

Objectifs

- Comprendre les concepts fondamentaux de DVC et ses fonctionnalités clés
- Maîtriser la gestion des versions des ensembles de données avec DVC
- Apprendre à collaborer efficacement avec d'autres membres de l'équipe grâce à DVC

Public visé

- Data scientists
- Ingénieurs en apprentissage automatique
- Data analysts

Pré-requis

- Connaissance de base en programmation
- Compréhension des concepts de base des systèmes de contrôle de version

Programme de notre formation DVC

Introduction à DVC

- Qu'est-ce que DVC ?
- Contrôle de version pour le code source vs contrôle de version pour les données
- Vue d'ensemble des fonctionnalités principales de DVC

Configuration de base

- Installation de DVC et de ses dépendances
- Initialisation d'un projet
- Configuration de l'emplacement de stockage des données

Gestion des ensembles de données

- Ajout d'un ensemble de données à DVC
- Suivi des modifications apportées à un ensemble de données
- Création de versions des ensembles de données

Collaboration et partage

- Utilisation de dépôts distants pour la collaboration
- Partage des ensembles de données
- Récupération des ensembles de données partagés

Expérimentation et reproductibilité

- Création d'une configuration d'expérience
- Enregistrement des résultats et des métriques expérimentales
- Reproduction des expériences précédentes

Ensembles de données volumineux

- Gestion des ensembles de données volumineux

- Stratégies de stockage efficace des données
- Fonctionnalités d'échantillonnage

Intégration d'outils externes

- Intégration de frameworks d'apprentissage automatique (TensorFlow, PyTorch)
- Plates-formes de gestion des expériences (MLflow)

Fonctions Avancé

- Utilisation de DVC avec des pipelines de traitement des données
- Branches et expériences parallèles
- Tags pour marquer des versions spécifiques des ensembles de données

Sécurité et gestion des autorisations

- Contrôle d'accès aux ensembles de données
- Gestion des secrets
- Audit et suivi des modifications
- Intégration avec les outils de sécurité existants

Déploiement et déploiement en production

- Ensembles de données en production
- Déploiement en production (Kubernetes, Docker)

Études de cas

- Projet d'apprentissage automatique
- Projet de recherche en sciences des données

Ressources supplémentaires

- Documentation officielle et ressources en ligne
- Communauté et forums

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.