



MLOps et DataOps

AWS - Concevoir, déployer et surveiller un pipeline MLOps

5 jours (35h00) | ★★★★★ 4,6/5 | MLOPS-AWS | Évaluation qualitative de fin de stage | Formation délivrée en présentiel ou distanciel

Formations Informatique › Big Data › MLOps et DataOps

Contenu mis à jour le 13/10/2023. Document téléchargé le 23/05/2024.

Objectifs de formation

A l'issue de cette formation, vous serez capable de :

- Concevoir et tester un modèle de ML (Machine Learning) sur AWS
- Concevoir et tester un pipeline MLOps robuste
- Déployer et automatiser votre pipeline MLOps de production et le passer à l'échelle
- Surveiller votre modèle de production et corriger les écarts.

Modalités, méthodes et moyens pédagogiques

Formation délivrée en présentiel ou distanciel* (blended-learning, e-learning, classe virtuelle, présentiel à distance).

Le formateur alterne entre méthode** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation).

Variables suivant les formations, les moyens pédagogiques mis en oeuvre sont :

- Ordinateurs Mac ou PC (sauf pour certains cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel)
- Environnements de formation installés sur les postes de travail ou en ligne
- Supports de cours et exercices

En cas de formation intra sur site externe à M2i, le client s'assure et s'engage également à avoir toutes les ressources matérielles pédagogiques nécessaires (équipements informatiques...) au bon déroulement de l'action de formation visée conformément aux prérequis indiqués dans le programme de formation communiqué.

* nous consulter pour la faisabilité en distanciel

** ratio variable selon le cours suivi

Prérequis

Avoir des connaissances en bases de données et/ou analytiques.

Public concerné

Data Scientists, Data Analysts, ingénieurs Data, ingénieurs DevOps, développeurs, architectes Data, chefs de projets, managers.

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence élargée par demi-journée par les stagiaires et le formateur.

Programme

Jour 1

Fondamentaux

- Qu'est-ce qu'une donnée ?
- Cycle de vie d'une donnée
- Cycle de vie de la Data Science
- Evolution des analytiques
- Apports des analytiques et du Cloud
- Qu'est-ce qu'une organisation Data Driven ?
- Gouvernance et qualité des données
- Lean, Agilité et DevOps, MLOps appliqués aux données
- Conteneurisation et architectures micro-services
- Apports de MLOps
- Workflow de MLOps
- DevOps vs DataOps vs MLOps
- Machine Learning et opérations : briser le mur de la confusion
- Constituer et former une équipe MLOps
- Amazon SageMaker
- Amazon AI Services et AutoML

Exemple de travaux pratiques (à titre indicatif)

- Mise en pratique : création d'un pipeline MLOps simple

Mise en place de l'infrastructure et des outils

- Ingestion, exploration, et préparation des données dans AWS
- Apprentissage et optimisation du modèle avec SageMaker
- Déploiement de modèles avec SageMaker et AWS Lambda Functions
- Streaming Analytics et Machine Learning sur AWS
- AWS Infrastructure et hardware construit sur-mesure
- Réduire les coûts avec les Tags, Budgets et Alertes

Jour 2

Machine Learning automatisé

- Machine Learning automatisé avec SageMaker Autopilot
- Suivre les expérimentations avec SageMaker Autopilot
- Entraîner et déployer un classificateur de texte avec SageMaker Autopilot
- Machine Learning automatisé avec Amazon Comprehend

Ingérer des données dans le Cloud

- Requêter Amazon S3 Data Lake avec Amazon Athena
- Ingérer des nouvelles données en continu avec AWS Glue Crawler
- Construire un Lakehouse avec Amazon Redshift Spectrum
- Choisir entre Amazon Athena et Amazon Redshift
- Réduire les coûts et augmenter la performance

Jour 3

Explorer le jeu de données

- Outils pour explorer les données dans AWS
- Visualiser le Data Lake avec SageMaker Studio
- Requêter le Data Warehouse
- Créer des tableaux de bord avec Amazon QuickSight
- Détecter des problèmes de qualité avec SageMaker et Apache Spark
- Détection du biais dans notre jeu de données
- Détecter différents types de drift avec SageMaker Clarity
- Analyser les données avec AWS Glue DataBrew
- Réduire les coûts et augmenter la performance

Préparer le jeu de données pour l'apprentissage

- Réaliser la sélection de variables et l'ingénierie des attributs
- Passer la sélection de variable à l'échelle avec SageMaker Processing Jobs
- Partager les attributs avec d'autres utilisateurs avec SageMaker Feature Store
- Ingérer et transformer des données avec SageMaker Data Wrangler
- Suivre le "lineage" des artéfacts et expérimentations avec SageMaker
- Ingérer et transformer des données avec Glue DataBrew

Entraîner un premier modèle

- Comprendre l'infrastructure de SageMaker
- Déployer un modèle BERT pré-entraîné avec SageMaker JumpStart
- Développer un modèle SageMaker
- Architecture de BERT
- Entraîner BERT en partant de rien
- Affiner un modèle BERT pré-entraîné
- Créer le script d'apprentissage
- Lancer le script d'apprentissage à partir de SageMaker Notebook
- Evaluer les modèles
- Déboguer et profiler l'apprentissage avec SageMaker Debugger
- Interpréter et expliquer les prédictions des modèles
- Réduire les coûts et améliorer la performance

Jour 4

Entériner et optimiser des modèles à l'échelle

- Trouver automatiquement les meilleurs hyperparamètres
- Utiliser Warm Start pour des jobs supplémentaires d'optimisation des hyperparamètres
- Passer à l'échelle avec SageMaker Distributed Training
- Réduire les coûts et augmenter la performance

Déployer des modèles en production

- Choisir entre des prédictions batch et temps réel
- Prédiction en temps réel avec SageMaker Endpoints
- Passer automatiquement à l'échelle Endpoints avec CloudWatch
- Stratégies pour le déploiement de modèles nouveaux ou mis à jour
- Tester et comparer des nouveaux modèles

- Monitorer la performance et le drift
- Monitorer la qualité du modèle et des données
- Réaliser des prédictions batch avec SageMaker Batch Transform
- AWS Lambda Functions et Amazon API Gateway
- Optimiser et gérer des modèles sur le bord (edge)
- Déployer des modèles PyTorch avec TorchServe
- Inférence TensorFlow-BERT avec AWS Deep Java Library
- Réduire les coûts et augmenter les performances

Jour 5

Pipelines et MLOps

- Pipelines logiciels
- Pipelines de ML
- Orchestration de pipelines avec SageMaker Pipelines
- Automatisation avec SageMaker Pipelines
- Workflows avec "humain dans la boucle" (human-in-the-loop)
- Réduire les coûts et augmenter la performance

Analytiques en streaming et Machine Learning

- Apprentissage Online vs Offline
- Applications de streaming
- Requêtes fenêtrées sur des données en streaming
- Classification avec Amazon Kinesis, AWS Lambda et Amazon SageMaker
- Implémenter l'ingestion de données en streaming avec Amazon Kinesis Data Firehose
- Analytiques en temps réel avec Streaming Analytics
- Mettre en place Amazon Kinesis Data Analytics
- Classification avec Apache Kafka, AWS Lambda et Amazon SageMaker

Meilleures pratiques de sécurité sur AWS

- Modèle de responsabilité partagée entre AWS et le client
- Appliquer AWS IAM
- Isoler les environnements de calcul et de réseau
- Sécuriser l'accès aux données Amazon S3
- Chiffrement au repos
- Chiffrement en transit
- Sécuriser les instances SageMaker Notebook
- Sécuriser SageMaker Studio
- Sécuriser SageMaker Jobs and Models
- Sécuriser Lake Formation
- Sécuriser Database Credentials avec AWS Secrets Manager
- Bonnes pratiques de gouvernance et d'auditabilité

Le contenu de ce programme peut faire l'objet d'adaptation selon les niveaux, prérequis et besoins des apprenants.

Modalités d'évaluation des acquis

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation

Accessibilité de la formation

Le groupe M2i s'engage pour faciliter l'accessibilité de ses formations. Les détails de l'accueil des personnes en situation de handicap sont consultables sur la page Accueil et Handicap.

Modalités et délais d'accès à la formation

Les formations M2i sont disponibles selon les modalités proposées sur la page programme.
Les inscriptions sont possibles jusqu'à 48 heures ouvrées avant le début de la formation.
Dans le cas d'une formation financée par le CPF, ce délai est porté à 11 jours ouvrés.