



Big Data - MLOps et DataOps

Google Cloud - Concevoir, déployer et surveiller un pipeline MLOps

5 jours (35h00) | 9 4,6/5 | MLOPS-GOO | Évaluation qualitative de fin de stage | Formation délivrée en présentiel ou distanciel

Formations Informatique > Data > Big Data - MLOps et DataOps

Contenu mis à jour le 13/10/2023. Document téléchargé le 27/07/2024.

Objectifs de formation

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

- Concevoir et tester un modèle de ML (Machine Learning) sur Google Cloud à partir d'un cas pratique fil rouge
- 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
- Google Cloud Platform

Exemple de travaux pratiques (à titre indicatif)

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

Ingérer des données dans Google Cloud

- Séparation du calcul et du stockage
- Ingérer des données
- Charger des données dans Google BigQuery
- Déployer, invoquer, et planifier Cloud Run

Jour 2

Données de streaming : publier et ingérer avec Pub/Sub et Dataflow

- Concevoir l'Event Feed
- Publier un Event Stream sur Cloud Pub/Sub
- Traitement en temps réel
- Tableau de bord en temps réel

Exploration des données interactive avec Vertex AI Workbench

- Analyse exploratoire des données avec SQL
- Analyse exploratoire avec Vertex AI Workbench
- Evaluation du modèle

Jour 3

Classification bayésienne avec Apache Spark sur Cloud Dataproc

- Introduction à Google Cloud Dataproc
- Utilisation de Spark SQL
- Classification bayésienne
- Orchestration
- Spark en serverless

Régression logistique avec Spark ML

- Régression logistique avec Spark
- Ingénierie des attributs
- Sélection de variables
- Transformation de variables
- Données catégorielles

Jour 4

Machine Learning avec BigQuery ML

- Régression logistique avec BigQuery ML
- Machine Learning non-linéaire avec XGBoost

ML avec TensorFlow dans Vertex AI

- Vers des modèles plus complexes
- Apprentissage et évaluation avec Keras
- Modèles larges et profonds avec Keras
- Déployer un modèle TensorFlow entraîné dans Vertex AI
- Invoquer le modèle déployé

Jour 5

Attributs à fenêtre temporelle pour du ML en temps réel

- Réaliser le fenêtrage temporel avec Apache Beam et Cloud Dataflow
- Apprentissage du modèle
- Streaming des prédictions
- Pipeline de streaming avec écriture dans BigQuery

Mettre en place MLOps dans Vertex AI

- Développer et déployer avec Python
- Utiliser AutoML
- Optimiser les hyperparamètres
- Expliquer le modèle avec des métadonnées
- Créer et déployer le modèle
- Surveiller le modèle

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.