

Google Cloud Big Data and Machine Learning Fundamentals

Les fondamentaux du Big Data et du Machine Learning sur Google Cloud

7h

Objectifs pédagogiques

- Identifier le cycle de vie données-vers-IA sur Google Cloud et les principaux produits du big data et du machine learning.
- Concevoir des pipelines de streaming avec Dataflow et Pub/Sub.
- Analyser le big data à grande échelle avec BigQuery.
- Identifier différentes options pour créer des solutions de machine learning sur Google Cloud.
- Décrire un flux de travail machine learning et les étapes clés avec Vertex AI.
- Créer un pipeline de machine learning à l'aide d'AutoML.

Public cible

- Analystes de données, data scientists et analystes commerciaux qui débutent avec Google Cloud
- Les personnes chargées de concevoir des pipelines et des architectures pour le traitement des données, de créer et de maintenir des modèles d'apprentissage automatique et statistiques, d'interroger des ensembles de données, de visualiser les résultats des requêtes et de créer des rapports
- Cadres et décideurs informatiques évaluant Google Cloud pour une utilisation par des data scientists

Prérequis

Des connaissances de base sur les sujets suivants aideront les participants à mieux appréhender ce cours:

- Langage de requête de base de données tel que SQL
- Workflow d'ingénierie des données, de l'extraction, de la transformation, du chargement à l'analyse, la modélisation et le déploiement
- Modèles de machine learning, tels que les modèles supervisés par rapport aux modèles non supervisés

Programme

Module 0: Introduction au cours

Sujets

- Cette section souhaite la bienvenue aux apprenants du cours Big Data and Machine Learning Fundamentals et fournit un aperçu de la structure et des objectifs du cours.

Objectifs

- Reconnaître le cycle de vie données-vers-IA sur Google Cloud
- Identifier le lien entre le data engineering et le machine learning

Module 1: Big Data et Machine Learning sur Google Cloud

Sujets

- Cette section explore les composants clés de l'infrastructure de Google Cloud. Nous présentons de nombreux produits et services de big data et de machine learning qui prennent en charge le cycle de vie données-vers-IA sur Google Cloud.

Objectifs

- Identifier les différents aspects de l'infrastructure de Google Cloud.
- Identifier les produits de big data et de machine learning sur Google Cloud.

Activités

- Lab : Explorer un ensemble de données public BigQuery

- Quizz

Module 2: Data Engineering pour le streaming de données

Sujets

- Cette section présente la solution de Google Cloud pour la gestion des données de streaming.
- Elle examine un pipeline de bout en bout, y compris l'ingestion de données avec Pub/Sub, le traitement des données avec Dataflow et la visualisation des données avec Looker et Data Studio.

Objectifs

- Décrire un flux de travail de flux de données de bout en bout, de l'ingestion à la visualisation des données.
- Identifier les défis du pipeline de données moderne et comment les résoudre à grande échelle avec Dataflow.
- Créer des tableaux de bord collaboratifs en temps réel avec des outils de visualisation de données.

Activités

- Lab : Créer un pipeline de flux de données pour un tableau de bord en temps réel avec Dataflow
- Quizz

Module 3: Big Data avec Big Query

Sujets

- Cette section présente aux participants BigQuery, l'entrepôt de données sans serveur entièrement géré de Google.
- Elle explore également BigQuery ML ainsi que les processus et commandes clés utilisés pour créer des modèles de machine learning personnalisés.

Objectifs

- Décrire les principes de base de BigQuery en tant qu'entrepôt de données.
- Expliquer comment BigQuery traite les requêtes et stocke les données.
- Définir les phases du projet BigQuery ML.
- Créez un modèle de machine learning personnalisé avec BigQuery ML.

Activités

- Lab : Prédire les achats des visiteurs à l'aide de BigQuery ML
- Quizz

Module 4 : Les options de Machine Learning sur Google Cloud

Sujets

- Cette section explore quatre options différentes pour créer des modèles de machine learning sur Google Cloud. Il présente également Vertex AI, la plate-forme unifiée de Google pour la création et la gestion du cycle de vie des projets ML.

Objectifs

- Identifier différentes options pour créer des modèles de ML sur Google Cloud.
- Définir Vertex AI et ses principales fonctionnalités et avantages.
- Décrire les solutions d'IA sur les marchés horizontaux et verticaux.

Activités

- Quizz

Module 5: Le flux de travail d'apprentissage automatique avec Vertex AI

Sujets

- Cette section se concentre sur les trois phases clés (préparation des données, entraînement du modèle et préparation du modèle) du flux de travail Machine Learning dans Vertex AI. Les participants peuvent s'entraîner à créer un modèle de machine learning avec AutoML.

Objectifs

- Décrire un flux de travail ML et les étapes clés.
- Identifier les outils et les produits pour soutenir chaque étape.
- Créer un flux de travail de ML de bout en bout à l'aide d'AutoML.

Activités

- Lab : Vertex AI : Prédire le risque de prêt avec AutoML

Module 6: Synthèse du cours

Sujets

- Cette section passe en revue les sujets abordés dans le cours et fournit des ressources supplémentaires pour un apprentissage plus approfondi.

Objectifs

- Décrire le cycle de vie données-vers-IA sur Google Cloud et identifier les principaux produits du big data et du machine learning.