+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Mis à jour le 27/07/2023

S'inscrire

Formation Datalake

2 jours (14 heures)

Présentation

L'augmentation phénoménale des données non structurées nous pousse à maîtriser l'analyse et le traitement des datalakes, aussi appelés « lac de données ». Le data lake est un ensemble regroupant une grande quantité de données hétérogènes. Ces données peuvent désormais être traitées de manière opérationnelle en temps réel. La maîtrise d'une approche data lake vous permettra de découvrir des données rares et diverses pouvant avoir un fort impact sur les performances de votre organisation. Cette formation datalake de 2 jours vous enseignera comment ingérer, modifier et stocker vos données. Vous apprendrez également comment implémenter une architecture lambda.

Objectifs

- Comprendre le métier de Data Engineer
- Monter un Datalake
- Mettre à disposition rapidement un important volume de données prêtes à l’emploi

Public visé

- Data Scientists
- Devops
- Développeurs
- Vous travaillez avec data engineers

Pré-requis

Base en programmation

Programme de notre formation Datalake

QU’EST-CE QUE LE BIG DATA ?

• Définition du Big Data • Définition du Data Engineering • Cartographie des compétences • Principaux use cases de Data Engineering Atelier pratique : mise en place de l’environnement

UTILITÉ D’UN DATALAKE

• Théorème de CAP • Qu’est-ce qu’un ETL et un ELT ? • Définition d’un datalake • Système de fichiers distribués Atelier pratique : manipulation de noeuds HDFS

INGESTION DES DONNÉES

• Manipulation de données en temps réel avec Kafka • Gestion des flux de données avec Nifi Atelier pratique : ingestion de données en temps réel dans un datalake avec Nifi et Kafka

TRANSFORMATION DES DONNÉES

• Différence entre données structurées et non structurées • Le paradigme MapReduce • L’écosystème Hadoop • Transformation de données avec Spark Atelier pratique : uniformisation du format et enrichissement des données sur un data lake

MISE À DISPOSITION DE LA DONNÉE

• Définition d’un Datawarehouse • Valeur ajoutée de Hyve pour l’exposition des données • Atelier pratique : mise à disposition des données structurées avec Hyve • Visualisation de données structurées

ARCHITECTURE LAMBDA

• Qu’est-ce qu’une architecture Lambda • Implémentation d’une architecture Lambda dans un Datalake Atelier pratique : implémentation d’une Architecture Lambda

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.

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.

Organisme de formation enregistré sous le numéro 11 75 54743 75. Cet enregistrement ne vaut pas agrément de l'État.
© Ambient IT 2015-2023. Tous droits réservés. Paris, France - Suisse - Belgique - Luxembourg