

Hadoop : Analytics

Référence : **PYCB036**

Durée : **2 jours (14 heures)**

Certification : **Aucune**

Connaissances préalables

- Connaissances des principes du BigData, d'un langage de programmation comme Java ou Scala ou Python

Profil des stagiaires

- Chefs de projet, développeurs, data scientists, architectes souhaitant mettre en œuvre des solutions analytics avec hadoop

Objectifs

- Savoir mettre en œuvre les frameworks analytics dans un environnement hadoop

Certification préparée

- Aucune

Méthodes pédagogiques

- Mise à disposition d'un poste de travail par participant
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques et de réflexions

Formateur·rice

- Consultant-Formateur expert Bigdata

Méthodes d'évaluation des acquis

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

Contenu du cours

1. Introduction

- Définitions : Analytics
- Arbres de décision, de régression, régression automatique
- Apprentissage supervisé, apprentissage automatique
- Présentation du data munging

2. Hadoop et les outils d'analyse

- Rôle des différents composants : socle hadoop, yarn, hdfs
- Frameworks analytics : Mahout, Flink, Spark ML

3. Mahout

- Principe de fonctionnement
- Sources de données, format de stockage des données
- Génération de recommandations, traitement, filtrage
- Exemples de base : génération de recommandations, traitement, filtrage
- Présentation des algorithmes les plus courants
- Compatibilité avec Hadoop Yarn, Spark, H2O, Flink

4. Flink

- Origine du projet, fonctionnalités
- Traitement distribué de flux de données, en temps réel ou batch
- APIs disponibles
- Mise en oeuvre avec des programmes Java/Scala
- Analyse de graphe avec l'API Gelly

5. Spark MLib

- Fonctionnalités : Machine Learning avec Spark, algorithmes standards, gestion de la persistance, statistiques
- Support de RDD
- Mise en oeuvre avec les DataFrames

6. GraphX

- Fourniture d'algorithmes, d'opérateurs simples pour des calculs statistiques sur les graphes
- Travaux pratiques : exemples d'opérations sur les graphes

Notre référent handicap se tient à votre disposition au [01.71.19.70.30](tel:01.71.19.70.30) ou par mail à referent.handicap@edugroupe.com pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.