

Hadoop : l'écosystème

Référence : **PYCB030**

Durée : **1 jour (7 heures)**

Certification : **Aucune**

Connaissances préalables

- Connaissances générales des systèmes d'information

Profil des stagiaires

- Chefs de projets, développeurs, et toute personne souhaitant comprendre les mécanismes Hadoop et le rôle de chaque composant

Objectifs

- Faire le point sur les différents éléments de l'écosystème Hadoop et leurs rôles respectifs

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

- 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

- Rappels sur NoSQL. Le théorème CAP
- Historique du projet hadoop
- Les fonctionnalités : stockage, outils d'extraction, de conversion, ETL, analyse, ...
- Exemples de cas d'utilisation sur des grands projets
- Les principaux composants : HDFS pour le stockage et YARN pour les calculs
- Les distributions et leurs caractéristiques (HortonWorks, Cloudera, MapR, GreenPlum, Apache, ...)

2. L'architecture

- Terminologie : NameNode, DataNode, ResourceManager
- Rôle et interactions des différents composants
- Présentation des outils : d'infrastructure (ambari, avro, zookeeper) - de gestion des données (pig, oozie, falcon, pentaho, sqoop, flume) - d'interfaçage avec les applications GIS - de restitution et requêtage : webhdfs, hive, hawq, impala, drill, stinger, tajo, mahout, lucene, elasticSearch, Kibana
- Les architectures connexes : spark, cassandra

3. Exemples interactifs

- Démonstrations sur une architecture Hadoop multi-noeuds.
- Mise à disposition d'un environnement pour des exemples de calcul
- Travaux pratiques : Recherches dans des données complexes non structurées

4. Applications

- Cas d'usages de Hadoop
- Les infrastructures hyperconvergées avec les appliances Hadoop
- Calculs distribués sur des clusters Hadoop

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.