

Hadoop : l'écosystème

Référence : PYCB030

Durée : 1 jour

Certification : **Aucune**

CONNAISSANCES PREALABLES

- 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 PREPAREE

Aucune

METHODES PEDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement

FORMATEUR

Consultant-Formateur expert Bigdata

METHODE D'EVALUATION DES ACQUIS

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

CONTENU DU COURS

Introduction

- Rappels sur NoSQL. Le théorème CAP
- Historique du projet hadoop
- Les fonctionnalités : stockage, outils '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, ...)

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

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

Applications

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

