

Hadoop l'écosystème

Détails

- **Code** : DB-HDE
- **Durée** : 1 jour (7 heures)

Public

- Chefs de projets
- Développeurs
- Toute personne souhaitant comprendre les mécanismes Hadoop et le rôle de chaque composant

Pré-requis

- Connaissances générales des systèmes d'information

Objectifs

- Faire le point sur les différents éléments de l'écosystème Hadoop et leurs rôles respectifs
- Comprendre l'architecture des applicatifs hadoop et savoir quels sont les apports et les cas d'usage des solutions hadoop

Programme

Introduction

- Rappels sur NoSQL
- Le théorème CAP
- Historique du projet hadoop
- 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, EMR, Apache, ...)

- Présentation des outils d'interfaçage avec les applications GIS
- Présentation des outils de restitution et requête : webhdfs, hive, hawq, impala, drill, stinger, tajo, mahout, lucene, elasticSearch, kibana
- Les architectures connexes : spark, cassandra

Exemples interactifs

- Démonstrations sur une architecture Hadoop multi-nœuds
- Mise à disposition d'un environnement pour des exemples de calcul

L'architecture

- Terminologie : NameNode, DataNode, ResourceManager
- Rôle et interactions des différents composants
- Présentation des outils d'infrastructure : ambari, avro, zookeeper
- Présentation des outils de gestion des données : pig, oozie, tez, falcon, sqoop, flume

Travaux pratiques

- Intégration de données avec Flume
- Calculs map/reduce avec Pig
- Hive
- Impala

Applications

- Cas d'usages de hadoop
- Calculs distribués sur des clusters hadoop

Modalités

- **Type d'action** :Acquisition des connaissances
- **Moyens de la formation** :Formation présentielle – 1 poste par stagiaire – 1 vidéo projecteur – Support de cours fourni à chaque stagiaire
- **Modalités pédagogiques** :Exposés – Cas pratiques – Synthèse
- **Validation** :Exercices de validation – Attestation de stages