

# **HADOOP : L'ÉCOSYSTÈME**

## **CODE STAGE : BD018**

### **OBJECTIFS**

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

### **DURÉE**

1 jour

### **PUBLIC**

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

### **PRÉ-REQUIS**

Connaissances générales des systèmes d'information.

### **PROGRAMME**

#### **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, tez, falcon, pentaho, sqoop, flume;

d'interface 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-nœuds.

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.

Calculs distribués sur des clusters hadoop

