

## Hadoop : l'écosystème

**Durée:** 1 jour

650 €

20 février

15 mai

4 septembre

13 novembre

### **Public:**

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. Comprendre l'architecture des applicatifs hadoop et savoir quels sont les apports et les cas d'usage des solutions hadoop.

### **Connaissances préalables nécessaires:**

Connaissances générales des systèmes d'information.

### **Programme:**

- 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, tez, 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.

## **Hadoop : l'écosystème**

Applications : Cas d'usages de hadoop.  
Les infrastructures hyperconvergées avec les appliances hadoop  
Calculs distribués sur des clusters hadoop