

## **Hadoop : développement avec MapReduce**

**Durée:** 2 jours

1060 € HT

5 au 6 février

12 au 13 juin

8 au 9 octobre

10 au 11 décembre

### **Public:**

Chefs de projets, développeurs, et toute personne souhaitant comprendre les techniques de développement avec MapReduce dans l'environnement Hadoop.

### **Objectifs:**

Connaître les principes du framework Hadoop et savoir utiliser la technologie MapReduce pour paralléliser des calculs sur des volumes importants de données.

### **Connaissances préalables nécessaires:**

Connaissance d'un langage de programmation objet comme Java.

### **Programme:**

- Introduction** : Les fonctionnalités du framework Hadoop  
Le projet et les modules :  
Hadoop Common, HDFS, YARN, MapReduce
- MapReduce** : Principe et objectifs du modèle de programmation MapReduce.  
Fonctions map() et reduce().  
Couples (clés, valeurs).  
Implémentation par le framework Hadoop.  
Etude de la collection d'exemples.  
Rédaction d'un premier programme et exécution avec Hadoop.
- Programmation** : Les interfaces principales : mapper, reducer, partitioner, reporter, outputcollector.  
Format des entrées et sorties d'un job MapReduce : InputFormat et OutputFormat.  
Configuration des jobs avec jobconf.  
Interaction avec le JobTracker par JobClient.
- Optimisation** : Gestion de la mémoire, des accès disque.  
Utilisation de la JVM.  
Mise en oeuvre du cache distribué.
- Securité** : Configuration des ACLs.  
Gestion de l'authentification.