

## Apache Mahout : analyse de données

**Durée:** 2 jours

1 180 €

16 au 17 mars

22 au 23 juin

12 au 13 octobre

14 au 15 décembre

### **Public:**

Chefs de projet, architectes, développeurs, data-scientists souhaitant utiliser Mahout et mettre en oeuvre des algorithmes d'apprentissage distribués.

### **Objectifs:**

Comprendre le fonctionnement de Mahout, connaître l'architecture et savoir implémenter les algorithmes de Machine Learning en local ou en environnement distribué.

### **Connaissances préalables nécessaires:**

Connaissance d'un langage de développement objet et connaissances générales sur le Bigdata, le data-mining, l'analyse de données.

### **Programme:**

- Introduction** : Présentation Mahout.  
Origine du projet, licence, positionnement dans l'offre BigData et Machine Learning : Hadoop, Spark,..  
Fonctionnalités.  
Définitions : apprentissage supervisé, apprentissage automatique  
Arbres de décision, de régression, régression automatique  
Classifieurs. Scoring
- Architecture** : Principe de fonctionnement.  
Sources de données, format de stockage des données,  
Génération de recommandations, traitement, filtrage  
Mode local ou distribué.
- Mise en oeuvre** : Installation en mode autonome .  
Exemples de base : génération de recommandations, traitement, filtrage  
Présentation des algorithmes les plus courants.  
Compatibilité avec Hadoop Yarn, Spark, H2O, Flink  
Installation en mode distribué sur une ferme Spark.  
Premiers pas avec le shell interactif REPL  
Exemple avec une classification bayésienne naïve