

Apache Mahout

Analyse des données avec Apache Mahout

Détails

- Code : DB-MAH
- Durée : 2 jours (14 heures)

Public

- Architectes
- Chefs de projets
- Data Scientist
- Développeurs

Pré-requis

Objectifs

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

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 œuvre

- 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

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