

## Pig : développement de tâches MapReduce

Durée: 2 jours

1320 €

### Public:

Chefs de projet, data scientists, développeurs souhaitant utiliser pig pour l'analyse de données

### Objectifs:

Savoir utiliser Pig pour développer des requêtes MapReduce

### Connaissances préalables nécessaires:

Connaissance de Java ou Python, des bases Hadoop, et notions de calculs statistiques

### Programme:

- Introduction : Le projet Apache Pig, fonctionnalités, versions  
Présentation de Pig Latin.
- Mise en oeuvre : Prérequis techniques, configuration de Pig  
Travaux pratiques:  
Exécution : les différents modes : interactif ou batch Scala
- Grunt : Principe de l'exécution de scripts Pig Latin avec Grunt  
Commandes HDFS  
Modèles de données avec Pig
- Pig Latin : Les requêtes Latin : chargement de données, instructions  
Les entrées/sorties, fonctions  
Intégration Pig avec MapReduce  
Contrôle d'exécution  
Présentation des outils de développement
- Extensions : Utilisation de Pig Latin depuis des programmes Python  
Fonctions d'évaluation et de filtres
- Liens avec l'écosystème Hadoop : Pig et Hive  
Les bases de données NoSQL  
Les métadonnées et Hadoop