

## HADOOP - CLUDERA POUR DÉVELOPPEURS

3 jours en présentiel (21 heures)

### Objectifs pédagogiques

Cette formation Hadoop - Cloudera pour développeurs permettra aux programmeurs Java une étude en profondeur dans les développements des applications Hadoop 2.0. Les participants apprendront à définir et à développer des applications MapReduce(Yarn), Hive et Pig efficaces et pertinentes avec Hadoop 2.0. Les participants à ce cours apprendront à exploiter toute la puissance d'Hadoop 2.0 pour manipulation, analyse et exécution des calculs sur leurs de cluster Hadoop.

### Population visée

Cette formation Hadoop - Cloudera pour développeurs est destinée aux développeurs qui développeront, maintiendront et optimiseront des applications pour Hadoop.

### Pré-requis

Cette formation Hadoop - Cloudera pour développeurs nécessite une expérience dans le développement, si possible avec Java. Une connaissance préliminaire d'Hadoop n'est pas exigée.

### Méthodes pédagogiques

1 poste et 1 support par stagiaire 8 à 10 stagiaires par salle Remise d'une documentation pédagogique papier ou numérique pendant le stage La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

### Formateur

Formateur consultant expert en BIG DATA

### Modalités de validation des acquis

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne Attestation de fin de stage remise au stagiaire

### Contenu

#### INTRODUCTION

- Présentation générale d'Hadoop
- Exemples d'utilisation dans différents secteurs
- Historique et chiffres clés : Quand parle-t-on de Big Data ?

#### HDFS : INTERFACAGE AVEC JAVA

- Le système de fichier Hadoop Distributed File System (HDFS)
- Lecture et écriture sur le HDFS
- Manipulation des types de fichier SequenceFile
- Qu'est-ce que le DistributedCache et comment partager les références ?

#### MAPREDUCE/YARN : PRINCIPE ET DEVELOPPEMENT JAVA

- Le paradigme MapReduce et l'utilisation à travers YARN
- Implémentation de problématiques réelles avec le paradigme MapReduce
- Implémentation de problématiques plus complexes
- Utilisation des combinateurs et partitionners
- Les bonnes pratiques pour le développement des applications MapReduce

#### MANIPULATION DES DONNEES AVEC UN QUERY LANGUAGE ET JAVA

- Présentation et fonctionnement de HBase, Hive et PIG
- HBase et Hive : JDBC et requêtes
- Extension du HiveQL avec des fonctionnalités développées : UDF (User-Defined Functions)
- Pig : Développement et intégration avec JAVA
- Extension de PIG avec des fonctionnalités développées : UDF (User-Defined Functions)