

# Formation Hadoop

## Informations

Durée : 4 jours (28h.)

Tarif\* : Nous consulter

Réf : HADO

Niveau : Difficile

intra

\*tarif valable jusqu'au 31/12/2024

## Prochaines sessions

Contactez-nous pour connaître nos futures sessions.

## Pré-requis

- Avoir des connaissances des principes de base Hadoop et des bases de données

## Objectifs

- Concevoir, exécuter et tester des programmes écrits avec Map/Reduce
- Entrer et sortir des données de formats variés pour les traiter avec Hadoop
- Utiliser Hive pour pouvoir interroger le système de fichiers HDFS avec un langage analogue à SQL
- Utiliser Pig pour produire facilement des programmes Map-Reduce en langage de haut niveau

## Programme

### Introduction

Problème des systèmes traditionnels à grande échelle

Qu'est-ce qu'Hadoop ?

Quels problèmes peut-on résoudre avec Hadoop ?

### Les concepts fondamentaux et HDFS

Le projet Hadoop et ses composants  
HDFS, le système de fichiers distribué

### MapReduce

L'utilisation de MapReduce  
L'analyse de données avec les outils Unix  
L'analyse de données avec Hadoop  
Mappers  
Reducers  
Combiners

### Clusters Hadoop et écosystème

Cluster Hadoop : concepts  
Jobs et tasks  
Systèmes de fichiers  
Programmation distribuée : MapReduce, Pig et Spark  
Bases NoSQL : HBase et Cassandra  
Accès SQL à Hadoop : Hive  
Ingestion de données : Flume, Kafka et Sqoop  
Planification des workflows Hadoop : Oozie  
Machine Learning : Mahout et Weka

### HDFS

Motivations et design  
Blocs et nœuds  
Interface en ligne de commande  
Interface Java  
Flux de données  
HBase

# Formation Hadoop

## Mise en place de clusters Hadoop

- Spécification du cluster
- Configuration et Installation
- Configuration d'Hadoop
- Configuration d'HDFS
- Monitoring et logging
- Maintenance

## Entrer et sortir des données d'Hadoop

- ingress et egress : éléments-clés
- Entrer des données de log avec Apache Flume
- Programmation des entrées de données avec Oozie
- Importer/Exporter des données depuis des SGBDR avec Sqoop
- MapReduce et XML
- MapReduce et JSON
- MapReduce et formats personnalisés

## L'API Hadoop pour Java

## Tests unitaires avec Hadoop

- Pertinence des tests unitaires
- Tester les mappers et reducers : JUnit et MRUnit
- Execution des tests
- LocalJobRunner

## Pig

- Faciliter l'écriture de programmes MapReduce avec Pig
- L'installation et l'exécution
- Le langage de script : Pig Latin
- Les fonctions Utilisateurs (UDF)
- Les opérateurs de traitement de données

## Hive

- Interroger et gérer de larges volumes de données avec Hive
- L'installation
- L'exécution
- La comparaison avec les bases de données traditionnelles
- HiveQL
- Tables
- L'interrogation des données
- La fonction utilisateurs

## Réalisation d'une application complète avec Hadoop, Pig et Hive