

# **HADOOP, DÉVELOPPER DES APPLICATIONS POUR LE BIG DATA**

**CODE STAGE : DB014**

## **OBJECTIFS**

Construire un programme à base de Map Reduce

Intégrer Hadoop HBase dans un workflow d'entreprise

Travailler avec Apache Hive et Pig depuis HDFS

Utiliser un graphe de tâches avec Hadoop

## **DURÉE**

4 jours

## **PUBLIC**

Concepteurs, développeurs.

## **PRÉ-REQUIS**

Bonne expérience en développement Java. Des connaissances en architecture Web constituent un plus.

## **PROGRAMME**

Le Big Data

Définition du périmètre du Big Data.

Le rôle du projet Hadoop.

Les concepts de base des projets Big Data.

Présentation du Cloud Computing.

Différence entre Cloud Computing privé et public.

Les architectures Big Data à base du projet Hadoop.

Démonstration

Usage d'Hadoop et de GoogleApp.

Collecte de données et application de Map Reduce

Analyse des flux de données dans l'entreprise.

Données structurées et non-structurées.



Les principes de l'analyse sémantique des données d'entreprise.

Graphes des tâches à base de MapReduce.

La granularité de cohérence des données.

Transfert de données depuis un système de persistance dans Hadoop.

Transfert de données d'un Cloud dans Hadoop.

Travaux pratiques

Gérer la collecte d'informations clientèles par Map Reduce. Configuration de

l'implémentation YARN. Développement d'une tâche basée sur Map Reduce.

Le stockage des données avec HBase

Plusieurs types de bases de données XML.

Patterns d'usages et application au Cloud.

Application de Hadoop database au sein d'un workflow.

Utilisation des projets Hive/Pig.

Utilisation du projet HCatalog.

L'API Java HBase.

Travaux pratiques

Gérer les modifications d'un catalogue de données fournisseur.

Le stockage des données sur HDFS

Patterns d'usages et application au Cloud.

Architecture et installation d'un système HDFS, journal, NameNode, DataNode.

Opérations, commandes et gestion des commandes.

L'API HDFS Java.

Analyse de données avec Apache Pig.

Le langage Pig Latin. Utiliser Apache Pig avec Java.

Requêtage avec Apache Hive.

Réplication de données. Partage de données sur une architecture HDFS.

Travaux pratiques

Administrer un référentiel client partagé sur Hadoop. Utilisation de la console de visualisation.

Spring Data Hadoop

Introduction à Spring et Spring Data.

Le namespace Hadoop pour Spring.

Utiliser Spring pour simplifier la configuration Hadoop.



Configuration du cache distribué.

Définition des Jobs et dépendance entre Jobs.

Intégration des outils (Pig, Hivecode2utf('8230',0)).

