

Programme de formation Apache Hadoop

● Objectifs

Cette formation Apache Hadoop permettra aux participants d'acquérir des compétences pratiques dans la mise en oeuvre d'un système Hadoop. Les participants seront à même d'interroger les données contenues sur cette plateforme.

● Pré requis

Connaissances en SQL.

● Durée

2 jours

● Public

Developpeurs, DSI,
Utilisateurs

● Plan de formation

Introduction

Les origines du Big Data
La donnée en tant que matière première

Les composants Hadoop

-

Écosystème Hadoop

Architecture HDFS
Architecture et principes de fonctionnement de
MapReduce
HBase

-

Hive

Pig
Sqoop

-

Présentation des distributions principales du marché

Hortonworks
MapR
Cloudera

Mise en place d'une infrastructure Hadoop

-

Organisation d'un cluster

Besoin matériel, réseau
Choisir le matériel adapté

-

Installation Hadoop et configuration initiale

Pré-requis

Cloudera Manager

-

Surveillez l'état de votre cluster

Développement Hadoop

-

Hive

Création de bases de données, tables, vues
Requêtes grâce au HiveSQL
Manipulations de fonctions définies par
l'utilisateur (UDF)
Gestion des droits utilisateurs

-

Pig

Pig Latin pour interroger vos données
Communication entre Pig et Hive

-

Oozie

Comment planifier ses travaux
Utilisation de workflows, coordinateurs,
bundles

Optimiser l'exécution de vos travaux

Partitionner vos données à l'aide d'Hive
Archivez vos données
Gestion de la mémoire (JVM)