

• **DURÉE**

3 jours / 21h

APPRÉCIATION

Évaluation qualitative de fin de stage

**MODALITÉS
ET MOYENS**

PÉDAGOGIQUES

Démonstrations, cas pratiques, synthèse et évaluation des acquis

► **Objectifs
pédagogiques**

- Comprendre ce que sont Hadoop et YARN
- Pouvoir manipuler les données sous Hadoop
- Savoir manipuler les données avec Spark
- Savoir analyser les données avec Hive

► **Public concerné**

Toute personne souhaitant manipuler et analyser des données dans un cluster Hadoop

► **Prérequis**

Connaissances systèmes d'information, connaissances bases de données, concepts de programmation

Big Data DATA ANALYSE PIG, HIVE & SPARK

2 030 € HT

Inter-Entreprises
Intra-entreprises sur devis

Le but de cette formation est de présenter des outils permettant de traiter et d'analyser des données sur Hadoop. Elle vous permettra de développer des compétences en analyse de données en se focalisant sur Pig, Hive et Spark, que vous serez en mesure d'exploiter.

Introduction

- Introduction au Big Data - Comprendre les concepts clés et les enjeux du Big Data
- Introduction à Hadoop - Principales distributions de Hadoop

La plateforme Hadoop

- Architecture et composants de la plateforme Hadoop
- HDFS
- NameNode / DataNode / ResourceManager
- Paradigme MapReduce et YARN
- Les technologies émergentes

Traitement des données avec Pig

- Description et caractéristiques de Pig
- Présentation de Pig
- Différence entre Pig et MapReduce
- Cas d'utilisation de Pig
- Traitement des données
- Modélisation des données
- Programmation avec Pig Latin
- Transformations dans la syntaxe Pig Latin
- Fonctions de chargement et de stockage
- Travaux pratiques

Requêtage des données avec Hive

- Description et caractéristiques de Hive
- Utilisation de HCatalog
- Analyse des données avec Hive
- Management des données Hive
- Formats de données Hive
- Création des bases de données et des tableaux de management
- Tableaux auto-managés
- Simplification des requêtes avec les vues
- Stockage des résultats de requêtes
- Contrôle d'accès aux données
- Traitement de texte avec Hive
- Fonctions String
- Utilisation des expressions habituelles dans Hive

Apache Spark SQL

- Présentation générale
- Caractéristiques – Architecture
- Les bases de Spark
- DataFrames et DataSets
- Les RDD
- Le SQL Context
- Opérations sur les DataFrames et les DataSets
- Comparaison entre Spark SQL et Hive

