Les Fondamentaux d'Hadoop

BIGDATA03

Hadoop étant la principale plateforme du Big Data sur le marché français la formation permet d'appréhender l'utilisation de l'outil pour le stockage et le traitement d'immense volume de données. Elle abordera également l'utilité de ses différents composants.

A distance - Synchrone

Objectifs

Ce module présente les grands outils de l'eìcosysteème Hadoop en se focalisant plus speìcifiquement sur HDFS, Map Reduce et Hive.
Il permet de découvrir également l'ingestion de données avec Nifi, les différents formats de fichier et la compression.

Le principal objectif est le deìveloppement de compeìtences de data engineer orienteìes acceès et traitement des donneìes

Pré Requis

Aucun - bases informatiques générales

Objectifs pédagogiques et d'évaluation

Ce module présente les grands outils de l'écosystème Hadoop en se focalisant plus spécifiquement sur HDFS, Map Reduce et Hive
Il permet de découvrir également l'ingestion de données avec Nifi, les différents formats de fichier et la compression
Le principal objectif est le développement de compétences de data engineer orientées accès et traitement des données

Méthodes pédagogiques

- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques et de réflexions

Parcours pédagogique

MAITRISER HADOOP

Charger les données dans HDFS
Comprendre le fonctionnement de Map Reduce
Maîtriser les composants de base d'Hadoop
Exercice: Hadoop shell

INGESTION DE DONNÉES

Présentation des différents outils d'ingestion de données
Exercice: Développement d'un flux sous Nifi
Exercice: Génération d'un fichier au format Parquet

Programmation Hive

Types et mots-clés dans Hive
Concept de table et base de données dans Hive
Principales fonctions de Hive
Jointures et vues
Application de cet outil dans l'analyse de données
Exercice: Exploitation d'un fichier de logs
Exercice: Analyse de données de la grande distribution

Format et compression

Présentation des différents formats de fichier

Ecole de la Data & de l'IA - Numéro de déclaration d'activité (ne vaut pas agrément de l'état) : 11 75 39027 75

Ecole de la Data & de l'IA
Cœur Défense A – 92931 Paris La Défense Cedex
Tél. : 01.58.35.19.37 01.58.35.19.38 – Email : bd.university@businessdecision.com
S.A.S. au capital de 237 725 € - B 381 837 764 R.C.S. Nanterre - NAF 6202A - N° de déclaration : 11 75 39027 75

Compression et codecs
Exercice: Exploitation des fichiers parquet sous Hive
Format et compression
Pig, Oozie, Spark, Solr
Présentation Cloudera, Cloudera Navigator
Présentation Hortonworks, Ranger

Méthodes et modalités d'évaluation

Une note globale égale à 70% de la note obtenue au QCM + 30% de la note obtenue au projet pédagogique supérieure à 10/20
Accomplissement des exercices pratiques de chacun des modules

Durée

14.00 Heures **2** Jours

Ecole de la Data & de l'IA - Numéro de déclaration d'activité (ne vaut pas agrément de l'état) : 11 75 39027 75

Ecole de la Data & de l'IA
Cœur Défense A – 92931 Paris La Défense Cedex
Tél. : 01.58.35.19.37 01.58.35.19.38 – Email : bd.university@businessdecision.com
S.A.S. au capital de 237 725 € - B 381 837 764 R.C.S. Nanterre - NAF 6202A - N° de déclaration : 11 75 39027 75