

Plan de formation Big Data

Formateur : Mohamed MANAA

Ce programme est susceptible de changer en fonction de l'avancement de la formation et du temps nécessaire pour chaque session

Jour 1 :

- 1- Introduction au Big Data
 - 1.1. Caractéristiques du Big Data (5 Vs)
 - 1.2. Cas d'utilisation du Big Data
 - 1.3. Big Data Vs Data Science Vs BI
- 2- Apache Hadoop
 - 2.1. Introduction à Hadoop
 - 2.2. Architecture Hadoop
 - 2.3. Avantages et limites de Hadoop
- 3- Atelier 1 (Installation et configuration d'un cluster Hadoop)

Jour 2 :

- 1- Hadoop Distributed File System (HDFS)
 - 1.1. Architecture HDFS
 - 1.2. Réplication
 - 1.3. Haute Disponibilité
- 2- MapReduce
 - 3.1. Architecture de MapReduce
 - 3.2. Les Tâches MapReduce
 - 3.3. Planification des Jobs
 - 3.4. Exemple : WordCount
- 3- Atelier 2 (Exécution d'un Job MapReduce : WordCount)

Jour 3 :

- 1- Hadoop Query Languages (HQL)
 - 1.1. Pig
 - 1.2. Hive
 - 1.3. JAQL
- 2- Bases de données NoSQL
 - 2.1. Modèle orienté Clé/Valeur
 - 2.2. Modèle orienté Colonnes
 - 2.3. Modèle orienté Document
 - 2.4. Modèle orienté Graphe
- 3- HBase
 - 3.1. Architecture de HBase
 - 3.2. Fonctionnement
- 4- Ecosystème de Hadoop
 - 4.1. Sqoop
 - 4.2. Flume
 - 4.3. Zookeeper
 - 4.4. Kafka
- 5- Atelier 3 (Bases de données NoSQL)

Jour 4 :

- 1- YARN
 - 1.1. Architecture de YARN
 - 1.2. Fonctionnement
- 2- Apache SPARK
 - 2.1. SPARK Vs MapReduce
 - 2.2. Architecture de SPARK
 - 2.3. RDD
- 3- Atelier 4 (SPARK)