

Introduction au Big Data

Volume horaire : 20h

Objectifs pédagogiques

- Comprendre et exploiter des ensembles de données volumineux
- Comprendre et être capable de résoudre un problème à l'aide de MapReduce
- Se familiariser avec les outils du BigData (Hadoop, Hive, Pig, etc.)

Programme

1. Introduction aux concepts et applications des solutions Big Data
2. Architecture Hadoop et HDFS (Hadoop Distributed File System)
TP : Installation d'un cluster Hadoop sous Linux
3. Administrer un Cluster Hadoop
TP : Configuration du cluster Hadoop
4. Principes d'exécution et de requêtage : Map/Reduce
TP : Insertion de données dans le cluster & requêtes
5. Introduction au langage Hive
TP : Insertion de données & requête Hive
6. Introduction au langage PIG
TP : requête PIG

Moyens et outils pédagogique

Cours théorique & TP de mise en situation à l'aide d'exemples types

Évaluation

Devoir sur table & Épreuve pratiques

Bibliographie

- *Big Data: High-Impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors* - **Roebuck**
- *Le Big Data* – **Delort**
- *La boîte à outils de la stratégie big data* - **Jouin, Rissoan**