



Hadoop & Python (Data Sciences)

Plan

0. Introduction

- Présentation de la formation
- Positionnement d'Hadoop & Python
- Installation Docker Desktop
- Installation des images Hadoop Diginamic

1. Hadoop, Découvertes Big Data

- Introduction de son écosystème
- Parcourir son écosystème
- Différencier ses distributions
- TP 01 : Utilisation d'Hadoop de base
- TP 02 : Voir les nœuds (Slave 1 et Slave 2)

2. Stockage HDFS

- Découverte HDFS
- Comprendre HDFS (Principe & les Services)
- Configurer et Administrer HDFS
- Stockage Fichiers HDFS
- TP 03 : Utilisation HDFS (utilisation des commandes)

3. MapReduce

- Présentation du MapReduce & Hadoop
- Fonctionnement de MapReduce
- Développer un MapReduce en Python
- Suivi d'un MapReduce
- TP 05 : MapReduce avec Hadoop & Python

4. HBASE

- Découverte & Comprendre le modèle de HBASE
- Administrer HBASE
- Concevoir une base de données dans HBASE
- Filtrer et Lire des données dans HBASE
- Utilisation dans le terminal HBASE
- TP 04 : Utilisation HBASE

5. Sqoop

- Présentation de Sqoop
- Collecter, importer et exporter avec Sqoop
- TP 07 : Utilisation de Sqoop

6. Oozie

- Présentation de l'ordonnanceur Oozie
- Mise en place d'un Workflow Oozie
- Lancer une Workflow Oozie
- TP 06 : Utilisation de Oozie