

Objectifs / Durée de la formation

Durée: 2 jours, soit 14 heures

- Découvrir les concepts et les enjeux liés à Hadoop
- Comprendre le fonctionnement de la plateforme et de ses composants

Participants / Pré-requis

- Administrateurs de cluster Hadoop, développeurs, data scientist
- Bonnes connaissances de l'administration Linux

Moyens pédagogiques

- Formation avec un formateur, dans une salle dédiée à la formation, équipée d'un vidéoprojecteur, d'un tableau blanc ou à distance dans le cas de solutions de Digital Learning
- 1 poste de travail par stagiaire adapté aux besoins de la formation + 1 support de cours et/ou un manuel de référence au format numérique ou papier
- Modalités pédagogiques : Exposés - Cas pratiques - Synthèse

Programme

1. Introduction à Hadoop

- Présentation générale d'Hadoop
- Exemples d'utilisation dans différents secteurs
- Historique et chiffres clés : Quand parle-t-on de Big Data ?

2. L'écosystème d'Hadoop

- Le système de fichier HDFS
- Le paradigme MapReduce et l'utilisation à travers YARN

3. Manipulation des données dans un cluster Hadoop

- Hue : Comment fonctionne cette interface web ?
- Hive : Pourquoi Hive n'est pas une base de données ?

4. Requête sur Hive

- Utilisation de HCatalog
- Utilisation avancée sur Hive
- Utilisation de fonctions utilisateurs
- Paramétrage de requête
- Pig : Fonctionnement de Pig

5. Programmation avec Pig Latin

- Utilisation du mode Local
- Utilisation de fonctions utilisateurs
- Tez : Qu'est-ce que Tez ?

6. Comment et quand l'utiliser ?

- Oozie : Fonctionnement de Oozie

7. Création de Workflows avec Oozie

- Manipulation des workflows
- Ajout d'éléments d'exploitation dans les workflows
- Ajout de conditions d'exécution
- Paramétrage des workflows
- Sqoop : A quoi sert Sqoop ?

8. Chargement des données depuis une base de données relationnelle

- Chargement des données depuis Hadoop
- Utilisation et paramétrage avancée
- Les particularités des distributions : Impala, Hawq
- Quelles sont les bonnes pratiques d'utilisation des différents outils ?