

Délai maximum : 2 mois.

Parcours concourant au développement des compétences. Action de formation réalisée en application des articles L 6313-1 et L 6313-2 du Code du travail.

Si vous êtes en situation de handicap, contactez-nous avant le début de votre formation pour que nous puissions vous orienter efficacement et vous accueillir dans les meilleures conditions.



Objectifs pédagogiques

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



Pré-requis

- Bonnes connaissances de l'administration Linux



Modalités pédagogiques

Modalités de formation:

- Formation réalisée en présentiel, à distance ou mixte,
- Toutes nos formations peuvent être organisées dans nos locaux ou sur site
- Feuille de présence signée en demi-journée, questionnaires d'évaluation de la satisfaction en fin de stage et 60 jours après, attestation de stage et certificat de réalisation.
- Horaires de la formation: 9h - 12h30 et 13h30 - 17h.
- Les horaires de la formation sont adaptables sur demande.



Moyens pédagogiques

- Formateur expert dans le domaine,
- Mise à disposition d'un ordinateur, d'un support de cours remis à chaque participant,
- Vidéo projecteur, tableau blanc et paperboard,
- Formation basée sur une alternance d'apports théoriques et de mises en pratique
- Formation à distance à l'aide du logiciel Teams pour assurer les interactions avec le formateur et les autres stagiaires, accès aux supports et aux évaluations. Assistance pédagogique afin de permettre à l'apprenant de s'approprier son parcours. Assistance technique pour la prise en main des équipements et la résolution des problèmes de connexion ou d'accès. Méthodes pédagogiques : méthode expositive 50%, méthode active 50%.

Public visé

- Administrateurs de cluster Hadoop, développeurs, data scientist

Modalités d'évaluation et de suivi

- Evaluation des acquis tout au long de la formation : QCM, mises en situation, TP, évaluations orales...



Programme de formation

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

Contacts



Notre centre à **Mérignac**

14 rue Euler
33700 MERIGNAC

☎ 05 57 92 22 00

✉ contact@afib.fr



Notre centre à **Périgueux**

371 Boulevard des Saveurs,
24660 COULOUNIEIX CHAMIERES

☎ 05 64 31 02 15

✉ contact@afib.fr

- 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 ?