

Administrer la plateforme Hadoop 2.X Hortonworks (niveau 1)

Code formation : EMPHH

Durée : 4 jours – 28h de cours

Format : Inter-entreprise*

4 jours

28 heures de cours

*Cette formation est également disponible en « Intra-entreprise », nous contacter pour plus d'infos.

| Description

Cette session prépare au rôle d'administrateur au sein d'un contexte technologique innovant et en particulier au cours d'un projet Big Data. A travers des exercices concrets, vous apprendrez à installer, configurer et maintenir un cluster Hadoop.

A la fin de cette formation, vous aurez une compréhension solide de comment Hadoop fonctionne avec le Big Data et, à travers nos mises en pratique, vous saurez déployer tout le cycle de vie pour des clusters multi-nœuds.

| Objectifs pédagogiques

- Dimensionner un cluster Hadoop
- Installer un cluster Hadoop
- Configurer un cluster Hadoop
- Sécuriser un cluster Hadoop
- Maintenir un cluster Hadoop

| Publics

Architecte, administrateur

| Pré-requis

Connaissances de l'environnement Linux.

| Méthode pédagogique

Formation avec apports théoriques, échanges sur les contextes des participants et retours d'expérience pratique du formateur, complétés de travaux pratiques et de mises en situation. Cette formation prépare à la certification éditeur Hortonworks.

| Programme détaillé

Jour 1

- **Big Data, Hadoop et la plateforme Hortonworks : les bases du Big Data**
 - | Les produits de la HDP
 - | Qu'est-ce que Hadoop ?
 - | Une architecture de cluster type
 - | Introduction à Ambari
- **Gestion des utilisateurs avec Ambari**
 - | Gérer les users et les groupes
 - | Gérer les permissions
 - | Mise en pratique : « Gestion des utilisateurs avec Ambari »
- **Gestion des services Hadoop via Ambari**
 - | Configuration des services
 - | Surveillance des services
 - | Maintenance des services
 - | Mise en pratique : « Gestion des services Hadoop »
- **Utiliser le stockage HDFS**
 - | Accéder aux données
 - | Gestion des fichiers
 - | Mise en pratique : « Utiliser le stockage HDFS »

Jour 2

- **Utiliser le stockage HDFS (suite)**
 - | Les web services d'HDFS
 - | Mise en pratique : « Utiliser WebHDFS »
 - | Protéger les accès
 - | Mise en pratique : « Utiliser les ACLs HDFS »
- **Gestion du stockage HDFS**
 - | Architecture HDFS
 - | Assurer l'intégrité de la donnée
 - | Mise en pratique : « Gestion du stockage sur HDFS »
 - | Les quotas HDFS
 - | Mise en pratique : « Gestion des quotas sur HDFS »

- **Gestion des ressources avec Yarn**
 - | Architecture de Yarn
 - | Utilisation de Yarn
 - | Les différentes façons de gérer Yarn
 - | Mise en pratique : « Configurer et gérer Yarn »
 - | Mise en pratique : « Gestion de Yarn sans Ambari »

Jour 3

- **Découverte des applications Yarn**
 - | Les bases d'une application Yarn
 - | Mise en pratique : « Démarrer une application Yarn »
- **Gestion des nœuds dans un cluster**
 - | Ajouter, enlever un nœud du cluster
 - | Déplacer des composants
 - | Mise en pratique : « Ajouter, décommissionner et recommissionner un nœud »
- **Le capacity scheduler de Yarn**
 - | Contrôler la répartition des ressources grâce aux queues Yarn
 - | Contrôler les accès sur les queues Yarn
 - | Mise en pratique : « Configuration des utilisateurs et des groupes pour Yarn »
 - | Mise en pratique : « Configurer les ressources avec les queues »
 - | Mise en pratique : « Tuning de la gestion des ressources »
- **Gestion des racks sur Hadoop**
 - | Les bénéfices de la « rack awareness »
 - | Configurer la « rack awareness »
 - | Mise en pratique : « Configurer la rack awareness »

Jour 4

- **Activer la haute disponibilité avec HDFS et Yarn**
 - | Les principes de la haute disponibilité
 - | Haute disponibilité du Namenode
 - | Haute disponibilité du Resource manager
 - | Mise en pratique : « Configurer la haute disponibilité du namenode »
 - | Mise en pratique : « Configurer la haute disponibilité du resource manager »
- **Surveillance de cluster**
 - | Surveillance avec Ambari
 - | Lever des alertes avec Ambari
 - | Mise en pratique : « Configurer les alertes avec Ambari »

- **Protéger ses données**

- | De l'importance des backups
- | Les snapshots HDFS
- | Utiliser DistCP
- | Mise en pratique : « Gestion des snapshots HDFS »
- | Mise en pratique : « Utiliser DistCP »

- **Installer la HDP**

- | Identifier les options de déploiement de cluster
- | Planifier un déploiement de cluster
- | Faire une installation avec Ambari
- | Mise en pratique : « Installer la HDP »