

## **Administrer la plateforme Hadoop 2.X Hortonworks : fondamentaux**

**Formation officielle Hortonworks \ADM 221 - HDP Operations: Administration Foundations\**

### 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 concevoir, installer, configurer et maintenir un cluster Hadoop.

A l'issue de cette formation, vous aurez grâce aux mises en pratique une solide compréhension d'Apache Ambari et vous appréhendez son utilisation comme outil de gestion de la plateforme Hortonworks.

### OBJECTIFS PEDAGOGIQUES

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

### PUBLIC CIBLE

- Architecte
- Administrateur

### PRE-REQUIS

- Connaissances de l'environnement Linux.
- Capacité à lire et exécuter des scripts shell Linux simples.
- Il est recommandé d'avoir des connaissances de base autour des requêtes SQL et de l'expérience sur les sujets opérationnels tels que la gestion des incidents et la gestion des versions.

### METHODE PEDAGOGIQUE

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.

**Stage pratique**  
Data Engineering

Code :  
**HWADM**

Durée :  
**4 jour(s) (28,00 heures)**

Exposés : **40.00 %**  
Cas pratiques : **50.00 %**  
Echanges d'expérience : **10.00 %**

**Inter-entreprises :**  
Prochaines sessions disponibles [sur notre site web](#).  
Tarif : 2 700,00 € HT / participant

**Intra-entreprise :**  
Tarifs et dates sur demande.

Cette formation prépare à la certification éditeur Hortonworks.

#### **PROFIL DES INTERVENANTS**

Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

#### **MODALITÉS D'ÉVALUATION ET FORMALISATION À L'ISSUE DE LA FORMATION**

L'évaluation des acquis se fait tout au long de la session au travers des ateliers et des mises en pratique. Une évaluation à chaud sur la satisfaction des stagiaires est réalisée systématiquement en fin de session et une attestation de formation est délivrée aux participants mentionnant les objectifs de la formation, la nature, le programme et la durée de l'action de formation ainsi que la formalisation des acquis.

#### **PROGRAMME PEDAGOGIQUE DETAILLE**

##### **Jour 1**

##### ***BIG DATA, HADOOP ET LA PLATEFORME HORTONWORKS : LES BASES DU BIG DATA***

- Les produits de la HDP
- Qu'est-ce que Hadoop ?
- Introduction à Ambari

##### ***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 »

##### ***GESTION DES UTILISATEURS AVEC AMBARI***

- Gérer les utilisateurs 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 »
- Les web services d'HDFS
- Mise en pratique : « Utiliser WebHDFS »
- Protéger les accès
- Mise en pratique : « Utiliser les ACLs HDFS »

### **Jour 2**

#### **GESTION DU STOCKAGE HDFS**

- Architecture HDFS
- Gestion d'HDFS à travers l'interface Ambari Web
- Gestion d'HDFS en ligne de commande
- Mise en pratique : « Gestion du stockage sur HDFS »
- Les quotas HDFS
- Mise en pratique : « Gestion des quotas sur HDFS »

#### **GESTION DES RACKS SUR HADOOP**

- Les bénéfices de la « rack awareness »
- Configurer la « rack awareness »
- Mise en pratique : « Configurer la rack awareness »

#### **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 »

#### **CONFIGURER LE STOCKAGE HÉTÉROGÈNE HDFS**

- Les principes du stockage hétérogène
- Mise en pratique "Configuration des règles de stockage HDFS"

#### **CONFIGURER LE CACHE CENTRALISÉ HDFS**

- De l'utilité d'un cache HDFS centralisé
- Définir et gérer des groupes et instructions de cache
- Mise en pratique « Configuration du cache centralisé HDFS »

#### **GATEWAY NFS SUR HDFS**

- Les cas d'utilisations d'une gateway NFS sur HDFS
- Architecture et opération de la gateway NFS
- Installer et configurer la gateway NFS
- Mise en pratique « configurer une gateway NFS sur HDFS »

### **Jour 3**

#### **GESTION DES RESSOURCES AVEC YARN**

- Architecture et Opération de YARN
- Les différentes façons de gérer YARN
- La gestion YARN des échecs de composants
- Mise en pratique : « Configurer et gérer YARN »
- Mise en pratique : « Gestion de YARN sans Ambari »

#### **DÉCOUVERTE DES APPLICATIONS YARN**

- Les bases d'une application YARN
- Mise en pratique : « Démarrer une application YARN »

#### **LE CAPACITY SCHEDULER DE YARN**

- Contrôler la répartition des ressources grâce aux queues YARN
- Configuration et gestion des queues YARN
- Contrôler les accès sur les queues YARN
- Mise en pratique : « Configurer le capacity scheduler »
- Mise en pratique : « Gérer les ressources et queues YARN »
- Mise en pratique : « Gérer les autorisations et les limites utilisateurs pour YARN »

#### **LES LABELS SUR LES NŒUDS YARN**

- Principes de base et application
- Activer et configurer les labels

- Gestion des labels (ajout, suppression et modification)
- Configurer les queues pour accéder aux ressources des labels
- Tester les labels pour valider leur comportement
- Mise en pratique : « Configurer les labels de nœuds YARN »

#### 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 »

##### **GESTION DES NŒUDS DANS UN CLUSTER**

- Ajouter, enlever un nœud du cluster
- Déplacer des composants
- Mise en pratique : « Ajouter, décommissionner et recommissionner un nœud »

##### **SURVEILLANCE DE CLUSTER**

- Surveillance avec Ambari
- Lever des alertes avec Ambari
- Mise en pratique : « Configurer les alertes avec Ambari »

##### **LES BLUEPRINTS AMBARI**

- Déploiement de cluster à la volée grâce aux blueprints
- Mise en pratique : "Déploiement de cluster avec les blueprints Ambari"

##### **MONTÉE DE VERSION HDP**

- Comprendre la stack HDP et sa version
- Les types et méthodes de montée de version avec HDP
- Le processus de montée de version, restrictions et prérequis
- Mise en pratique : "Faire une montée de version HDP"

