

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

Code formation : EMPHA

Durée : 3 jours – 21h de cours

Format : Inter-entreprise*

3 jours

21 heures de cours

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

| Description

Cette formation est destinée aux administrateurs de plateforme HDP qui souhaiteraient approfondir leurs connaissances. Le focus est mis sur la haute disponibilité des divers outils, la gouvernance de la donnée, les réglages avancés de la plateforme ainsi que l'automatisation de déploiement de celle-ci.

| Objectifs pédagogiques

- Mettre en place une gestion avancée du stockage sur HDFS
- Piloter une gestion avancée de l'utilisation des ressources avec YARN et Slider
- Réaliser un paramétrage avancé des outils de développement
- Assurer une gouvernance de la donnée avec Falcon et Oozie
- Automatiser le déploiement avec Ambari Blueprint

| Publics

Administrateur et architecte de la plateforme HDP cherchant à approfondir et à développer leurs compétences

| Pré-requis

- Avoir suivi la formation "Administrer la plateforme Hadoop 2.X Hortonworks 1" serait un plus.
- Expérience en ligne de commande.
- Expérience en administration de la HDP.

| 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

Upgrade de version HDP

- Compatibilités de version
- Les prérequis à la montée de version
- Installer la nouvelle version
- Mise en pratique : "Faire une montée de version HDP"

Configuration du stockage hétérogène HDFS

- Les différents stockages physiques supportés
- Stockage asymétrique
- Politique de stockage de la donnée
- Configurer ses politiques
- Mise en pratique : "Configurer les politiques de stockage HDFS"

Gateway NFS sur HDFS

- Monter une gateway NFS
- Comment scaler son gateway ?
- Authentification des utilisateurs
- Installation de la gateway
- Mise en pratique : "Configurer la gateway NFS"

Configurer le cache centralisé HDFS

- Qu'est-ce que le cache centralisé ?
- Mécanisme interne du cache
- Utilisation du cache
- Gestion du cache
- Mise en pratique : "Configurer le cache centralisé HDFS"

Gestion de la compression de la donnée

- Trade-off de la compression
- Quel niveau de compression pour vos données ?
- Codecs et sérialisation

Jour 2

Les labels sur les nœuds YARN

- Principes de base et application
- Gestion des labels
- Activer et configurer les labels
- Mise en pratique : "Configurer et tester les labels de nœuds YARN"

Déploiement d'applications avec apache slider

- Quels usages pour Slider ?
- L'intégration avec YARN
- Slider pour les développeurs
- Déployer une application Slider
- Mise en pratique : "Déployer HBase avec Slider"

Intégration du ldap avec Ambari

- Configurer le LDAP avec Ambari
- Synchronisation des utilisateurs
- Mise en pratique : "Intégrer un LDAP avec Ambari"

Paramétrage de HIVE

- Vers des requêtes Hive interactive
- Améliorer Hive avec Tez

Haute disponibilité du serveur HIVE

- L'architecture de Hive
- Les prérequis pour la HA
- Installer et configurer la HA
- Mise en pratique : "Configurer la HA Hive"

Gestion des workflows avec Oozie

- Retour sur les possibilités offertes
- L'architecture d'Oozie
- Déploiement de jobs sur Oozie
- Mise en pratique : "Gestion de workflow avec Oozie"

Jour 3

Haute disponibilité Oozie

- Prérequis à la HA
- Architecture HA pour Oozie
- Configurer la HA
- Mise en pratique : "Configurer la HA Oozie"

Introduction à Falcon

- Les challenges de la gouvernance de la donnée
- Le cycle de vie de la donnée
- Lineage de la donnée
- Architecture Falcon
- Déploiement de Falcon
- Les entités Falcon
- Mirroring avec Falcon
- Mise en pratique : "Configurer les entités et le mirroring Falcon"

Provisionnement de cluster avec les blueprints Ambari

- Déploiement de cluster à la volée grâce aux blueprints
- Capturer les blueprints
- Déploiement de template de création
- Les fichiers de configuration logique
- Les templates de création pour les noeuds
- Les bonnes pratiques pour le déploiement avec les blueprints
- Mise en pratique : "Déploiement de cluster et ajouts de nœuds avec les blueprints Ambari"