

# ADMINISTRATION DE HADOOP® POUR SOLUTIONS BIG DATA



Nombre de jours -> 4

**Participants** Aux administrateurs Hadoop, aux administrateurs de systèmes Linux, aux administrateurs de bases de données,  
-> aux administrateurs de réseaux et aux développeurs qui ont besoin de savoir comment installer et gérer leurs clusters de développement Hadoop.

**Pré-requis** -> Une expérience en programmation est utile mais pas obligatoire.

**Objectifs** Au cours de cette formation, vous apprendrez à installer, configurer et gérer la plateforme Hadoop d'Apache et son  
-> écosystème, mais aussi à surveiller des tâches Hadoop avec des fonctionnalités intégrées et des outils dédiés, comme par exemple Ganglia.

**Moyens pédagogiques, techniques et d'encadrement**

- 1 poste de travail complet par personne
- De nombreux exercices d'application
- Mise en place d'ateliers pratiques
- Remise d'un support de cours
- Passage de certification(s) dans le cadre du [CPF](#)
- > ● Mise en place de la Charte contrôle et qualité OPCA
- [Notre plateforme d'évaluation](#) :
  - Evaluation des besoins et objectifs en pré et post formation
  - Evaluation technique des connaissances en pré et post formation
  - Evaluation générale du stage

---

## 1 - Installation du système de fichiers distribué Hadoop (HDFS)

- Définir l'architecture globale et ses principaux composants
- Configurer et installer le système de fichiers
- Exécuter des commandes depuis la console
- Accéder aux fichiers en lecture et en écriture

## 2 - Préparer l'environnement pour MapReduce

- Revoir les principes de MapReduce
- Découvrir les démons informatiques
- Examiner une tâche MapReduce

## 3 - Planification de l'architecture

- Choisir le matériel adéquat
- Concevoir un cluster évolutif

## 4 - Développement du cluster

- Installer les démons Hadoop
- Optimiser l'architecture du réseau

## 5 - Préparation du système HDFS

- Paramétrer les principales options de configuration
- Configurer les redondance, réplication et affectation des blocs

## 6 - Déploiement de MapReduce

- Installer et paramétrer l'environnement MapReduce
- Réaliser l'équilibrage de charge par la connaissance des racks

## 7 - Créer un système de fichiers tolérant aux pannes

- Isoler les éléments dysfonctionnels
- Garantir la haute disponibilité
- Déclencher manuellement le basculement
- Automatiser le basculement avec Zookeeper

## 8 - Optimiser la fédération des nœuds de nommage

- Développer des ressources HDFS
- Gérer les volumes des espaces de noms

## 9 - Présentation de YARN

- Analyser l'architecture de YARN
- Identifier les nouveaux démons

## 10 - Affectation des ressources

- Définir des quotas pour limiter l'utilisation du système HDFS
- Planificateurs pour hiérarchiser l'accès à MapReduce

## **11 - Gestion de HDFS**

- Démarrer et arrêter des démons Hadoop
- Surveiller l'état du système HDFS
- Ajouter et supprimer des nœuds de données

## **12 - Administration de MapReduce**

- Gérer les tâches MapReduce, suivre l'avancement avec des outils d'analyse, mise en service et arrêt des nœuds de calcul

## **13 - Utiliser les outils standards intégrés**

- Gérer et déboguer les processus avec les mesures JVM
- Vérifier l'état d'Hadoop

## **14 - Utiliser des outils de personnalisation complémentaires**

- Évaluer les performances avec Ganglia et Chukwa
- Benchmarking pour garantir des performances continues

## **15 - Simplifier l'accès aux informations**

- Activer l'envoi de requêtes de type SQL avec Hive
- Installer Pig pour créer des tâches MapReduce

## **16 - Intégrer des éléments supplémentaires de l'écosystème**

- Afficher le système HDFS sous forme de tableau avec HBase
- Configurer Oozie pour la planification des workflows

## **17 - Faciliter l'entrée / la sortie de données génériques**

- Déplacer des blocs de données depuis ou vers Hadoop
- Transférer des données HDFS via HTTP avec WebHDES

## **18 - Recueillir des données d'application**

- Récupérer des fichiers journaux provenant de différentes sources avec Flume, importer et exporter des informations sur les bases de données relationnelles avec Sqoop

## **19 - Planification des stratégies de sauvegarde, de récupération et de sécurité**

- Gérer les pannes matérielles courantes
- Sécuriser le cluster Hadoop