

Durée : 4 jours soit 28 heures

Référence : IF-CLOUDERA



Public visé :

- Administrateurs systèmes et toutes personnes qui gèrent et maintiennent les clusters Apache Hadoop en production ou qui évoluent dans un environnement de développement.



Pré-requis :

- Avoir une expérience dans l'administration système Linux.
- Aucune expérience avec Hadoop n'est nécessaire



Objectifs pédagogiques :

- Poser les bases des clusters Hadoop : choix du matériel, configuration des noeuds
- Configurer, déployer et gérer des Clusters : utiliser "Cloudera Manager for Easy Installation", "Configuration Management Tools"
- Ajout et suppression des noeuds, rebalancing et optimisation d'un cluster.
- Gestion des ressources : Le "FIFO Scheduler", Le "FairScheduler"
- Gérer les "Log files" d'Hadoop
- Utiliser le "NameNode" et "Job Tracker Web UI"
- Troubleshooting
- Comment « Benchmarker » son cluster dans le temps
- Poser les bases de l'environnement Hadoop et HDFS



Compétences acquises à l'issue de la formation :



Modalités pédagogiques :

Session dispensée en présentiel ou téléprésentiel, selon la modalité inter-entreprises ou intra-entreprises sur mesure. La formation est animée par un(e) formateur(trice) durant toute la durée de la session et présentant une suite de modules théoriques clôturés par des ateliers pratiques validant l'acquisition des connaissances. Les ateliers peuvent être accompagnés de Quizz. L'animateur(trice) présente la partie théorique à l'aide de support de présentation, d'animation réalisée sur un environnement de démonstration. En présentiel comme en téléprésentiel, l'animateur(trice) accompagne les participants durant la réalisation des ateliers.

Note relative au programme : Les ateliers, démonstrations, travaux pratiques et scénarios présentés dans ce programme sont fournis à titre d'exemples illustratifs et ne sont pas exhaustifs. Afin de garantir une adéquation permanente avec les évolutions technologiques, le contenu pédagogique ainsi que les environnements de travaux pratiques (TP) sont susceptibles d'être adaptés, ajustés ou mis à jour au cours du temps.



Moyens et supports pédagogiques :

Cadre présentiel

Salles de formation équipées et accessibles aux personnes à mobilité réduite.

- Un poste de travail par participant
- Un support de cours numérique ou papier (au choix)
- Un bloc-notes + stylo
- Vidéoprojection
- Connexion Internet
- Accès extranet pour partage de documents et émargement électronique

Cadre téléprésentiel

Session dispensée via notre solution iClassRoom s'appuyant sur Microsoft Teams.



- Un compte Office 365 par participant
- Un poste virtuel par participant
- Un support numérique (PDF ou Web)
- Accès extranet pour partage de documents et émargement électronique



Modalités d'évaluation et de suivi :

Avant

Afin de valider le choix d'un programme de formation, une évaluation des prérequis est réalisée à l'aide d'un questionnaire en ligne ou lors d'un échange avec le formateur(trice), qui valide la base de connaissances nécessaires.

Les stagiaires disposent également, depuis leur extranet, d'un **outil d'auto-positionnement** leur permettant d'évaluer leur niveau de maîtrise des compétences visées par la formation. Cette auto-évaluation est réalisée avant le démarrage de la formation à l'aide d'un curseur de positionnement allant du niveau le plus faible au niveau le plus élevé.

Pendant

Après chaque module théorique, un ou plusieurs ateliers pratiques permettent de valider progressivement l'acquisition des connaissances. Un quiz peut accompagner les ateliers pratiques afin de mesurer la compréhension des notions abordées.

Après

À l'issue de la formation, les stagiaires réalisent une nouvelle **auto-évaluation** sur les mêmes compétences que celles évaluées avant la formation. Cette comparaison permet de mesurer leur perception de la progression réalisée et de visualiser les compétences développées au cours de la formation.

Lorsque le programme de formation le prévoit, un examen de certification est organisé conformément aux modalités définies par l'éditeur ou le centre de certification (TOSA, Pearson VUE, ENI, PeopleCert...).

Le formateur réalise également une **évaluation individuelle des compétences** de chaque stagiaire sur l'ensemble des compétences visées par la formation. Pour chacune d'elles, il indique si la compétence est :

- Acquis ;
- En cours d'acquisition ;
- Non acquise.

Lorsque la compétence est évaluée comme **en cours d'acquisition** ou **non acquise**, le formateur formule des observations précisant les difficultés rencontrées ainsi que les actions ou recommandations permettant au stagiaire de poursuivre sa montée en compétences.

Enfin

Un questionnaire de satisfaction permet au participant d'évaluer la qualité de la prestation de formation.



Informations sur l'accessibilité :

Nos formations sont, dans la mesure du possible, conçues pour être accessibles à toutes et à tous. Afin de garantir les meilleures conditions d'accueil et d'apprentissage pour les personnes en situation de handicap, nous vous invitons à contacter notre référente handicap certifiée :

Céline SOLATGES – 05 61 34 39 80 – csolatges@iform.fr

Nous vous remercions de bien vouloir nous communiquer toute information utile à ce sujet en amont de la formation, afin de mettre en place les adaptations nécessaires et d'assurer un accompagnement optimal.

Pour en savoir plus sur les dispositifs d'accompagnement existants, vous pouvez consulter les sites suivants :

- [AGEFIPH](#)
- [FIPHFP](#)
- MDPH de votre département

Description / Contenu

Module 1 : Introduction – L'argument en faveur d'Apache Hadoop

- Pourquoi Hadoop ?
- Concepts fondamentaux

- Composants Core Hadoop

Module 2 : Installation du cluster Hadoop

- Caractéristiques de Cloudera Manager
- Installation de Cloudera Manager
- Installation de Hadoop (CDH)

Module 3 : Le système de fichiers distribué Hadoop (HDFS)

- Caractéristiques de HDFS
- Ecrire et lire des fichiers
- Gestion de la mémoire du NameNode
- Présentation de la sécurité HDFS
- UI Web pour HDFS
- Utilisation d'Hadoop File Shell

Module 4 : MapReduce et Spark sur YARN

- YARN : le gestionnaire de ressources de cluster
- Concepts MapReduce
- Concepts Apache Spark
- Exécution des traitements sur YARN
- Explorer les applications YARN à travers les interfaces utilisateurs Web et le Shell
- Journaux d'application YARN

Module 5 : Configuration Hadoop et Daemon Logs

- Gestion des configurations avec Cloudera Manager
- Gestion des nœuds et ajout de services
- Configuration du service HDFS
- Configuration des Hadoop Daemon Logs
- Configuration du service YARN

Module 6 : Obtenir des données dans HDFS

- Sqoop : Ingestion des données issues des bases relationnelles
- Interfaces REST
- Les meilleures pratiques pour l'importation des données

Module 7 : Planification de votre cluster Hadoop

- Considérations générales
- Choisir le bon matériel
- Options de virtualisation
- Considérations sur le réseau
- Configurations des nœuds

Module 8 : Installation et configuration de Hive, Impala et

Pig

- Hive
- Impala
- Pig

Module 9 : Clients Hadoop incluant Hue

- Les clients Hadoop
- Installation et configuration des clients Hadoop
- Installation et configuration de Hue
- Authentification et autorisation de Hue

Module 10 : Configuration avancée du cluster

- Paramètres de configurations avancés
- Configuration des ports Hadoop
- Data locality : Configuration de HDFS
- Configuration de la haute disponibilité HDFS

Module 11 : Sécurité Hadoop

- Concepts du système de sécurité Hadoop
- Qu'est ce que Kerberos et comment ça marche ?
- Sécurisation d'un cluster Hadoop avec Kerberos
- Autres concepts de sécurité

Module 12 : Gestion des ressources

- Configuration des cgroups
- Le Fair Scheduler
- Configuration du Dynamique Resource Pools
- Configuration de la RAM et CPU pour YARN
- Gestion des requêtes Impala