Hadoop Cloudera : architecture et administration

Lien : https://innov-systems.com/formation/hadoop-cloudera-architecture-et-administration

DURÉE
5 jours (35h)

RÉFÉRENCE
BSI27

CATÉGORIE
Big Data : développement, langages et NOSQL

OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- Maîtriser l'administration des clusters Cloudera
- Exploiter et de gérer un cluster Hadoop
- Configurer Hadoop et ses outils
- Gérer la maintenance des noeuds dans un cluster

POUR QUI ?

- Administrateurs système

Innov Systems

Programme détaillé

1 / Introduction

- Big Data et Hadoop
- Présentation générale d'Hadoop
- Les possibilités d'implantation de l'infrastructure et les coûts associés
- Exemples d'utilisation dans différents secteurs
- NoSQL

2 / L'écosystème d'Hadoop

- Le système de fichier HDFS
- Le paradigme MapReduce et l'utilisation à travers YARN
- Le garant des informations : Zookeeper
- Le système de stockage de la donnée basé sur le socle : HBase
- La présentation de l'architecture globale : où installer les services ?
- La définition du dimensionnement pour un cluster Hadoop
- Les particularités de MapR

3 / Installation d'un cluster Cloudera

- Cluster : vue d'ensemble
- Justification d'une solution de gestion de cluster
- Caractéristiques du Cloudera Manager
- Installation du Cloudera Manager
- Installation de Hadoop (CDH)

4 / Gestion des utilisateurs

- Gestion des utilisateurs et des groupes avec Cloudera Manager

Innov Systems

5 / Configuration des services Hadoop via Cloudera Manager

- Configuration des services
- Surveillance et maintenance des services

6 / Le système de fichiers distribués Hadoop (HDFS)

- Caractéristiques de HDFS
- Ecriture et lecture de fichiers
- Gestion de la mémoire du NameNode
- Présentation de la sécurité HDFS
- IU Web pour HDFS
- Utilisation de Hadoop File Shell

7 / MapReduce et Spark avec Yarn

- Le rôle des frameworks de calcul
- YARN : le gestionnaire de ressources de cluster
- Concepts du MapReduce
- Concepts d'Apache Spark
- Exécution des traitements sur YARN
- Explorer des applications YARN via l'interface Web et Shell
- Logs applicatifs YARN

8 / Configuration de hadoop et daemon logs

- Gestion des configurations avec Cloudera Manager
- Localisation des configurations
- Changements de configuration
- Gestion des instances de rôles
- Gestion des nœuds et ajout de services

9 / Configuration du service HDFS

Inov Systems

- Configuration des journaux Hadoop Daemon
- Configuration du service YARN
- Ingestion de données dans HDFS
- Ingestion de données à partir de sources externes avec Flume
- Sqoop : Ingestion des données issues des bases relationnelles
- Interfaces REST
- Les meilleures pratiques pour l'importation des données

10 / Planification du cluster Hadoop

- Considérations générales
- Choisir le bon matériel
- Options de virtualisation
- Considérations sur le réseau
- Configuration des noeuds

11 / Installation et configuration de Hive, Impala et Pig

- Hive
- Impala
- Pig

12 / Clients Hadoop incluant Hue

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

13 / Configuration avancée du cluster

- Paramètres de configuration avancés
- Configuration des ports Hadoop

Innov Systems

- Configuration de HDFS pour la sensibilisation au rack
- Configuration de haute disponibilité HDFS

14 / Sécurité Hadoop

- Importance de la sécurité Hadoop
- Concepts du système de sécurité de Hadoop
- Présentation et fonctionnement de Kerberos
- Sécurisation d'un cluster Hadoop avec Kerberos
- Autres concepts de sécurité

15 / Gestion des ressources

- Configuration des Cgroups
- Planification des ressources
- Configuration des pools de ressources dynamiques
- Configuration de la RAM et CPU pour YARN
- Exploration du moteur de requêtes Impala

16 / Entretien du cluster

- Vérification de l'état HDFS
- Copier les données entre clusters
- Ajout et suppression de noeuds de cluster
- Rééquilibrage du cluster snapshots de répertoires
- Mise à niveau du cluster

17 / Surveillance et dépannage du cluster

- Fonctionnalités de surveillance du Cloudera Manager
- Surveillance des clusters Hadoop
- Dépannage des clusters Hadoop

Innov Systems

Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

Prochaines dates programmées

15 au 19 Fév. 2027	Présentiel - Casablanca
12 au 16 Avr. 2027	Présentiel - Casablanca
07 au 11 Juin 2027	Présentiel - Casablanca
02 au 06 Août 2027	Présentiel - Casablanca
27 Sep. au 01 Oct. 2027	Présentiel - Casablanca

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Réservation & Renseignements

Téléphone : +212 522 247 210
Email : contact@innov-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés