Hadoop Hortonworks pour Architectes et Administrateurs

Lien :
https://innov-systems.com/formation/hadoop-hortonworks-pour-architectes-et-administrateurs-

DURÉE
4 jours (28h)

RÉFÉRENCE
BSI29

CATÉGORIE
Big Data : développement, langages et NOSQL

Innov Systems

OBJECTIFS DE LA FORMATION

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

- Elaborer une architecture Hadoop
- Comprendre le fonctionnement de HDFS et YARN/MapReduce
- Exploiter et gérer un cluster Hadoop
- Être en mesure de configurer Hadoop et ses outils
- Configurer et administrer un cluster Hortonworks
- Gérer la maintenance des noeuds dans un cluster

POUR QUI ?

- Administrateurs systèmes
- Toutes personnes qui gèrent et maintiennent les clusters Apache Hadoop en production

Programme détaillé

1 / Introduction

- Présentation générale d'Hadoop
- Exemples d'utilisation dans différents secteurs
- Les possibilités d'implantation de l'infrastructure et les coûts associés
- Les fonctionnalités du framework Hadoop
- Les différentes versions
- Les distributions et leurs caractéristiques (HortonWorks, Cloudera, MapR, EMR, Apache, ...)
- Spécificités de chaque distribution
- Architecture et principe de fonctionnement
- Terminologie : NameNode, DataNode, ResourceManager, NodeManager
- Rôle des différents composants
- Le projet et les modules : Hadoop Common, HDFS, YARN, Spark, MapReduce, Oozie, Pig, Hive, HBase, ...

2 / Les outils Hadoop

- Infrastructure/Mise en oeuvre : Avro, Ambari, Zookeeper, Pig, Tez, Oozie, Falcon, Pentaho
- Gestion des données
- Exemple de sqoop
- Restitution : webhdfs, hive, Hawq, Mahout,ElasticSearch ...
- Outils complémentaires : Spark, SparkQL, SparkMLib, Storm, BigTop, Zebra
- Outils de développement : Cascading, Scalding, Flink/Pachyderm
- Outils d'analyse : RHadoop, Hama, Chukwa, kafka

3 / L'écosystème d'Hadoop

- Le système de fichier HDFS
- MapReduce et l'utilisation à travers Yarn

Innov Systems

- Zookeeper
- HBase
- Présentation de l’architecture globale
- La définition du dimensionnement pour un cluster Hadoop
- Les particularités de MapR

4 / Installation de Hadoop

- Les différents modes d'installation : local, pseudo-distribué, distribué
- Installation d’Hadoop / HDFS
- Installation des principaux composants (Hive, Pig, HBase, Flume...)
- Installation d’Hadoop et de quelques composants

5 / Configuration d’Hadoop

- Fichiers de configuration :
- core-site.xml, hdfs-site.xml
- mapred-site.xml
- yarn-site.xml
- capacity-scheduler.xml
- Création des users pour les daemons hdfs et Yarn, droits d'accès sur les exécutables et répertoires
- Lancement des services
- Démarrage des composants : hdfs, hadoop-daemon, Yarn-daemon, etc ..

6 / Cluster Hadoop

- Gestion des noeuds du cluster Hadoop
- TaskTracker, JobTracker pour MapReduce
- Gestion des tâches via les schedulers
- Administration des logs
- Utiliser un manager

Innov Systems

7 / Administration hadoop

- Outils complémentaires à Yarn et hdfs : jConsole, jconsole yarn
- Exemples sur le suivi de charges, l'analyse des journaux
- Principe de gestion des noeuds, accès JMX
- Administration HDFS :
- Outils de stockage des fichiers, fsck, dfsadmin
- Exemples simples de récupération de fichiers
- Gestion centralisée de caches avec Cacheadmin
- Déplacement d'un NameNode
- Mise en mode maintenance

8 / Concepts d'architecture complexe

- Architecture en rack : les avantages et fonctionnement et paramétrage du rack awareness
- Scalabilité et performances : le Namenode Federation
- High Availability des serveurs maitres
- Utilisation des snapshots : sauvegarde et restauration

9 / Haute disponibilité

- Mise en place de la haute disponibilité sur une distribution Ambari

10 / Sécurité

- Mécanismes de sécurité et mise en oeuvre pratique :
- Activer la sécurité avec Kerberos dans core-site.xml, et dans hdfs-site.xml pour les NameNode et DataNode
- Sécuriser de Yarn avec la mise en oeuvre d'un proxy et d'un Linux Container Executor

11 / Exploitation

- Installation d'une grappe Hadoop avec Ambari
- Tableau de bord

Innov Systems

- Lancement des services
- Principe de la supervision des éléments par le NodeManager
- Monitoring graphique avec Ambari
- Présentation de Ganglia et Kibana

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

23 au 26 Fév. 2027	Présentiel - Casablanca
20 au 23 Avr. 2027	Présentiel - Casablanca
15 au 18 Juin 2027	Présentiel - Casablanca
10 au 13 Août 2027	Présentiel - Casablanca
05 au 08 Oct. 2027	Présentiel - Casablanca

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

Innov Systems

Réservation & Renseignements

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

Innov Systems — Tous droits réservés