Hadoop : présentation de l’écosystème

Lien : https://innov-systems.com/formation/hadoop-presentation-de-lecosysteme-

DURÉE
3 jours (21h)

RÉFÉRENCE
BSI24

CATÉGORIE
Big Data : développement, langages et NOSQL

OBJECTIFS DE LA FORMATION

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

- Apprendre à élaborer une architecture Hadoop
- Connaître les différents éléments de l'écosystème Hadoop et leurs rôles respectifs
- Configurer et superviser les services
- Identifier et manipuler les outils de développement et de supervision Hadoop
- Exploiter et orchestrer les développements

POUR QUI ?

- Chefs de projets
- Développeurs
- Toute personne souhaitant comprendre les mécanismes Hadoop et le rôle de chaque composant

Innov Systems

Programme détaillé

1 / Introduction à l'écosystème Hadoop

- Rappels sur NoSQL
- Le théorème CAP
- Historique du projet Hadoop
- Fonctionnalités : stockage, outils d'extraction, de conversion, ETL, analyse, ...
- Exemples de cas d'utilisation sur des grands projets
- Les principaux composants : HDFS pour le stockage et YARN pour les calculs
- Les distributions et leurs caractéristiques (HortonWorks, Cloudera, MapR, EMR, Apache, ...)

2 / De projet et les modules

- Hadoop Common, HDFS, YARN, MapReduce Oozie, Pig, Hive, HBase
- Les fonctionnalités du framework Hadoop
- Les différentes versions
- Les principales distributions et leurs spécificités : Apache, Cloudera, Hortonworks, EMR, MapR
- Architecture et principe de fonctionnement
- Terminologie : NameNode, DataNode, ResourceManager, NodeManager
- Rôle des différents composants
- Présentation des outils d'infrastructure : ambari, avro, zookeeper
- Présentation des outils d'interfaçage avec les applications GIS
- Présentation des outils de restitution et requêtage : webhdfs, hive, hawq, impala, drill, stinger, tajo, mahout, lucene, elasticSearch, Kibana
- Les architectures connexes : spark, cassandra

3 / Administration Hadoop

- Etude des fichiers de configuration core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml et capacity-scheduler.xml
- Gestion des utilisateurs pour les daemons hdfs et yarn

Innov Systems

- Droits d'accès sur les exécutables et répertoires
- Architecture et gestion des services généraux Hadoop des utilisateurs pour les daemons hdfs et yarn
- HDFS
- YARN
- MapReduce
- HBase
- Monitoring du cluster Hadoop
- Suivi de charges
- Journaux (jConsole)
- Gestion des nœuds
- Accès JMX
- Mise en œuvre d'un client JMX
- Administration HDFS
- Stockage des fichiers : fsck, dfsadmin
- Gestion centralisée de caches avec Cacheadmin
- Sécurité
- Activation de la sécurité avec Kerberos dans core-site.xml, et hdfs-site.xml pour les NameNode et DataNode
- Gestion de la sécurité avec Apache Sentry
- Exploitation
- Supervision des éléments par le NodeManager
- Monitoring graphique avec Ambari, Kibana, Cloudera Manager
- Visualisation des alertes en cas d'indisponibilité d'un nœud
- Configuration des logs avec log4j

Inov Systems

4 / Développement Hadoop

- HDFS
- Architecture
- Commandes SHELL
- MapReduce
- Architecture de MapReduce

- Exécuter un code MapReduce
- Hue
- Introduction
- Fonctionnalités et utilisation
- HBase
- Architecture
- Commandes SHELL de création et d'interrogation de données
- Hive
- Architecture
- Création de bases de données, tables, vues
- Requêtes grâce au HiveSQL
- Manipulations de fonctions définies par l’utilisateur (UDF)
- Gestion des droits utilisateurs
- Exploration de l’outil Apache Pig
- Introduction à Pig
- Caractéristiques du langage : latin
- Installation / lancement
- Ecriture d'un script Pig
- Les fonctions de bases
- Ajouts de fonctions personnalisées
- Les UDF
- Mise en oeuvre
- Communication entre Pig et Hive
- Sqoop
- Introduction
- Usages
- Méthodes d'utilisation
- Import et export des données
- Oozie
- Comment planifier ses travaux
- Utilisation de workflows, coordinateurs, bundles

Innov Systems

5 / Applications

- Cas d'usages de Hadoop
- Calculs distribués sur des clusters Hadoop

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

19 au 23 Nov. 2026	Distanciel
20 au 22 Jan. 2027	Présentiel - Casablanca
17 au 19 Mars 2027	Présentiel - Casablanca
12 au 14 Mai 2027	Présentiel - Casablanca
07 au 09 Juil. 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