Apache Hadoop

Lien : https://innov-systems.com/formation/apache-hadoop

DURÉE
4 jours (28h)

RÉFÉRENCE
BSI23

CATÉGORIE
Big Data : développement, langages et NOSQL

OBJECTIFS DE LA FORMATION

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

- Identifier et définir les différents composants de l’écosystème Hadoop
- Installer et configurer un cluster Hadoop
- Développer des applications de manipulation de données en utilisant les différents composants de l’écosystème Hadoop
- Utiliser Hdfs, MapReduce, Yarn, Hive, Sqoop, Hbase, Oozie

POUR QUI ?

- Développeurs
- DSI
- Utilisateurs

Innov Systems

Programme détaillé

1 / Introduction

- Définition du périmètre du Big Data
- Les origines du Big Data
- La donnée en tant que matière première
- Le rôle du projet Hadoop

2 / Les composants Hadoop

- Écosystème 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
- Architecture et principes de fonctionnement de MapReduce
- HBase

3 / Principales distributions d'Hadoop

- Hortonworks
- MapR
- Cloudera

4 / Mise en place d'une infrastructure Hadoop

5 / Organisation d'un cluster

- Besoin matériel, réseau
- Choisir le matériel adapté

Innov Systems

6 / Installation Hadoop et configuration initiale

- Type de déploiement
- Installation d'Hadoop
- Installation d'autres composants (Hive, Pig, HBase, Flume...)

7 / Configuration de l’environnement Hadoop

- Étude des fichiers de configuration
- modes de configurations : ligne de commandes, API Rest, serveur http intégré, APIS natives
- Configuration des services: HDFS, YARN, Logs, etc.
- Configuration des droits d’accès

8 / Gestion d'un cluster Hadoop

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

9 / Manipulation de la base de données Hive

- Création de bases de données, tables, vues
- Requêtes grâce au HiveSQL
- Manipulations de fonctions définies par l’utilisateur (UDF)
- Gestion des droits utilisateurs

10 / Exploration de l’outil apache pig

- Pattern et best practices MapReduce
- Introduction à Pig
- Caractéristiques du langage : latin

Inov Systems

- Installation / lancement
- Ecriture d'un script Pig
- Les fonctions de bases
- Ajouts de fonctions personnalisées
- Les UDF
- Mise en oeuvre

11 / OOZIE

- Comment planifier ses travaux
- Utilisation de workflows, coordinateurs, bundles

12 / Optimiser l’exécution de vos travaux

- Partitionner vos données à l’aide d’Hive
- Archivez vos données
- Gestion de la mémoire (JVM)

13 / Sécurité en environnement hadoop

- Mécanisme de gestion de l'authentification
- Configuration des ACL

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

Inov Systems

Prochaines dates programmées

09 au 12 Fév. 2027	Présentiel - Casablanca
06 au 09 Avr. 2027	Présentiel - Casablanca
01 au 04 Juin 2027	Présentiel - Casablanca
27 au 30 Juil. 2027	Présentiel - Casablanca
28 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

Innov Systems