Hadoop Cloudera/Hortonworks pour Développeurs

Lien : https://innov-systems.com/formation/hadoop-clouderahortonworks-pour-developpeurs

DURÉE
4 jours (28h)

RÉFÉRENCE
BSI28

CATÉGORIE
Big Data : développement, langages et NOSQL

OBJECTIFS DE LA FORMATION

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

- Connaître Hadoop 2.0 et Hadoop Distributed File System
- Comprendre comment mettre en oeuvre le nouveau Framework YARN dans Hadoop 2.0
- Apprendre à programmer avec « MapReduce »
- Savoir optimiser une tache MapReduce en suivant les meilleures pratiques
- Etre en mesure de créer des workflows avec Oozie
- Comprendre comment créer et éxécuter des requêtes Hive sur des données HDFS
- Créer des tâches MapReduce permettant d'analyser des données avec Pig
- Savoir débugger des programmes « MapReduce » pour les fiabiliser

POUR QUI ?

- Développeurs

Innov Systems

Programme détaillé

1 / Introduction

- Big Data et Hadoop
- Présentation générale d'Hadoop
- Les fonctionnalités du framework Hadoop
- Exemples d'utilisation dans différents secteurs
- Le projet et les modules : Hadoop Common, HDFS, YARN, MapReduce

2 / MapReduce

- Principe et objectifs du modèle de programmation MapReduce
- La granularité de cohérence des données
- Principe et objectifs du modèle de programmation MapReduce
- Fonctions map() et reduce()
- Couples (clés, valeurs)
- Implémentation par le framework Hadoop
- Etude de la collection d'exemples
- Les bonnes pratiques pour le développement des applications MapReduce

3 / HDFS : interfaçage avec java

- Le système de fichier Hadoop Distributed File System (HDFS)
- Lecture et écriture sur le HDFS
- Manipulation des types de fichier SequenceFile
- DistributedCache et comment partager les références ?

4 / MapReduce /Yarn : principe et développement java

- MapReduce et l'utilisation à travers Yarn

Innov Systems

- Implémentation de problématiques réelles avec le paradigme MapReduce
- Implémentation de problématiques plus complexes
- Utilisation des combinateurs et partitionners
- Les bonnes pratiques pour le développement des applications MapReduce

5 / Programmation

- Configuration des jobs, notion de configuration
- Les interfaces principales : mapper, reducer
- La chaîne de production : entrées, input splits, mapper, combiner, shuffle/sort, reducer, sortie
- Partitioner, outputcollector, codecs, compresseurs..
- Format des entrées et sorties d'un job MapReduce : InputFormat et OutputFormat

6 / Outils complémentaires

- Mise en oeuvre du cache distribué
- Paramétrage d'un job : ToolRunner, transmission de propriétés
- Accès à des systèmes externes : S3, hdfs, har, ...

7 / Streaming

- Définition du streaming map/reduce
- Création d'un job map/reduce en python
- Répartition sur la ferme
- Avantage et inconvénients
- Liaisons avec des systèmes externes
- Introduction au pont HadoopR

8 / PIG

- Introduction: principes et cas d'utilisation
- Pig versus MapReduce
- Pig : développement et intégration avec Java

Innov Systems

- Extension de Pig avec des fonctionnalités développées : UDF (User-Defined Functions)

9 / HIVE

- Introduction: principes et cas d'utilisation
- Environnement et configuration
- Hive et métadonnées
- Keywords pour Hive
- Extension du HiveQL avec des fonctionnalités développées : UDF (User-Defined Functions)

10 / HBASE

- Histoire et concepts
- Architecture
- HBase versus RDBMS
- HBase shell
- HBase Java API
- Lecture / Ecriture
- Conception de schéma

11 / Compétences en Big Data du monde réel et un Hackathon

- Les modèles de conception NoSQL: allant de SQL à NoSQL
- Collecte de données Smart Meter avec Flume
- HDFS et HBase
- Analyser les données des compteurs intelligents avec Pig and Hive
- Analyse intelligente avec Mahout
- Planification du flux de travail complet avec Oozie

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

24 au 27 Nov. 2026	Distanciel
19 au 22 Jan. 2027	Présentiel - Casablanca
16 au 19 Mars 2027	Présentiel - Casablanca
11 au 14 Mai 2027	Présentiel - Casablanca
06 au 09 Juil. 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