Hadoop : développer des applications pour le Big Data

Lien : https://innov-systems.com/formation/hadoop-developper-des-applications-pour-le-big-data

DURÉE
5 jours (35h)

RÉFÉRENCE
BSI25

CATÉGORIE
Big Data : développement, langages et NOSQL

OBJECTIFS DE LA FORMATION

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

- Connaître les principes du framework Hadoop
- Développer des applications MapReduce (Yarn), Hive et Pig efficaces et pertinentes avec Hadoop
- Charger des données non structurées des systèmes HDFS et HBase
- Etre en mesure d'exploiter toute la puissance d'Hadoop

POUR QUI ?

- Concepteurs
- Développeurs

Innov Systems

Programme détaillé

1 / Introduction

- Les problématiques du Big Data
- Historique du projet Hadoop
- Comprendre les spécificités propres aux projets Big Data
- Présentation du Cloud Computing
- Clouds publics et privés
- Les architectures Big Data à base du projet Hadoop

2 / Collecte de données et application de MapReduce

- Analyse des flux de données dans l'entreprise
- Données structurées et non structurées
- 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

3 / MapReduce / Yarn: principe et développement java

- Le paradigme MapReduce et l’utilisation à travers Yarn
- 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
- Rédaction d'un premier programme et exécution avec Hadoop

Innov Systems

4 / 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
- OutputFormat
- Type personnalisé : création d'un Writable spécifique
- Utilisation
- Contraintes

5 / Outils complémentaires

- Introduction à Spring et Spring Data
- Le namespace Hadoop pour Spring
- Utiliser Spring pour simplifier la configuration Hadoop
- Mise en oeuvre du cache distribué
- Paramétrage d'un job : ToolRunner, transmission de propriétés
- Accès à des systèmes externes : S3, hdfs, har, ...

Innov Systems

6 / 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

7 / PIG

- Pattern et best practices MapReduce
- 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

8 / HIVE

- Simplification du requêtage
- Syntaxe de base
- Création de tables
- Ecriture de requêtes
- Comparaison Pig / Hive

9 / Le stockage des données sur HDFS

- Patterns d'usages et application au Cloud
- Architecture et installation d'un système HDFS, journal, NameNode, DataNode

Innov Systems

- Opérations, commandes et gestion des commandes
- L'API HDFS Java
- Analyse de données avec Apache Pig
- Le langage Pig Latin
- Utiliser Apache Pig avec Java
- Requêtage avec Apache Hive
- Réplication de données
- Partage de données sur une architecture HDFS

10 / Sécurité en environnement Hadoop

- Configuration des ACLs
- Gestion de l’authentification

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

01 au 05 Fév. 2027	Présentiel - Casablanca
29 Mars au 02 Avr. 2027	Présentiel - Casablanca
24 au 28 Mai 2027	Présentiel - Casablanca
19 au 23 Juil. 2027	Présentiel - Casablanca
13 au 17 Sep. 2027	Présentiel - Casablanca

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

Inov Systems

Réservation & Renseignements

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

Document généré le 25/11/2026 — Réf : BSI25
Innov Systems — Tous droits réservés

Innov Systems