



# Hadoop : installation et administration

 DURÉE  
**5 jours (35h)**

 RÉFÉRENCE  
**BSI26**

 CATÉGORIE  
**Big Data :  
développement,  
langages et NOSQL**

## OBJECTIFS DE LA FORMATION

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

- ✓ Appréhender les concepts et les enjeux liés à Hadoop
- ✓ Comprendre le fonctionnement de la plateforme et de ses composants
- ✓ Savoir installer Hadoop
- ✓ Etre capable d'administrer un cluster
- ✓ Optimiser la disponibilité et le débit des données

## POUR QUI ?

- ✓ Administrateurs de cluster Hadoop
- ✓ Développeurs



## ☰ Programme détaillé

### 1/ Vue d'ensemble d'apache Hadoop

- Big Data et Hadoop
- Architecture et principe de fonctionnement Hadoop
- Principales Distributions d'Hadoop : Apache, Cloudera, Hortonworks, EMR, MapR, DSE
- Spécificités de chaque distribution
- Terminologie : NameNode, DataNode, ResourceManager, NodeManager
- Rôle des différents composants Hadoop
- Avantages/inconvénients de la plateforme
- Le projet et les modules : Hadoop Common, HDFS, YARN, Spark, MapReduce, Oozie, Pig, Hive, HBase, ...
- Exemples d'utilisations d'Hadoop

### 2/ Les outils Hadoop

- Infrastructure/Mise en oeuvre : Avro, Ambari, Zookeeper, Pig, Tez, Oozie, Falcon, Pentaho
- Vue d'ensemble
- Gestion des données
- Exemple de sqoop
- Restitution : webhdfs, hive, Hawq, Mahout, ElasticSearch ...

### 3/ Préparations et configuration du cluster Hadoop

- Architecture de Hadoop Distributed File System (HDFS)
- Architecture de MapReduce
- Conception du cluster
- Considération sur le Hardware

### 4/ Installation de Hadoop

- Les différents type de déploiement de Hadoop
- Installation d'Hadoop / HDFS
- Installation des principaux composants (Hive, Pig, HBase, Flume...)
- Mise en œuvre : installation d'Hadoop et de quelques composants

## 5/ 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.

## 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

## 7/ Administration de HDFS

- Gestion des démons Hadoop
- Ajout/Suppression des nœuds
- Gestion des données dans HDFS :
- Import de données externes
- Utilisation des fichiers HDFS

## 8/ Administration de MapReduce

- Gestion des taches MapReduce
- Gestion des nœuds de calcul

## 9/ Configuration avancée

- Configuration de la sécurité
- Reprise sur échec
- NameNode high availability (MRV2/YARN)

## 10/ Monitoring et optimisation tuning

- Monitoring (Ambari, Ganglia...)
- Benchmarking/profiling d'un cluster
- Les outils Apache GridMix, Vaaidya
- Choisir la taille des blocs
- Autres options de tuning

## 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 Juin 2026	📍 Casablanca - Maroc
📅 27 au 31 Juil. 2026	📍 Casablanca - Maroc
📅 21 au 25 Sep. 2026	📍 Casablanca - Maroc
📅 16 au 20 Nov. 2026	📍 Casablanca - Maroc

📅 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](mailto:contact@innov-systems.com)  
🌐 **Web** : <https://www.innov-systems.com>

Scannez pour accéder  
à la fiche en ligne

Innov Systems