

## Hadoop - Développement

### Présentation

Pour faire face à un volume croissant de données à analyser, et fournir aux internautes des résultats de recherche pertinents, le framework Hadoop voit le jour au milieu des années 2000.

Devant le succès rencontré par la solution, Hadoop passera chez Cloudera qui créera alors la distribution Hadoop la plus utilisée à ce jour.

Durée : 21,00 heures (3 jours)

Tarif INTRA : [Nous consulter](#)

### Objectifs de la formation

- Connaître les différents éléments de l'écosystème Hadoop et leurs rôles respectifs.
- Comprendre l'architecture des applicatifs Hadoop.
- Comprendre les apports et cas d'usage des solutions Hadoop.

### Prérequis

- Avoir des connaissances en matière de développement d'applications d'entreprise, d'intégration de systèmes commerciaux et/ou de conception, d'interrogation et de rapports de bases de données.

### Public

- Développeurs de logiciels,
- Administrateurs de bases de données,
- Analystes BI,
- Architectes logiciels,
- Chefs de projet.

## Programme de la formation

### Architecture Hadoop

- Histoire de Hadoop - Facebook, Dynamo, Yahoo, Google
- Le noyau Hadoop



- Architecture YARN, Hadoop 2.0

### **Systeme de fichiers distribués Hadoop (HDFS)**

- Clusters HDFS - NameNodes, DataNodes et clients
- Métadonnées
- Administration en ligne

### **MapReduce**

- Traitement et génération de grands ensembles de données
- Fonctions Map
- Programmation de MapReduce à l'aide de SQL / Bash / Python
- Traitement parallèle
- Failover

### **Entreposage de données avec Hive**

- Synthétisation des données
- Requêtes ad-hoc
- Analyser de grands ensembles de données
- HiveQL (langage de requête de type SQL)
- Intégration avec les bases de données SQL
- Analyse des n-grammes

### **Traitement parallèle avec Pig**

- Évaluation parallèle
- Interface du langage de requête
- Algèbre relationnelle

### **Extraction de données avec Mahout**

- Regroupement ou Clustering
- Classification
- Filtrage collaboratif par lots

### **Recherche avec Elastic Search**

- Concepts de recherche Elastic
- Installation, importation des données
- Démonstration de l'API, exemples de requêtes

**Stockage de données structurées avec HBase**

- Big Data : échelle de taille d'un grand ensemble de données.
- Optimiser l'accès en lecture/écriture en temps réel

**Base de données multi-master Cassandra**

- Le modèle de données de Cassandra
- Cohérence éventuelle
- Quand utiliser Cassandra

**Redis**

- Modèle de données Redis
- Quand utiliser Redis

**MongoDB**

- Modèle de données MongoDB
- Installation de MongoDB
- Quand utiliser MongoDB

**Kafka**

- L'architecture de Kafka
- Installation
- Exemple d'utilisation
- Quand utiliser Kafka

**Architecture Lambda**

- Concept
- Hadoop + Intégration du traitement des flux
- Exemples d'architecture

**Big Data dans le Cloud**

*(Si le temps le permet)*

- Amazon Web Services
- Concepts : modèle de paiement à l'utilisation
- Amazon S3, EC2, EMR
- Plateforme Google Cloud
- Google Big Query

## Organisation

### Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

### Moyens pédagogiques et techniques

- Apports didactiques pour apporter des connaissances communes.
- Mises en situation de réflexion sur le thème du stage et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel dans les locaux de Softeam, les stagiaires sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc ou paperboard. Un carnet de notes est offert. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

### Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation

En amont de la formation :

- Recueil des besoins : permet de récolter des informations sur le stagiaire (profil, formation, attentes particulières, ...).
- Auto-positionnement des stagiaires afin de mesurer le niveau de départ.

Tout au long de la formation :

- Évaluation continue des acquis via des questions orales, exercices / projet fil rouge, des QCM, des cas pratiques et mises en situation.

A la fin de la formation :

- Auto-positionnement des stagiaires afin de mesurer l'acquisition des compétences.
- Evaluation du formateur des compétences acquises par les stagiaires.
- Questionnaire de satisfaction à chaud : permet de connaître le ressenti des stagiaires à l'issue de la formation.
- Questionnaire de satisfaction à froid : permet d'évaluer les apports réels de la formation et leurs mises en application au quotidien.

### **Accessibilité**

Cette formation est **accessible** aux personnes en situation de handicap, consultez-nous pour plus d'informations.