



# ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

## Formation Talend Open Studio for Big Data, exploiter ses données massives

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

### Objectifs

- | Ecrire des données sur HDFS et dans des BDD NoSQL avec des jobs Talend
- | Adopter des bonnes pratiques et concevoir des Systèmes d'informations flexibles et robustes
- | Réaliser des jobs de transformation à l'aide de Pig et Hive
- | Utiliser Scoop pour faciliter la migration de bases de données relationnelles dans Hadoop

### Public

- | Gestionnaires de données, architectes, consultants en informatique décisionnelle.

### Prérequis

- | Expérience dans l'utilisation de l'outil Talend Open Studio For Data Integration
- | ou compétences acquises durant la formation "Talend Open Studio intégration de données"

### Programme de la formation

#### Présentation de Talend Open Studio for Big Data

- | Problématique du Big Data : le modèle de 3V, les cas d'usage.
- | L'écosystème Hadoop (HDFS, MapReduce, HBase, Hive, Pig...).
- | Données non structurées et bases de données NoSQL.
- | TOS for Big Data versus TOS for Data Integration.
- | Travaux pratiques Installation/configuration de TOS for Big Data et d'un cluster Hadoop (Cloudera ou Hortonworks), vérification du bon fonctionnement.

#### Intégration de données dans un cluster et des bases de données NoSQL

- | Définition des métadonnées de connexion du cluster Hadoop.
- | Connexion à une base de MongoDB, Neo4j, Cassandra ou Hbase et export de données.
- | Intégration simple de données avec un cluster Hadoop.
- | Capture de tweets (composants d'extension) et importation directe dans HDFS.
- | Travaux pratiques Lire des tweets et les stocker sous forme de fichiers dans HDFS, analyser la fréquence des thèmes abordés et mémorisation du résultat dans HBase.

#### Import / Export avec SQOOP

- | Utiliser Scoop pour importer, exporter, mettre à jour des données entre systèmes RDBMS et HDFS.
- | Importer/exporter partiellement, de façon incrémentale de tables.
- | Importer/Exporter une base SQL depuis et vers HDFS.
- | Les formats de stockage dans le Big Data (AVRO, Parquet, ORC...).
- | Travaux pratiques Réaliser une migration de tables relationnelles sur HDFS et réciproquement.

#### Effectuer des manipulations sur les données

- | Présentation de la brique PIG et de son langage PigLatin.
- | Principaux composants Pig de Talend, conception de flux Pig.
- | Développement de routines UDF.

Référence	TAD
Durée	2 jours (14h)
Tarif	1 620 €HT

### SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

du 21 au 22 octobre 2024

[VOIR TOUTES LES DATES](#)

| Travaux pratiques Dégager les tendances d'utilisation d'un site Web à partir de l'analyse de ses logs.

### **Architecture et bonnes pratiques dans un cluster Hadoop**

- | Concevoir un stockage efficient dans HADOOP.
- | Datalake versus Datawarehouse, doit-on choisir ?
- | HADOOP et le Plan de Retour d'Activité (PRA) en cas d'incident majeur.
- | Automatiser ses workflows.
- | Travaux pratiques Créer son datalake et automatiser son fonctionnement.

### **Analyser et entreposer vos données avec Hive**

- | Métadonnées de connexion et de schéma Hive.
- | Le langage HiveQL.
- | Conception de flux Hive, exécution de requêtes.
- | Mettre en oeuvre les composants ELT de Hive.
- | Travaux pratiques Stocker dans HBase l'évolution du cours d'une action, consolider ce flux avec Hive de manière à matérialiser son évolution heure par heure pour une journée donnée.

## **Méthode pédagogique**

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

## **Méthode d'évaluation**

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

---

## **Suivre cette formation à distance**

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
- | Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
- | Privilégier une connexion filaire plutôt que le Wifi.
- | Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
- | Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
- | Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
- | Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
- | Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
- | Horaires identiques au présentiel.

---

## **Accessibilité**

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.

Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.