

## Talend : intégration de données

CB060

Durée: 3 jours

2 430 €

10 au 12 février  
19 au 21 mai

27 au 29 août  
24 au 26 novembre

### Public :

Consultants BI, Architectes, Chefs de projets, Gestionnaires de données ou toute personnes devant gérer des flux de données.

### Objectifs :

A l'issue de la formation, le stagiaire sera capable d'utiliser Talend pour intégrer et manipuler les principaux flux de données.

### Connaissances préalables nécessaires :

Maîtrise des SGBDR et de SQL.

### Objectifs pédagogiques :

Maîtriser Talend dans un environnement Big Data  
Se servir de Talend comme lien entre les fichiers, applications et bases de données  
Acquérir la philosophie de l'outil  
Adopter des bonnes pratiques et concevoir des Systèmes d'informations flexibles et robustes  
Être capable d'implémenter ses Jobs  
Lire et écrire des données sur HDFS et dans des bases de données NoSQL avec des Jobs Talend  
Réaliser des Jobs de transformation à l'aide de Pig et Hive  
Gérer la qualité de la donnée avec Talend  
Utiliser Sqoop pour faciliter la migration de bases de données relationnelles dans Hadoop  
Maîtriser l'utilisation de la bibliothèque de composants  
Effectuer des traitements ETL (Extraction, Transform and Load) simple et complexes de bout en bout

### Programme :

#### Acquérir la philosophie de l'outil

Augmenter les attentes en matière de gestion des données.  
Rappels sur les solutions ETL.  
Ecosystème Talend : Data integration, big data, realtime big datas, data fabric. Intégration, intégrité et gouvernance des données dans une seule et même plateforme unifiée.  
Présentation de TMC (talend management cloud).  
Présentation Talend OpenStudio : installation, configuration des préférences utilisateurs. Documentation.  
Concevoir des jobs simples avec Talend OpenStudio.  
Comprendre le fonctionnement de Talend : schéma et propagation de schémas au travers de composants, flux de composants.



# Phirio

---

## Se servir de Talend comme lien entre les fichiers, applications et bases de données

---

Intégrer les liens avec les bases de données : présentation des bases de données supportées.  
Opérations sur les tables, connexion à un schéma de bases de données.  
Gérer les transactions  
Utilisation de SQLBuilder pour créer des requêtes

Atelier : réaliser une transformation d'un fichier plat vers une table SQL

---

## Maîtriser l'utilisation de la bibliothèque de composants

---

Utiliser le Designer, la palette. Zoom sur les composants.  
Tour d'horizon des différentes familles de composants

Atelier : mise en oeuvre des composants jointures, tests, associations, multi-sorties

Création d'un composant. Paramétrage d'un composant. Création de liaison "Main".

Atelier : mise en oeuvre des composants tFileInputDelimited et tMap.

---

## Être capable d'implémenter ses Jobs

---

Création d'un job. Création dans le référentiel.  
Analyse du code et exécution des jobs.  
Optimiser les jobs, utilisation des joblets  
Utilisation des métadonnées, import/export, propagation sur les jobs, configuration de connexions réutilisables  
Utiliser des contextes. Exploiter le composant tContextLoad  
Gérer le traitement des jobs : les composants tPreJob et tPostJob. Recherche de jobs.  
Activer/désactiver un élément.  
Importer/exporter un élément

Atelier : création d'un flux à embranchements multiples avec exploitation du composant tRunIf

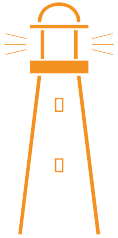
---

## Gérer la qualité de la donnée avec Talend

---

Mise en oeuvre des principales connexions.  
Intégration de fichiers XML, CSV et JSON. Filtrer et orienter la donnée avec le composant regex.  
Etude des composants de transformation.  
Les composants tAddCRCRow, tChangeFileEncoding, tUniqueRow et tIntervalMatch.  
Normaliser la donnée avec tNormalize

Atelier : création d'un flux de transformation de données et enrichissement



# Phirio

---

## Adopter des bonnes pratiques et concevoir des Systèmes d'informations flexibles et robustes

---

Modélisation des applications.

Présentation des outils : Business Modeler, JobDesigner

Partager avec les équipes une vue non technique des besoins en terme de flux, opérations, gouvernance des données.

Atelier : création d'un modèle à partir de diagrammes fonctionnels retraçant les fonctionnalités à prendre en compte, à implémenter dans le projet, développement réalisés par les Jobs.

---

## Maîtriser Talend dans un environnement Big Data

---

Intégration de Talend dans un environnement Hadoop. Parallélisation des tâches.

Comprendre les mécanismes BigData dans la suite Talend. Dénormalisation BigData et normalisation SGBD.

Présentation des composants BigData, connexion HDFS, mappage et stockage HDFS.

Les composants BigData disponibles.

---

## Lire et écrire des données sur HDFS et dans des bases de données NoSQL avec des Jobs Talend

---

Liens avec les bases de données : présentation des bases de données supportées.

Lecture et écriture HDFS et bases Nosql supportées.

Comprendre les composants tHDFSInput, tHDFSOutput et tHDFSConnection.

Atelier : création d'un Job ETL de base de transformation.

---

## Utiliser Sqoop pour faciliter la migration de bases de données relationnelles dans Hadoop

---

Comprendre le composant tSqoopExport et le projet Sqoop pour transférer des données du système de fichiers distribué Hadoop (Hadoop Distributed File System, HDFS) vers un système de gestion de base de données relationnelle (SGBDR).

Atelier : Mise en oeuvre de tSqoopExport / tSqoopImport entre une base SQL et Hadoop HDFS

---

## Réaliser des Jobs de transformation à l'aide de Pig et Hive

---

Jointure Pig. Comprendre les composants tPigLoad, tPigStoreResult et tPigMap

Comprendre les composants tHiveConnection et tHiveRow

Couplage HBase.

Atelier : réalisation d'un ETL Pig sur un lac Hadoop distribué, entre des fichiers plats et des fichiers parquets.

Atelier : réalisation d'un requêteur Hive les fichiers structurés Hadoop



# — Phirio —

---

Effectuer des traitements ETL (Extraction, Transform and Load) simple et complexes de bout en bout

---

Traitement de données multi-sources.

Le composant tMap

Génération de sorties multiples

Extensions :

décomposition de jobs, mise en oeuvre du tRunJob,

debugging, analyse des statistiques d'exécution,

reporting avec jJasperOutput

Atelier : création de jointures, transformations à l'aide des variables, expressions et jointures, qualification des données à l'aide de filtres.