

## Formation Talend Open Studio Data Integration : Initiation

<b>Durée :</b>	3 jours
<b>Public :</b>	Administrateurs systèmes/BDD, Chefs de projets, Développeurs
<b>Pré-requis :</b>	Notions de Java et fondamentaux SQL
<b>Objectifs :</b>	Comprendre le processus d'intégration de données - Maîtriser l'outil Talend Open Studio et la bibliothèque de composants
<b>Sanction :</b>	Attestation de fin de stage mentionnant le résultat des acquis
<b>Taux de retour à l'emploi:</b>	Aucune donnée disponible
<b>Référence:</b>	BUS875-F
<b>Note de satisfaction des participants:</b>	4,59 / 5

### Introduction

Problématique d'intégration de données  
Intégration Opérationnelle et Décisionnelle  
Outils ETL (Extract Transform Load) disponibles  
Talend Open Studio : présentation, licences, lien avec Eclipse  
Support disponible : documentation, communauté, ...

### Atelier : Installation et présentation de l'environnement Talend Open Studio

### Découvrir l'interface

Les perspectives et les vues Talend  
Utilisation du Business Modeler  
Les jobs : grille, palette de composants  
Les flux principaux (main)  
La notion de schéma : types des colonnes  
Premiers composants : entrée, sortie, génération de flux

### Atelier pratique : création et exécution d'un job simple, génération de contenu dans un fichier

### Les principaux composants de Talend

Composants d'entrée (tFileInput, tMysqlInput, ...)  
Génération de contenu (tRowGenerator, tFixedFlowInput, ...)  
Sorties et affichage (tLogRow, tFileOutput, tMysqlOutput, ...)  
Logique et chaînage des composants  
Suivi pas à pas des traces de l'exécution du job  
Gestion des rejets  
Transformations simples (filtre, tri, remplacement, jointure)

## **Atelier pratique : génération de contenu dans une base de données, extraction et traitement des données pour affichage en console**

### **Garantir la pérennité des jobs**

Versions et statuts des jobs

Les métadonnées : propriétés et schémas.

La découverte de schéma pour les fichiers CSV

Récupération des schémas de bases de données

Mises à jour des jobs depuis les métadonnées

### **Atelier pratique : insérer les sources et cibles de données dans les métadonnées du référentiel**

### **Variables et contextes**

Utiliser des variables dans un job (tSetGlobalVar)

Mise en place d'un contexte local au job

Groupes de contexte dans le référentiel

Les métadonnées comme contextes

### **Atelier pratique : utiliser une variable dans un job, contextualiser les paramètres d'une métadonnées, créer un contexte de test, et de production**

### **Le mappage de données : tMap**

Le composant ETL par excellence

Gérer les entrées : flux principal, secondaires, nommage des flux

Gérer les sorties : l'éditeur intégré de la tMap

Filtrer les données : en entrée, sortie

Sorties conjointes, capture des rejets de sortie, de jointure interne

L'éditeur d'expression

Les performances

### **Atelier pratique : extraction multi-sources, multi-destinations avec une tMap, sur un exemple complet**

### **Orchestration des jobs, gestion d'erreurs**

Exécutions parallèles de sous-jobs

Déclencheurs Talend (triggers)

Enchaînement de sous-jobs et de composants (OnSubjobOK, OnComponentOK)

Gestion des erreurs de composants, sous-job (OnComponentError, OnSubjobError)

Composant de départ, et de fin (tPreJob, tPostJob)

Attraper et lever des erreurs : tWarn, tDie, tLogCatcher

### **Atelier pratique : assembler plusieurs jobs en un seul, factoriser la gestion d'erreurs pour générer un fichier de compte-rendu.**