

Talend Niveau 1

Code : EBI-TALEND-100

Durée : 3 jours (21 heures)

Talend Data Integration est l'un des ETL open source les plus utilisés pour les échanges de données inter-applicatifs et l'alimentation de Data Warehouses. Cette formation de 3 jours vous guide pas à pas dans la prise en main de l'environnement, la création de jobs et la manipulation des composants essentiels. À l'issue, vous serez capable de concevoir, tester et déployer vos propres flux de données.

Public visé

Professionnels de l'informatique et développeurs ayant besoin d'implémenter des flux de données inter-applications ou devant alimenter un entrepôt de données.

Objectifs pédagogiques

À l'issue de cette formation, les apprenants seront capables de :

- Identifier l'environnement de Talend et son fonctionnement
- Préciser la philosophie des ETL (compilation de type ETL ou ELT)
- Créer des flux de données
- Transformer et gérer la donnée
- Identifier et corriger les erreurs

Programme

Jour 1

Introduction

- Définition du besoin
- EAI / ETL
 - Échange inter-application
 - L'ETL dans un projet décisionnel
- Mise en place de l'outil
 - Installation du JDK (kit de développement Java)
 - Astuces et remarques (premier démarrage)

L'environnement de développement

- Découverte de l'interface graphique Eclipse
 - Écran de démarrage
 - Connexion au workspace
 - Création d'un projet
 - Ouverture d'un projet
 - Présentation des éléments d'interface
- Définition des éléments de référentiel projet
- Initiation aux composants
 - La palette
 - Énumération des grandes familles de composants

Travaux pratiques :

- Générer un fichier Excel avec des données aléatoires et structurer un fichier

Jour 2

Développer les premiers flux

- La modélisation de flux par le Business Model
- Le développement de jobs
- Conventions et règles de développement
- `tRowGenerator`
- `tFileOutputExcel`
- `tFileInputExcel`
- `tMsgBox`
- `tSendMail`
- Lier les composants entre eux
- Exécuter un job

Transformer et gérer les données

- `tFilterRow`
- `tUniqRow`
- `tAggregateRow`
- `tSortRow`
- `tMap`

Mise en place et utilisation des métadonnées

- Définition du concept
- Créer une métadonnée Excel
- Utiliser une métadonnée Excel
- Créer une métadonnée de base de données
- Récupérer les informations d'une base de données

Manipuler une base de données avec Talend

- Définition des composants liés aux bases de données

- tAccessInput / tMysqlInput
- tAccessOutput / tMysqlOutput

Notions et composants avancés

- Définition et utilisation des variables globales
- tBuffer
- tSystem
- tServerAlive
- tUnite
- tReplicate

Travaux pratiques :

- Créer une métadonnée à l'aide de la base de données sample2007 et récupérer les tables suivantes

Jour 3

Les boucles dans Talend

- tFileList
- tInfiniteLoop
- tLoop
- tPOP
- tFlowToIterate

Gestion des logs / erreurs

- tLogRow
- tDie et tLogCatcher
- tWarn
- tStatCatcher
- tFlowMeter et tFlowMeterCatcher

Gestion des environnements et transformations réutilisables

- Définition de la notion de contexte
- Créer une variable de contexte
- Utiliser une variable de contexte
- Modifier une variable de contexte
- Développement et utilisation de routines

Préparation au déploiement de la solution

- Définition et développement d'un séquenceur
- Génération de la documentation du projet
- Construction et automatisation du projet

Travaux pratiques :

- Construire un état des ventes par pays et par mois
- Créer une cascade de jobs

Modalités d'évaluation des acquis

En cours de formation, par des travaux pratiques. En fin de formation, par un questionnaire d'auto-évaluation.

Pour aller plus loin

Cette formation donne accès à cette formation :

- Talend Niveau 2

Voir aussi

Vous serez peut-être intéressés par les formations suivantes :

- HBase - Base NoSQL Big Data
- Modéliser un système d'information décisionnel
- Oracle Data Integrator
- Talend Niveau 2

Notre accompagnement

Au-delà de la formation, nous intervenons sur ce domaine :

- Business Intelligence