



MANAGEMENT INFORMATIQUE
TELEPHONE + 06 78 77 97 78
<mailto:contact@managementinformatique.com>

ETL : TALEND OPEN STUDIO v5 – v6 - v7.3 : Développer des flux d'alimentation avec Talend Data Integrator.

Date : 11/02/2022

Réf. interne : MI 2014 - 02

Durée : 3 j

Objectifs : savoir réaliser des jobs d'alimentation à l'aide de Talend Open Studio v7.3

Public: Business Analystes, Développeurs, Consultants techniques

Moyens humains engagés :

Formateur Expert ETL Talend

Moyens pédagogiques :

Accès à ETL TALEND sur Machine virtuelle Oracle Virtual Box.

Supports de cours et d'exercices présentés sous forme de slides PDF

Prérequis :

Langage SQL, Java/Perl

(*) Ces formations restent ouvertes aux personnes en situation de handicap.

Tarif : Consulter nos CGV.

Planning détaillé :

Tour de table - Présentation

1. Modélisation & Documentation

Préférences

Le Business Modeler et l'approche Top/down

Exercice 1 / faire 1 Business Model

2. Réalisation de Job : fichiers & bases

1^{ère} création de job

Composants tMsgBox tRowGenerator tLogRow

Exercice 3 : réaliser un job (tRowGenerator vers tLogRow)

Composants tFileInputDelimited, tFileOutputDelimited

Exercice 4 : réaliser 2 jobs qui chargent 2 fichiers de type délimité et positionné

Composants DataBase tMysqlInput tMysqlOutput tMysqlRow tMysqlCommit
tMysqlRollback



MANAGEMENT INFORMATIQUE
TELEPHONE + 06 78 77 97 78
<mailto:contact@managementinformatique.com>

L'interface sqlBuilder

Exercice 5 : réaliser un job (tRowGenerator vers tDBOutput)

Composant tCreateTable

Autres composants : Bulk pour le chargement de masse

ELT les données restent dans le même SGBD

OutputBulk BulkExec OutputBulkExec

Les Méta-données : Bases de données, Fichiers

Exercice 6 : Définir les méta-données pour 2 fichiers délimités

Exercice 7 : Remplacer le composant tRowGenerator par le composant tFileInputDelimited

Le composant tMap : possibilités

Exercice 8 : charger le résultat d'une jointure entre les tables customers_in et states_in dans 3 tables de sortie ohio (customers_in.idState = 35) Nebraska(customers_in.idState = 27)

Erreurs (error_on_filter)

Exercice 9 : fonctions de transformation

(customers_in.CustomerName !=null) ? customerName.toUpperCase() : ""

Lien Main / Lookup

Exercice 10 : composant tMap,

10.1 réaliser une jointure (inner join) entre Customers (Main) et States(Lookup)

10.2 gérer des sorties multiples (ohio / nebraska / inner_join_error)

10.3 stocker les records de Customers sans correspondance sont stockés dans la table inner_join_error (option Catch lookup_inner_join_reject à true)

10.4 stocker les records ne correspondant à aucun des filtres déclarés dans tous les flux.

Error_on_filter (option Catch output_reject à true)

Exercice 11 : les filtres

Exercice 12 : collecter les données rejetées

Révisions

3. Les autres composants Processing (Transformation)

tMap

tJoin : jointure

tFilterColumns : enlever certaines colonnes

tUnite : un regroupement de flux à structure identique (équivalent union)

tUniqRow : dédoublement

tSortRow : tri des données

tExternalSortRow : appel à programme externe pour trier

tAggregateRow : regroupement et agrégation d'indicateurs

tNormalize : sépare les colonnes dans le flux en sortie avec caractère spécial

tDenormalize : dénormalise les entrées basées sur une colonne.

tBuffer : composant utilisé pour communiquer un flux à un autre job.

4. Les autres composants System (Système)

tRunJob : lancer un job

tSystem : lancer une commande système (exemple ls)



MANAGEMENT INFORMATIQUE
TELEPHONE + 06 78 77 97 78
<mailto:contact@managementinformatique.com>

tSSH : exécuter une commande système sur un serveur distant (via SSH)
tSetEnv : affecter une variable d'environnement

5. Les autres composants

File Management (Fichier)

Internet / FTP

Data Quality

XML

Business (Alfresco, CentricCRM, HealthCare, Microsoft_CRM, Open Bravo, SAP, Salesforce, SugarCRM)

BI (Jasper, OLAP Cibe, SPSS)

6. Les itérations

Lien Iterate

tFor, tPOP, tFileList

tFlowToIterate

Lister tous les fichiers d'un répertoire donné : File / tFileList

tLoop / Iterate

Récupérer tous les mails sous forme de fichiers : tPOP

7. Enchaînement de composants

Liens conditionnels : If

On Subjob OK/ On Subjob Error

On Component OK/ On Component Error

Gestion d'erreur : tLogCatcher à l'écoute des tDie et tWarn

Exercice 13 : gestion d'erreur avec le composant tLogCatcher

Exercice 14 : personnaliser les logs d'erreur

Exercice 15 : créer les tables pour stocker les logs

8. Les contextes d'exécution

Créer un groupe de contextes

Affecter des valeurs aux variables de contexte globales

F5 pour déclarer une nouvelle variable.

Exercice 17 : créer des variables de contexte et les utiliser.

9. Travaux Pratiques

Où utiliser les fonctions Perl/java ?



MANAGEMENT INFORMATIQUE
TELEPHONE + 06 78 77 97 78
<mailto:contact@managementinformatique.com>

Rassembler vos routines Perl/Java
Présenter les composants partagés tPerl*/tJava*
Injecter son propre code tPerl* / tJava* - Editeur Java/Perl intégré
Créer une classe ou une sub
Composant tPerlFlex/tJavaFlex
Composant tPerl / tPerlRow
Composant tJava / tJavaRow

10. Étendre Talend Open Studio

Maquetter un composant simple à l'aide de tPerlFlex / tJavaFlex
Concevoir un composant spécifique

11. Debugging, déploiement et optimisation

Mode pas à pas

Suivi des statistiques : tStatCatcher / tFlowMeter

Gestion des logs : tLogCatcher / tStatCatcher / tFlowMeterCatcher

Les 3 types de log :

- tLogCatcher : les log tDie, tWarn et les erreurs Java
- tStatCatcher : les stats et la durée d'exécution
- tFlowMeterCatcher : les métriques

Gestion des logs / préférences : pour ne pas à avoir à utiliser les composants tCatch, paramétrer la vue Job Settings

(sélectionner Catch Components statistics (tStatCatcher Statistics)

Fichier/Editer les paramètres du projet

Gestion des logs/préférences

12. S'appuyer sur la communauté Talend

Activity Monitoring Console/PE

Déployer vos jobs sur une machine de prod

Déploiement et optimisation

Déployer un webservice

Générer la documentation

Organisation de la formation :

- Chaque module est présenté sous forme de démonstration simple, intuitive et rapide.
- A la fin de chaque journée, la dernière demi-heure sera consacrée à des révisions ou à des questions liées à certains modules abordés dans la journée.
- Les cours seront ponctués d'Ateliers (ou workshops) permettant au cours d'exercices de concrétiser les connaissances acquises.