

Talend Open Studio Data Integration - Fonctions principales

3 j (21 heures)

Ref : TALEND1

Public

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

Pré-requis

Etre familiarisé avec le langage SQL ainsi que les modèles relationnels d'architecture de tables de base de données

Moyens pédagogiques

Formation réalisée en présentiel ou à distance selon la formule retenue
Exposés, cas pratiques, synthèse, assistance post-formation pendant trois mois
Un poste par stagiaire, vidéoprojecteur, support de cours fourni à chaque stagiaire

Modalités de suivi et d'évaluation

Feuille de présence émargée par demi-journée par les stagiaires et le formateur
Exercices de mise en pratique ou quiz de connaissances tout au long de la formation permettant de mesurer la progression des stagiaires
Questionnaire d'évaluation de la satisfaction en fin de stage
Auto-évaluation des acquis de la formation par les stagiaires
Attestation de fin de formation

Cette formation présentera aux stagiaires comment implémenter une solution ETL (Extract – Transform – Load) d'entreprise avec Talend Open Studio. Le cours expliquera comment développer, déployer et gérer des jobs Talend

Objectifs

- Comprendre la philosophie d'un ETL
- Créer des flux de données
- Transformer et gérer la donnée
- Comprendre et corriger les erreurs

Programme détaillé

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)

ENVIRONNEMENT DE DEVELOPPEMENT

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 (palette, Énumération des grandes familles de composants)

DEVELOPPER 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 GERER LES DONNEES

tFilterRow

tUniqRow

tAggregateRow

tSortRow

tMap

MISE EN PLACE ET UTILISATION DES METADONNEES

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 DONNEES AVEC TALEND

Définition des composants liés aux bases de données

tAccessInput / tMySQLInput

tAccessOutput / tMySQLOutput

NOTIONS ET COMPOSANTS AVANCES

Définition et utilisation des variables globales

tBuffer

tSystem

tServerAlive

tUnite

tReplicate

LES BOUCLES DANS TALEND

tFileList

tInfiniteLoop

tLoop

tPOP

tFlowTolterate

MISE EN PLACE DE LA GESTION LOGS/ERREURS

tLogRow

tDie et tLogCatcher

tWarn

tStatCatcher

tFlowMeter et tFlowMeterCatcher

GESTION DES ENVIRONNEMENTS ET TRANSFORMATIONS REUTILISABLES

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

PREPARATION AU DEPLOIEMENT DE LA SOLUTION

Définition et développement d'un séquenceur

Génération la documentation du projet

Construction et automatisation du projet