



STERE

Programme de Formation Talend Open Studio -Data Quality

talend

Notre contact :

Jean-Noël SAUNOIS

 contact@stere-informatique.fr

 46, rue de Lagny
93100 Montreuil

PRESENTATION : Open Studio for Data Quality, un des principaux outils de profilage de données open source du marché. Vous apprendrez à manipuler efficacement cet outil de Talend afin d'évaluer le niveau de qualité des données de votre système d'information.

OBJECTIFS

A l'issue de cette formation, les participants seront capables de :

- Se connecter aux sources de données, produire des statistiques, identifier les données à profiler.
- Choisir les différents types d'indicateurs et d'analyses adaptés aux données à contrôler.
- Mettre en œuvre des analyses complexes permettant de vérifier des règles métiers
- Définir des stratégies de correction des données erronées via des jobs de Talend Data Integration.

DURÉE DE LA FORMATION

- 2 jours, soit 14 heures de formation.

PRÉREQUIS À LA FORMATION

- Bonnes connaissances des bases de données relationnelles et du SQL.
- Connaissances de base de Talend Open Studio for Data Integration.

PUBLIC VISÉ

- Analystes métiers, intégrateurs de données, gestionnaires de données.

CONTENU DE LA FORMATION

1. La problématique de la qualité des données

- L'évaluation de la qualité des données d'un système d'information.
- Critères fondamentaux : exhaustivité, précision et intégrité des données.
- Positionnement du produit Talend Open Studio for Data Quality dans la suite Talend.

Travaux pratiques

Installation du produit, configuration des préférences.

2. Les concepts fondamentaux de TOS for Data Quality

- Les métadonnées : connexions aux bases, fichiers délimités et fichiers Excel.
- Présentation des différents types d'analyses.
- Les outils et indicateurs d'aide à la réalisation des analyses.
- L'explorateur de données.

Travaux pratiques

Effectuer une première analyse de colonne sur des données issues d'un fichier csv, exploitation des résultats obtenus.

3. Les analyses simples

- Recherche de doublons, respect de contraintes d'intervalle, format de date, d'email...
- Métriques d'une table, dépendances fonctionnelles entre colonnes.
- Identification des redondances de valeurs.
- Contrôles de cohérence entre clefs étrangères et primaires.
- Utiliser les indicateurs, les modèles, les règles et les fichiers sources.

Travaux pratiques

Réaliser une analyse de chaque type sur un jeu de données partiellement erronées.

4. Les analyses avancées

- Analyse de schéma et de la structure des tables via l'explorateur de données.
- Analyse multitable et multicolonne, respect de règles métiers.
- Recherche et visualisation de corrélation entre colonnes.
- Créer ses propres indicateurs et fichiers sources.
- Gérer les analyses.

Travaux pratiques

Créer une règle métier complexe impliquant plusieurs tables et l'associer à une tâche. Publier la règle dans la forge Talend.

5. Eléments avancés

- Utiliser des variables de contexte.
- Créer des modèles basés sur des expressions régulières.
- Exporter/importer les analyses et les données analysées.
- Corriger les données erronées avec Talend Data Integration.

Travaux pratiques

Paramétrer les métadonnées et les analyses à l'aide de variables de contexte, exporter les données analysées pour les corriger dans Talend Data Integration.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Accueil des stagiaires dans une salle dédiée à la formation.
- Supports de formation projetés sur écran.
- Etude de cas concrets.
- Mise à disposition en ligne de documents supports à la suite de la formation.

Formation à distance : utilisation de la plateforme de visioconférence client et partage d'écran pour le support de formation.

Attention : Nous ne fournissons pas le matériel informatique pour la formation. Les stagiaires doivent être équipés d'ordinateur et d'une connexion internet.

MODALITÉS D'ÉVALUATION

- Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...
- Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MODALITÉS ET DÉLAIS D'ACCES

- L'inscription doit être finalisée 24 heures avant le début de la formation.

TARIFICATION

Inter entreprise :

- 1200€ HT par personne pour 2 jours de formation.

Intra entreprise :

- Contactez-nous pour un devis personnalisé.