Talend Open Studio for Data Quality, gérer la qualité des données

Cours Pratique de 2 jours - 14h
Réf : TDQ - Prix 2024 : 1 620€ HT

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. Vous mettrez en œuvre des analyses, et mesurerez la conformité des données aux normes internes ou externes et définirez des stratégies de nettoyage des données erronées.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure 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

MÉTHODES PÉDAGOGIQUES

70 % du temps est consacré à la pratique de l'outil. Chaque participant possède son poste de travail.

LE PROGRAMME

dernière mise à jour : 06/2021

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.

PARTICIPANTS

Analystes métiers, intégrateurs de données, gestionnaires de données.

PRÉREQUIS

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

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

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.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

• Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
• À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
• Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

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

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

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.

LES DATES

CLASSE À DISTANCE
2024 : 12 déc.