



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

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

Ce cours vous apprendra à exploiter efficacement l'outil Talend Open Studio for Data Quality de manière à évaluer le niveau de qualité des données d'un système d'information. Vous mettrez en oeuvre des analyses, vérifierez des règles métiers et définirez des stratégies de correction pour les données erronées.

Objectifs

- | 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 oeuvre des analyses complexes permettant de vérifier des règles métiers
- | Définir des stratégies de correction des données erronées via des jobs de Talend Data Integration

Public

- | 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.

Programme de la formation

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.

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.

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

Référence	TDQ
Durée	2 jours (14h)
Tarif	1 550 €HT
Repas	repas inclus

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

du 12 au 13 décembre 2024

[VOIR TOUTES LES DATES](#)

partiellement erronées.

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.

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.

Méthode pédagogique

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

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
- | Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
- | Privilégier une connexion filaire plutôt que le Wifi.
- | Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
- | Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
- | Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
- | Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
- | Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
- | Horaires identiques au présentiel.

Accessibilité

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.

Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.