Intégration de données avec Talend Data Fabric

Référence : **BDDTAL002** Durée : **3 jours (21 heures)** Certification : **Aucune**

CONNAISSANCES PRÉALABLE

- Avoir des connaissances en SQL et ETL de base

PROFIL DES STAGIAIRES

- Intégrateurs, data engineers

OBJECTIFS

- Concevoir et automatiser des flux ETL
- Transformer et nettoyer les données
- Déployer et superviser des jobs Talend
- Connecter Talend à différentes sources de données

CERTIFICATION PRÉPARÉE

- Aucune

MÉTHODES PÉDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- Remise d'une documentation pédagogique numérique pendant le stage
- La formation est constituée d'apports théoriques interactif, démonstrations en direct, travaux dirigés, ateliers collaboratif, mises en situation professionnelles, projet fil rouge, études de cas réels, ateliers d'optimisation et d'industrialisation.
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement par demi-journée de formation

FORMATEUR

- Consultant-formateur expert Talend Data Fabric

MÉTHODES D'ÉVALUATION DES ACQUIS

- Validation des prérequis envoyée en amont de la formation. En l'absence de réponse d'un ou plusieurs participants, un temps sera consacré en ouverture de session pour prendre connaissance du positionnement de chaque stagiaire sur les objectifs pédagogiques évalués
- Validation progressive des compétences tout au long de la formation via exercices pratiques, quiz techniques, études de cas, réalisation de workflows ETL, corrections collectives et restitution des travaux réalisés dans le cadre du projet fil rouge
- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

ACCESSIBILITÉ DE LA FORMATION

- Si vous avez besoin d'un aménagement pour suivre votre formation dans de bonnes conditions, nous vous invitons à contacter notre référent handicap au 01 71 19 70 30 ou à l'adresse referent.handicap@edugroupe.com. Un échange en amont nous permettra d'étudier avec vous les adaptations possibles.

CONTENU DU COURS

1. JOUR 1 - Prise en main de Talend et conception des flux ETL (7h)

2. Comprendre l'architecture Talend Data Fabric (2h)

- Présentation de Talend Data Fabric
- Concepts ETL et ELT
- Architecture de la plateforme
- Talend Studio
- Talend Management Console
- Référentiel de métadonnées
- Organisation des projets
- Cycle de vie d'un flux d'intégration
- 💡 *Mises en pratique : Prise en main de Talend Studio ; Exploration de l'environnement ; Création d'un premier projet d'intégration*

3. Se connecter aux sources de données (2h)

- Connexion aux bases relationnelles
- Fichiers CSV, Excel et JSON
- Services Web et API
- Connecteurs Cloud
- Gestion des métadonnées
- Paramétrage des connexions
- Bonnes pratiques de configuration
- 💡 *Mises en pratique : Connexion à plusieurs sources de données ; Importation de métadonnées ; Validation des accès et des échanges*

4. Concevoir un premier workflow ETL (3h)

- Composants Talend
- Flux de données
- Mapping
- Chargement de données
- Gestion des erreurs
- Contrôle des traitements
- Structuration des jobs
- 💡 *Mises en pratique : Création d'un workflow ETL complet ; Chargement de données vers une base cible ; Tests et validation du flux*

5. JOUR 2 - Transformation, qualité et automatisation des données (7h)

6. Transformer et enrichir les données (3h)

- Normalisation des données
- Conversion de formats
- Jointures
- Agrégations
- Filtrage
- Calculs et enrichissements
- Gestion des données de référence
- 💡 *Mises en pratique : Création de traitements de transformation ; Enrichissement de jeux de données ; Contrôle de cohérence des résultats*

7. Mettre en œuvre des contrôles qualité (2h)

- Contrôle de qualité
- Validation des formats
- Détection des doublons
- Gestion des valeurs manquantes
- Gestion des erreurs
- Journalisation des traitements
- 💡 *Mises en pratique : Mise en œuvre de contrôles qualité ; Traitement d'anomalies ; Analyse des rapports d'exécution*

8. Automatiser et paramétrer les workflows (2h)

- Variables de contexte
- Paramétrage des jobs
- Sous-jobs
- Réutilisation des composants
- Planification des traitements
- Bonnes pratiques d'industrialisation
- 💡 *Mises en pratique : Création de jobs paramétrables ; Mise en place de traitements réutilisables ; Automatisation d'un workflow*

9. JOUR 3 - Déploiement, supervision et projet de synthèse (7h)

10. Déployer les jobs Talend (2h)

- Génération des exécutables
- Déploiement des jobs
- Gestion des environnements
- Gestion des versions
- Bonnes pratiques de déploiement
- 💡 *Mises en pratique : Génération d'un package de déploiement ; Déploiement d'un job ; Validation du fonctionnement en environnement cible*

11. Superviser et administrer les traitements (2h)

- Talend Management Console
- Monitoring des traitements
- Journaux d’exécution
- Gestion des erreurs
- Alertes
- Indicateurs d’exploitation

12. Projet fil rouge : mise en œuvre d’une chaîne d’intégration de données complète (3h)

- Connexion à plusieurs sources
- Transformation et nettoyage des données
- Contrôle qualité
- Chargement vers une cible
- Automatisation
- Déploiement
- Supervision
- 💡 *Mises en pratique : Réalisation d’un projet complet d’intégration de données ; Construction d’un workflow ETL industrialisé ; Déploiement et supervision du traitement ; Restitution des travaux et débriefing collectif*