



Réalisé par

Faissal Mouflla

Présenté à

Moad Tajmouati

---

**TALEND**

**EXTRACTION,  
TRANSFORMATION ET  
CHARGEMENT (ETL) DE  
DONNEES A L'AIDE DE  
TALEND**

# CONTENU

- 01 Introduction
- 02 Planification
- 03 Création d'un nouveau projet et un premier Job Talend
- 04 Extraction des Données et Traitement Initial
- 05 Transformation des données
  - Les valeurs manquantes**
  - Les valeurs dupliquées**
  - Vérification**
- 06 Chargement dans la base de données cible
- 07 Problèmes rencontrés et les solutions
- 08 Conclusion



# INTRODUCTION

Dans ce projet d'intégration de données, nous avons utilisé Talend Open Studio pour automatiser le processus d'extraction, de transformation et de chargement (ETL) des données à partir d'un fichier source CSV vers une base de données SQL. Les objectifs principaux étaient de gérer les valeurs manquantes, de supprimer les doublons et de préparer les données pour une interrogation et une analyse efficaces.



# PLANIFICATION

@Faissal-00's ETL de données avec Talend

View 1

+ New View

Filter by keyword or by field

2

Todo

This item hasn't been started

Draft

Test et validation

Draft

Rédaction du rapport

+ Add item

1

In Progress

This is actively being worked on

Draft

Chargement dans la base de données cible

+ Add item

4

Done

This has been completed

Draft

Compréhension des données

Draft

Création d'un nouveau projet Talend

Draft

Création d'un premier Job Talend

Draft

Transformation des données

+ Add item

Discard

Save

@Faissal-

View 1

Filter by keyword

2

Todo

This item hasn't been started

Draft

Test et validation

Draft

Rédaction du rapport

+ Add item

Transformation des données

Draft

Faissal-00 opened 3 days ago

Faissal-00 16 hours ago (edited)

Edit

Utilisez les composants et les transformations de Talend pour effectuer les actions suivantes :  
-Supprimer les doublons.  
-Traiter les valeurs nulles (remplissage des valeurs manquantes, suppression des lignes avec des valeurs manquantes, etc.)

Assignees

Add assignees...

Status

Done

Convert to issue

Copy link in project

Archive

Delete from project



# CREATION D'UN PREMIER JOB TALEND

Nous avons débuté par la création d'un Job Talend qui englobe l'ensemble du processus ETL. Ce Job coordonne les composants nécessaires pour extraire, transformer et charger les données. Le design du Job assure un flux fluide des données à travers chaque étape.

The screenshot shows the 'Nouveau Job' (New Job) dialog box in Talend Studio. The dialog has a title bar with a Talend logo and the text 'Nouveau Job'. Below the title bar, there is a subtitle 'Nouveau Job' and a button 'Ajouter un Job dans le Référentiel'. The main area contains several fields:

- Nom: TVV
- Objectif: Amélioration l'efficacité et la qualité des données
- Description: TV Chargement des données à partir d'un fichier source CSV vers une base de données SQL.
- Créé par : user@talend.com
- Verrouillé par:
- Version: 0.1
- Statut:
- Chemin d'accès: Sélectionner

At the bottom right, there are two buttons: 'Finish' and 'Cancel'.



# EXTRACTION DES DONNEES ET TRAITEMENT INITIAL

Une fois le Job créé, nous avons utilisé le composant "tFileInputDelimited" pour lire les données sources à partir du fichier CSV. Les données comprenaient des enregistrements de vente de téléviseurs avec divers attributs. Les données CSV étaient structurées avec des colonnes telles que Brand, Resolution, Size, Selling price, Original price, Operating system et Rating.





**tFileInputDelimited** : Lit des données à partir d'un fichier délimité (comme CSV) pour les utiliser dans le Job Talend.

**tv(tFileInputDelimited\_1)**

**Paramètres simples**

Type de propriété: Référentiel | DELIM:tv

Schéma: Référentiel | DELIM:tv - metadata | Modifier le schéma

"When the input source is a stream or a zip file, footer and random shouldn't be bigger than 0."

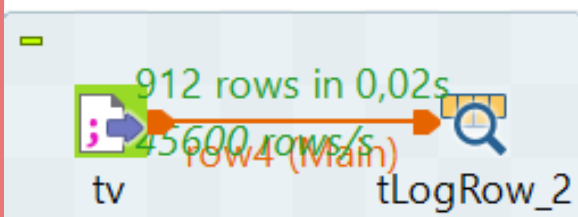
Nom de fichier/Flux: "C:/Users/Youcode/Desktop/8 Months/sprint 2/first week\_ETL avec Talend/TV.csv"

Séparateur de lignes: "\n" | Séparateur de champs: ","

☐ Options CSV

En-tête: 1 | Pied de page: 0 | Limite:

☐ Ignorer les lignes vides | ☐ Décompresser en tant que fichier ZIP | ☐ Arrêter en cas d'erreur



**tLogRow** affichant les données nettoyées : Affiche les données à des fins de débogage et de validation.

## Exécution

Exécuter Arrêter Effacer

[statistics] connecting to socket on port 3600  
[statistics] connected

tLogRow_1						
Brand	Resolution	Size	Selling_Price	Original_Price	Operating_System	Rating
TOSHIBA	Ultra HD LED	55	37999	54990	VIDAA	4.3
TCL	QLED Ultra HD	55	52999	129990	Android	4.4



# TRANSFORMATION DES DONNEES

## Les valeurs manquantes

Pour gérer les valeurs manquantes dans les colonnes "Operating System" et "Rating", nous avons utilisé le composant "tMap". À l'intérieur de "tMap", nous avons appliqué les transformations suivantes :

- Pour les valeurs manquantes de "Operating System", nous avons attribué la valeur 'Unspecified'.
- Pour les valeurs manquantes de "Rating", nous avons défini la valeur à 0.





## Exécution



Exécuter

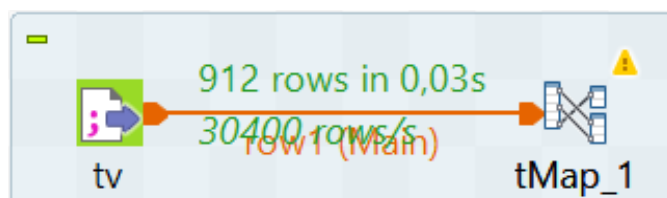


Arrêter



Effacer

Lloyd	Ultra HD LED	30	44999	64990	Linux	4.3	
Lloyd	HD LED	32	18689	26957	Linux	null	
Lloyd	Full HD LED	39	22990	39990	Linux	3.8	
Lloyd	Full HD LED	48	39999	50990	Linux	null	
Lloyd	Ultra HD LED	43	33999	69990	Linux	4.2	
Lloyd	HD LED	32	14990	19990	Linux	4.3	
Lloyd	Ultra HD LED	42	25990	59900	Linux	3.5	
Lloyd	Full HD LED	55	114990	114990	Linux	null	
Lloyd	Full HD LED	46	55990	55990	Linux	5.0	
Lloyd	HD LED	32	11000	20990	Linux	3.7	
Lloyd	HD LED	24	8103	15990	Linux	3.7	
Lloyd	Ultra HD LED	55	51999	99990	Linux	3.7	
Lloyd	HD LED	24	13200	14990	Linux	2.5	
Lloyd	Full HD LED	43	25490	39000	Linux	null	
LumX	HD LED	24	7290	17900	Android	4.3	
LumX	HD LED	40	29990	29990	Android	null	
LumX	Ultra HD LED	55	65900	65900	Android	null	
MarQ by Flipkart	HD LED	24	7499	14999		4.2	
MarQ by Flipkart	Full HD LED	24	6999	11499		4.3	
MarQ by Flipkart	Full HD LED	32	9499	16499		4.3	
MarQ by Flipkart	HD LED	32	13499	15499	Android	4.1	
MarQ by Flipkart	Full HD LED	40	18999	29499	Linux	4.3	
MarQ by Flipkart	HD LED	32	10499	18499	Linux	4.3	
Micromax	Ultra HD LED	43	91990	91990	Android	4.0	
Micromax	Full HD LED	40	29990	29990	Android	4.0	
Micromax	Full HD LED	49	65990	65990	Android	4.4	



**tMap avec les Transformations :**  
**Transforme les données en**  
**appliquant des règles spécifiques**  
**définies dans le "tMap".**



Talend Open Studio for Data Integration - tMap - Map\_1

Find :

Var

Auto map !

**row1**

Column
Brand
Resolution
Size
Selling_Price
Original_Price
Operating_System
Rating

**out1**

Expression	Column
row1.Brand	Brand
row1.Resolution	Resolution
row1.Size	Size
row1.Selling_Price	Selling_Price
row1.Original_Price	Original_Price
row1.Operating_System.equals("") ? "" : "U..."	Operating_System
row1.Rating == null ? 0 : row1.Rating	Rating

Éditeur de schéma Éditeur d'expression

row1

Colonne	C...	Type	N.	Modèle de date...	Longueur	Précision	Par dé...	Comme...
Brand		String	<input checked="" type="checkbox"/>		16	0		
Resolution		String	<input checked="" type="checkbox"/>		13	0		
Size		Integer	<input checked="" type="checkbox"/>		2	0		
Selling_Price		Integer	<input checked="" type="checkbox"/>		5	0		
Original_Price		Integer	<input checked="" type="checkbox"/>		6	0		
Operating_System		String	<input checked="" type="checkbox"/>		7	0		
Rating		Float	<input checked="" type="checkbox"/>		3	2		

out1

Colonne	C...	Type	N.	Modèle de date...	Longueur	Précision	Par dé...	Comme...
Brand		String	<input checked="" type="checkbox"/>		16	0		
Resolution		String	<input checked="" type="checkbox"/>		13	0		
Size		Integer	<input checked="" type="checkbox"/>		2	0		
Selling_Price		Integer	<input checked="" type="checkbox"/>		5	0		
Original_Price		Integer	<input checked="" type="checkbox"/>		6	0		
Operating_System		String	<input checked="" type="checkbox"/>		7	0		
Rating		Float	<input checked="" type="checkbox"/>		3	2		

Appliquer Ok Annuler

Constructeur d'expression

Expression

☒ Wrap Undo(Ctrl + Z) Effacer

row1.Rating == null ? 0 : row1.Rating

Test

Test Effacer

Var	Valeur
row1.Brand	null
row1.Res...	null
row1.Size	null
row1.Selli...	null

Ajouter Supprimer

Constructeur d'expression

Expression

☒ Wrap Undo(Ctrl + Z) Effacer

row1.Operating\_System.equals("") ? "Unspecified" : row1.Operating\_System

Test

Test Effacer

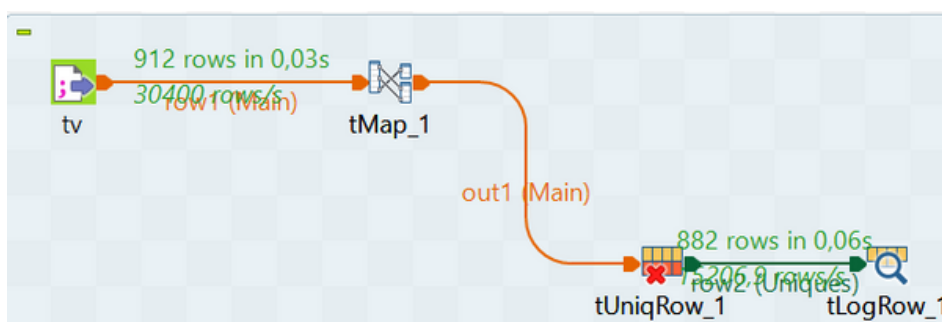
Var	Valeur
row1.Brand	null
row1.Res...	null
row1.Size	null
row1.Selli...	null

Ajouter Supprimer



## Les valeurs dupliquées

Afin d'assurer la qualité des données, nous avons utilisé le composant "tUniqueRow" pour supprimer les lignes en double de l'ensemble de données. Cela était essentiel pour éviter que des enregistrements redondants ne soient chargés dans la base de données cible.



**tUniqueRow** : Supprime les doublons de l'ensemble de données pour éviter la redondance.



## Vérification

Afin de valider la propreté des données, nous avons utilisé le composant "tLogRow". Cela nous a permis d'inspecter les premières lignes de l'ensemble de données nettoyé et de nous assurer que les transformations étaient appliquées comme prévu.



## Exécution



Exécuter



Arrêter



Effacer

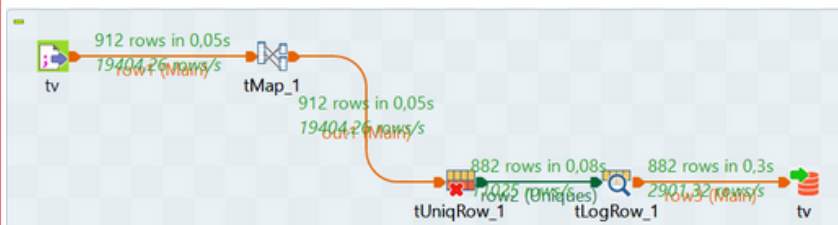
OnePlus	Ultra HD LED	65	66999	69999	Android	4.5
OnePlus	QLED Ultra HD	55	62899	69900	Android	4.6
Onida	HD LED	24	8359	13490	Android	4.2
Onida	HD LED	32	22990	22990	Android	4.7
Onida	Ultra HD LED	43	33190	52900	Android	4.2
Onida	Full HD LED	32	56990	56990	Android	5.0
Onida	Full HD LED	32	35990	35990	Android	4.7
Onix	HD LED	32	19790	19790	Android	0.0
Onix	HD LED	39	14990	26349	Android	3.7
Onix	Full HD LED	39	35990	35990	Android	0.0
Oxygen	Ultra HD LED	65	56990	124990	Android	0.0
Oxygen	Ultra HD LED	43	27999	42990	Android	0.0
Panasonic	Ultra HD LED	43	30001	79900	HomeOS	4.3
Panasonic	Ultra HD LED	49	43259	66900	HomeOS	0.0
Panasonic	Ultra HD LED	55	55890	82400	HomeOS	0.0
PHILIPS	Full HD LED	43	28999	35990	Linux	4.0
PHILIPS	HD LED	32	18999	20990	Linux	4.0
PHILIPS	Full HD LED	43	33459	36990	Android	4.5
PHILIPS	Full HD LED	22	6649	12990	Unspecified	3.9
Power Guard	HD LED	24	10990	12990	HomeOS	0.0
Power Guard	HD LED	24	11999	14990	Android	0.0
Power Guard	HD LED	32	13300	19990	HomeOS	0.0
Power Guard	HD LED	32	15650	28990	Android	0.0
realme	HD LED	32	11999	21999	Linux	4.3
realme	Ultra HD LED	55	46999	69999	Android	4.3
RGL	Full HD LED	20	4849	8990	HomeOS	3.9



# CHARGEMENT DANS LA BASE DE DONNEES CIBLE

Pour la dernière étape, nous avons utilisé le composant "tDbOutput" pour insérer les données nettoyées et transformées dans une base de données SQL. Ce composant nous a permis d'établir une connexion à la base de données cible et de charger les données dans les tables appropriées.





**tDbOutput : Insère, met à jour ou supprime des données dans une base de données cible.**

**tv(tDbOutput\_1)(Microsoft SQL Server)**

**Paramètres simples** Database: Microsoft SQL Server Appliquer

**Paramètres avancés** Type de propriété: Référentiel Bases de données (MSSQL):tv

**Paramètres dynamiques** ☐ Utiliser une connexion existante

**Vue** Fournisseur JDBC: JTDS open source

**Documentation** Hôte: "localhost" Port: "1433" Schéma: ""

Base de données: "Talend"

Utilisateur: "" Mot de passe: "\*\*\*\*\*"

Table: "TVs"

Action sur la table: ☐ Supprimer la table si elle existe et la créer ☐ Activer les insertions Identity Action sur les données: **Insert**

☐ Spécifier le champ de l'identité

Schéma: Built-in Modifier le schéma ☐ Sync colonnes

**Schéma de tv**

**tLogRow\_1 (Input - Main)**

Colonne	C...	Type	<input checked="" type="checkbox"/> N.	Modèle d...	Long...	Préci...	Par...	Com...
Brand	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		16	0		
Resolution	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		20	0		
Size	<input type="checkbox"/>	Inte...	<input checked="" type="checkbox"/>		2	0		
Selling_Price	<input type="checkbox"/>	Inte...	<input checked="" type="checkbox"/>		5	0		
Original_Price	<input type="checkbox"/>	Inte...	<input checked="" type="checkbox"/>		6	0		
Operating_Sys...	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		20	0		
Rating	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>		3	2		

**tv (Output)**

Colonne	Db Column	C...	Type	Typ...	<input checked="" type="checkbox"/> N.	Modèle...	Lon...	Pré...	Pa...	Co...
Brand	Brand	<input type="checkbox"/>	Stri...	VAR...	<input checked="" type="checkbox"/>		16	0		
Resolution	Resolution	<input type="checkbox"/>	Stri...	VAR...	<input checked="" type="checkbox"/>		20	0		
Size	Size	<input type="checkbox"/>	Int...	INT	<input checked="" type="checkbox"/>		2	0		
Selling_Price	Selling_Price	<input type="checkbox"/>	Int...	INT	<input checked="" type="checkbox"/>		5	0		
Original_Pri...	Original_Price	<input type="checkbox"/>	Int...	INT	<input checked="" type="checkbox"/>		6	0		
Operating_Sy...	Operating_Sy...	<input type="checkbox"/>	Stri...	VAR...	<input checked="" type="checkbox"/>		20	0		
Rating	Rating	<input type="checkbox"/>	Float	FLO...	<input checked="" type="checkbox"/>		3	2		

SQLQuery1.sql - L...GU3\Youcode (61))\*

```
use talend;
select * from TVs;
```

90 %

**Results** **Messages**

	Brand	Resolution	Size	Selling_Price	Original_Price	Operating_System	Rating
1	TOSHIBA	Ultra HD LED	55	37999	54990	VIDAA	4,3
2	TCL	QLED Ultra HD	55	52999	129990	Android	4,4
3	realme	HD LED	32	13999	17999	Android	4,3
4	Mi	HD LED	32	14999	19999	Android	4,4
5	realme	HD LED	32	12999	21999	Android	4,3
6	OnePlus	HD LED	32	15999	19999	Android	4,3
7	OnePlus	Full HD LED	43	25999	29999	Android	4,3
8	TCL	Ultra HD LED	65	57999	119990	Android	4,2
9	Mi	Full HD LED	40	21999	29999	Android	4,3
10	realme	Ultra HD LED	43	26999	32999	Android	4,3
11	Mi	Ultra HD LED	43	27999	34999	Android	4,3

Query executed successfully.

LAPTOP-1US3GU3\SQLEXPRESS ... LAPTOP-1US3GU3\Youcod... Talend 00:00:00 882 rows



# PROBLEMES RENCONTRES ET LES SOLUTIONS

## **Installation et Configuration de Talend :**

**Problème :** L'installation de Talend Open Studio a exigé des prérequis, notamment l'installation du kit de développement Java (JDK).

**Solution :** Pour résoudre ce problème, j'ai téléchargé et installé la version appropriée du JDK sur mon système.

## **Familiarité avec Talend :**

**Problème :** Au début du projet, je n'avais pas de connaissances préalables sur Talend, ce qui a impacté ma capacité à aborder efficacement les tâches.

**Solution :** J'ai entrepris un apprentissage actif en explorant les ressources en ligne, les tutoriels et la documentation de Talend.

## **Configuration de la Connexion à la Base de Données :**

**Problème :** Lors de la configuration de la connexion à la base de données, j'ai rencontré un problème d'incompatibilité du pilote JDBC.

**Solution :** Pour surmonter ce problème, j'ai identifié que l'utilisation du pilote JTDS était nécessaire pour établir une connexion réussie avec la base de données SQL Server. J'ai alors téléchargé et installé le package JTDS pour assurer la compatibilité avec Talend.





# CONCLUSION

En réalisant ce projet d'intégration de données ETL à l'aide de Talend, nous avons transformé des données brutes et non structurées en un format propre et bien structuré, prêt pour l'analyse. Le projet a démontré notre capacité à utiliser efficacement les composants Talend pour la manipulation, la transformation et le chargement de données, améliorant ainsi la qualité et la convivialité des données globales.





# MERCI

## QUESTION & INFORMATION



(212) 07 72 10 47 62



faysalfayfa@gmail.com



Faissal Mouflla