

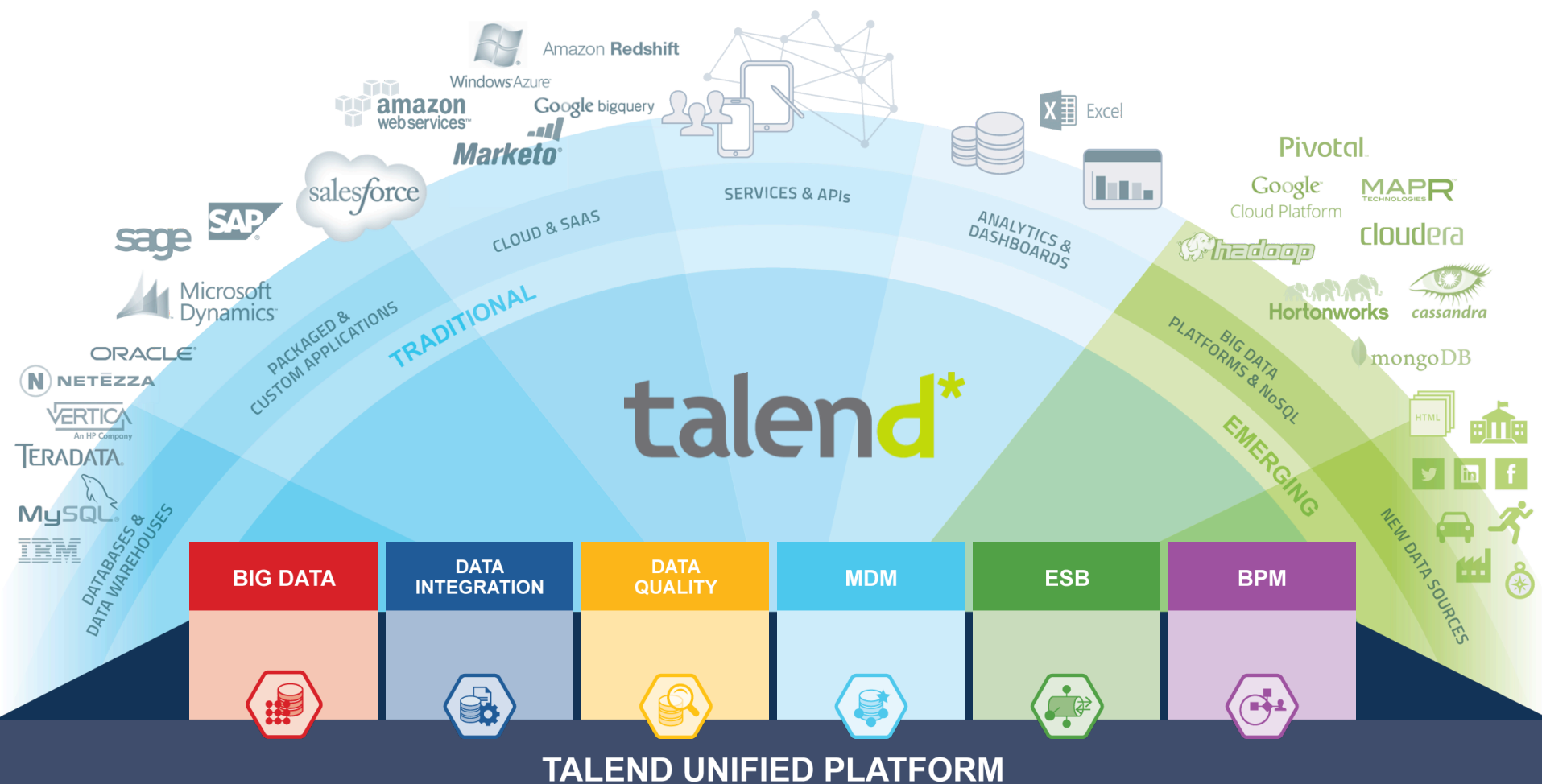
# Talend Data Intégration



**CGI**

La force de l'engagement<sup>MD</sup>

# Talend l'offre complète



# Talend Data Intégration

L'ETL de Talend permettent d'accéder, de transformer et d'intégrer des données de tout système en temps réel ou en batch afin de répondre aux besoins d'intégration de données opérationnelles et analytiques.

Avec plus de 800 composants, Talend intègre presque toutes les sources possibles de données.

Les trois principaux composants sont :

- **Le Business Modeler :**  
Outil graphique pour créer des logiques métier.
- **Le Job Designer**  
Le Job Designer sert à créer des diagrammes, déterminer des développements de données et des séquençages de flux de manière visuelle.
- **Le Metadata Manager**  
Le Metadata Manager stocke et gère toutes les métadonnées des projets, notamment les données contextuelles, comme les informations de connexion aux bases de données et les chemins d'accès aux fichiers.

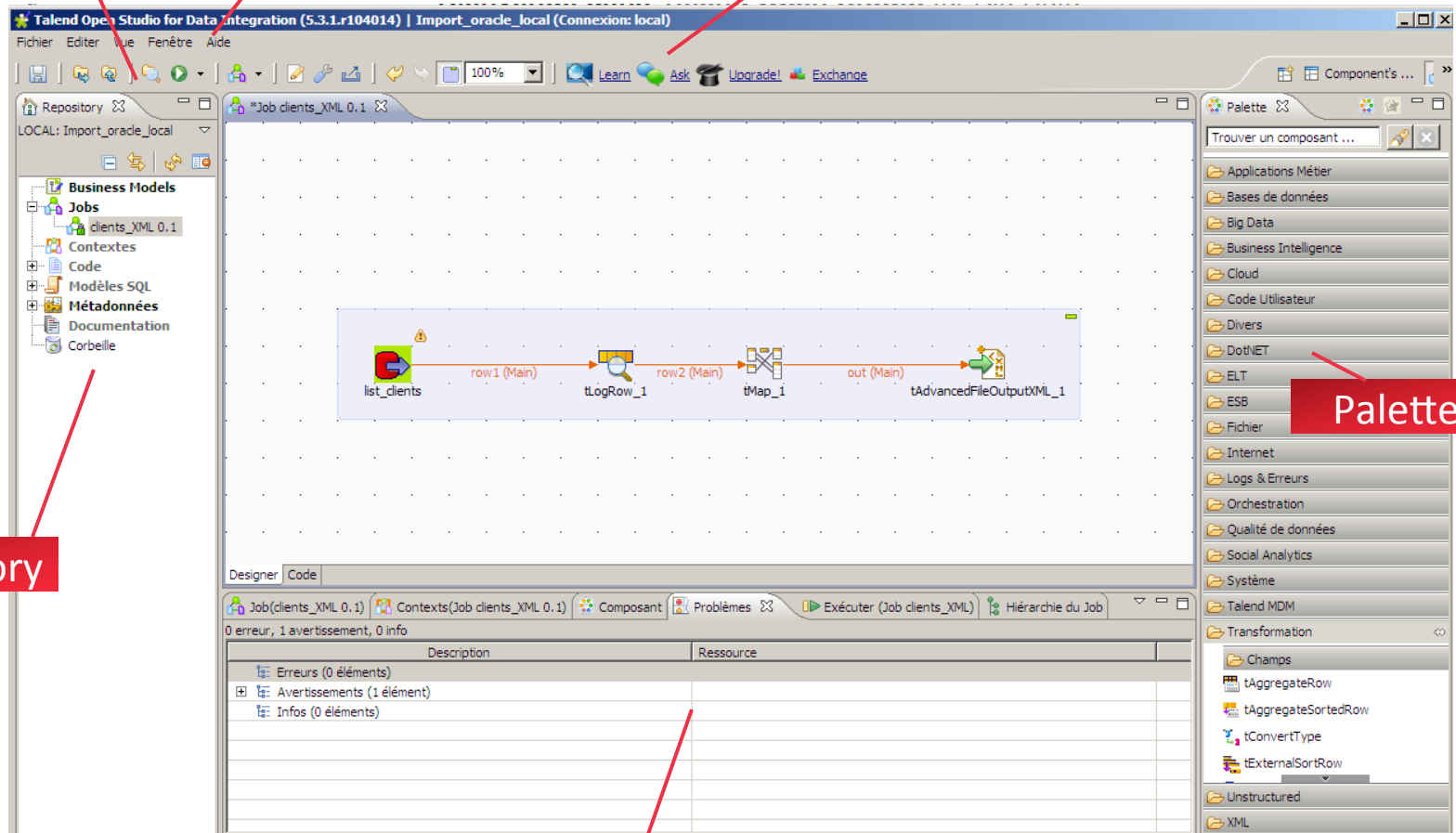


# Interface de Talend Open Studio

Barre d'outils

Menu

Espace de modélisation graphique



Repository

Palette

Onglets de configuration

# ETL vs ELT



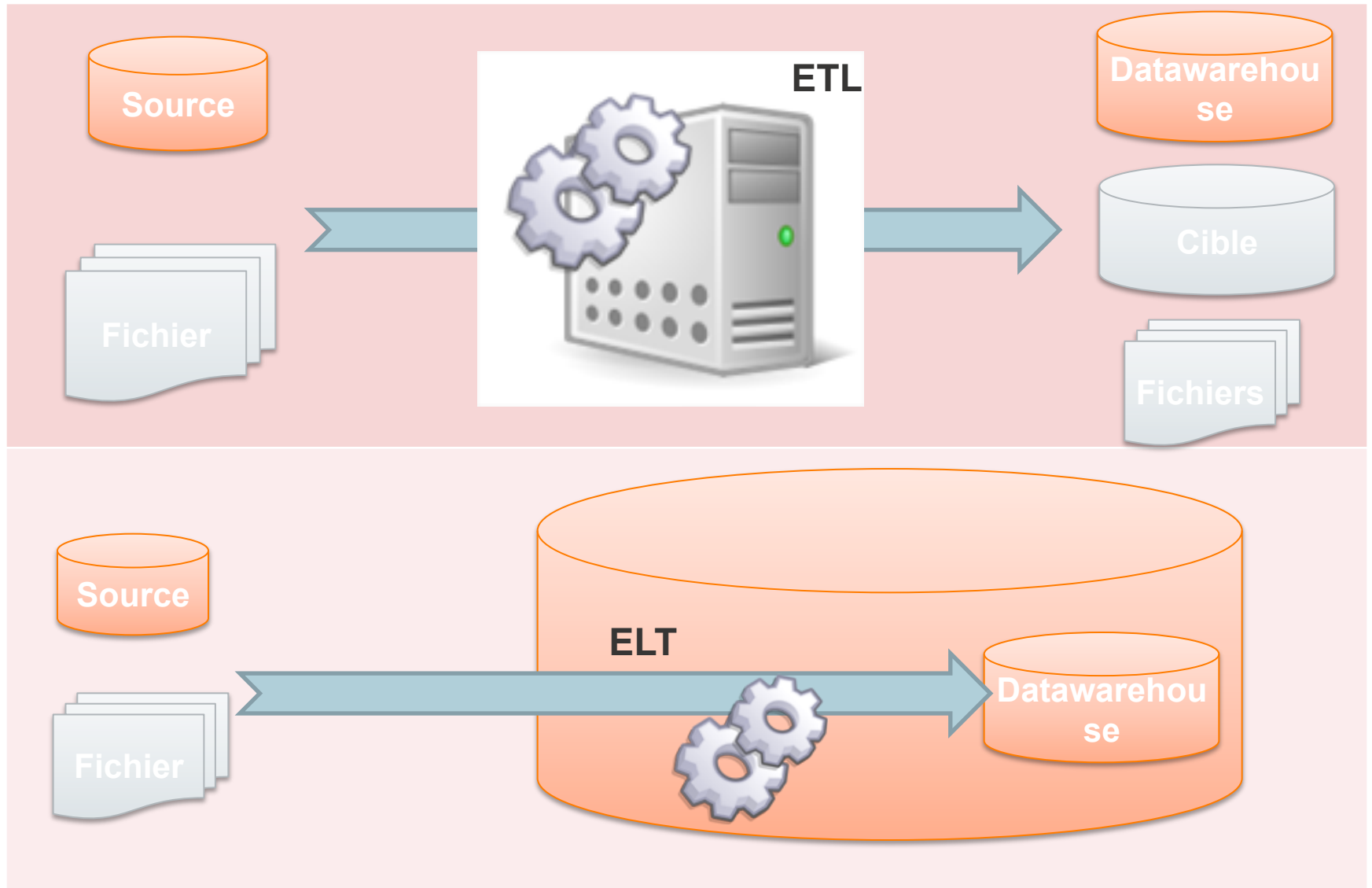
**CGI**

La force de l'engagement<sup>MD</sup>

# Qu'est ce qu'un ETL

- ETL est l'acronyme de Extract Transform Load. Un ETL combine en elle seule trois fonctionnalités :
  - Extraction depuis une source de données
  - Transformation des données issues de l'étape précédente
  - Chargement dans une nouvelle source de données
- On utilise les ETL pour les projets d'intégration de données (BI), de synchronisation de sources de données et de migration de données...

# ETL vs ELT





# Les principaux concepts



**CGI**

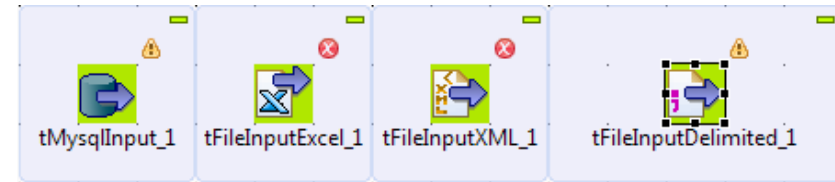
La force de l'engagement<sup>MD</sup>



# Les principaux concept

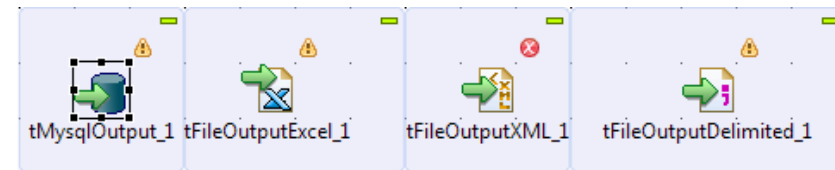
- **Les composants d'entrée :**

Lecture de fichiers (CSV, XML, positionnels, ...), bases de données ou services web, ils sont nommés en général txxxInput.



- **Les composants de sortie**

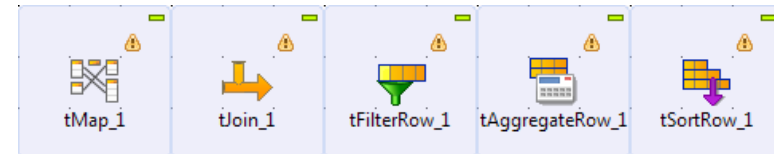
Ecriture/Mise à jour dans fichiers, bases de données ou services web, ils sont nommés en général txxxOutput.



- **Les composants de transformation**

Permettent de gérer graphiquement la transformation des données

Ex : Taggregate, tfilter, tmap, tjoin, ...



- **Les contextes**

Permettent de variabiliser les chemins d'accès, connexions, filtres ou autres données de configuration, en fonction des environnements



# Création d'un contexte

The image shows the Talend Studio interface. On the left, the 'Repository' tree is visible under 'LOCAL: Miage'. The 'Contextes' folder is selected, and a context menu is open with the following options:

- Créer un groupe de contextes
- Créer un dossier
- Importer des éléments
- Exporter des éléments

The main window displays the 'Créer / Editer un groupe de contextes' dialog box, which is divided into two steps:

**Etape 1 sur 2**  
Il est recommandé de ne pas laisser le champ Objectif vide.

Fields in Step 1:

- Nom: general
- Objectif: (empty)
- Description: (empty)
- Créé par: test@talend.com
- Verrouillé par: (empty)
- Version: 0.1
- Statut: (empty)
- Chemin d'accès: (empty)

**Etape 2 sur 2**  
Définir les contextes, les variables et les valeurs

Fields in Step 2:

- Créer un groupe de contextes
- Créer un dossier
- Importer des éléments
- Exporter des éléments

The main window displays the 'Créer / Editer un groupe de contextes' dialog box, which is divided into two steps:

**Etape 2 sur 2**  
Définir les contextes, les variables et les valeurs

Fields in Step 2:

- Créer un groupe de contextes
- Créer un dossier
- Importer des éléments
- Exporter des éléments

	Name	Type	Production	
			Value	Value
1	data_directory	String	D:/miage	D:/miage/production

Default context environment: Default

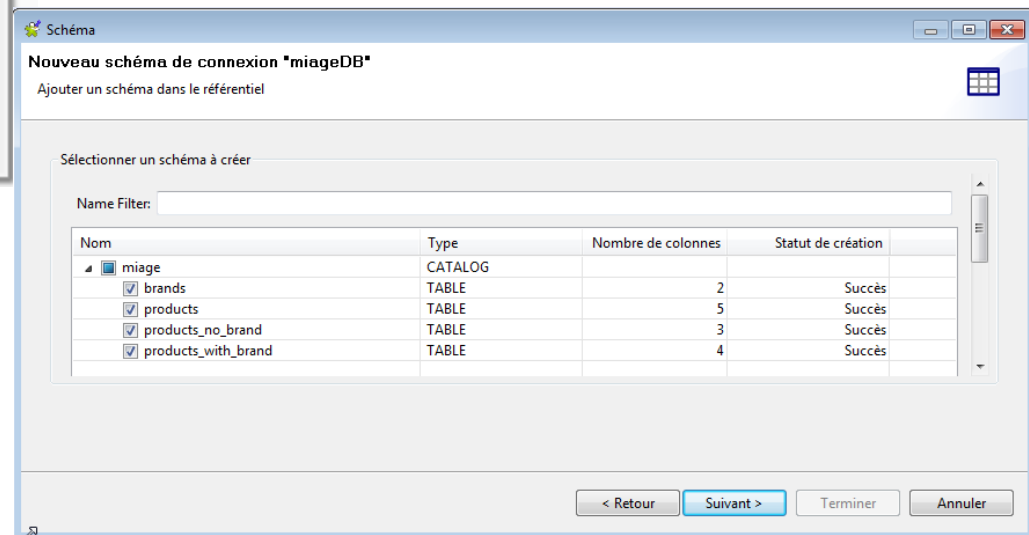
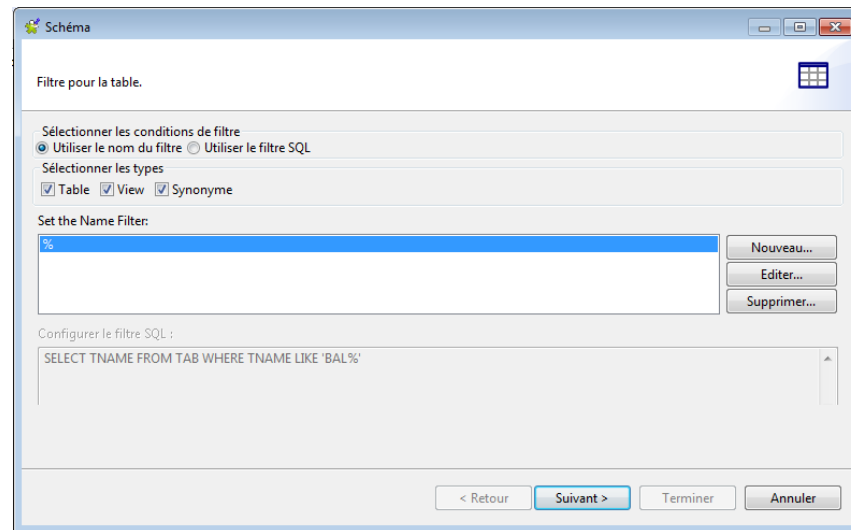
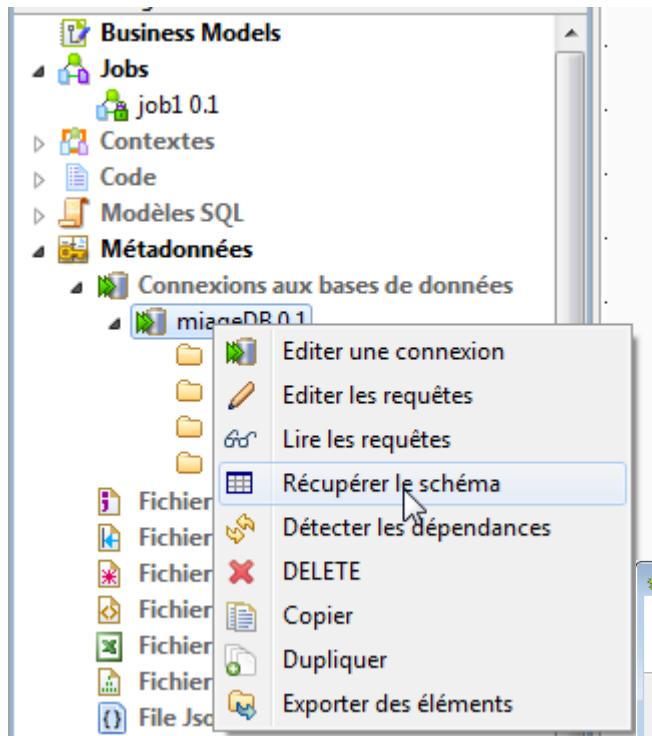
# Création d'un métadonnée

The screenshot shows the Talend Studio interface. On the left, the 'Business Models' tree is expanded to 'Métadonnées', and the 'Connex' menu is open, showing options like 'Créer une connexion', 'Créer un dossier', 'Importer des éléments', and 'Exporter des éléments'. The 'Créer une connexion' option is highlighted. In the center, a dialog box titled 'Nouvelle connexion à une base de données dans le référentiel - Etape2/2' is open, showing the 'Définir les paramètres de connexion' section. The 'DB Type' is set to 'MySQL', and the 'Version de la base de données' is set to 'MySQL 5'. The 'Chaine de caractères de connexion' is 'jdbc:mysql://localhost:3306/miage?noDatetimeStringSync=true'. The 'Login' is 'root', and the 'Mot de passe' is masked with dots. The 'Server' is 'localhost', the 'Port' is '3306', and the 'DataBase' is 'miage'. The 'Paramètres supplémentaires' are 'noDatetimeStringSync=true'. The 'Vérifier' button is visible. Below this, the 'Propriétés de la base de données' section shows 'Syntaxe SQL' set to 'SQL 92', 'Séparateur de chaîne de caractères' set to '|', and 'Caractère Null' set to '000'. The 'Exporter en tant que contexte' button is highlighted. At the bottom, there are buttons for '< Retour', 'Suivant >', 'Terminer', and 'Annuler'. On the right, another dialog box titled 'Connexion à la base de données' is open, showing the 'Nouvelle connexion à une base de données dans le référentiel - Etape1/2' section. The 'Nom' is 'miageDB', and the 'Objectif' is empty. The 'Description' is empty. The 'Créé par' is 'test@talend.com', and the 'Verrouillé par' is empty. The 'Version' is '0.1', and the 'Statut' is 'M'. The 'Chemin d'accès' is empty, and the 'Sélectionner' button is visible. At the bottom, there are buttons for '< Retour', 'Suivant >', 'Terminer', and 'Annuler'.

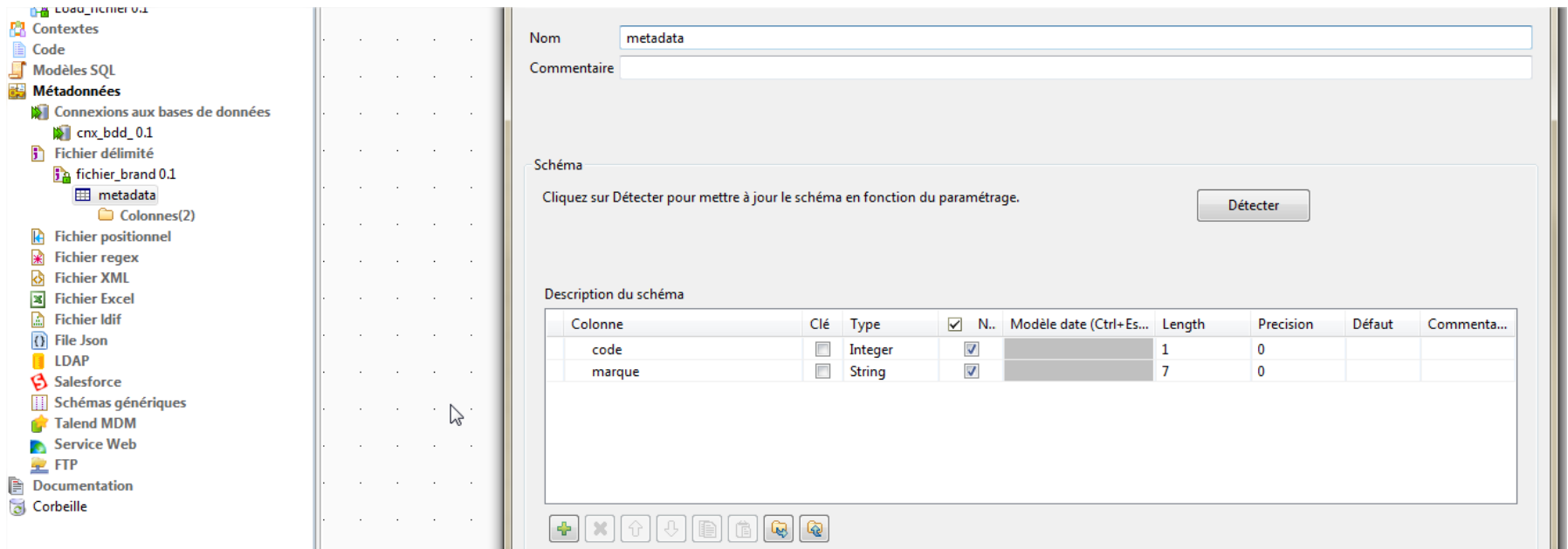
Create a new context or reuse the existing one

- ☒ Create a new repository context
- ☐ Reuse a existing repository context

# Récupération des schémas des tables



# Récupération du format d'un fichier



Nom: metadata

Commentaire:

Schéma

Cliquez sur Détecter pour mettre à jour le schéma en fonction du paramétrage. Détecter

Description du schéma

Colonne	Clé	Type	<input checked="" type="checkbox"/>	N..	Modèle date (Ctrl+Es...	Length	Precision	Défaut	Commenta...
code	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>			1	0		
marque	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			7	0		

Création d'une métadonnée : format d'un fichier de marque.

A faire pour tous les fichiers.

# Manipulation de fichiers



**CGI**

La force de l'engagement<sup>MD</sup>

# Job 1 : Lecture des fichiers de données

## Objectif :

Lire les fichiers `data_brand.csv` et `data_products.csv` et les afficher dans la console d'exécution

Importer le contexte general

## Utilisation :

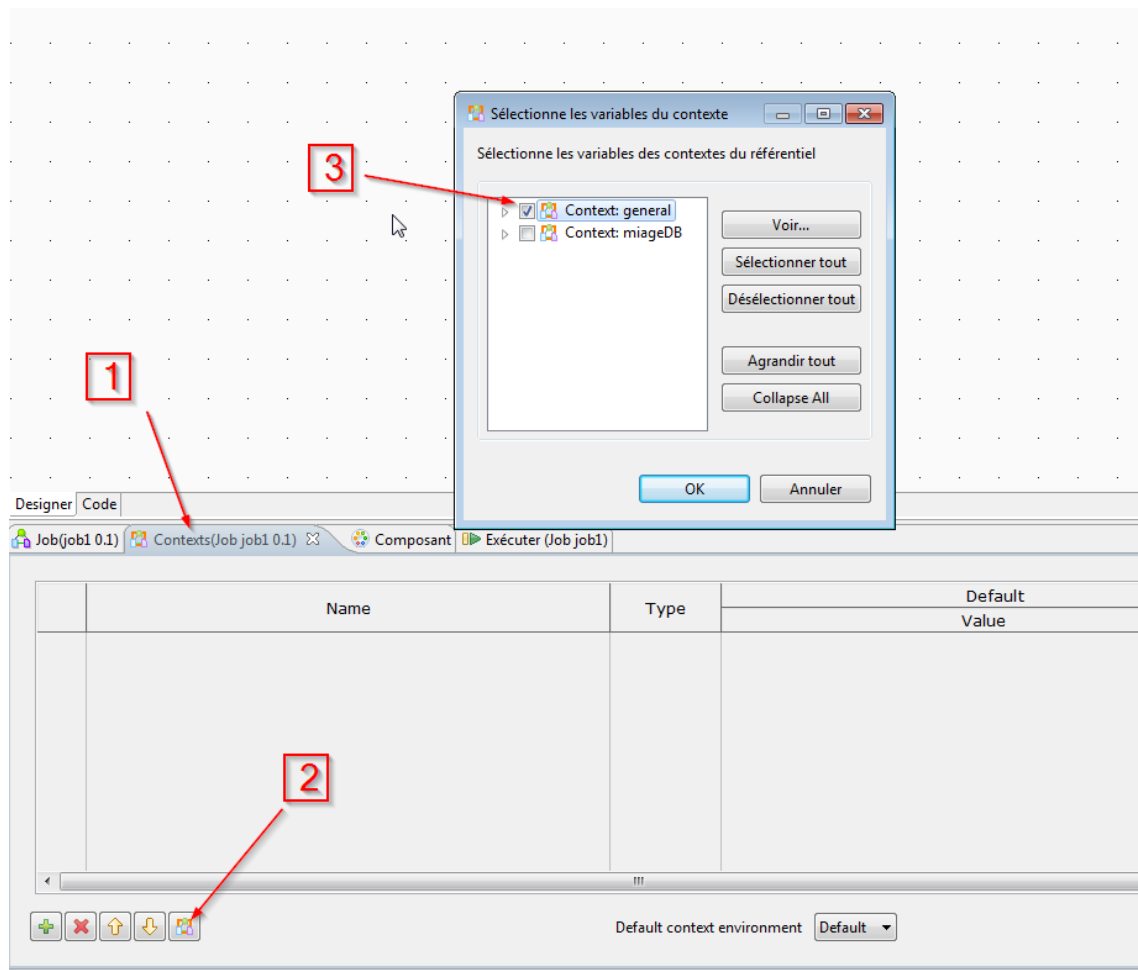
`tfileInputDelimited`

`tlogrow`





# Job 1 : Lecture des fichiers de données



# Job 1 : Lecture des fichiers de données

The screenshot displays a data processing job configuration interface. At the top, two parallel flows are shown on a grid background. The first flow, labeled 'row1 (Main)', connects a 'tFileInputDelimited\_1' component to a 'tLogRow\_1' component, with performance metrics of '11 rows in 0,01s' and '1100 rows/s'. The second flow, labeled 'row2 (Main)', connects a 'tFileInputDelimited\_2' component to a 'tLogRow\_2' component, with performance metrics of '3 rows in 0,01s' and '300 rows/s'. Below the flows, the 'Designer' tab is active, showing the configuration for 'tFileInputDelimited\_1'. The configuration includes a 'Paramètres simples' section with a 'Type de propriété' dropdown set to 'Built-In'. A warning message states: 'When the input source is a stream or a zip file, footer and random shouldn't be bigger than 0.' The 'Nom de fichier/Flux' is set to 'context.data\_directory + "/data\_products.csv"'. The 'Séparateur de lignes' is '\n' and the 'Séparateur de champs' is ','. The 'Options CSV' section includes 'En-tête' (0), 'Pied de page' (0), and 'Limite' (empty). The 'Schéma' is set to 'Référentiel' and 'Bases de données (MYSQL):miageDB - proc\*'. The 'Ignorer les lignes vides' checkbox is checked, while 'Décompresser en tant que fichier zip' and 'Arrêter en cas d'erreur' are unchecked.

Designer Code

Job(job1 0.1) Contexts(Job job1 0.1) Composant Exécuter (Job job1)

**tFileInputDelimited\_1**

Type de propriété Built-In

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

Nom de fichier/Flux context.data\_directory + "/data\_products.csv"

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

☐ Options CSV

En-tête 0 Pied de page 0 Limite

Schéma Référentiel Bases de données (MYSQL):miageDB - proc\* Editer le schéma

☒ Ignorer les lignes vides ☐ Décompresser en tant que fichier zip ☐ Arrêter en cas d'erreur

# Job 2 : Générer des données de vente

## **Objectif :**

Générer un fichier de données de vente ayant la forme :

Product\_id;qty

Avec 100 lignes

Importer le contexte general

## **Utilisation :**

trowgenerator

tfileOutputDelimited

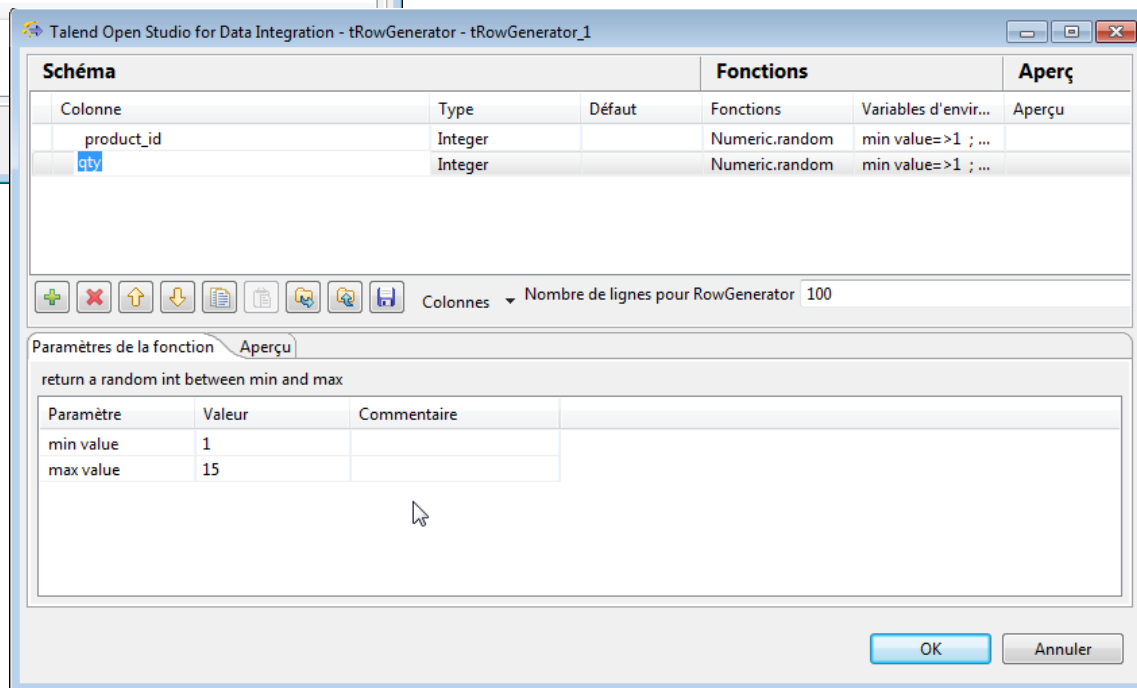
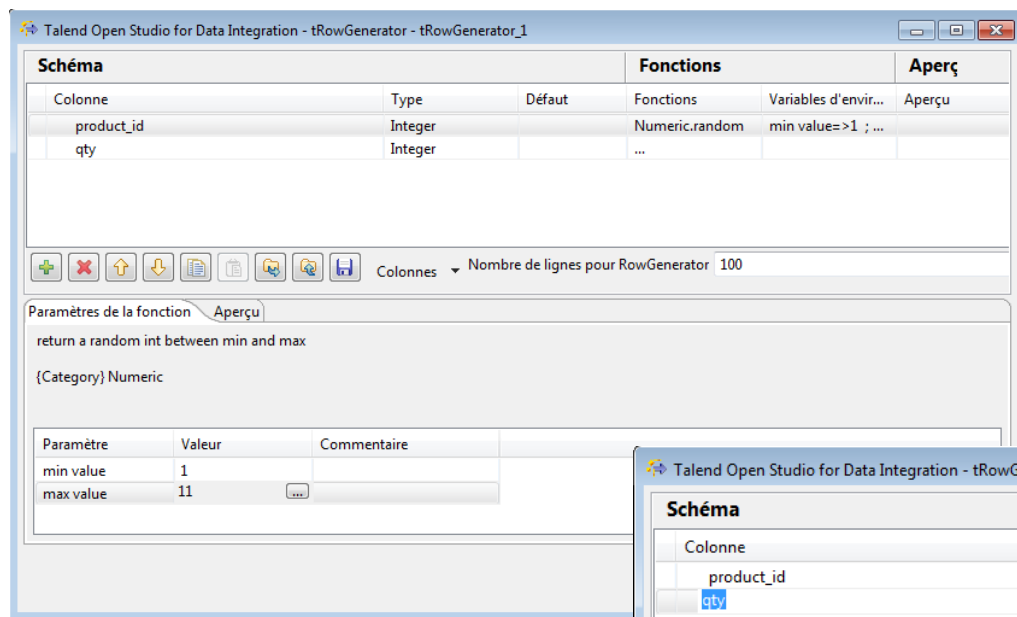


## Job 2 : Générer des données de vente

The screenshot shows the configuration window for 'tRowGenerator\_1' in a software interface. The window is titled 'Schéma detRowGenerator\_1' and contains a table with columns: Colonne, Clé, Type, N., Modèle d..., Len..., Prec..., Dé..., and Co... The table has two rows: 'product\_id' and 'qty', both with 'Inte...' as the type and checked 'N.' boxes. The 'Clé' column has checkboxes that are not checked. Below the table is a toolbar with icons for adding, deleting, and moving rows, as well as buttons for 'OK' and 'Annuler'. A red box labeled '1' points to the 'tRowGenerator\_1' component on the canvas, and a red box labeled '2' points to the 'Editer le schéma' button in the bottom toolbar.

Colonne	Clé	Type	<input checked="" type="checkbox"/> N.	Modèle d...	Len...	Prec...	Dé...	Co...
product_id	<input type="checkbox"/>	Inte...	<input checked="" type="checkbox"/>					
qty	<input type="checkbox"/>	Inte...	<input checked="" type="checkbox"/>					

# Job 2 : Générer des données de vente



# Job 2 : Générer des données de vente

The screenshot displays the Talend Studio interface for configuring a job. The main workspace shows a flow from the **tRowGenerator\_1** component to the **tFileOutputDelimited\_1** component. The flow is labeled with performance metrics: "100 rows in 0,01s", "10000 rows/s", and "row1 (Main)".

Below the workspace, the **Designer** tab is active, showing the configuration for the **tFileOutputDelimited\_1** component. The configuration includes:

- Paramètres simples**: Type de propriété: Built-In.
- Advanced settings**: ☐ Utiliser le flux de sortie.
- Paramètres dynamiques**: Nom de fichier: `context.data_directory + "/data_sales.csv"`.
- View**: Séparateur de lignes: `"\n"`; Séparateur de champs: `","`.
- Documentation**: ☐ Ecrire après; ☐ Inclure l'en-tête; ☐ Compresser en tant que fichier zip.
- Schéma**: Built-In; Boutons: Editer le schéma, Synchroniser les colonnes.



# Lecture – écriture en base



**CGI**

La force de l'engagement<sup>MD</sup>



# Job 3 : Insertion des données en base et Lecture

## Objectif :

Lire les fichiers data\_brand.csv et data\_products.csv et les insérer dans les tables products et brand

Puis lire les données et les afficher dans la console

Importer le contexte general et miageDB

## Utilisation :

tfileInputDelimited

tMysqlOutput

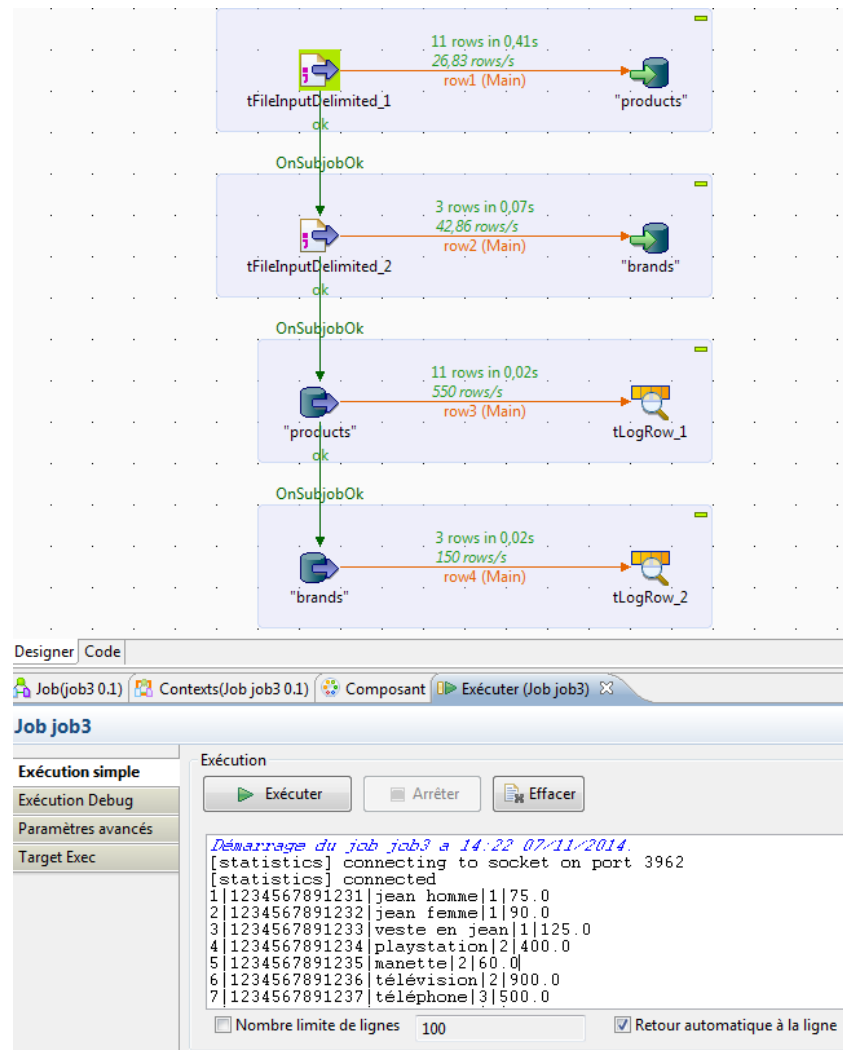


# Job 3 : Insertion des données en base et Lecture

The image illustrates the configuration of a data job in a tool, showing three sequential steps:

- Step 1:** In the left-hand tree view, the user navigates to **Métadonnées** > **Connexions aux bases de données** > **miageDB 0.1** > **Schémas des tables**, where the **brands** table is selected.
- Step 2:** In the **Composants** window, the user selects the **tMysqlOutput** component from the list of available components.
- Step 3:** In the **Contexte ajouté** dialog, the user confirms the addition of context parameters for the **miageDB** variable. The parameters listed are: **miageDB\_Database**, **miageDB\_Login**, **miageDB\_Port**, **miageDB\_AdditionalParams**, **miageDB\_Server**, and **miageDB\_Password**. The **Oui** (Yes) button is highlighted.

# Job 3 : Insertion des données en base et Lecture



# tMap



# CGI

La force de l'engagement<sup>MD</sup>

# Job 4 : Jointure données de marques et produits

## Objectif :

Lire les tables products et brand, et compléter le fichier produits avec le libellé des marques : affichage dans la console et dans un fichier complémentaire.

## Utilisation :

tfileInputDelimited ou tMySQLInput

tMap

tfileOutputDelimited.



# tMap et tagregate



**CGI**

La force de l'engagement<sup>MD</sup>

# Job 5 : Jointure données de vente, marques et produits

## Objectif :

Lire les tables products et brand, lire le fichier data\_sales.csv

Joindre les données, agréger au niveau produit, alimenter la table PRODUCTS\_WITH\_BRAND pour les produits avec marque connue et la table PRODUCTS\_NO\_BRAND pour les produits dont la marque est inconnue.

## Flux sortie intermédiaire :

Marque / nom du produit / vente = prix \* nb de produit.

## Utilisation :

tfileInputDelimited

tMysqlInput

tMap

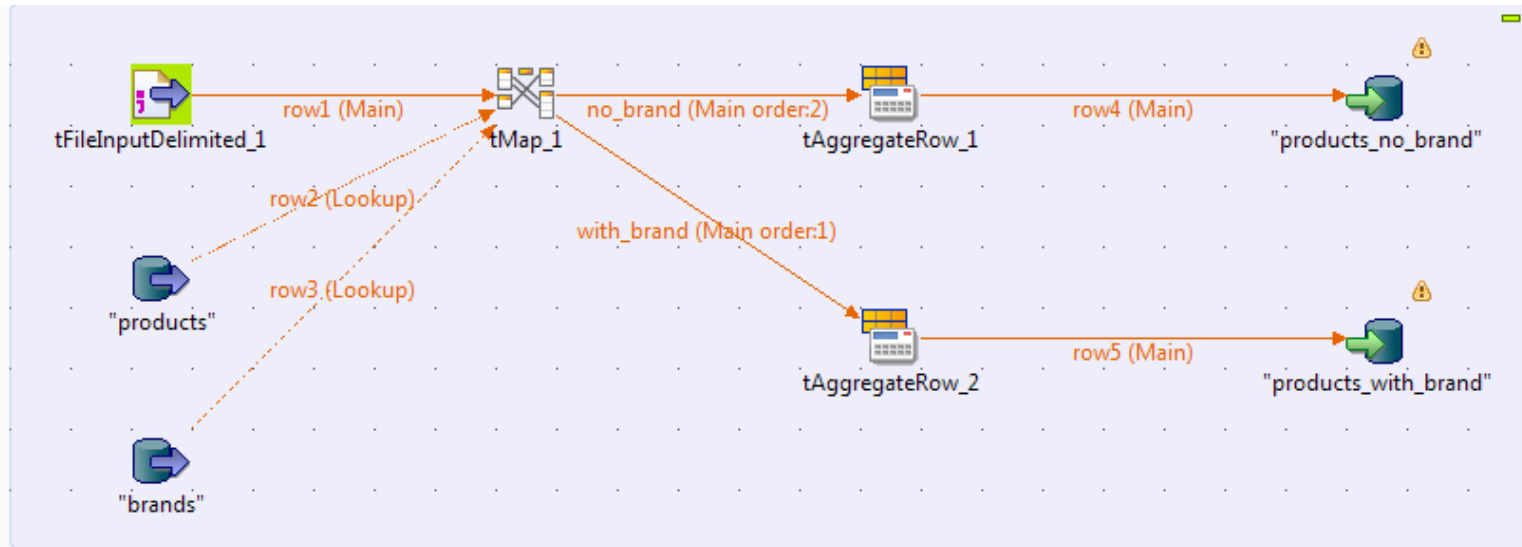
taggregate

tMysqlOutput





# Job 5 : Jointure données de vente, marques et produits



# Job 5 : Jointure données de vente, marques et produits

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

Find :

Var

Auto map !

**row1**

Column

product\_id  
qty

**row2**

Clé d'expr. Column

row1.product\_id ROWID  
EAN  
NAME  
BRAND\_ID  
UNIT\_PRICE

**row3**

Clé d'expr. Column

row2.BRAND\_ID ROWID  
NAME

**out1**

!Relational.ISNULL(row3.NAME)

Expression Column

row2.ROWID ROWID  
row2.EAN EAN  
row2.NAME NAME  
row2.BRAND\_ID BRAND\_ID  
row2.UNIT\_PRICE UNIT\_PRICE  
row3.NAME NAME\_1  
row2.UNIT\_PRICE \* row1.qty AMOUNT

**out2**

!Relational.ISNULL(row3.NAME)

Expression Column

row2.ROWID ROWID  
row2.EAN EAN  
row2.NAME NAME  
row2.BRAND\_ID BRAND\_ID  
row2.UNIT\_PRICE UNIT\_PRICE  
row3.NAME NAME\_1  
row2.UNIT\_PRICE \* row1.qty AMOUNT

Éditeur de Schéma Éditeur d'expression

**row2**

Colonne	Clé	Type	✓	N.	Modèle date (Ctrl+E...	Length	Precision	Défaut	Comment...
ROWID	<input checked="" type="checkbox"/>	int	<input checked="" type="checkbox"/>	<input type="checkbox"/>		10	0		
EAN	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>		20	0		
NAME	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>		255	0		Nom du p...
BRAND_ID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>	<input type="checkbox"/>		10	0		Marque
UNIT_PRICE	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>	<input type="checkbox"/>		8	2		Marque

**out2**

Colonne	Clé	Type	✓	N.	Modèle date (Ctrl+E...	Length	Precision	Défaut	Comment...
ROWID	<input checked="" type="checkbox"/>	int	<input checked="" type="checkbox"/>	<input type="checkbox"/>		10	0		
EAN	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>		20	0		
NAME	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>		255	0		Nom du p...
BRAND_ID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>	<input type="checkbox"/>		10	0		Marque
UNIT_PRICE	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>	<input type="checkbox"/>		8	2		Marque
NAME_1	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input type="checkbox"/>		255	0		
AMOUT	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>	<input type="checkbox"/>					

# Merci !



# CGI

La force de l'engagement<sup>MD</sup>