



Formation Talend

Caroline Cardon-Weber
Mahmed AIT MOUHOU
04/11/2015

© Groupe CGI inc. CONFIDENTIEL

CGI

La force de l'engagement^{MD}

Sommaire

- ❑ Installation de l'environnement
- ❑ Présentation de Talend DI
- ❑ Les principaux concepts
- ❑ Manipulation de fichiers
- ❑ Lecture – écriture en base
- ❑ Le Tmap

Installation de l'environnement



CGI

La force de l'engagement^{MD}

Installation

- Mettre server.zip dans C: ou D:
- Faire clic droit et extraire ici
- Démarrer c:\server\nginx-start54.cmd
- Pour accéder à la base de donnée : <http://localhost/adminer>
- Login : root / Pass : password
- Copier coller TOS_DI-20141024_1545-V5.6.0.zip dans C: ou D:
- Faire clic droit et extraire ici



Installation

Sélectionner la base de données

MySQL » 127.0.0.1

[Créer une base de données](#) [Privilèges](#) [Liste des processus](#) [Variables](#) [Statut](#)

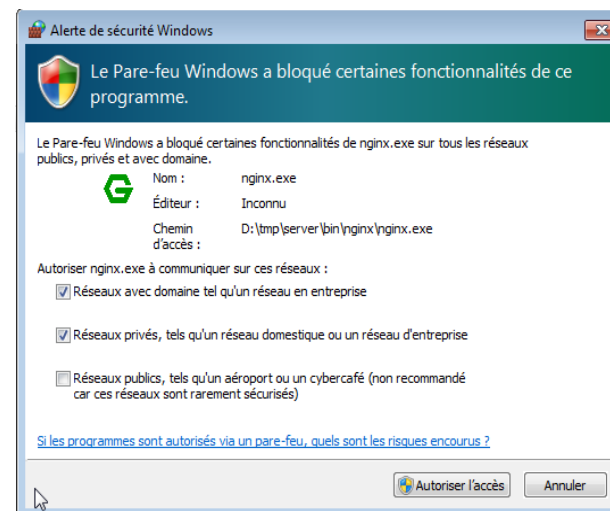
Version de MySQL : **5.5.5-10.0.13-MariaDB** via l'extension PHP **MySQLi**

Authentifié en tant que : **root@localhost**

	Base de données - Rafraîchir	Interclassement	Tables	Taille - Calcul
<input type="checkbox"/>	information_schema	utf8_general_ci	?	?
<input type="checkbox"/>	mysql	latin1_swedish_ci	?	?
<input type="checkbox"/>	performance_schema	utf8_general_ci	?	?

Sélectionnée(s) (0)

Supprimer



Créer une base de données

MySQL » 127.0.0.1 » Créer une base de données

miage utf8_general_ci



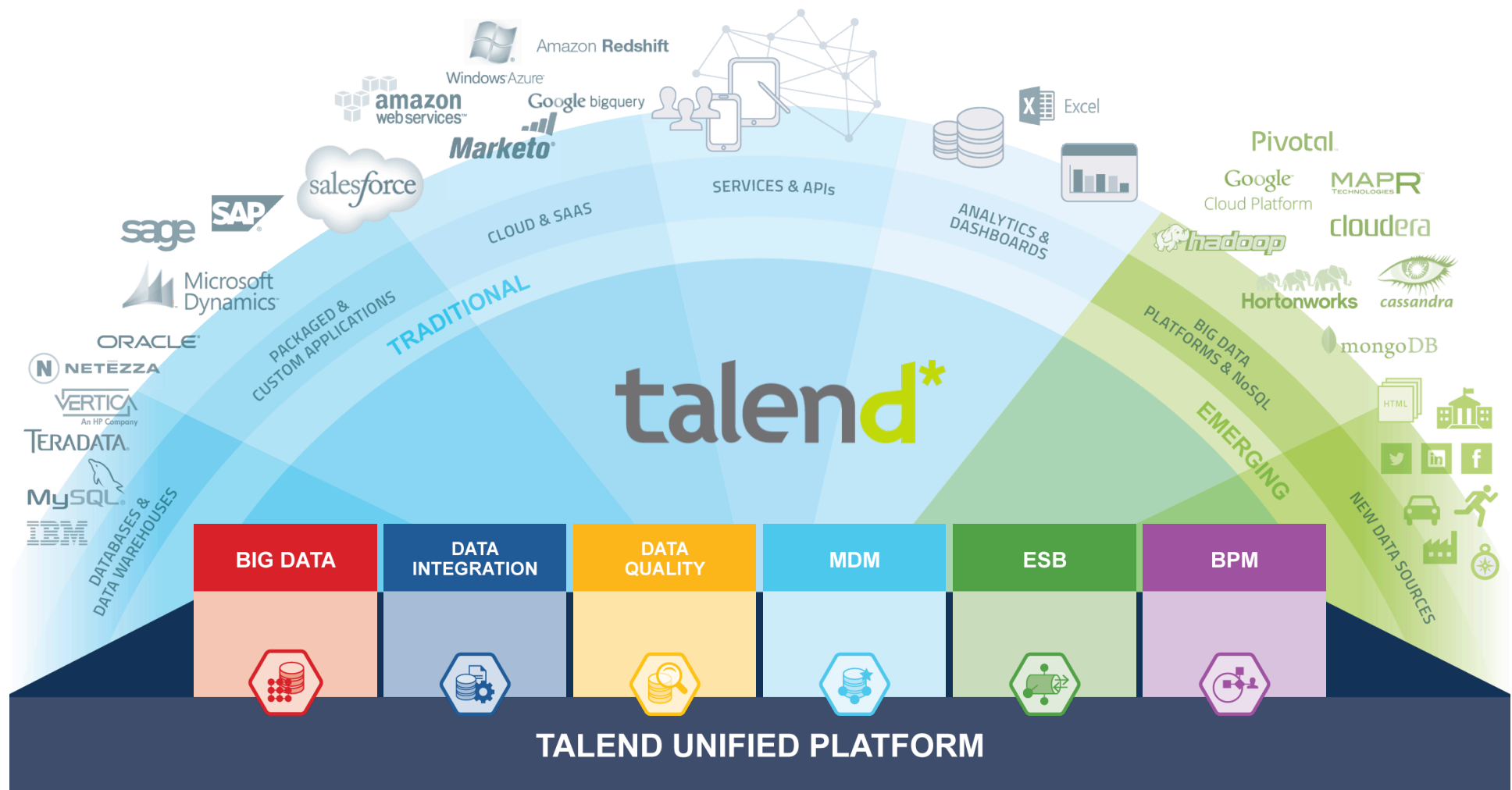
Talend Data Intégration



CGI

La force de l'engagement^{MD}

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

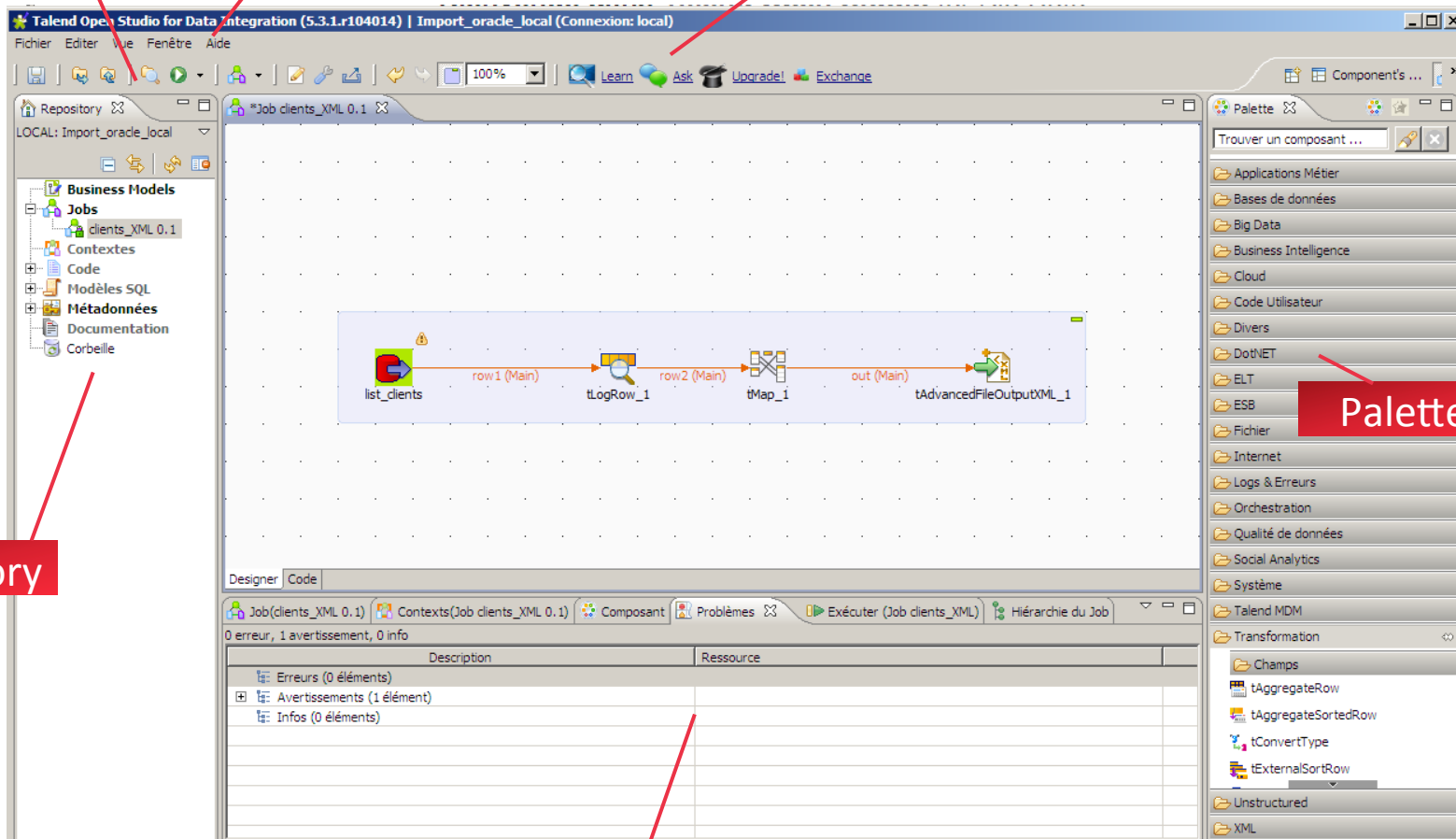
Menu

Espace de modélisation graphique

Repository

Palette

Onglets de configuration



ETL vs ELT



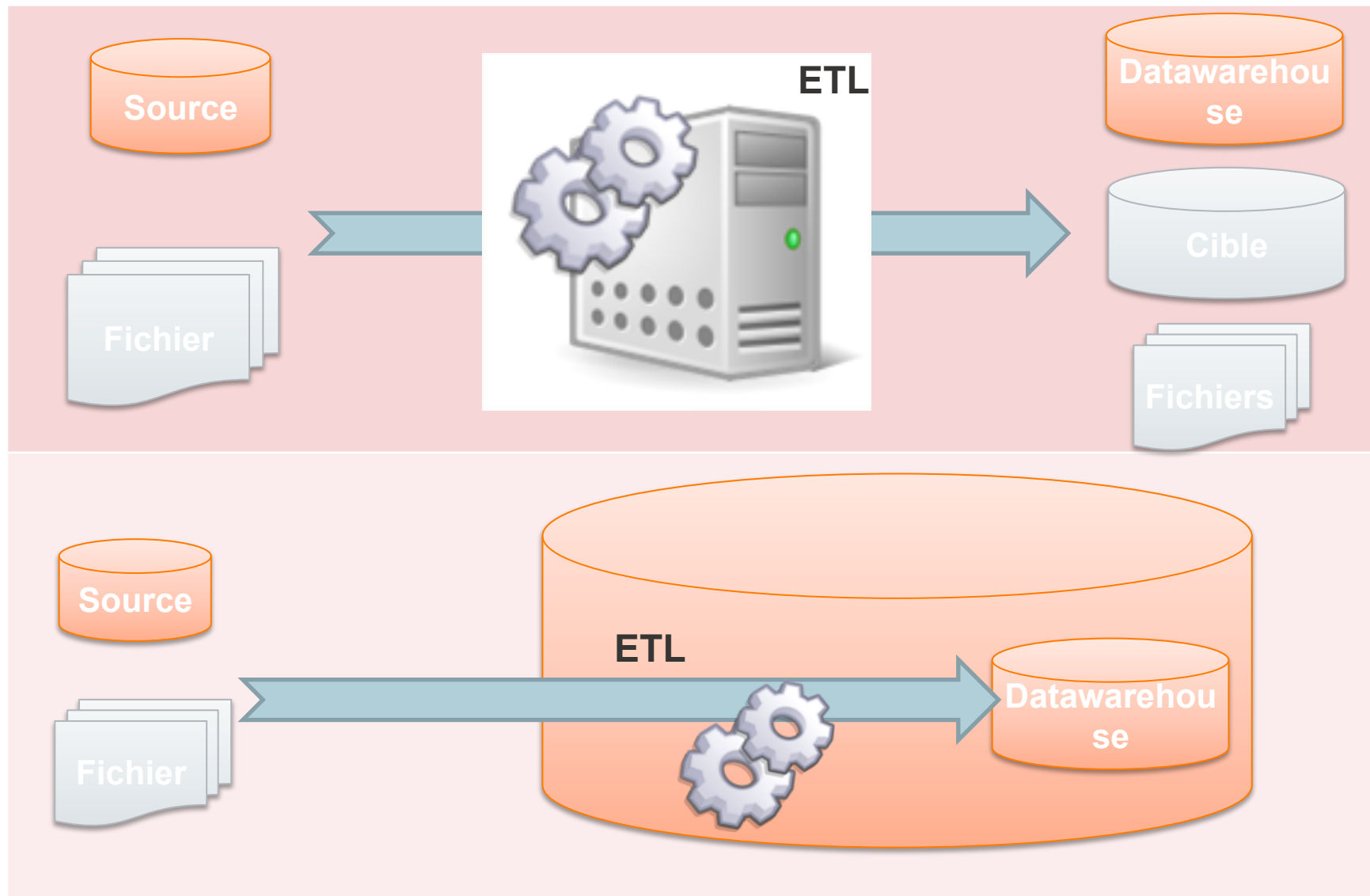
CGI

La force de l'engagement^{MD}

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



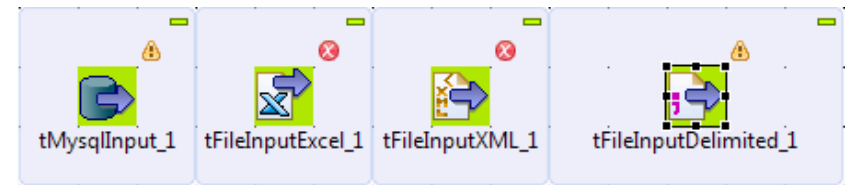
CGI

La force de l'engagement^{MD}

Les principaux concept

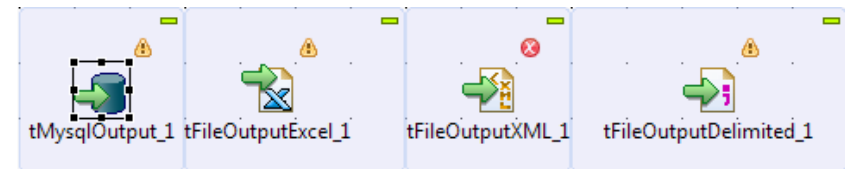
- **Les composants d'entrée :**

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



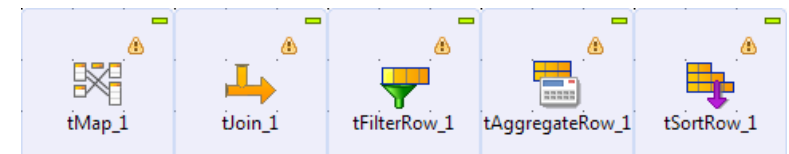
- **Les composants de sortie**

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, ...

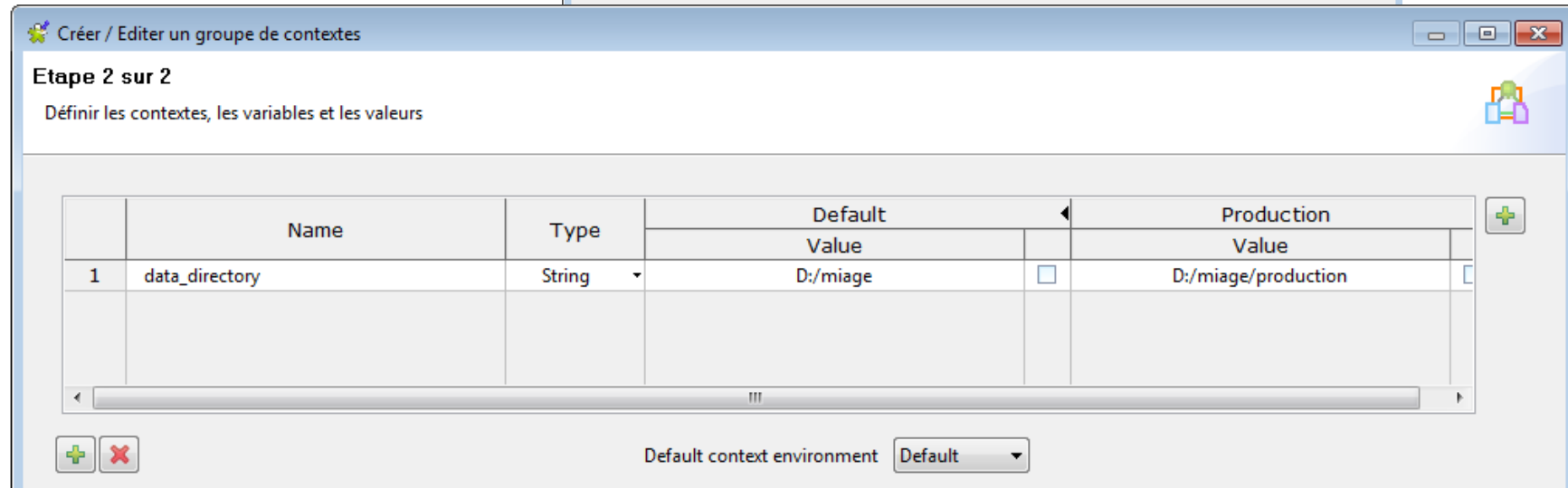
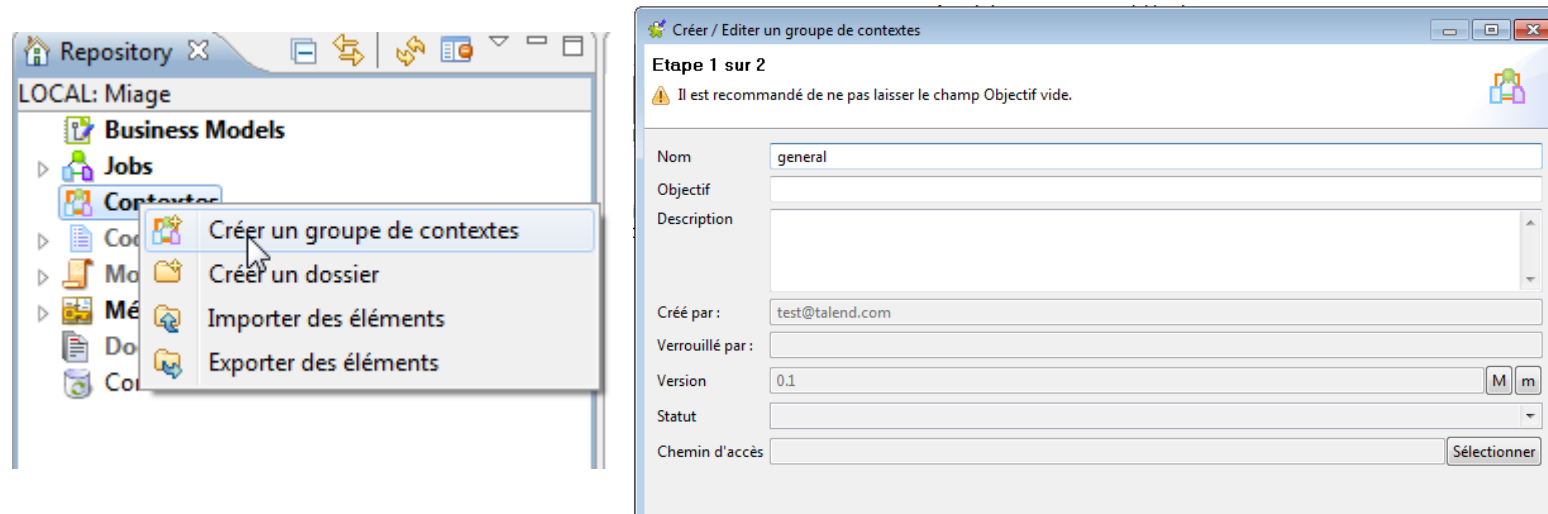


- **Les contextes**

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



Création d'un contexte



Création d'un métadonnée

Business Models

- Jobs
- Contextes
 - general 0.1
- Code
- Modèles SQL
- Métadonnées
 - Connex
 - Créer une connexion
 - Créer un dossier
 - Importer des éléments
 - Exporter des éléments
 - Fichier
 - Fichier
 - Fichier
 - Fichier
 - Fichier
 - Fichier Excel

Connexion à la base de données

Nouvelle connexion à une base de données dans le référentiel - Etape1/2

Il est recommandé de ne pas laisser le champ Objectif vide.

Nom: miageDB

Objectif:

Description:

Créé par: test@talend.com

Verrouillé par:

Version: 0.1

Statut:

Chemin d'accès: Sélectionner

< Retour Suivant > Terminer Annuler

Connexion à la base de données

Nouvelle connexion à une base de données dans le référentiel - Etape2/2

Définir les paramètres de connexion

DB Type: MySQL

Version de la base de données: MySQL 5

Chaine de caractères de connexion: jdbc:mysql://localhost:3306/miage?noDatetimeStringSync=true

Login: root

Mot de passe:

Server: localhost

Port: 3306

DataBase: miage

Paramètres supplémentaires: noDatetimeStringSync=true

Vérifier

Propriétés de la base de données

Syntaxe SQL: SQL 92

Séparateur de chaîne de caractères: |

Caractère Null: 000

Exporter en tant que contexte

Revenir au contexte précédent

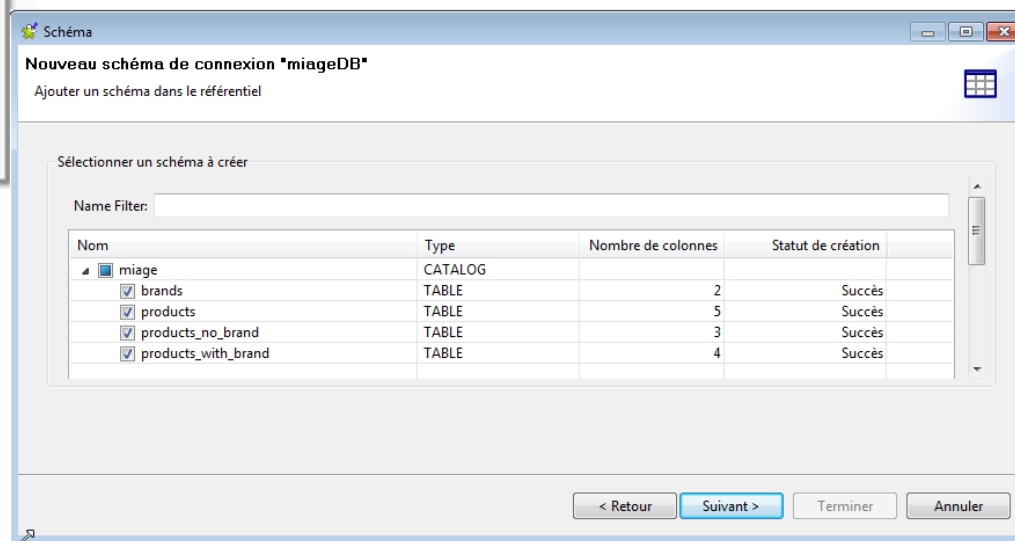
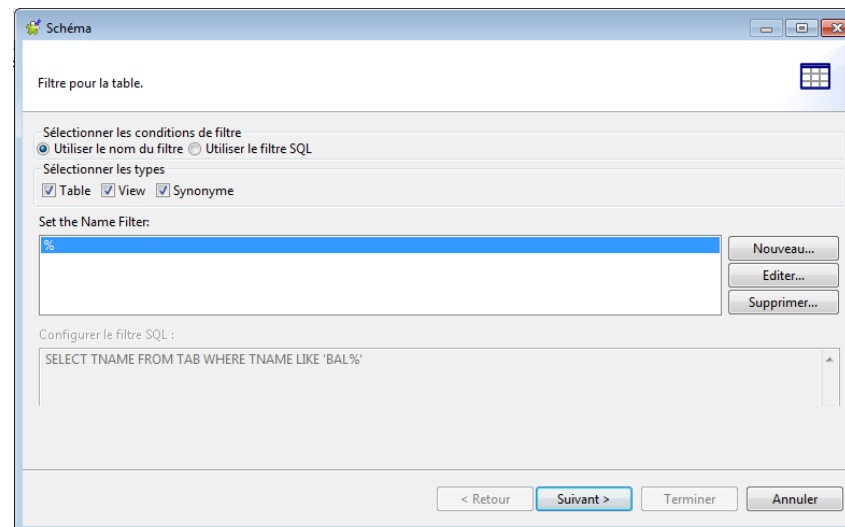
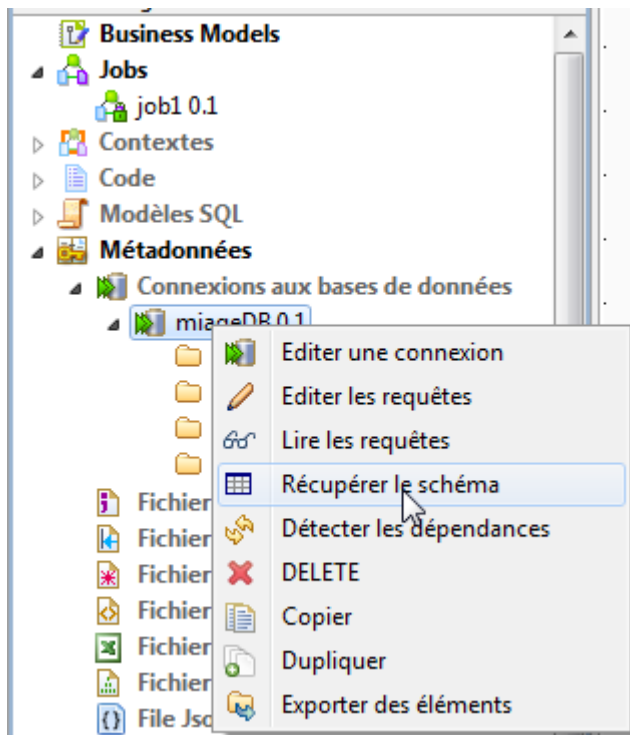
Comment installer un pilote

< Retour Suivant > Terminer 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^{MD}

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

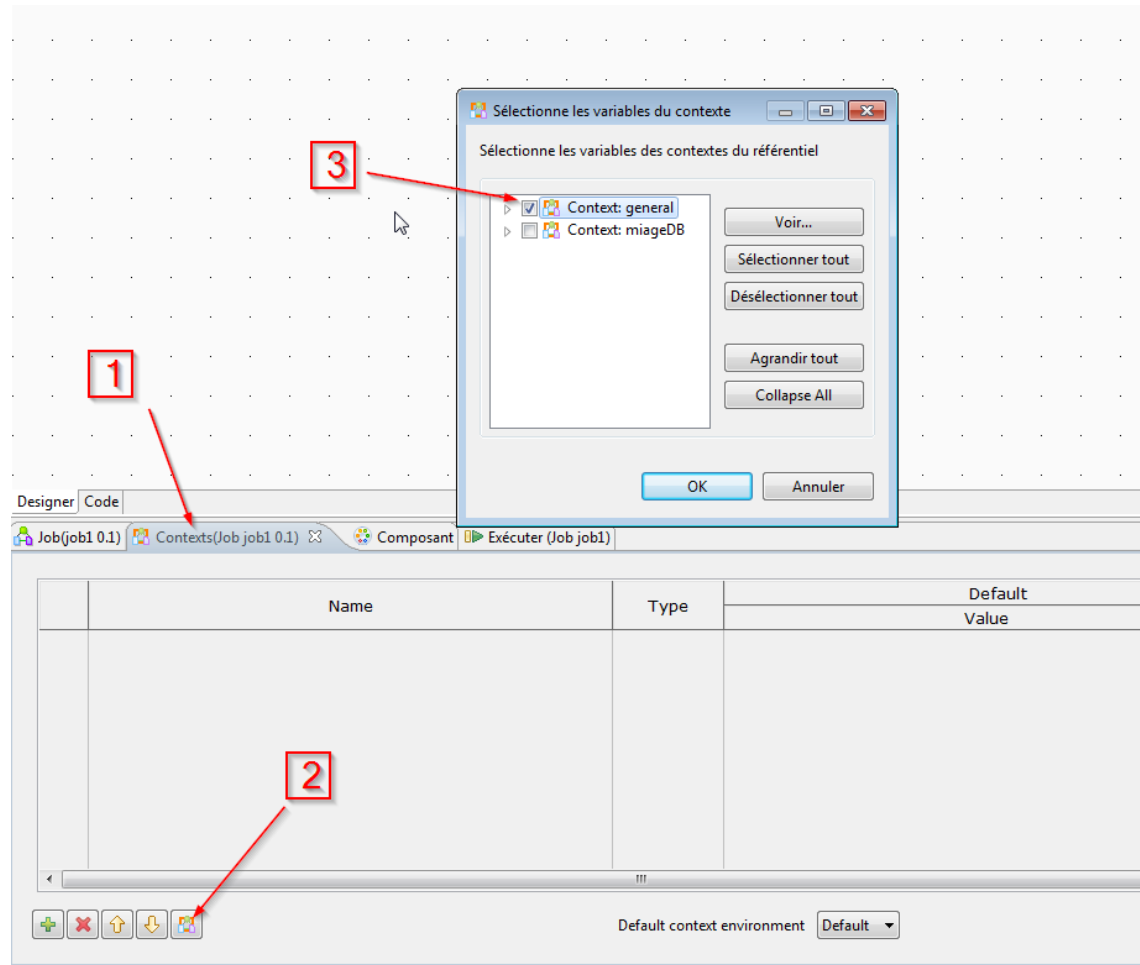
Utilisation :

tfileInputDelimited

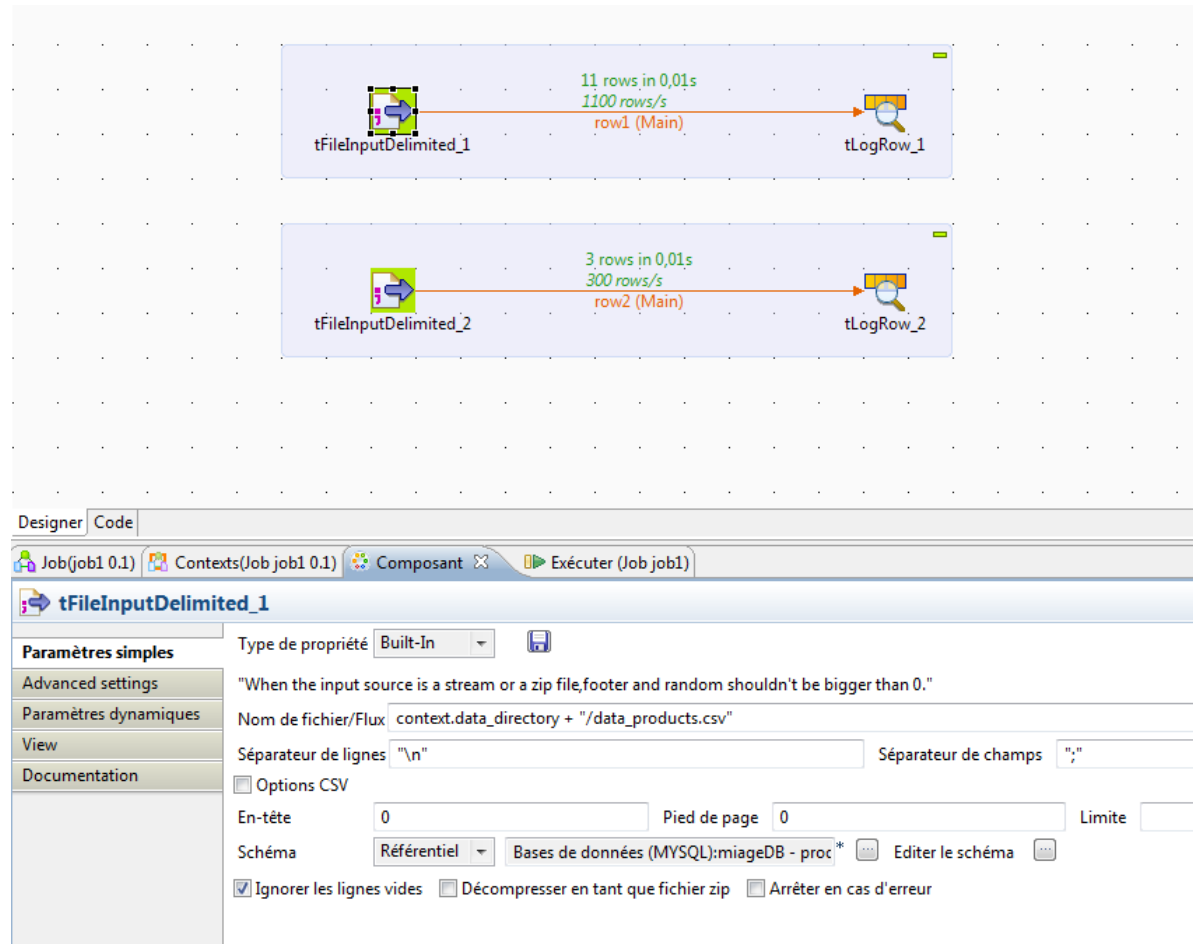
tlogrow



Job 1 : Lecture des fichiers de données



Job 1 : Lecture des fichiers de données



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

1

2

Schéma detRowGenerator_1

tRowGenerator_1

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"/>					

OK Annuler

Designer Code

Job(job2 0.1) Contexts(Job job2 0.1) Composant Exécuter (Job job2)

tRowGenerator_1

Paramètres simples Schéma Built-In Editer le schéma

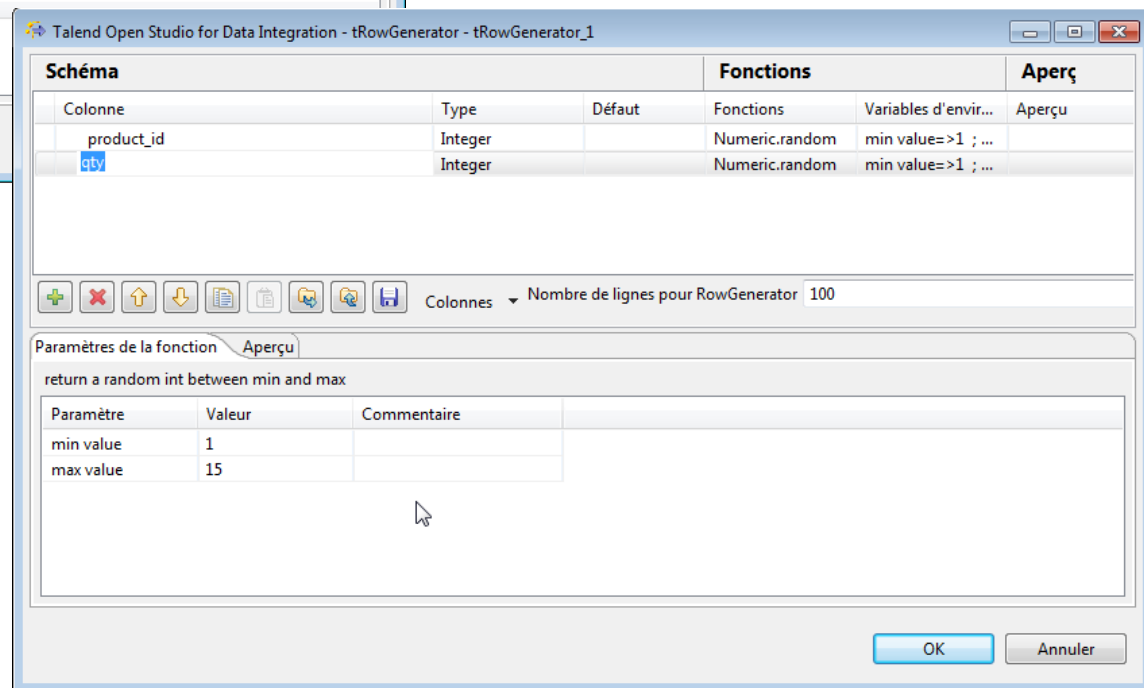
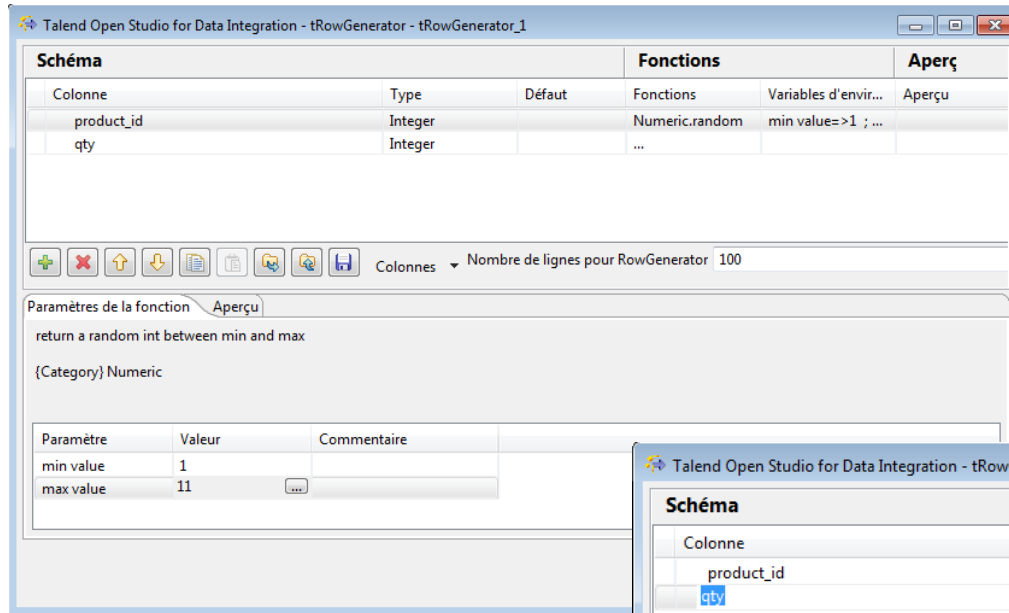
Advanced settings Editeur RowGenerator

Paramètres dynamiques

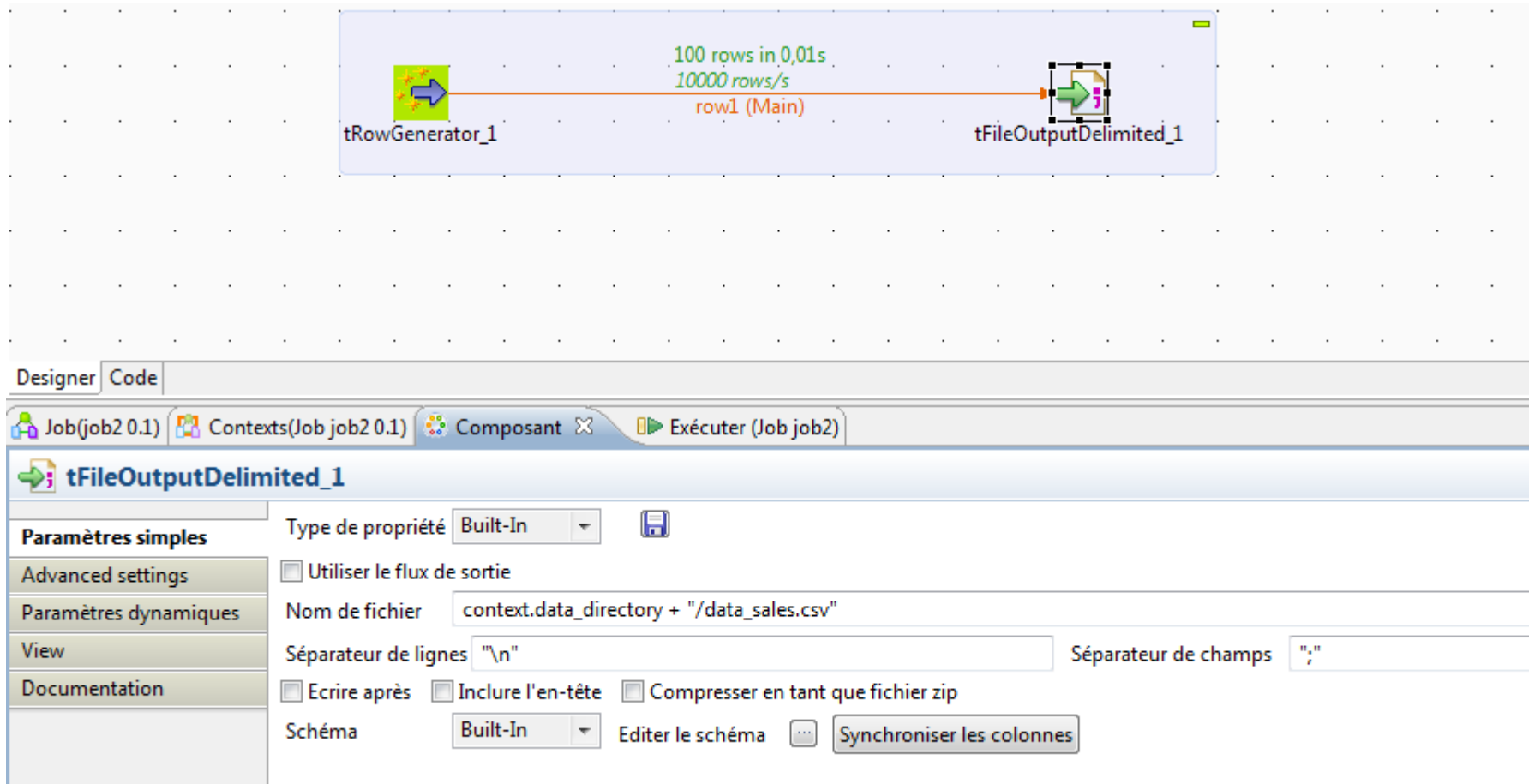
View

Documentation

Job 2 : Générer des données de vente



Job 2 : Générer des données de vente



The screenshot displays the Talend Designer interface for Job 2. The main workspace shows a data 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)".

The bottom panel shows the configuration for the **tFileOutputDelimited_1** component. The configuration is organized into sections:

- 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, Editer le schéma, Synchroniser les colonnes.



Lecture – écriture en base



CGI

La force de l'engagement^{MD}

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

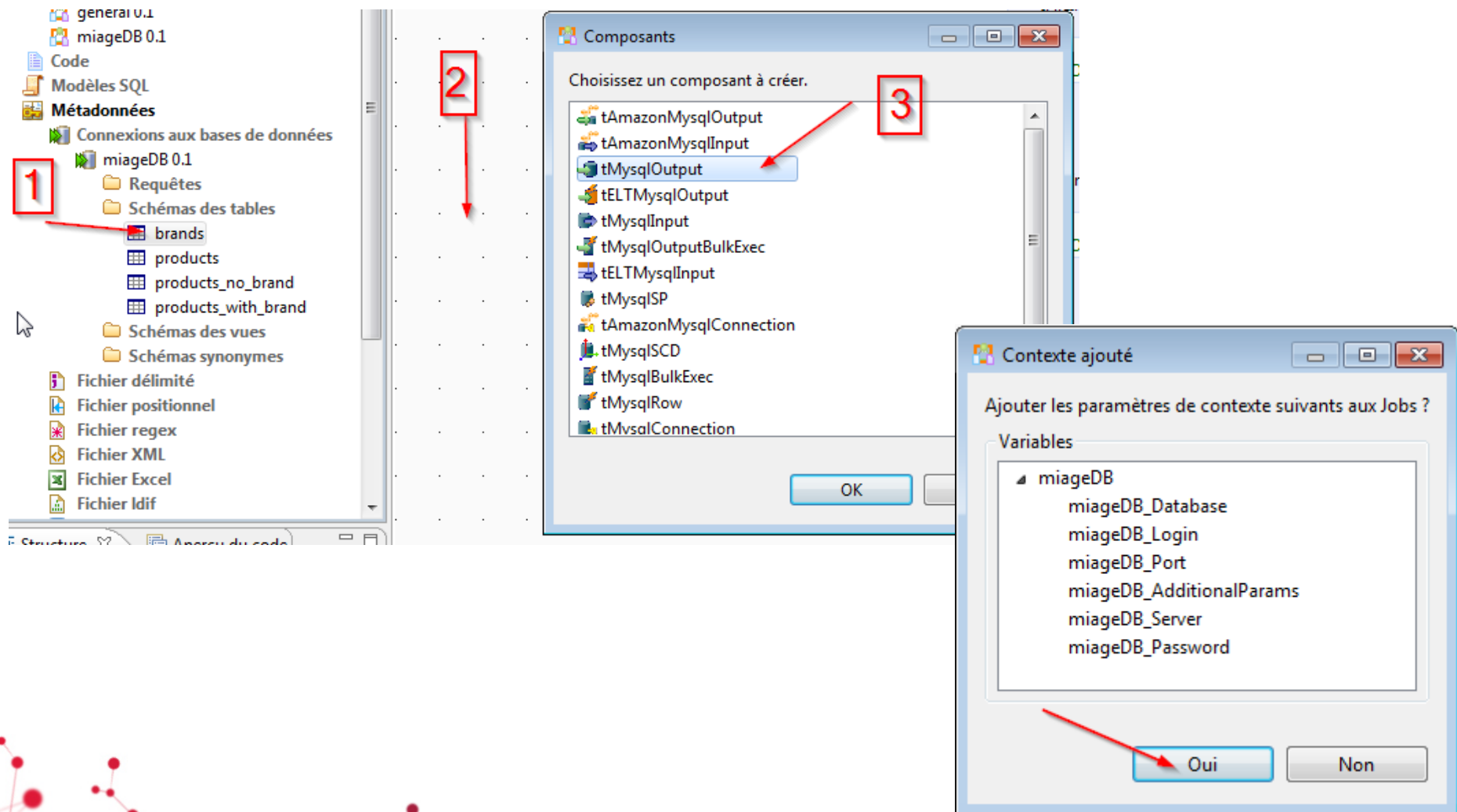
Utilisation :

tfileInputDelimited

tMysqlOutput



Job 3 : Insertion des données en base et Lecture

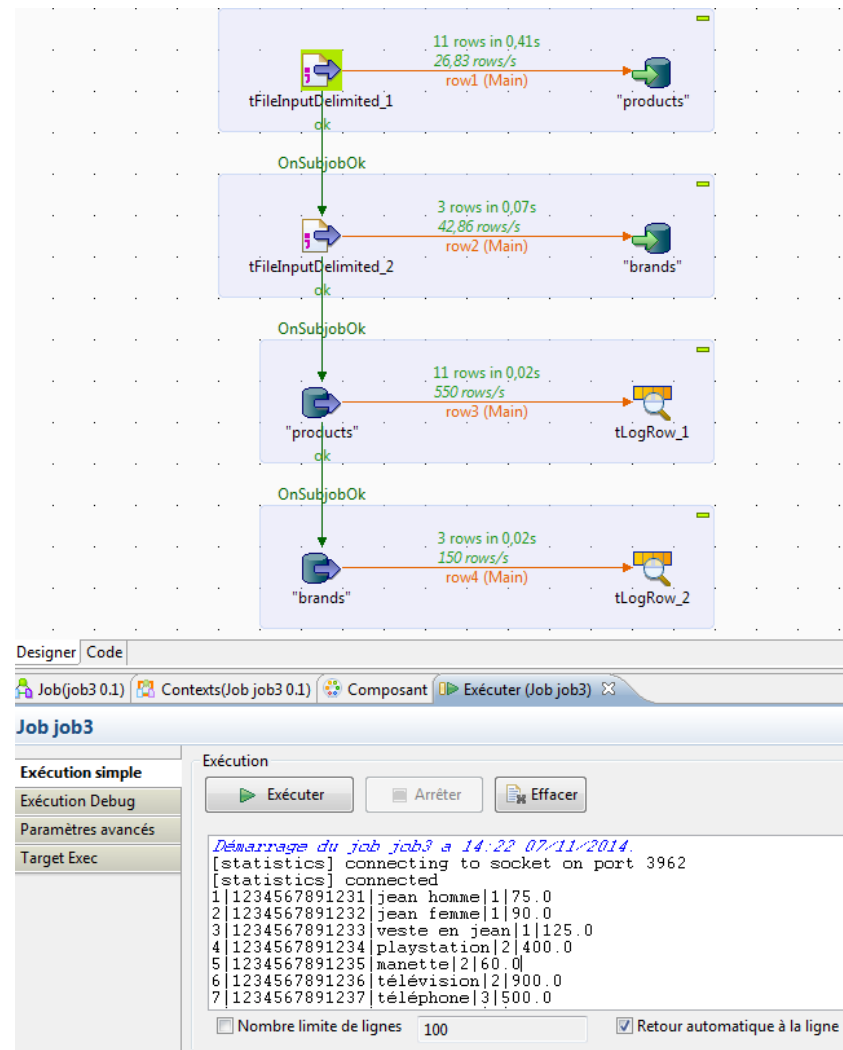


The image illustrates the configuration of a data job in a tool, showing three sequential steps:

- Step 1:** Selecting a data source table. In the 'Métadonnées' pane, the 'brands' table is selected under the 'miageDB 0.1' connection.
- Step 2:** Choosing a component to create. In the 'Composants' pane, 'tMysqlOutput' is selected from the list of available components.
- Step 3:** Adding context parameters. The 'Contexte ajouté' dialog box is shown, asking to add the following variables to the job:
 - miageDB
 - miageDB_Database
 - miageDB_Login
 - miageDB_Port
 - miageDB_AdditionalParams
 - miageDB_Server
 - miageDB_PasswordThe 'Oui' (Yes) button is highlighted.



Job 3 : Insertion des données en base et Lecture



tMap



CGI

La force de l'engagement^{MD}

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^{MD}

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 :

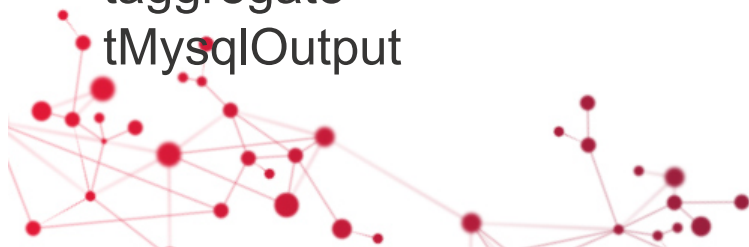
tfileInputDelimited

tMysqlInput

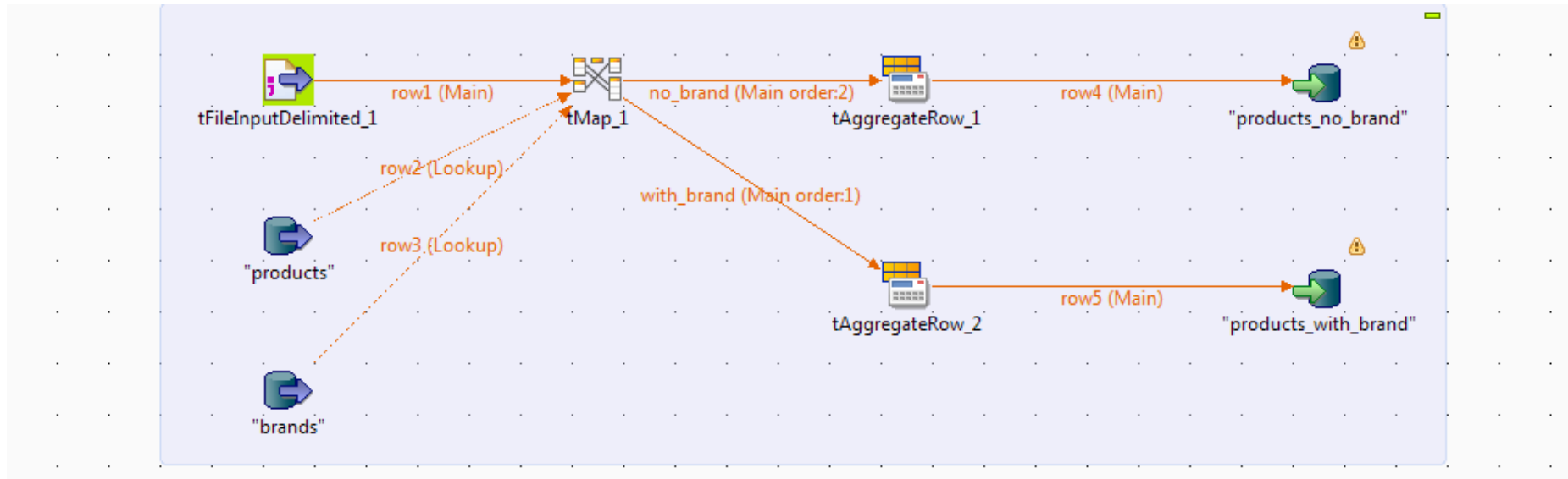
tMap

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

row1

Column

product_id

qty

row2

Clé d'expr.

row1.product_id

Column

ROWID

EAN

NAME

BRAND_ID

UNIT_PRICE

row3

Clé d'expr.

row2.BRAND_ID

Column

ROWID

NAME

out1

!Relational.ISNULL(row3.NAME)

Expression

row2.ROWID

row2.EAN

row2.NAME

row2.BRAND_ID

row2.UNIT_PRICE

row3.NAME

row2.UNIT_PRICE * row1.qty

Column

ROWID

EAN

NAME

BRAND_ID

UNIT_PRICE

NAME_1

AMOUNT

out2

Relational.ISNULL(row3.NAME)

Expression

row2.ROWID

row2.EAN

row2.NAME

row2.BRAND_ID

row2.UNIT_PRICE

row3.NAME

row2.UNIT_PRICE * row1.qty

Column

ROWID

EAN

NAME

BRAND_ID

UNIT_PRICE

NAME_1

AMOUNT

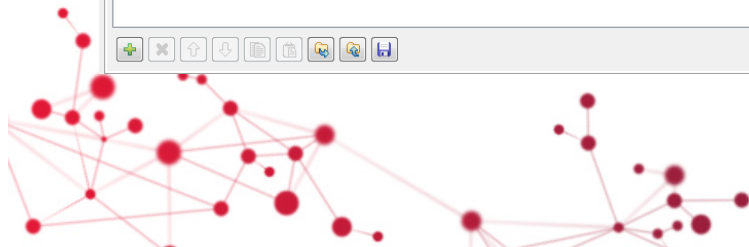
Éditeur de Schéma Éditeur d'expression

row2

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



Merci !



CGI

La force de l'engagement^{MD}