

TALEND DATA INTEGRATION





Agenda

- ❑ Introduction à Talend Data integration
- ❑ **Prise en main de Talend Data integration**
- ❑ Utilisation avancée de Talend Data intégration



2.

PRISE EN MAIN

DE TALEND

OPEN STUDIO FOR DATA INTEGRATION



Objectifs

- ❑ Avantages de Talend pour Data Integration
- ❑ Les principaux composants de Talend DI
- ❑ Groupes de contextes
- ❑ Gestions des erreurs
- ❑ Les paramètres de contexte
- ❑ Ecrire et lire des fichiers
- ❑ Consolider plusieurs fichiers dans un seul
- ❑ Nettoyer les fichiers sources



Talend et l'intégration des données

- ❑ La collecte et l'intégration des données est une phase nécessaire dans un projet BI
- ❑ La présence d'un outil ETL dans un projet BI/Big Data est devenue obligatoire
- ❑ Talend est devenu le leader des ETL pour le BI/Big Data



Talend et l'intégration des données

Les points forts de Talend Data Integration sont :

- ❑ Génération automatique de programmes BI/Big Data: Java, Spark
- ❑ Facilite la vie d'un développeur BI/Big Data:
interface graphique offre plusieurs composants (plus que 900) qui répondent aux besoins
- ❑ Rapidité de compilation et d'exécution : Talend génère du java
- ❑ Produit abordable et pas chère sur le marché : En plus des produits payants, Talend offre le TOS For Data Integration



Avantages du TOS For Data Integration

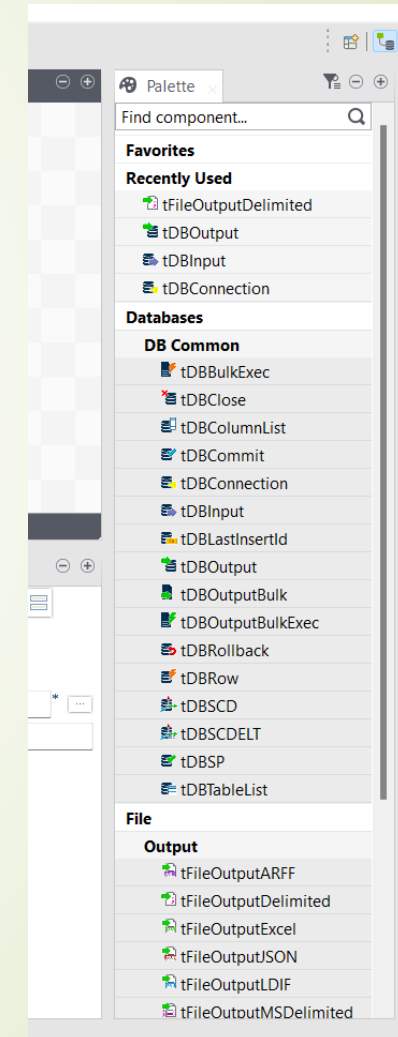
- ❑ Conception plus rapide : un environnement de développement graphique simple et facile à utiliser
- ❑ Meilleure collaboration : communauté active pour améliorer et créer les composants BI
- ❑ Portabilité : le TOS peut être utilisé sur n'importe quel OS
- ❑ Scalabilité : l'évolution des jobs Talend est facile.
- ❑ Personnalisable : un développeur peut ajouter des nouveaux composants

Les composants du TOS For Data Integration



Les composants du TOS For Data Integration

- ❑ Une pièce fonctionnelle permettant d'effectuer une seule opération
- ❑ Les composants sont classifiés par besoin fonctionnel ou technique dans la palette
- ❑ C'est un extrait de code java généré lors de l'exécution d'un job
- ❑ Talend fournit plus de 900 composants



Les composants du TOS For Data Integration

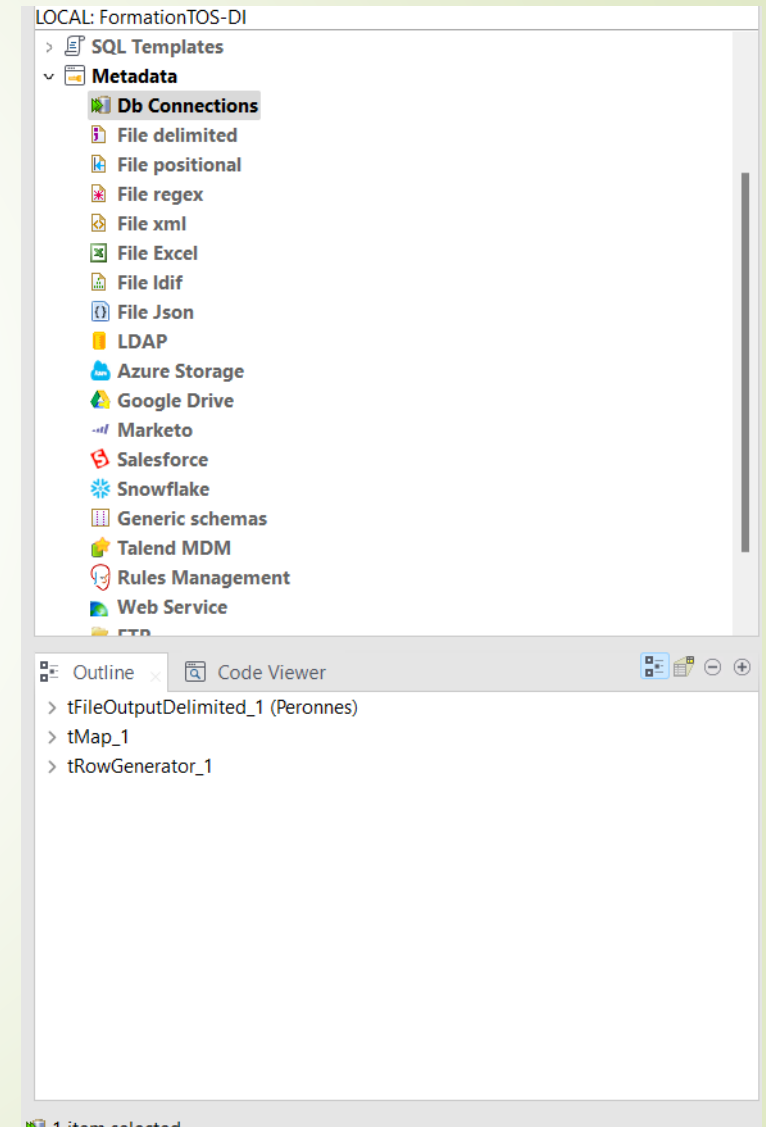
Type de composants	Description	Exemples
Connexion	<ul style="list-style-type: none">✓ Configure et initialise une connexion à un environnement ou à une base de données.✓ Pas de connecteurs d'entrée ou sortie	TDBConnexion TDBCclose
Flux entrée	<ul style="list-style-type: none">✓ Lire à partir d'un fichier ou une table et met à disposition des lignes des données pour le composant qui succède✓ Pas de connecteur d'entrée	TfileinputDelimited TDBinput
Flux en Sortie	<ul style="list-style-type: none">✓ Ecrire le flux dans un fichier ou une table✓ Pas de connecteur de sortie	TfileoutputDelimited TDBoutput
Traitement des flux	<ul style="list-style-type: none">✓ Transformation, jointure, agrégation, tri, filtrage ...✓ Connecteurs en entrée et en sortie	tMap, tSortRow, tAggregateRow ...

Les composants du TOS For Data Integration

Type de composants	Description	Exemples
Traitement des fichiers	✓ Lister, copier, supprimer, renommer des fichiers • Existent pour HDFS et Local	tHDFSList, tFileList, tFileCopy, tHDFSCopy ...
Composants Java	✓ Composants personnalisables en écrivant du code java	Java, tJavaRow, tJavaFlex
Orchestration	✓ Orchestration des composants dans un job Talend : lancement d'un job, iteration ...	TfileoutputDelimited TDBoutput

Déclaration des Métadonnées

- ✓ Talend fournit des connecteurs permettant d'extraire les métadonnées des fichiers, base de données ...
- ✓ Ces métadonnées sont mise à disposition au développeur dans l'onglet référentiel projet
- ✓ Les métadonnées peuvent être utilisés par plusieurs jobs

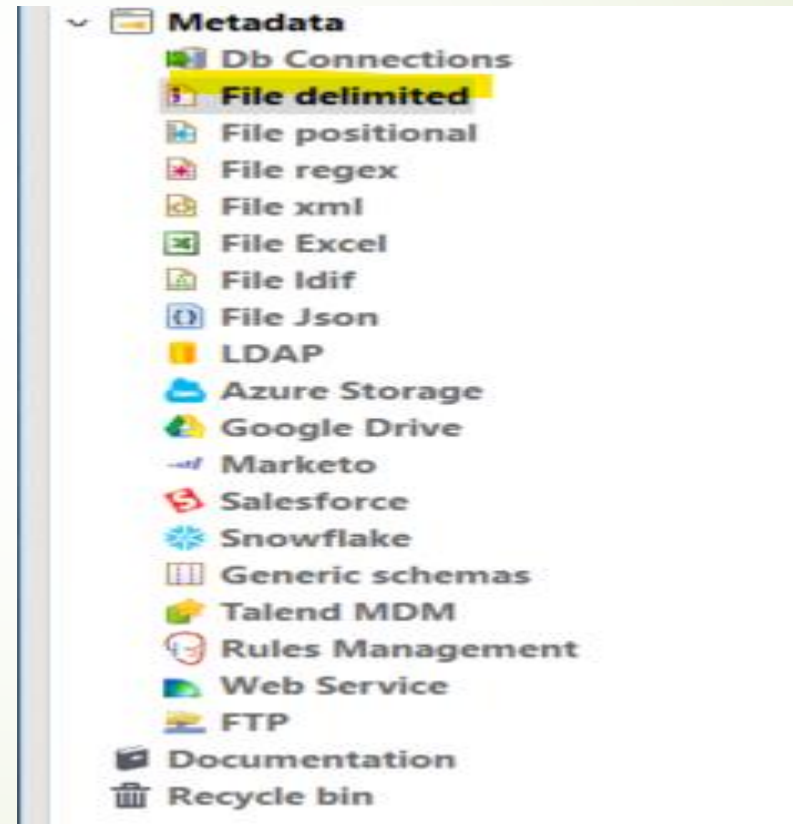




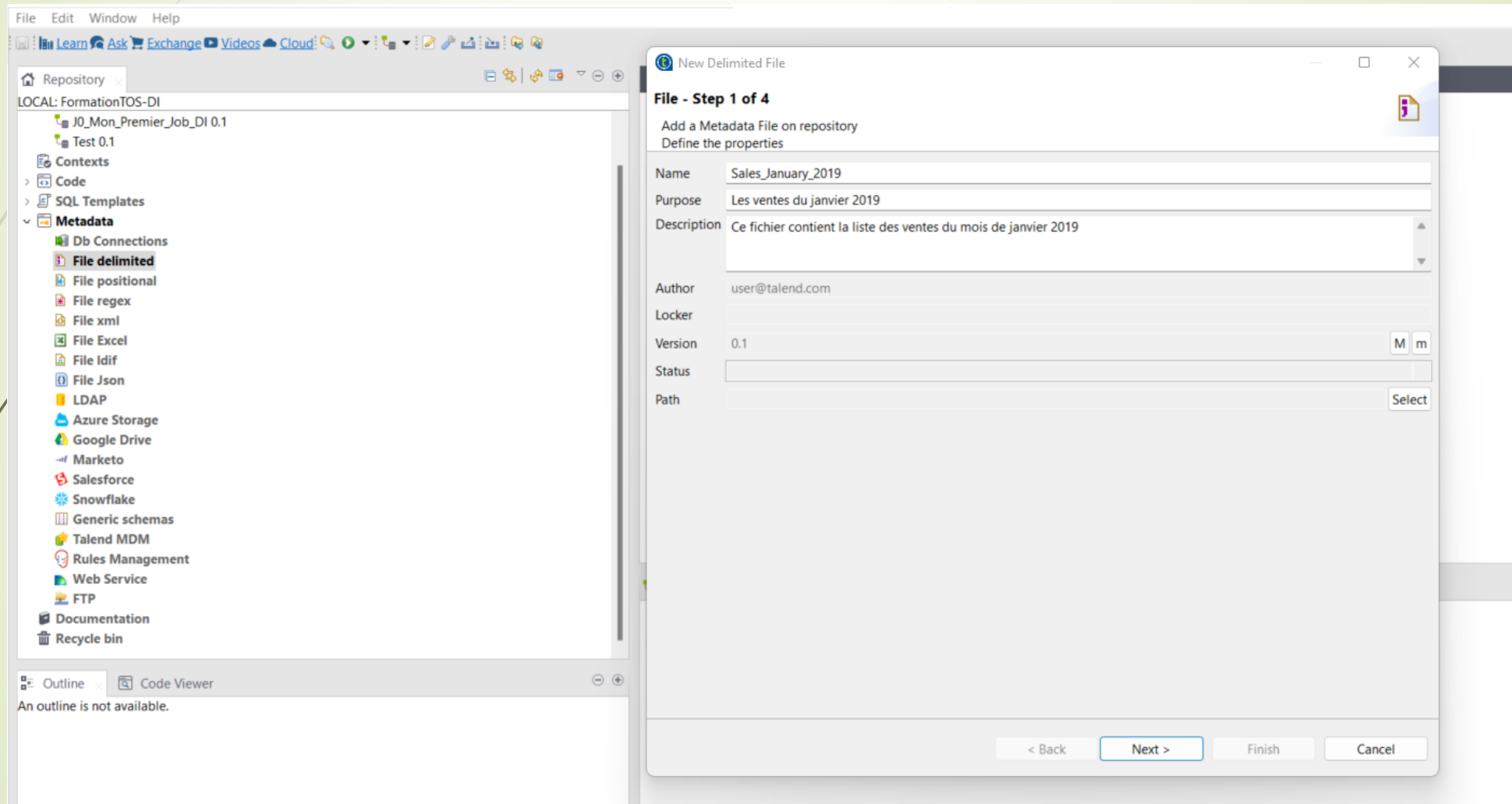
Lab !

Lab1: Prise en main de métadonnées Fichier délimité

Création d'une métadonnée fichier délimité



Lab1: Prise en main de métadonnées Fichier délimité



Lab1: Prise en main de métadonnées Fichier délimité

New Delimited File

File - Step 2 of 4

Add a Metadata File on repository
Define the path of the file and the format settings

File Settings

Server: Localhost 127.0.0.1

File: C:/Formation TOS DI/Data/Input/Ventes/Sales_January_2019.csv Browse...

Format: WINDOWS

File Viewer

Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
141234	iPhone	1	700	01/22/19 21:25	"944 Walnut St, Boston, MA 02215"
141235	Lightning Charging Cable	1	14.95	01/28/19 14:15	"185 Maple St, Portland, OR 97035"
141236	Wired Headphones	2	11.99	01/17/19 13:33	"538 Adams St, San Francisco, CA 94016"
141237	27in FHD Monitor	1	149.99	01/05/19 20:33	"738 10th St, Los Angeles, CA 90001"
141238	Wired Headphones	1	11.99	01/25/19 11:59	"387 10th St, Austin, TX 73301"
141239	AAA Batteries (4-pack)	1	2.99	01/29/19 20:22	"775 Willow St, San Francisco, CA 94016"
141240	27in 4K Gaming Monitor	1	389.99	01/26/19 12:16	"979 Park St, Los Angeles, CA 90001"
141241	USB-C Charging Cable	1	11.95	01/05/19 12:04	"181 6th St, San Francisco, CA 94016"
141242	Bose SoundSport Headphones	1	99.99	01/01/19 10:30	"867 Willow St, Los Angeles, CA 90001"
141243	Apple AirPods Headphones	1	150	01/22/19 21:20	"657 Johnson St, San Francisco, CA 94016"
141244	Apple AirPods Headphones	1	150	01/07/19 11:29	"492 Walnut St, San Francisco, CA 94016"

< Back Next > Finish Cancel

Lab1: Prise en main de métadonnées Fichier délimité

New Delimited File

File - Step 3 of 4

Add a Metadata File on repository
Define the setting of the parse job

File Settings

Encoding: UTF-8

Field Separator: Comma Corresponding Character: ","

Row Separator: Standard EOL Corresponding Character: "\n"

Escape Char Settings

☒ CSV ☐ Delimited

Escape Char: \"

Text Enclosure: \"

☐ Split row before field

Rows To Skip

If any rows must be ignored, specify the following parameters

Header: ☒ 1

Footer: ☐

☐ Skip empty row

Limit Of Rows

If the number of lines must be limited, specify this number

Limit: ☐

Preview | **Output**

☒ Set heading row as column names Refresh Preview

Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
141234	iPhone	1	700	01/22/19 21:25	944 Walnut St, Boston, MA 02215
141235	Lightning Charging Cable	1	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035
141236	Wired Headphones	2	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016
141237	27in FHD Monitor	1	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001
141238	Wired Headphones	1	11.99	01/25/19 11:59	387 10th St, Austin, TX 73301
141239	AAA Batteries (4-pack)	1	2.99	01/29/19 20:22	775 Willow St, San Francisco, CA 94016
141240	27in 4K Gaming Monitor	1	389.99	01/26/19 12:16	979 Park St, Los Angeles, CA 90001
141241	USB-C Charging Cable	1	11.95	01/05/19 12:04	181 6th St, San Francisco, CA 94016
141242	Bose SoundSport Headphones	1	99.99	01/01/19 10:30	867 Willow St, Los Angeles, CA 90001

Export as context Revert Context

< Back Next > Finish Cancel

Lab1: Prise en main de métadonnées Fichier délimité

New Delimited File

File - Step 4 of 4

Add a Schema on repository
Define the Schema

Name

Comment

Schema

Click to update schema preview

Description of the Schema

Column	K...	Type	<input checked="" type="checkbox"/> N..	Date Pattern (Ctrl+Spac...	Length	Precision	Default	Comment
Order_ID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		6	0		
Product	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		26	0		
Quantity_Ordered	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		1	0		
Price_Each	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>		6	4		
Order_Date	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		14	0		
Purchase_Address	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		40	0		

< Back Next >

Lab2: Création d'un job qui affiche la liste des ventes du mois de janvier

- ❑ L'objectif de ce lab est la création d'un job Talend qui affiche les ventes du mois de janvier 2019
- ❑ Pour réaliser ce job, vous devez utiliser deux composants FileinputDelimited et TlogRow
- ❑ Télécharger le dossier Ventes.zip
- ❑ Dézipper l'archive dans un répertoire local.

Lab3: Création d'un job qui calcul la quantité vendue par produit

- ❑ L'objectif de ce lab est de calculer la quantité vendue par produit en mois de janvier 2019
- ❑ Le job a aussi pour objectif de vérifier le schéma d'entrée des données
- ❑ Rejeter les lignes qui ne respecte pas les métadonnées d'entrée
- ❑ Supprimer l'espace de début et de fin de chaque colonne
- ❑ Les composants dont vous avez besoin sont les suivants:
 - ✓ TFileInputDelimited
 - ✓ tSchemaComplianceCheck
 - ✓ tAggregateRow
 - ✓ TlogRow

Lab3: Création d'un job qui calcul la quantité vendue par produit.

tLogRow_1	
Product	Quantity_Ordered
Lightning Charging Cable	1126
iPhone	375
27in FHD Monitor	414
27in 4K Gaming Monitor	312
ThinkPad Laptop	215
Bose SoundSport Headphones	646
USB-C Charging Cable	1262
AA Batteries (4-pack)	1407
Apple Airpods Headphones	807
Google Phone	311
LG Dryer	37
34in Ultrawide Monitor	312
Wired Headphones	1067
Flatscreen TV	243
20in Monitor	210
LG Washing Machine	42
Macbook Pro Laptop	229
AAA Batteries (4-pack)	1574
Vareebadd Phone	123
	null

Lab4: Création d'un job qui affiche les produits « iPhone »

- ❑ L'objectif de ce lab est d'afficher tous les produit « iPhone» vendues en mois janvier 2019
- ❑ Pour faire ce lab, vous devez utilisez le composant tFilterRow

Job J03_Afficher_Produit

Execution

Run Kill Clear

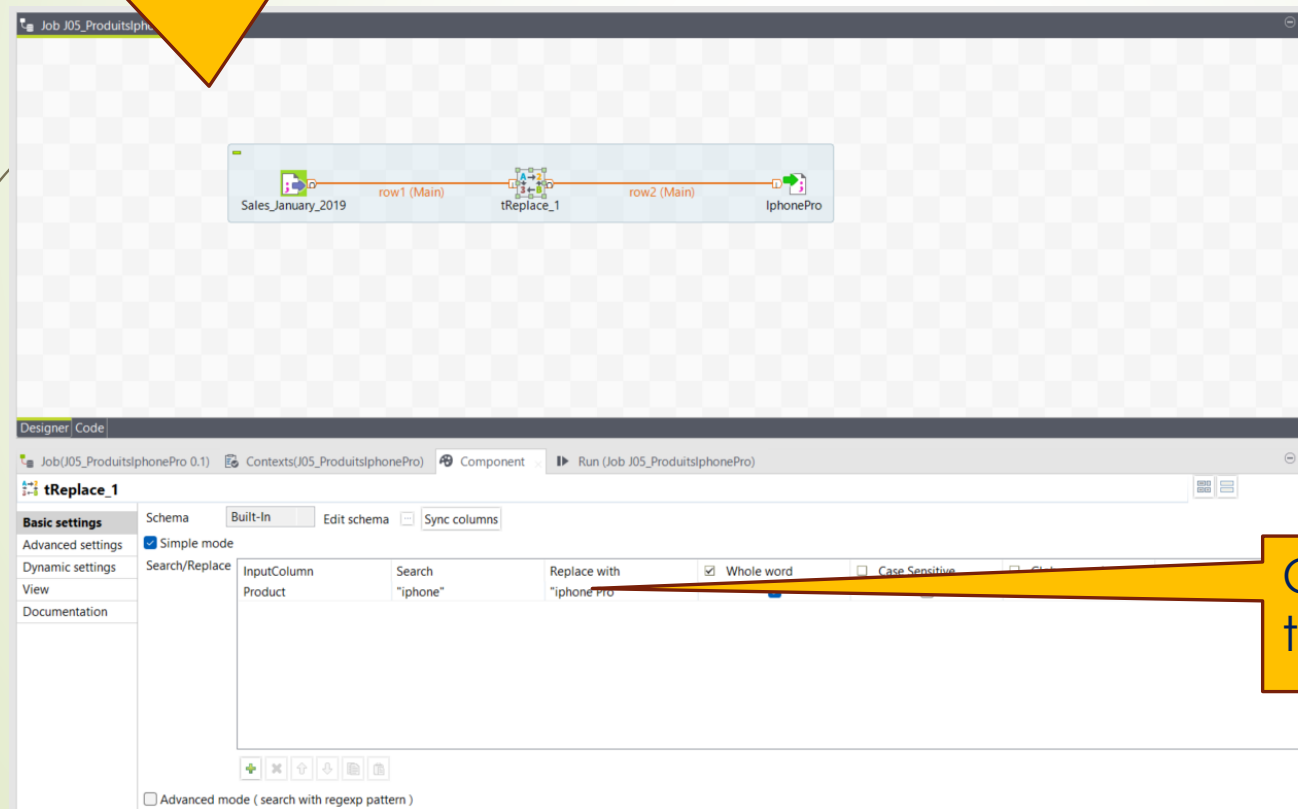
Order_ID	Product	Quantity_Ordered	Price_Each	Order_Date	Purchase_Address
141234	iPhone	1	700.0	01/22/19 21:25	944 Walnut St, Boston, MA 02215
141336	iPhone	1	700.0	01/09/19 18:23	811 Hickory St, Portland, OR 97035
141394	iPhone	1	700.0	01/06/19 16:54	534 12th St, San Francisco, CA 94016
141437	iPhone	1	700.0	01/10/19 15:40	377 Meadow St, New York City, NY 10001
141457	iPhone	1	700.0	01/09/19 22:11	820 Jackson St, Seattle, WA 98101
141458	iPhone	1	700.0	01/29/19 01:12	497 Park St, San Francisco, CA 94016
141460	iPhone	1	700.0	01/26/19 17:29	855 2nd St, New York City, NY 10001
141472	iPhone	1	700.0	01/27/19 07:05	853 9th St, San Francisco, CA 94016
141473	iPhone	1	700.0	01/11/19 16:48	768 Maple St, San Francisco, CA 94016
141476	iPhone	1	700.0	01/06/19 15:38	295 Hickory St, New York City, NY 10001
141504	iPhone	1	700.0	01/13/19 13:37	855 11th St, Seattle, WA 98101
141517	iPhone	1	700.0	01/13/19 14:14	326 Maple St, Los Angeles, CA 90001
141520	iPhone	1	700.0	01/15/19 11:24	23 Cherry St, Atlanta, GA 30301
141550	iPhone	1	700.0	01/31/19 10:58	399 Church St, Boston, MA 02215
141582	iPhone	1	700.0	01/15/19 01:45	416 9th St, San Francisco, CA 94016
141632	iPhone	1	700.0	01/21/19 05:00	964 Hill St, Seattle, WA 98101
141663	iPhone	1	700.0	01/01/19 11:01	738 8th St, San Francisco, CA 94016
141670	iPhone	1	700.0	01/31/19 12:52	166 10th St, Atlanta, GA 30301
141702	iPhone	1	700.0	01/27/19 18:33	776 10th St, Boston, MA 02215
141703	iPhone	1	700.0	01/29/19 11:04	371 11th St, Portland, OR 97035
141732	iPhone	1	700.0	01/01/19 06:13	446 Pine St, Atlanta, GA 30301
141738	iPhone	1	700.0	01/14/19 20:53	183 Cherry St, Atlanta, GA 30301
141795	iPhone	1	700.0	01/19/19 20:31	383 Jefferson St, San Francisco, CA 94016
141886	iPhone	1	700.0	01/10/19 06:37	94 7th St, New York City, NY 10001
141888	iPhone	1	700.0	01/09/19 11:01	17 Spruce St, Los Angeles, CA 90001
141961	iPhone	1	700.0	01/12/19 17:39	512 Hickory St, Los Angeles, CA 90001
141967	iPhone	1	700.0	01/25/19 12:02	397 Meadow St, Los Angeles, CA 90001
141972	iPhone	1	700.0	01/24/19 15:25	52 Main St, San Francisco, CA 94016
141993	iPhone	1	700.0	01/06/19 09:19	422 Jackson St, Seattle, WA 98101
142009	iPhone	1	700.0	01/15/19 13:59	411 11th St, San Francisco, CA 94016
142010	iPhone	1	700.0	01/04/19 06:28	511 2nd St, Portland, OR 97035
142035	iPhone	1	700.0	01/17/19 11:06	890 North St, Atlanta, GA 30301
142037	iPhone	1	700.0	01/11/19 21:48	230 North St, Austin, TX 73301
142050	iPhone	1	700.0	01/22/19 21:25	944 Walnut St, Boston, MA 02215

Lab4: Création d'un job qui remplace les produits « iPhone » par « iPhone Pro »

- ❑ L'objectif de ce lab est de remplacer le produit « iPhone » par « iPhone Pro »
- ❑ Nous souhaitons récupérer le résultat sous forme d'un fichier plat (csv)
- ❑ Proposer deux solutions différentes:
 - ✓ Utilisation du composant tReplace
 - ✓ Utilisation du tMap

Lab4: Création d'un job qui remplace les produits « iPhone » par « iPhone Pro » Solution 1

1. Créer un job j04_ProduitsIphonePro et déposez les composants suivants :



Configurer le composant tReplace

Lab4: Création d'un job qui remplace les produits « iPhone » par « iphone Pro » Solution 2

1. Créer un job j04_ProduitsIphonePro_V2 et déposez les composants suivants :

The screenshot shows the Talend Open Studio for Data Integration interface. At the top, a job design is visible with the following components: **Sales_January_2019** (Source) → **row1 (Main)** (Table Input) → **tMap_1** (Transformation) → **Out_IphonePro (Main)** (Table Output) → **IphoneProV2** (Destination).

The **tMap_1** component is selected, and its configuration window is open. The window has two tabs: **Schema editor** and **Expression editor**. The **Schema editor** tab is active, showing the mapping between the input schema **row1** and the output schema **Out_IphonePro**.

Input Schema (row1):

Column	K...	Type	✓	N.	Date Patter...	Length	Precis...	Def...	Com...
Order_ID		String	✓			255	0		
Product		String	✓			255	0		
Quantity_Ordered		String	✓			255	0		
Price_Each		String	✓			255	4		
Order_Date		String	✓			255	0		
Purchase_Address		String	✓			255	0		

Output Schema (Out_IphonePro):

Column	K...	Type	✓	N.	Date Patter...	Length	Precis...	Def...	Com...
Order_ID		String	✓			255	0		
Product		String	✓			255	0		
Quantity_Ordered		String	✓			255	0		
Price_Each		String	✓			255	4		
Order_Date		String	✓			255	0		
Purchase_Address		String	✓			255	0		

The **Expression editor** tab is also visible, showing the following expressions:

- row1.Order_ID
- row1.Product.Equals("iPhone")?"iPhone Pro":row1.Product
- row1.Quantity_Ordered
- row1.Price_Each
- row1.Order_Date
- row1.Purchase_Address

A yellow callout box points to the **Product** column in the input schema and the corresponding expression, with the text "If else then".



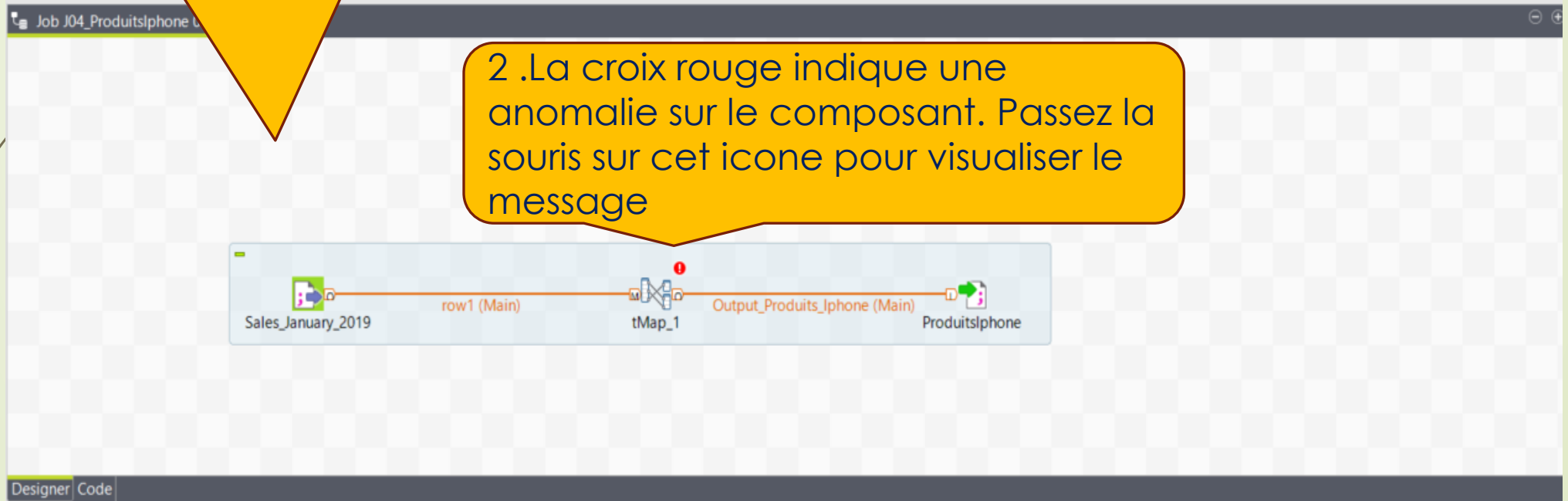
Lab5: Création d'un job qui envoie vers un fichier les ventes du produit iPhone.

- ☐ L'objectif de ce lab est de récupérer les produits « iphone » vendues en mois de janviers
- ☐ Supprimer l'espace au début et la fin de chaque colonne
- ☐ Utiliser le composant Tmap

Lab5: Création d'un job qui envoie vers un fichier les ventes du produit iPhone.

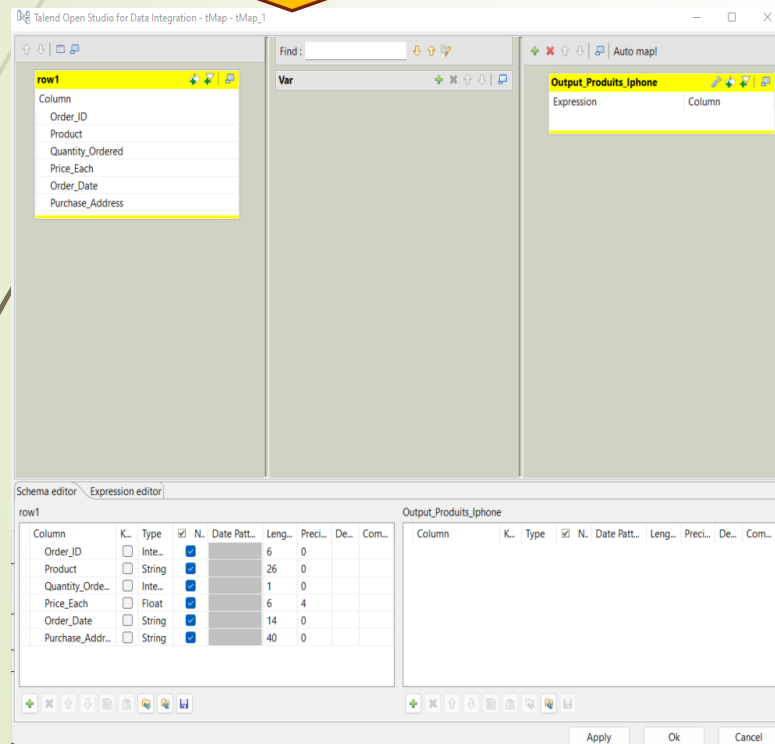
1. Créer un job j04_ProduitsIphone et déposez les composants suivants :

2 .La croix rouge indique une anomalie sur le composant. Passez la souris sur cet icone pour visualiser le message

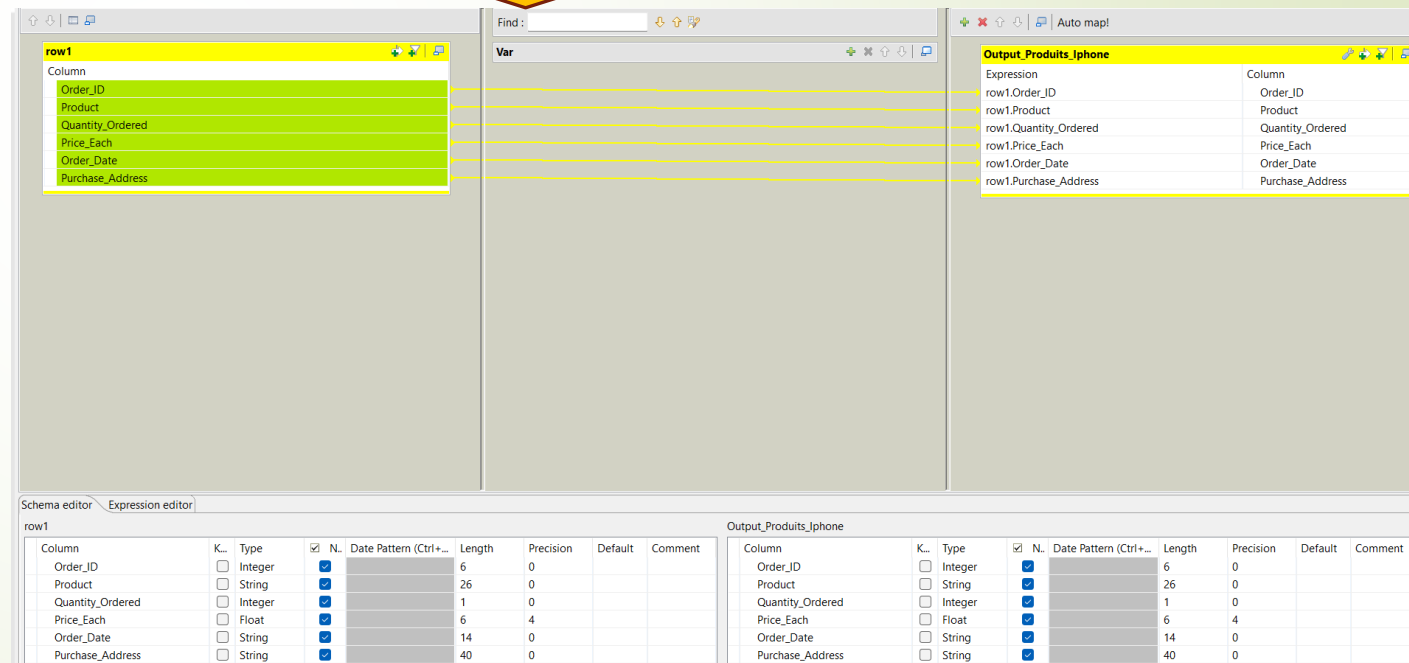


Lab5: Création d'un job qui envoie vers un fichier les ventes du produit iPhone.

3. Un double clic sur le tMap permet de l'ouvrir



4. Par drag & drop, effectuez le paramétrage suivant



Lab6 : Contrôle d'intégrité

- ❑ L'objectif de ce Job est de récupérer la liste des ventes référencées, c'est-à-dire les ventes que leur code de produit existe dans le fichier(table) produit.
- ❑ L'étape 1 consiste à définir le fichier produit dans la liste métadonnées

Lab6 : Définir le fichier Produits dans les metadata

- Etape1: consiste à définir le fichier produit dans les métadonnées

Edit an existing Delimited File

File - Step 3 of 3

Update an existing Metadata File on repository
Define the setting of the parse job

File Settings

Encoding: UTF-8

Field Separator: Comma

Row Separator: Stanc

Rows To Skip
If any rows must be ignored, specify the following parameters

Escape Char Settings

☒ CSV

Escape Char: \"

Text Enclosure: \"

☐ Split row before field

Preview | **Output**

☒ Set heading row as column names Refresh Preview

Product	Description
Lightning Charging Cable	Lightning Charging Cable
iPhone	iPhone apple
27in FHD Monitor	27in FHD Monitor
27in 4K Gaming Monitor	27in 4K Gaming Monitor
ThinkPad Laptop	ThinkPad Laptop

Export as context Revert Context

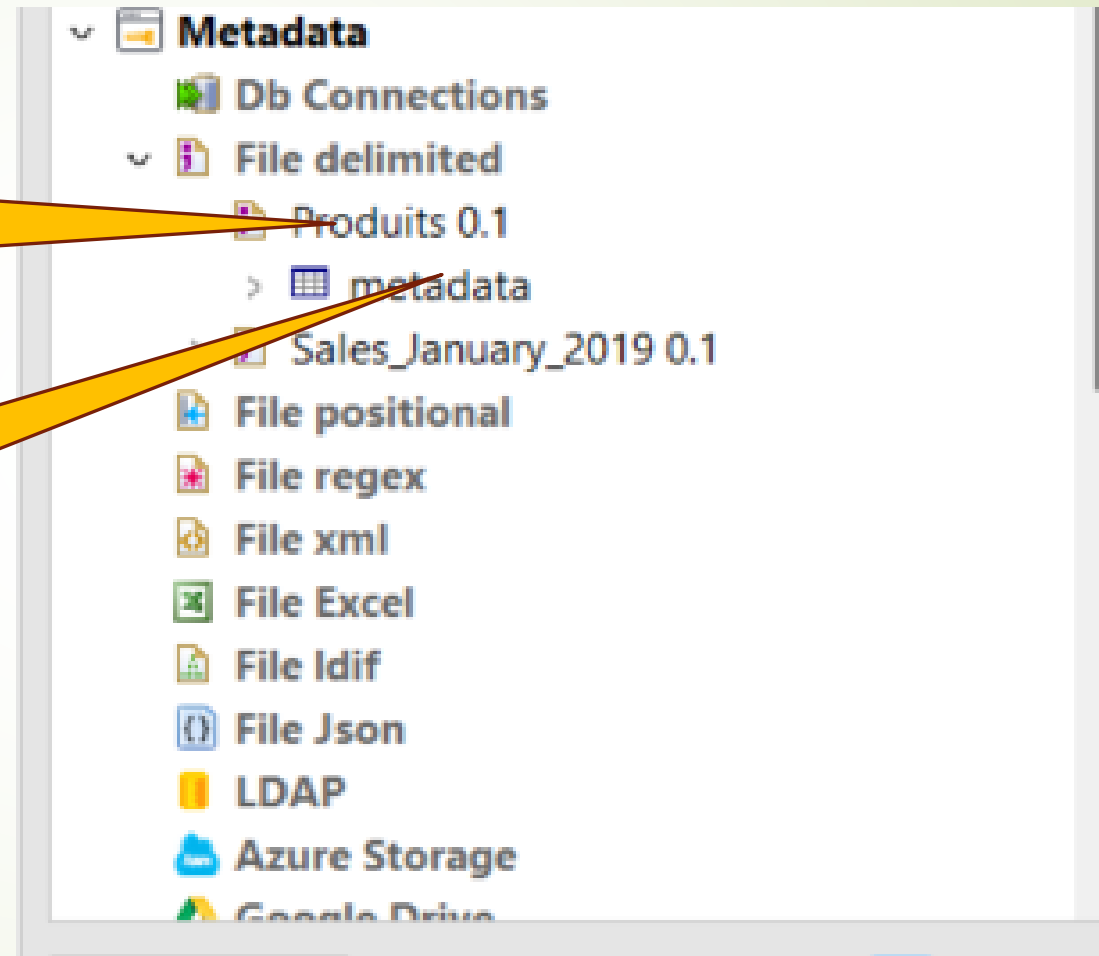
< Back Next > Finish Cancel

De la même façon que le fichier Ventes, définissez le fichier Produits à partir du fichier Produits.csv fourni pour la formation

Lab6 : Définir le fichier Produits dans les metadata

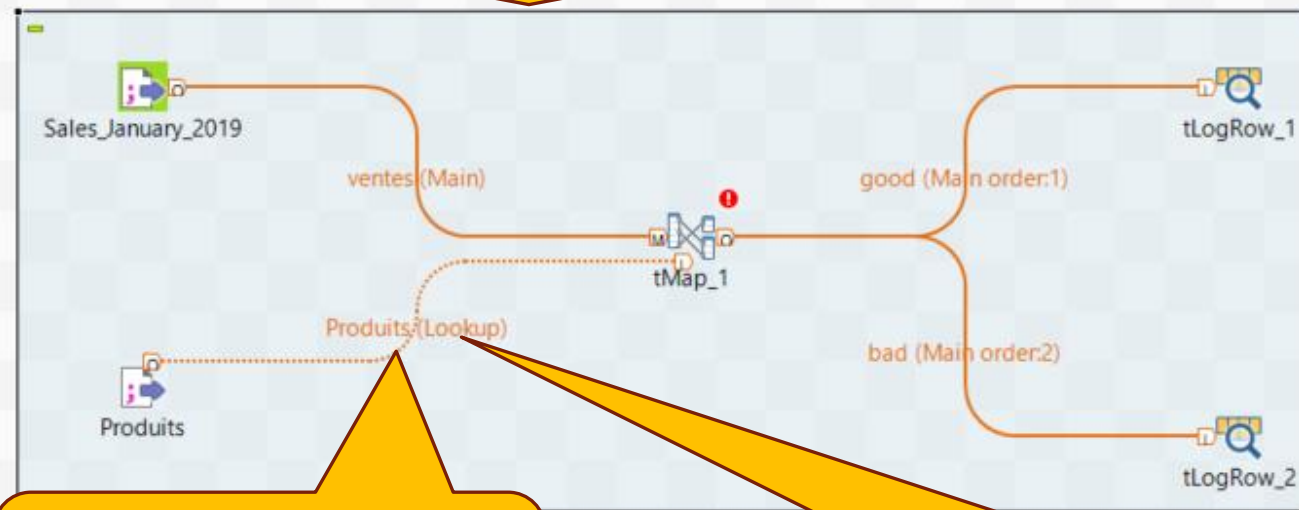
1. Comprenez bien la différence entre la définition du fichier (chemin, séparateur, encodage, etc...)

2. Comprenez bien la différence entre la définition du format du fichier (nom des colonnes et types des données)



Lab6: Contrôle d'intégrité

Créez un job_005 avec les composants suivants. Nous cherchons à identifier les ventes qui concernent des produits connus (lien good) et les anomalies (lien bad)



Donnez des noms explicites aux liens

Notez qu'un des liens en entrée du tMap est marqué « main » et l'autre « Lookup ».

Lab6: Contrôle d'intégrité

1. Ouvrez le tMap et définissez la jointure entre les deux tables sources

3. Par drag&drop, définissez la clé de jointure

ventes

Column
Order_ID
Product
Quantity_Ordered
Price_Each
Order_Date
Purchase_Address

Produits

Property	Value
Lookup Model	Load once
Match Model	Unique match
Join Model	Inner Join
Store temp data	false
Expr. key	Column
ventes.Product	Product
	Description

Find :

Var

2. Puis les options

Lab6: Contrôle d'intégrité

1. Sélectionnez toutes les colonnes de la table ventes

2. et drag&drop dans la cible good

The screenshot shows a data integration tool interface with three main panels. The left panel displays the 'ventes' table with columns: Order_ID, Product, Quantity_Ordered, Price_Each, Order_Date, and Purchase_Address. Below this is the 'Produits' section with properties: Lookup Model (Load once), Match Model (Unique match), Join Model (Inner Join), Store temp data (false), and Expr. key (ventes.Product). The middle panel is a mapping area labeled 'Var'. The right panel shows two target tables: 'good' and 'bad'. The 'good' table has columns: Order_ID, Product, Quantity_Ordered, Price_Each, Order_Date, and Purchase_Address. The 'bad' table has columns: Order_ID, Product, Quantity_Ordered, Price_Each, Order_Date, and Purchase_Address. The 'bad' table also has properties: Catch output reject (false), Catch lookup inner join reject (true), and Schema Type (Built-In). Arrows indicate the mapping of columns from 'ventes' to both 'good' and 'bad'.

4. Cliquez sur « enable lookup inner join reject » pour indiquer que ce flux « bad » contiendra les enregistrements qui ne satisfont pas à la jointure

3. même opération dans la cible bad

Lab6: Contrôle d'intégrité

1. Modifier les options suivantes des deux tLogRow

Job(J06_VentesProduits_CT 0.1) Contexts(J06_VentesProduits_CT) Component x Run (Job J06_VentesProduits_CT)

tLogRow_2

Basic settings

Schema Built-In Edit schema ... Sync columns

Mode

☒ Basic

☐ Table (print values in cells of a table)

☐ Vertical (each row is a key/value list)

Field Separator "|"

☐ Print header

☒ Print component unique name in front of each output row

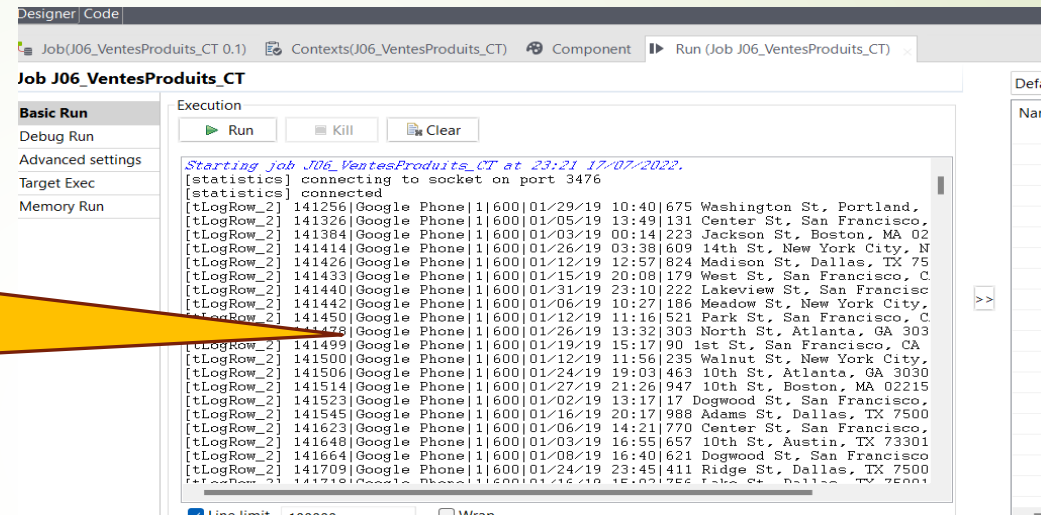
☐ Print schema column name in front of each value

☐ Use fixed length for values

☐ Print content with log4j

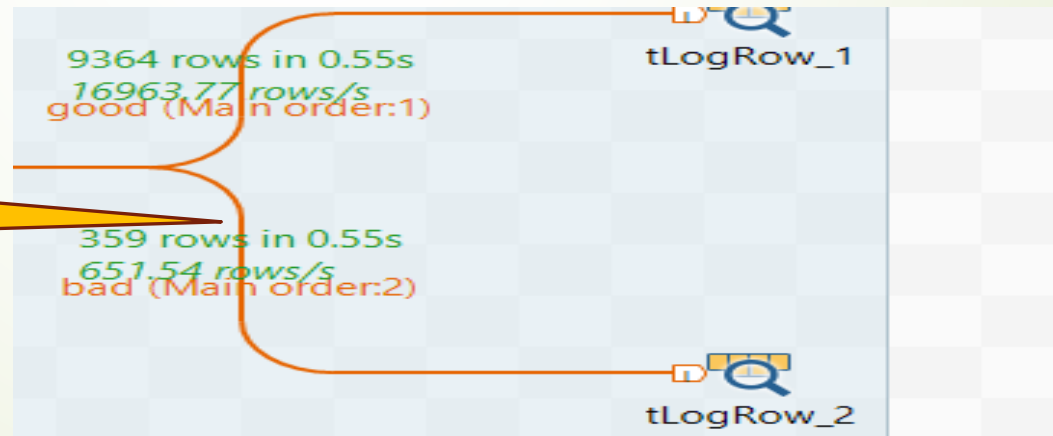
Lab6: Contrôle d'intégrité

1. On peut visualiser les rejets dans la sortie de la console. Le produit « Google Phone » est inconnu dans le fichier produits en effet



```
Starting job J06_VentesProduits_CT at 23:21 17/07/2022.
[statistics] connecting to socket on port 3476
[statistics] connected
[tLogRow_2] 141256|Google Phone|1|600|01/29/19 10:40|675 Washington St, Portland,
[tLogRow_2] 141326|Google Phone|1|600|01/05/19 13:49|131 Center St, San Francisco,
[tLogRow_2] 141384|Google Phone|1|600|01/03/19 00:14|223 Jackson St, Boston, MA 02
[tLogRow_2] 141414|Google Phone|1|600|01/26/19 03:38|609 14th St, New York City, N
[tLogRow_2] 141426|Google Phone|1|600|01/12/19 12:57|824 Madison St, Dallas, TX 75
[tLogRow_2] 141433|Google Phone|1|600|01/15/19 20:08|179 West St, San Francisco, C
[tLogRow_2] 141440|Google Phone|1|600|01/31/19 23:10|222 Lakeview St, San Francisc
[tLogRow_2] 141442|Google Phone|1|600|01/06/19 10:27|166 Meadow St, New York City,
[tLogRow_2] 141450|Google Phone|1|600|01/12/19 11:16|521 Park St, San Francisco, C
[tLogRow_2] 141478|Google Phone|1|600|01/26/19 13:32|303 North St, Atlanta, GA 303
[tLogRow_2] 141499|Google Phone|1|600|01/19/19 15:17|90 1st St, San Francisco, CA
[tLogRow_2] 141500|Google Phone|1|600|01/12/19 11:56|235 Walnut St, New York City,
[tLogRow_2] 141506|Google Phone|1|600|01/24/19 19:03|463 10th St, Atlanta, GA 3030
[tLogRow_2] 141514|Google Phone|1|600|01/27/19 21:26|947 10th St, Boston, MA 02215
[tLogRow_2] 141523|Google Phone|1|600|01/02/19 13:17|17 Dogwood St, San Francisco,
[tLogRow_2] 141545|Google Phone|1|600|01/16/19 20:17|988 Adams St, Dallas, TX 7500
[tLogRow_2] 141623|Google Phone|1|600|01/06/19 14:21|770 Center St, San Francisco,
[tLogRow_2] 141648|Google Phone|1|600|01/03/19 16:55|657 10th St, Austin, TX 73301
[tLogRow_2] 141664|Google Phone|1|600|01/08/19 16:40|621 Dogwood St, San Francisco
[tLogRow_2] 141709|Google Phone|1|600|01/24/19 23:45|411 Ridge St, Dallas, TX 7500
[tLogRow_2] 141719|Google Phone|1|600|01/16/19 15:03|756 Lake St, Dallas, TX 75001
```

2. 359 lignes ont été rejetées

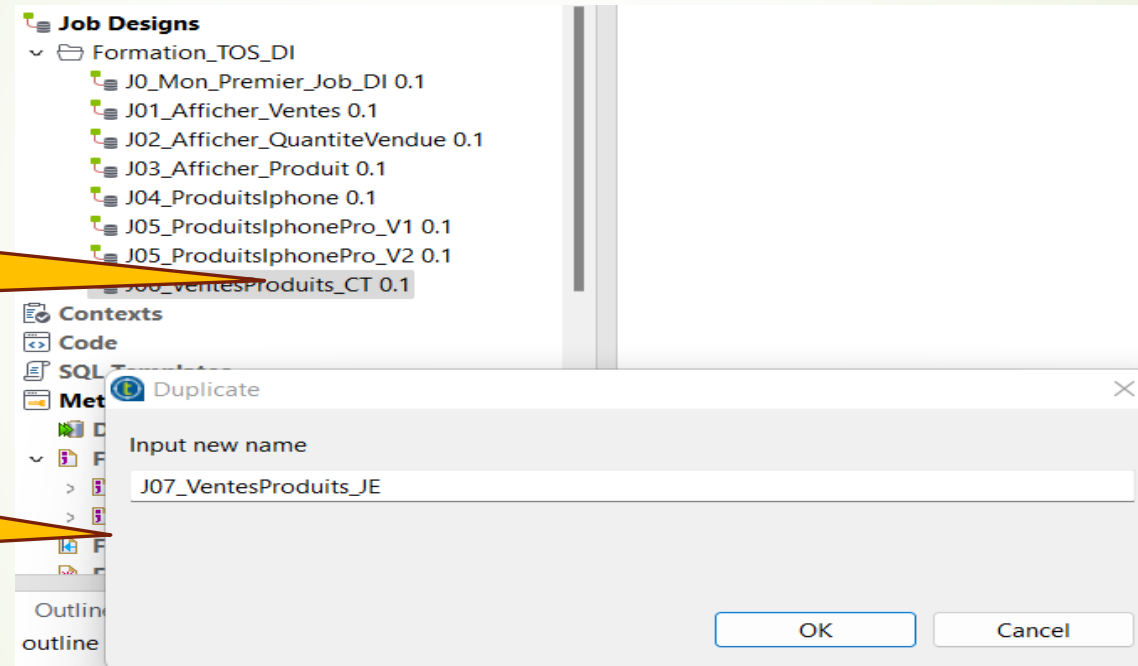


Lab7: Dupliquer un job

❑ L'objectif de ce lab est de dupliquer le job précédent

1. Enregistrez et fermez le job puis clic droit/
Duplicate

2. Appelez le nouveau job
j07_ventesProduits
_JE puis OK



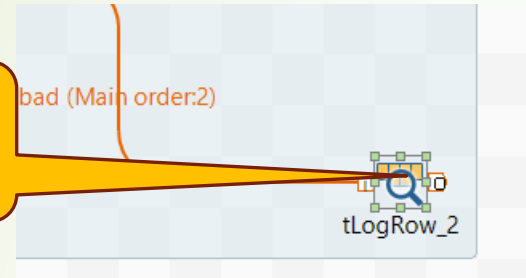


Lab8: Jointure externe

- ❑ L'objectif du job numéro 6 nous permet de récupérer les ventes des produits existants, la direction de l'entreprise souhaite avoir toutes les ventes des produits, que ce soit des produits référencés ou non.
- ❑ Adapter le job que vous venez de dupliquer afin de répondre aux besoins de la direction.

Lab8: Jointure externe

1. Clic droit puis supprimer le composant TlogRow2



3. Ajouter en sortie good la colonne description de la table du produit

2. Choisir l'option Left Outer join

Talend Open Studio for Data Integration - tMap - tMap_1

Find :

Var

ventes

Column
Order_ID
Product
Quantity_Ordered
Price_Each
Order_Date
Purchase_Address

Produits

Property	Value
Lookup Model	Load once
Match Model	Unique match
Join Model	Left Outer Join
Store temp data	false

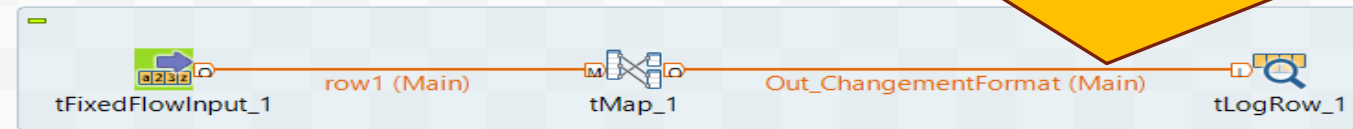
Expr. key	Column
ventes.Product	Product
	Description

good

Expression	Column
ventes.Order_ID	Order_ID
ventes.Product	Product
ventes.Quantity_Ordered	Quantity_Ordered
ventes.Price_Each	Price_Each
ventes.Order_Date	Order_Date
ventes.Purchase_Address	Purchase_Address
Produits.Description	Description

Lab9: Changement du format d'un champ

1. Créez un job_007 avec les composants suivants. Le composant tFixedFlowInput (catégorie Misc) permet de générer des données (utile pour tester les composants notamment)



2. Allez dans les propriétés du composant tFixedFlowInput et cliquez sur « Edit Schéma ». Ce composant permet de générer des données. Il est utile notamment pour les tests (ou les formations !)

Column	K...	Type	<input checked="" type="checkbox"/> N.	Date Pa...	Len...	Pre...	D...	Co...
Produit	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>					
Prix	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>					

Buttons: +, -, up, down, print, save, load, export, import, disk

OK Cancel

Lab9: Changement de format

1. Le résultat de votre job sur la console

```
Starting job J08_ChangementFormat at 00:35 18/07/2022.  
[statistics] connecting to socket on port 3693  
[statistics] connected  
PC Portable|1300  
[statistics] disconnected  
  
Job J08_ChangementFormat ended at 00:35 18/07/2022. [Exit code = 0]
```

Lab9: Gestions des erreurs

1. Changement du format de la colonne Prix de string à integer

à integer

Schema editor

Expression editor

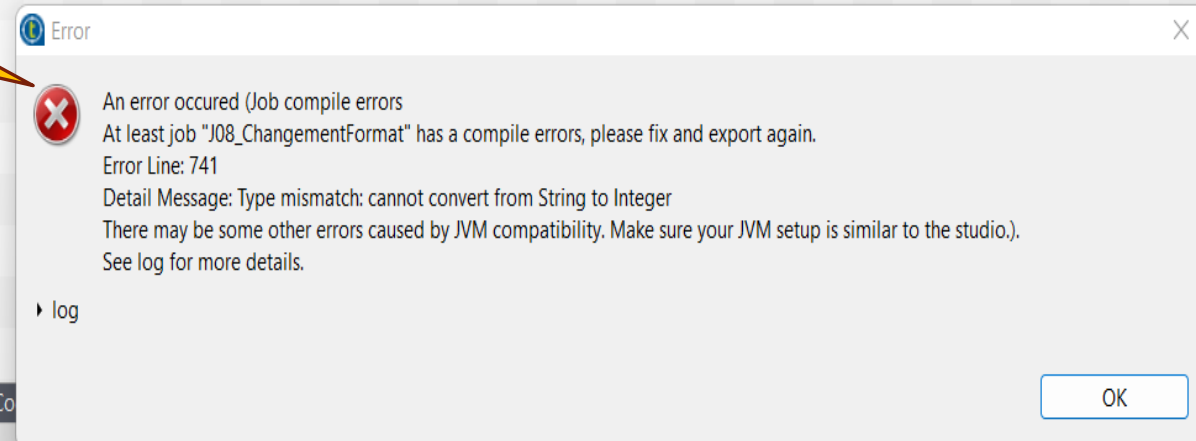
row1

Column	K...	Type	<input checked="" type="checkbox"/> N..	Date Pattern (Ctrl+...	Length	Precision	Default	Comment
Produit		<input type="checkbox"/> String	<input checked="" type="checkbox"/>					
Prix		<input type="checkbox"/> String	<input checked="" type="checkbox"/>					

Out_ChangementFormat

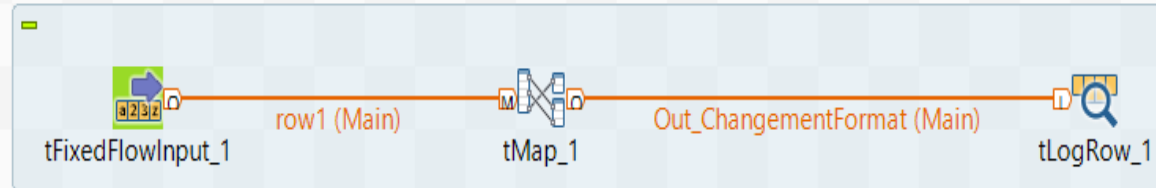
Column		<input checked="" type="checkbox"/> N..	Date Pattern (Ctrl+...	Length
Produit	<input type="checkbox"/> String	<input checked="" type="checkbox"/>		
Prix	<input type="checkbox"/> Integer	<input checked="" type="checkbox"/>		

2. En exécutant le job on obtiens le message suivant:



Lab9: Gestion des erreurs

1. Ajouter et configurer les deux composants :



2. Vous pouvez ainsi tracer les erreurs de votre chaine de traitement dans un fichier (ou une table) journal des erreurs

Lab9: Changement de format

- Autres conversions de format (se référer au langage java):

Format source	Format cible	Expression
Integer	String	String.valueOf(row1.prix) Row1.prix + ""
Integer	Float	New Float(row1.prix)
Integer	Long	New Long(row1.prix)
Mois String Annee String	Date	TalendDate.parseDate("yyyy-MM-dd", row1.Annee+"-"+row1.Mois+"-01")
String	Integer	Integer.parseInt(row1.prix) New Integer(row1.prix)

Gestions des contextes

- ❑ Pour passer facilement d'un environnement de développement à un environnement de production sans modifier les paramètres « en dur » (par exemple, le nom du serveur de données cible, ou les chemins vers les répertoires sources)
- ❑ Nous allons créer un « Context group » qui contiendra:
 - ❑ des « contextes » (par exemple, « dev », « test » et « prod »)
 - ❑ des variables (par exemple, « server_name », « sources_path »)
- ❑ Ensuite, nous donnerons des valeurs à chaque variable pour chaque contexte (par exemple, pour le contexte « dev », la variable « source_path » vaut « c:/temp »)

Gestions des contextes

- ❑ Enfin, nous choisirons à l'exécution du job le contexte dans lequel nous souhaitons qu'il s'exécute: « dev », « test » ou « prod »

1. Clic droit sur le répertoire contexte, nommez le ...

Create / Edit a context group

Step 1 of 2

⚠ It is inadvisable to leave the purpose blank.

Name: InfoCentre_Ventes

Purpose:

Description:

Author: user@talend.com

Locker:

Version: 0.1 M m

Status:

Path: Select

< Back Next > Finish Cancel

Gestions des contextes

Create / Edit a context group

Step 2 of 2

Define the contexts, variables and values

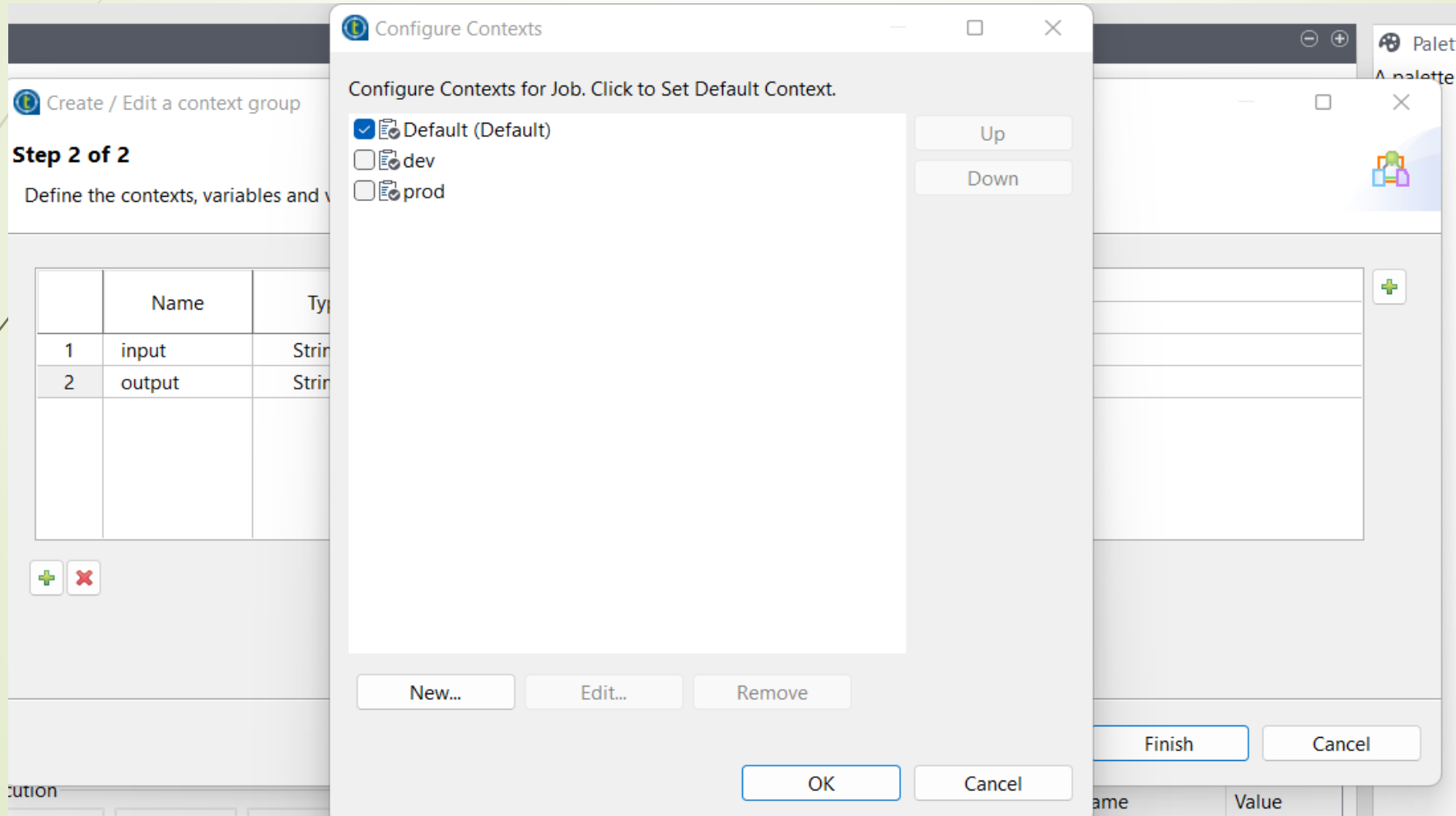
	Name	Type	Comment	Default Value	
1	input	String		"C:/Formation TOS DI/Data/Input/"	<input type="checkbox"/>
2	output	String		"C:/Formation TOS DI/Data/Output/"	<input type="checkbox"/>

+ -

Default context environment Default

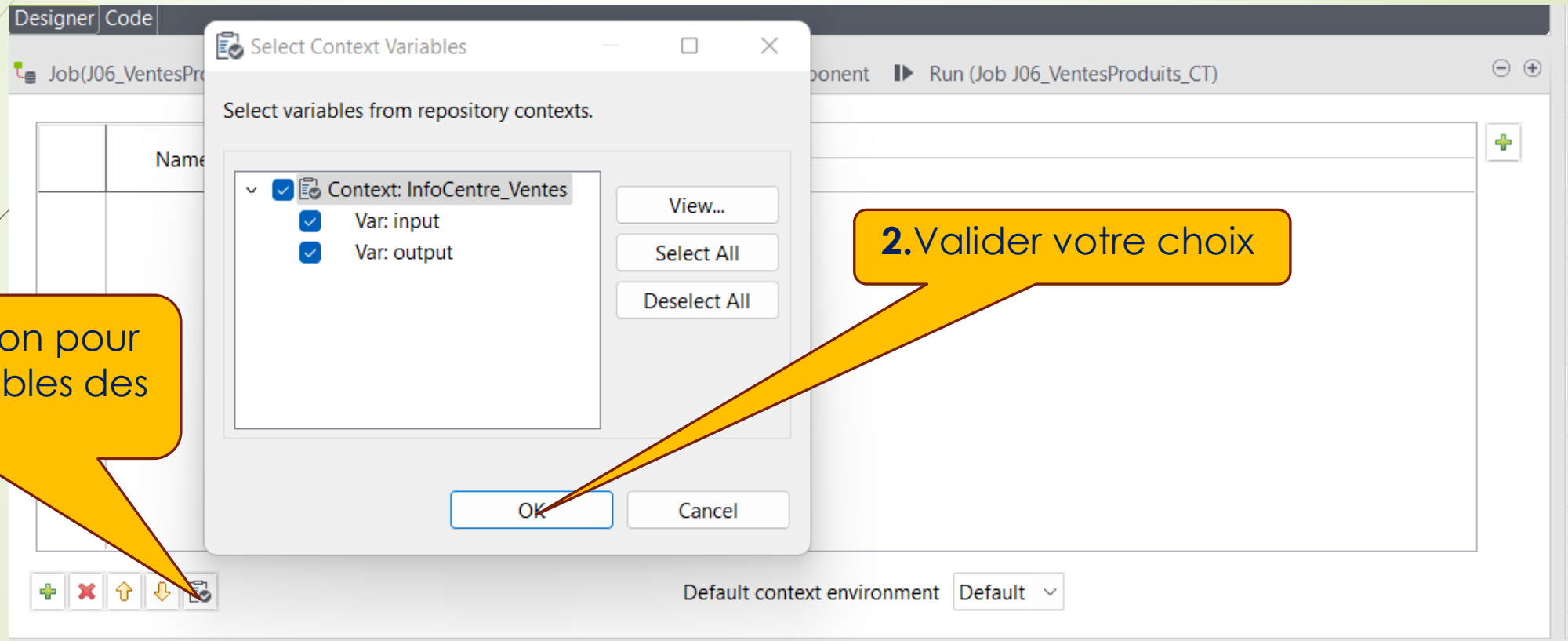
< Back Next > Finish Cancel

Gestions des contextes : ajouter l'environnement de dev et de prod



Utilisation des variables : chargement des variables des contextes dans un job

1. Clic sur le bouton pour charger les variables des contextes



Utilisation des variables de contexte

1. Clic sur ctrl espace
Afin d'afficher les
variables des contexte

The screenshot displays the Talend Studio interface. At the top, a data flow diagram shows a 'Produits' component connected to a 'Produits(Lookup)' component via a dashed orange line. Below the diagram, the 'Designer' tab is active, showing the configuration for the 'Produits(tFileInputDelimited_3)' component. The 'Basic settings' tab is selected, and the 'File name/Stream' field is highlighted in yellow, containing the expression 'context.input + "Produits.csv"'. Other settings include 'Property Type' set to 'Built-In', 'Schema' set to 'Repository', and 'DELIM:Produits - metadata'. The 'CSV Row Separator' is 'LF("\n")', 'Field Separator' is ',', 'CSV options' are checked, 'Escape char' is '\\', 'Header' is 1, and 'Footer' is 0. The 'Skip empty rows', 'Uncompress as zip file', and 'Die on error' options are unchecked.

Produits(Lookup)

Produits

Designer Code

Job(J06_VentesProduits_CT 0.1) Contexts(J06_VentesProduits_CT) Component

Produits(tFileInputDelimited_3)

Basic settings

Property Type Built-In

Schema Repository DELIM:Produits - metadata

File name/Stream context.input + "Produits.csv"

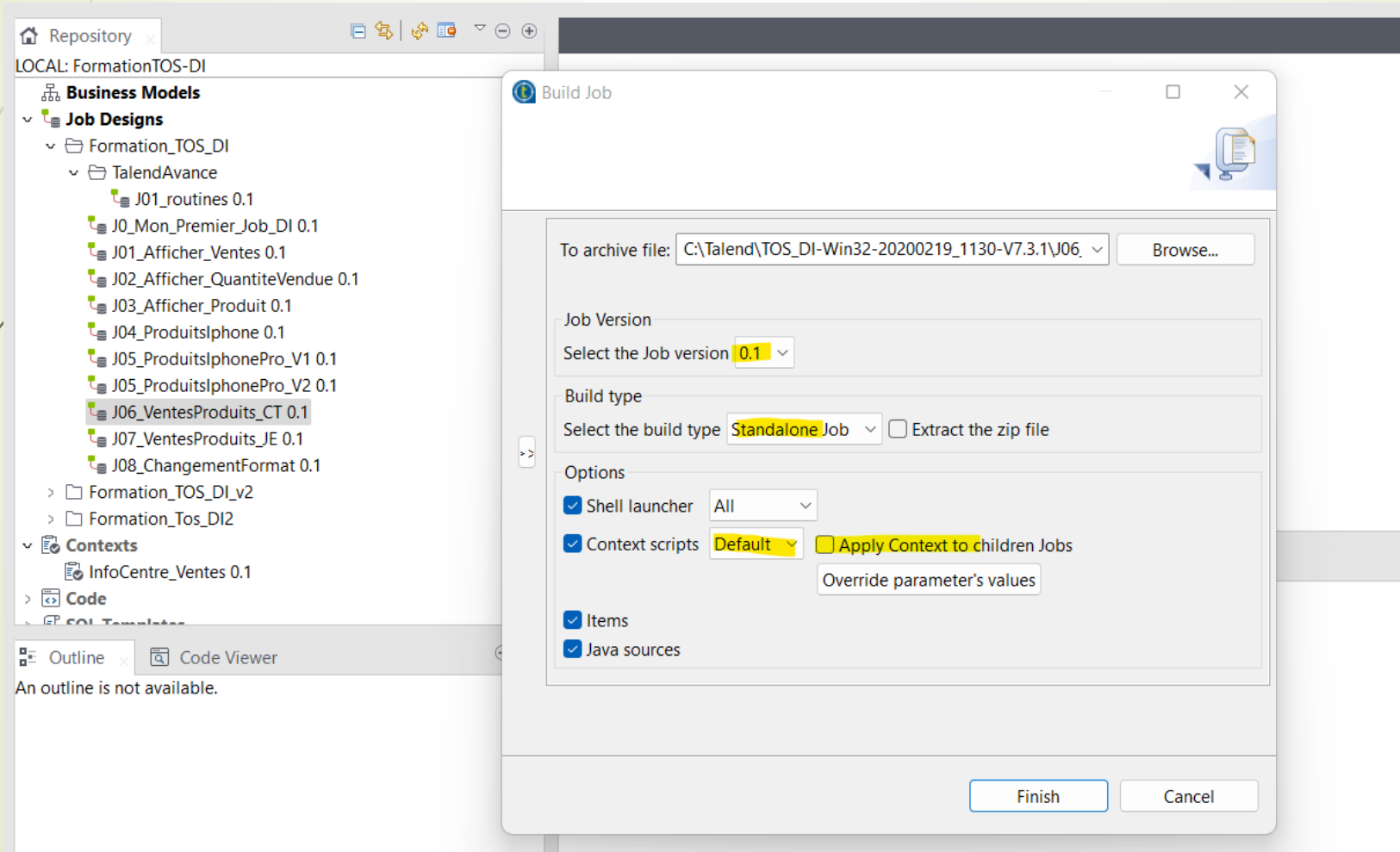
CSV Row Separator LF("\n") Field Separator ,

☒ CSV options Escape char \"

Header 1 Footer 0

☐ Skip empty rows ☐ Uncompress as zip file ☐ Die on error

Utilisation des variables de contextes : exécution du code généré





”

