

Renouvaud

Réseau vaudois
des bibliothèques

Introduction au logiciel Talend

Talend Open Studio for Data Integration



Août 2023 - Journée LibreABC - Genève

Noémie Payot
Coordination Renouvaud

Talend Data Fabric

Intégration de données →

Unifiez toutes vos données avec des outils simples de collecte, transformation et mapping de données

Data Quality →

Nettoyez et profilez automatiquement des données en temps réel afin de ne pas vous soucier des données de mauvaise qualité entrant dans vos systèmes

Intégrité et gouvernance des données →

Garantissez la fiabilité et l'excellence des données tout au long de leur cycle de vie et évaluez leur qualité avec le Talend Trust Score™

Intégration d'API et d'applications →

Partagez et générez de la valeur à partir de données fiables en interne et en externe, grâce à une multitude de fonctionnalités en self-service

<https://www.talend.com/fr/products/data-fabric/>

Talend Open Studio for Data Integration

Outil ETL : Extract transform and Load

Fonctionnalités :

- Gestion de fichiers sans script : ouvrir, déplacer, supprimer, décompresser
- Mapping, fusion et transformation de données, tri, dédoublonnage, etc.
- Connexion serveur, import de tables, appels API

<https://www.talend.com/fr/products/talend-open-studio/>


Plan de la présentation

- Exemple d'utilisation
 - Transformation d'un export BiblioMaker pour permettre l'import des données dans Alma
- L'interface de Talend
- Démonstration et exercices
 - Création d'un job simple : logique et chaîne des composants
 - Utilisation des métadonnées : créer un modèle de fichier délimité
 - Quelques composants simples : tFileInputDelimited, tMap, tLog, tReplace, tFilterRow, tFileOutputDelimited
 - Gestion des erreurs et documentation

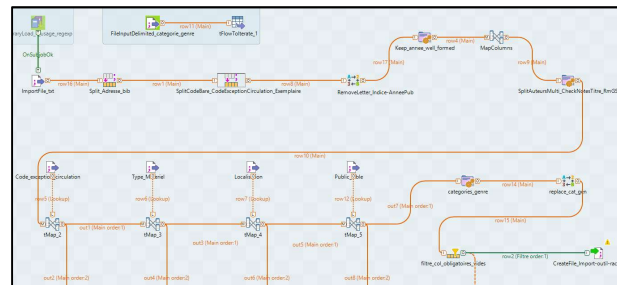
Exemple d'utilisation

Transformation d'un fichier avec des données bibliographiques pour permettre d'importer ces dernières dans Alma

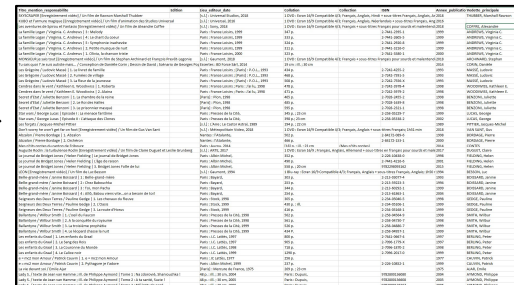
Export BiblioMaker



Job Talend



Fichier transformé



Export BiblioMaker

Numero	Type	Code masque	Jour	Mois	Annee	Cote	Titre	Adresse bibliographique	Zone de notes	ISBN	Annee de publication	Vedette	Exemplaires code barre duree prêt quantité
5330	Monographie	C1	22	2	2005	789.0	Cirque Knie : Une dynastie du cirque	Librairie Marguerat Journée LibreABC - Genève			1975	HAWS, Duncan	CV000D31/30/1
134	Monographie	C1	10	3	2006	771	La fête des vigneron Vevey 1977	Editions Payot Journée LibreABC - Genève	prêt réservé		199	HAWS, Duncan	CV00018F/30/1
247	BD	C1	11	4	2007	772	Guide de la technique 3 : L' énergie Les bateaux et la mer : Histoire de la navigation et des combats sur mer	Presses polytechniques et universitaires romandes		28807422	1993	HAWS, Duncan	CV0001C3/30/1
558	BD	C1	12	5	2008	773		Editions Plantyn			1975	HAWS, Duncan	CV0001LU/30/1
1080	BD	C1	13	6	2009	774	Un siècle à toute vapeur : Chemins de fer des Rochers-de-Naye : 1892-1992	Groupe MOB, Clarens, 1992	gratuit		1992	HAWS, Duncan STYGER, EdgarKOLLROS, Jean- Charles	CV00026L/30/1
1081	BD	C1	14	7	2010	775	A l'assaut du roc de Naye	Groupe MOB, Clarens	gratuit		1985	STYGER, EdgarWIDMER, RobertKOLLROS, Jean- Charles	CV00026M/30/1
1082	BD	C1	15	8	2011	776	MOB : 75 ans Jahre : 1901 - 1976	Groupe MOB, Clarens	gratuit		1983	STYGER, EdgarWIDMER, RobertKOLLROS, Jean- Charles	CV00026N/30/1
4065	BD	C1	28	9	2002	777	L'ABCdaire du Chocolat : Le chocolat blanc, Le chocolat noir	Flammarion, Paris	reserve	208012479x	1997	KHODOROWSKY, KatherineROBERT, Hervé	CV0008SI/30/1
4066	BD	C1	28	10	2002	778	CH comme Chocolat	Slatkine, Journée LibreABC - Genève	réserve	2051018618	2001	BOUGARD, Alain J.	CV0008SK/30/1
4344	BD	C1	2	11	2003	779	Les Marionnettes	Bordas, Paris	réserve	204163964	1982		CV0009R8/30/1
4345	Article	C3	3	12	2004	80	Venedig Menschen hinter Masken	Benteli		371650580	204		CV0009R9/30/1
4347	Article	C3	4	1	2005	80	Encyclopédie des instruments de musique	Gründ, Paris		2700013166	1980		CV0009RC/30/1
4348	Article	C3	5	1	2006	80	Les fêtes des vigneron de Vevey	Editions Hermès R.	couverture abîmée		1956		CV0009RD/30/1
4350		C4	6	5	2007		DVD jeune Le monde fascinant des bateaux	Gründ, Paris		2700054059	1977		CV0009RH/30/1
4352	DVD	C4	1	6	2008		DVD ado Instruments de musique	Gallimard, Paris	don	2070564401	1989		CV0009RL/30/1
5070	Manga	C1	28	8	2003		DVD jeune Le livre du chocolat	Flammarion		2082004988	2		CV000C00/30/1
5498	DVD	C4	8	9	2005		DVD jeune Le café	Nestlé SA, Vevey CH	livre reserve		1980		CV000DHF/30/1
7767	DVD	C4	15	9	2014		DVD jeune Victor Desarzens : The musician and the man, Portrait 1908-1986	Nestlé, Vevey			1992	PITTIER, Jacques- Michel	CV000JIO/30/1
2504	DVD	C4	15	10	1999		DVD jeune Harrap's Shorter : Dictionnaire Anglais- Français / Français-Anglais	Harrap, 1996		024550351x	1996		CV0003JE/30/1

Données nettoyées et transformées

Numero	Titre_mention_responsabilite	Lieu_editeur_date	Collation	ISBN	Annee_publication	Vedette_principale	Code_barres	Type_Materiel	Code_exception_circulation	Date_reception
8466	Plan d'études général pour les collèges et les gymnases, ainsi que pour les écoles supérieures de jeunes filles du canton de Vaud, du 30 décembre 1909	Journée LibreABC - Genève : Impr. Giesser et Held, 1910	102 p. ; 21 cm		1910	VAUD. DEPARTEMENT DE L'INSTRUCTION PUBLIQUE ET DES CULTES	07818	BOOK	3000	
9558	Plaidoyer pour la forêt tropicale : sommet de la diversité / Francis Hallé	Arles : Actes Sud, 2014	210 p. : ill. ; 25 cm	9782330021467	2014	HALLE, Francis	EN000NC K	BOOK	3000	4.7.2014
9559	Le radeau des cimes / Francis Hallé	Paris : Bayard, 2013	61 p. ; 18 cm	9782227486447	2013	HALLE, Francis	EN000NIL	BOOK	3000	3.9.2014
9578	Manifeste pour les grands singes / Christophe Boesch, Emmanuelle Grundmann, Blaise Mulhauser	Journée LibreABC - Genève : Presses polytechniques et universitaires romandes, 2011	143 p. : ill. ; 18 cm	9782880749149	2011	BOESCH, Christophe; GRUNDMANN, Emmanuelle; MULHAUSER, Blaise	EN000IV5	BOOK	3000	15.11.2011
9581	De la fourmi domestique à la fourmi des bois : petit guide myrmécologique à travers la ville de Journée LibreABC - Genève / Daniel Chérix, Cécile Maddalena-Feller	Journée LibreABC - Genève : Service des forêts, domaines et vignobles, 1988	42 p. : ill. ; 30 cm		1988	CHERIX, Daniel; MADDALENA-FELLER, Cécile	11975	BOOK	3000	
9588	La vie des fourmis / Laurent Keller, Elisabeth Gordon	Paris : Odile Jacob, 2006	303 p. : ill. ; 22 cm	9782738118257	2006	KELLER, Laurent; GORDON, Elisabeth	EN000GS P	BOOK	3000	27.9.2010
9590	Au fond du zoo à droite : découvertes récentes et intéressantes sur le règne animal / Edouard Launet	Paris : Ed. du Seuil, 2012	169 p. : ill ; 20 cm	9782757829752	2012	LAUNET, Edouard	EN0011D A	BOOK	3000	9.1.2017
11392	Vie et mystère des mots / Maurice Bossard	Morges : Editions Cabédita, 1990	205 p. : ill. ; 22 cm		1990	BOSSARD, Maurice	16406	BOOK	3000	23.4.1990
11397	Parlons vaudois / Jean-Pierre Cuendet	Chavannes-de-Bogis : Editions Slatkine, 1991	189 p. ; 22 cm		1991	CUENDET, Jean-Pierre	06591	BOOK	3000	18.8.1993
11399	Les formes de l'écriture impliquée / dossier coordonné par Kareen Illiade & Remi Hess	Paris : Téraèdre, 2009	131 p. : ill. ; 21 cm	9782912868824	2009	ILLIADE, Kareen; HESS, Remi	EN000GU 7	BOOK	3000	30.9.2010
11403	Langues et territoires / [directeur: Yves Lacoste ; auteurs: Béatrice Giblin ... et al.]	Paris : La Découverte, 2002	200 p. : ill. ; 20 cm	270713757X	2002	LACOSTE, Yves; GIBLIN, Béatrice	EN0005S2	BOOK	3000	14.6.2002
11405	Le langage des Vaudois / Henri Perrochon ; préf. et glossaire de Jean-Pierre Chuard	Journée LibreABC - Genève : Editions 24 Heures, 1979	149 p. ; 15 cm		1979	PERROCHON, Henri; CHUARD, Jean-Pierre	04725	BOOK	3000	22.4.1980
11408	Le patois vaudois : grammaire et vocabulaire / Jules Reymond, Maurice Bossard	Journée LibreABC - Genève : Payot, 1979	263 p. ; 21 cm		1979	REYMOND, Jules; BOSSARD, Maurice	10675	BOOK	3000	
11409	Le langage : introduction à l'étude de la parole / Edward Sapir ; (trad. de l'ang. par S.M. Guillemin)	Paris : Payot, 1970	231 p. ; 18 cm		1970	SAPIR, Edward; GUILLEMIN, S.M	04733	BOOK	3000	26.2.1980
11411	L'oreille et le langage / Alfred Tomatis	Paris : Seuil, 1991	186 p. : ill. ; 18 cm		1991	TOMATIS, Alfred	00424	BOOK	3000	19.5.1992
11412	Le bégaiement : la parole désorchestrée / Elisabeth Vincent	Paris : Editions Milan, 2004	63 p. : ill. ; 18 cm	2745915088	2004	VINCENT, Elisabeth	EN0009S2	BOOK	3000	16.11.2004
11414	Catalogue des idées reçues sur la langue / Marina Yaguello	Paris : Editions du Seuil, 1988	157 p. : ill. ; 18 cm	2020099136	1988	YAGUELLO, Marina	EN000AL9	BOOK	3000	8.9.2005

Pourquoi Talend et non Excel ?

Avantages de Talend

- Grand gain de temps si on doit transformer régulièrement des fichiers reçus dans un format A vers un format B.
- Plus lisible que des macros Excel et facile à réutiliser par d'autres personnes.
- Liberté presque complète, car basé sur le langage Java : si aucun composant ne correspond à nos attentes, possibilité d'en programmer un.

Désavantage

- Logiciel complexe à prendre en main

Présentation de l'interface

Référentiel

The screenshot displays the Talend Studio interface. On the left, a tree view shows the project structure under 'Référentiel' and 'LOCAL: Talend3'. The main workspace shows a job design for 'Job ouvrir_fichier 0.1' with components: A_ExportBiblioMaker, tMap_1, tReplace_1, tFilterRow_1, and tFileOutputDelimited_1. The 'Designer' tab is active, showing the configuration for 'A_ExportBiblioMaker(tFileInputDelimited_1)'. The configuration includes parameters for file path, line and field separators, and CSV options.

Paramètres simples

Type de propriété	Built-In
Schéma	Built-In
Nom de fichier/Flux	"C:/Users/npayot/Desktop/importBiblioMaker.txt"
Séparateur de lignes	"\n"
Séparateur de champs	"\t"
Options CSV	<input type="checkbox"/>
En-tête	1
Pied de page	0
Limite	
Ignorer les lignes vides	<input type="checkbox"/>
Décompresser en tant que fichier zip	<input type="checkbox"/>
Arrêter en cas d'erreur	<input type="checkbox"/>

Espace de modélisation graphique

The screenshot displays the Talend graphical modeling environment. The main workspace shows a workflow for 'Job ouvrir_fichier 0.1' with the following components: A_ExportBiblioMaker (row1 (Main)), tMap_1 (out1 (Main)), tReplace_1 (row2 (Main)), tFilterRow_1 (row3 (Filter order:1)), tFileOutputDelimited_1 (row4 (Reject order:2)), and tFileOutputDelimited_2. The 'Designer' tab is active, showing the configuration for 'A_ExportBiblioMaker(tFileInputDelimited_1)'. The configuration includes:

- Type de propriété: Built-In
- Paramètres avancés: Schéma Built-In, Modifier le schéma
- View: "When the input source is a stream or a zip file, footer and random shouldn't be bigger than 0."
- Documentation: Nom de fichier/Flux "C:/Users/npayot/Desktop/importBiblioMaker.txt"
- Séparateur de lignes: "\n", Séparateur de champs: "\t"
- 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:

The left sidebar shows the project structure under 'Business Models' and 'Jobs', with 'ouvrir_fichier 0.1' selected. The right sidebar contains a 'Palette' with a search bar and a list of categories including Favoris, Récemment utilisé, Applications Métier, Bases de données, Big Data, Business Intelligence, Business, Cloud, Code Utilisateur, Databases, Divers, DotNET, ELT, ESB, Fichier, Gestion, Lecture, NamedPipe, Écriture, Internet, Logs & Erreurs, Orchestration, Qualité de données, Système, Technique, Transformation, Unstructured, XML, and Misc.

Palette des composants

The screenshot displays the Talend Open Studio interface. On the left, a tree view shows the project structure under 'Business Models' and 'Jobs'. The main workspace shows a job design for 'Job ouvrir_fichier 0.1' with components: A_ExportBiblioMaker, tMap_1, tReplace_1, tFilterRow_1, tFileOutputDelimited_1, and tFileOutputDelimited_2. The 'Designer' tab is active, showing the configuration for 'A_ExportBiblioMaker(tFileInputDelimited_1)'. The configuration includes parameters for file path, line and field separators, and CSV options. On the right, the 'Palette' window is open, listing various component categories such as 'Favoris', 'Récemment utilisé', 'Applications Métier', 'Bases de données', 'Big Data', 'Business Intelligence', 'Business', 'Cloud', 'Code Utilisateur', 'Databases', 'Divers', 'DotNET', 'ELT', 'ESB', 'Fichier', 'Gestion', 'Lecture', 'NamedPipe', 'Écriture', 'Internet', 'Logs & Erreurs', 'Orchestration', 'Qualité de données', 'Système', 'Technique', 'Transformation', 'Unstructured', 'XML', and 'Misc'.

Onglet de configuration des composants

The screenshot shows the Talend Studio interface. On the left is a project tree with 'Jobs' > 'ouvrir_fichier 0.1' selected. The main canvas displays a workflow with components: A_ExportBiblioMaker, tMap_1, tReplace_1, tFilterRow_1, tFileOutputDelimited_1, and tFileOutputDelimited_2. A configuration window for 'A_ExportBiblioMaker(tFileInputDelimited_1)' is open, showing the following settings:

- Type de propriété: Built-In
- Schéma: Built-In
- Message: "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/npayot/Desktop/importBiblioMaker.txt"
- Séparateur de lignes: "\n" | Séparateur de champs: "\t"
- 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:

On the right, a component palette lists various categories such as 'Fichiers', 'Gestion', 'Lecture', 'NamedPipe', 'Écriture', 'Internet', 'Logs & Erreurs', 'Orchestration', 'Qualité de données', 'Système', 'Technique', 'Transformation', 'Unstructured', 'XML', and 'Misc'.

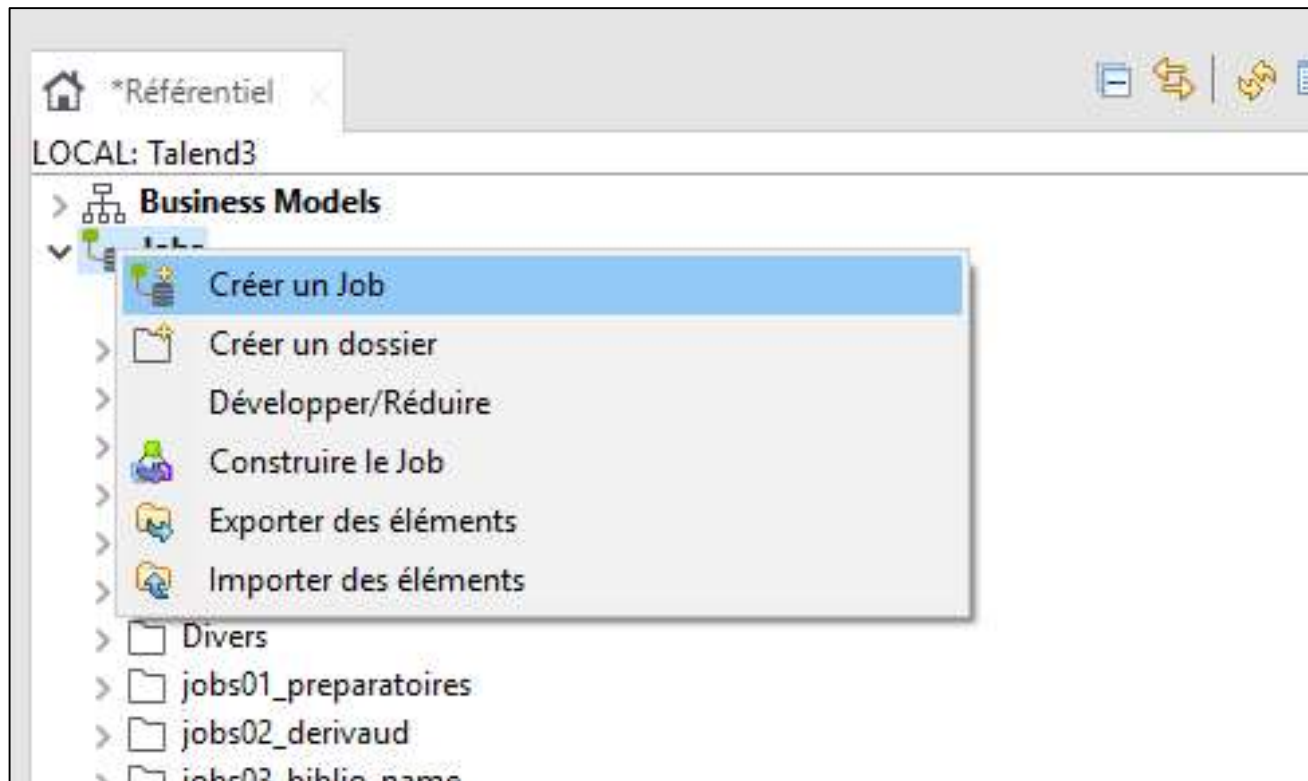
Onglet d'exécution

The screenshot shows the Talend Studio interface with the execution tab selected for the job 'Job ouvrir_fichier 0.1'. The main workspace displays a job design with several components: A_ExportBiblioMaker, tMap_1, out1 (Main), tReplace_1, row2 (Main), tFilterRow_1, row3 (Filtre order:1), tFileOutputDelimited_1, and tFileOutputDelimited_2. The execution tab is highlighted with a red box and contains the following elements:

- Job ouvrir_fichier**: Title of the job.
- Exécution simple**: Tab selected for simple execution.
- Exécution**: Sub-tab with buttons for **Exécuter** (highlighted with a red box), **Arrêter**, and **Effacer**.
- Paramètres avancés**: Tab for advanced parameters.
- Exéc distante**: Tab for remote execution.
- Exéc. test mémoire**: Tab for memory test execution.
- Table**: A table with columns 'Nom' and 'Valeur' for monitoring execution details.
- Options**: Checkboxes for 'Nombre limite de lignes' (set to 100) and 'Retour automatique à la li'.

Etape 1

Créer un job



Décrire le job

Nouveau Job

⚠ Il est recommandé de ne pas laisser le champ Description vide.

Nom: generer_lignes

Objectif: générer 2 lignes aléatoires et les afficher

Description:

Créé par: user@talend.com

Verrouillé par:

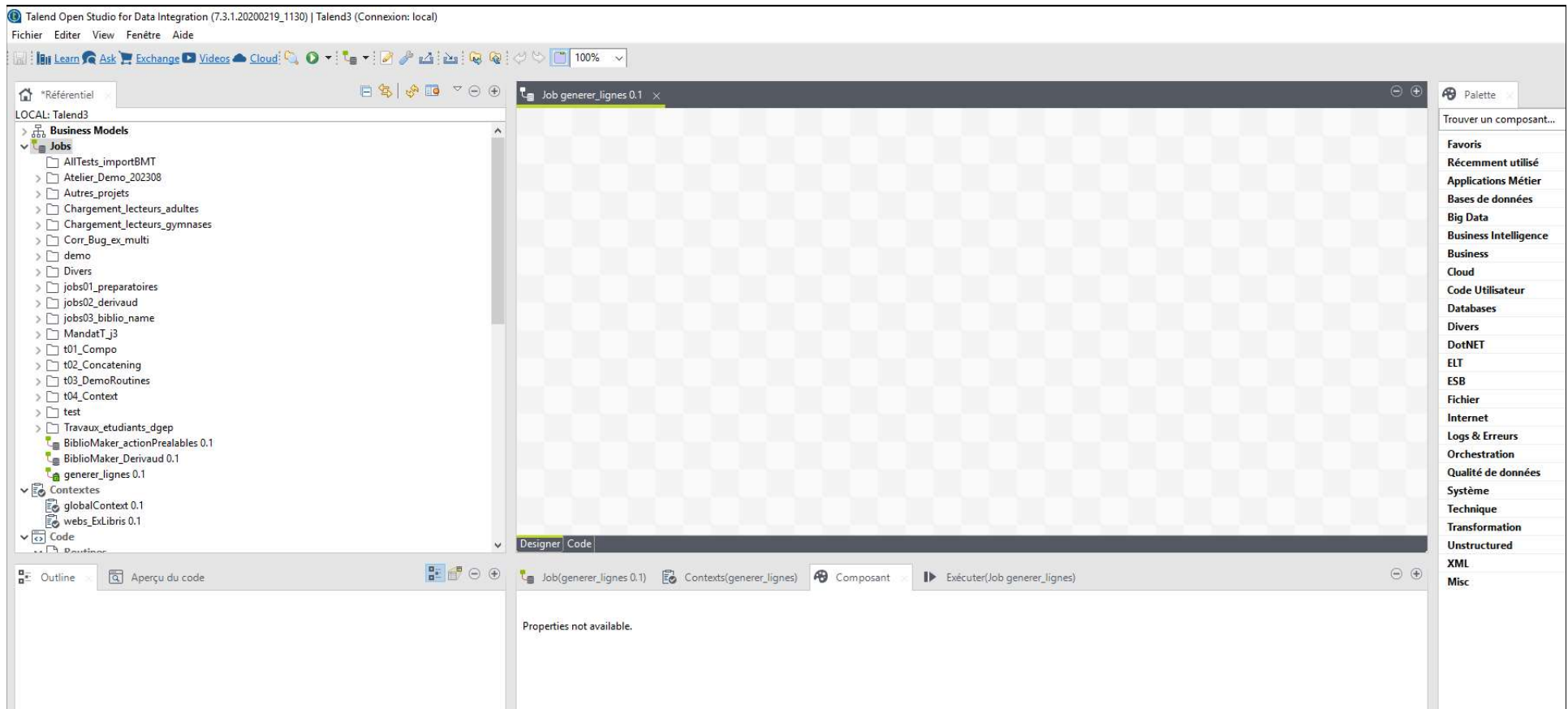
Version: 0.1 M m

Statut:

Chemin d'accès: Sélectionner

Finish Cancel

Interface prête pour ajouter les composants



Etape 2

- Ajouter les composants
- Lier les composants
- Paramétrer chacun des composants

The screenshot displays the Talend Studio interface. On the left, the 'Business Models' tree shows a job named 'ouvrir_fichier_temp 0.1'. The main workspace shows a job design with three components: 'A_ExportBiblioMaker', 'row1 (Main)', and 'tLogRow_1'. The 'tLogRow_1' component is selected, and its properties are shown in the bottom panel. The 'Paramètres simples' section is expanded, showing the 'Mode' set to 'Tableau (afficher les valeurs dans les cellules d'un tableau)' and the checkbox 'Afficher le contenu avec Log4j' checked. The 'tLogRow_1' component is also visible in the 'Composant' palette on the right.

Exercice introductif pour appliquer l'étape 2

- Ouvrir un fichier tsv (valeurs séparées par tabulations)
- Afficher le contenu de ce fichier

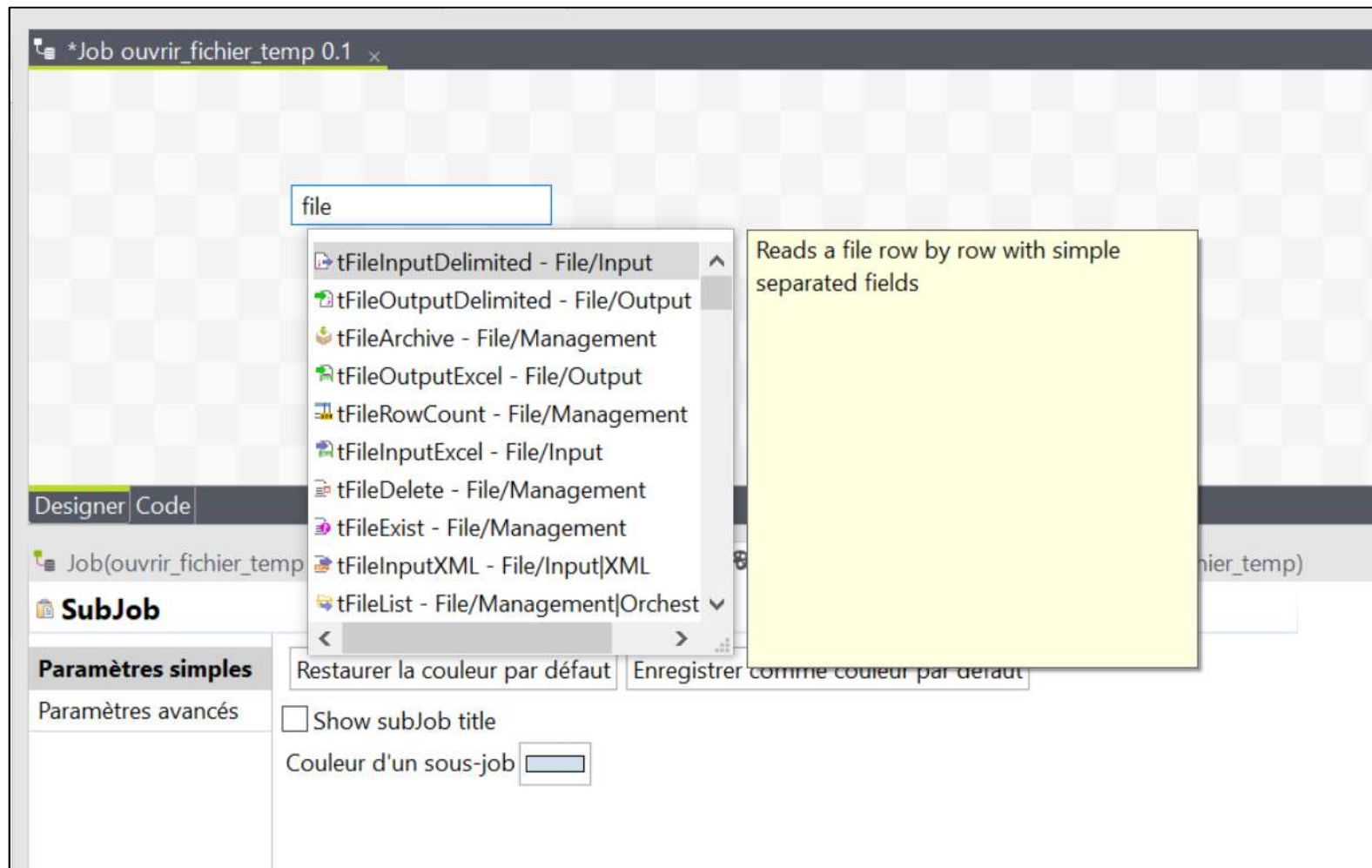
The screenshot shows the Talend Studio interface. On the left, the 'Business Models' tree is expanded to 'Jobs' > 'Atelier_Demo_202308' > 'ouvrir_fichier_temp 0.1'. The main workspace shows a job graph with two components: 'A_ExportBiblioMaker' and 'tLogRow_1', connected by a line labeled 'row1 (Main)'. Below the workspace, the 'Designer' tab is active, showing the configuration for the 'tLogRow_1' component. The 'Paramètres simples' section is expanded, showing the 'Mode' section with the following options:

- Mode
- Basique
- Tableau (afficher les valeurs dans les cellules d'un tableau)
- Vertical (chaque ligne est une liste de clé/valeur)
- Afficher le contenu avec Log4j

The 'Schéma' section is set to 'Built-In' and 'Sync colonnes' is checked. The 'Code' tab is also visible below the Designer tab.

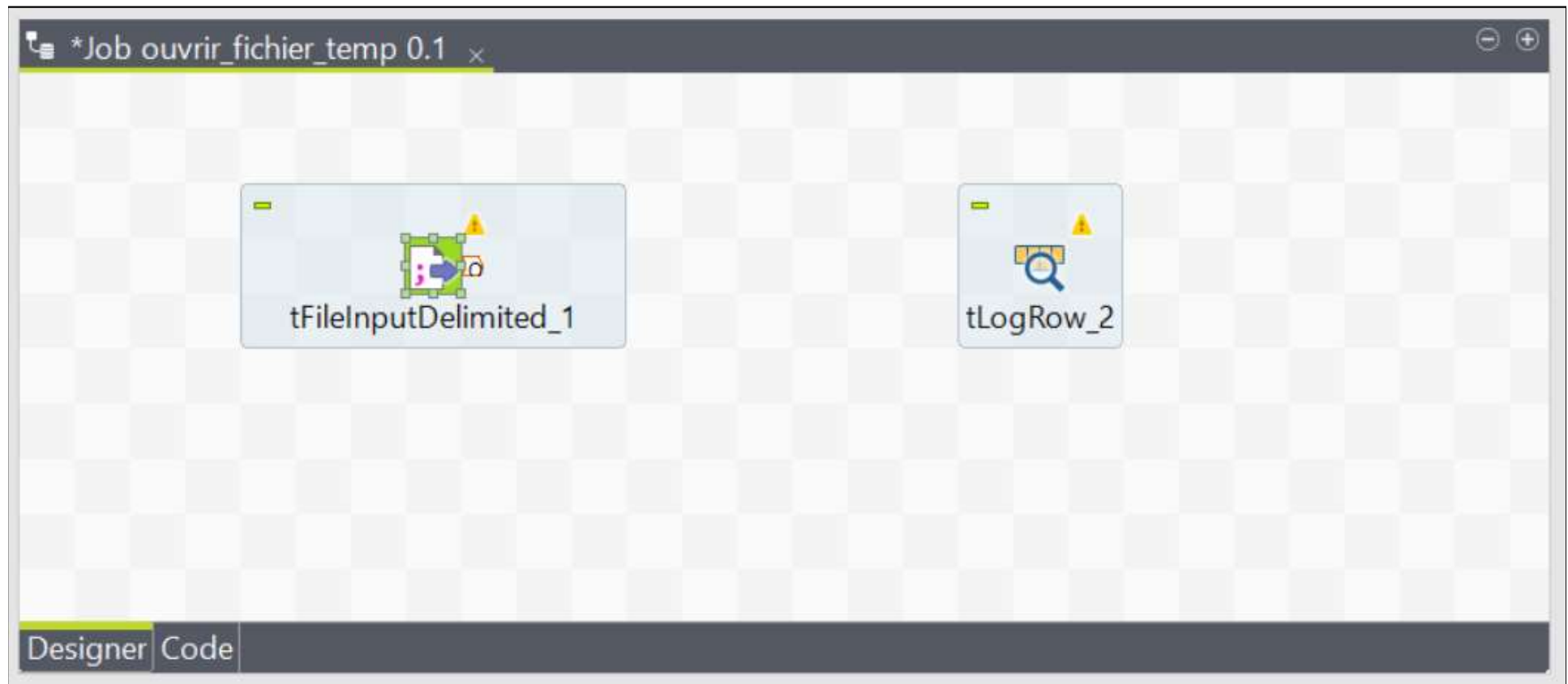
Comment ajouter un composant?

Ecrire le nom du composant dans l'espace de modélisation graphique (pas besoin de cliquer)

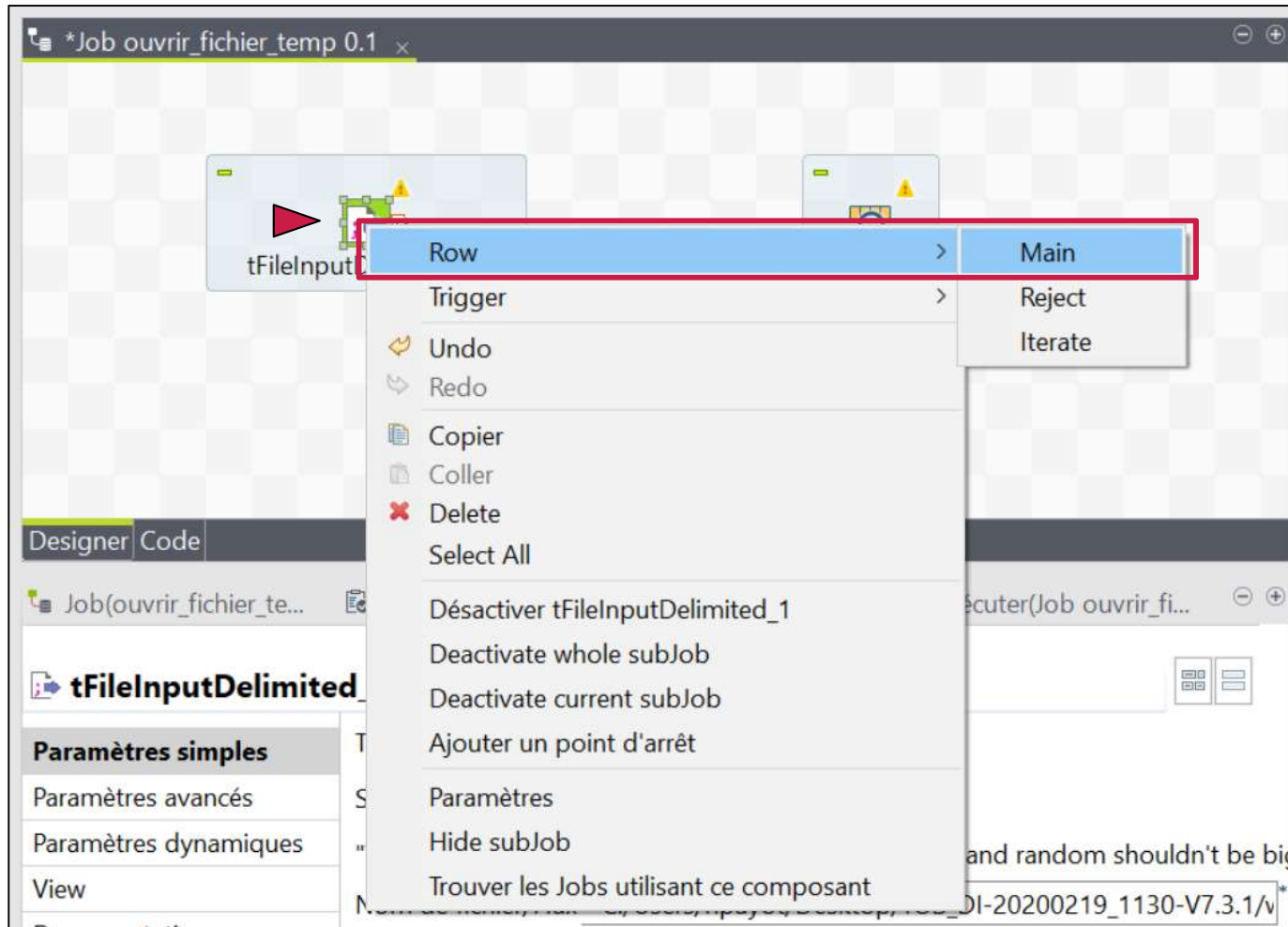


Ajouter les composants

- tFileInputDelimited
- tLogRow

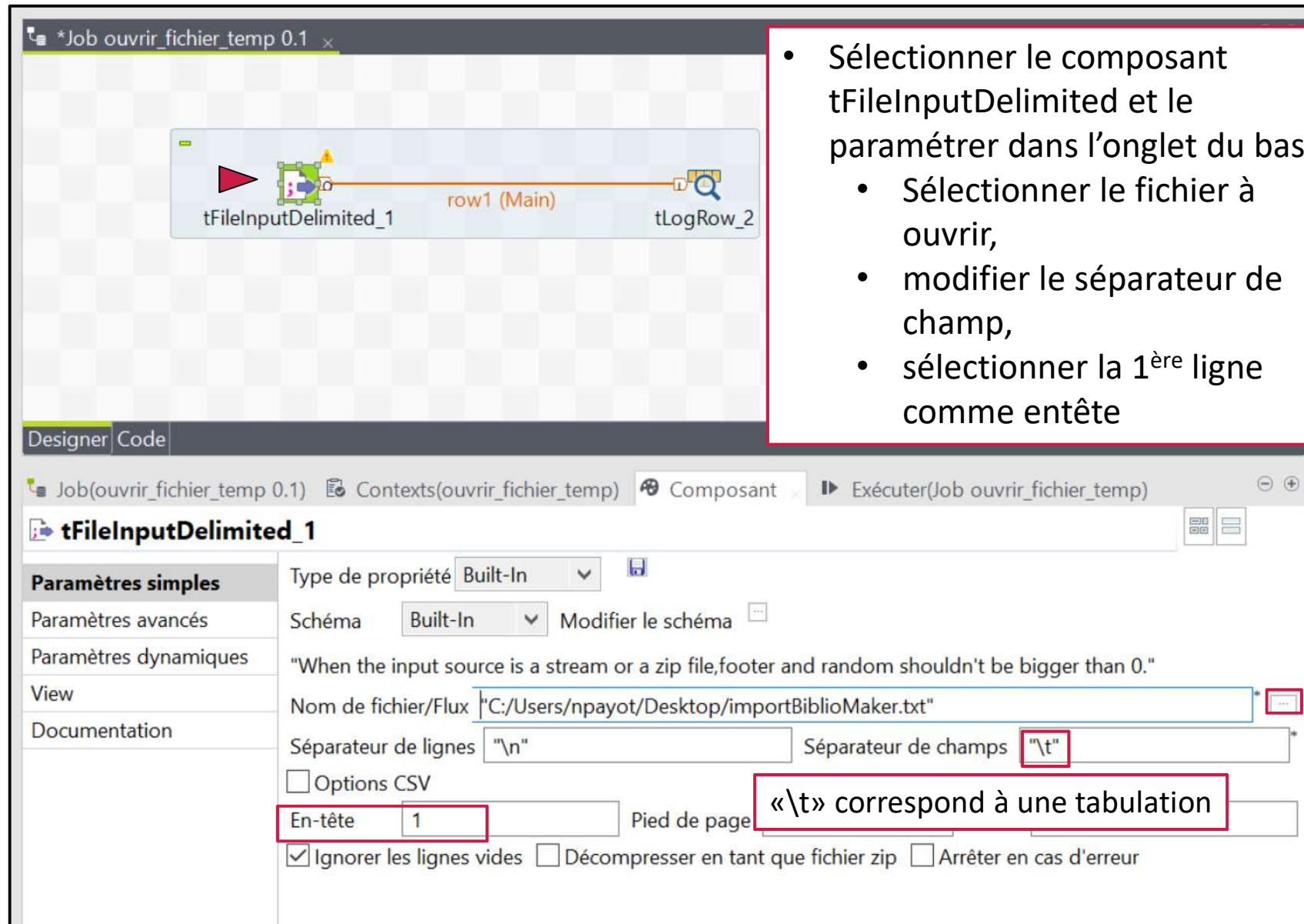


Lier les composants entre eux par un flux



- Faire un clic droit sur le composant source
- Sélectionner > Row > Main et cliquer sur le composant cible

Paramétrer le composant tFileInputDelimited 1/2




The screenshot displays the Talend Studio interface. At the top, a workflow diagram shows a red triangle icon for the `tFileInputDelimited_1` component, connected to a `tLogRow_2` component via a connector labeled `row1 (Main)`. Below the diagram, the configuration panel for `tFileInputDelimited_1` is visible. The panel includes a sidebar with tabs for **Paramètres simples**, Paramètres avancés, Paramètres dynamiques, View, and Documentation. The main configuration area shows the following settings:

- Type de propriété: Built-In
- Schéma: Built-In
- Nom de fichier/Flux: "C:/Users/npayot/Desktop/importBiblioMaker.txt"
- Séparateur de lignes: "\n"
- Séparateur de champs: "\t"
- Options CSV:
- En-tête: 1
- Pied de page:
- Ignorer les lignes vides:
- Décompresser en tant que fichier zip:
- Arrêter en cas d'erreur:

A red box highlights the `\t` separator, with a callout stating: «\t» correspond à une tabulation.

- Sélectionner le composant `tFileInputDelimited` et le paramétrer dans l'onglet du bas
 - Sélectionner le fichier à ouvrir,
 - modifier le séparateur de champ,
 - sélectionner la 1^{ère} ligne comme entête

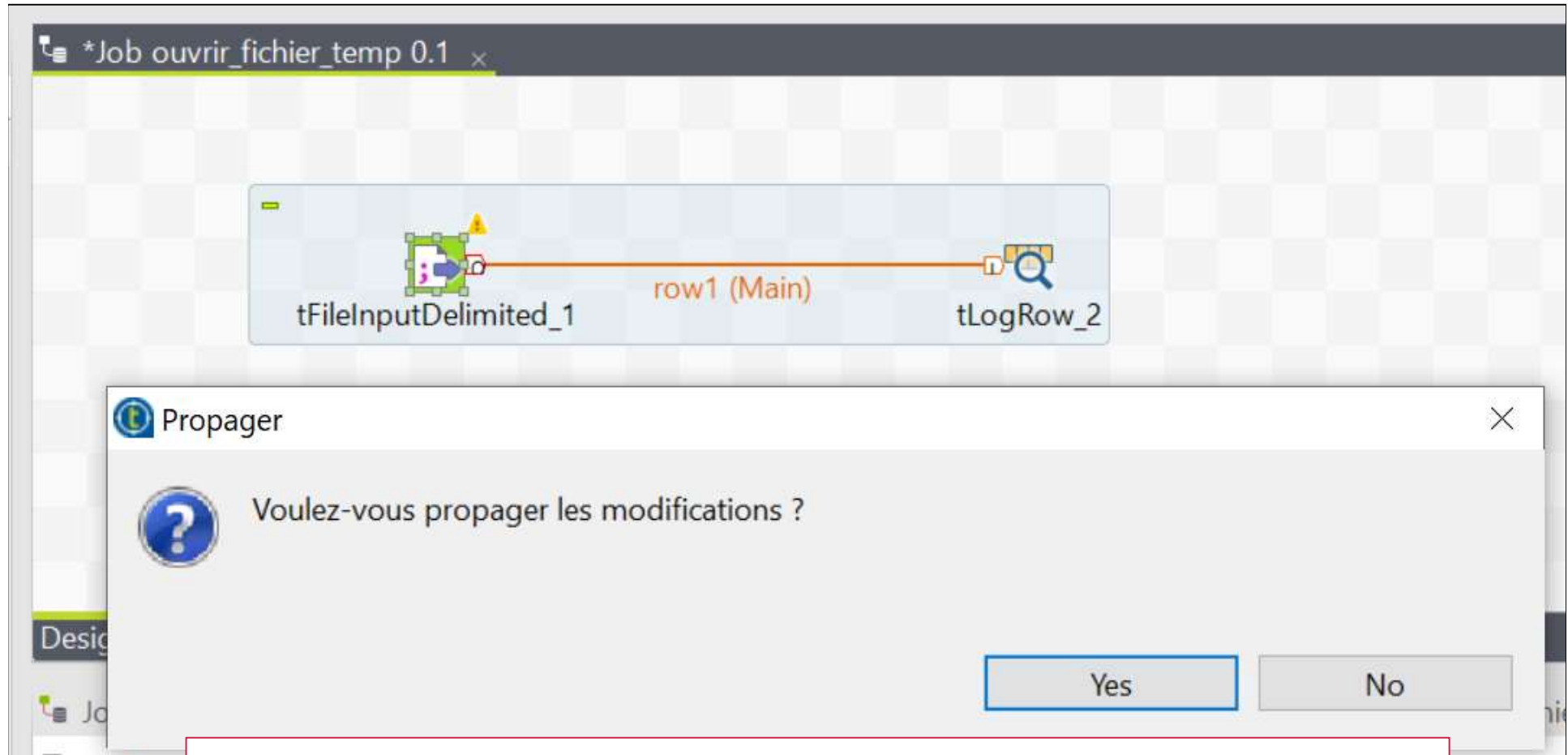
Paramétrer le composant tFileInputDelimited 2/2

• Cliquer sur les  pour modifier le schéma

• Dans la fenêtre qui s'ouvre, cliquer sur «+» pour ajouter 2 lignes

Colonne	C...	Ty...	N.	Modè...	Le...	Pr...	D...	C...
Numero	<input type="checkbox"/>	St...	<input checked="" type="checkbox"/>					
Type	<input type="checkbox"/>	St...	<input checked="" type="checkbox"/>					

Accepter de propager les modifications



Cliquer sur OK afin que les deux colonnes créées précédemment soit propagées par le flux de données *row1* vers le composant tLogRow

Exécuter le job

Job ouvrir_fichier_temp 0.1

19 rows in 0.02s
863.64 rows/s
tFileInputDelimited_1 → tLogRow_2

Designer Code

Job(ouvrir_fichier_temp 0.1) Contexts(ouvrir_fichier_temp) Composant ▶ Exécuter(Job ouvrir_fichier_temp)

Job ouvrir_fichier_temp

Exécution simple

Exécution Debug

Paramètres avancés

Exéc distante

Exéc. test mémoire

Exécution

▶ Exécuter Arrêter Effacer

Nom	Valeur
4065	BD
4066	BD
4344	BD
4345	Article
4347	Article
4348	Article
4350	
4352	DVD
5070	Manga
5498	DVD
7767	DVD
2504	DVD

Seules les 2 premières colonnes du fichier apparaissent car les autres n'ont pas été ajoutées lors du paramétrage de tFileInputDelimited à l'étape 2.

Exercice 2

Créer un modèle de métadonnées et l'utiliser dans le job précédemment créé

Avantages :

- Si le fichier source contient beaucoup de colonnes, il n'est pas nécessaire de toutes les ajouter manuellement.
- On crée un modèle de fichier réutilisable dans plusieurs jobs

Importer les métadonnées d'un fichier tsv

Depuis le bloc Référentiel :

- Ouvrir Métadonnées > Fichier délimité > Créer un fichier délimité

Nommage du fichier et description

Nouveau fichier délimité

Fichier - Etape 1 de 4

Ajouter un fichier de métadonnées au référentiel
Définir les propriétés

Nom: ExportBiblioMaker1

Objectif: Export de données bibliographiques à partir de BiblioMaker

Description: Export de données bibliographiques à partir de BiblioMaker

Créé par: user@talend.com

Verrouillé par:

Version: 0.1 M m

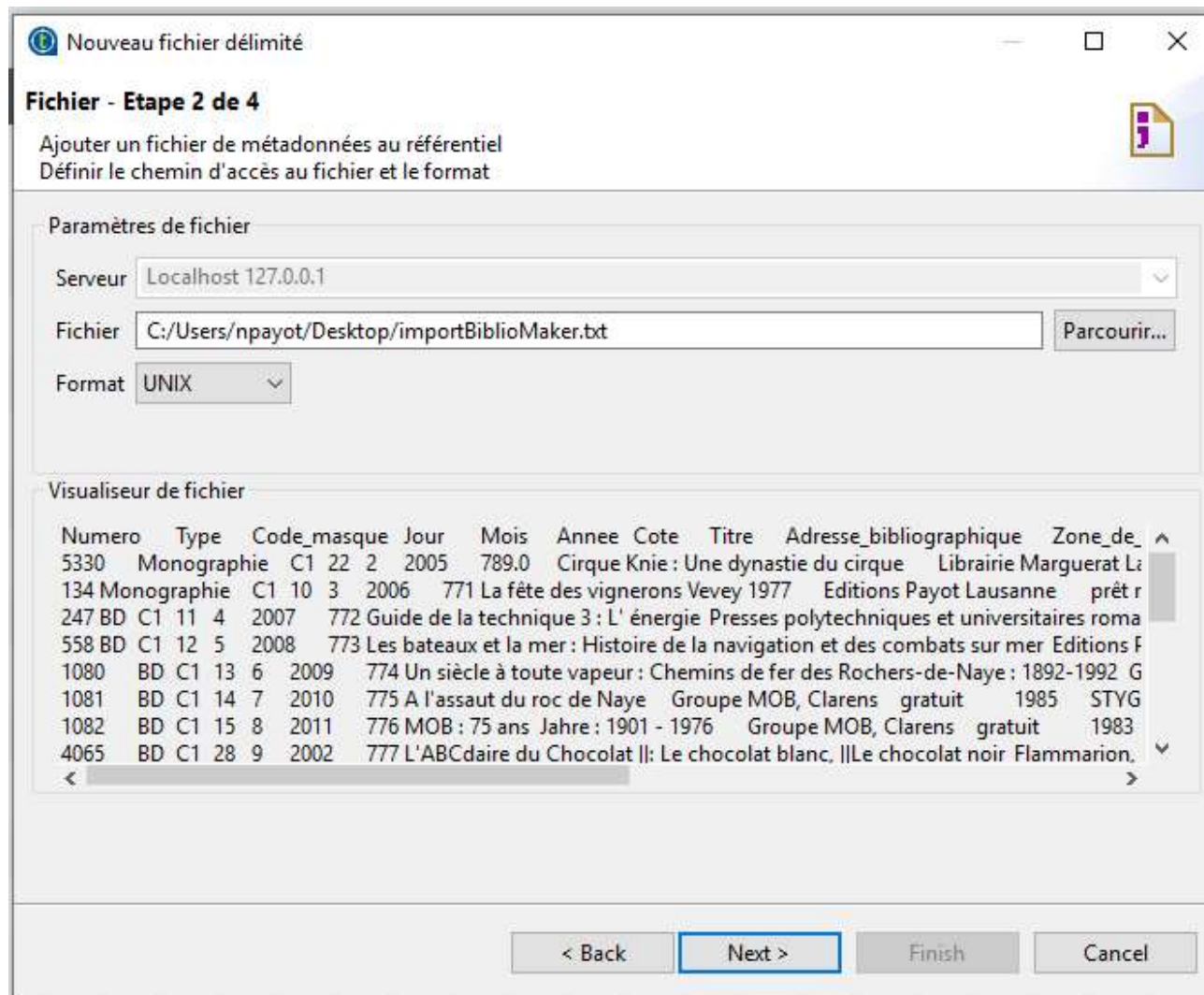
Statut:

Chemin d'accès: Sélectionner

< Back Next > Finish Cancel

Sélection du fichier

!Attention le fichier doit avoir des entêtes



Paramétrer les options selon le fichier choisi

Nouveau fichier délimité

Fichier - Etape 3 de 4

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

Paramètres de fichier

Encodage: UTF-8

Séparateur de champs: Tabulation (highlighted) Caractère correspondant: "\t"

Séparateur de lignes: Standard E Caractère correspondant: "\n"

Paramètres du caractère d'échappement

CSV Délimité

Caractère d'échappement: Vide

Entourage du texte: Vide

Découper la ligne avant le champ

Lignes à ignorer

Si des lignes doivent être ignorées, s

En-tête: 1

Pied de page:

Ignorer les lignes vides

Limite de lignes

Si le nombre de lignes doit être limit

Limite:

Aperçu (highlighted) Sortie

Définir les lignes d'en-tête comme nom de colonnes (highlighted)

< Back Next > Finish Cancel

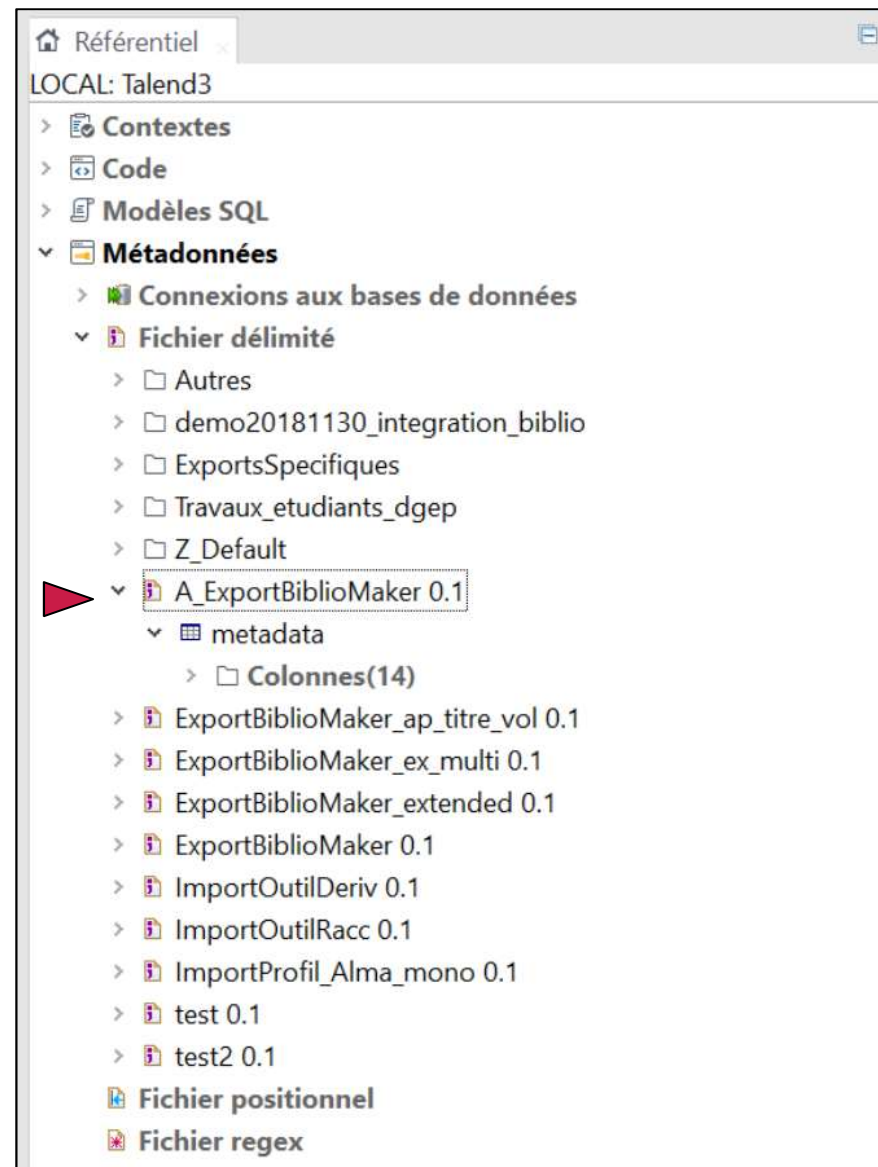
Vérifier que les noms des colonnes ainsi que le type de données

String = chaîne de caractères
Integer = nombre entier

Modifier le type des données pour
Annee_de_publication où le choix Integer sera utile pour
la suite des exercices

Colonne	Clé	Type	<input checked="" type="checkbox"/> N..	Modèle de date (Ctrl...	Length	Precision	Défaut	Comment...
Titre	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		75	0		
Adresse_bibliographique	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		49	0		
Zone_de_notes	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		19	0		
ISBN	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		10	0		
Annee_de_publication	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>			0		
Vedette	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		50	0		
Exemplaires_code_barre_duree...	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		13	0		

Une fois le paramétrage terminé, le modèle apparaît dans la liste



Modifier le job créé lors de l'exercice précédent

The screenshot displays the Talend Studio interface. The main window shows a job named '*Job ouvrir_fichier_temp 0.1' with a data flow from 'tFileInputDelimited_1' to 'tLogRow_2'. The job is running, showing '19 rows in 0.02s' and '863.64 rows/s'. Below the job canvas, the 'Designer' tab is active, showing the configuration for 'tFileInputDelimited_1'. The 'Paramètres simples' section is expanded, and the 'Schéma' dropdown is set to 'Référentiel'. A red box highlights the 'Référentiel' option and the dropdown arrow. To the right, a dialog box titled 'Contenu du référentiel' is open, showing a tree view of available models. The 'A_ExportBiblioMaker 0.1' folder is expanded, and the 'metadata' sub-folder is selected. A red box highlights this selection. The dialog has 'OK' and 'Cancel' buttons at the bottom.

- Choisir l'option Référentiel en face de Schéma
- Cliquer sur les  et sélectionner le modèle précédemment créé

Propager les modifications et exécuter le job

The screenshot displays a job execution window titled "Job ouvrir_fichier_temp 0.1". The main area shows a data flow from "tFileInputDelimited_1" to "tLogRow_2". The flow is labeled with "19 rows in 0.05s" and "404.26 rows/s". Below the flow, the "Job ouvrir_fichier_temp" execution window is visible, showing a log and a table of data.

Toutes les colonnes du fichier sont affichées

```
Starting job ouvrir_fichier_temp at 21:30:29 on 08/2025
[statistics] connecting to socket on port 3853
[statistics] connected
-----+-----+-----+-----+-----+-----+-----+
|Numero|Type          |Code_masque|Jour|Mois|Annee|Cot|
|-----+-----+-----+-----+-----+-----+-----+
|5330  |Monographie  |C1         |22  |2   |2005  |789
|134   |Monographie  |C1         |10  |3   |2006  |771
|247   |PB           |C1         |11  |4   |2007  |772
```

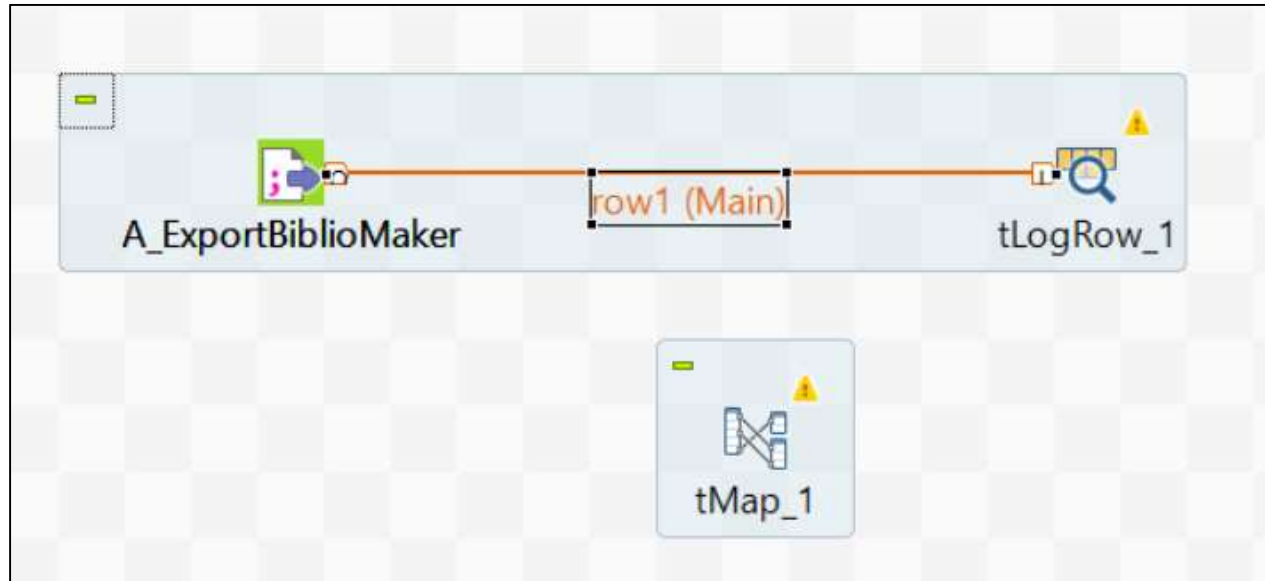
Exercice 3

Copier le job précédent et modifier les données en intégrant le composant tMap : objectif découvrir ce composant !



tMap est un composant très puissant qui permet de modifier l'ordre des colonnes, de les renommer, de concaténer ou séparer le contenu des colonnes, d'ajouter des conditions sur le contenu des colonnes, de fusionner des données à partir de plusieurs fichiers ou encore de créer plusieurs flux de sorties selon des filtres choisis.

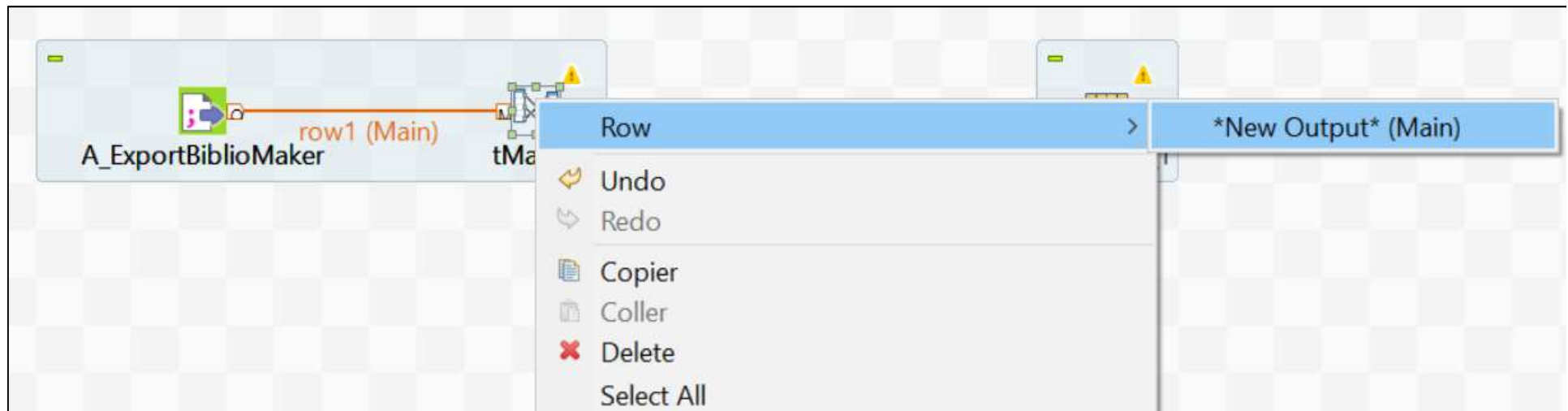
Intégration du composant tmap dans le job



Sélection du flux *row1* et suppression avec touche DEL

Ajout d'un nouveau flux entre les composants tFileInputDelimited et tMap

Création d'un flux de sortie depuis tMap vers tLogRow (clic droit > Row > New Output)



Ajout et paramétrage du composant tMap

Variables à ajouter dans les colonnes de sortie : glisser déposer les variables à conserver du flux *row1* vers *out1*

Le nom, l'ordre, etc. des colonnes de sortie peut être modifié à partir de l'éditeur de schéma

row1	out1
Numero	Titre
Type	Auteur
Code_masque	Type
Jour	Adresse_bibliogra...
Mois	ISBN
Annee	Annee_de_publica...
Cote	Code_barres
Titre	
Adresse_bibliographique	
Zone_de_notes	
ISBN	
Annee_de_publication	
Vedette	
Exemplaires_code_barre_duree_pret_quantite	

Colonne	C...	Type	N.	Modèle de...	Length	Precis...	Déf...	Com...
Type	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		11	0		
Code_masque	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		2	0		
Jour	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		0	0		

Colonne	C...	Type	N.	Modèle de...	Length	Precis...	Déf...	Com...
Titre	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		75	0		
Auteur	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		50	0		
Type	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		11	0		

Ajout d'une condition pour la colonne ISBN

The screenshot displays the Talend Open Studio interface. The 'Constructeur d'expression' (Expression Builder) window is open, showing the following expression:

```
row1.ISBN.length() == 10 || row1.ISBN.length() == 13?  
row1.ISBN :  
""
```

A red box highlights the expression with the following text:

Si l'ISBN compte 10 ou 13 caractères (condition à renseigner avant le « ? »), conserver l'ISBN, sinon (après les « : ») laisser vide

The 'Mapping auto' window is also visible, showing a table with columns 'Expression' and 'Column':

Expression	Column
row1.Titre	Titre
row1.Vedette	Auteur
row1.Adresse_bibliographique	Adresse_bibliogr...
row1.ISBN	ISBN
row1.Annee_de_publication	Annee_de_public...
row1.Exemplaires_code_barre_du...	Exemplaires_cod...

A red box highlights the 'row1.ISBN' row in the mapping auto window with the text:


Cliquer sur les ... afin de modifier la variable : la fenêtre qui s'ouvre permet d'utiliser des conditions et des fonctions java

At the bottom, the 'Éditeur de Schéma' (Schema Editor) window shows two tables: 'row1' and 'out1'.

Colonne	Clé	Type	✓	N..	Modèle de...	Leng...	Preci...	Dé...	Com...
Numero	<input type="checkbox"/>	Inte...	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		4	0		
Type	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		11	0		
Code_masque	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2	0		
Jour	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			0		
Mois	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			0		
Annee	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			0		

Colonne	Clé	Type	✓	N..	Modèle de...	Leng...	Preci...	Dé...	Com...
Titre	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		75	0		
Auteur	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		50	0		
Adresse_biblio...	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		49	0		
ISBN	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		10	0		
Annee_de_publ...	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	"yyyy"		0		
Exemplaires co...	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		13	0		

Filtrage des lignes retenues en fonction de la date

Cliquer sur  afin d'ajouter un filtre en fonction de la valeur d'une variable choisie : ici on ne conserve dans le flux de sortie que les lignes dont la date est supérieure à 1990

Expression	Column
row1.Titre	Titre
row1.Vedette	Auteur
row1.Type	Type
row1.Adresse_bibliographique	Adresse_bibliographique
row1.ISBN.length() == 10 row1.ISBN.lengt...	ISBN
row1.Annee_de_publication	Annee_de_publication
row1.Exemplaires_code_barre_duree_pret_q...	Code_barres

Colonne	Clé	Type	N.	Modèle de dat...	Length	Precision	Défaut	Comm...
Número	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		
Type	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		11	0		
Code_masque	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		2	0		
Jour	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0		

Colonne	Type	Clé	N.	Modèle de dat...	Length	Precision	Défaut	Comm...
Titre		<input type="checkbox"/>	<input checked="" type="checkbox"/>					
Auteur		<input type="checkbox"/>	<input checked="" type="checkbox"/>					
Type		<input type="checkbox"/>	<input checked="" type="checkbox"/>					
Adresse_bibliographiq...	String	<input type="checkbox"/>	<input checked="" type="checkbox"/>		49	0		
ISBN	String	<input type="checkbox"/>	<input checked="" type="checkbox"/>		10	0		

Une fois les modifications terminées, cliquer sur OK

Exécuter le job

19 rows in 0.03s
672.9 rows/s
row1 (Main)

tMap_1

6 rows in 0.03s
193.55 rows/s
out1 (Main)

tLogRow_1

Contexts(ouvrir_fichier_Copy) Composant

Copy

Exécution

▶ Exécuter ■ Arrêter 🗑 Effacer

Seules les lignes avec des dates supérieures à 1990 ont été conservées, soit 6 sur les 19 que compte le fichier d'origine

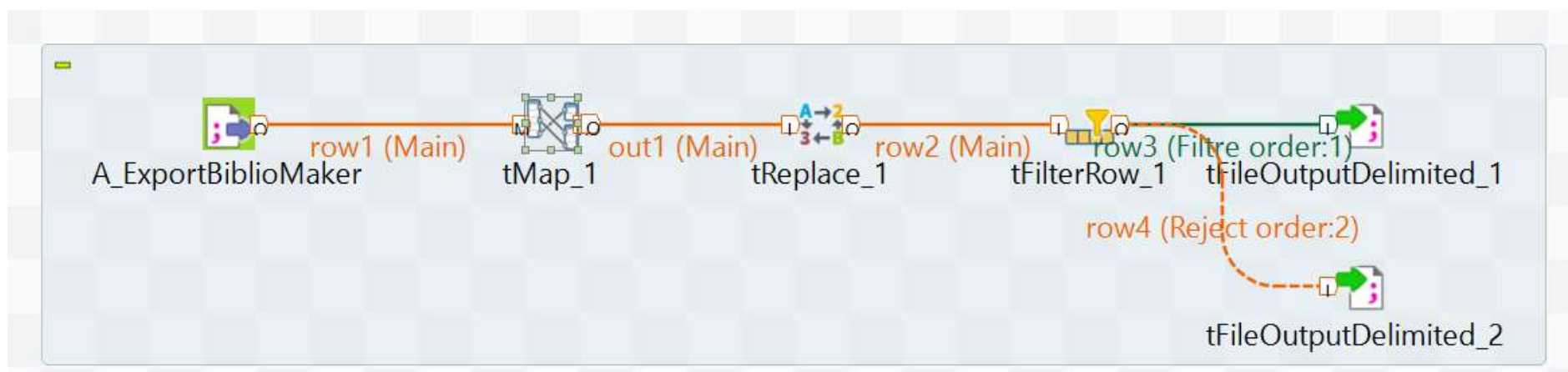
Seuls les ISBN avec 10 ou 13 caractères sont affichés

ISBN	Annee_de_publication	Code_barres
mandes	1993	CV0001C3/30/1
	1992	CV00026L/30/1
208012479x	1997	CV0008SJ/30/1
2051018618	2001	CV0008SK/30/1
	1992	CV000JIO/30/1
024550351x	1996	CV0003JE/30/1

Exercice supplémentaire

Reprise du job précédent et ajout des composants tReplace, tFilterRow et tFileOutputDelimited

Objectif : remplacer des valeurs à l'aide d'une expression simple, puis d'une expression régulière (regex), séparer les données en deux flux selon leur type et créer les 2 fichiers de sortie correspondants



Ajout et paramétrage du composant tReplace

The screenshot shows the Talend Studio interface. At the top, a data pipeline is visible with components: A_ExportBiblioMaker, tMap_1, and tReplace_1. Below the pipeline, the configuration for the tReplace_1 component is shown. The 'Paramètres simples' section has 'Mode simple' checked. The 'Chercher/Remplacer' table is as follows:

Colonne d'entr...	Rechercher	Remplacer par	Tout le m...	Sensible ...	Expressio...	Comr
Type	"BD"	"BOOK"	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

The 'Paramètres avancés' section has 'Mode avancé (recherche par expression régulière "regexp")' checked. The 'Expr.Régulières' table is as follows:

Source	Modèle	Remplacer	Commentaire
Code_barres	"\\w+"	"	remplacer tous les caractères qui suivent le "/"

L'expression «`\\w`» inclut n'importe quel caractère : ici on remplace par rien le caractère «/» et tous ceux qui le suivent

Ajout et paramétrage du composant tFilterRow

The screenshot displays the Talend Studio interface. At the top, a job named "Job ouvrir_fichier 0.1" is shown in the Designer view. The job flow consists of four components: A_ExportBiblioMaker, tMap_1, tReplace_1, and tFilterRow_1. Performance statistics are visible above each component: A_ExportBiblioMaker (19 rows in 0.02s, 791.67 rows/s), tMap_1 (6 rows in 0.02s, 250 rows/s), tReplace_1 (6 rows in 0.02s, 250 rows/s), and tFilterRow_1 (6 rows in 0.02s, 250 rows/s). Below the job flow, the configuration panel for the tFilterRow_1 component is open. The "Paramètres simples" tab is selected, showing the "Schéma" set to "Built-In" and the "Opérateur logique utilisé pour combiner des conditions" set to "and". A table of conditions is displayed, with one condition highlighted in red:

Colonne d'entrée	Fonction	Opérateur	Valeur
Type	Vide	Vaut	"BOOK"

Below the table, there are icons for adding (+), deleting (x), moving up (↑), moving down (↓), and refreshing (🔄) conditions. The checkbox "Utiliser le mode avancé" is unchecked.

A red-bordered box contains the text: "Filtre appliqué pour retenir les données de type «BOOK» uniquement".

Ajout et paramétrage du composant tFileOutputDelimited

The screenshot displays the Talend Studio interface. At the top, a job named "Job ouvrir_fichier 0.1" is shown in the Designer view. The job flow consists of several components: A_ExportBiblioMaker, tMap_1, tReplace_1, tFilterRow_1, and tFileOutputDelimited_1. Performance metrics are visible above each component, such as "19 rows in 0.04s" and "500 rows/s".

Below the job flow, the configuration for the selected component, tFileOutputDelimited_1, is shown. The configuration includes:

- Type de propriété: Built-In
- Utiliser le flux de sortie:
- Nom de fichier: |C:/Users/npayot/Desktop/importBiblioMaker_out.txt
- Séparateur de lignes: "\n"
- Séparateur de champs: "\t"
- Ecrire après:
- Inclure l'en-tête:
- Compresser en tant que fichier zip:
- Schéma: Built-In
- Modifier le schéma: ...
- Sync colonnes: Sync

A red box highlights the "Nom de fichier" field. Another red box highlights the "Séparateur de champs" field, with a callout box stating: «\t» correspond à une tabulation.

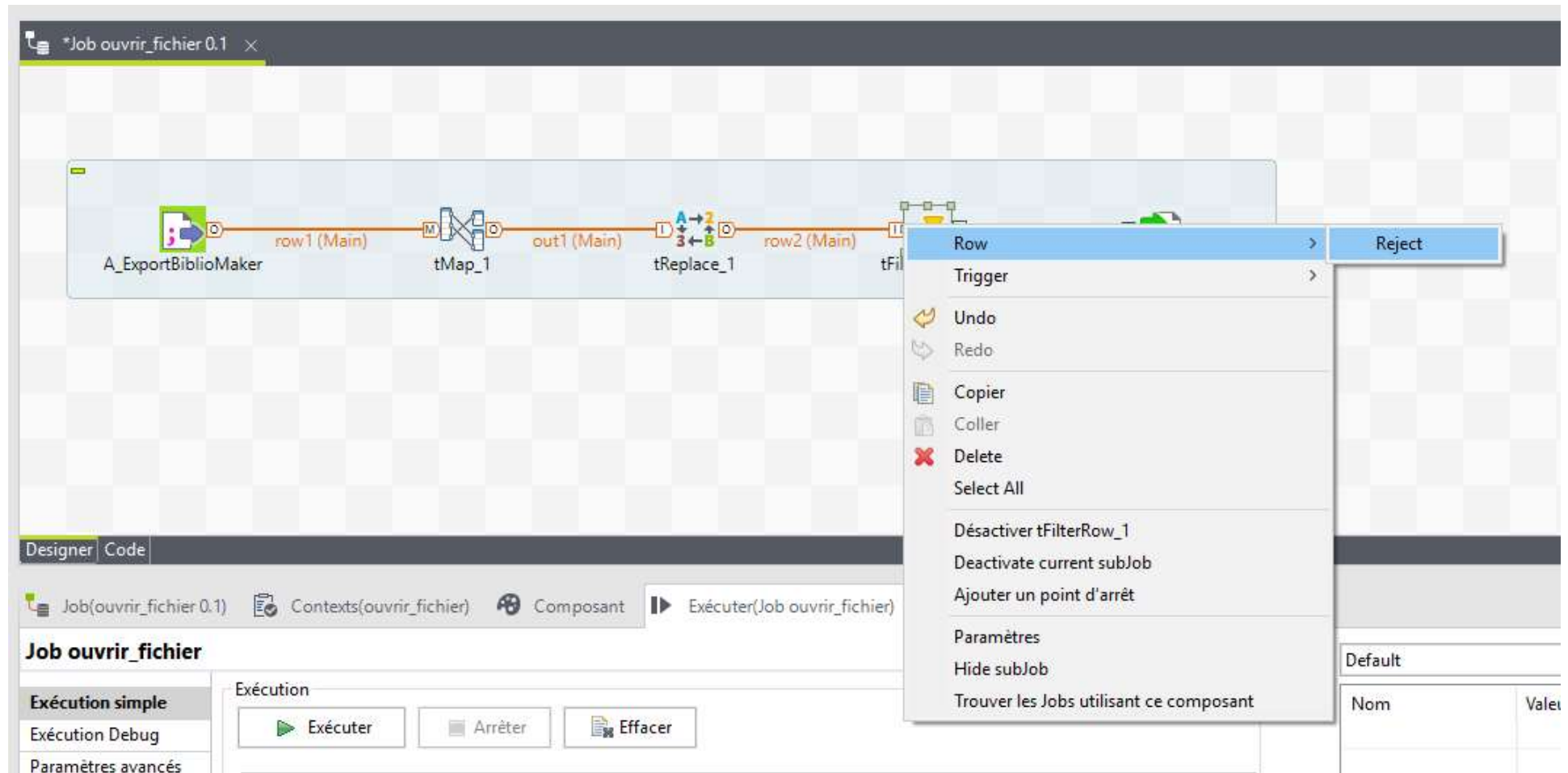
Modification des paramètres avancés de tFileOutputDelimited

The screenshot displays the Talend Studio interface for a job named '*Job ouvrir_fichier 0.1'. The job design is shown in the Designer view, featuring a sequence of components: A_ExportBiblioMaker, tMap_1, tReplace_1, tFilterRow_1, and tFileOutputDelimited_1. Performance statistics are visible above each component, such as '19 rows in 0.02s' for A_ExportBiblioMaker and '6 rows in 0.03s' for tFileOutputDelimited_1.

The configuration panel for 'tFileOutputDelimited_1' is shown below the job design. The 'Paramètres avancés' (Advanced parameters) tab is selected and highlighted with a red box. The following options are visible:

- Séparateur avancé (pour les nombres)
- Options CSV
- Créer le répertoire s'il n'existe pas.
- Diviser la sortie dans plusieurs fichiers
- Personnaliser la taille de la mémoire tampon
- Sortie en mode ligne
- Encodage: UTF-8
- Ne pas générer de fichier vide
- Retourner une erreur si le fichier existe déjà
- Statistiques du tStatCatcher

Capture des lignes rejetées par tFilterRow en ajoutant un 2^{ème} fichier de sortie



Clic droit sur le composant tFilterRow et sélection de la 2^{ème} option « Reject »

Exécuter le job final

The screenshot displays a data integration tool interface with a job execution flow and a log window. The flow consists of several components: A_ExportBiblioMaker (19 rows in 0.02s, 1187.5 rows/s), tMap_1 (6 rows in 0.02s, 375 rows/s), tReplace_1 (6 rows in 0.02s, 375 rows/s), tFilterRow_1 (4 rows in 0.03s, 139.7 rows/s), tFileOutputDelimited_1 (2 rows in 0.02s, 125 rows/s), and tFileOutputDelimited_2 (2 rows in 0.02s, 125 rows/s). A text box highlights the tFilterRow_1 component, stating: "On peut constater que le tFilterRow sépare les données en retenant 4 lignes dans le flux principal (de type BOOK) tandis que les 2 dernières lignes sont capturées dans le second fichier". The log window shows the execution details for the job "ouvrir_fichier", including the start time (15:30 30/08/2023), connection statistics, and the end time (15:30 30/08/2023) with an exit code of 0.

On peut constater que le tFilterRow sépare les données en retenant 4 lignes dans le flux principal (de type BOOK) tandis que les 2 dernières lignes sont capturées dans le second fichier

Job ouvrir_fichier

Exécution

▶ Exécuter ■ Arrêter 🗑 Effacer

```
Starting job ouvrir_fichier at 15:30 30/08/2023.  
[statistics] connecting to socket on port 3441  
[statistics] connected  
[statistics] disconnected  
  
Job ouvrir_fichier ended at 15:30 30/08/2023. [exit code = 0]
```

Correction des erreurs

The screenshot displays a software interface with a code editor and a job execution log. The code editor is in the 'Code' tab, and the log shows the job completed successfully.

```
out1 = null;

// # Output table : 'out1'
out1_tmp.Titre = row1.Titre ;
out1_tmp.Auteur = row1.Vedette ;
out1_tmp.Type = row1.Type ;
out1_tmp.Adresse_bibliographique = row1.Adresse_bibliographique ;
out1_tmp.ISBN = row1.ISBN.length() == 10 || row1.ISBN.length() == 13?
row1.ISBN
"" ;
out1_tmp.Annee_de_publication = row1.Annee_de_publication ;
out1_tmp.Code_barres = row1.Exemplaires_code_barre_duree_pret_quantite ;
out1 = out1_tmp;
// #####

} // end of Var scope

rejectedInnerJoin_tMap_1 = false;
```

The interface includes a 'Designer' tab and a 'Code' tab. The 'Code' tab is active, and the log shows the job completed successfully.

Job ouvrir_fichier ended at 04:06 29/08/2023. [exit code = 0]

L'onglet «Code» permet de visualiser les lignes en erreur dans le job. Les erreurs ne peuvent par contre être corrigées que dans l'onglet «Designer»

Problème possible

100 rows in 0.02s
4545.45 rows/s
row1 (Main)

tRowGenerator_1 tLogRow_1

Contexts(generer_lignes) Composant Exécuter(Job gen

Schéma Built-In Modifier le schéma Sync colonnes

Mode

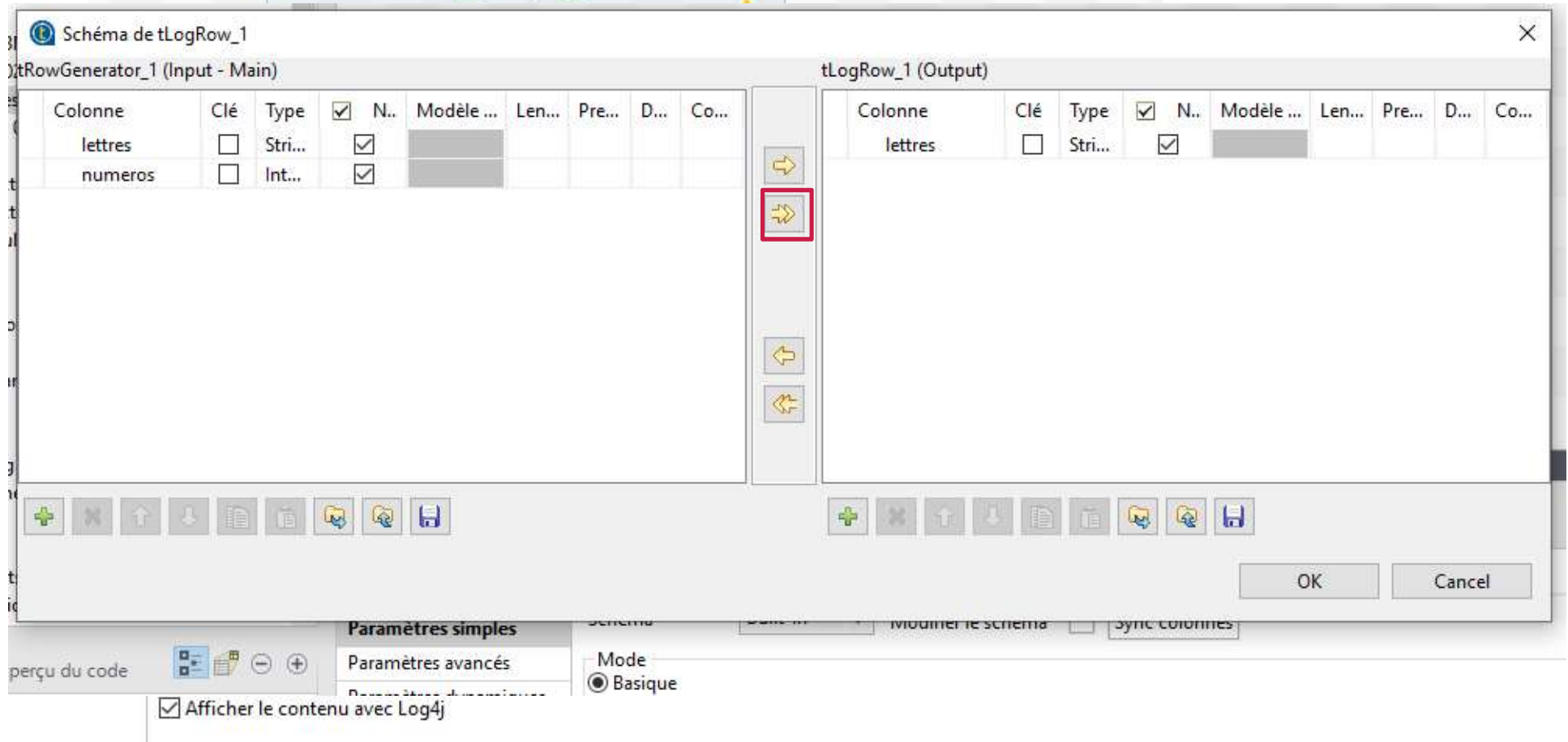
- Basique
- Tableau (afficher les valeurs dans les cellules d'un tableau)
- Vertical (chaque ligne est une liste de clé/valeur)

Séparateur de champs "|"

- Imprimer l'en-tête
- Afficher le nom unique du composant en face de chaque ligne de sortie
- Affiche le nom des colonnes du schéma en face de chaque valeur
- Utiliser une longueur fixe pour les valeurs
- Afficher le contenu avec Log4j

Si l'icône «Attention» apparaît près d'un composant, cela signifie qu'un élément est mal paramétré. Dans cet exemple, le flux de sortie ne correspond pas à celui en entrée. Pour résoudre le problème, cliquer sur «Modifier le schéma»

Problème possible : résolution



Ajouter les colonnes manquantes au flux de sortie

Documentation

Documentation officielle :

- <https://help.talend.com>
 - Rechercher les composants
 - Tester les scénarios

Tutoriaux web et pdf :

- <https://talend.developpez.com/cours/>

