

# Pipeline Talend

# Contexte

3 tables provenant de différentes databases :

- CRM
- Ecommerce
- ERP

=> mêmes clients, mais **id\_client** différents

**Objectif** : Récupérer les informations de chaque client depuis ces 3 DBs, unifier, formater et nettoyer.

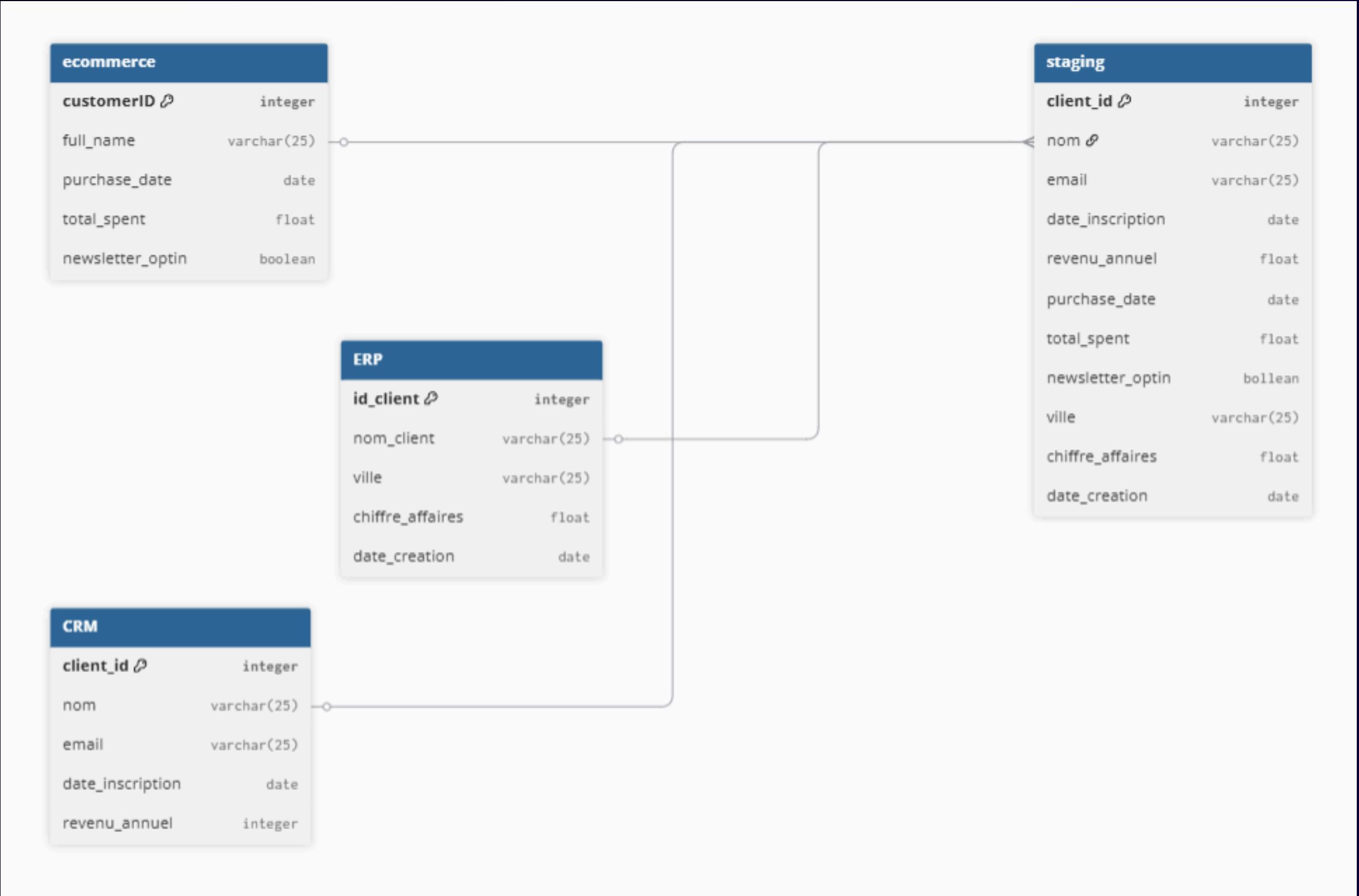
talend

client_id	nom	email	date_inscription	revenu_annuel
1	Client_1	client1@gmail.com	6/3/2022	
2	Client_2	client2@gmail.com	15/03/2022	35000
3	Client_3	client3@gmail.com	8/1/2023	35000
4	Client_4	client4@gmail.com	28/04/2022	28000€
5	Client_5	client5@gmail.com	25/09/2022	42 000 €

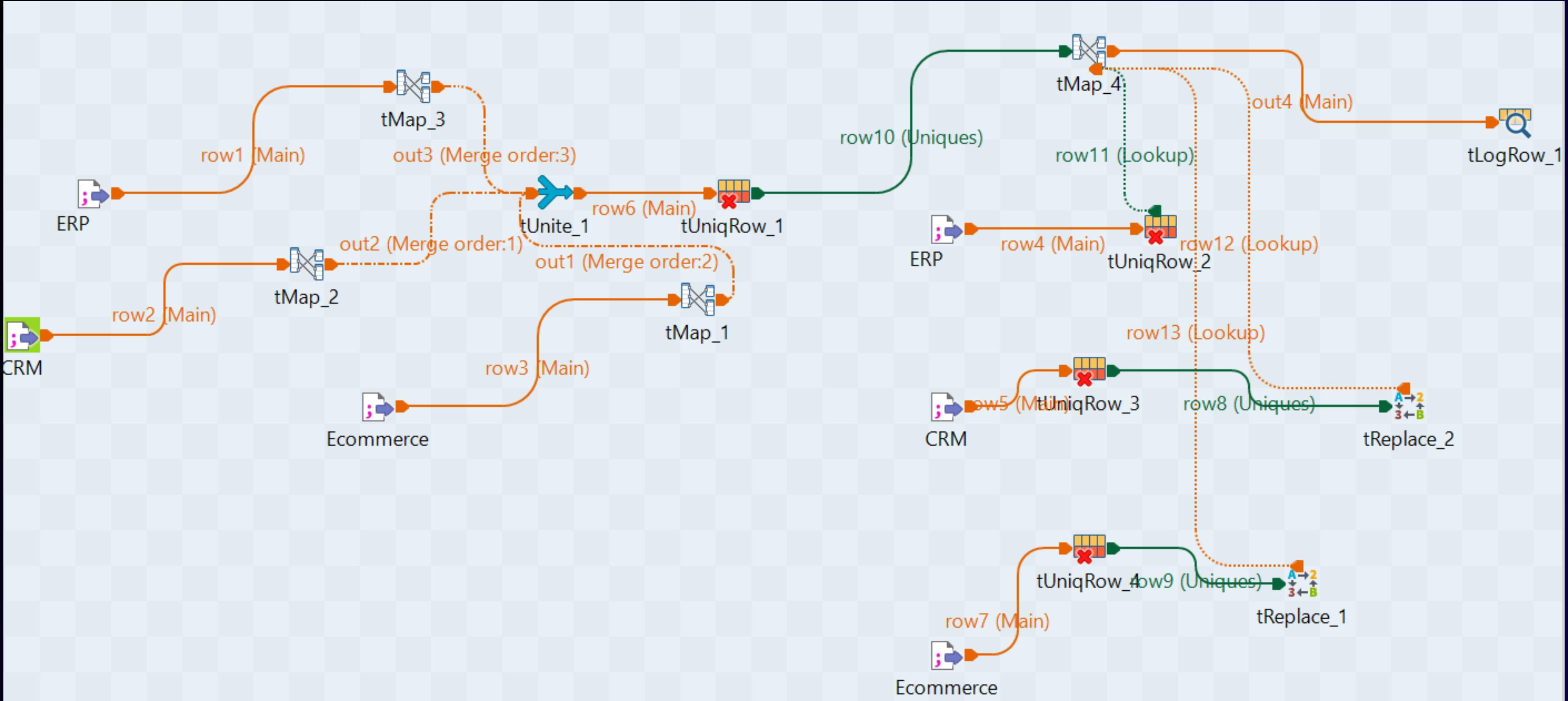
customerID	full_name	purchase_date	total_spent	newsletter_optin
13	Client_11	8/10/2023	436.86€	Yes
10	Client_8	24-02-2023	238.32€	
18	Client_18	12/1/2023	92.62€	
21	Client_9	7/1/2023	394.42€	No
9	Client_17	22-03-2023	184.17€	Yes

id_client	nom_client	ville	chiffre_affaires	date_creation
17	Client_4	Paris	9547.21	10/13/2023
15	Client_3	Marseille	3022.8	10/24/2021
5	Client_11	Paris	2484.1	3/20/2022
6	Client_18	Paris	7221.65	6/26/2022
11	Client_14	Marseille	6062.34	9/22/2023

# Diagramme






Union des 3 tables













Pipeline complet




# À l'intérieur de tMap1 : Préparation pour le tUnite

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




Find :    

Var     

Auto map!     

**row3**   

Column
customerID
name
purchase_date
total_spent
newsletter_optin

**out1**   

Expression	Column
row3.name	name









Schema editor Expression editor









row3

Column	K...	Type	<input checked="" type="checkbox"/> N.	Date Patter...	Length	Precisi...	Def...	Com...
customerID	<input checked="" type="checkbox"/>	int	<input type="checkbox"/>					
name	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					
purchase_date	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"dd-MM-y...				
total_spent	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					
newsletter_optin	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>				"No"	

out1

Column	K...	Type	<input checked="" type="checkbox"/> N.	Date Patter...	Length	Precisi...	Def...	Com...
name	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					

Apply Ok Cancel



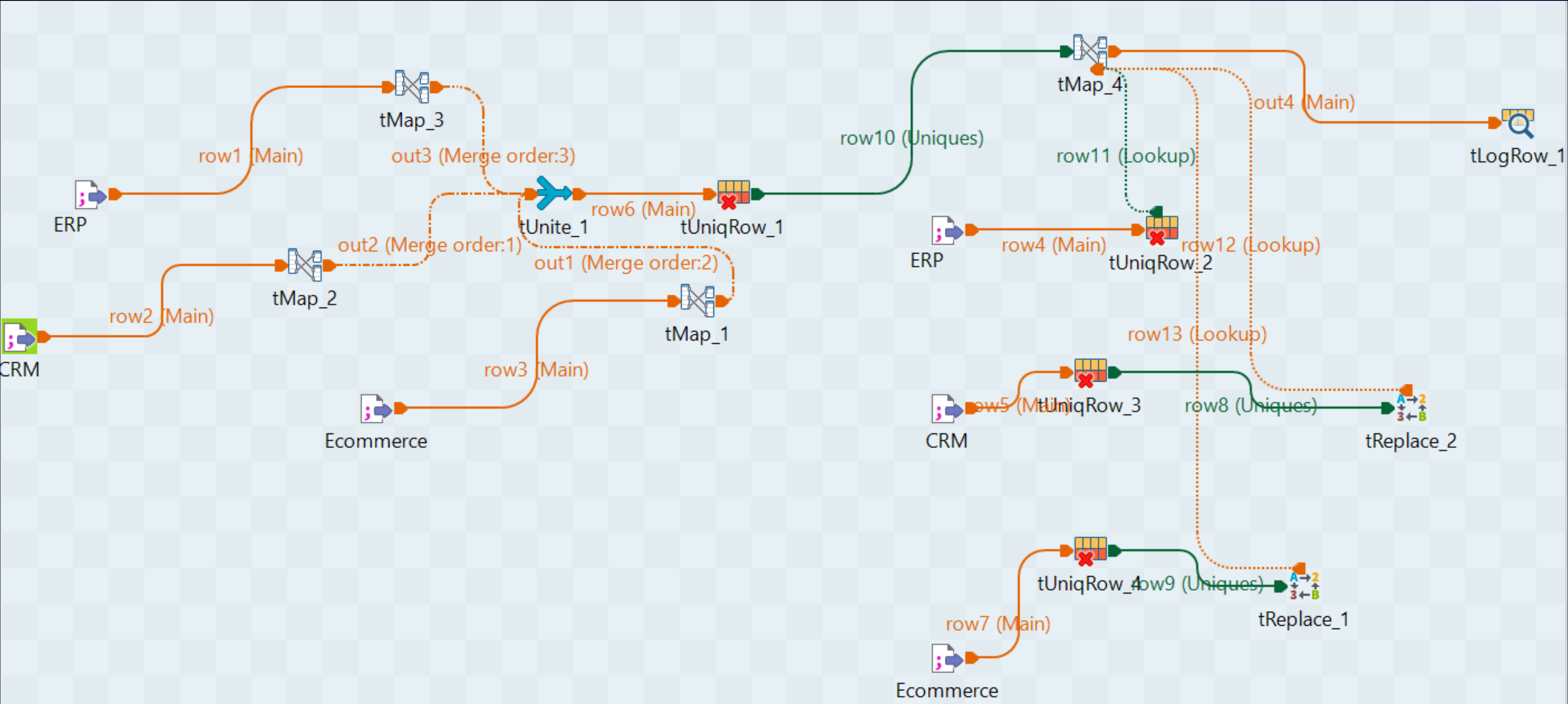
# tReplace : remplacement des valeurs inadéquates

The screenshot shows the configuration interface for the 'tReplace\_2' component. The top bar includes tabs for 'Job(Exo1 0.1)', 'Contexts(Exo1)', 'Component', and 'Run (Job Exo1)'. The left sidebar has a menu with 'Basic settings' (selected), 'Advanced settings', 'Dynamic settings', 'View', and 'Documentation'. The main area shows the 'Search/Replace' configuration. It includes a 'Schema' dropdown set to 'Built-In', an 'Edit schema' button, and a 'Sync columns' button. Below this, the 'Simple mode' checkbox is checked. The 'Search/Replace' table has two rows, both for the 'revenu\_annuel' column. The first row has a search value of '"â¬"' and an empty 'Replace with' field. The second row has a search value of '" "' and an empty 'Replace with' field. To the right of the table are checkboxes for 'Whole word' (checked), 'Case sensitive', and 'Glob express...', all of which are currently unchecked. A 'Comment' column is also present.

Search/Replace	InputColumn	Search	Replace with	<input checked="" type="checkbox"/> Whole word	<input type="checkbox"/> Case sensitive	<input type="checkbox"/> Glob express...	Comment
	revenu_annuel	"â¬"	""	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	revenu_annuel	" "	""	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Exemple avec tReplace2, pour la table ecommerce

**Ne pas oublier de décocher "Whole word"**



Pipeline complet

# À l'intérieur de tMap4

Union et gestion des valeurs vides avec tMap

row10

Column
name

row11

Expr.	globalMap Key
Expr. key	Column
	client_id
row10.name	name
	ville
	chiffre_affaires
	date_creation

row12

Expr.	globalMap Key
Expr. key	Column
	client_id
row10.name	name
	email
	date_inscription
	revenu_annuel

row13

Expr.	globalMap Key
Expr. key	Column
	customerID
row10.name	name
	purchase_date
	total_spent
	newsletter_optin

out4

Expression	Column
row10.name	name
row12.email	email
row11.ville == null    StringHandli...	ville
row11.date_creation == null ? Tal...	date_creation
row11.chiffre_affaires == null    S...	chiffre_affaires
row12.date_inscription == null ? ...	date_inscription
row12.revenu_annuel == null    St...	revenu_annuel
row13.purchase_date == null ? Ta...	purchase_date
row13.total_spent == null   String...	total_spent
row13.newsletter_optin == null    ...	newsletter_optin

Schema editor

Expression editor

row10

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

out4

Column	K...	Type	<input checked="" type="checkbox"/> N.	Date P...	Le...	Pre...	D...	Co...
name	<input type="checkbox"/>	Str...	<input checked="" type="checkbox"/>					
email	<input type="checkbox"/>	Str...	<input checked="" type="checkbox"/>					
ville	<input type="checkbox"/>	Str...	<input checked="" type="checkbox"/>					
date_crea...	<input type="checkbox"/>	Da...	<input checked="" type="checkbox"/>	"yyyy-...				
chiffre_aff...	<input type="checkbox"/>	Str...	<input checked="" type="checkbox"/>					
date_inscr...	<input type="checkbox"/>	Da...	<input checked="" type="checkbox"/>	"dd/M...				
revenu_an...	<input type="checkbox"/>	Str...	<input checked="" type="checkbox"/>					
purchase_...	<input type="checkbox"/>	Da...	<input checked="" type="checkbox"/>	"dd-M...				
total spent	<input type="checkbox"/>	Str...	<input checked="" type="checkbox"/>					



# tMap : remplacement des valeurs manquantes

```
row4.date_inscription == null ? TalendDate.parseDate("yyyy-MM-dd", "1900-01-01") : row4.date_inscription
```

```
row4.revenu_annuel == null || StringHandling.TRIM(row4.revenu_annuel).equals("") ? "0" : row4.revenu_annuel
```

```
row5.newsletter_optin == null || StringHandling.TRIM(row5.newsletter_optin).equals("") ? "No" : row5.newsletter_optin
```

**Format “si valeur nulle, remplir par [valeur par défaut], sinon, laisser la valeur présente**

# Sortie tMap4

	=	+	+	+	+	+	+	+	=
	name		email		ville		date_creation		chiffre_affaires
							date_inscription		revenu_annuel
	=	+	+	+	+	+	+	+	=
	Client_1		client1@gmail.com		Paris		1900-01-01		0
							06/03/2022		0
	Client_2		client2@gmail.com		Paris		1900-01-01		0
							15/03/2022		35000
	Client_3		client3@gmail.com		Marseille		2021-10-24		3022.8
							08/01/2023		35000
	Client_4		client4@gmail.com		Paris		2023-10-13		9547.21
							28/04/2022		28000
	Client_5		client5@gmail.com		Toulouse		2022-12-27		3714.65
							25/09/2022		42000
									01-01-1900
									0
									No

# Conclusion



# Démonstration