

Business Intelligence

Emine BOUCHIBA

DIA2

Rapport Talend TP Noté

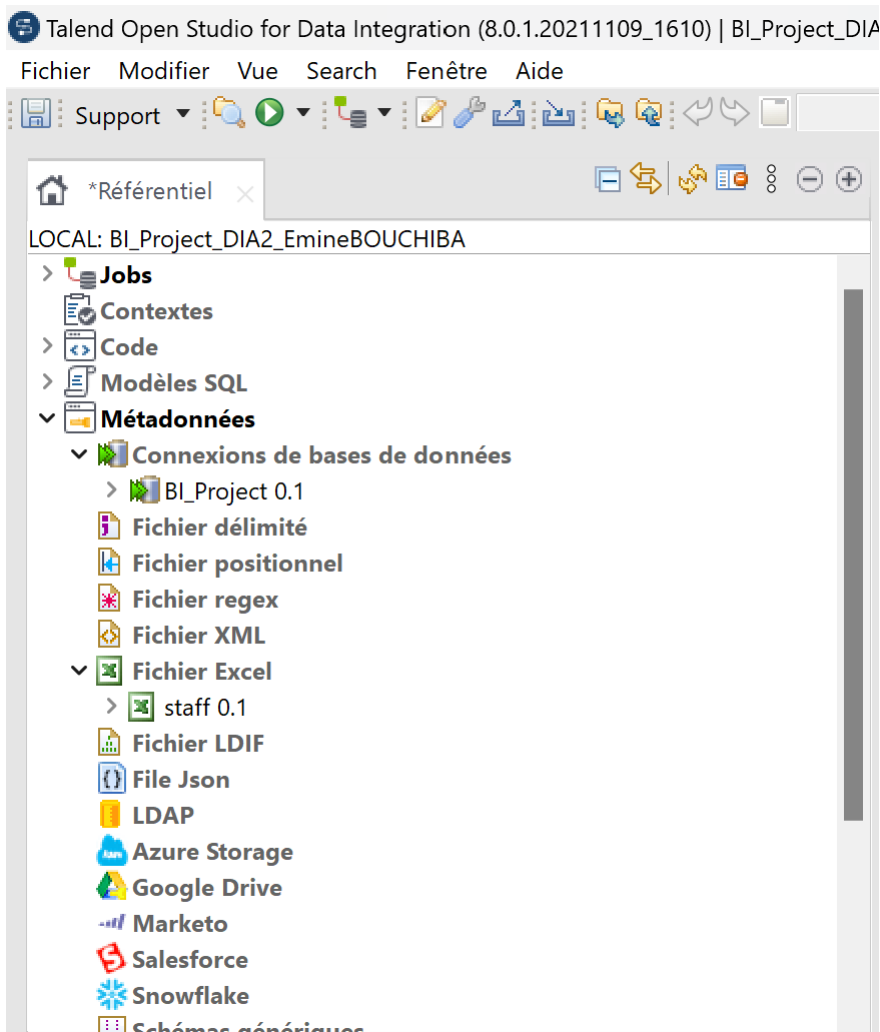


TABLE DES MATIERES

1.1 & 1.2 Accessing the data sources	2
1.3 Reading and displaying data	5
2. Data Transformation.....	7
2.1 Flow Joins	9
2.2 Writting to a Database	11

1.1 & 1.2 Accessing the data sources

➤ Importing the data :



I.1 & I.2 Accessing the data sources

3

Modifier un fichier Excel existant

Fichier - Étape 2 de 4

Ajouter un fichier de métadonnées au référentiel
Définissez le chemin d'accès au fichier et les paramètres de format

Paramètres du fichier

Serveur

Localhost 127.0.0.1

Fichier

C:/Users/emine/Documents/Ecole/A4/S8/Business Intelligence/BI_Project/Data/staff.xls

Parcourir...

☐ Lire le format Excel 2007 (xlsx)

Mode de génération

Memory-consuming(User mode)

Prévisualisation du fichier et paramètres des feuilles

Configurez les paramètres de la feuille

☒ All sheets/DSelect sheet

☒ _t1

Sélectionnez une feuille (structure de feuille comme schéma)

t1

A	B	C	D	E	F	G	H	I
Employee...	Start_Da...	End_Date	Job_Title	Salary	Gender	Birth_Da...	Emp_Hir...	Emp_Ter...
120101	01/07/1...	31/12/9...	Director	163040	M	18/08/1...	01/07/1...	
120102	01/06/1...	31/12/9...	Sales M...	108255	M	11/08/1...	01/06/1...	
120103	01/01/1...	31/12/9...	Sales M...	87975	M	22/01/1...	01/01/1...	
120104	01/01/1...	31/12/9...	Adminis...	46230	F	11/05/1...	01/01/1...	
120105	01/05/1...	31/12/9...	Secretar...	27110	F	21/12/1...	01/05/1...	
120106	01/01/1...	31/12/9...	Office A...	26960	M	23/12/1...	01/01/1...	
120107	01/02/1...	31/12/9...	Office A...	30475	F	21/01/1...	01/02/1...	
120108	01/08/2...	31/12/9...	Wareho...	27660	F	23/02/1...	01/08/2...	
120109	01/10/2...	31/12/9...	Wareho...	26495	F	15/12/1...	01/10/2...	
120110	01/11/1...	31/12/9...	Wareho...	28615	M	20/11/1...	01/11/1...	

< Back

Next >

Finish

Cancel

Installer un pilote

I.3 Reading and displaying data

➤ Reading the displayed data :

The screenshot shows the Talend Open Studio for Data Integration (8.0.1.20211109_1610) interface. The main workspace displays a job configuration for 'Job Job01_Read_display 0.1'. The job consists of two rows: 'row1 (Main)' and 'row2 (Main)'. 'row1 (Main)' reads from a 'continent' input and outputs to 'tLogRow_1', showing 8 rows in 0.37s at 21.74 rows/s. 'row2 (Main)' reads from a 'staff' input and outputs to 'tLogRow_2', showing 1048 rows in 0.21s at 4897.2 rows/s. The left sidebar shows the project structure, including 'Connexions de bases de données' and 'Fichiers'. The bottom panel shows the 'Exécution' (Execution) tab for 'Job Job01_Read_display', displaying a table of execution results.

Employee_ID	Start_Date	End_Date	Job_Title
120101.0	01-07-1999	31-12-9999	Director

The screenshot shows the configuration window for 'Job Job01_Read_display'. The 'Paramètres avancés' (Advanced parameters) tab is selected. The 'Exécution' (Execution) section shows the following settings:

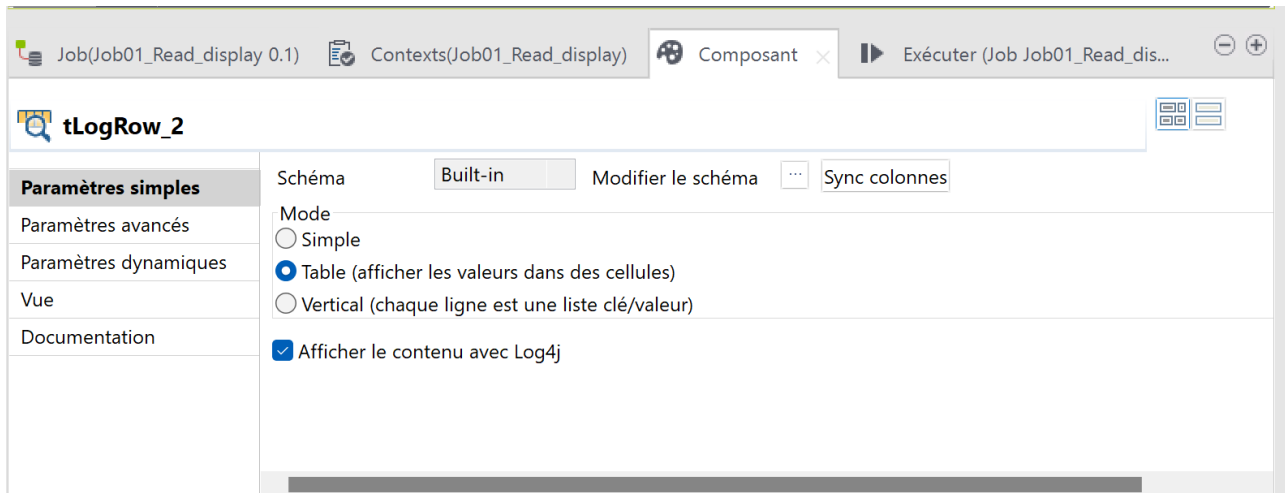
- ☒ Statistiques
- ☒ Enregistrer le Job avant exécution
- ☐ Temps d'exéc
- ☒ Effacer avant l'exécution

The 'JVM Setting' section shows the following arguments:

- Xms256M
- Xmx1024M

The 'Cible d'exécution' (Execution target) is set to 'Default'.

I.3 Reading and displaying data



2. Data Transformation

➤ Creating the job and using a tmap for the country table :

The screenshot shows the Talend Open Studio interface. On the left, the 'Référentiel' (Repository) pane shows a project named 'LOCAL: BI_Project_DIA2_EmineBOUCHIBA' with a job 'Job02_Filter_value 0.1'. The main workspace displays a job design with a 'country' input, a 'tMap_1' component, and a 'tLogRow_1' output. The 'tMap_1' component is configured with a 'row1 (Main)' input and a 'log123 (Main)' output. The 'tLogRow_1' component is configured with a 'Nom' output. The 'Job02_Filter_value' component is configured with a 'filter_value' expression: 'iter_value < 17:53 20-03-2023' and a 'socket on port 3957'.

The 'Job02_Filter_value' component is also shown in the 'Exécution simple' (Simple Execution) view, displaying a table of data:

population	country_id	continent_id	country
140.0	93.0	93.0	country
109.0	93.0	93.0	country
110.0	93.0	93.0	country
149.0	93.0	93.0	country
142.0	93.0	93.0	country
150.0	93.0	93.0	country

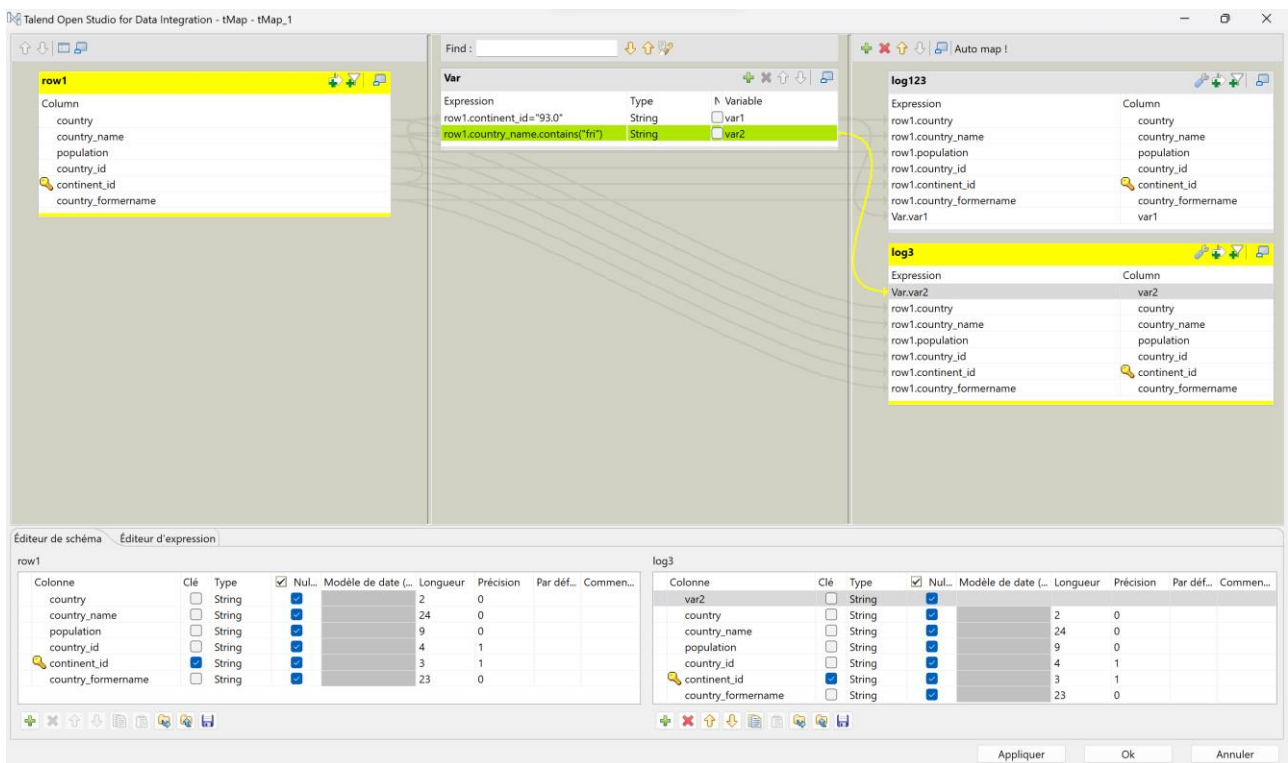
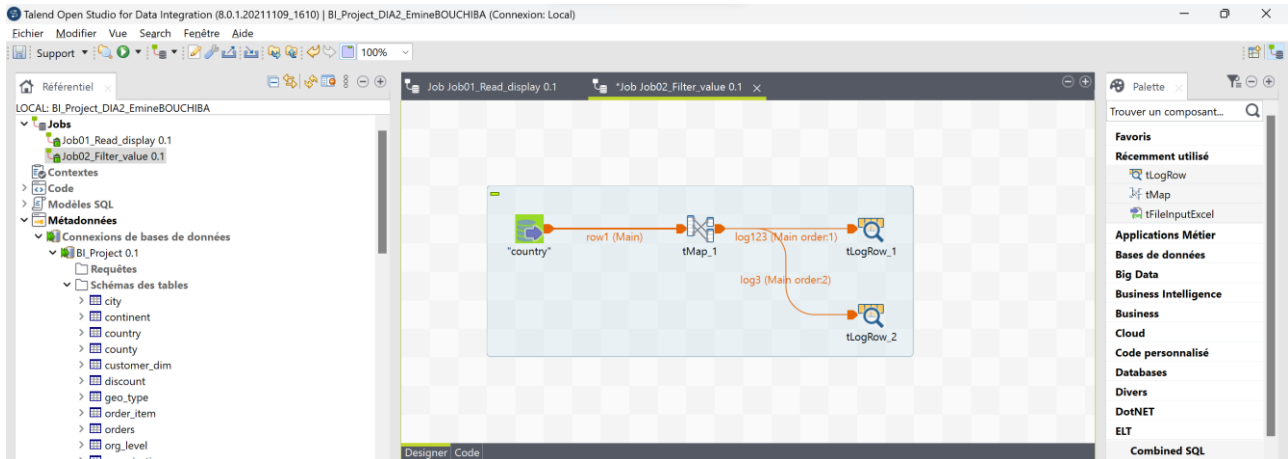
The screenshot shows the 'tMap' component configuration in Talend Open Studio. The 'tMap' component is configured with a 'row1' input and a 'log123' output. The 'row1' input is configured with a 'country' column and a 'continent_id' column. The 'log123' output is configured with a 'country' column, a 'continent_id' column, and a 'country_formname' column. The 'tMap' component is also configured with a 'Var' variable, which is a 'String' type and is used to store the 'continent_id' value.

The 'tMap' component is also shown in the 'Éditeur de schéma' (Schema Editor) view, displaying a table of data:

Colonne	Clé	Type	Nul.	Modèle de date (...)	Longueur	Précision	Par déf.	Commen.
country		String			24	0		
country_name		String			9	0		
population		String			4	1		
continent_id		BigDecimal			3	1		
country_formname		String			23	0		

2. Data Transformation

➤ Making filters :



2.1 Flow Joins

► Displaying « Continent » and « Country » :

The screenshot shows the Talend Open Studio interface. On the left, the 'Référéntiel' tree shows the project structure. The main workspace displays two data flows. The first flow reads from 'country' and outputs to 'tLogRow_1'. The second flow reads from 'continent' and outputs to 'tLogRow_2'. The 'Job Job03_Joins' window is open, showing the execution results for both flows.

country	continent
91 0	North America
92 0	South America
93 0	Europe
94 0	Africa
95 0	Asia
96 0	Australia/Pacific
97 0	Fangde

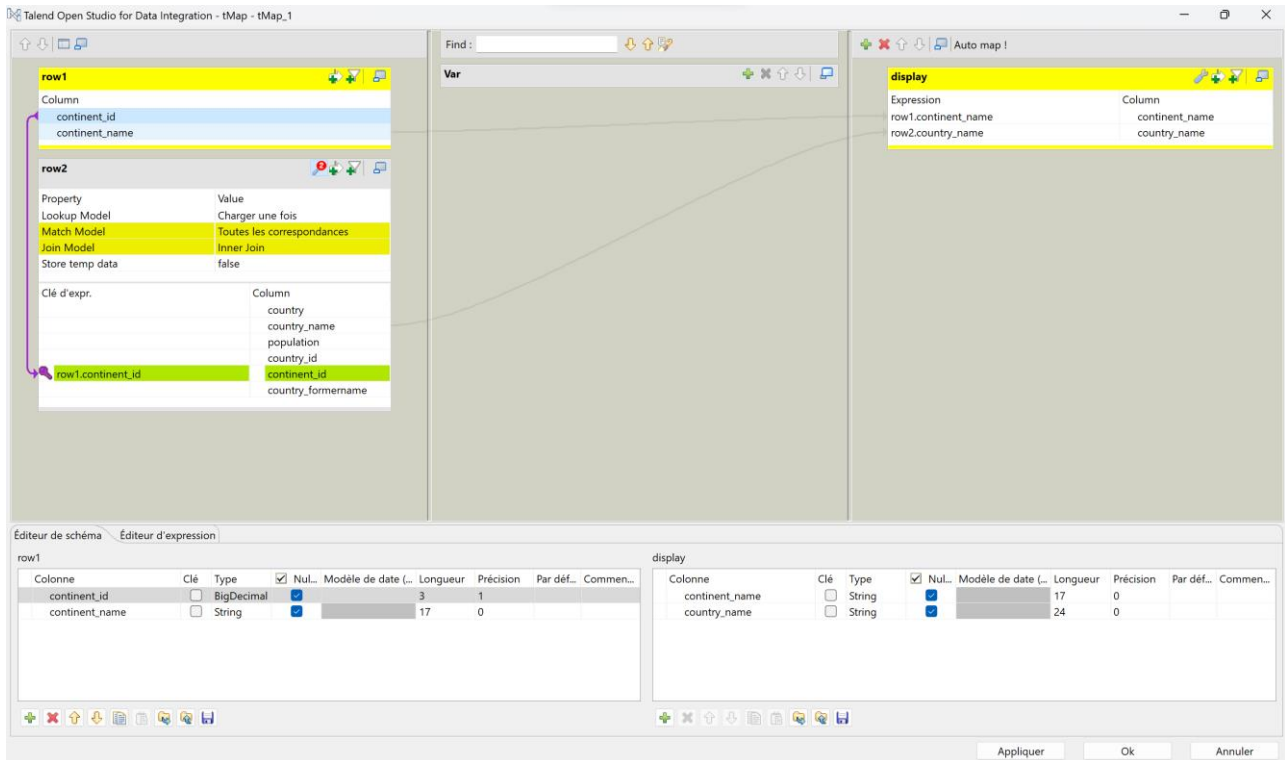
[statistics] disconnected
Job Job03_Joins terminé à 18:56 20-03-2023. [C]

► Making the join :

The screenshot shows the Talend Open Studio interface. The main workspace displays the 'country' and 'continent' flows joined together using a 'tMap_1' component. The 'Job Job03_Joins' window is open, showing the execution results for the joined flows.

continent_name	country_name
Antarctica	Antarctica
North America	Aruba
North America	Bermuda
North America	Bahamas
North America	Belize
North America	Canada
North America	Costa Rica
North America	Cuba
North America	Dominican Republic

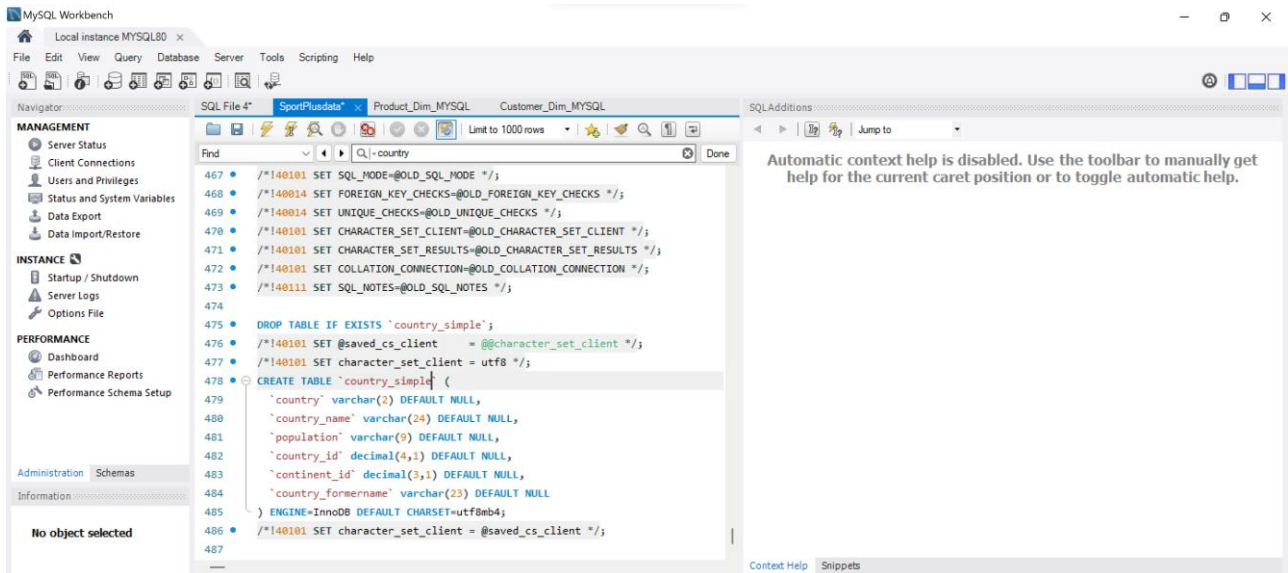
2.1 Flow Joins



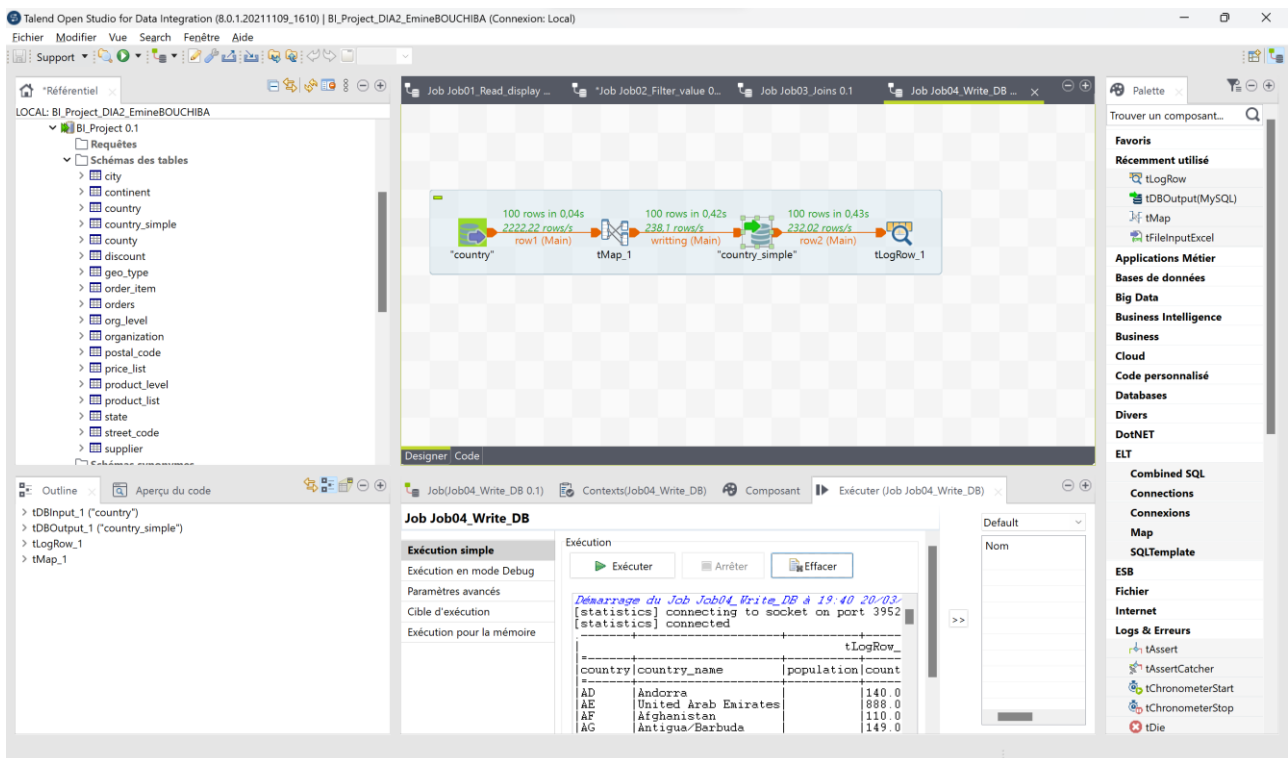
- ✓ The difference is that the « All matches » option enable us to display all results including occurrences.

2.2 Writing to a Database

➤ MySQL Database « country_simple » creation :



➤ Writing in the « country_simple » database :



2.2 Writing to a Database

12

Talend Open Studio for Data Integration - tMap - tMap_1

row1

Colonne	Clé	Type	Nul.	Modèle de date (...)	Longueur	Précision	Par déf.	Commentaire
country_name		String			24	0		
country_name		String			9	0		
population		String			4	1		
country_id		BigDecimal			3	1		
continent_id		BigDecimal			23	0		
country_formername		String						

writing

Colonne	Clé	Type	Nul.	Modèle de date (...)	Longueur	Précision	Par déf.	Commentaire
country_name		String			24	0		
country_name		String			9	0		
population		String			4	1		
country_id		BigDecimal			3	1		
continent_id		BigDecimal			23	0		
country_formername		String						

Appliquer Ok Annuler

➤ Checking the table in MySQL :

MySQL Workbench

Local instance MySQL80 x

File Edit View Query Database Server Tools Scripting Help

Navigator: SQL File 4* SportPlusdata* Product_Dim_MYSQL Customer_Dim_MYSQL

MANAGEMENT

- Server Status
- Client Connections
- Users and Privileges
- Status and System Variables
- Data Export
- Data Import/Restore

INSTANCE

- Startup / Shutdown
- Server Logs
- Options File

PERFORMANCE

- Dashboard
- Performance Reports
- Performance Schema Setup

Administration Schemas

Information

No object selected

SQL: File 4*
1 • show tables;
2 • select * from country_simple;
3

Result Grid

country	country_name	population	country_id	continent_id	country_formername
AD	Andorra	140.0	93.0		
AE	United Arab Emirates	888.0	95.0		
AF	Afghanistan	110.0	95.0		
AG	Antigua/Barbuda	149.0	92.0		
AI	Anguilla	142.0	92.0		
AL	Albania	120.0	93.0		
AM	Armenia	135.0	93.0		
AN	Netherlands Antilles	640.0	92.0		Soviet Union
AO	Angola	141.0	94.0		
AQ	Antarctica	11.0	90.0		
AR	Argentina	150.0	92.0		
AS	American Samoa	60.0	96.0		

country_simple 13 x

SQLAdditions

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

Read Only Context Help Snippets