SSIS & SQL Server Project.

Group: Safa, Ekaterina, Zhassulan, Charles.

Chosen model: star.

Schema:

Diagram, schematic

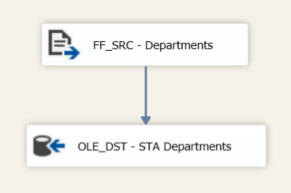
Description automatically generated

SQL scripts to create tables: in Tables\_creation.sql.

1. **STA : load source data.**

First we create STA\_project database in SQL.

Departments :

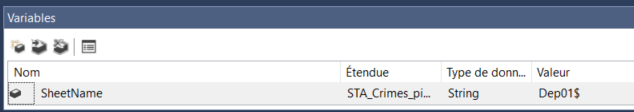


Result file:



Crimes :

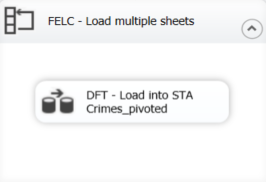
Creating variables to hold each worksheet name:



Add a loop Foreach Loop Container:

* Foreach ADO.NET Schema Rowset Enumerator
* Connection: MS Jet 4.0 OLE DB Provider
* Extended Properties “Excel 8.0”
* Variable Mapping “User::SheetName 2”

Add a data flow task within the loop.

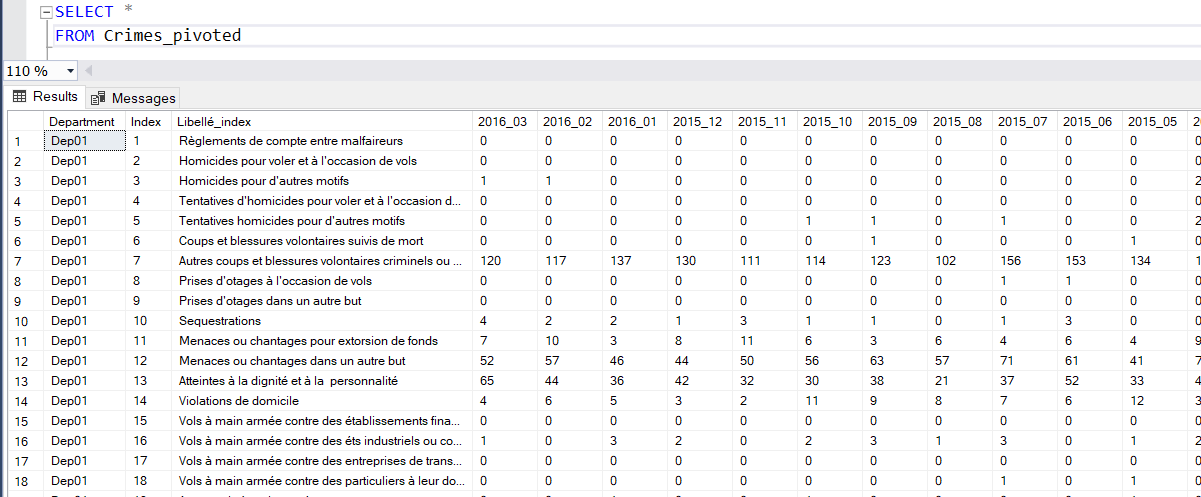


In the data flow:

* Add derived column with “(DT\_WSTR,50)(SUBSTRING(@[User::SheetName],1,LEN(@[User::SheetName]) - 1))” to transform sheets names on column.



Result table:

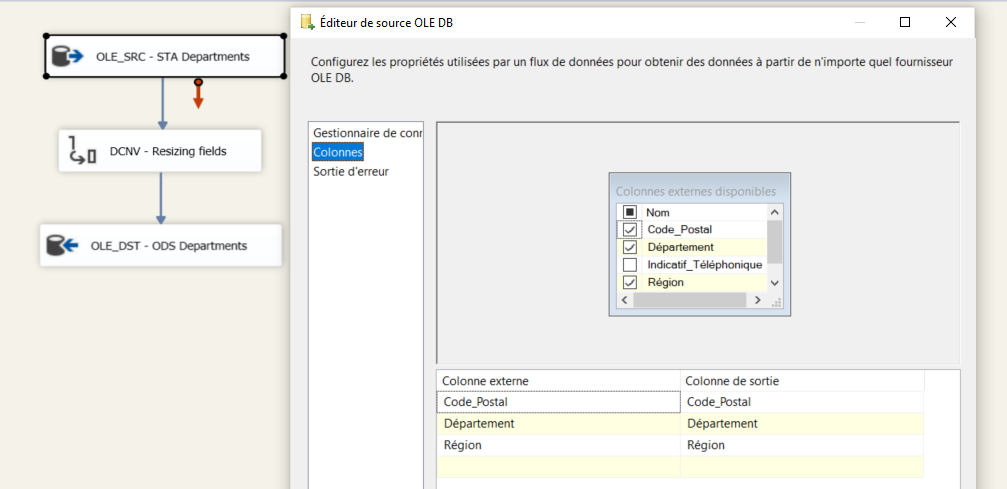


1. **ODS : transforamtions.**

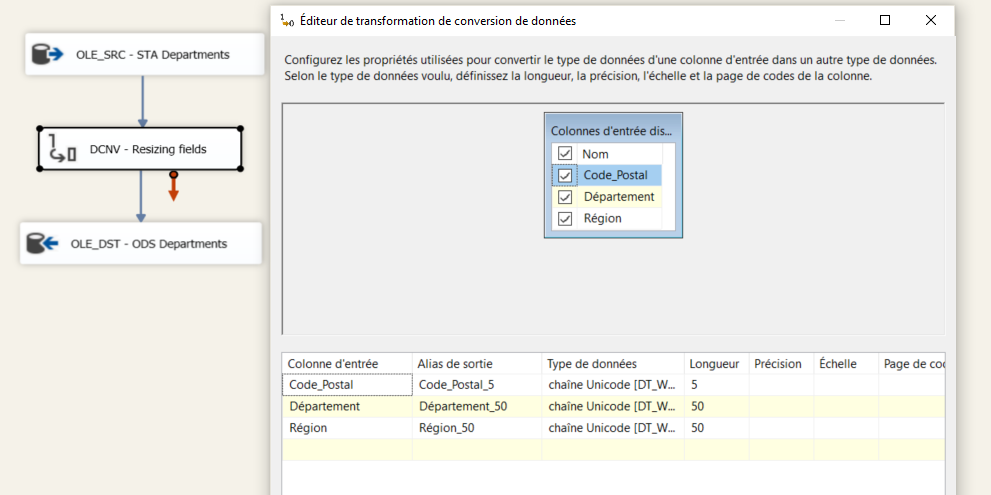
Create ODS\_Project database in SQL.

Departments :

We will not keep Phone code and Vacations zone as not related to crimes.

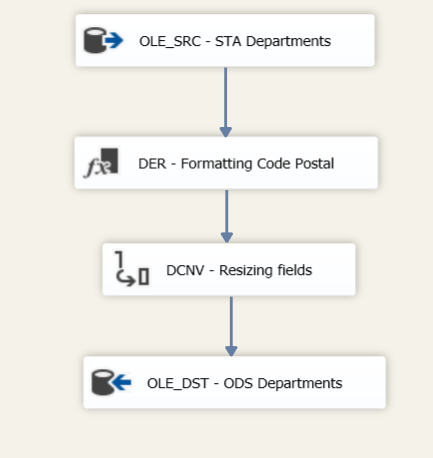


Resizing:

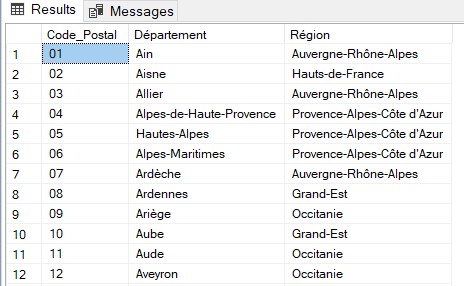


Derived column to format Department (Code\_Postal): “RIGHT("00"+[Code\_Postal]**,2**)” – to add 0 for first 9 (as I don’t know how to delete this 0 from Crimes table and we need coherence)

All transformations:

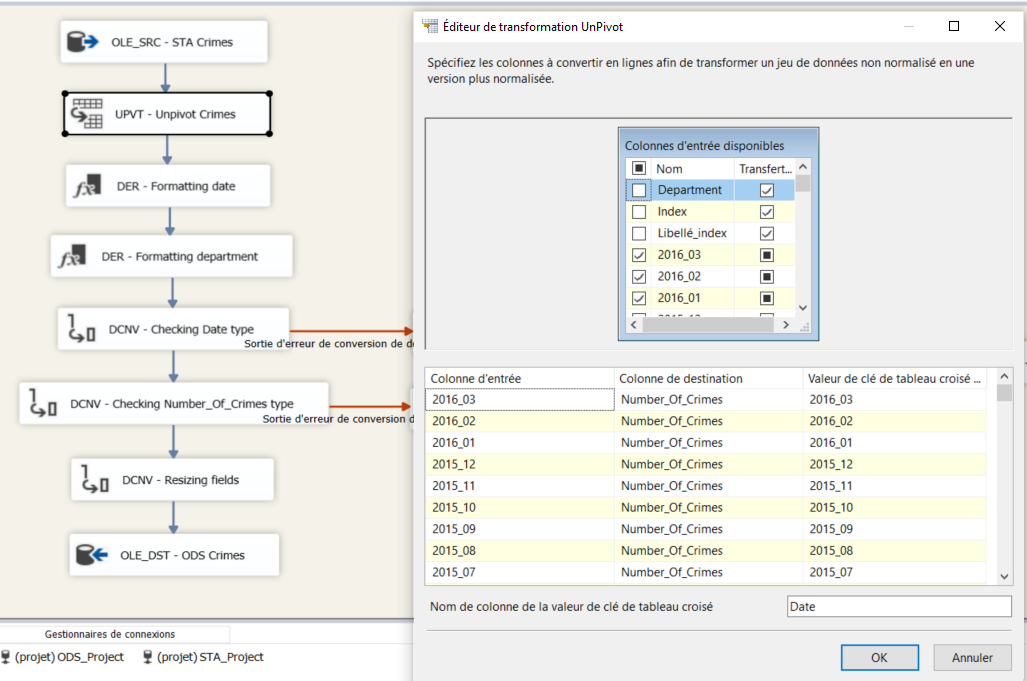


Result file:



Crimes :

Unpivot:



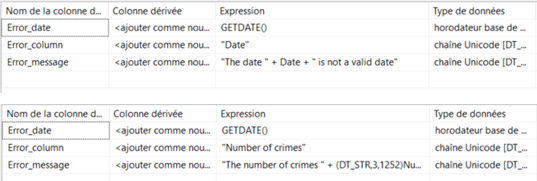
(If you know the way to put destination column name of all input column automatically – tell me!)

Derived column to format date: “(DT\_DATE)(SUBSTRING(Date,1,4) + "-" + SUBSTRING(Date,6,2) + "-" + "01")”;

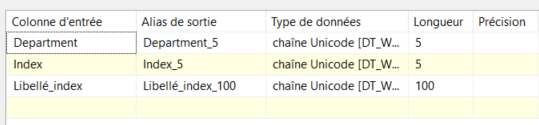
Derived column to format department: “(DT\_WSTR,50)(DT\_I8)SUBSTRING(Department,4,2)” (Delete Dep and leading 0);

Transform date in “date [DT\_DATE]” and number of crimes in “INT”.

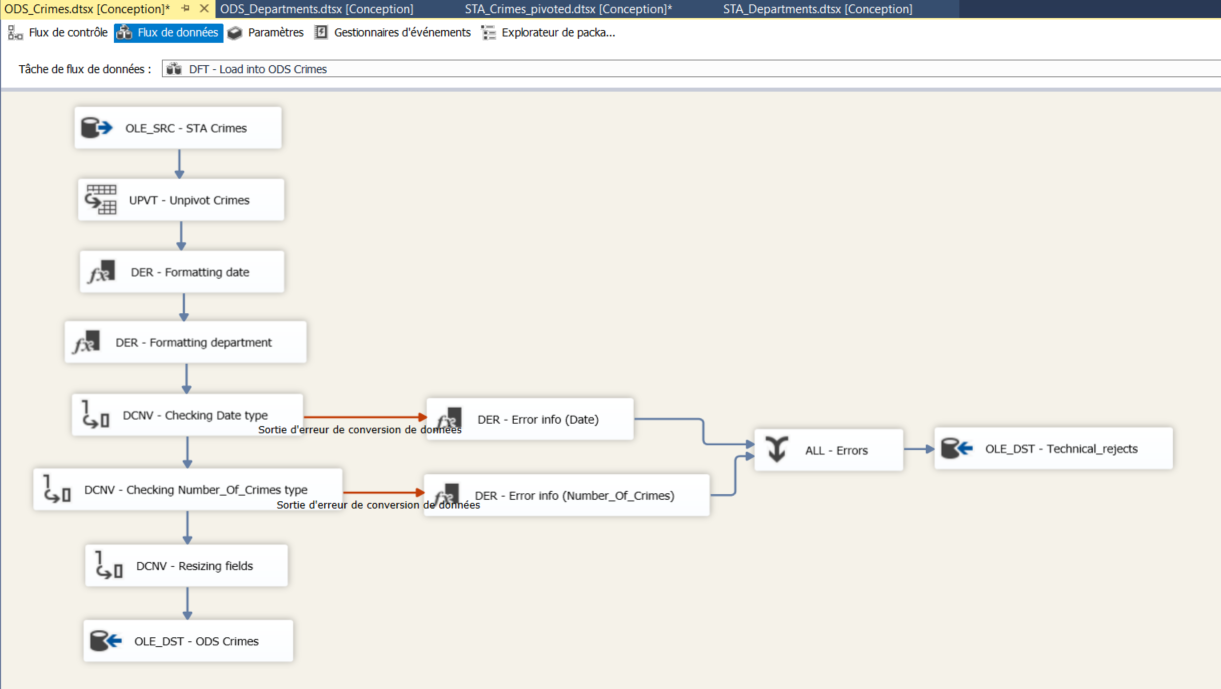
Put date errors and non-numeric crimes numbers in Technical rejects table:



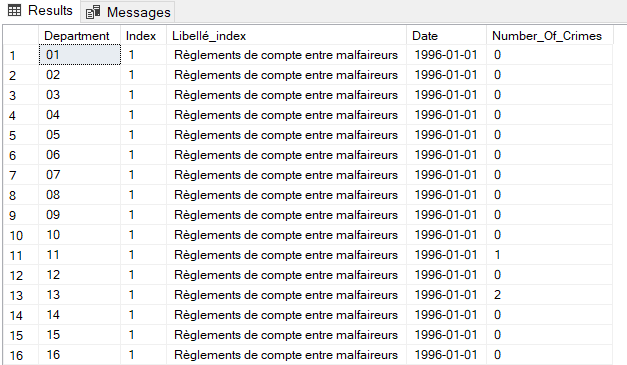
Resizing the rest fields:



All transformations:



Result file:



Technical rejects: (there were not created any).

1. **DWH**

Create DWH\_Projet database in SQL.

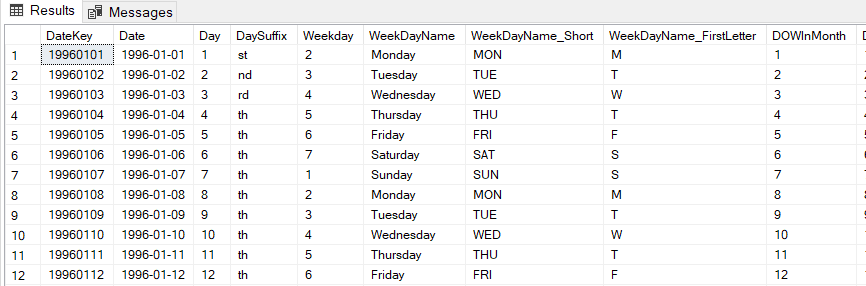
Dim tables: Date, Libellé (name of crimes), Department.

Fact table: Crimes (keys from dim tables + number of crimes)

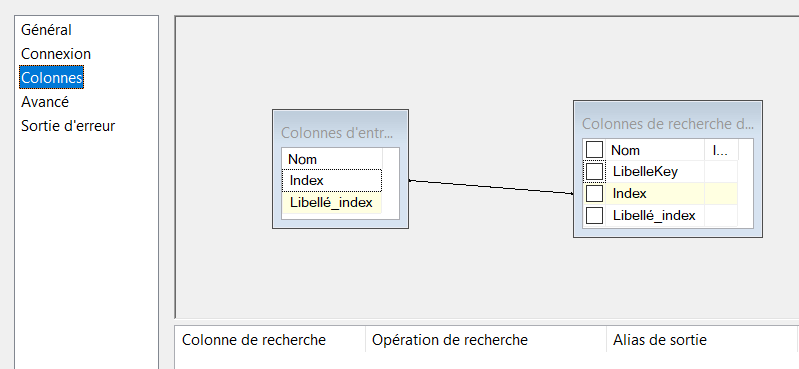
DimDate :

SQL script in SQL Script – DimDate.sql.

Result file:



DimLibelle :



For the case if libellé did not existed or were modified:

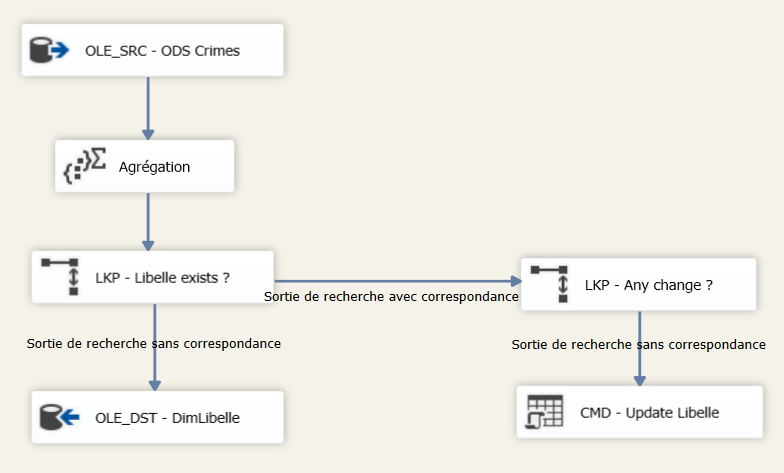
UPDATE [dbo].[DimLibelle]

SET

[Libellé\_index] = ?

WHERE [Index] = ?

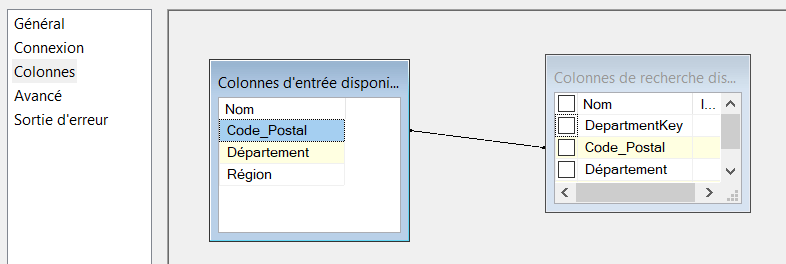
All transformations:



Result file:



DimDepartments :



For occasional modified records:

UPDATE [dbo].[DimDepartments]

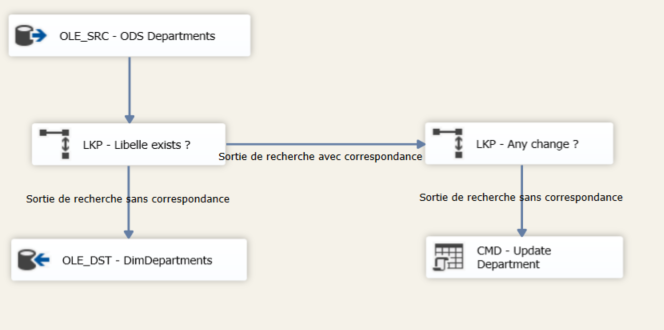
SET

[Département] = ?

[Région] = ?

WHERE [Code\_Postal] = ?

All transformations:

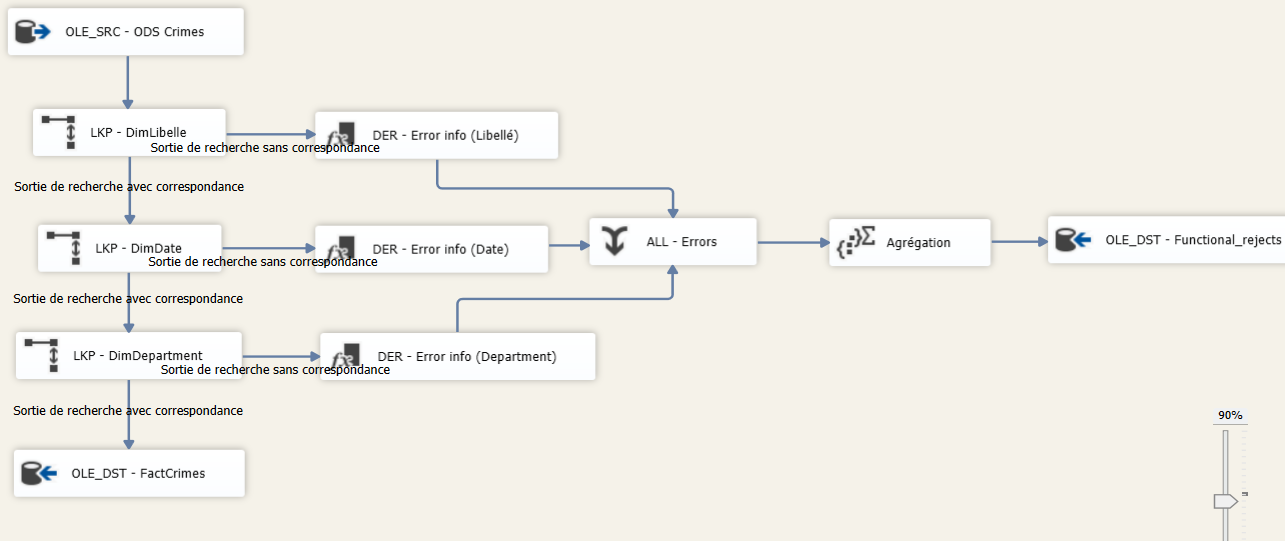


Result file:

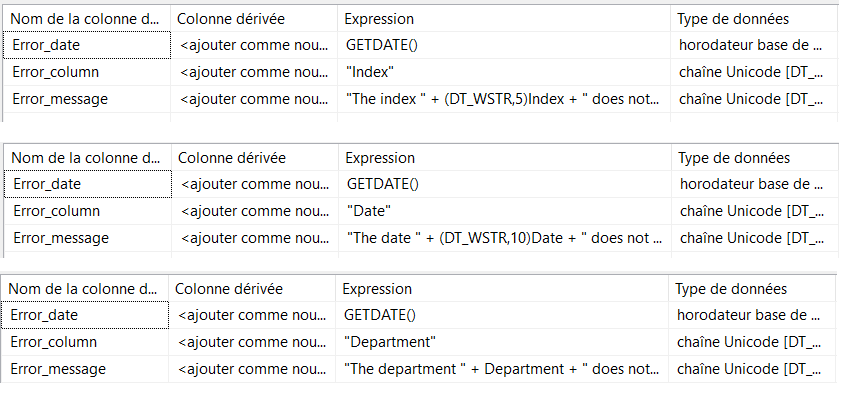


FactCrimes :

All transformations:

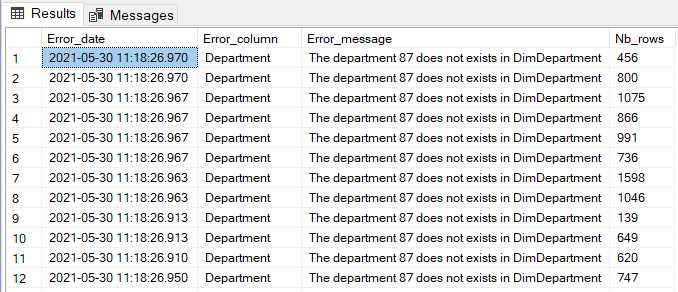


Functional rejects:



(Some rows rejected, 1 department was in Crimes and not in Departments source)

I don’t understand why every time it rejects different number of rows.. And it is not just an addition (I truncate table before test).



Result file: (LibelleKey IS NOT Libellé index, DepartmentKey IS NOT Department code, it’s ok 😊 )

