BPM Conseil

Vanilla

BI Gateway

Data & Document management services

for Vanilla BI Platform

Patrick Beaucamp
BPM Conseil
Contact : patrick.beaucamp@bpm-conseil.com

March-April 2009
Document : BPM_Vanilla_BIGateWay_User_Guide.odt
Document Status : Draft, early version

Internal Version

Index

Introduction

Vanilla BI GateWay is an innovative package to manage both data & documents and stands as Vanilla's favorite component to manage & transform data coming from various databases.

BI GateWay objective is to run in standalone (for local data manipulation) or embeded in the Vanilla platform, to provide Vanilla with data transformation services.

BI GateWay development addresses specific BI requirements and leverage all the resources available in the Vanilla platform, such as database connection, Vanilla globales variables, FreeMetadata documents or group security definition.

Bi GateWay stands ad the major component of Vanilla version 2, as an integrated ETL/ECM services that leverage database information and perform document manipulation, using secured freemetadata documents to access to database information (database tables & metadata repositories)

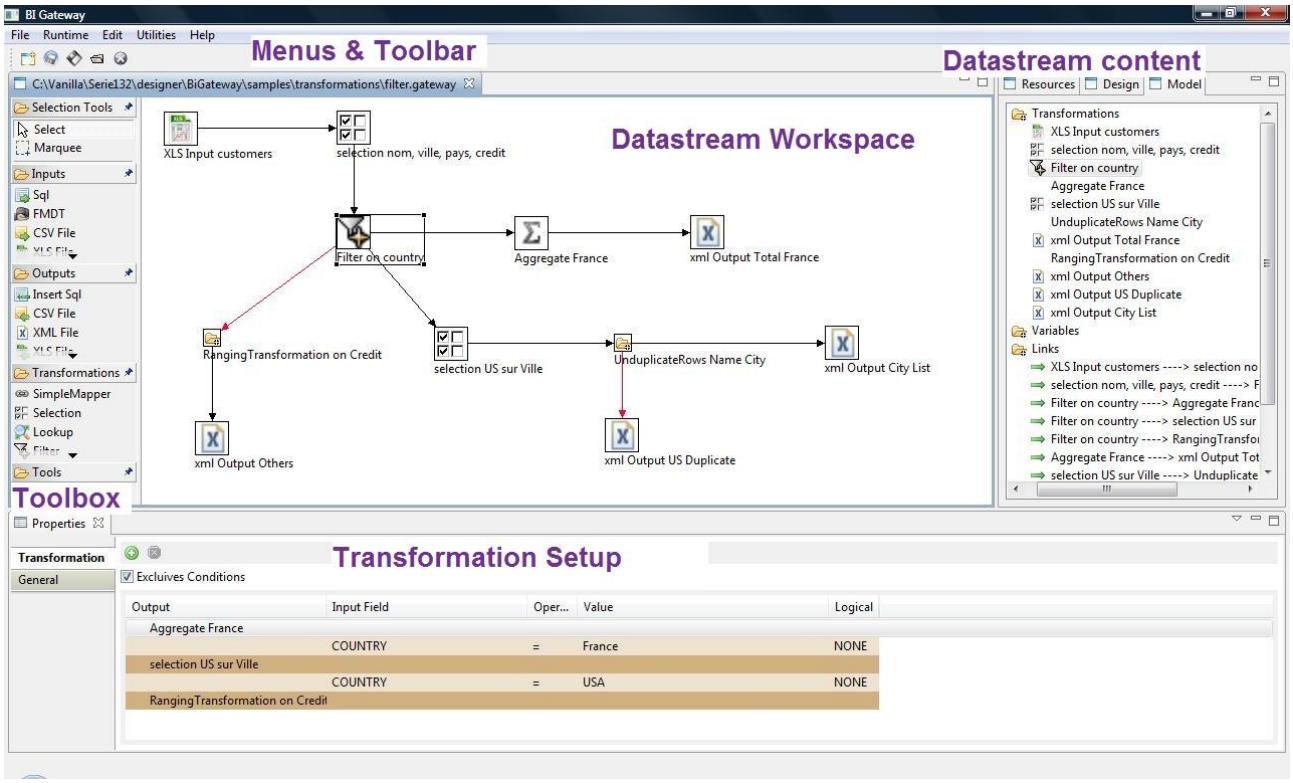
BI GateWay stands as a data service for Vanilla Platform

- tight integration with Vanilla requirements in terms of data management, such as the possibility to load FreeMetrics KPI database using a standard transformation
- Support of Enterprise Content Management functions to manipulate documents to be loaded in external ECM Servers such as Alfresco
- Clean and tight integration with other Vanilla components, with usage of common ressources and providing usefull embeded functions :
  - Use of Vanilla repository to store and deploy datastream transformations
  - Use of Vanilla repository ressources, such as :
    - database definition
    - Authentification provider definition (LDAP, DBMS)
    - Global variables definition
    - external ressources declaration, such as mail server or FTP server
  - BI Scheduler to schedule Bi GateWay datastream
  - BI Workflow to create complexe workflow with datastreams
  - FreeMetadata to access or load database using business views on physical data structure
  - Birt Plugin to create reports using a specific « BI Gateway » datasource definition
  - future SQL Designer module to create and load database structures & data

Bi GateWay Overview

Bi GateWay Interface contains different panels, treeview, tools & icons to facilitate access to ressources and smooth the development process.

Overview of a datastream from Bi GateWay interface



BI GateWay Approach

There are major difference between Bi GateWay and other ETL available on the market.

- Established BI editors (commercial or Open Source) have either bought an ETL Platform, or signed agreement with a market leader. None of those 2 approachs have given significant result in terms of tight integration. In vast majority of case, ETL still stands as an external services, not sharing BI Platform resources (security, repository, ressources) or BI even Platform components (scheduler, portal)

  BI GateWay approach is different, because initial objective is to build a component that stands as a dataservice link in the hearth of Vanilla BI Platform. It is the same approach we successfully had with FreeMetadata documents that unify data access, join strategy and security definition for Dashboard, Reporting and Olap development.

- **BI Gateway will be the first data manipulation component with a secured metadata driven approach, opening the door to new type of developers : business analysts who don't have the knowledge of physical database structure, but are able to use a Metadata document to manipulate and access to their business data**

- BI GateWay functions are divided into 2 types :
  - **Data Management** (transcodification, conversion functions, pivot transformation, slow changing dimension ...)
  - **Document Management / Injection** (ECM scope) : providing features to manipulate documents is important for Vanilla platform, because Vanilla produce a lot of documents (PDF, Excel formats) using various types of data. Some of those documents will be accessible or stored in enterprise ECM services

- BI GateWay does not target to compete in the « pure ETL players », but rather stands as one of the services (data services) for the Vanilla platform. Knowing this, objectives are differents from other ETL, and comparison is difficult & does not make sense. In terms of functions, we more focused of « ETL functions for BI process »

- Key features for BI GateWay will reproduce Vanilla reporting features, such as :
  - dynamic database connection definition (for both source and target database),
  - column content security management (to hide the content of columns based on FreeMetadata security definition),
  - Secured & reliable package definition for deployment between servers

Integration with other Vanilla components

BI Gateway is abble to communicate with other Vanilla components and use Vanilla ressources definitions. One of the objectives of Bi GateWay is to provides « data services »

Please understand that with this very first version of Bi GateWay, all those features described bellow may not yet be available

- Integration with Vanilla BI Platform
  - Use of Vanilla global variables, Vanilla User Defined Interfaces ...
- Integration with BI Scheduler
  - schedule of Bi GateWay Datastreams
- Integration with BI Workflow Designer
  - Use of additional ressources available in Workflow Designer (interfaces)
  - Datastream integration, to create complex workflow
- Integration with FreeAnalysisSchemaDesigner
  - management & schedule for aggregate tables
- FreeMetaData
  - Snapshot creation, using business views
  - Dataset structure are saved as table in FreeMetadata
- Reporting Integration
  - Possibility to use Datastream XML Output as datasource in Birt
  - Possibility to use Datastream XML Output as datasource in FreeMetadata

Architecture & Installation

Bi GateWay download

Download Bi GateWay package at the following URL :

http://www.bpm-conseil.org/bigateway

Additional downloads

Vanilla 1.31 is available for download at http://www.bpm-conseil.org/vanilla131

In order to develop Vanilla BI application, you need to download & install the various design packages : http://www.bpm-conseil.org/vanilla131/designer-1.31/

In order to manage Vanilla server & application, you need to download & install the various admin packages : http://www.bpm-conseil.org/vanilla131/administrator-1.31/

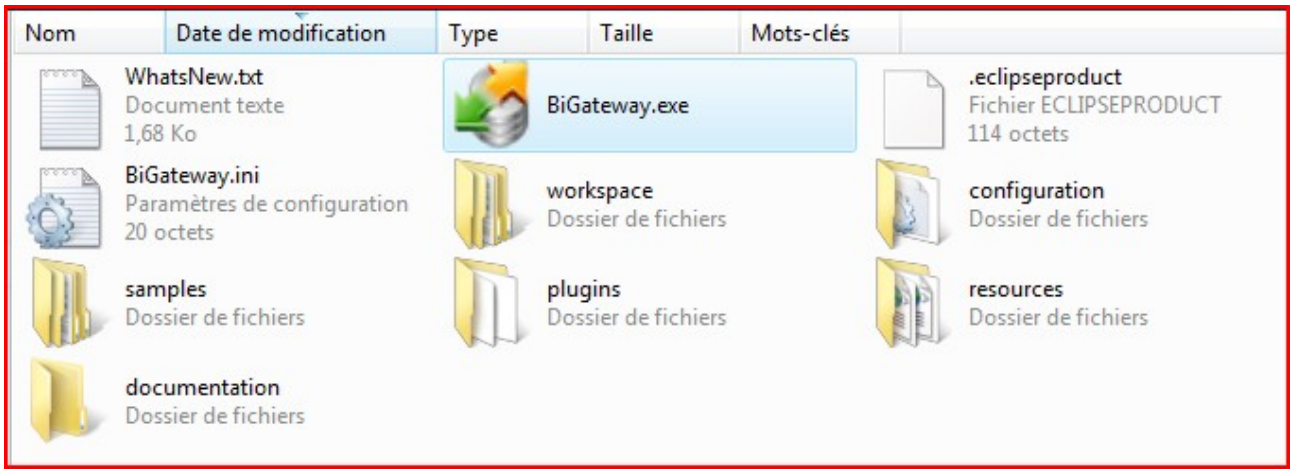
If you want to test Vanilla 1.31 outside of MySQL area, simply download, unzip and run the embeded HSQL version : http://www.bpm-conseil.org/vanilla131/vanilla-tomcat-hsql-1.31-ce.zip

Installation

Unzip the Bi GateWay package or run the installer (when available).

Don't unzip in a directory that contains space in name.

Run BiGateWay.exe (windows or linux version) to start a session



Architecture

BI GateWay can run either in standalone or along with a Vanilla BI server installation

Bi GateWay development are stored in XML documents that can be edited outside of BI GateWay interface.

BI GateWay is a full functional Eclipse package, you need JVM 1.5.10 minimum

Vanilla Version

To be abble to use FreeMetadata documents as datasource, you need to have a Vanilla server up & running, version 1.31 minimum

To be abble to save your datastream into Vanilla repository, you need a Vanilla server up & running, version 1.40 minimum.

Directories

Once unzip, Bi GateWay is delivered with a common set of folders, where you can find usefull samples & configuration files.

Resources

Ressources folder contains XML documents that references external ressources such as database connexion, jdbc driver setup, Vanilla server connection …

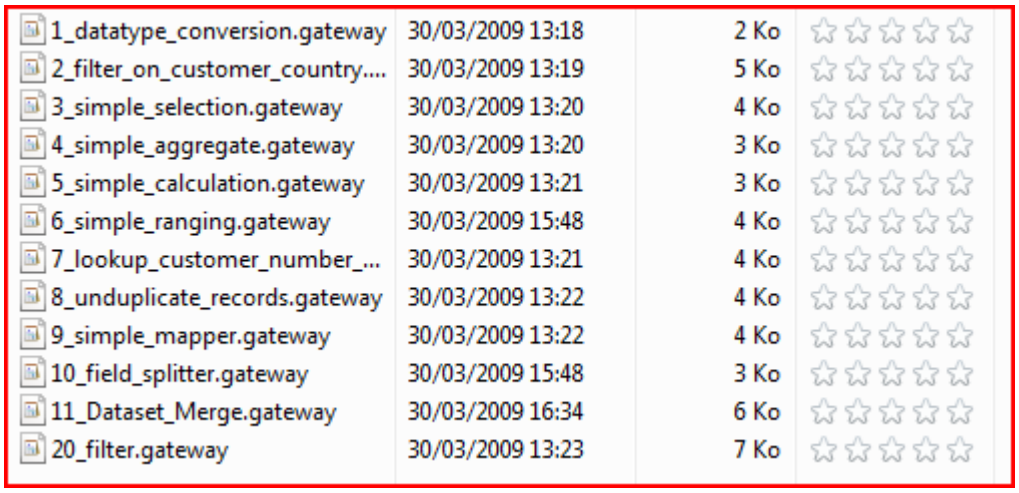
All those ressources all declared and stored in xml file : BiGateway\resources

resources.xml
connections.xml
driverjdbc.xml

Jdbc drivers are available in subfolder «jdbc »

Samples

BI Gateway is provided with many samples that developer can use to view & learn how to develop DataStream with Bi GateWay. This samples folder are detailed in the annexe of this document (use the number of the sample to get direct access)



1_datatype_conversion.gateway	30/03/2009 13:18	2 Ko
2_filter_on_customer_country....	30/03/2009 13:19	5 Ko
3_simple_selection.gateway	30/03/2009 13:20	4 Ko
4_simple_aggregate.gateway	30/03/2009 13:20	3 Ko
5_simple_calculation.gateway	30/03/2009 13:21	3 Ko
6_simple_ranging.gateway	30/03/2009 15:48	4 Ko
7_lookup_customer_number_...	30/03/2009 13:21	4 Ko
8_unduplicate_records.gateway	30/03/2009 13:22	4 Ko
9_simple_mapper.gateway	30/03/2009 13:22	4 Ko
10_field_splitter.gateway	30/03/2009 15:48	3 Ko
11_Dataset_Merge.gateway	30/03/2009 16:34	6 Ko
20_filter.gateway	30/03/2009 13:23	7 Ko

Documentation

Documentation folder contains the different PDF for documentation

How to start using Bi GateWay

Use the samples availables

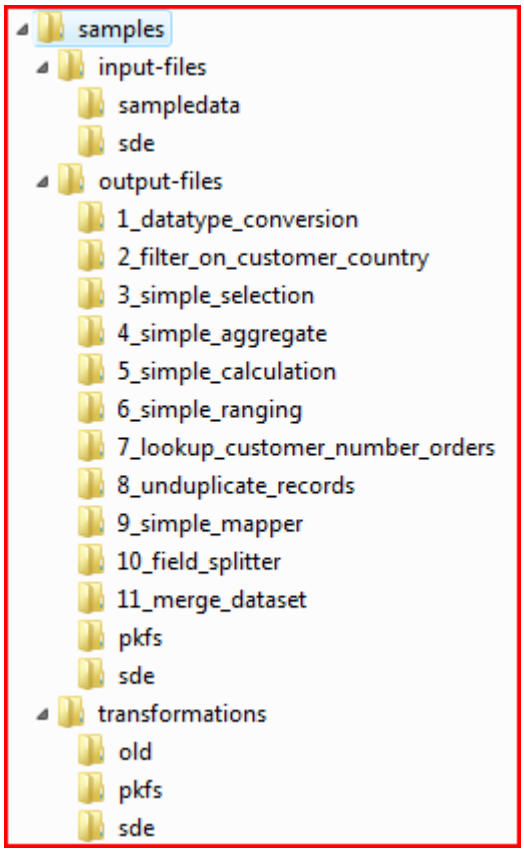
One simple method to start using BI GateWay is to browse through the installation folder and open the differents samples provided with the default installation. All those samples describe in detail each of the transformation type available, using the welknown sampledata database, in an Excel spreadsheet format (no database installation).

For a complete overview of the available samples, please reference to the last section of this document, in annexe : « browsing the samples »

Bi GateWay is delievred with a lot of transformations samples (see bellow)

All the samples available in the /transformations are using data available as text imput in the sampledata folder. All the output are available in a folder whose name is the transformation's name.

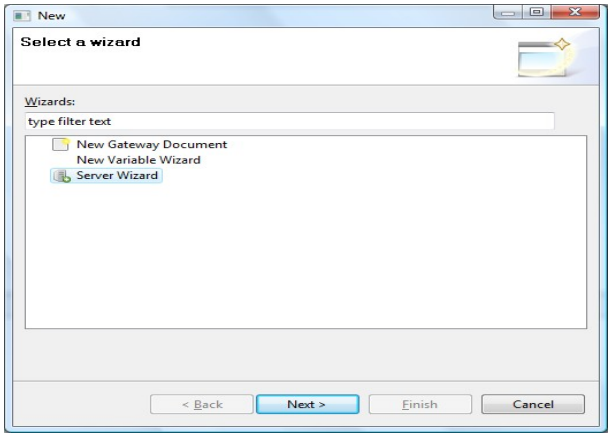
« sde » folder references Sylvain Decloix's standard tests, in reference with job done by Sylvain when trying to compare other ETL on a performance perspective.



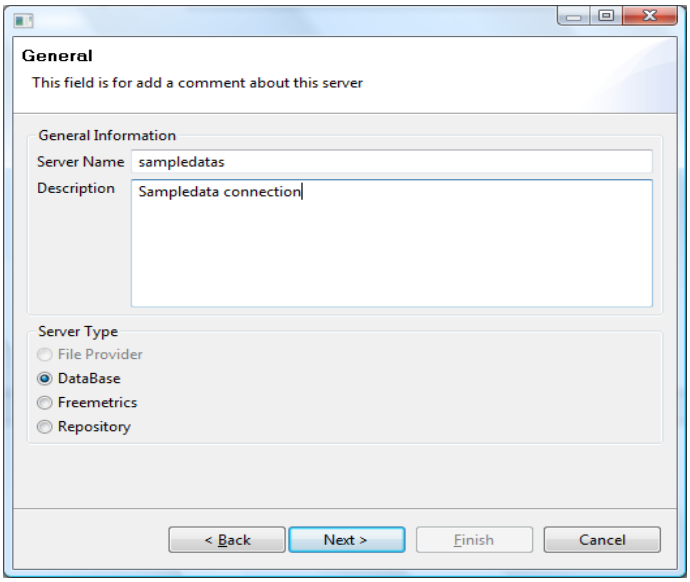
Define your resources

In order to access your own data, you need to define your database connection profile

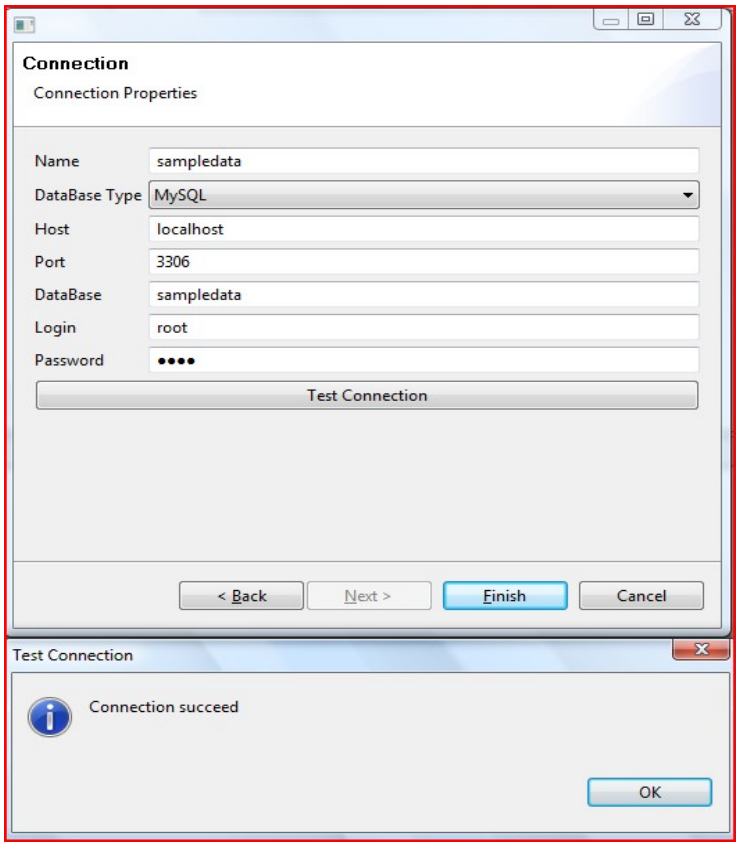
To define your database, choose menu « File + New Object », and choose « Server Wizard »



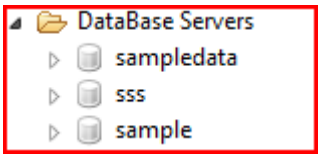
Select « Database » as Server Type



Define the connection profile



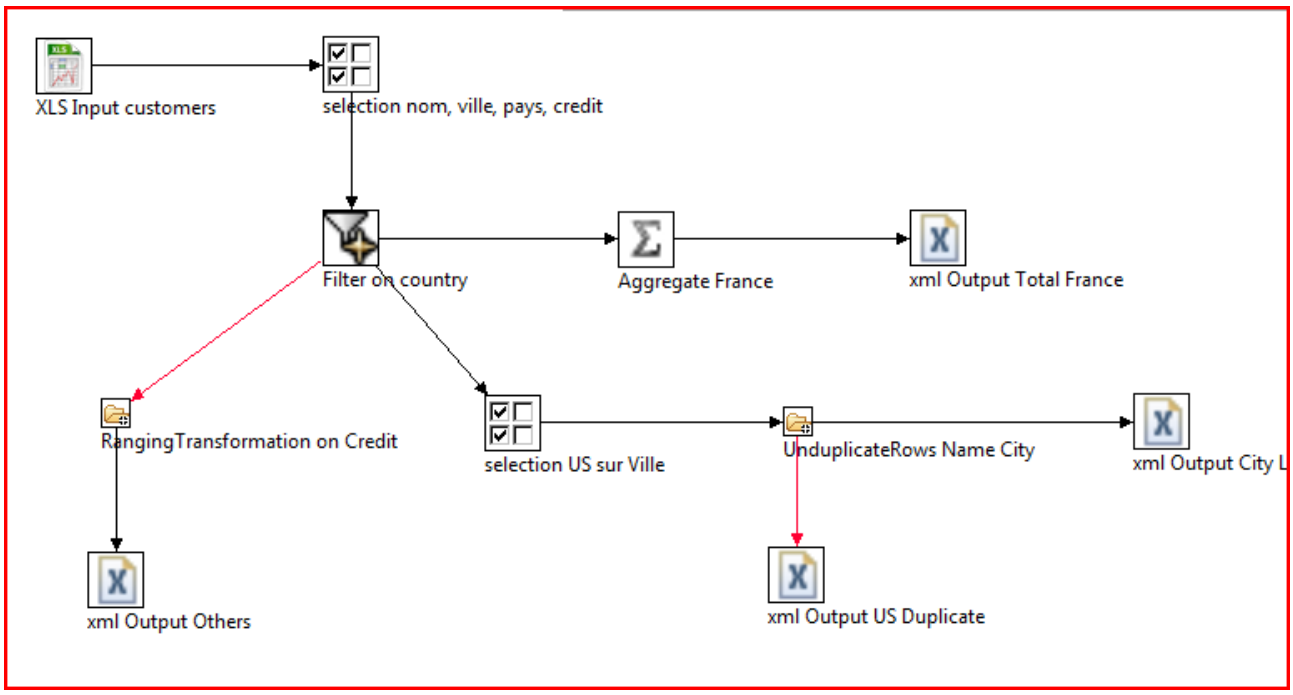
Your database connection will be available in your database server list, and those database profile can be used when you select « SQL Input » datasources



Bi Gateway tools & ressources

This part gives you an overview of the different panels available when using Bi GateWay

Workspace

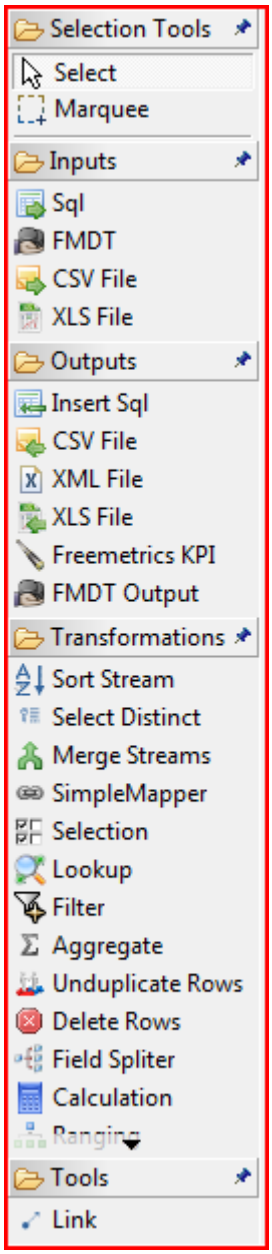


ToolBox

Toolbox is explained in detail later in this document

Toolbox contains all the object type you need while developing a datastream

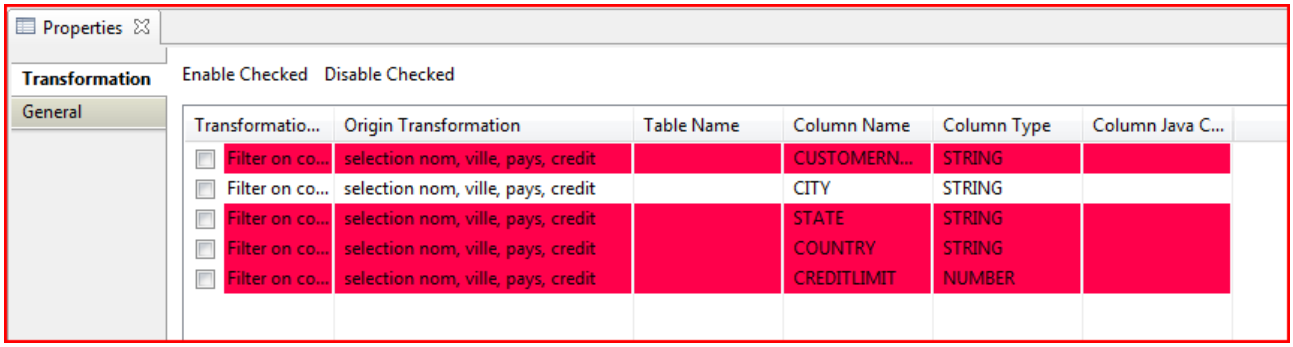
Toolbox are arranged in different subjecvt/folder : Imputs, Outpus, Transformations



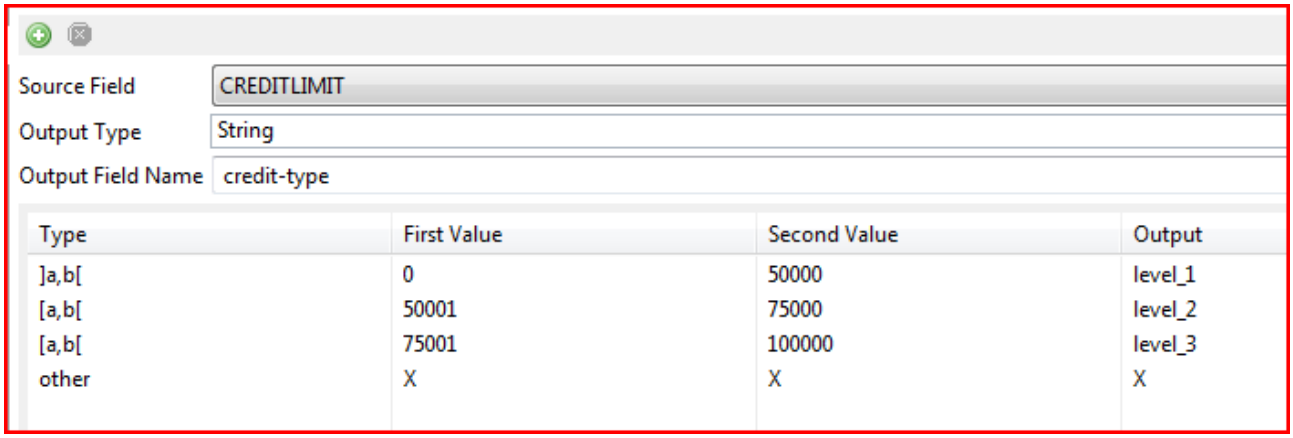
Properties

Properties are interfaces where developers set their properties for every toolbox

example of property for a mapping



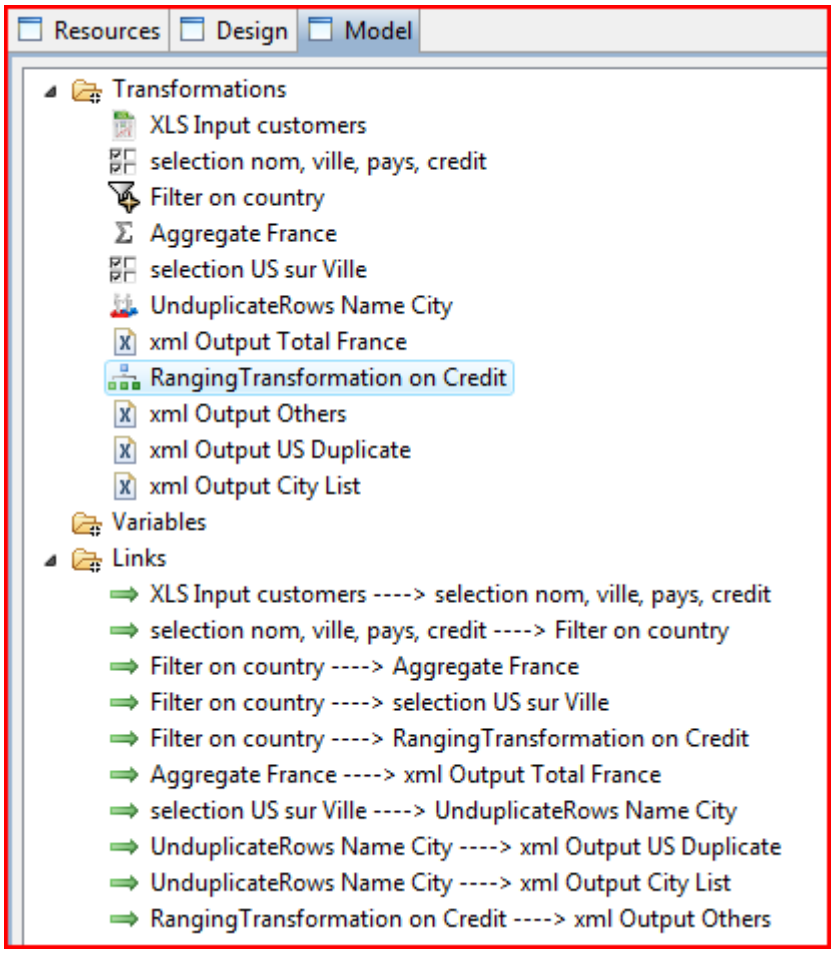
example of property for a ranging



Datastream list of components

One the right side of your interface, you can access to a treeview that contains the different boxes and links definied on your workspace.

This panel gives you an overview of the structure of your datastream



How To Develop DataStream

Bi GateWay can be used to develop simple or advanced datastream.

Objective is to take data from one place (Extract process), perform some transformations on the dataset (Transform process), and store the different resultsets in differents places (Load process)

Each datastream contains :

- one or many datasource input
- none to many transformations (in case there is no transformation, we call the datastream a simple data transportation, from one place to another one)
- one or many target output

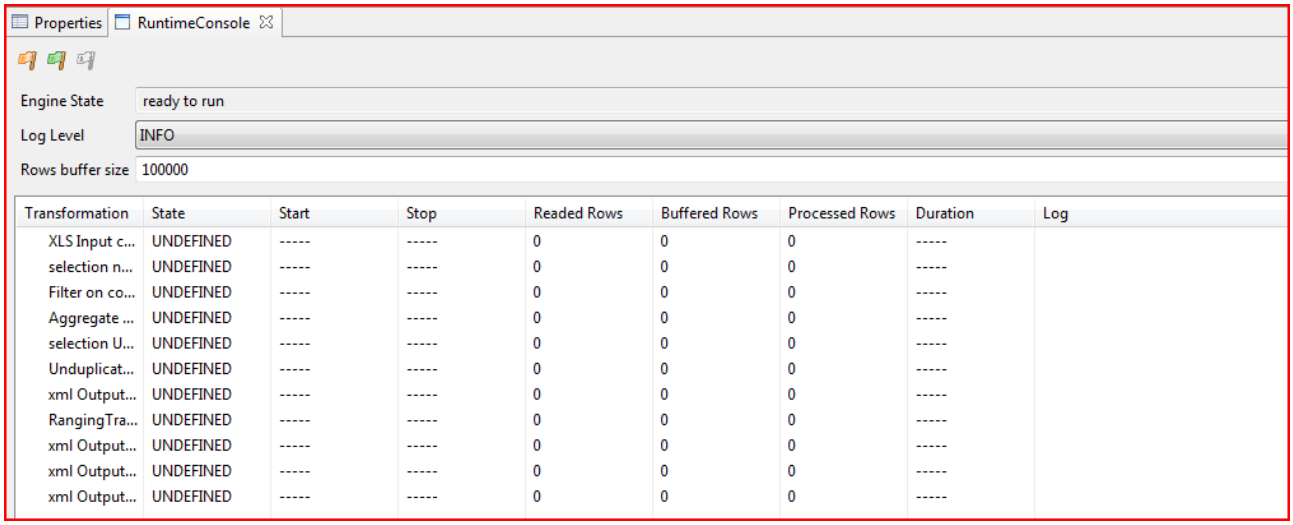
Development of a Datastream is very simple, you just create a new BiGateWay document and drag & drop boxes from Toolbox to the workspace, and enter the different parameters to setup your transformation.

Data Flow is drawn by Links, that indicates the direction for data from boxes to boxes.

Concept behind BiGateWay is to be abble to manage various kind of database type (Oracle, MySQL …) or data input type (XML, Excel, FreeMetadata) and use the transformation boxes to manipulate datasets (which means that every transformation boxes are available for any type of data input)

Once completed, use the « Runtime / Prepare Runtime» menu to access to the console from where you can initialise and run every datastream.

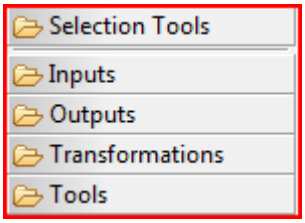
Use Init, to prepare the Runtime execution, and then Run to start the Datastream



Bi GateWay Toolbox

Bi GateWay provides developers with already configured boxes that represent various kind of data manipulation & transformation, such as database request, filter or aggregate transformation, or excel spreadsheet creation

There are 5 groups of boxes that group actions by actions type.
Each item for every group is explain in detail later in this chapter



- Selection
  Contains 2 boxes : Select (turn back the mouse to select mode) and Marquee (available in future release for multi boxes selection)

- Inputs
  Contains differents kinds of objects supported as « data input »

- Outputs
  Contains different kinds of objects supported as « data output »

- Transformations
  Contains the different transformations that developer can apply on any datastream

- Tools
  Contains the Link object, used to link 2 boxes together along with indicating the datastream orientation (from where are data coming, and what are their destination)

Boxes contain common tabulations, such as general (to set the name of the object) or transformation (for the transformation boxes, to define the transformation formula), or server (for input and output dataset definition, to define the database server).

Based on the box type, you can also have addition tabulation

Inputs

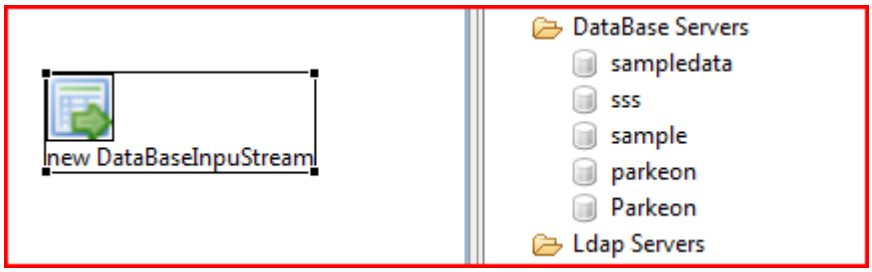
Inputs boxes manage data input, and allow various type of data

Icon	Name	Comment
	SQL request	SQL datasource (jdbc drivers)
	FreeMetadata document	FreeMetadata datasource
	Text File	Text file Input
	Excel Spreadsheet	Excel spreadsheet

The following datasources are not yet available

- XML file
- Web Services request
- Access Database

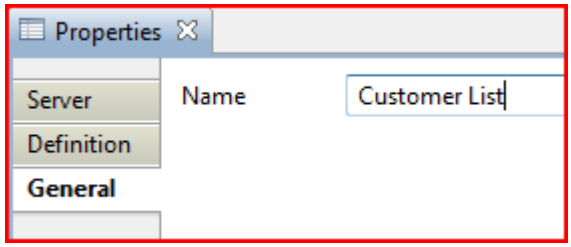
SQL Request



Use this toolbox to access to any kind of SQL datasource (jdbc drivers).
You must first have defined your database connection in the database list

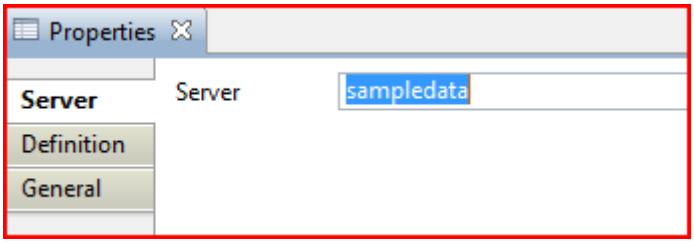
General

Enter name for the box



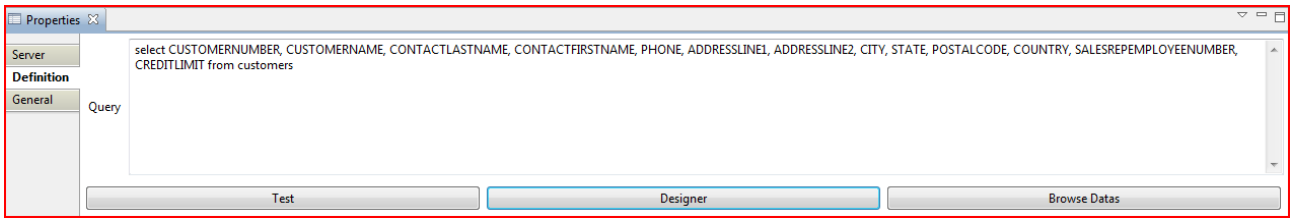
Server

Enter the database server name

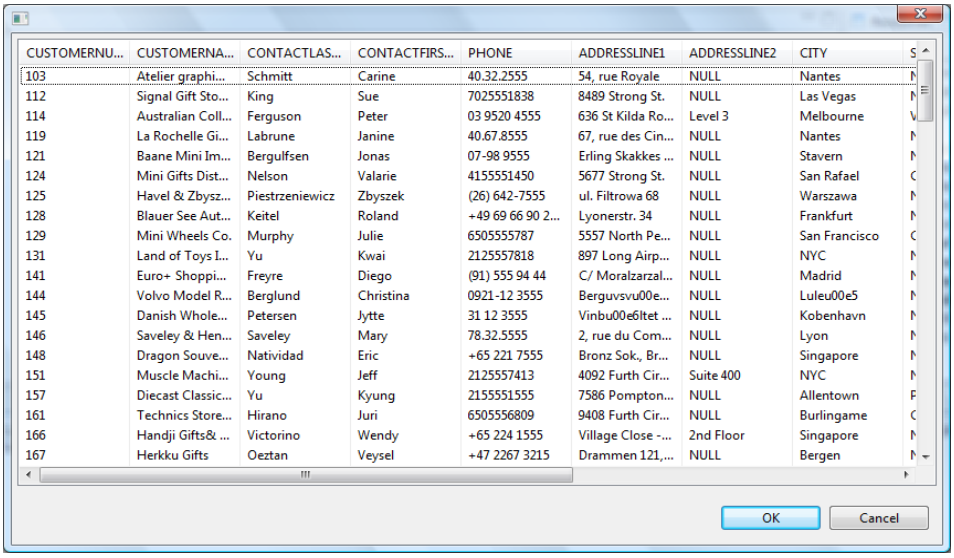


Definition

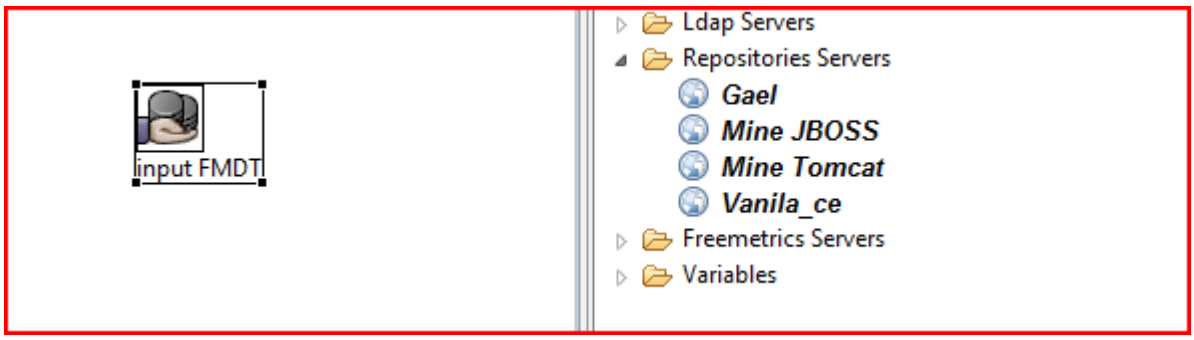
Defined the SQL statment. You can use the Designer interface to access to a common SQL interface where you can visualize your schema, tables & views and create your SQL Statment



Browse data feature



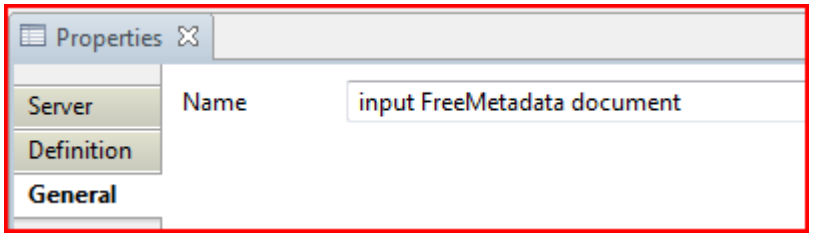
FreeMetadata



Use this toolbox to access to any FreeMetadata datasource, available in Vanilla Servers
You must first have defined your Vanilla connection in the Repositories Servers list

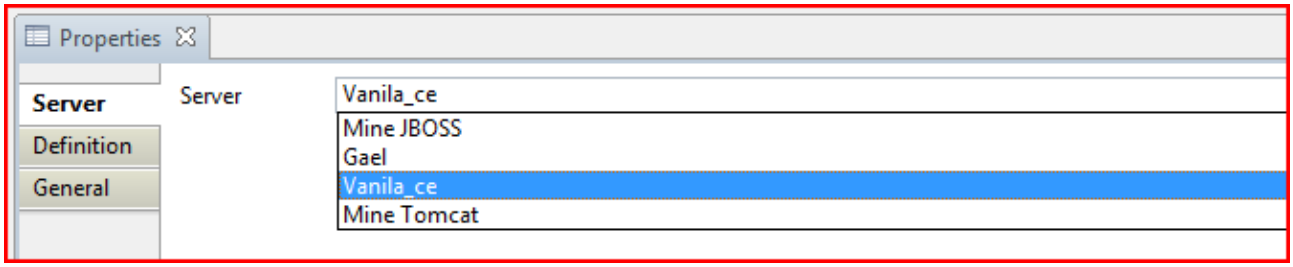
General

Enter name for the box



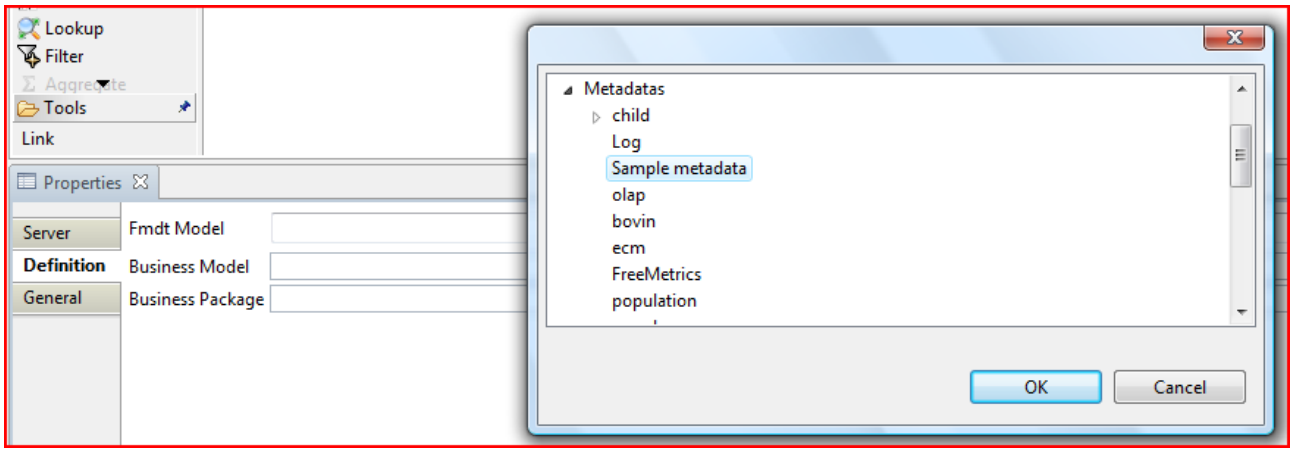
Server definition

Enter the Vanilla server name

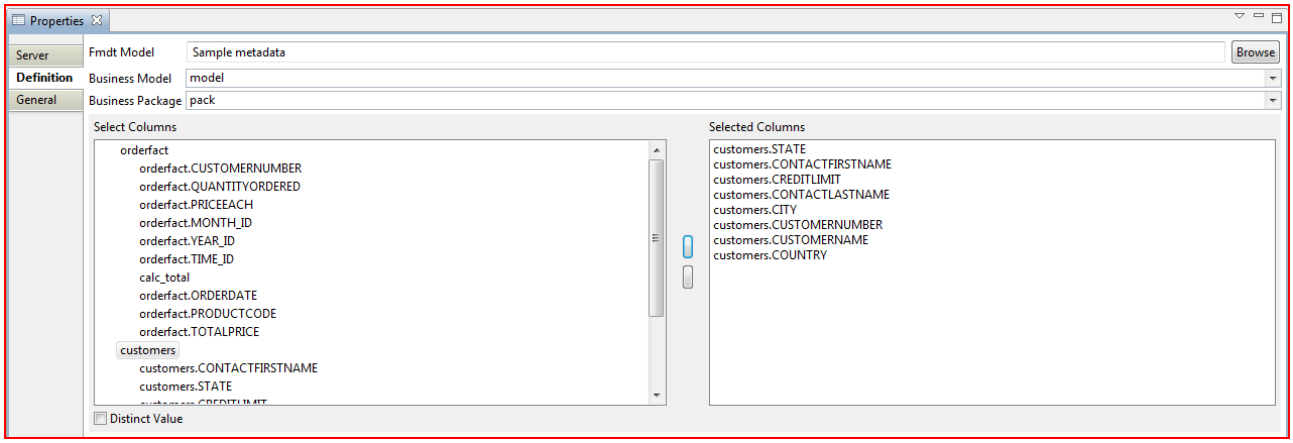


Datasource Definition

Enter references for FreeMetadata document, FreeMetadata Business Model and FreeMetadata Package. For additional information about FreeMetadata, please refer to the FreeMetadata documentation.

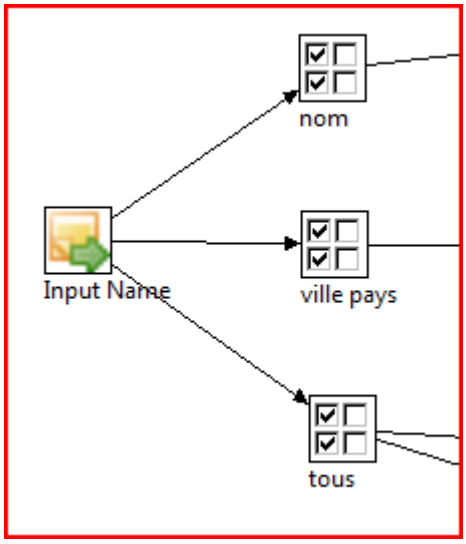


Select your business items (columns)



(be carefull with the 2 buttons, they don't contains 'arrows')

CSV File

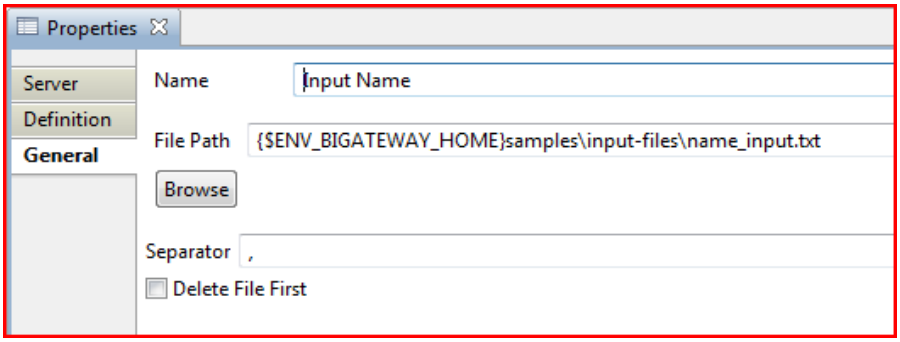


Use this toolbox to access to any Text file

General

Enter name for the box
Enter path for the file
Enter separator for columns in the file

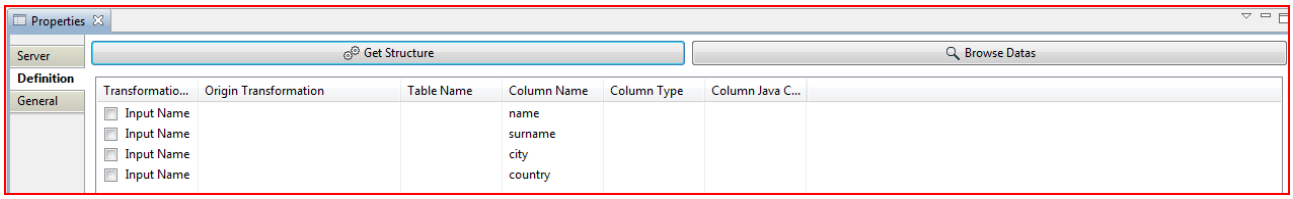
Use variable to prevent hard coded references



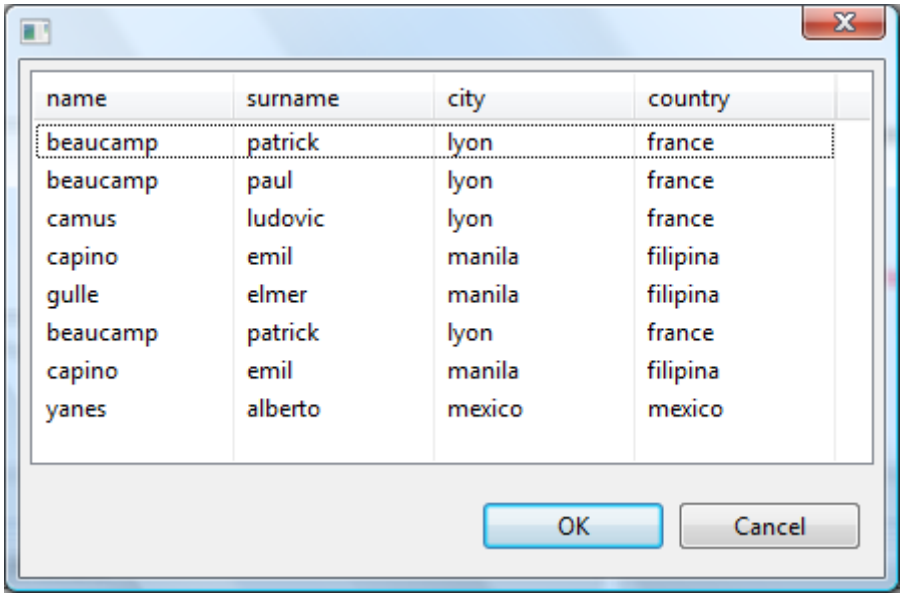
Limitation : first column should contain column's name

Definition

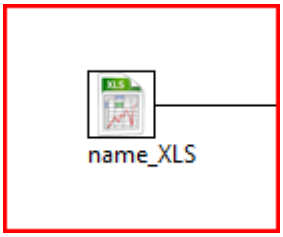
Use the Get Structure button to retrieve the structure of your file



Browse data feature



XLS File

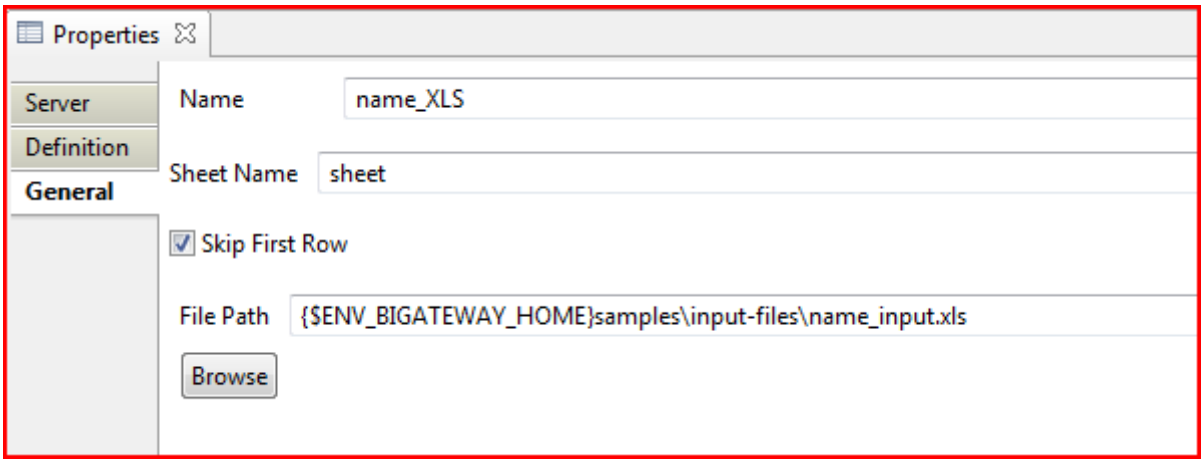


Use this toolbox to access to any Excel file

General

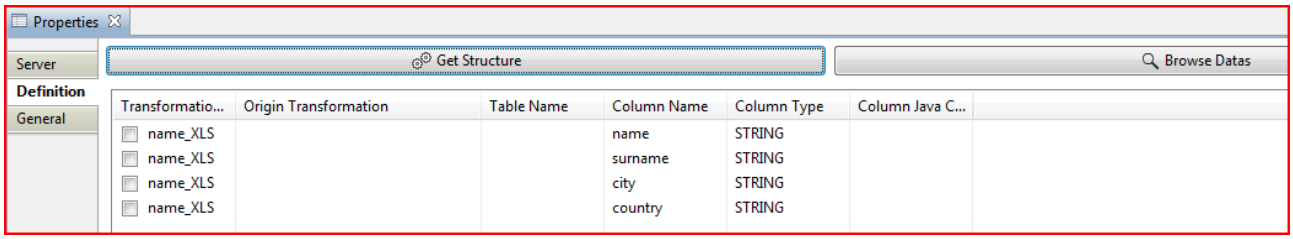
Enter name for the box
Enter path for the file
Select Sheet name in the sheet list
Check or uncheck « skip first row » whether or not first row contains column labels

Use variable to prevent hard coded references

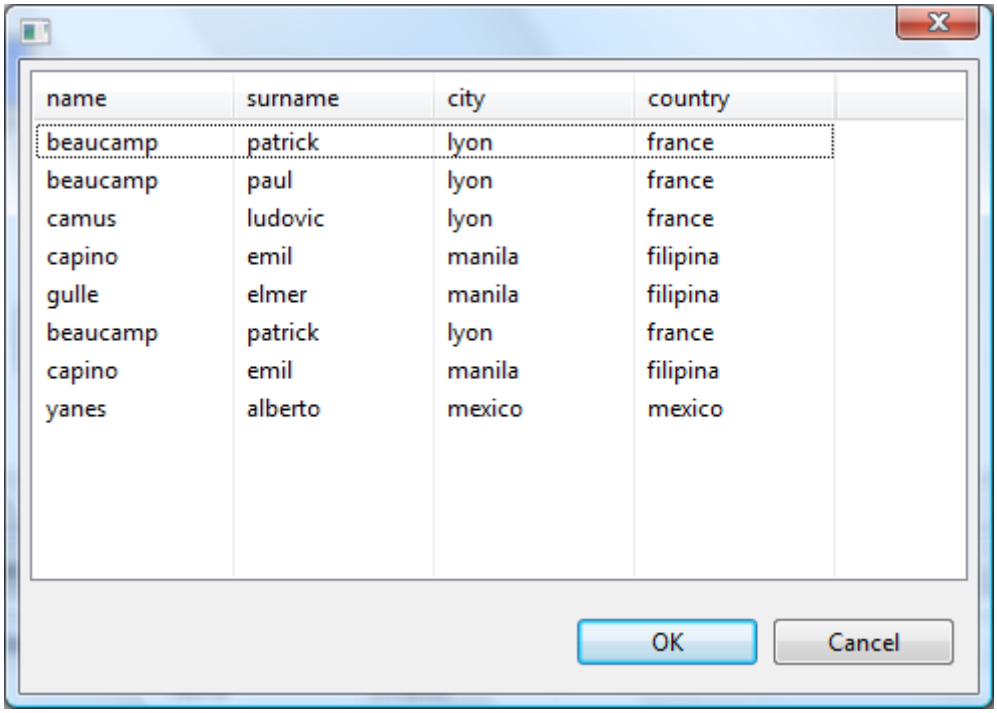


Definition

Use the Get Structure button to retrieve the structure of your file



Browse data feature



Outputs

Outputs boxes manage data output, and allow various type of data or objects to be manage as « target » for your dataset

Icon	Name	Comment
	SQL	SQL datasource (jdbc drivers)
	Text File	Text file Output
	XML File	XML File Output
	XLS Spreadsheet	Excel spreadsheet
	FreeMetrics KPI	FreeMetrics KPI target. Usefull to load Metrics values using BiGateWay
	FreeMetadata document	FreeMetadata target

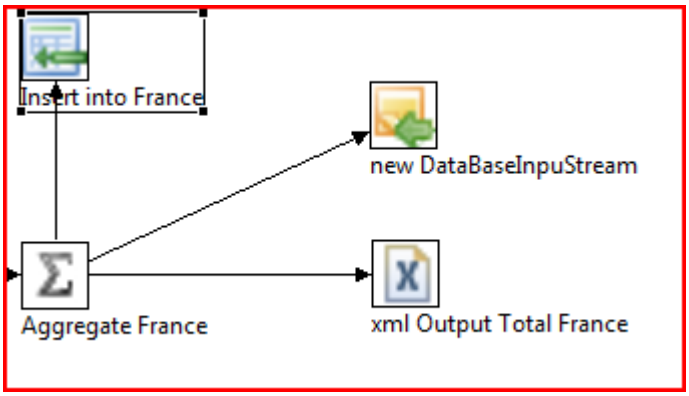
The following target types are not yet running in version 0.21

- FreeMetrics KPI
- Freemetadata document

The following target types are not yet available

- Access Database

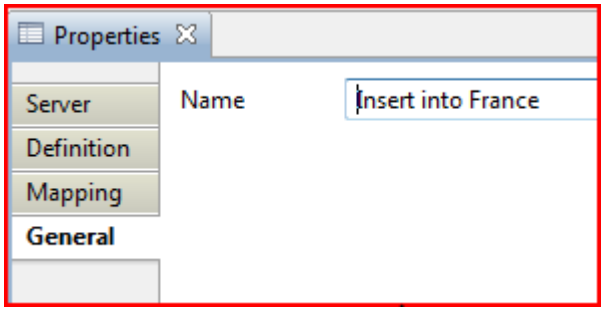
Insert SQL



Use this toolbox to access to any kind of SQL datasource (jdbc drivers) and save your data in tables
You must first have defined your database connection in the database list

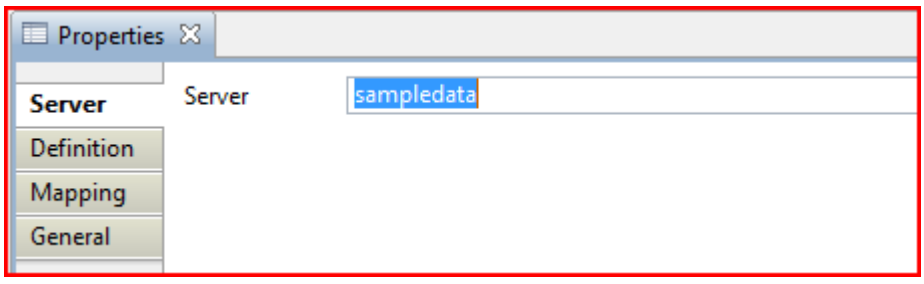
General

Enter name for the box



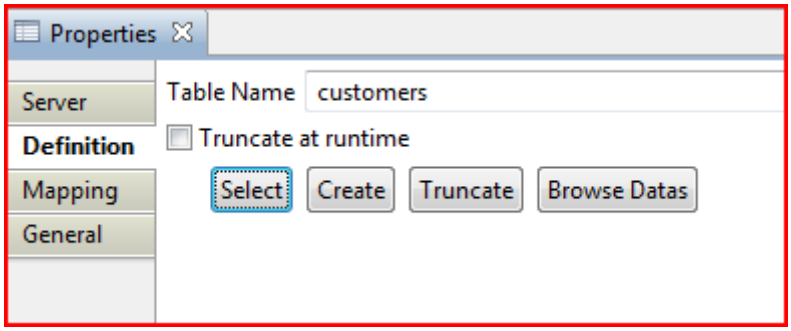
Server

Enter the database server name



Target Definition

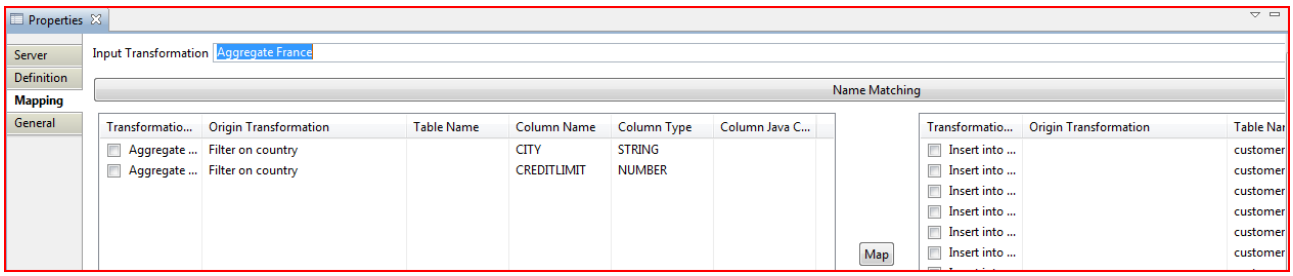
Select the table where you want to store your dataset.
You can notice you have functions to manipulate your data (truncate table at runtime for example)



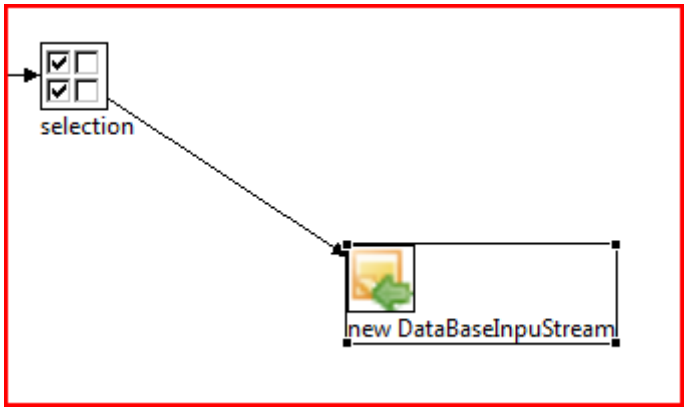
Mapping

Using Mapping, you can associate your dataset column with columns in your target table.
Insert statment will be a « one to one » insert statment, meaning that every row in your dataset will be (or try to be) inserted one by one in your table.

Be carefull, some insert statment can be rejectec due to your table's constraints (such as foreign keys or null values)



CSV File

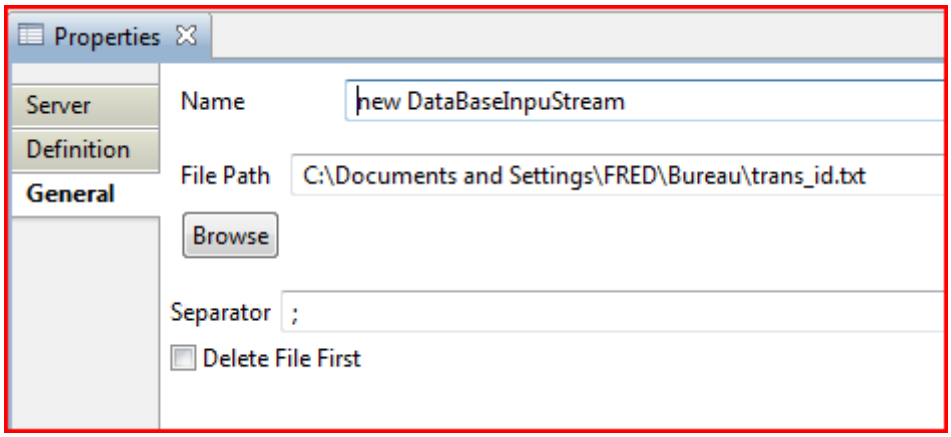


Use this toolbox to save your data in simple files

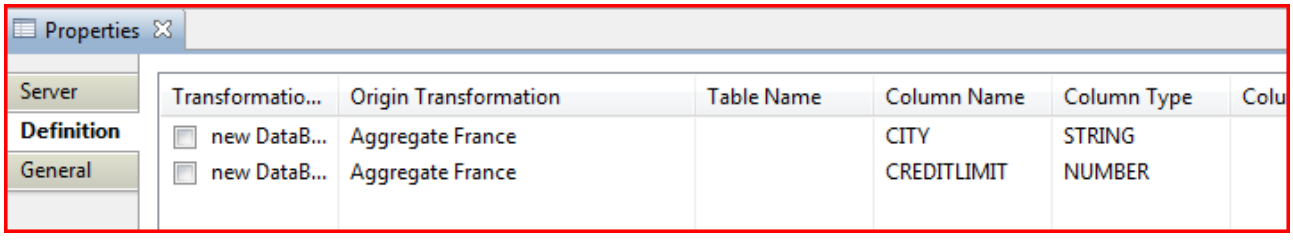
General

Enter name for the box
Enter your separator
Indicate if you want to delete the file first (when exists)

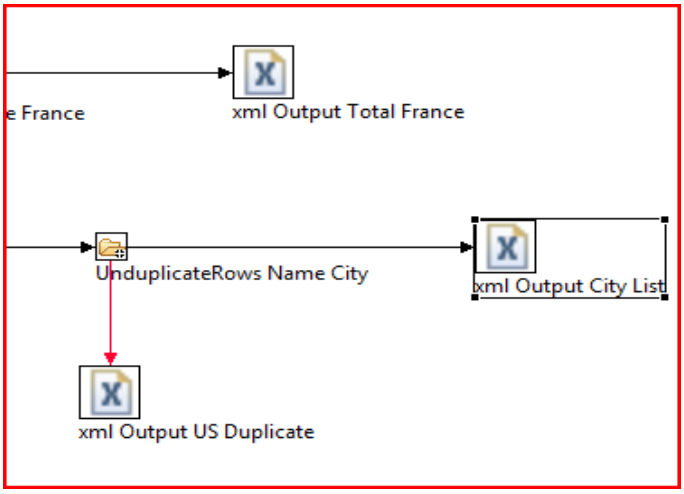
Use variable to prevent hard coded references



Definition
You can see which columns will be saved (with their type)



XML File

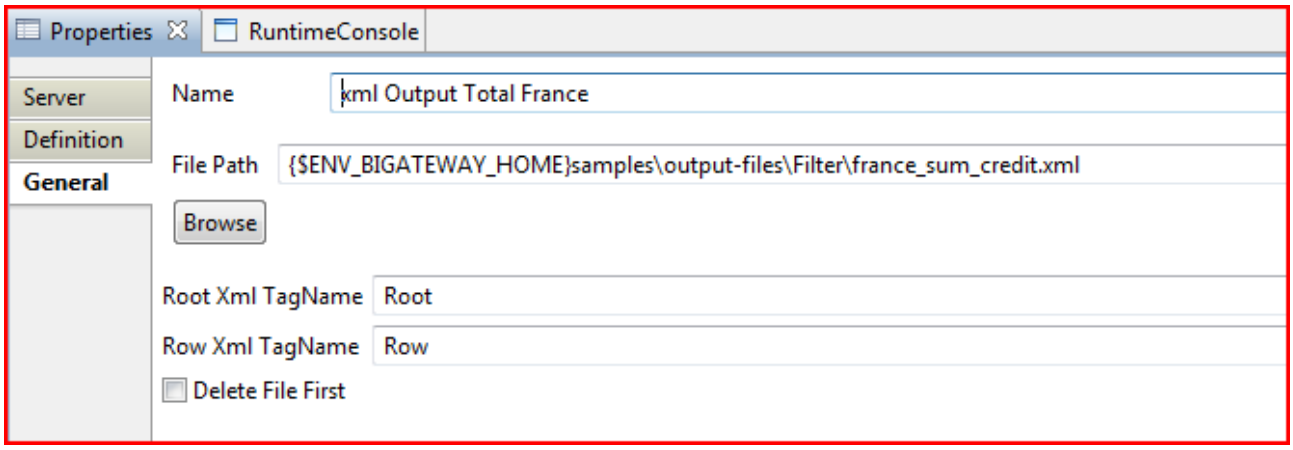


Use this toolbox to save your data in XML files

General

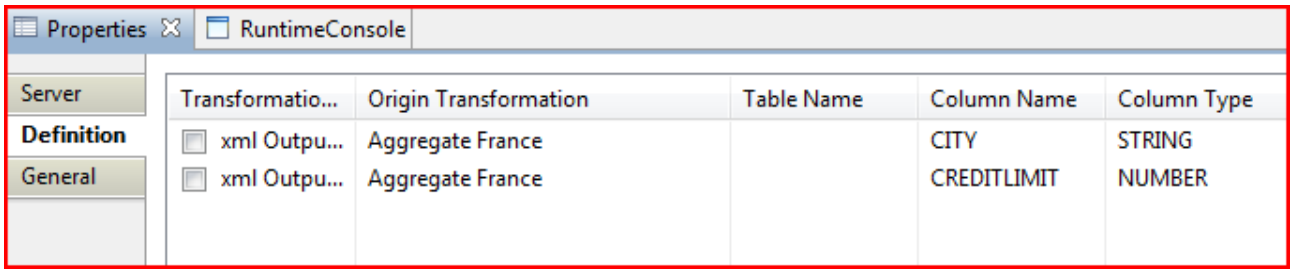
Enter name for the box
Enter Structures indication (Root & Row) to give structure instruction for your XML file
Indicate if you want to delete the file first (when exists)

Use variable to prevent hard coded references

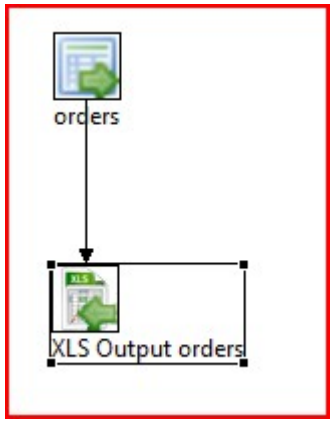


Definition

You can see which columns will be saved (with their type)



XLS File

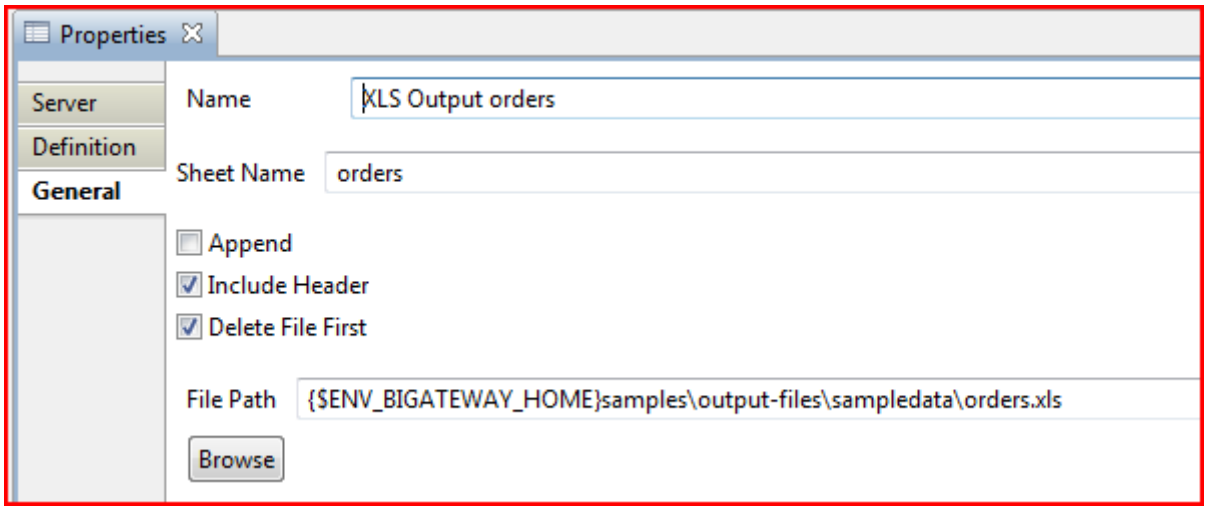


Use this toolbox to save your data in Excel files

General

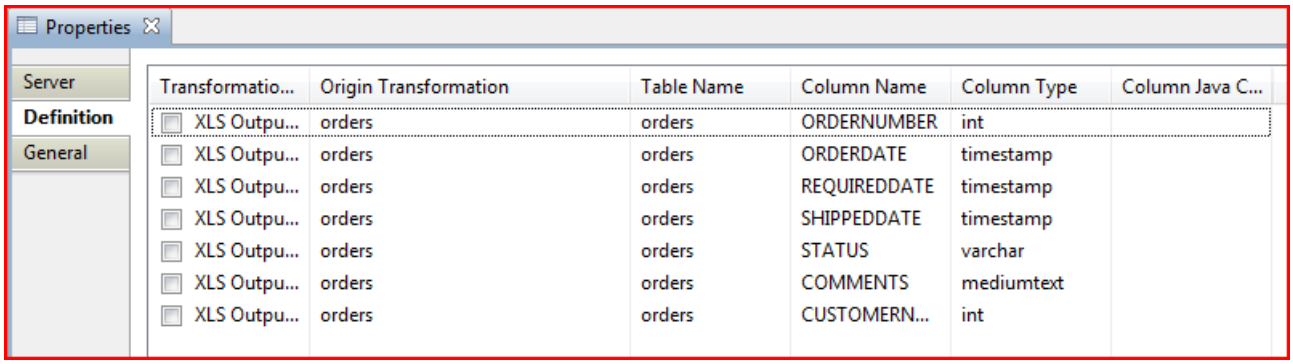
Enter name for the box
Enter Sheet name for Excel spreadsheet
Indicate if you want to delete the file first (when exists)
Indicate additional informations (Include Header, append in file)

Use variable to prevent hard coded references



Definition

You can see which columns will be saved (with their type)



BPM Conseil

FreeMetrics KPI

To Be Completed

FreeMetadata Output

To Be Completed

Transformations

Transformations boxes manage data transformation tasks, and provide numerous kind of data manipulation, based on datasets declared in the datastream

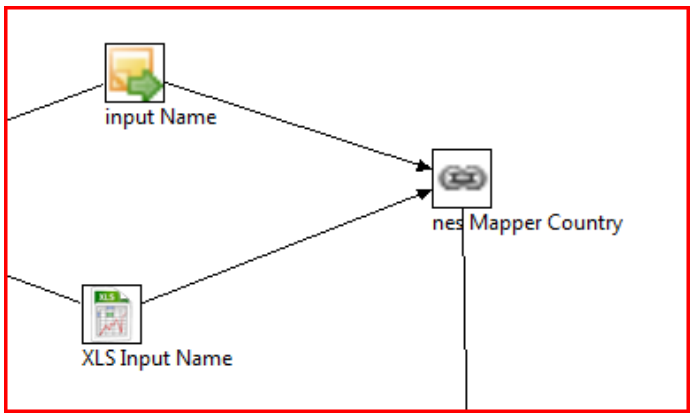
Icon	Name	Comment
	Sort Stream	Allow developer to sort a datastream
	Select Distinct	Allow developer to set a « select distinct » statment on a datastream
	Merge Streams	Allow developer to merge 2 datasets, assuming those datasets have the same structure
	Simple Mapper	Allow developer to join 2 datasets, using join strategy on columns from each dataset
	Selection	Allow developer to unselect some columns from a dataset
	Lookup	Provide Lookup interface to call a lookup table
	Filter	Perform filter operation, with multi conditions and multi dataset as target
	Aggregate	Perform aggregate operation, based on group of data
	Unduplicate Rows	Provide a nice and easy interface to extract duplicate records and separate the dataset in 2 different subset
	Delete rows	Delete rows based on a condition
	Field Splitter	Provide an easy to use interface to split a field into different fields
	Calculation	Box where you can define calculation on columns of your dataset
	Ranging	Box where you can transform a numeric column, using its value and get a string as a result.

Under development :

The following transformation are not yet available in version 0.31, but will be available in 0.34

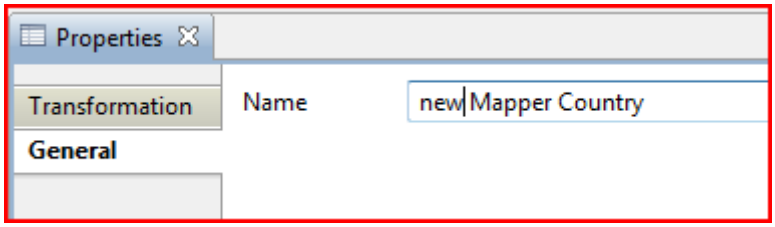
- Subtransformation

Mapper



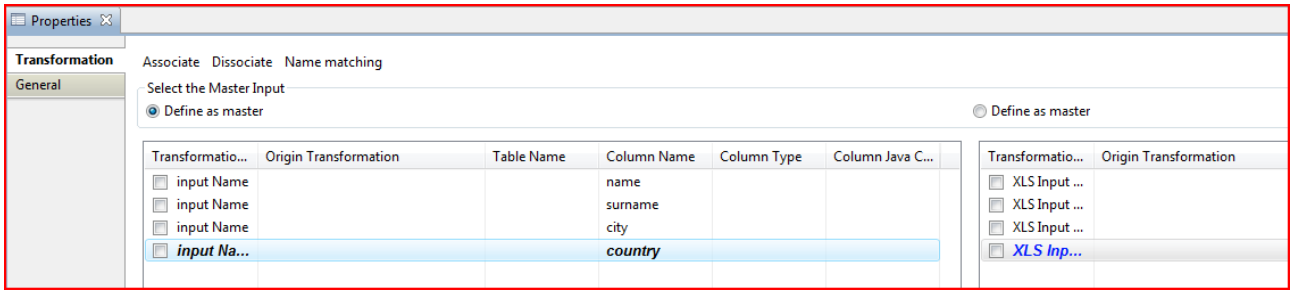
Allow developer to join 2 datasets, using join strategy on columns from each dataset. Mapper is similar to issue a SQL statment thta includes 2 tables, where you need to define the join strategy

General



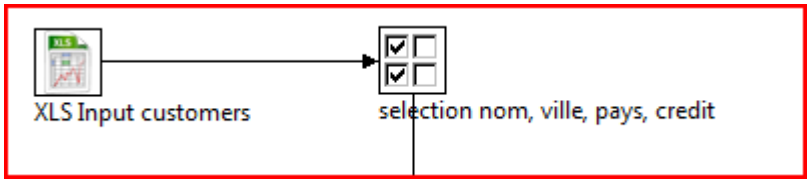
Transformation

In the transformation, you define the join strategy between the 2 recordsets.



Limitation : only join strategy with « = » are available

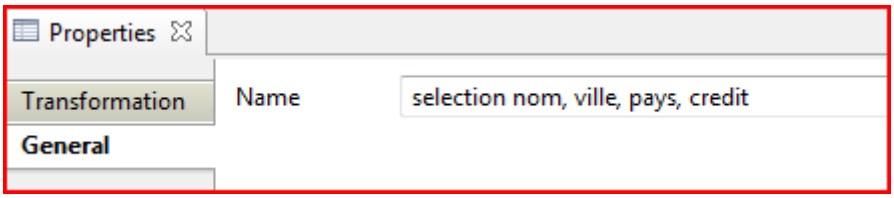
Selection



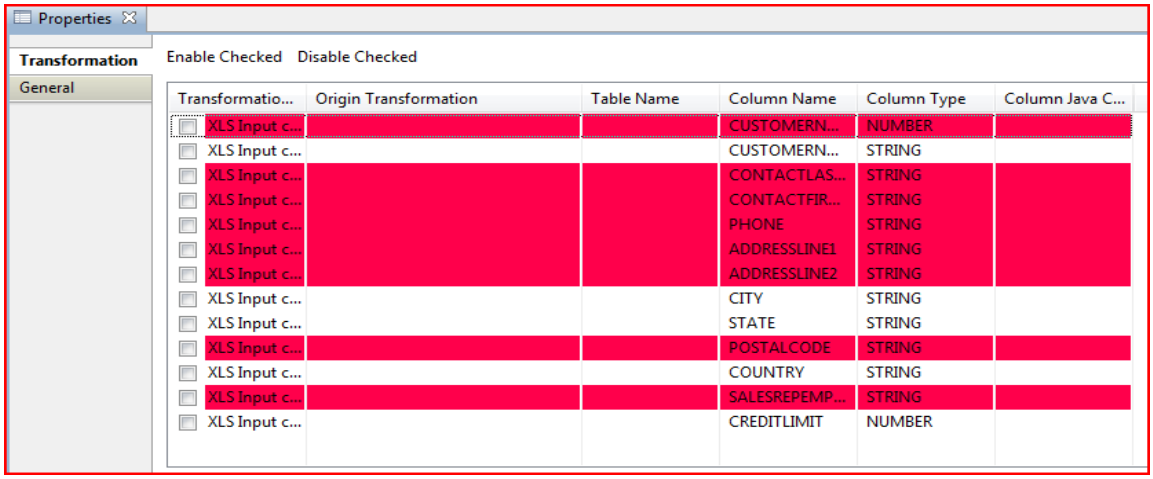
Selection allows developer to unselect some columns from a dataset.
Business usage : when you want to suppress some columns from your dataset, to perfom operation such as select unduplicate, or simply because your first SQL statment contains too many columns.
It's also usefull when you run a lookup, because all the columns from the lookup table are embeded in the resultset

Selection box has 2 tabulations :

General
Definition of the name of the box

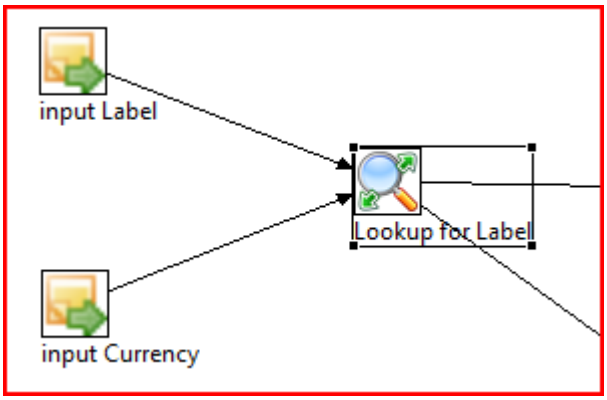


Transformation
Definie the columns you want to remove from the dataset, simply by selecting the column and choosing the On/Off buttons : Enable Checked or Disabled Checked



Current limitation
You can't rename the columns in this stage

Lookup



Lookup is a method to add additional information (colulmns) for each row of your dataset, assuming developer send the valuable parameters to retrieve only one record from the lookup dataset.

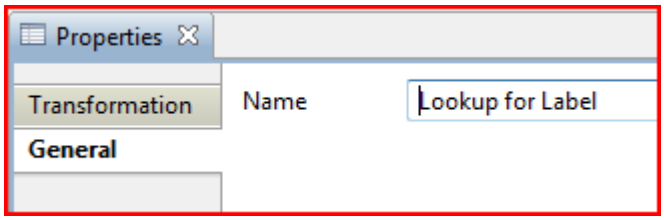
Lookup can be usefull to look for exchange rate (you send 2 currencies & a date, and lookup send you in return you daily exchange rate), or to unify some column's content (you have entry such as « Mr », « Mister », « M. » … and you perform a lookup in a table to unify this with « Mr »)

Like Mapper box, Lookup box uses 2 datasets in entry :

- Major difference between mapper and lookup is that mapper can result in a n * m join strategy, which can produce additional records, as opposed to lookup, which only add columns to existing records.
- For Lookup, you have to define which dataset is your main dataset, and the number of records after your data passed through the lookup box will be exactly the same as the number of records from your main dataset before the lookup process

Lookup box has 2 tabulations :

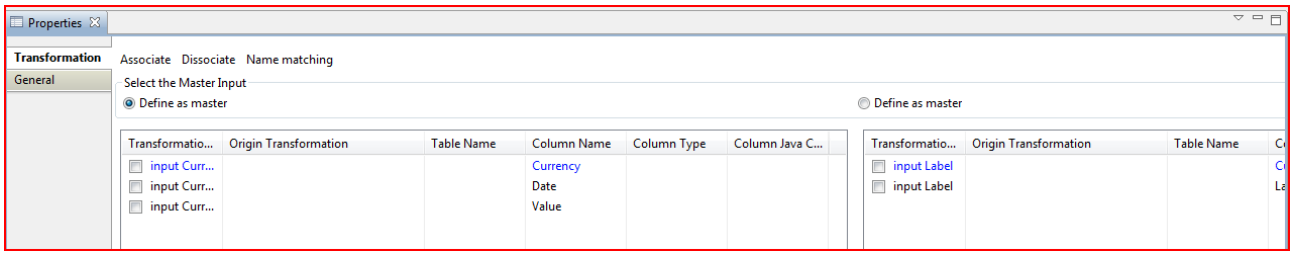
General
Definition of the name of the box



Transformation

In the transformation, you define which resultset is your master resultset, which turns automatically the other resultset as a lookup resultset.

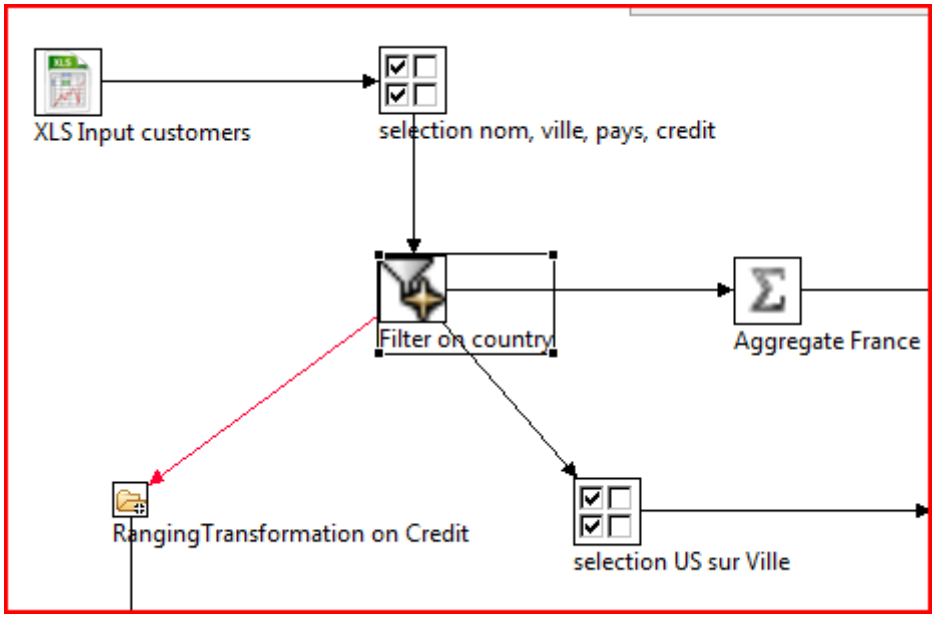
You select columns from the master resultset and from the lookup resultset and use the associate function to setup an « = » operation. You can use name matching feature to automatically display the columns from the 2 resultsets that have the same columns name.



Limitation

- Lookup are only available with dataset, you can't call any database stored procedure to retrieve your addditional columns
- All the columns from the lookup will be embeded in the resultset, which means that you have to apply a selection box to removed additional unexpected columns
- Retrieve conditions are based on test « = » - conditions such as « < » or « > » are not supported

Filter



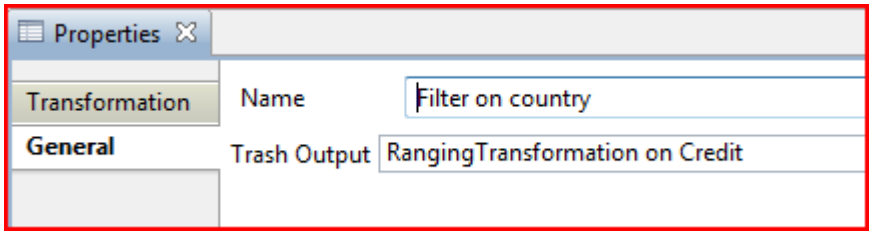
Filter allow developers to perform filter operation, with multi conditions and multi dataset as target, using a single dataset as major based dataset.

Filter is usefull to split a dataset into subdataset, then apply different transformation and even produce differents columns using those dataset

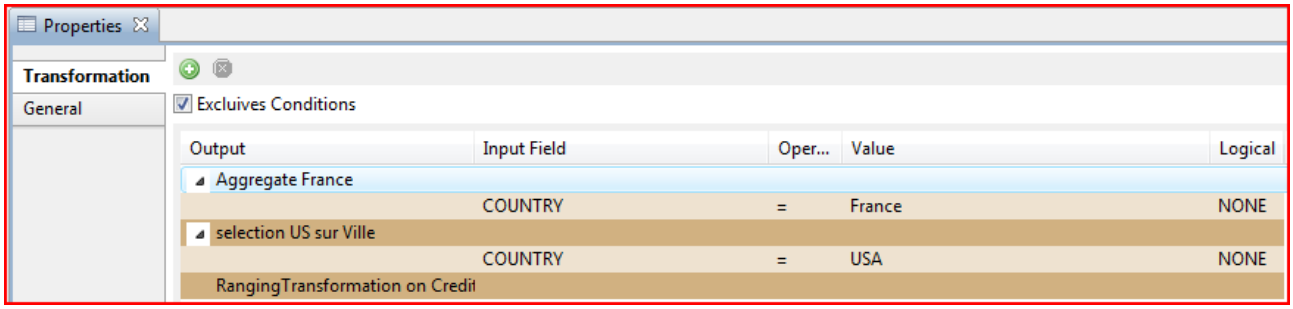
Filter box has 2 tabulations :

General
Definition of the name of the box

Define also the « Trash Output », which stands as the « by default » target when none of the previous conditions apply on the current record.



Transformation



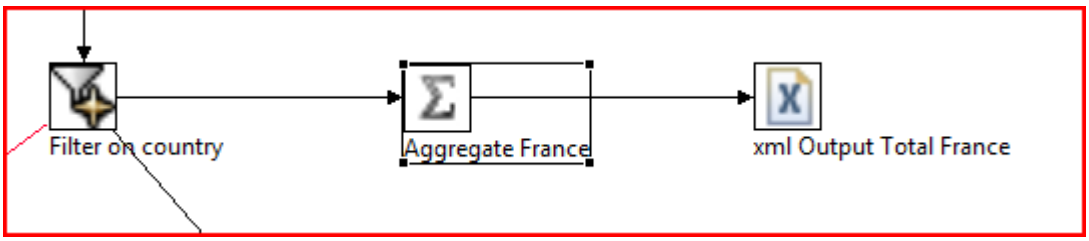
Define the differents conditions along with the target box/dataset that will receive the filtered data

Conditions are evaluated in the order they are displayed in the interface

For each Output, you can define multiples tests that combine together using logical operators such as AND, OR or NONE (if only one condition)

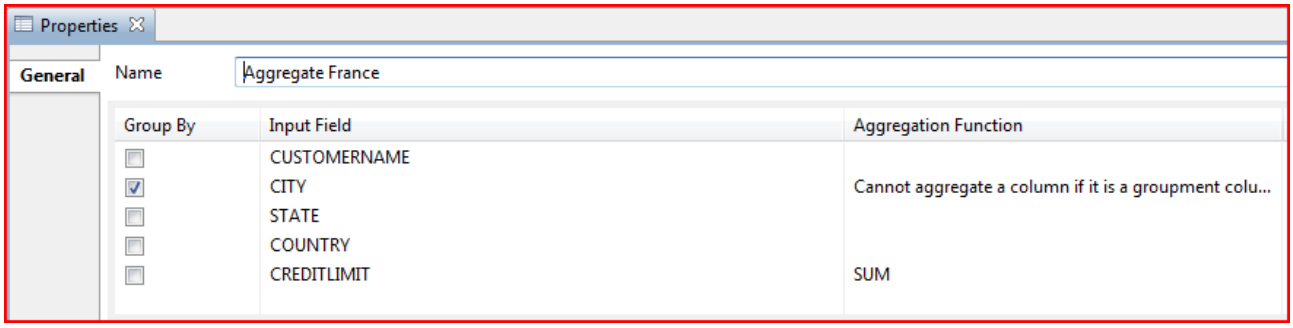
Exclusives condition check box indicates that each record can only be sent to one and one only output. When this check box is not validated, every tests will be run on each record, and record can be sent to many outputs (leading to data duplication, but this can also be usefull for business reasons).

Aggregate



Aggregate performs aggregate operation, based on group of data

There is only one interface to define both : box name and transformation

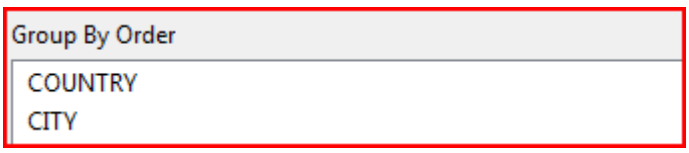


You select the columns you want to group, understanding that the columns which are not grouped and are not aggregate will be automatically removed from the resultset.

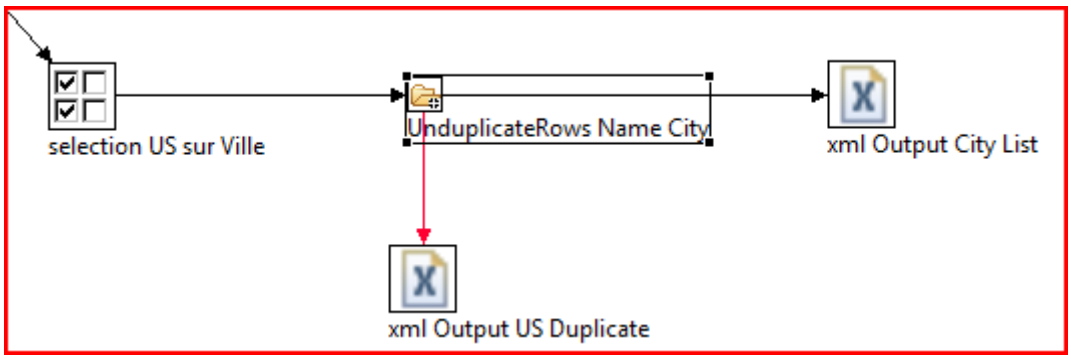
Available Aggregate functions are :

- SUM
- MAXIMUM
- MINIMUM
- AVERAGE
- COUNT
- DISTINCT COUNT

At the bottom in this interface, you can drag & drop to influence the order of the « group by »



Unduplicate

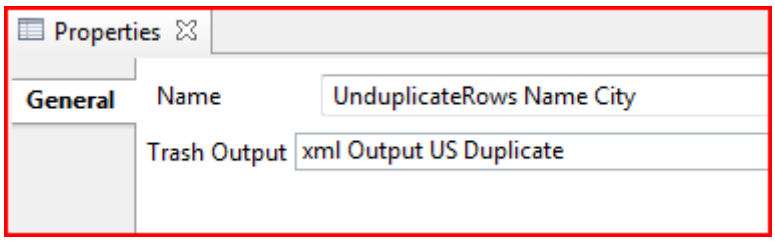


Provide a nice and easy interface to extract duplicate records and separate the dataset in 2 different subsets :

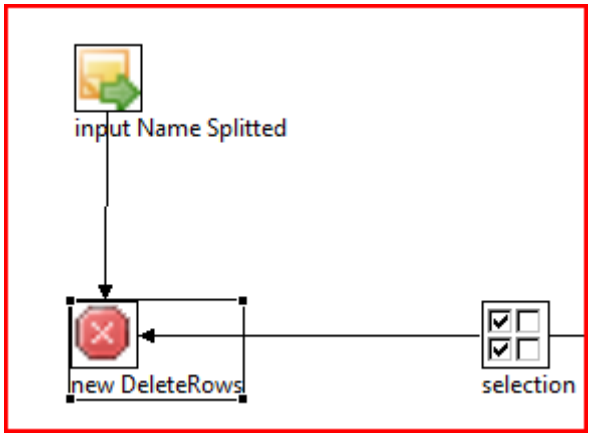
- One subset (black arrow) will contains unduplicate records, understanding that if originally a record is available 3 times before the unduplicate box, it will be available only 1 time in the black target
- One subset (red arrow) will contains all the duplicate removed records, which means that if originally a record is available 3 times before the unduplicate box, it will be available 2 times in the red target (because it has been deleted 2 times).

The number of records coming from your dataset before the unduplicate box is the sum of records from the 2 boxes.

There is only one interface to define both : box name and transformation
« Trash Output », which stands as the target that manage the duplicate records

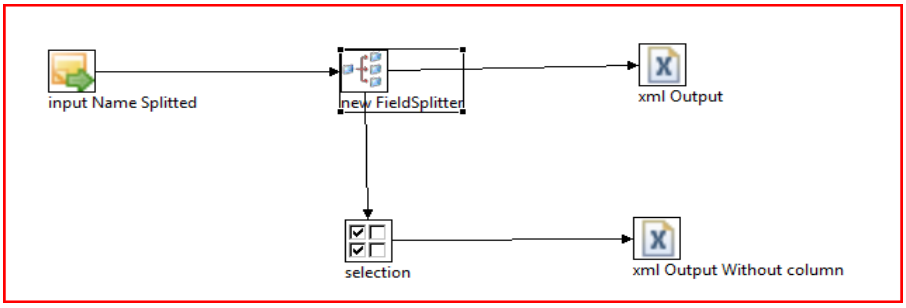


Delete Rows



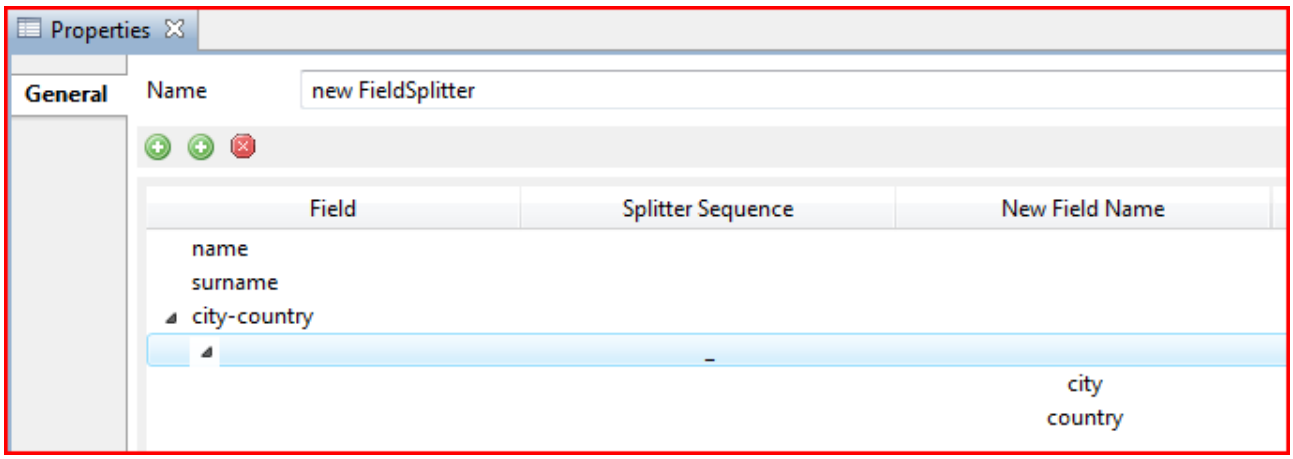
To Be Completed

Field Splitter

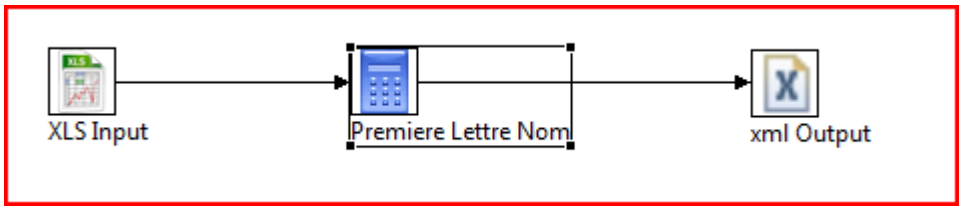


Field Splitter provide a nice and easy interface to split a fied using a separator.
By splitting a field, you will add 2 additional columns (the 2 part of your splited field) to your recordset.

In the example bellow, the original column 'city-country' will be splitted into 2 new column using the separator '-'

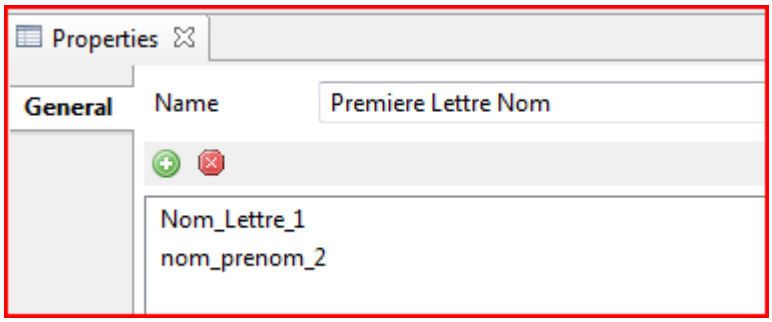


Calculation



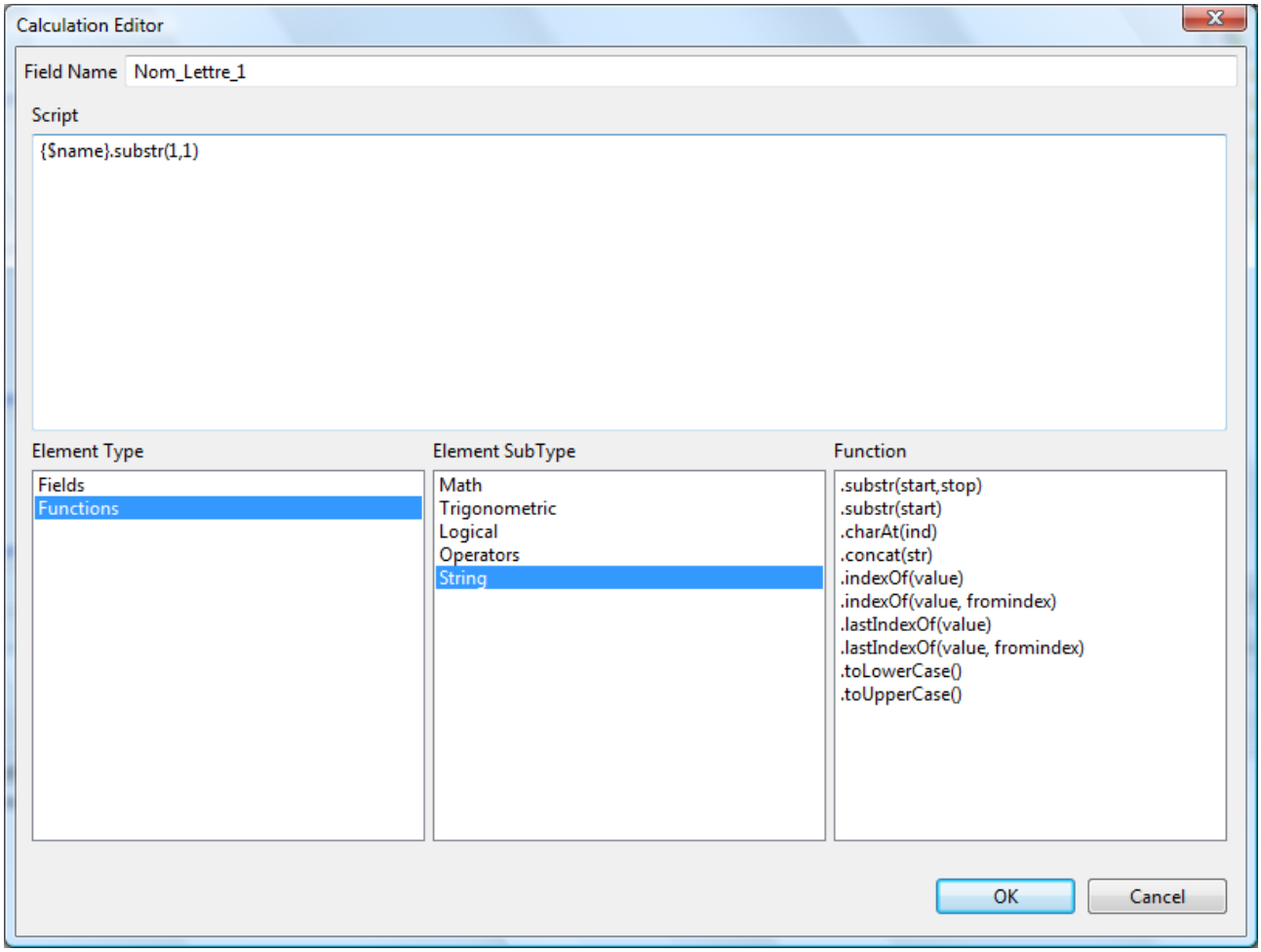
Calculation provide an easy interface to calculate new fieds using original columns, additional formula (*, - , / …), functions (math, string, date) and parameters (number, string)

In the calculator box, you can define many new columns and then access the formula definition

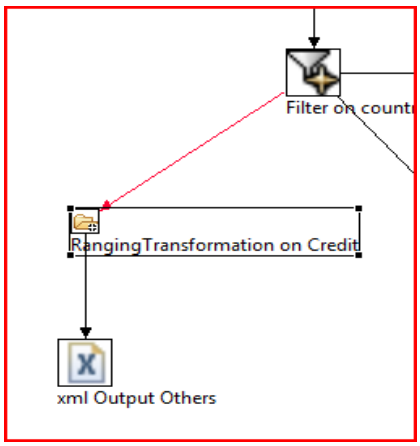


Calcul definition interface

In this interface, Birt-alike, you will be abble to use many functions to create complex functions.

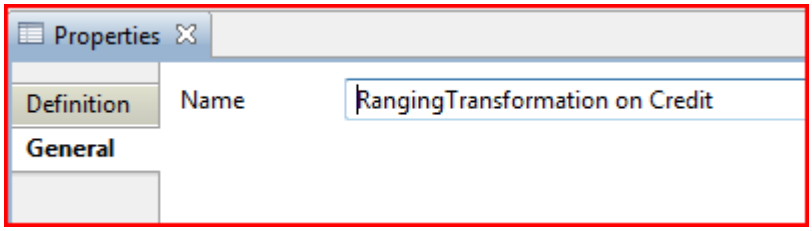


Ranging



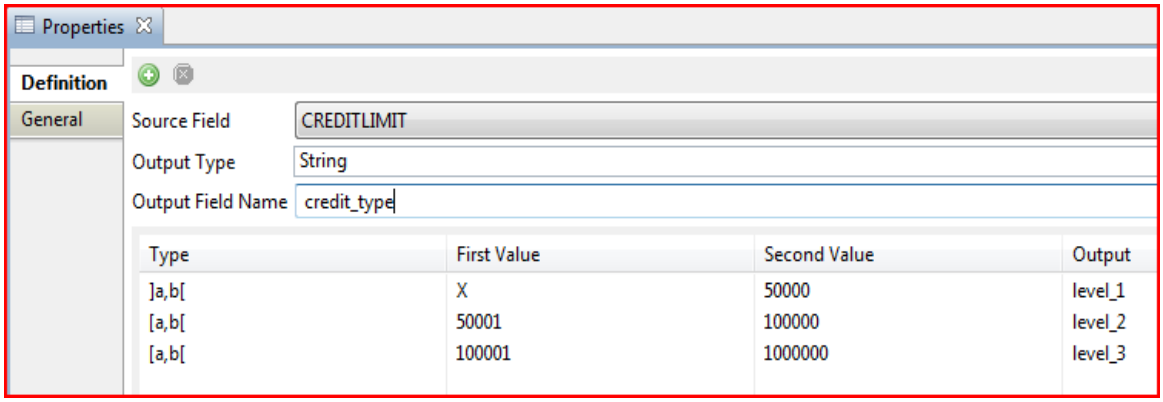
Ranging is a box where you can transform a numeric column, using its value and get a string as a result.

General



Definition

In the example bellow, you create a new column in your dataset, credit-type, and content for every row will be evaluated by using the creditlimit column and its range of values



Sort

To Be Completed

Select Distinct

To Be Completed

Annexes

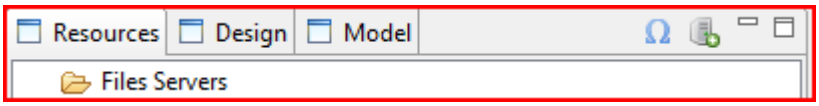
Variables management

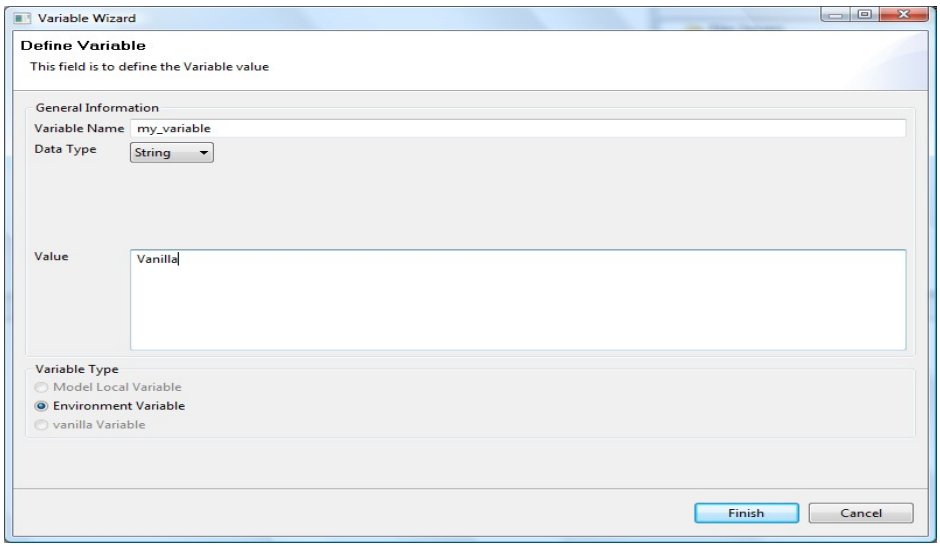
BI GateWay provides a clean and easy way to define and use variables.

Variables are stored in file BiGateway\resources\resources.xml

```
    <variable>
      <name>BIGATEWAY_HOME</name>
      <value>/C:/Vanilla/Serie132/designer/BiGateway/</value>
      <dataType>0</dataType>
      <scope>0</scope>
    </variable>
    <variable>
      <name>test_sde</name>
      <value>C:\Temp\SDE\datas\</value>
      <dataType>0</dataType>
      <scope>0</scope>
    </variable>
  </sharedResources>
```

Variables are defined either directly using a text editor in this ressource file, or using the create variable option





Important : variable **$ENV_BIGATEWAY_HOME** is not editable, and references the Bi GateWay home, used in all the samples to reference the location of files

Roadmap

Information in this chapter are subject to change without prior notification, based on priority & workload

April

Multi threads issues
Mechanism to manage input parameters
Integration with Vanilla Server
Developement of features : Undo/Redo, CheckIn/CheckOut, Deployment package, command line support to start/stop datastreams

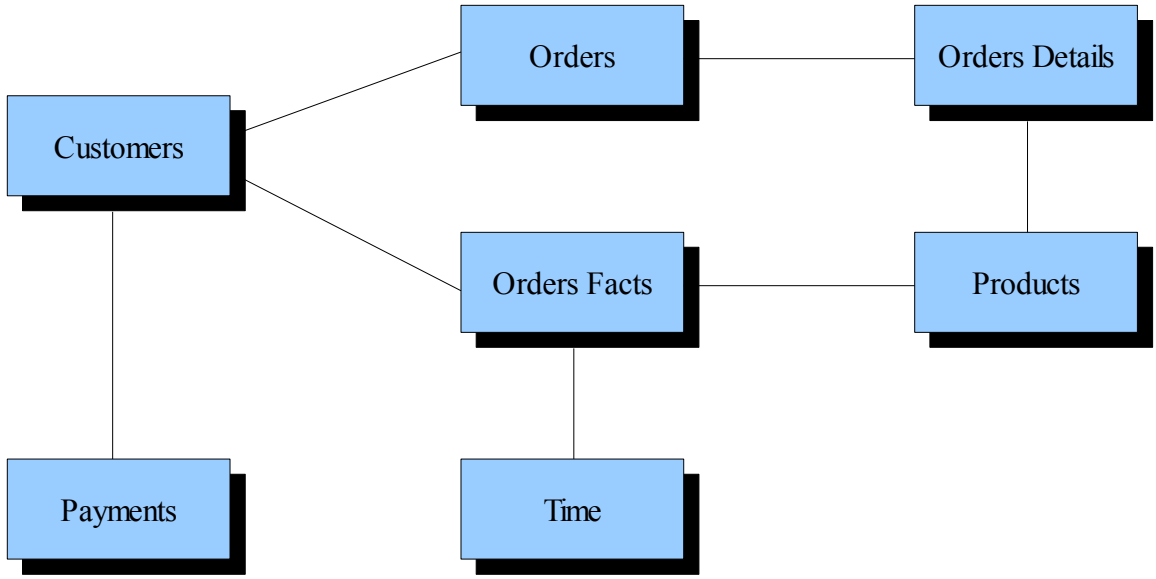
May – June

- Corrections & enhancements based on user feedback
- New transformations : Slow Changing Dimension, other transformations to be defined
- Various Wizard : for database migration, for some transformations
- Integration with Vanilla/Orbeon to create advanced user interfaces
- Vanilla Scheduler integration (to be finished)

Summer 09

- Integration with FreeAnalysis Schema Designer (Aggregate table)
- PlugIn fort BIRT : BI GateWay as datasource for reporting
- Datastream remote support (Web Services Call)
- Integration ad Datasource for FreeMetadata documents
- SDK to develop, publish and run Datastream using java applications

Database schema for sampledata



Browsing the samples

In all the samples bellow, we will use Excel files as input files and XML files as Output files, in order to facilitate the use of BI GateWay outside of any database.

Only last samples will contain MySQL connection (Input and Output), in order to show how to setup a connection and how to create target tables and load data into databases

All the transformations are located in the /transformation folders

Excel Load & XML Conversion

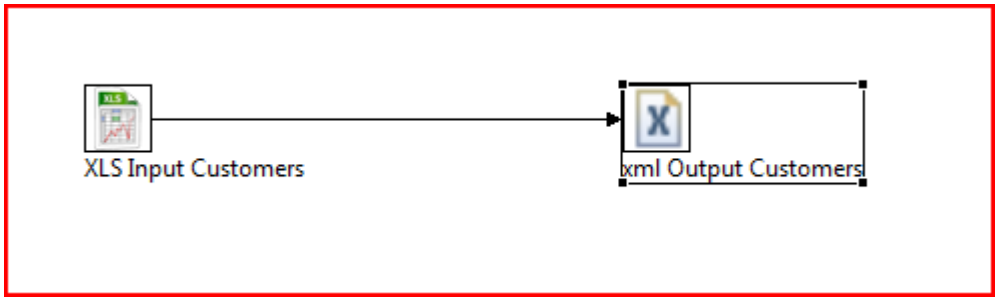
Reference : 1_datatype_conversion.gateway

This example demonstrate how to connect to an Excel file and convert this Excel file into an XML document

Interest of the example :

- show how to connect to an Excel files
- show how to create an XML document
- no transformation

Global datastream outlook



Simple Filter

Reference : 2_filter_on_customer_country.gateway

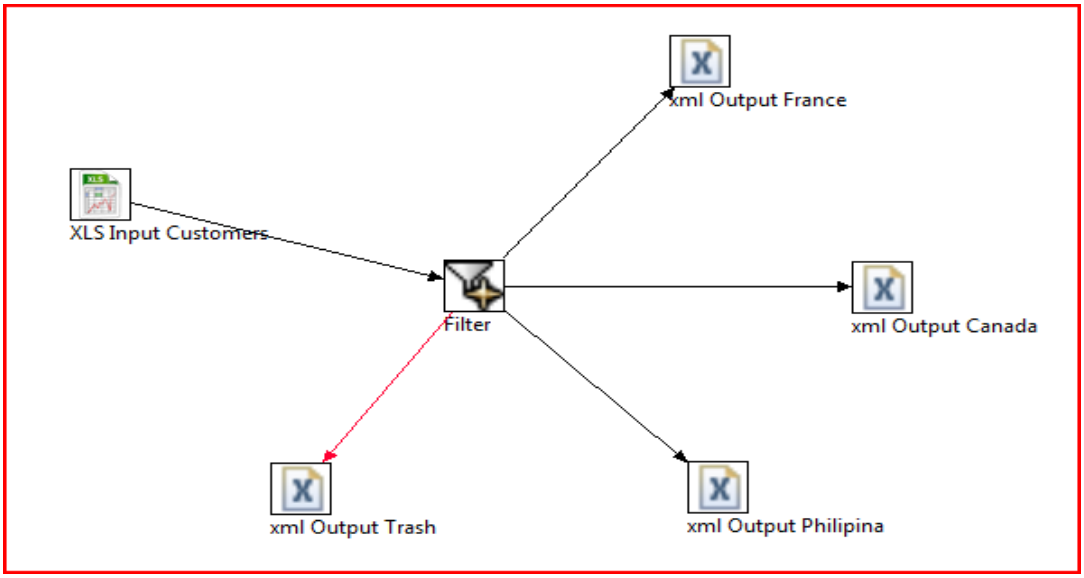
This example contains simple filtre instructions to separate records based on country code

Using values for Canada, France, Philippines and « rest of the world », it creates 4 XML documents

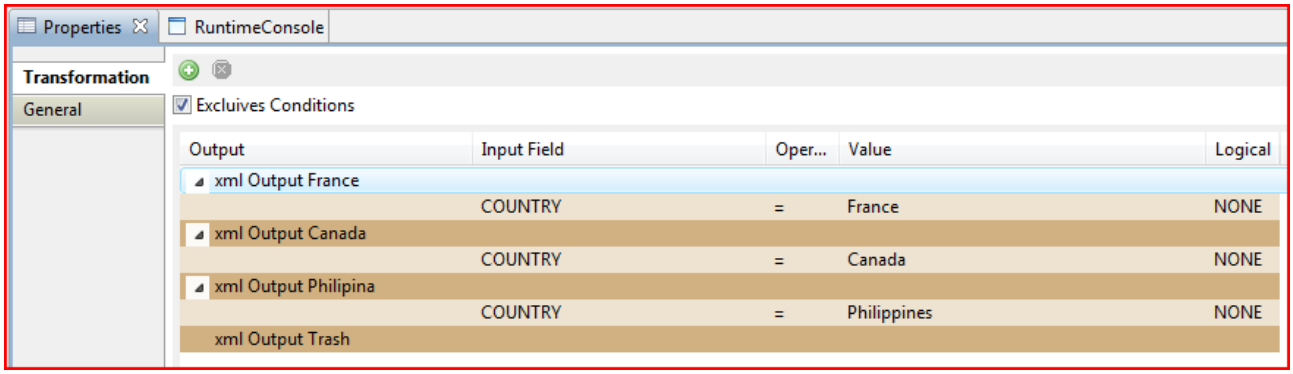
Keep attention on the « exclusive » check box, that indicate that each row has only one and one taregt

Keep attention on « trash », as defined in the filter box. « Trash » indicates which dataset will received the record that don't match any defined conditions. On the workplan, « trash » is in red.

Global datastream outlook



Filter definition



Simple Selection

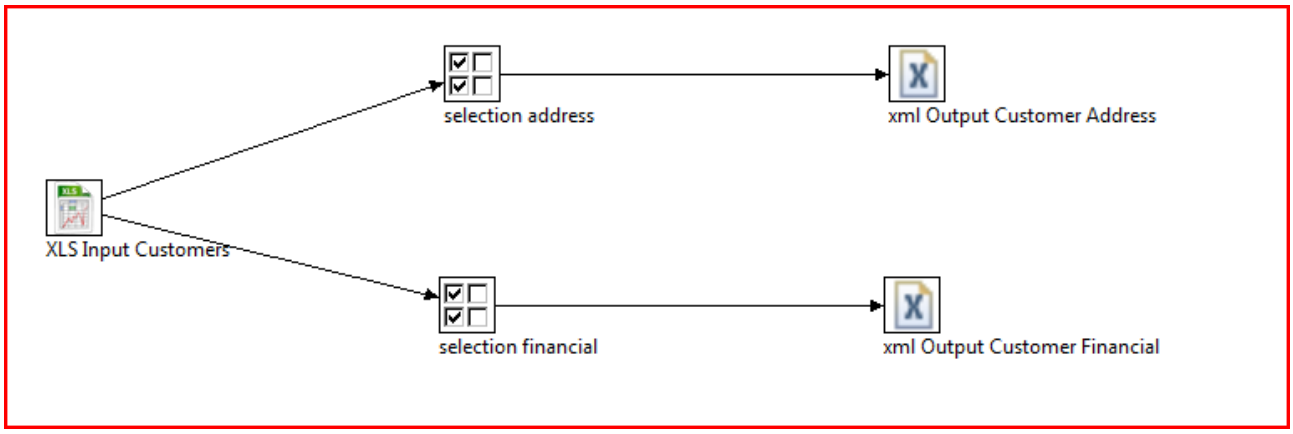
Reference : 3_simple_selection.gateway

This example shows how a selection can be used with an input file to create multiples files.

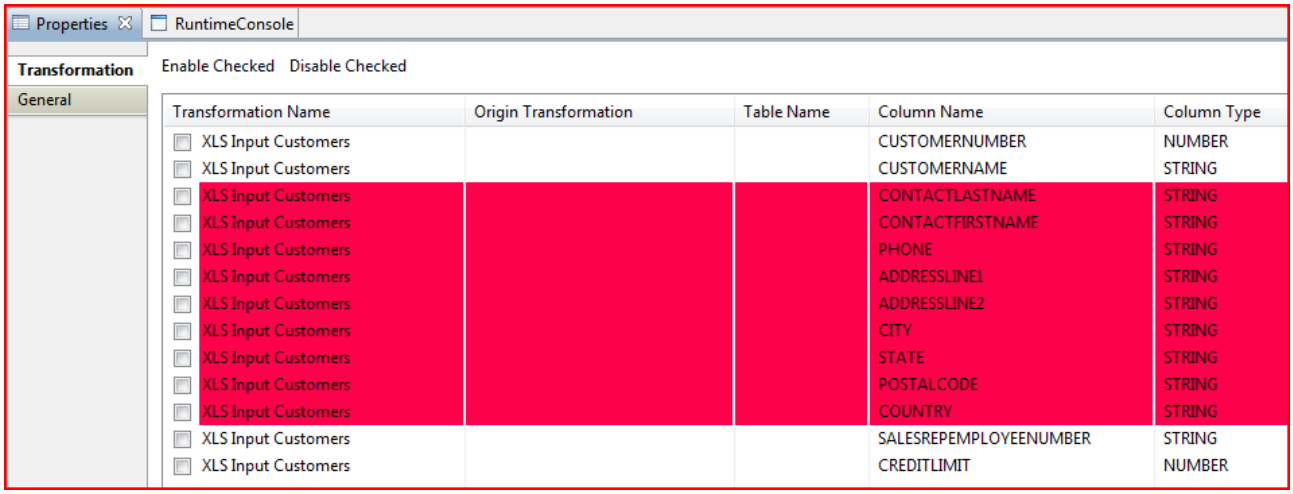
Selection can be used to remove columns from a dataset (to remove unexpected columns from a lookup, for example)

It this example, we create 2 consistents XML files that repeats columns for customers, with one file that contains address, and the other one that contains financial informations

Global datastream outlook



Selection for the Financial Stream



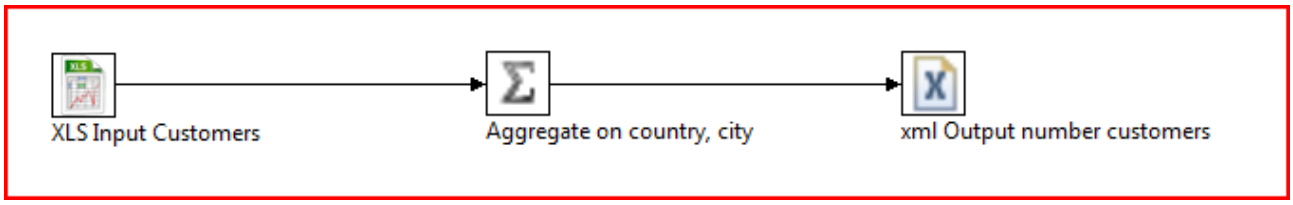
Simple Aggregate

Reference : 4_simple_aggregate.gateway

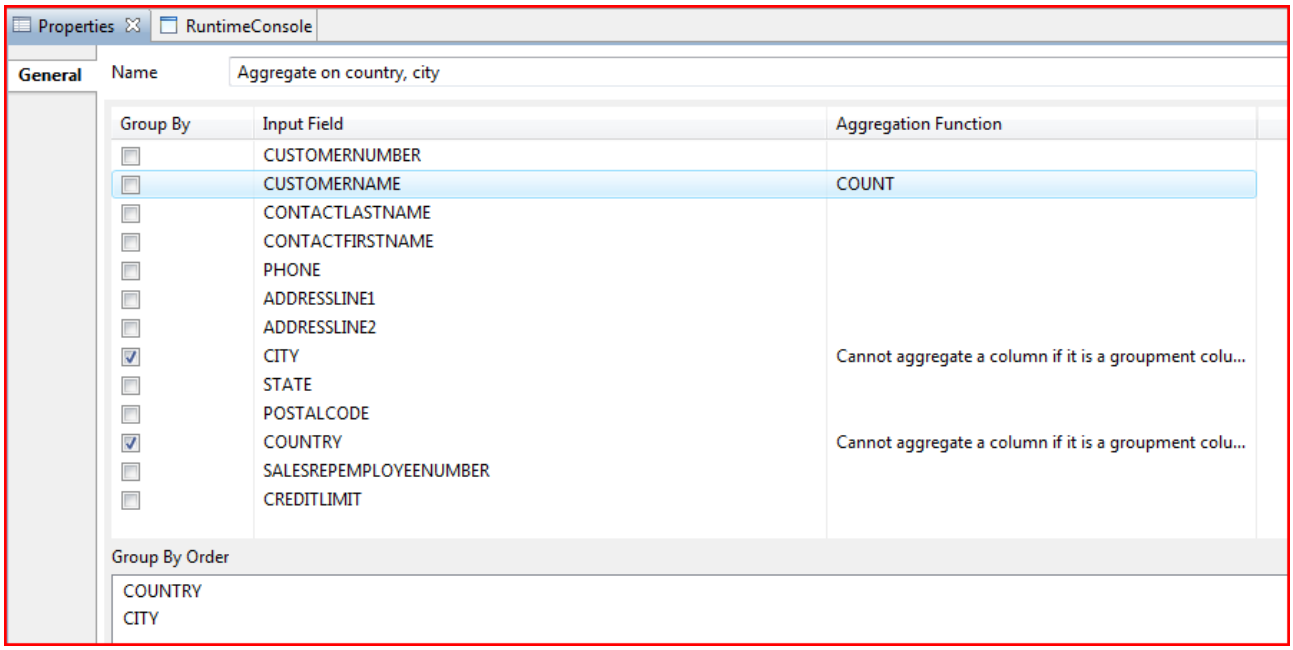
This example shows how to create an aggregate using columns from a dataset

It this example, we create an aggregate on country + city that count the number of customer in each city and stores the results in an XML document.

Global datastream outlook



Aggregate definition



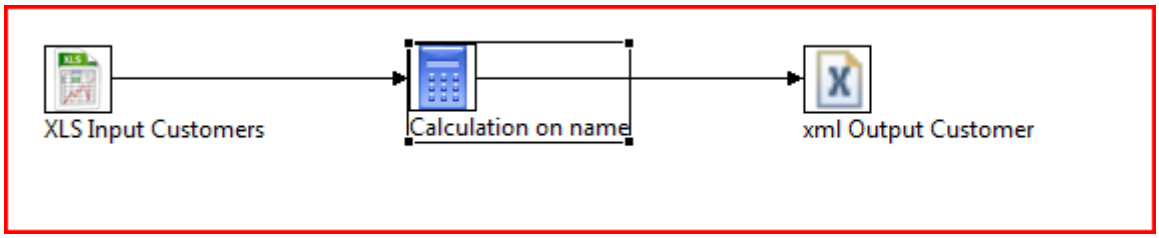
Simple Calculation

Reference : 5_simple_calculation.gateway

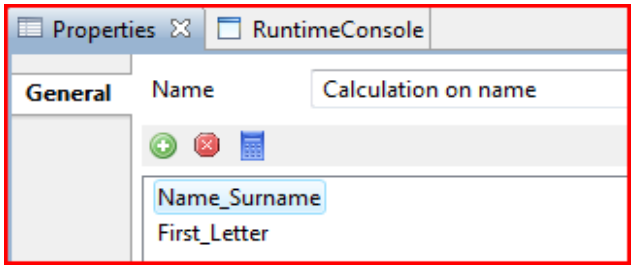
This example shows how to create a calcul using columns from a dataset

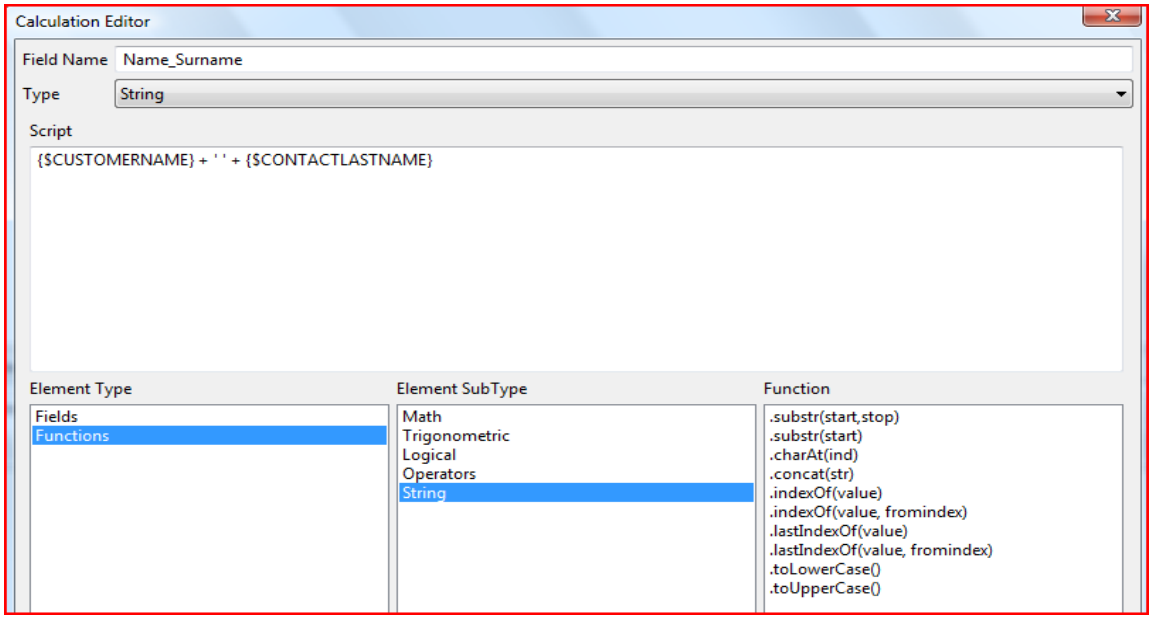
It this example, we create 2 calculated columns « first letter of name », « name + surname » and stores the results in an XML document.

Global datastream outlook



Detail for new calculated column





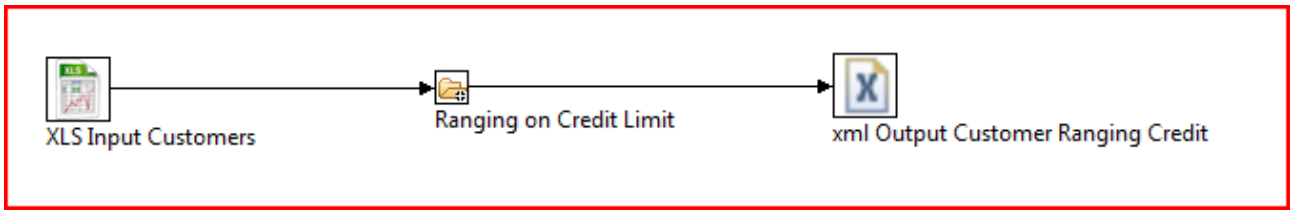
Ranging on columns

Reference : 6_simple_ranging.gateway

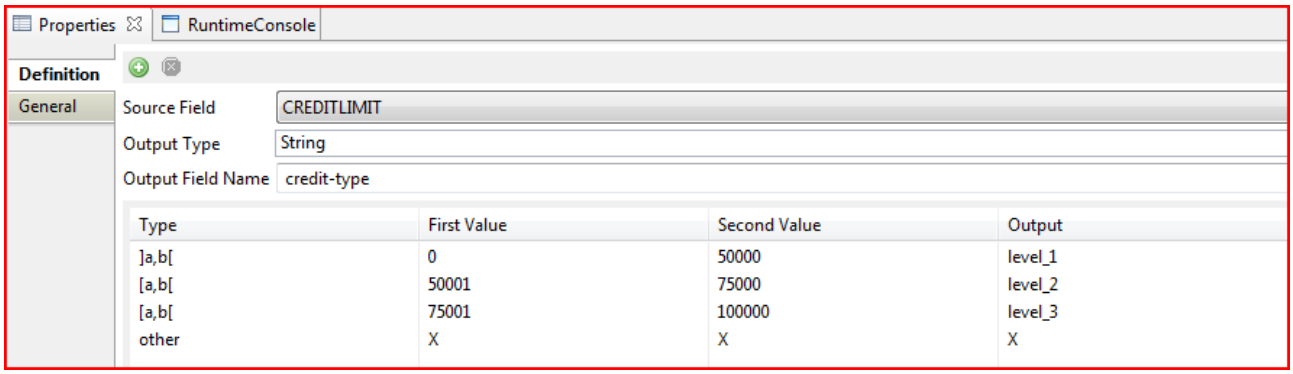
This example shows how to create a calcul using another column value

It this example, we create a new column « credit-type», based on the value (range) of the column « CREDITLIMIT » and stores the results in an XML document.

Global datastream outlook



Detail for ranging column



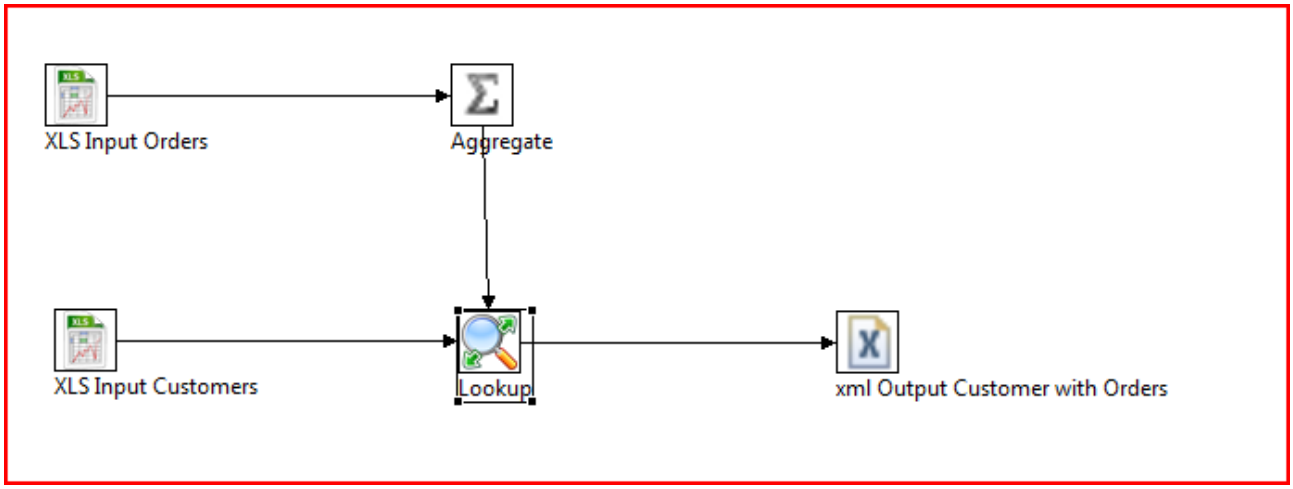
Lookup

Reference : 7_lookup_customer_number_orders.gateway

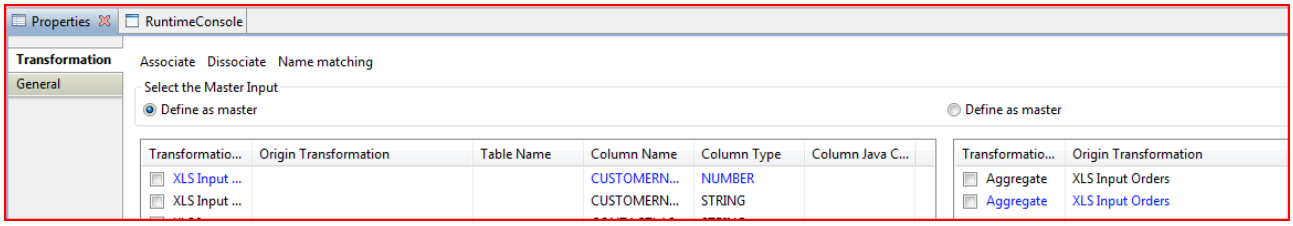
This example shows how to create a lookup to get additional data in your recordset

It this example, we first create an aggregate in the order table to count the number of orders per customers, then return to the main recordset to add those aggregate columns and stores the results in an XML document.

Global datastream outlook



Detail for Lookup definition



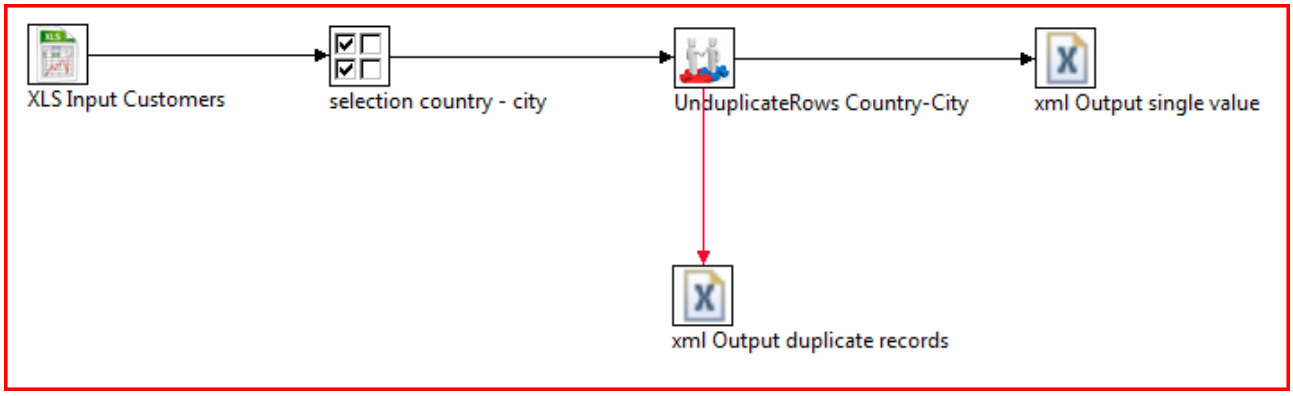
Unduplicate

Reference : 8_unduplicate_records.gateway

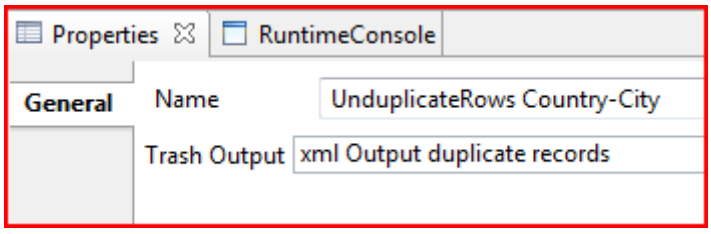
This example shows how to manage duplicate records from a dataset

It this example, we first create an select on countr and city, and then apply an unduplicate action to create 2 recordset : one with single value, and one with all the duplicated-removed records. The results are stored in 2 XML documents.

Global datastream outlook



detail for unduplication definition



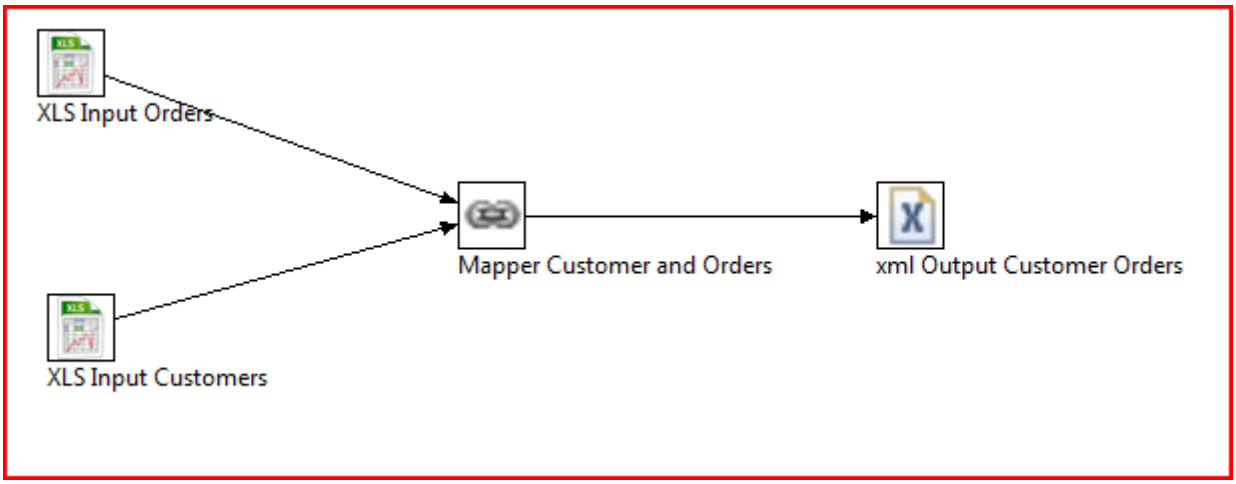
Mapper on dataset

Reference : 9_simple_mapper.gateway

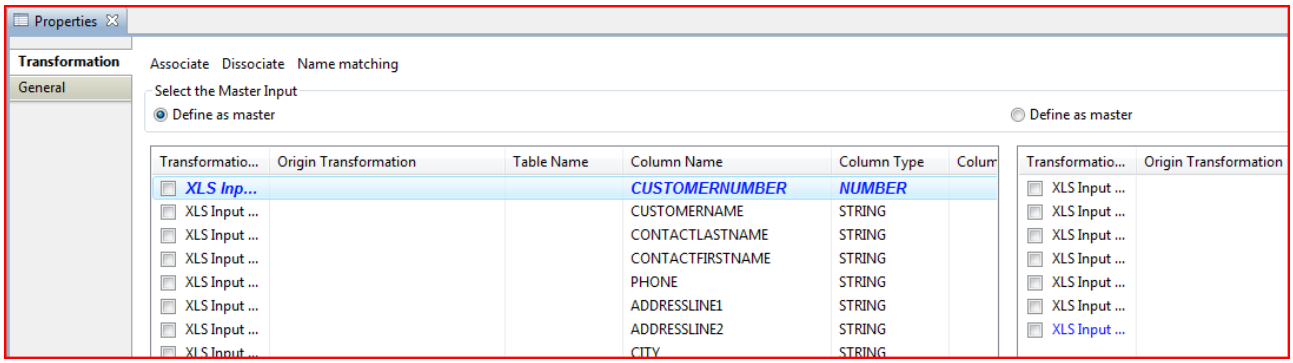
This example shows how to create join between 2 datasets

In this example, we create a join between 2 dataset : customer and order, based on a join on customer number, to create a larger dataset. The results are stored in 2 XML documents.

Global datastream outlook



detail for mapping definition



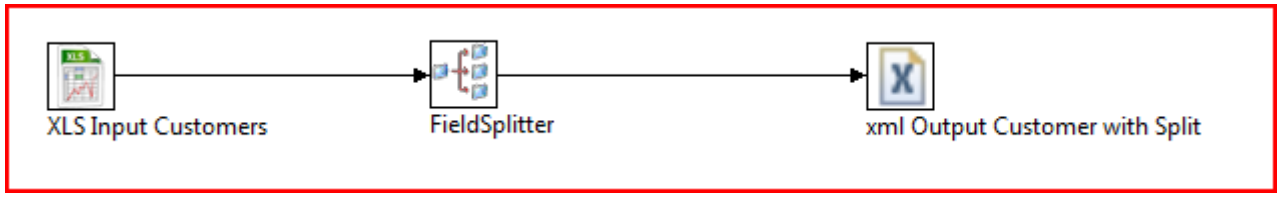
Splitting Field

Reference : 10_field_splitter.gateway

This example shows how to create new columns using a column content and a splitter sequence

In this example, we create 2 new columns for customer name that contains « & », with result stores in an XML documents.

Global datastream outlook



Merging Dataset

Reference : 11_Dataset_Merge.gateway

This example shows how to merge 2 datasets using the merge dataset box

In this example, we separate the customers into 3 groups : France, Canada and Others, and we merge France and Canada, with result stores in an XML documents.

It's an evidence that this example has no sense, since it's possible to build a filter that directly extract customer from France and Canada and send those records in a specific output

MergeStreams box assumes the different input dataset have the same structure

Important : MergeStreams box can have more than 2 datasets as input, there is no limitation on this subject

Global datastream outlook