DATA SOLUTIONS DE SPIRLET

DV Generator (3.0) for

SAP Sybase PowerDesigner 16 SP6

A DATA SOLUTIONS DE SPIRLET White Paper

Revised April 2019

All information also available on http://www.data-solutions-de-spirlet.com

DATA SOLUTIONS DE SPIRLET

Executive Summary

In September 2013, we release our first version of “EDW DV Generator for SAP Sybase PowerDesigner”.

Since that moment Data Vault approach evolved and, now, ware are proud to announce our new version of our tool under the name “DV Generator 3.0” which will permit to deliver database models which can be deployed immediately on premise or in the cloud.

This white paper goes a bit deeper and describes the capabilities of the latest edition of “DV Generator Extension for SAP Sybase PowerDesigner 16 SP6”:

- We introduced tables which can be not processed by Data Vault such as:
  - Referential tables
  - Referential tables with historization
  - Tables not to be processed
- Capability to specify
  - your hash function
  - generate, or not, hash function in mappings or data movement models
  - add, or not, source record identification in hubs, links or satellites
  - add, or not, delete flag in hubs, links or satellites
  - add, or not, load end date in hubs, links or satellites
  - add, or not, validity period in satellites
  - Specific color in order to understand quickly a model
  - Database types of technical fields
  - default values of technical fields
  - Specific prefix or suffix
- By using hash technique, capacity to make massive data load parallelization
- Two version of our tool:
  - Command line
  - Power Designer Add-in (development still ongoing, Available soon)
- Available editions:
  - Generator of physical data models, staging and data vault, that are ready to be deployed on premise or in the cloud without modification
  - Optional data movement models providing the foundation of ETL, ready to be deployed on SAP Data Services or any ETL tool if you have the appropriate extended model
  - No more Orchestration model is needed, everything can be parallelized

Data Solutions de Spirlet

DATA SOLUTIONS DE SPIRLET

Why such solution

Initial idea of this product was suggested by several usual questions when building data warehouse:

- How to elaborate data vault data model from classical model using automation
- How to generate all steps (staging and final data vault model) considering classical technical aspects without drowning end user
- How to generate full traceability (mapping) between all models to keep full impact and linage powerdesigner's capabilities
- How to pre-optimize data model considering target physical database type (option)
- How to generate state of the art data vault model that fit your business
- How to generate fast prototyping or final models **in few minutes** while manual work takes weeks or months
- How to generate fast coding while manual work takes weeks or months

Our solution

« DV Generator 3.0 » is our solution permitting generating a Data Vault data model from original physical data model:

- Several PowerDesigner extension models to support analysis and automation and delivery
- Automatic model generation
    - Complete data models with linage from source to data vault model, ideal for impact and linage analysis which is usually difficult to maintain.
    - Data movement models showing all components and how to sequence data load (edition with DMM generation)
    - customization of data types and default values used for generated columns.

If you have a doubt on our solution capabilities, contact us at sales@dsds-it.net and we will provide you complete example.

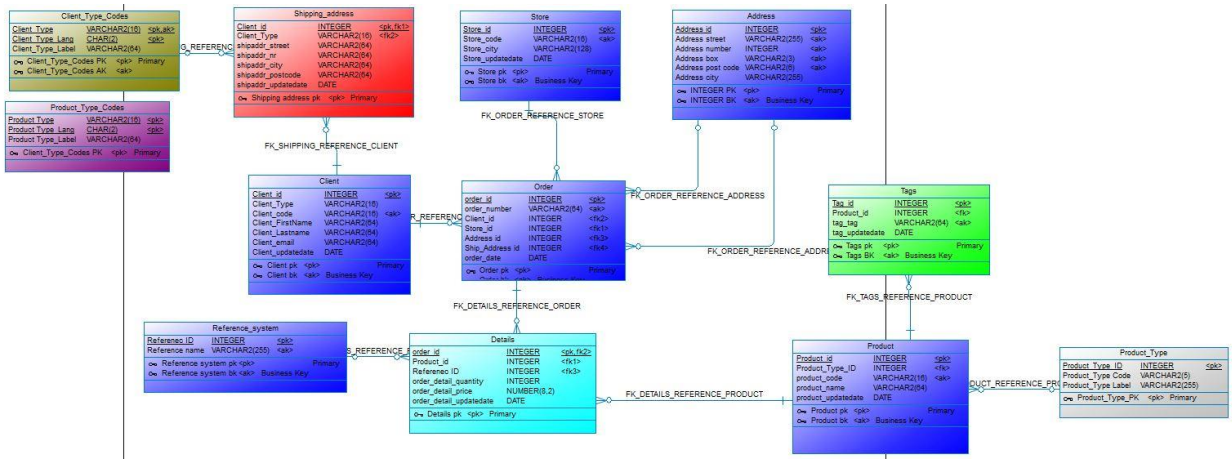
Data Solutions de Spirlet

DATA SOLUTIONS DE SPIRLET

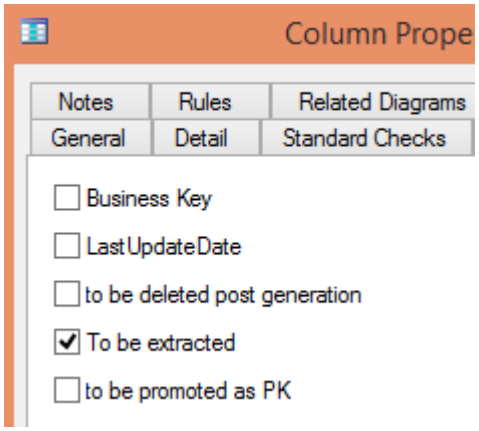
Quick description of our approach

1. Source analysis:

   As usual in Business Intelligence area, all sources must be well understood from Business and Technical point of view. First choose a source model which is usually representative of an operational model (for example classical ERP/CRM or extractor model):



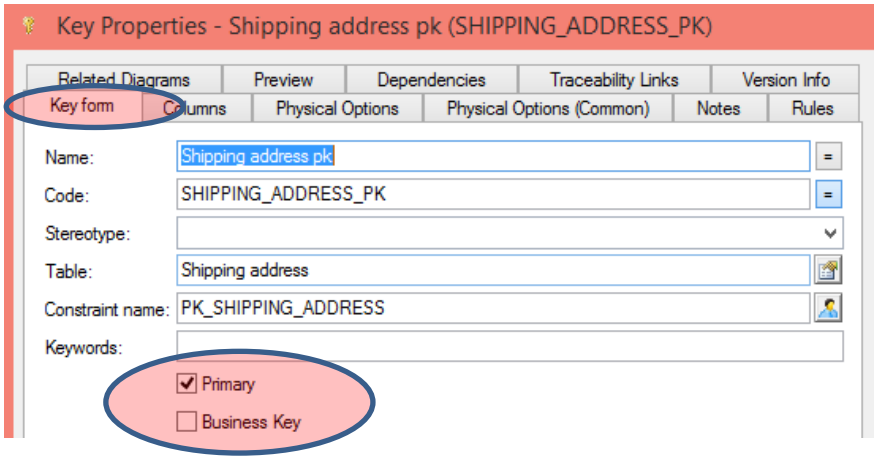
   We provide extended definition to permit identification of Business keys which will be used during automation:



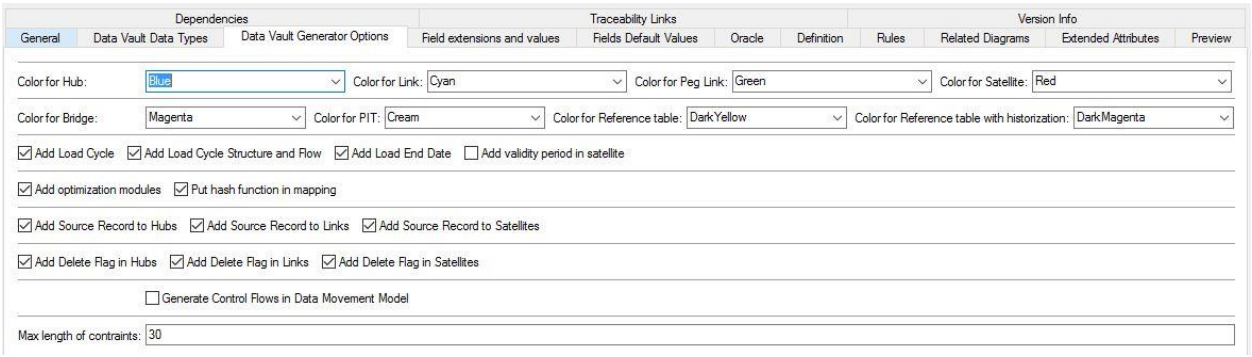
Data Solutions de Spirlet

2. Complex (multi-fields) primary and business keys are identified:

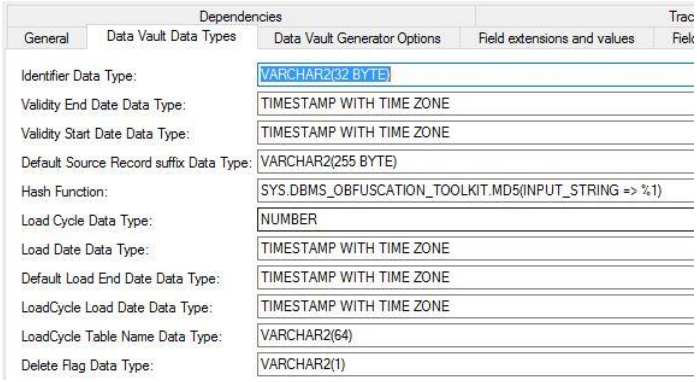
   business keys can be very complex even it is not often the case:



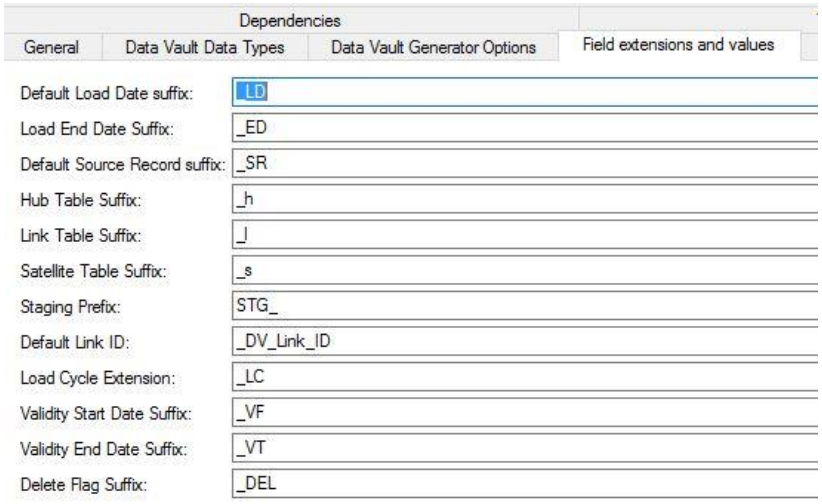
3. Many options available at model level, in multiple forms, to customize model generation:



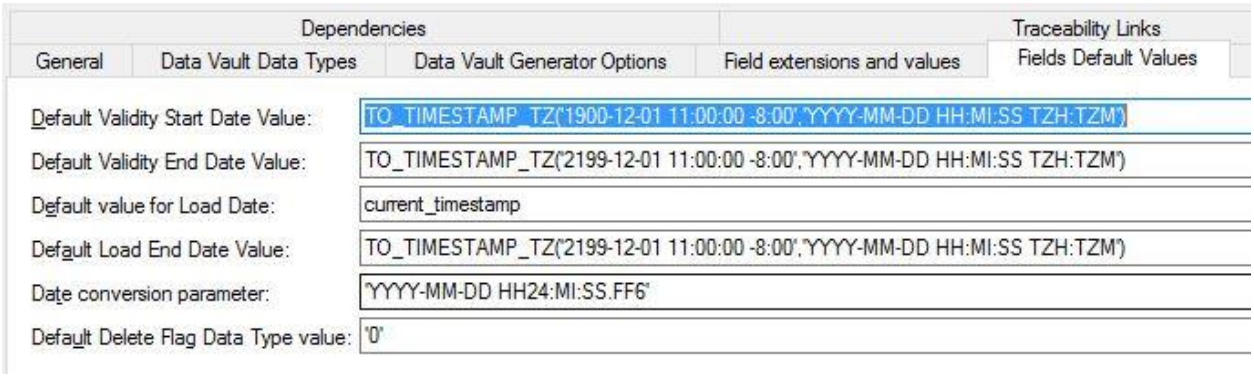
4. Capacity to specify data types for technical fields with default values and hash function to be used:



5. Capacity to specify extensions:

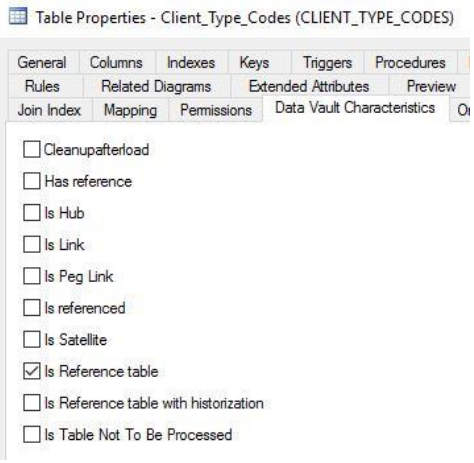


6. Some additional default values:



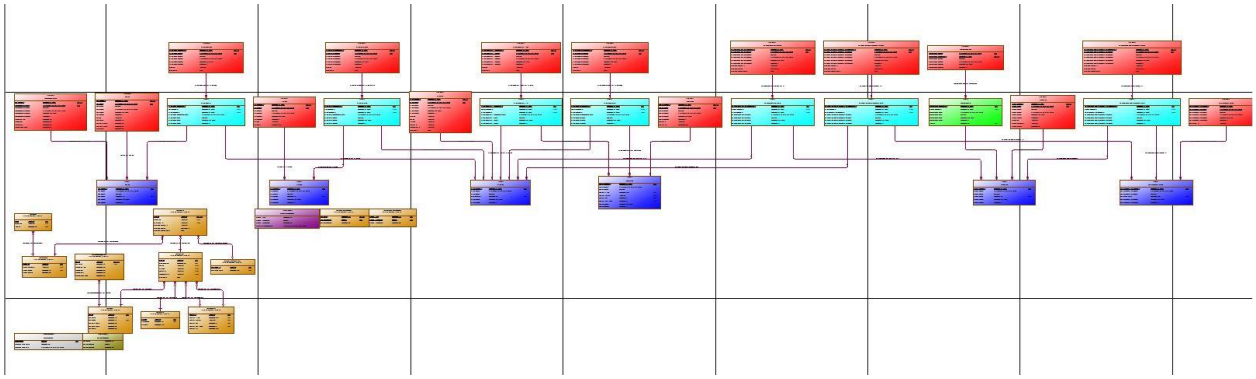
Data Solutions de Spirlet

7. Capacity to specify specific tables such as reference tables, reference tables with historization or tables which are not to be processed:



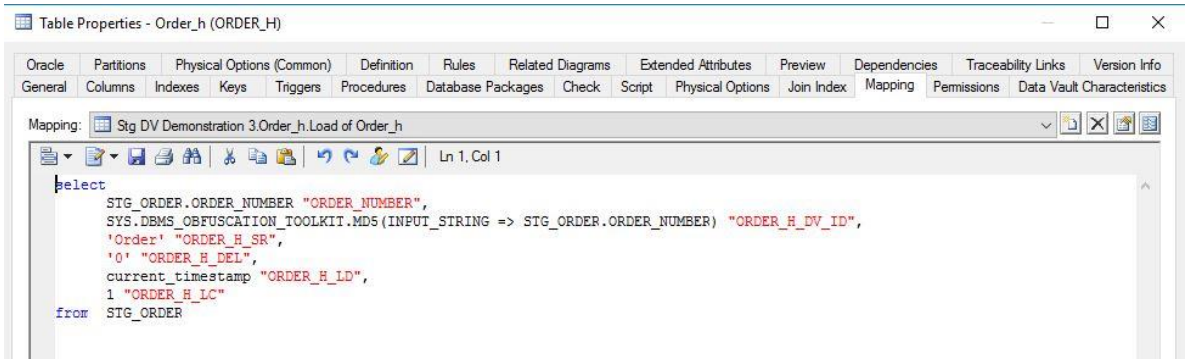
8. Target model:

   Target model will be automatically draw in a well understandable layout with color support to quickly visualize final objects (see parameters at point 3):

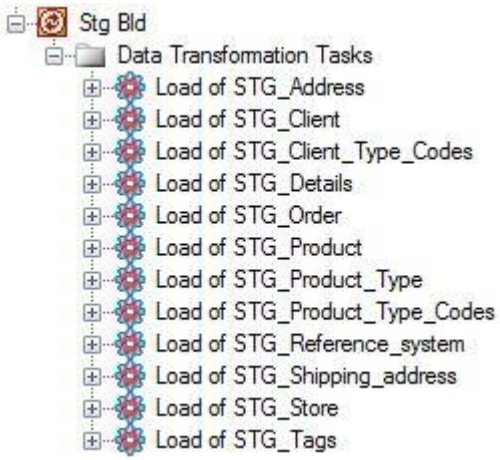


Data Solutions de Spirlet

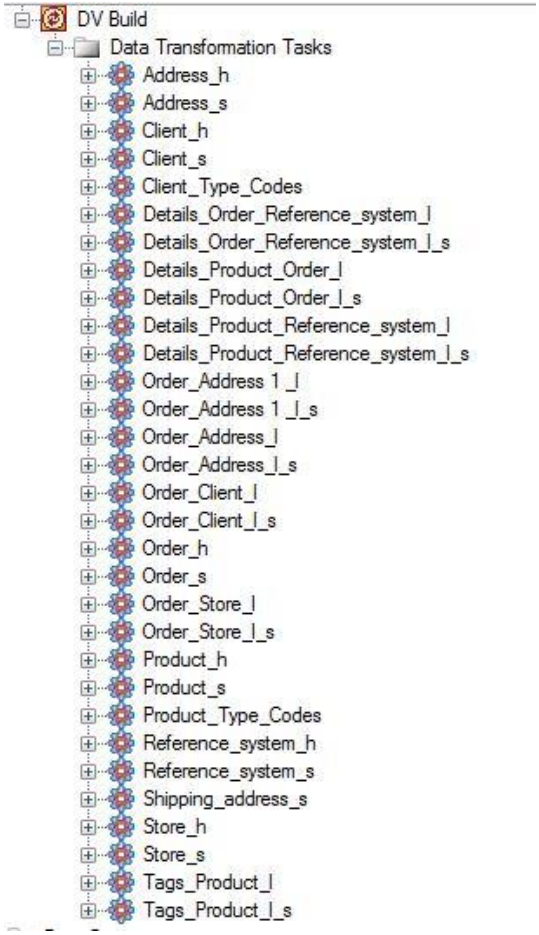
9. Mappings integrating hash technique directly in the SQL:



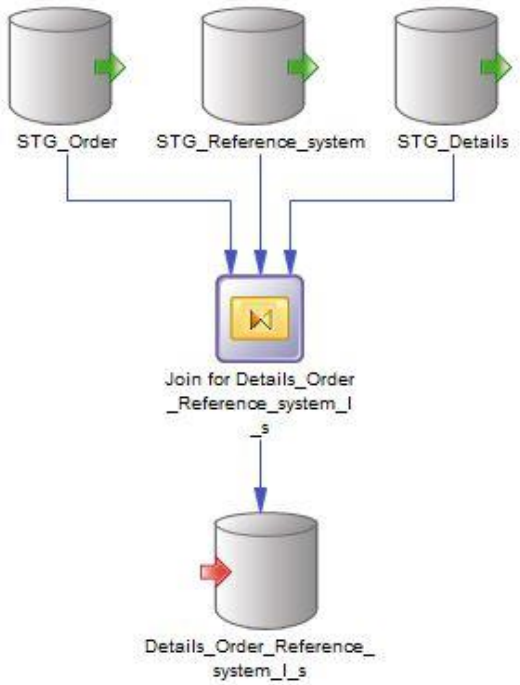
10. All data movement tasks to load the staging:



11. All data movement tasks to load data vault model including reference tables:



12. Description of simple or complex flows:



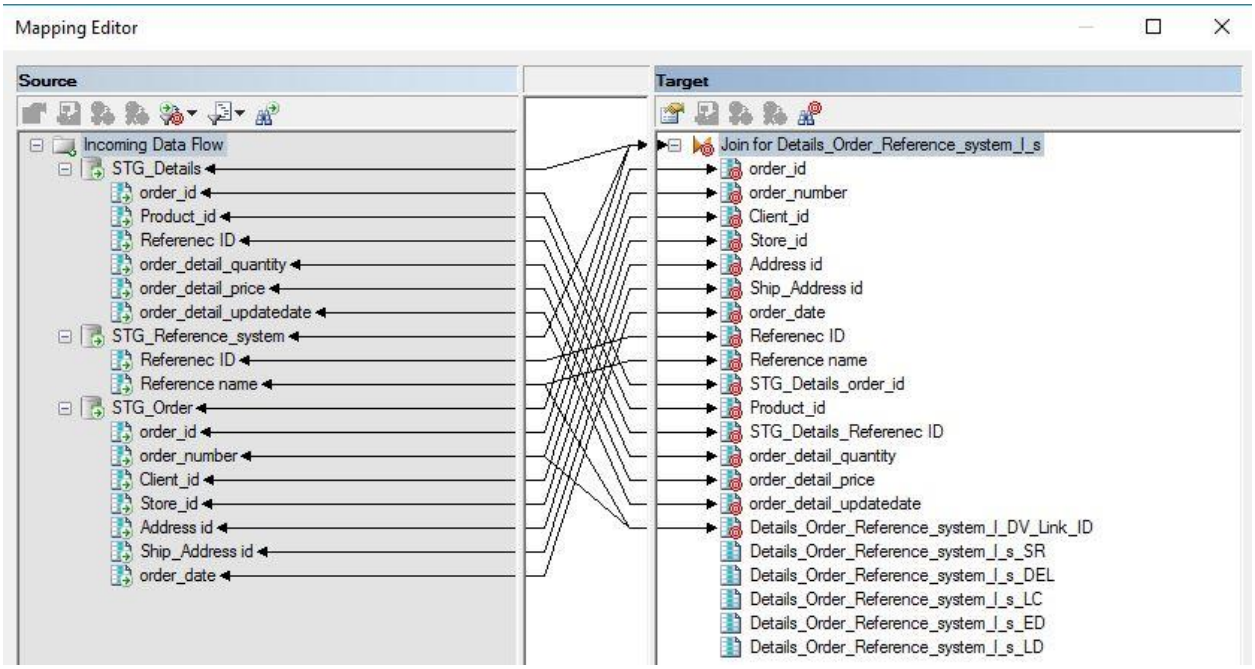
13. Integration of functions or default values directly in data transformation tasks:

Data Query Transform Properties - Join for Details_Order_Reference_system_l_s (Join_for_Details_Order_Reference_system_l_s)

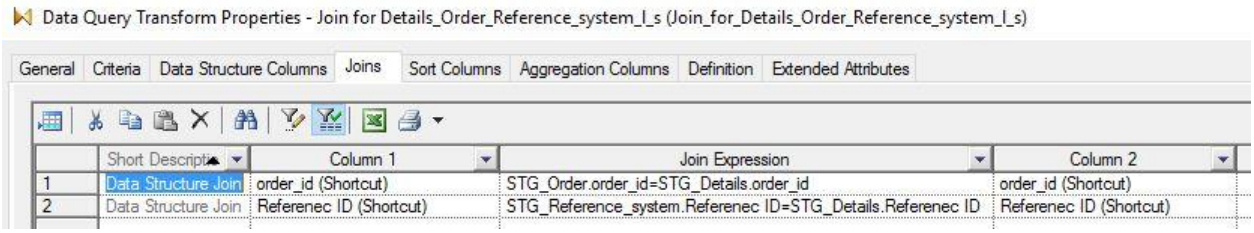
General | Criteria | Data Structure Columns | Joins | Sort Columns | Aggregation Columns | Definition | Extended Attributes

	Name	Code	Data	Source Expression
1	order_id	ORDER_ID	I	STG_Order.ORDER_ID
2	order_number	ORDER_NUMBER	VA64	STG_Order.ORDER_NUMBER
3	Client_id	CLIENT_ID	I	STG_Order.CLIENT_ID
4	Store_id	STORE_ID	I	STG_Order.STORE_ID
5	Address id	ADDRESS_ID	I	STG_Order.ADDRESS_ID
6	Ship_Address id	SHIP_ADDRESS_ID	I	STG_Order.SHIP_ADDRESS_ID
7	order_date	ORDER_DATE	DT	STG_Order.ORDER_DATE
8	Referenec ID	REFERENEC_ID	I	STG_Reference_system.REFERENEC_ID
9	Reference name	REFERENCE_NAME	VA255	STG_Reference_system.REFERENCE_NAME
10	STG_Details_order_id	STG_Details_ORDER_ID	I	STG_Details.ORDER_ID
11	Product_id	PRODUCT_ID	I	STG_Details.PRODUCT_ID
12	STG_Details_Referenec	STG_Details_REFERENEC_ID	I	STG_Details.REFERENEC_ID
13	order_detail_quantity	ORDER_DETAIL_QUANTITY	I	STG_Details.ORDER_DETAIL_QUANTITY
14	order_detail_price	ORDER_DETAIL_PRICE	N8,2	STG_Details.ORDER_DETAIL_PRICE
15	order_detail_updatedate	ORDER_DETAIL_UPDATEDATE	DT	STG_Details.ORDER_DETAIL_UPDATEDATE
16	Details_Order_Reference	DETAILS_ORDER_REFERENCE_SYSTEM	VA32	SYS.DBMS_OBFUSCATION_TOOLKIT.MD5(INPUT_STRING => STG_Order.ORDER_NUMBER+STG_Reference_system.REFERENCE_NAME)
17	Details_Order_Reference	DETAILS_ORDER_REFERENCE_SYSTE5	VA255	'Details'
18	Details_Order_Reference	DETAILS_ORDER_REFERENCE_SYSTE6	VA1	'0'
19	Details_Order_Reference	DETAILS_ORDER_REFERENCE_SYSTE3	N	1
20	Details_Order_Reference	DETAILS_ORDER_REFERENCE_SYSTE4	TS	TO_TIMESTAMP_TZ('2199-12-01 11:00:00 -8:00','YYYY-MM-DD HH:MI:SS TZH:TZM')
21	Details_Order_Reference	DETAILS_ORDER_REFERENCE_SYSTE2	TS	current_timestamp

14. Integrated mapping (note that technical fields are loaded from previous step):



15. Specification of joins to be used:



DATA SOLUTIONS DE SPIRLET

In summary

	Classical working	***Our Solution***
Time to design	Classically weeks or months	Immediate
Production of models	Manual	Fully automated (after Business Model)
Time to produce physical models	Classically weeks or months	Models are generated in **minutes**
Linage between physical models	Often inexistent or manually defined	Fully automated
Production of data movement models	Often absent or made in other tools (visio, powerpoint) …)	Fully integrated and provide information for external scheduler
Time to produce data movement model	Days or weeks or months	Produced in minutes
Linage between physical models	Often inexistent or manually defined	Fully automated
Data Service Code	Depends on the model size but can be several man months	Fully automated, solid foundation are ready after generation.

Data Solutions de Spirlet