DATA ANALYTICS
(Data Warehouse)
Pentaho Data Integration

Luca Cinelli, PhD

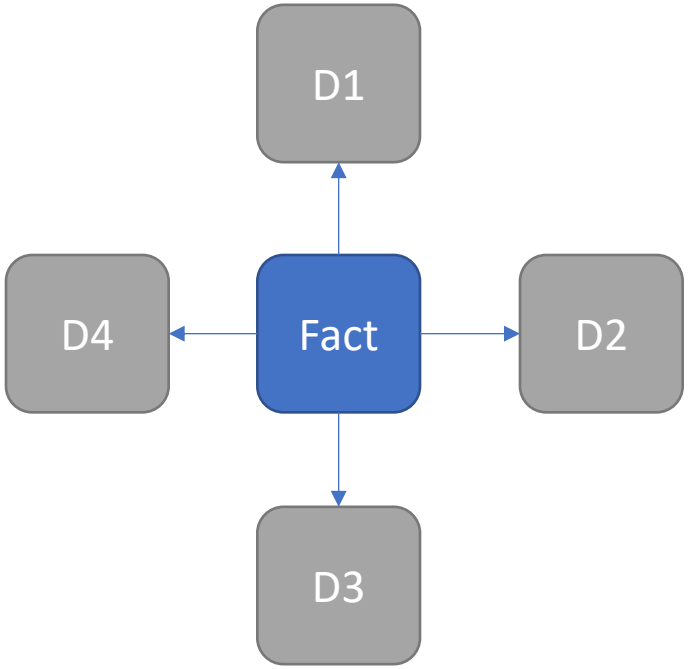
luca.cinelli@unical.it

From data to Dashboards

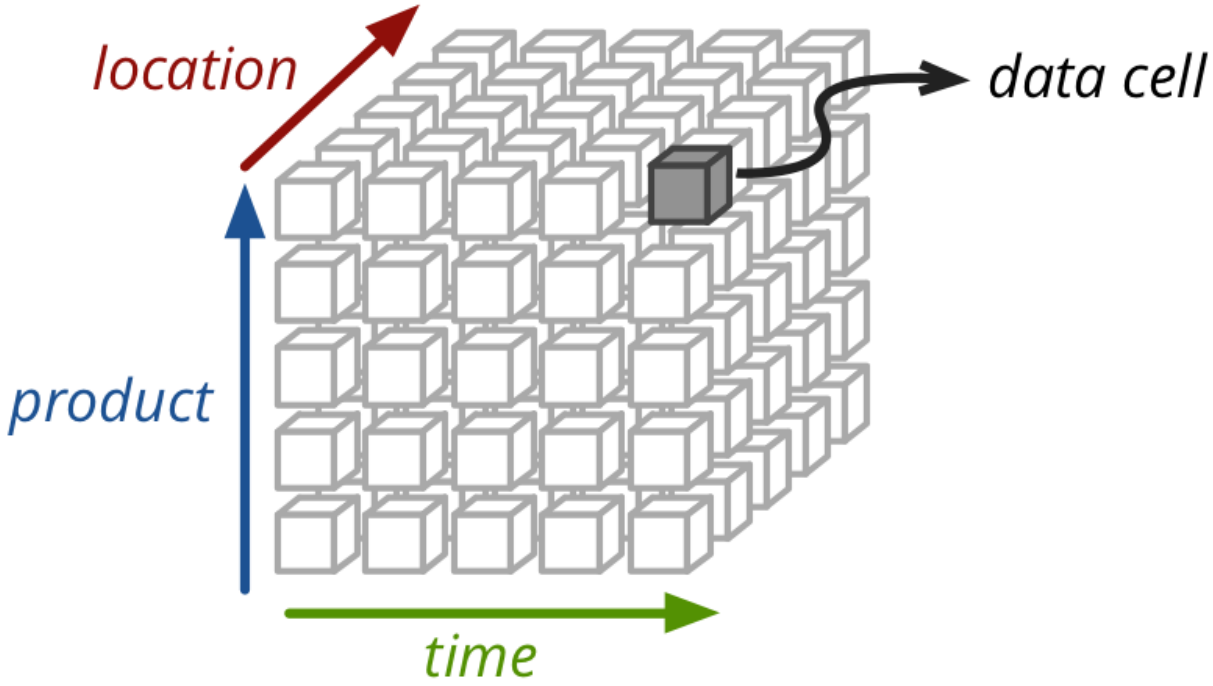
From Data to Data Warehouse

From Star Schema to Data cube

The process of creation the Star Schema is accomplished in PDI

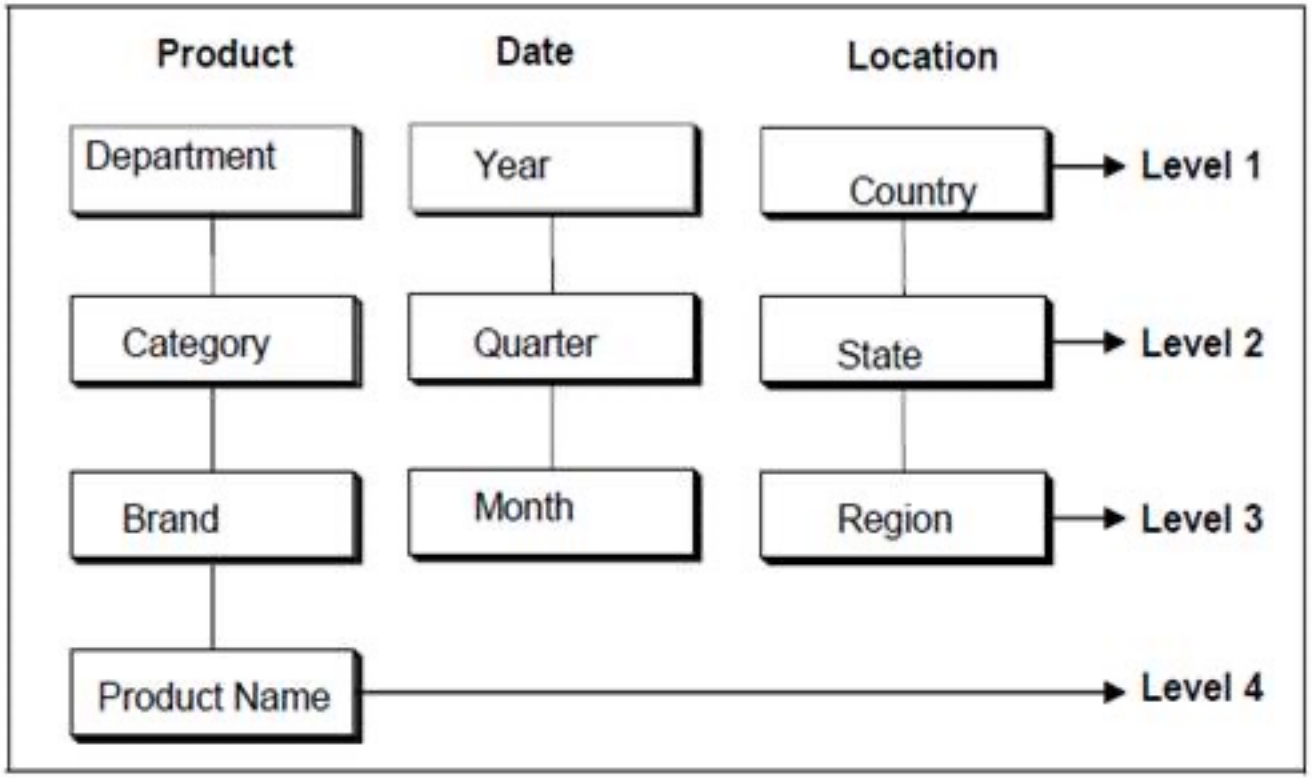


The creation of OnLine Analytical process (OLAP) Cube happens in Mondrian



OLAP Cube is a cube assembled from the Star Schema, where the dimensions are the sides and the measurements are the center.

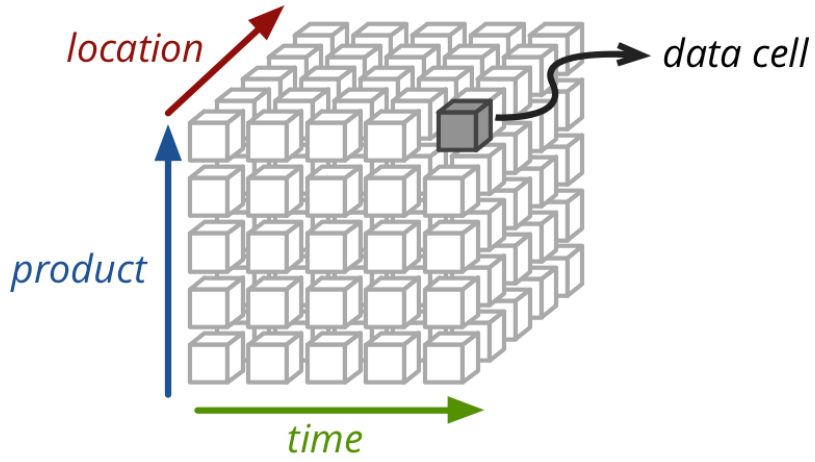
Hierarchy of Dimensions



More levels into dimensions enables us to make a deeper analysis and dashboard allowing, for example, more filters

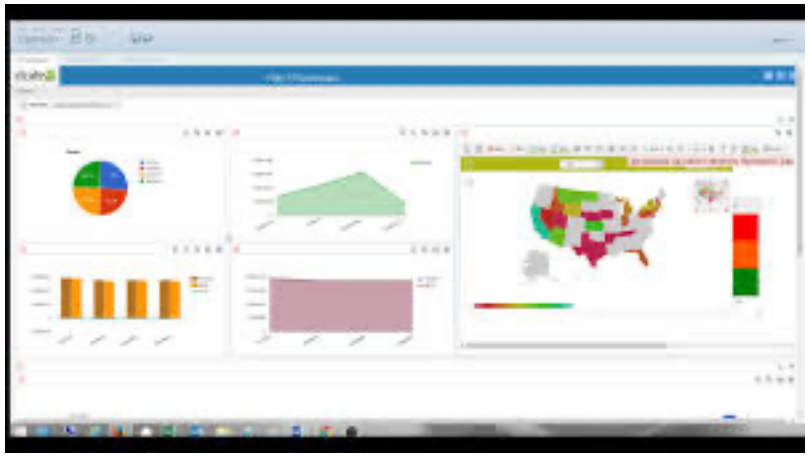
From Data cube to Dashboards

OLAP Cube



publish

Dashboard
making graphs, tables, and filters

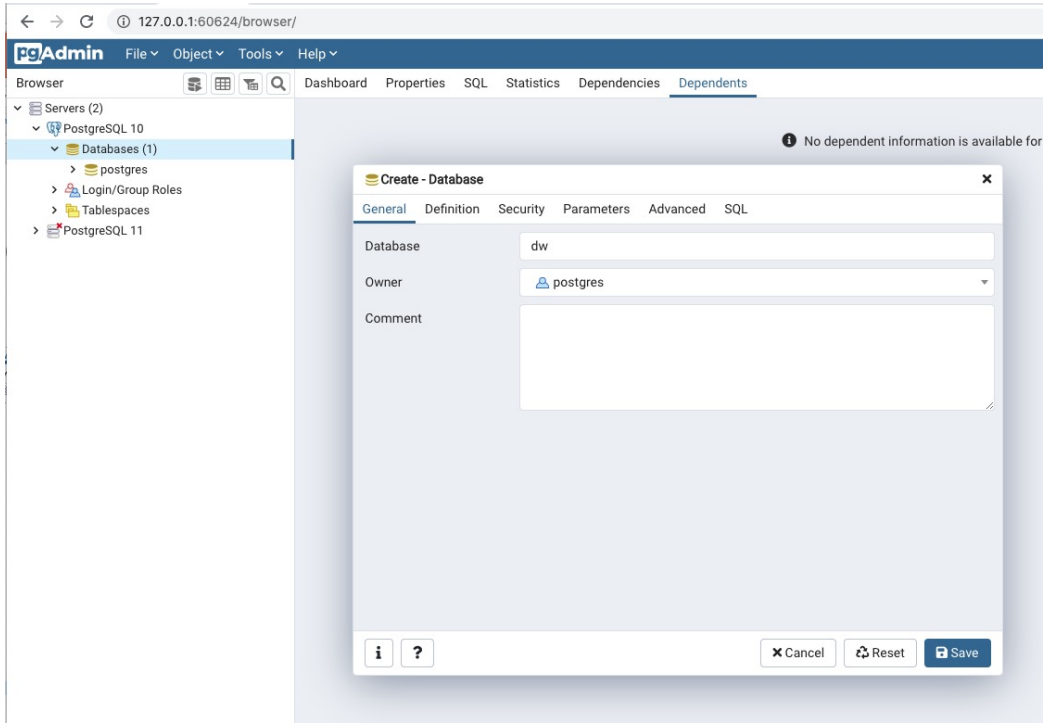


Installed ETL tool + Database

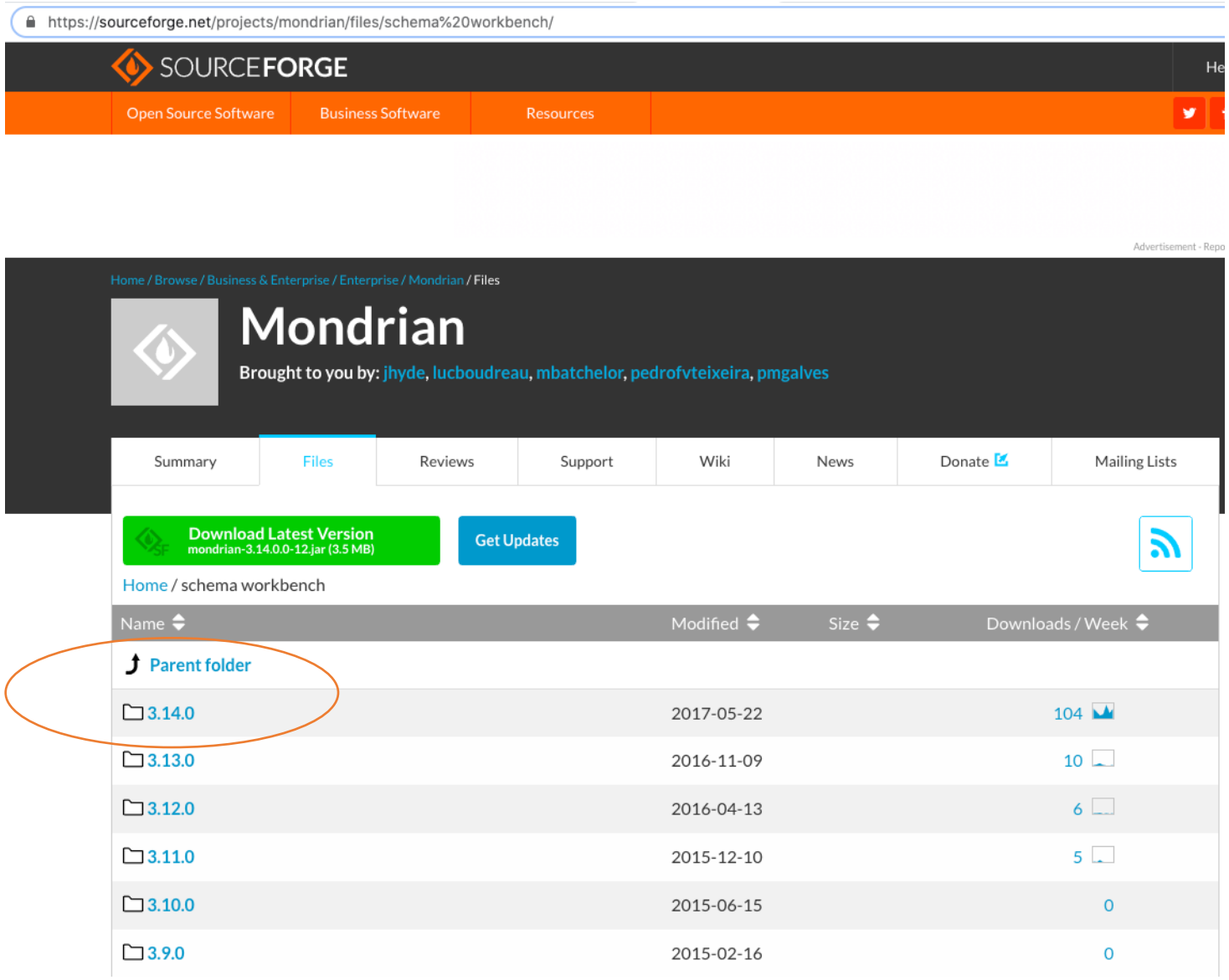
- We have already installed
  - Pentaho Data Integration to perform our ETL process
  - PostgreSQL to store our Data Warehouse and its pgAdmin graphical interface

http://127.0.0.1:60624/browser/

Create new database: dw



Download and install Mondrian



https://sourceforge.net/projects/mondrian/files/schema%20workbench/

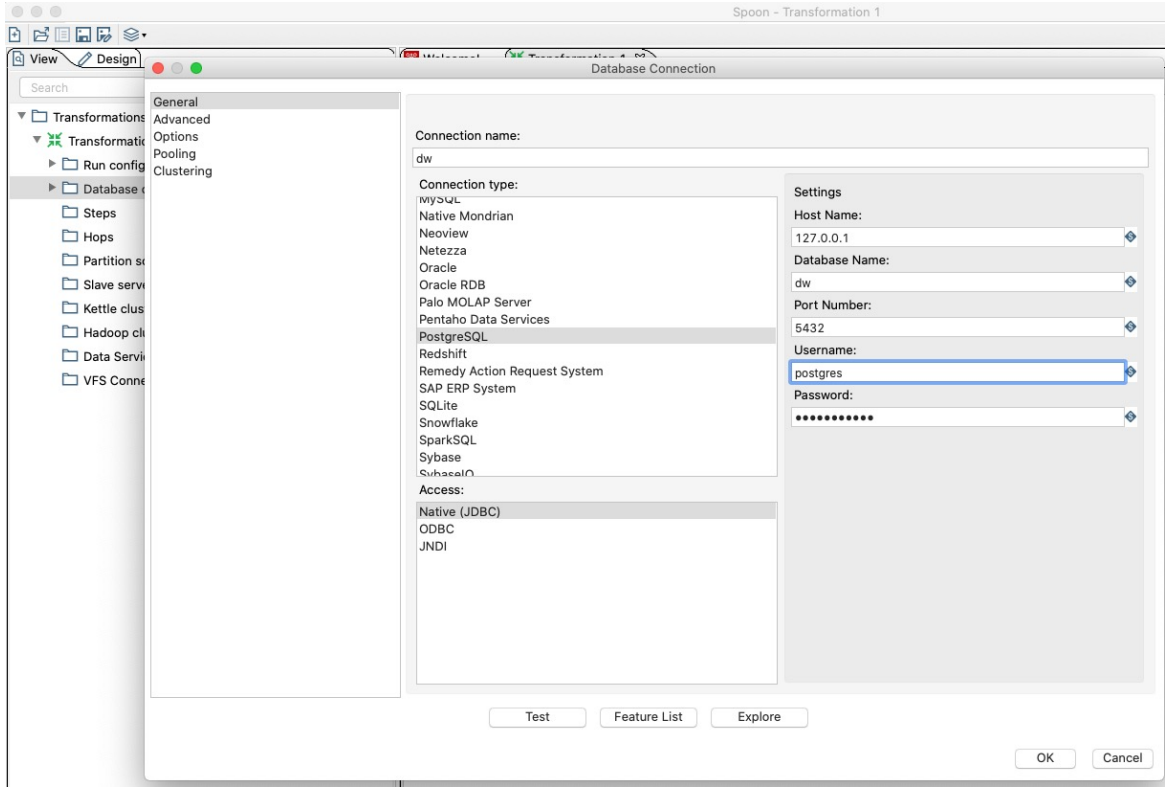
psw-ce-3.14.0.0-12.zip

Unzip to obtain the folder: schema-workbench

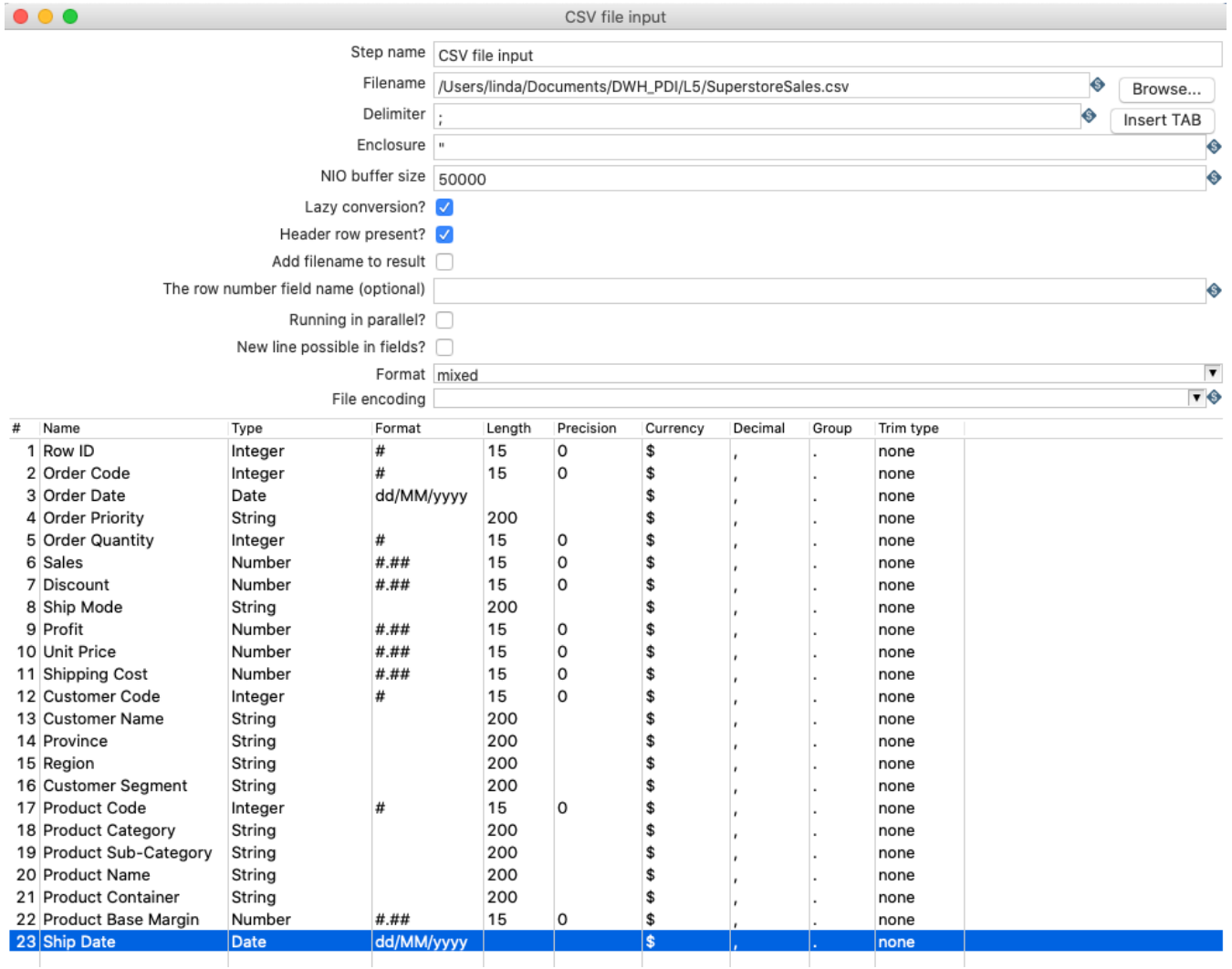
Example of ETL and Data Warehouse

Database connection in PDI

- New transformation
- New database connection in View
  - Host server address/location: localhost 127.0.0.1 take a look to your url http://127.0.0.1:60624/
  - Name database: dw
  - Password
  - User: postgres
- Share



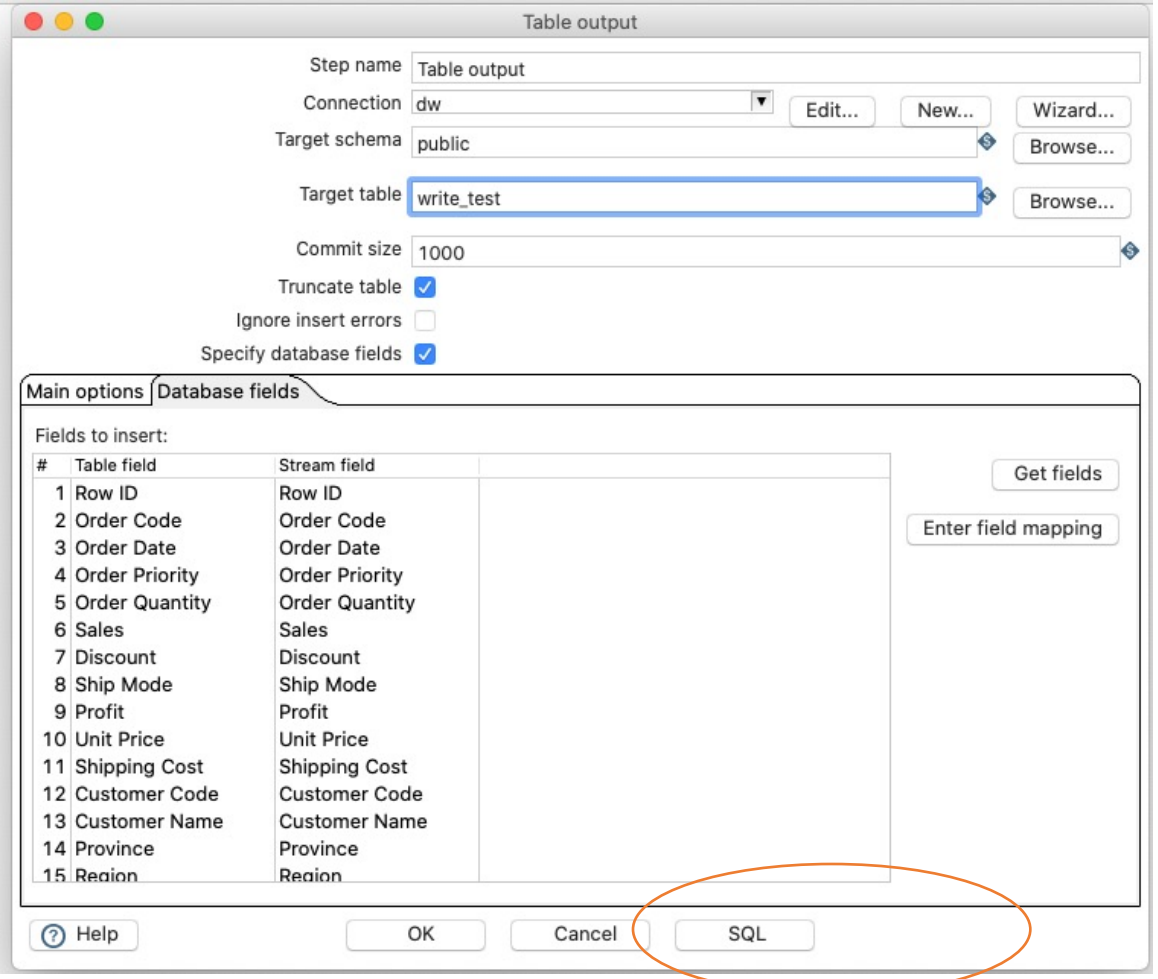
Example Sales: import Table by PDI

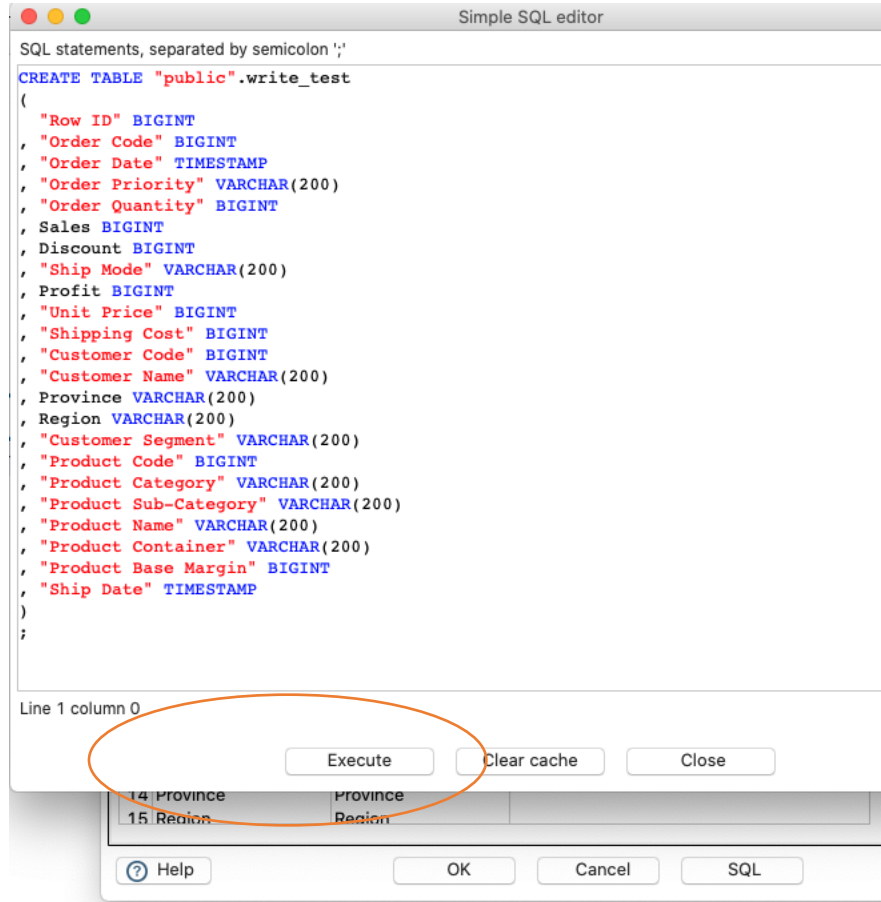


#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Row ID	Integer	#	15	0	$	,	.	none
2	Order Code	Integer	#	15	0	$	,	.	none
3	Order Date	Date	dd/MM/yyyy			$	,	.	none
4	Order Priority	String		200		$	,	.	none
5	Order Quantity	Integer	#	15	0	$	,	.	none
6	Sales	Number	#.##	15	0	$	,	.	none
7	Discount	Number	#.##	15	0	$	,	.	none
8	Ship Mode	String		200		$	,	.	none
9	Profit	Number	#.##	15	0	$	,	.	none
10	Unit Price	Number	#.##	15	0	$	,	.	none
11	Shipping Cost	Number	#.##	15	0	$	,	.	none
12	Customer Code	Integer	#	15	0	$	,	.	none
13	Customer Name	String		200		$	,	.	none
14	Province	String		200		$	,	.	none
15	Region	String		200		$	,	.	none
16	Customer Segment	String		200		$	,	.	none
17	Product Code	Integer	#	15	0	$	,	.	none
18	Product Category	String		200		$	,	.	none
19	Product Sub-Category	String		200		$	,	.	none
20	Product Name	String		200		$	,	.	none
21	Product Container	String		200		$	,	.	none
22	Product Base Margin	Number	#.##	15	0	$	,	.	none
23	Ship Date	Date	dd/MM/yyyy			$	,	.	none

Check Length and Format for Number type will be #.##

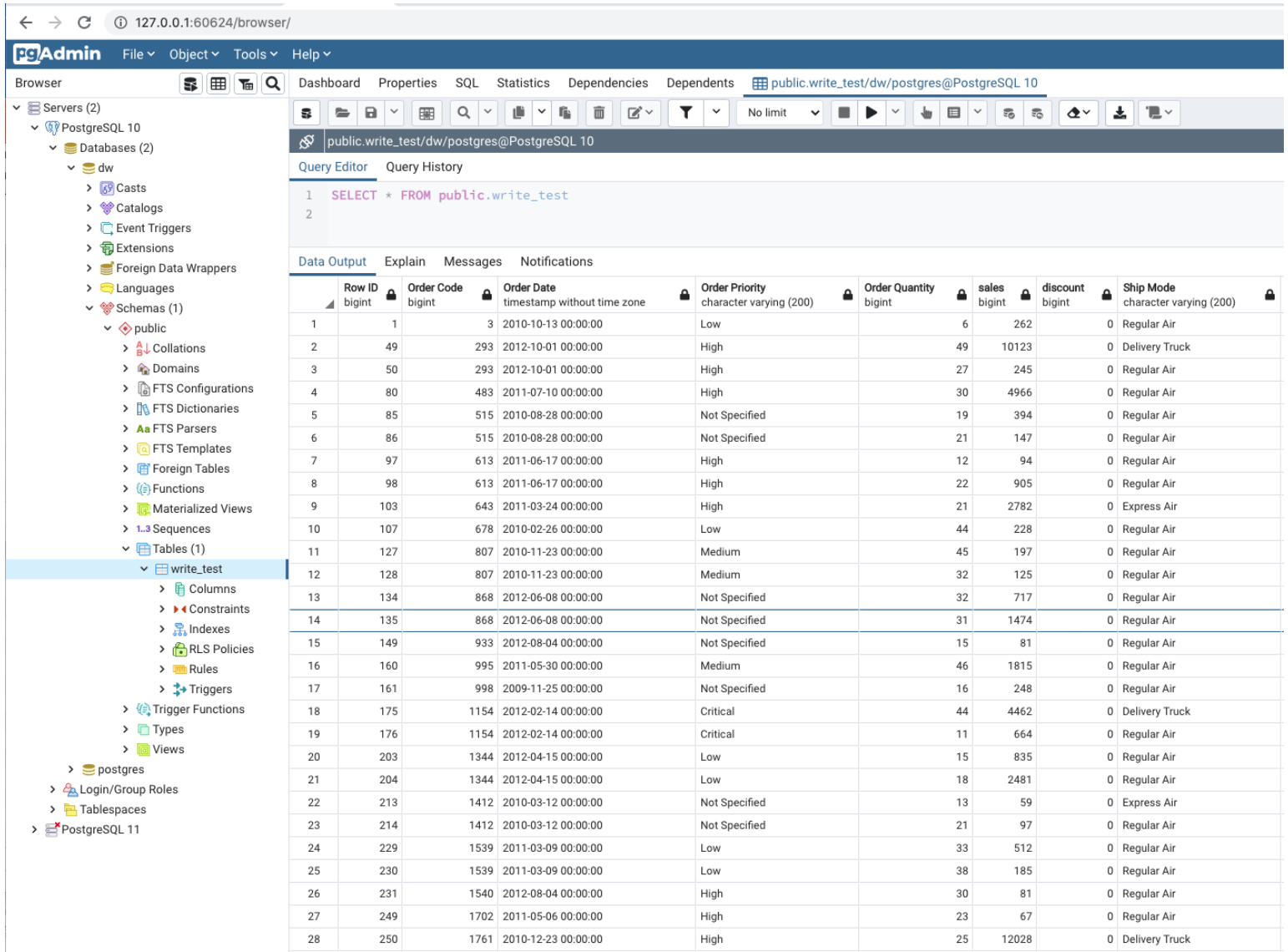
Example Sales: import and create Table by PDI





createTableSales.ktr

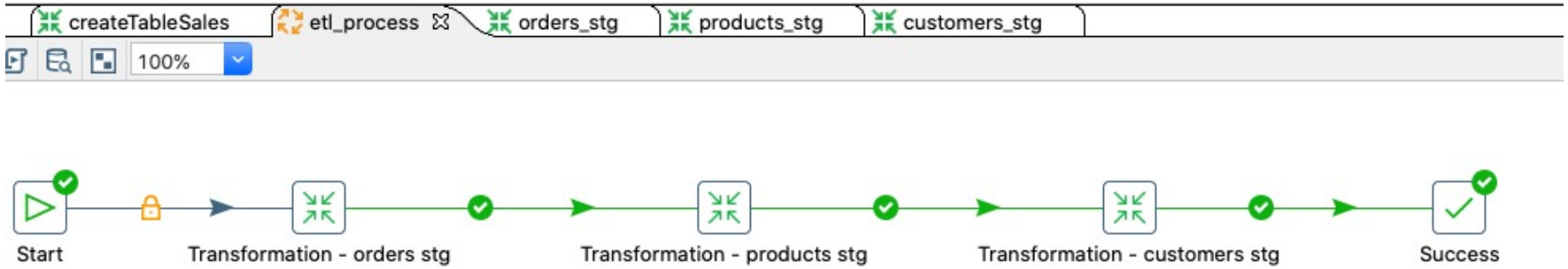
Example Sales: the table in the DW



	Row ID bigint	Order Code bigint	Order Date timestamp without time zone	Order Priority character varying (200)	Order Quantity bigint	sales bigint	discount bigint	Ship Mode character varying (200)
1	1	3	2010-10-13 00:00:00	Low	6	262	0	Regular Air
2	49	293	2012-10-01 00:00:00	High	49	10123	0	Delivery Truck
3	50	293	2012-10-01 00:00:00	High	27	245	0	Regular Air
4	80	483	2011-07-10 00:00:00	High	30	4966	0	Regular Air
5	85	515	2010-08-28 00:00:00	Not Specified	19	394	0	Regular Air
6	86	515	2010-08-28 00:00:00	Not Specified	21	147	0	Regular Air
7	97	613	2011-06-17 00:00:00	High	12	94	0	Regular Air
8	98	613	2011-06-17 00:00:00	High	22	905	0	Regular Air
9	103	643	2011-03-24 00:00:00	High	21	2782	0	Express Air
10	107	678	2010-02-26 00:00:00	Low	44	228	0	Regular Air
11	127	807	2010-11-23 00:00:00	Medium	45	197	0	Regular Air
12	128	807	2010-11-23 00:00:00	Medium	32	125	0	Regular Air
13	134	868	2012-06-08 00:00:00	Not Specified	32	717	0	Regular Air
14	135	868	2012-06-08 00:00:00	Not Specified	31	1474	0	Regular Air
15	149	933	2012-08-04 00:00:00	Not Specified	15	81	0	Regular Air
16	160	995	2011-05-30 00:00:00	Medium	46	1815	0	Regular Air
17	161	998	2009-11-25 00:00:00	Not Specified	16	248	0	Regular Air
18	175	1154	2012-02-14 00:00:00	Critical	44	4462	0	Delivery Truck
19	176	1154	2012-02-14 00:00:00	Critical	11	664	0	Regular Air
20	203	1344	2012-04-15 00:00:00	Low	15	835	0	Regular Air
21	204	1344	2012-04-15 00:00:00	Low	18	2481	0	Regular Air
22	213	1412	2010-03-12 00:00:00	Not Specified	13	59	0	Express Air
23	214	1412	2010-03-12 00:00:00	Not Specified	21	97	0	Regular Air
24	229	1539	2011-03-09 00:00:00	Low	33	512	0	Regular Air
25	230	1539	2011-03-09 00:00:00	Low	38	185	0	Regular Air
26	231	1540	2012-08-04 00:00:00	High	30	81	0	Regular Air
27	249	1702	2011-05-06 00:00:00	High	23	67	0	Regular Air
28	250	1761	2010-12-23 00:00:00	High	25	12028	0	Delivery Truck

Example create a job – create staging tables

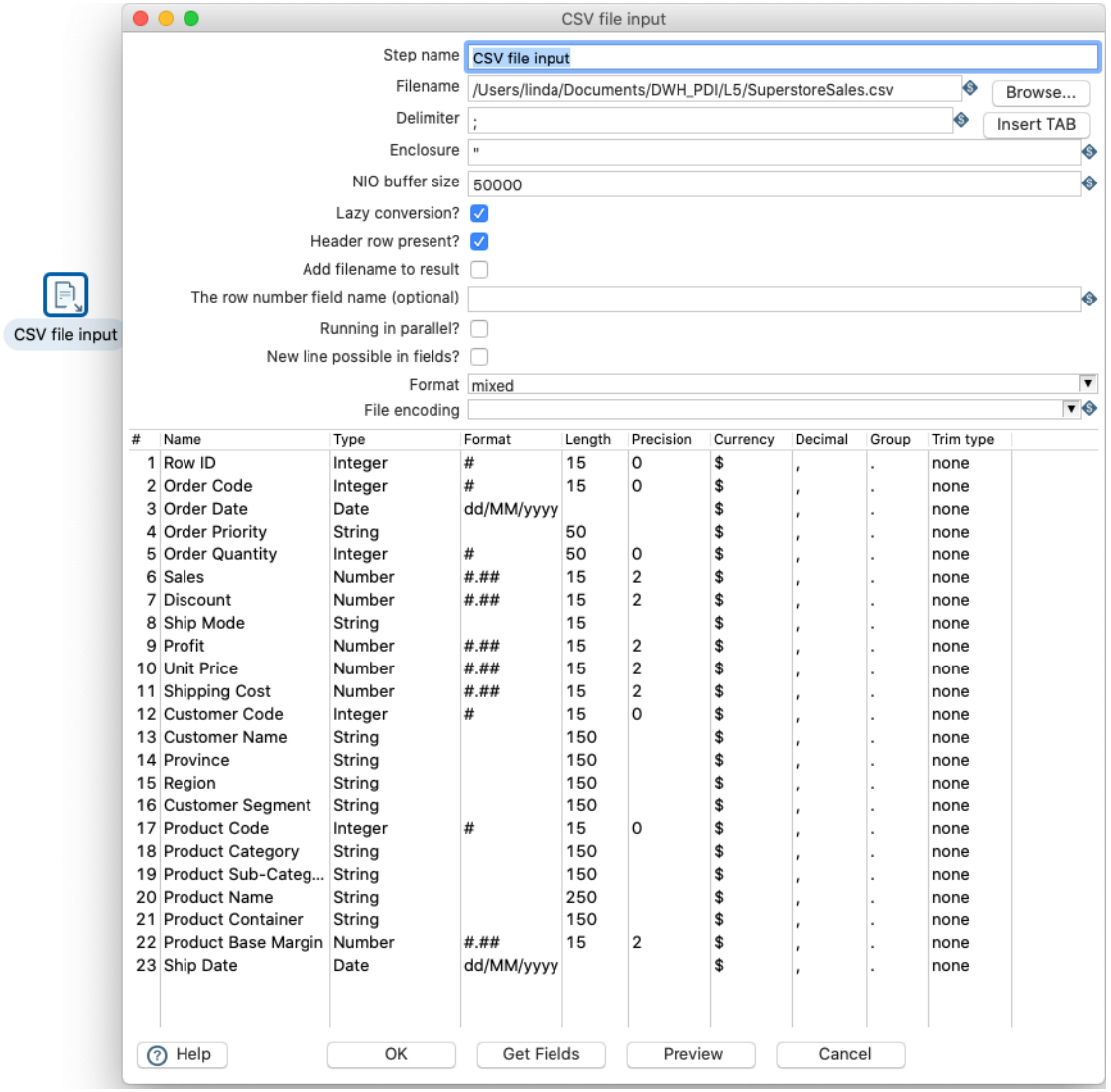
etl_process.kjb



orders_stg.ktr products_stg.ktr customers_stg.ktr

SQL --> Execute
in the transformations to create the tables

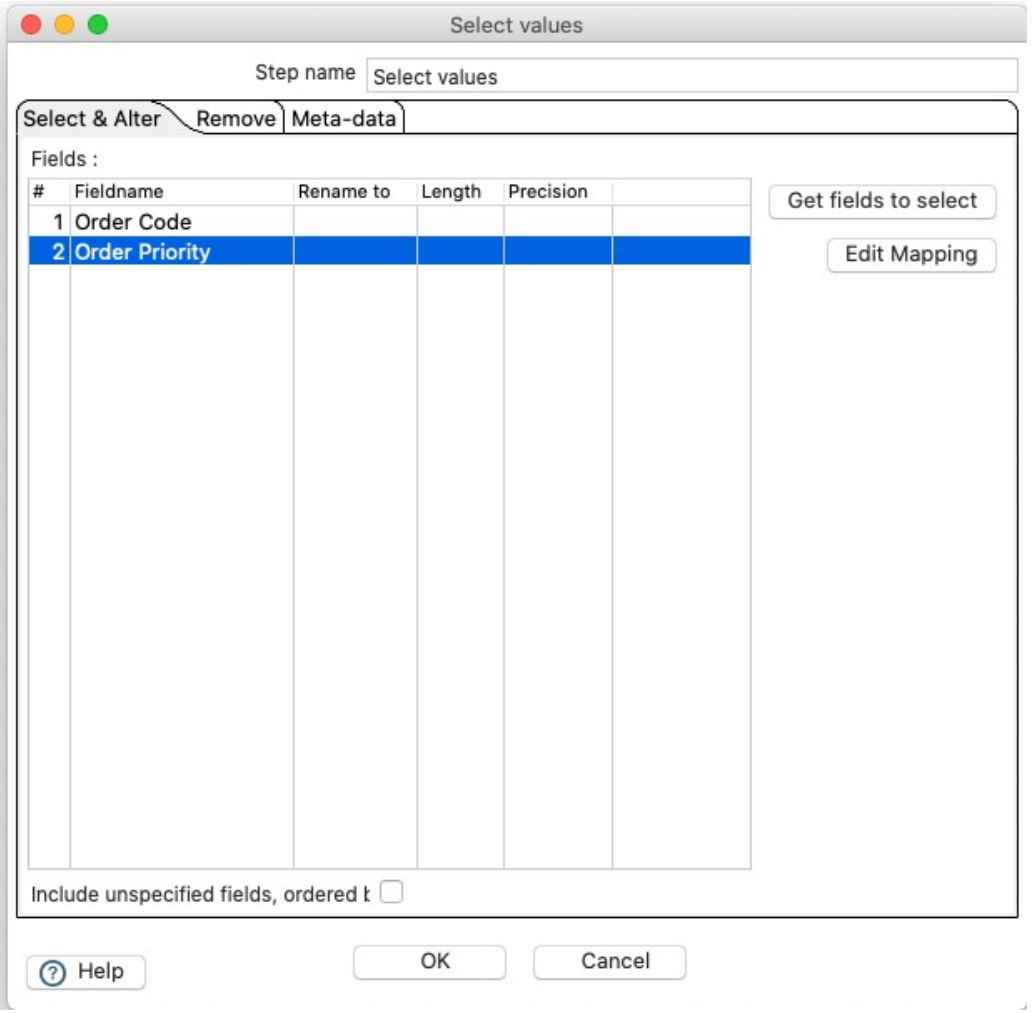
Example – orders staging table



#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Row ID	Integer	#	15	0	$	,	.	none
2	Order Code	Integer	#	15	0	$	,	.	none
3	Order Date	Date	dd/MM/yyyy			$	,	.	none
4	Order Priority	String		50		$	,	.	none
5	Order Quantity	Integer	#	50	0	$	,	.	none
6	Sales	Number	#.##	15	2	$	,	.	none
7	Discount	Number	#.##	15	2	$	,	.	none
8	Ship Mode	String		15		$	,	.	none
9	Profit	Number	#.##	15	2	$	,	.	none
10	Unit Price	Number	#.##	15	2	$	,	.	none
11	Shipping Cost	Number	#.##	15	2	$	,	.	none
12	Customer Code	Integer	#	15	0	$	,	.	none
13	Customer Name	String		150		$	,	.	none
14	Province	String		150		$	,	.	none
15	Region	String		150		$	,	.	none
16	Customer Segment	String		150		$	,	.	none
17	Product Code	Integer	#	15	0	$	,	.	none
18	Product Category	String		150		$	,	.	none
19	Product Sub-Categ...	String		150		$	,	.	none
20	Product Name	String		250		$	,	.	none
21	Product Container	String		150		$	,	.	none
22	Product Base Margin	Number	#.##	15	2	$	,	.	none
23	Ship Date	Date	dd/MM/yyyy			$	,	.	none

orders_stg.ktr

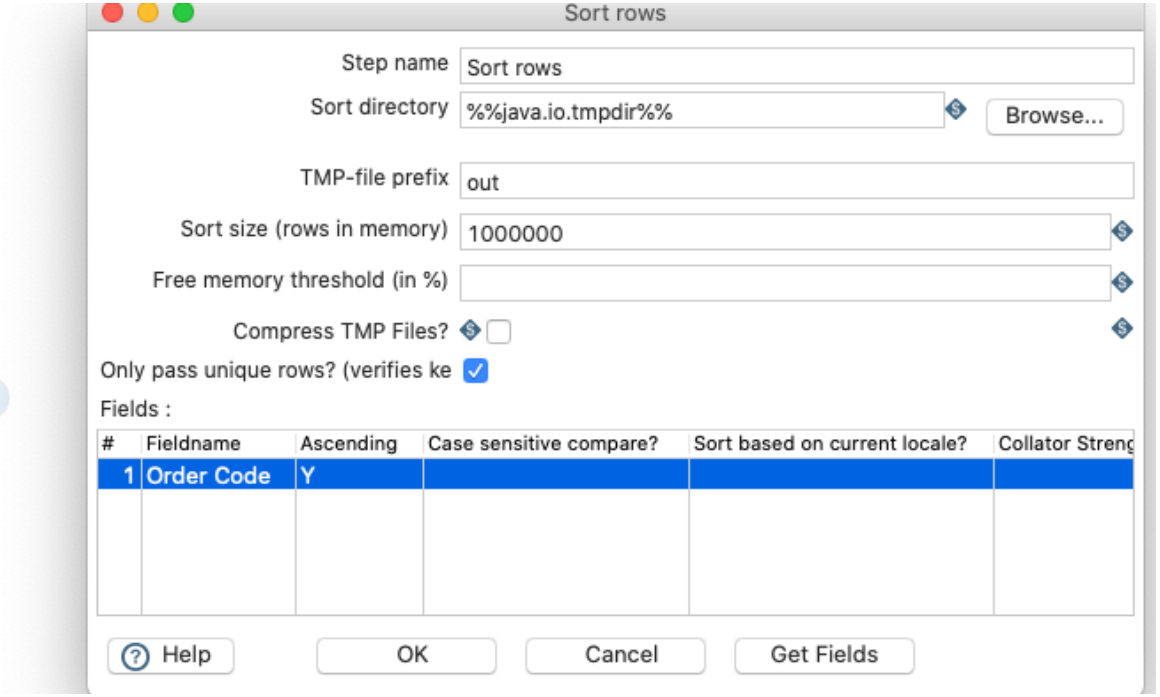
Example – orders staging table



orders_stg.ktr

Example – orders staging table

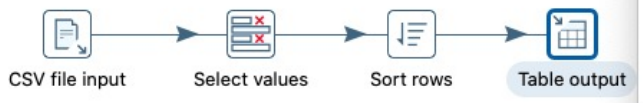
orders_stg.ktr



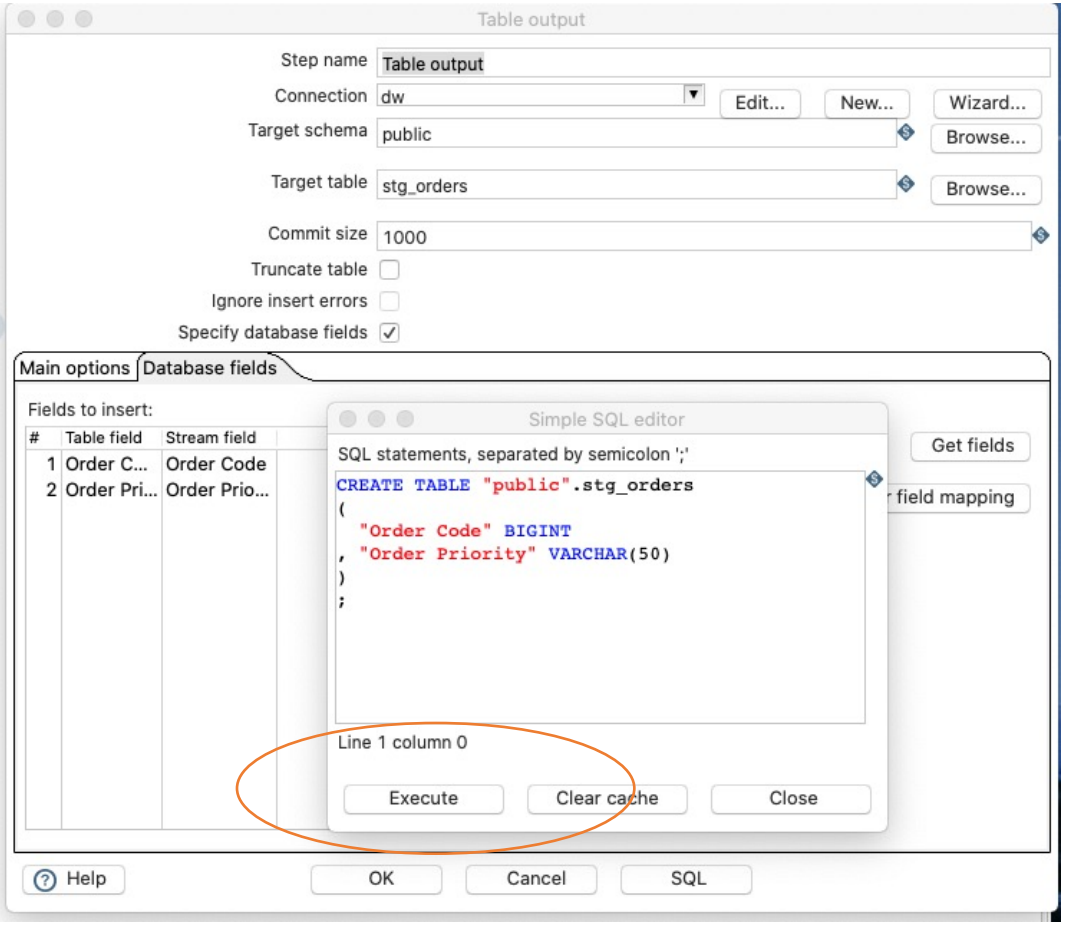
Make order code unique

Example – orders staging table

orders_stg.ktr



Create the table orders dimension



Remember:
SQL --> Execute to create the table!

Example – products and customers staging tables

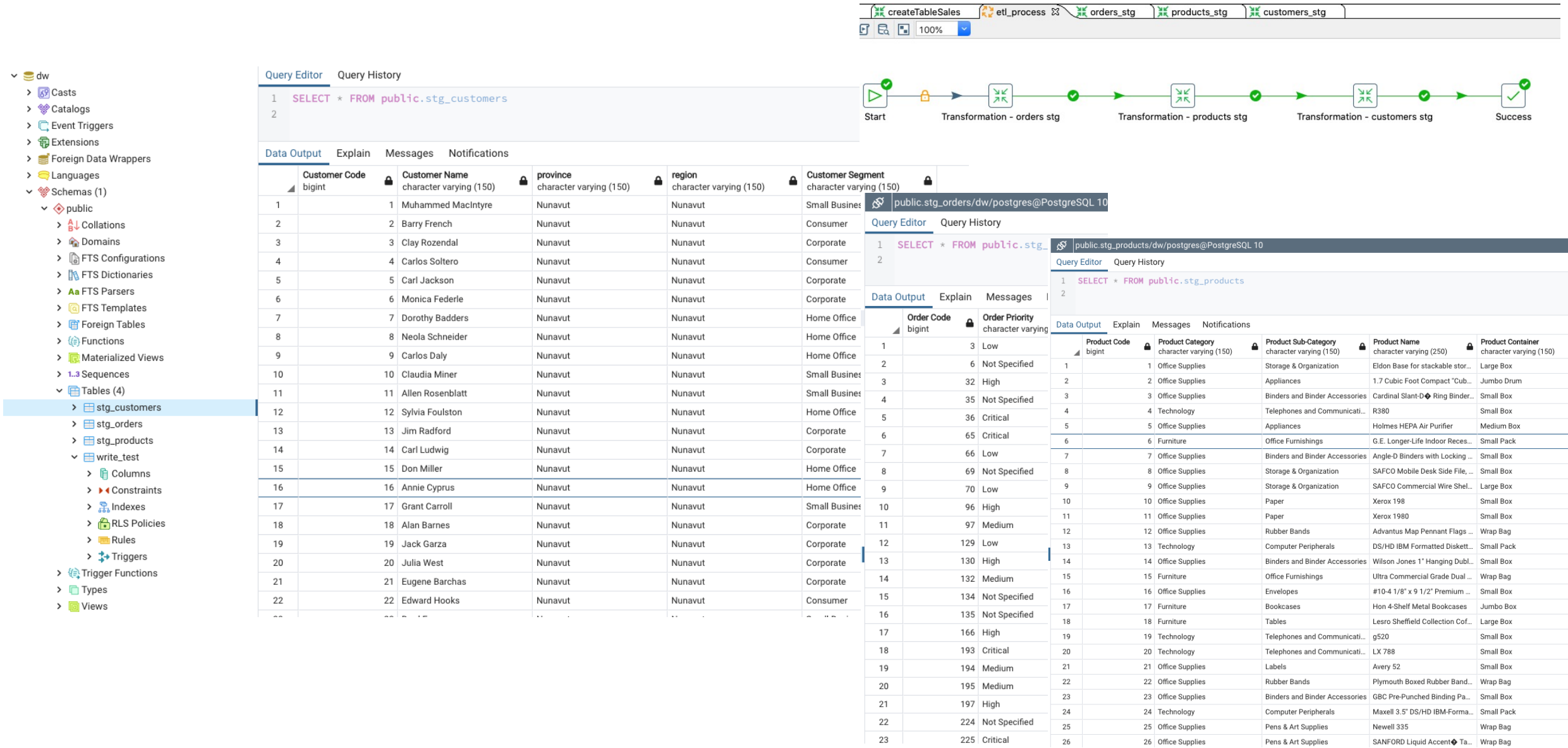
- Similar processes to the previous for the creation of the orders staging table

products_stg.ktr

customers_stg.ktr

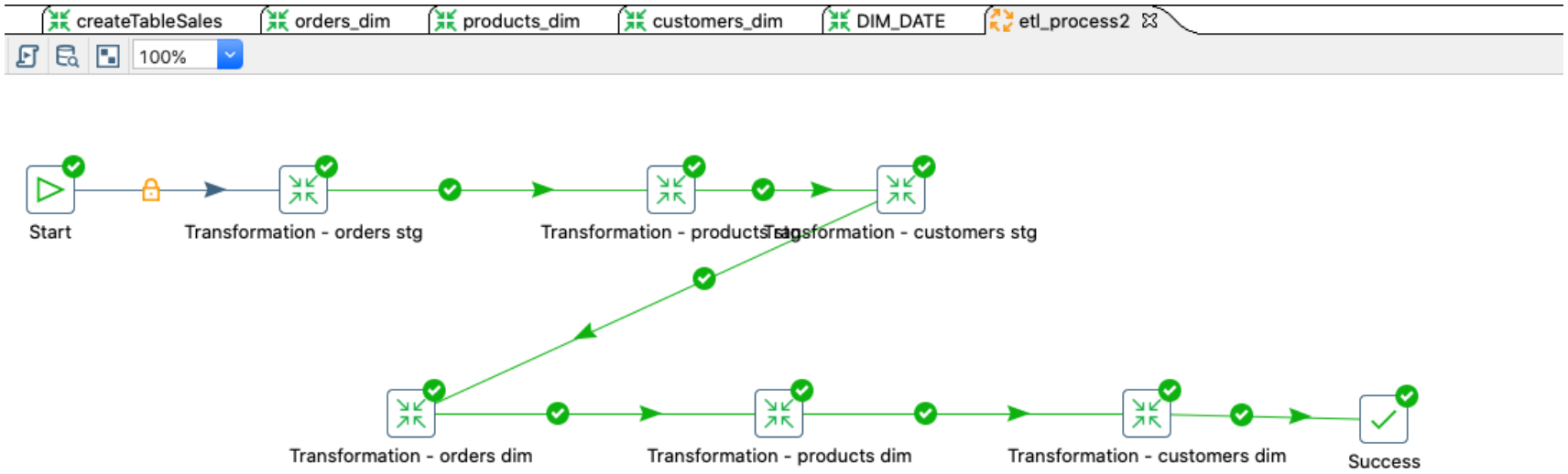
Remember:
SQL --> Execute to create the table!

Example create a job – creating staging tables



Example create a job – creating dimension tables

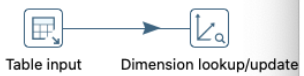
etl_process2.kjb



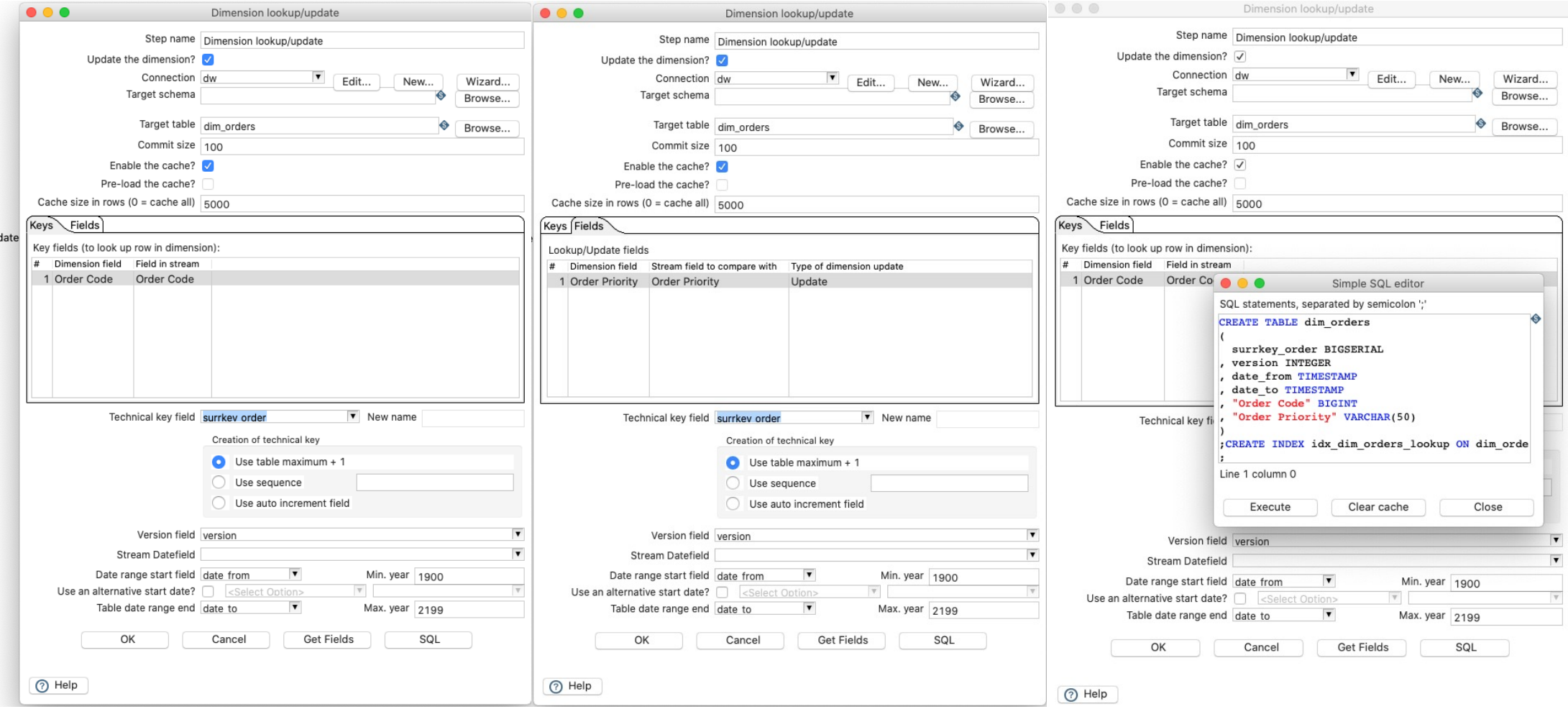
orders_dim.ktr products_dim.ktr customers_dim.ktr

SQL --> Execute
in the transformations to create the tables

Example – orders dimension



Remember:
SQL -->
Execute to create the table!



Example – products and customers dimension

- Similar processes as for the orders dimension tables

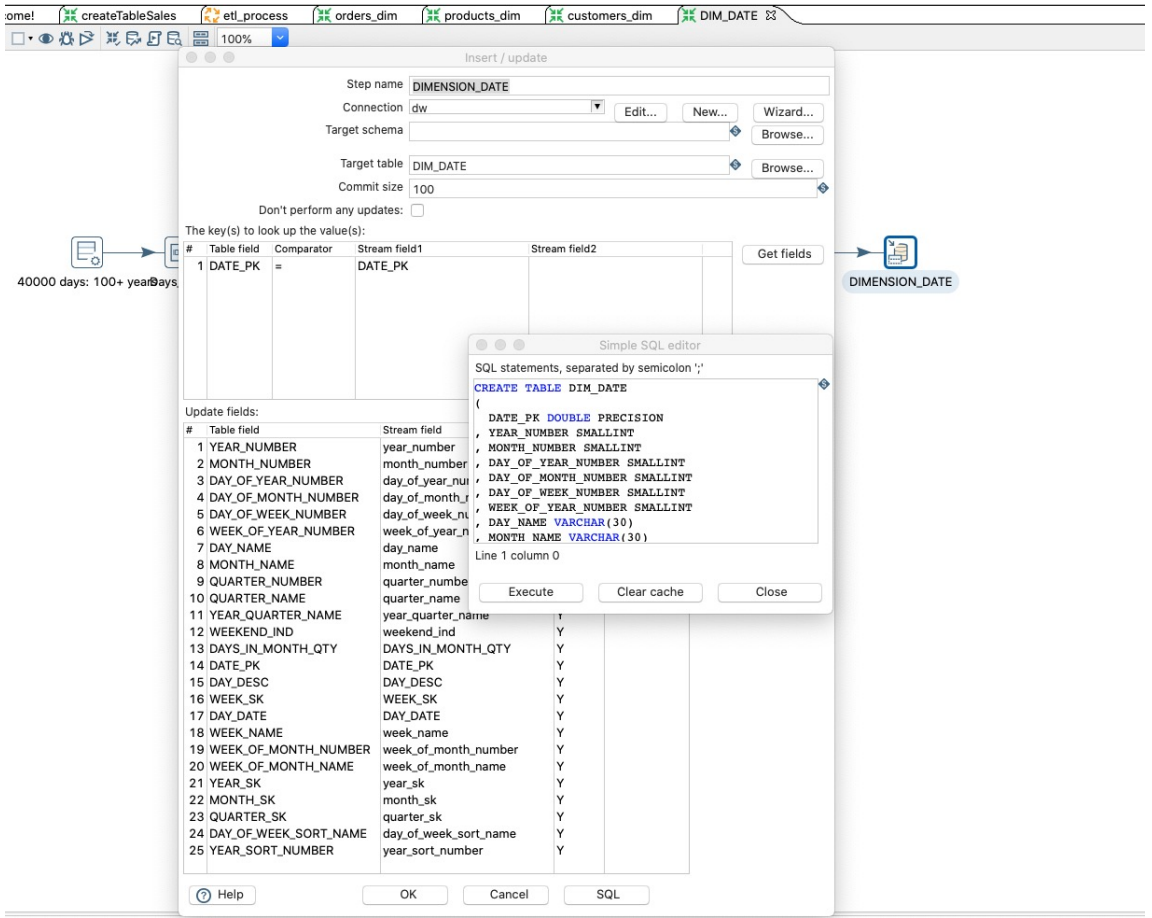
Remember:
SQL --> Execute to create the table!

Example – time dimension

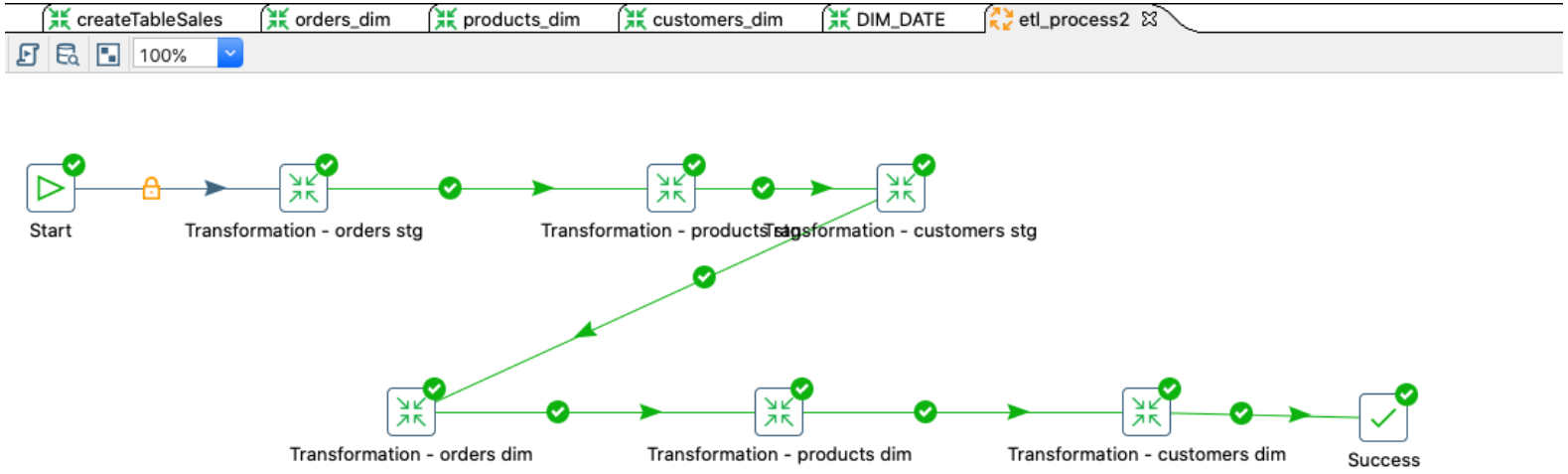
- In addition we use a dim_date transformation
- The time doesn't change so we will not execute this function anymore, it is only execute one time

DIM_DATE.ktr

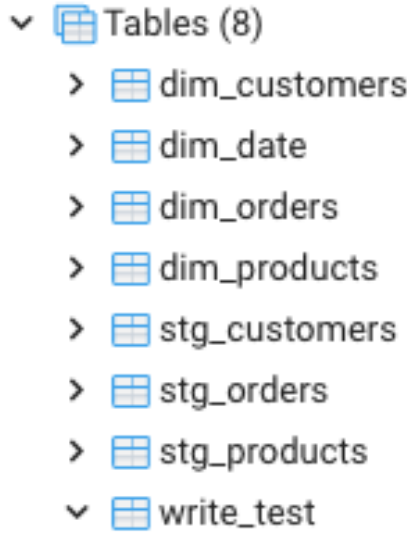
Remember:
SQL --> Execute to create the table!



Example create a job – creating dimension tables

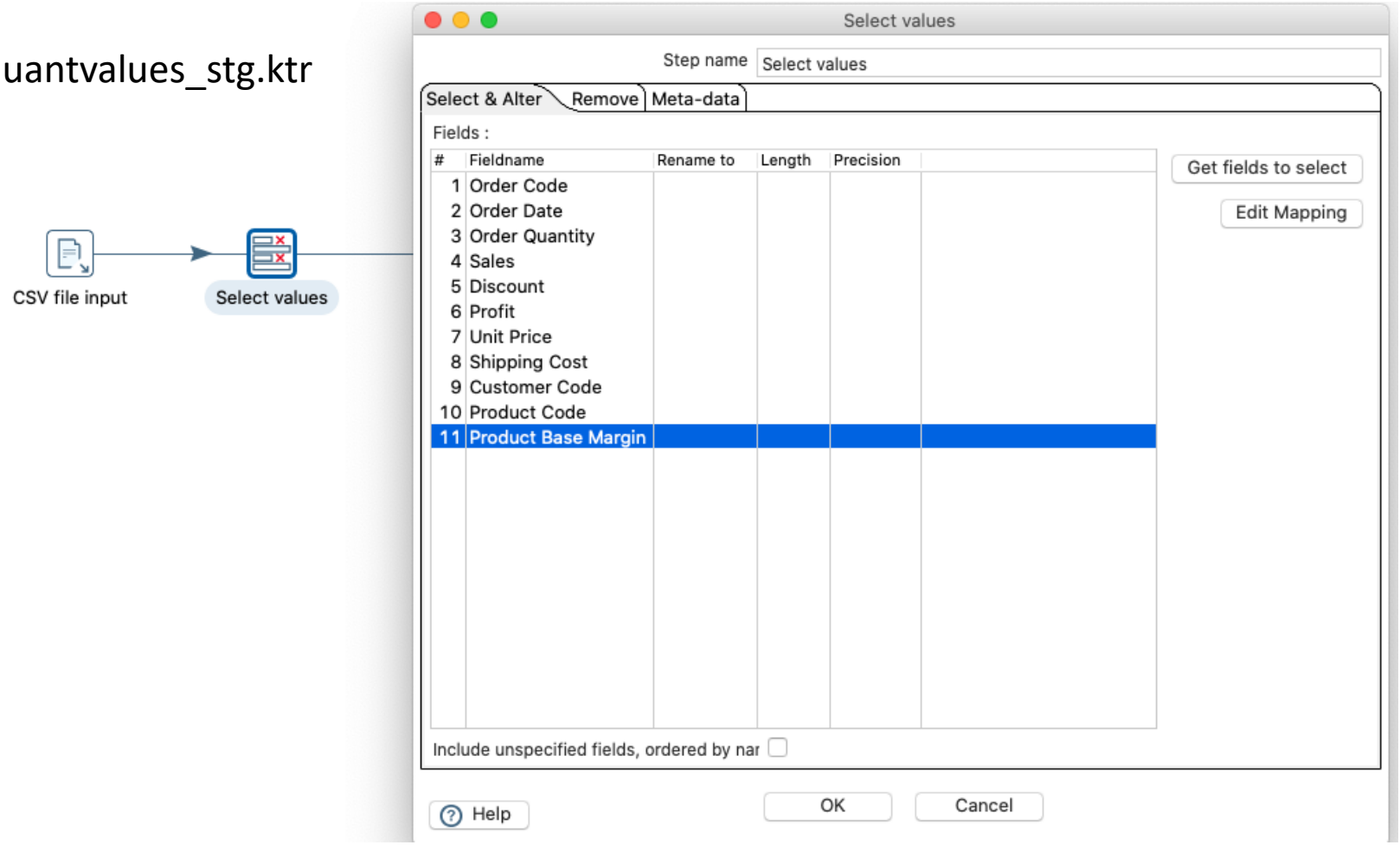


+ DIM_DATE.ktr



Example – creating the staging quantitative values table

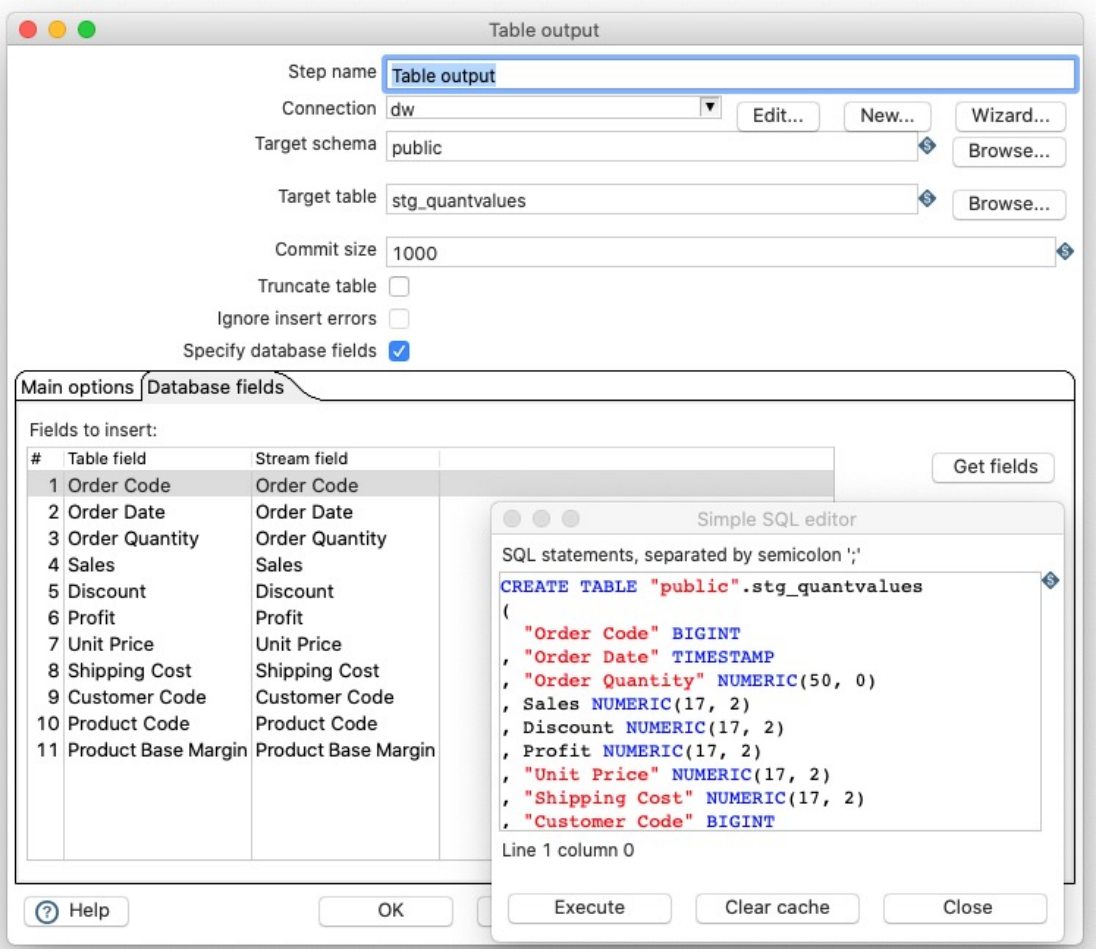
quantvalues_stg.ktr



We keep only quantitative values and id for dimensions

Example – creating the staging quantitative values table

quantvalues_stg.ktr

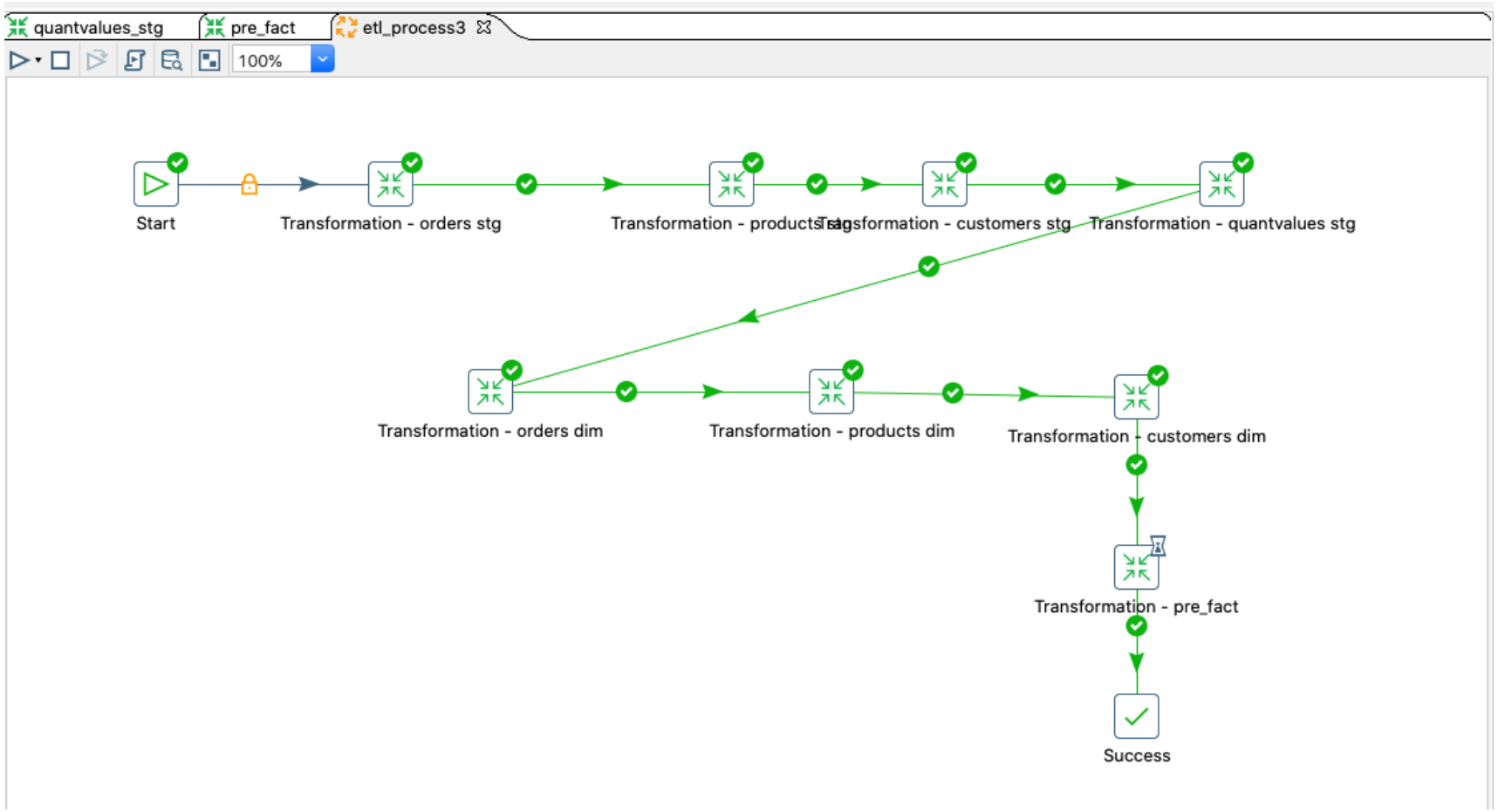


We keep only quantitative values and id for dimensions

Remember:
SQL --> Execute to create the table!

Example create a job – creating a pre-fact table

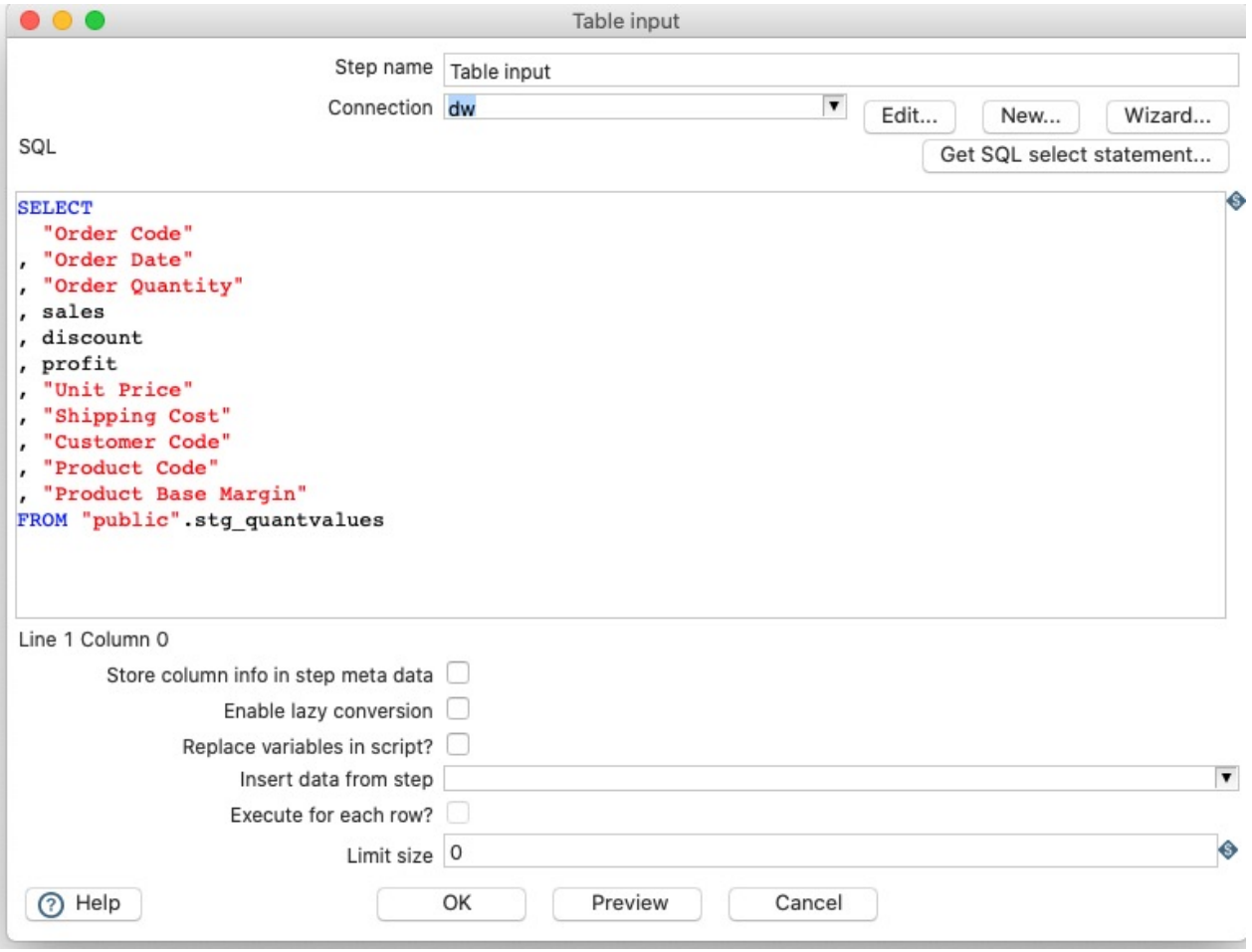
etl_process3.kjb



pre_fact.ktr

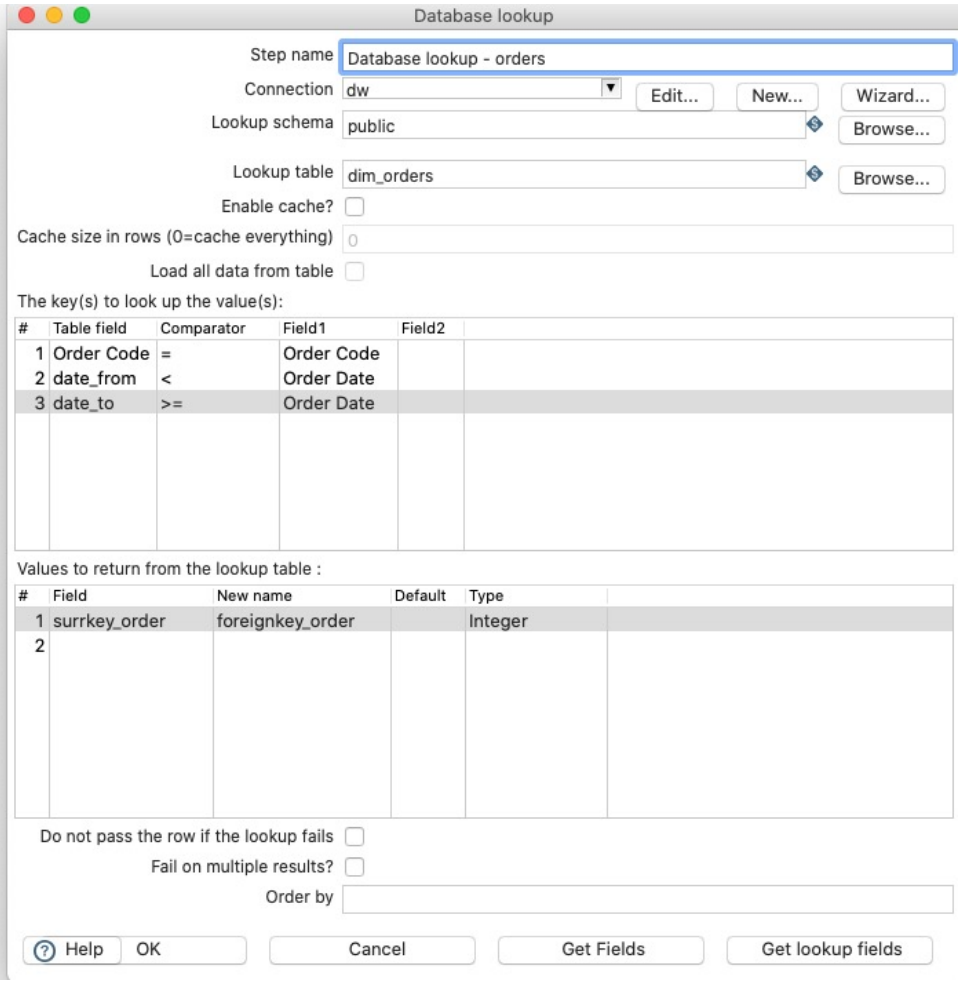
Remember:
SQL --> Execute to create the table!

Example – creating a pre-fact table



pre_fact.ktr

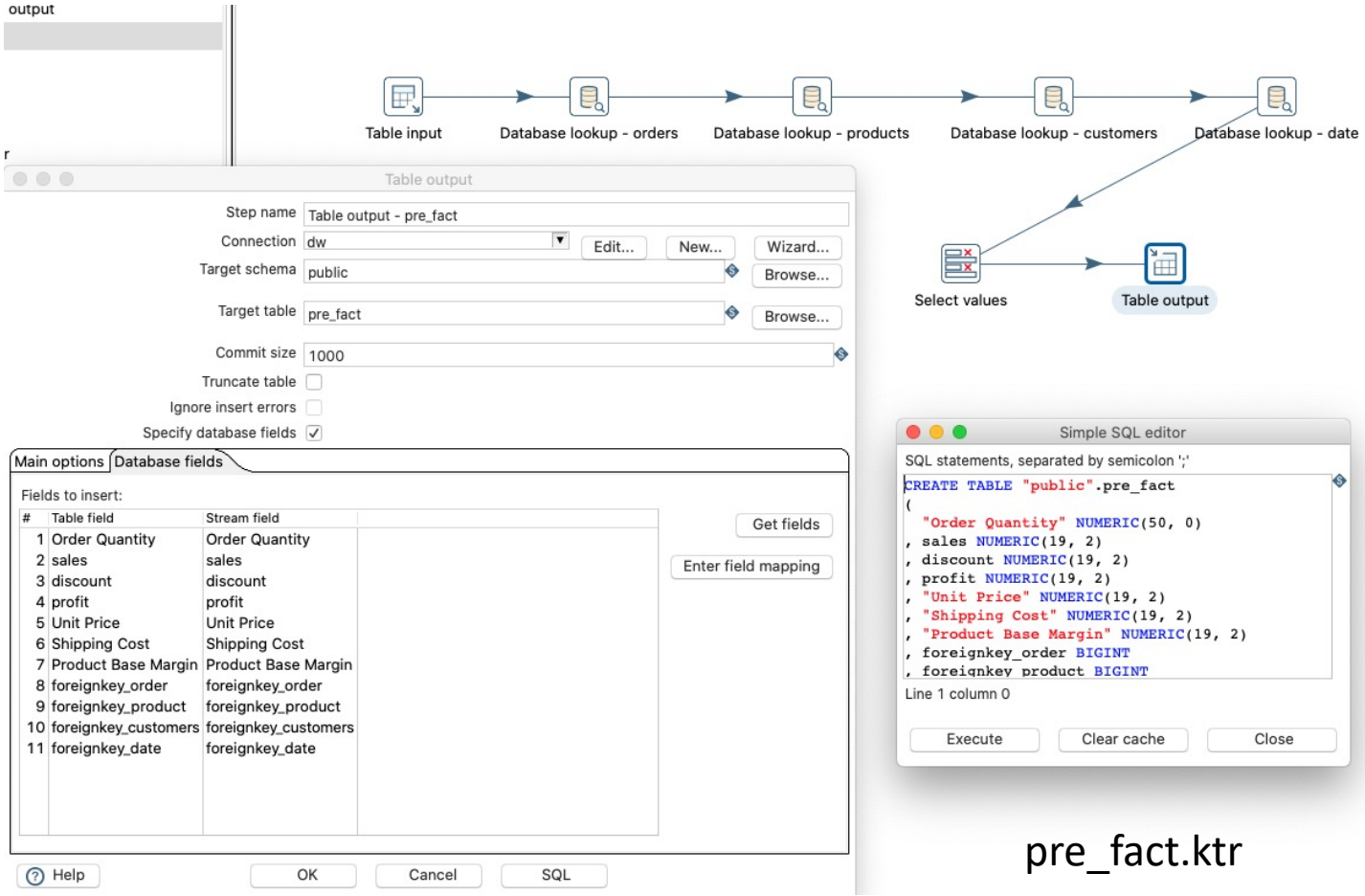
Example – creating a pre-fact table



Similar database lookup for dim_orders, dim_products, dim_customers, dim_date

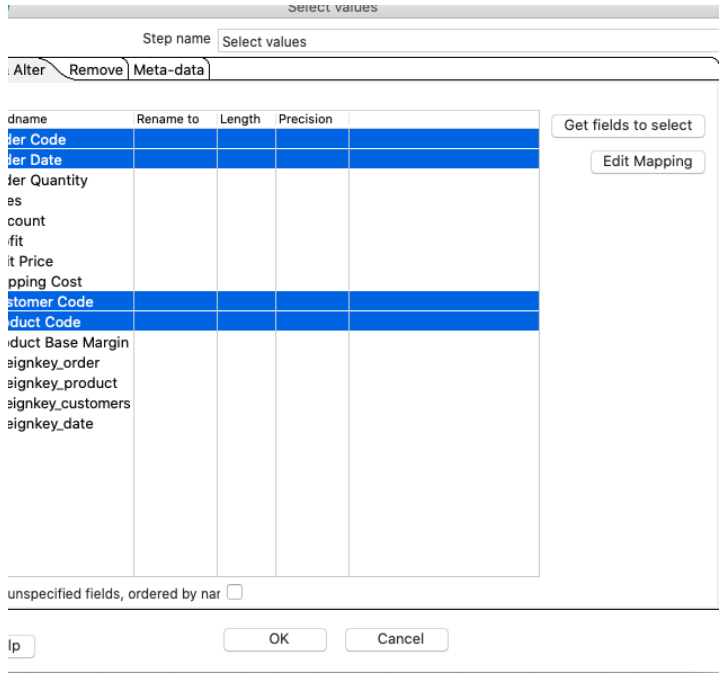
pre_fact.ktr

Example – creating a pre-fact table



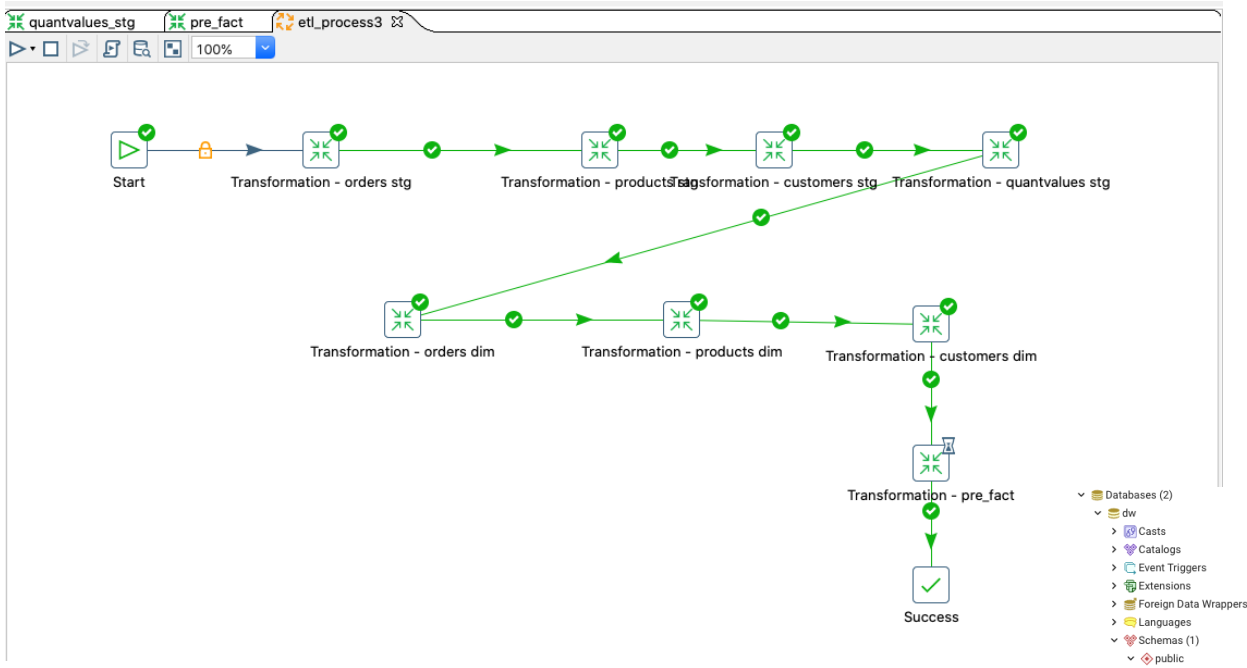
pre_fact.ktr

Select only the foreign keys and the quantitative values in Select values

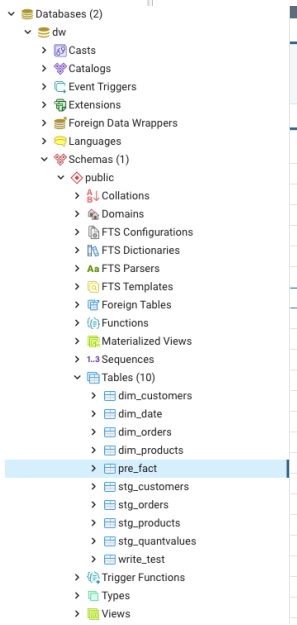


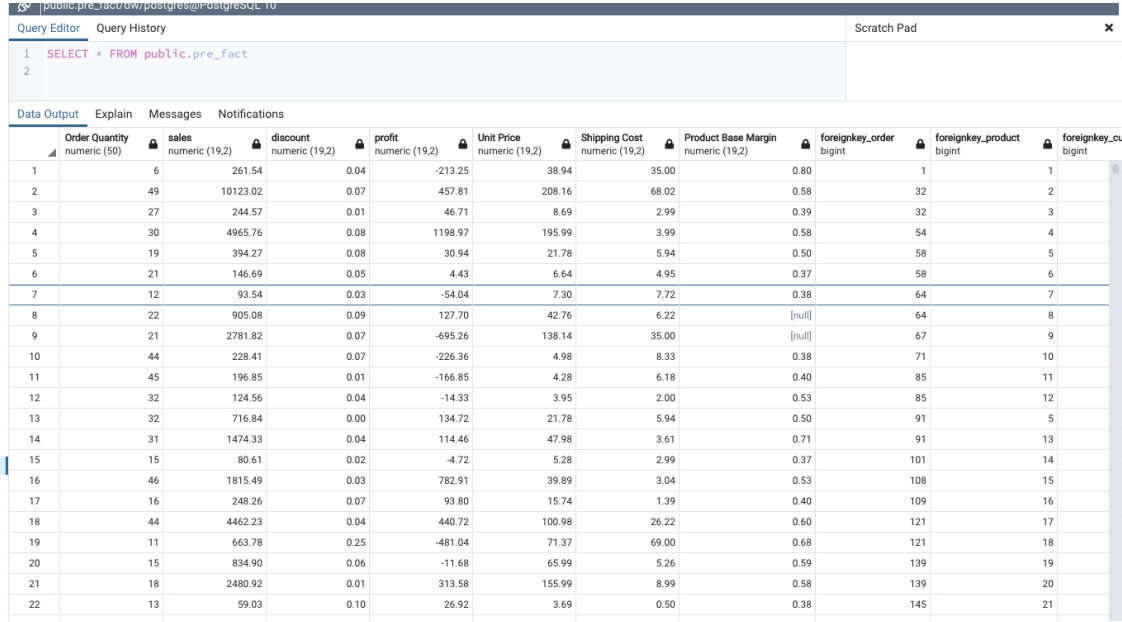
Remember:
SQL --> Execute to create the table!

Example create a job – creating pre-fact table



etl_process3.kjb

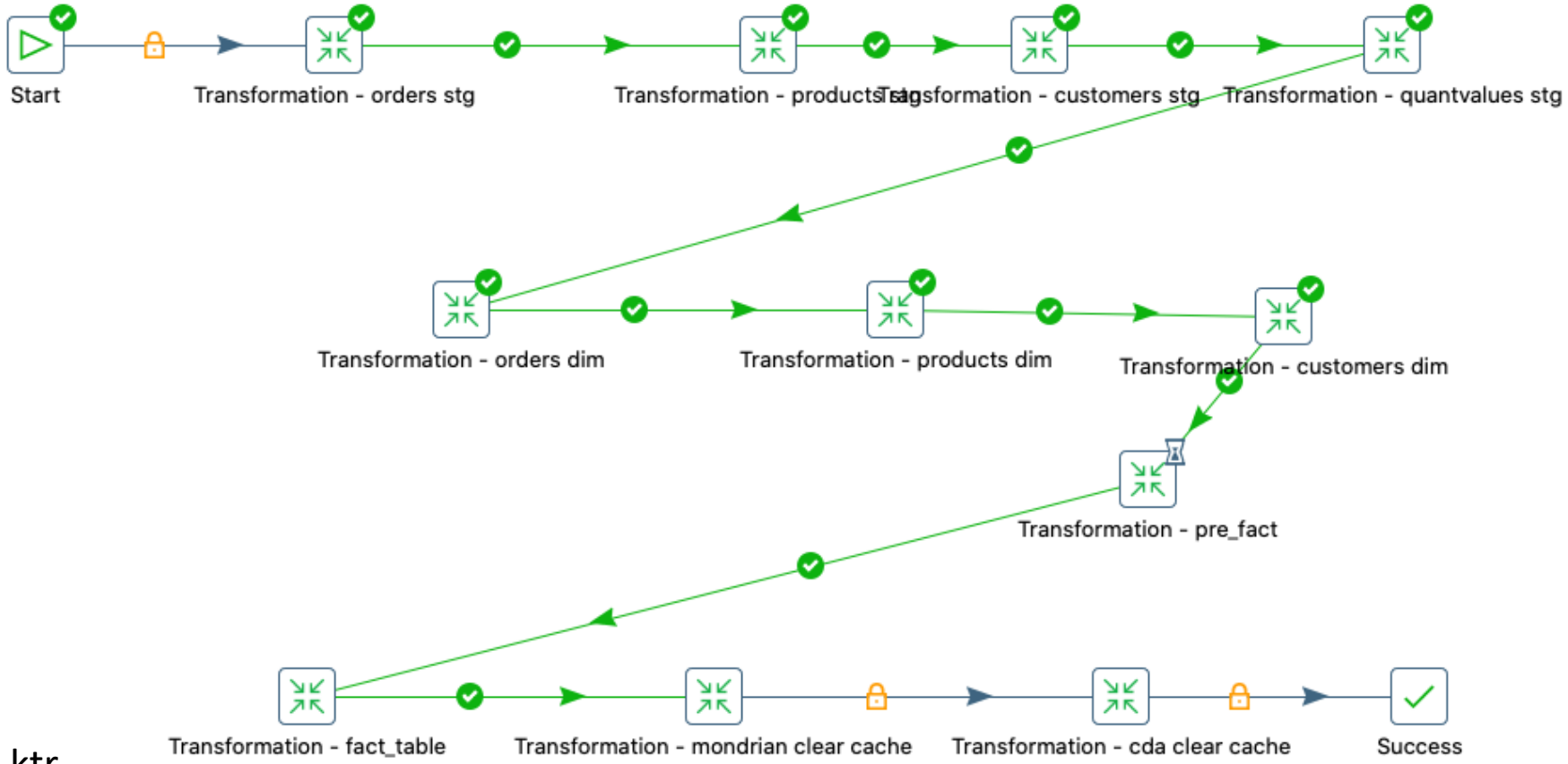




	Order Quantity numeric (50)	sales numeric (19,2)	discount numeric (19,2)	profit numeric (19,2)	Unit Price numeric (19,2)	Shipping Cost numeric (19,2)	Product Base Margin numeric (19,2)	foreignkey_order bigint	foreignkey_product bigint	foreignkey_cu bigint
1	6	261.54	0.04	-213.25	38.94	35.00	0.80	1	1	
2	49	10123.02	0.07	457.81	208.16	68.02	0.58	32	2	
3	27	244.57	0.01	46.71	8.69	2.99	0.39	32	3	
4	30	4965.76	0.08	1198.97	195.99	3.99	0.58	54	4	
5	19	394.27	0.08	30.94	21.78	5.94	0.50	58	5	
6	21	146.69	0.05	4.43	6.64	4.95	0.37	58	6	
7	12	93.54	0.03	-54.04	7.30	7.72	0.38	64	7	
8	22	905.08	0.09	127.70	42.76	6.22	[null]	64	8	
9	21	2781.82	0.07	-695.26	138.14	35.00	[null]	67	9	
10	44	228.41	0.07	-226.36	4.98	8.33	0.38	71	10	
11	45	196.85	0.01	-166.85	4.28	6.18	0.40	85	11	
12	32	124.56	0.04	-14.33	3.95	2.00	0.53	85	12	
13	32	716.84	0.00	134.72	21.78	5.94	0.50	91	5	
14	31	1474.33	0.04	114.46	47.98	3.61	0.71	91	13	
15	15	80.61	0.02	-4.72	5.28	2.99	0.37	101	14	
16	46	1815.49	0.03	782.91	39.89	3.04	0.53	108	15	
17	16	248.26	0.07	93.80	15.74	1.39	0.40	109	16	
18	44	4462.23	0.04	440.72	100.98	26.22	0.60	121	17	
19	11	663.78	0.25	-481.04	71.37	69.00	0.68	121	18	
20	15	834.90	0.06	-11.68	65.99	5.26	0.59	139	19	
21	18	2480.92	0.01	313.58	155.99	8.99	0.58	139	20	
22	13	59.03	0.10	26.92	3.69	0.50	0.38	145	21	

Example create a job – creating a fact table

etl_process4.kjb



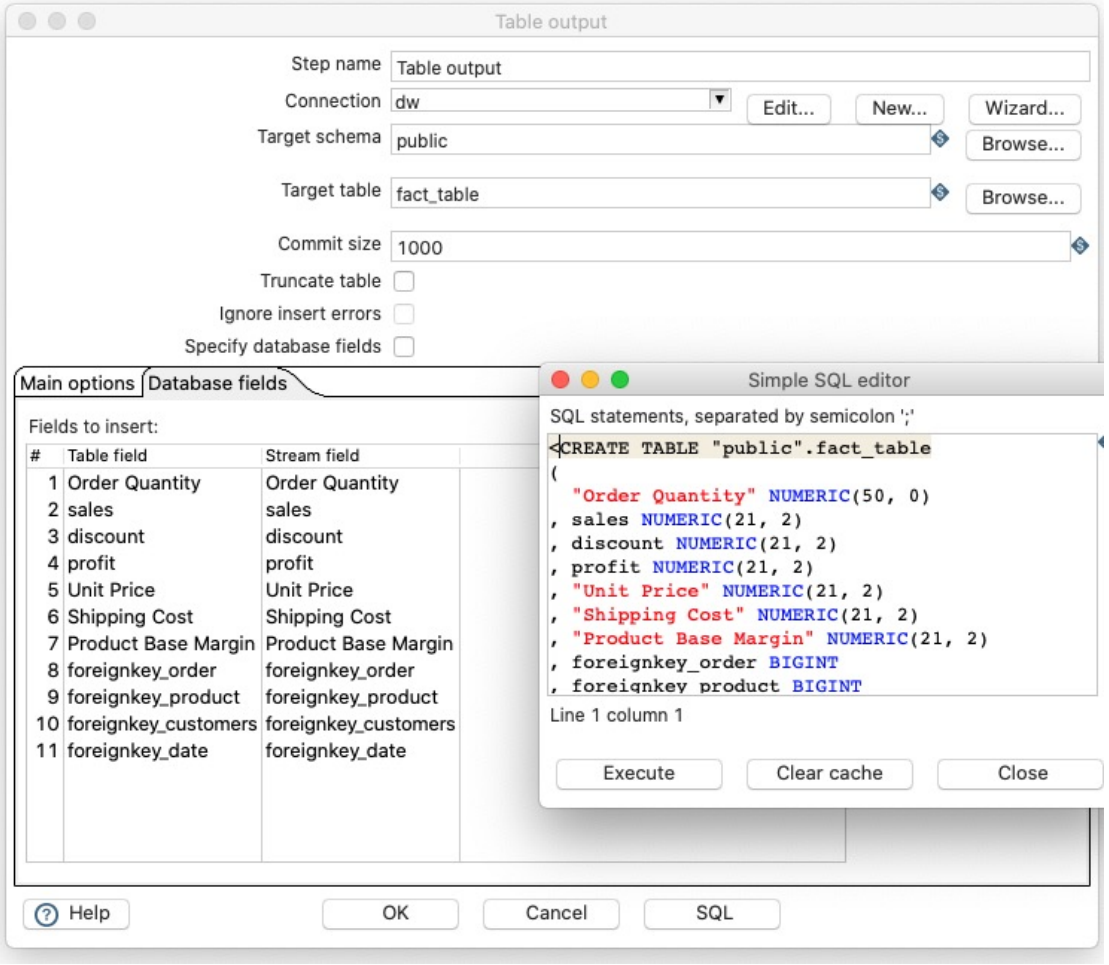
fact_table.ktr

mondrian_clear_cache.ktr

cda_clear_cache.ktr

Remember:
SQL --> Execute to create the table!

Example – creating a **fact table**

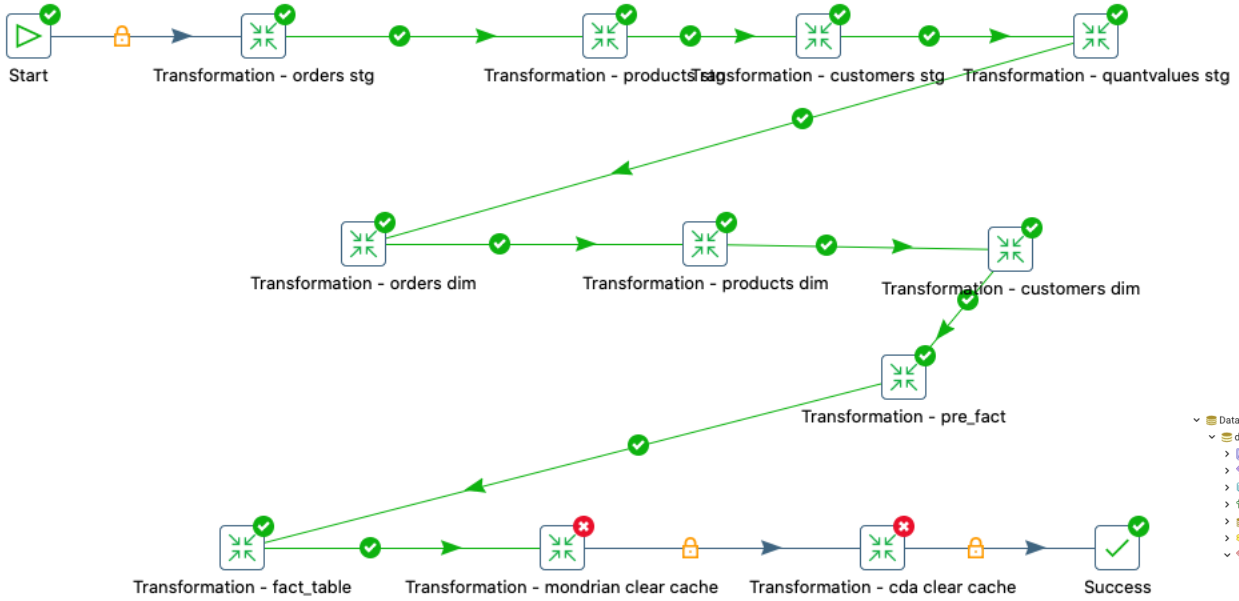


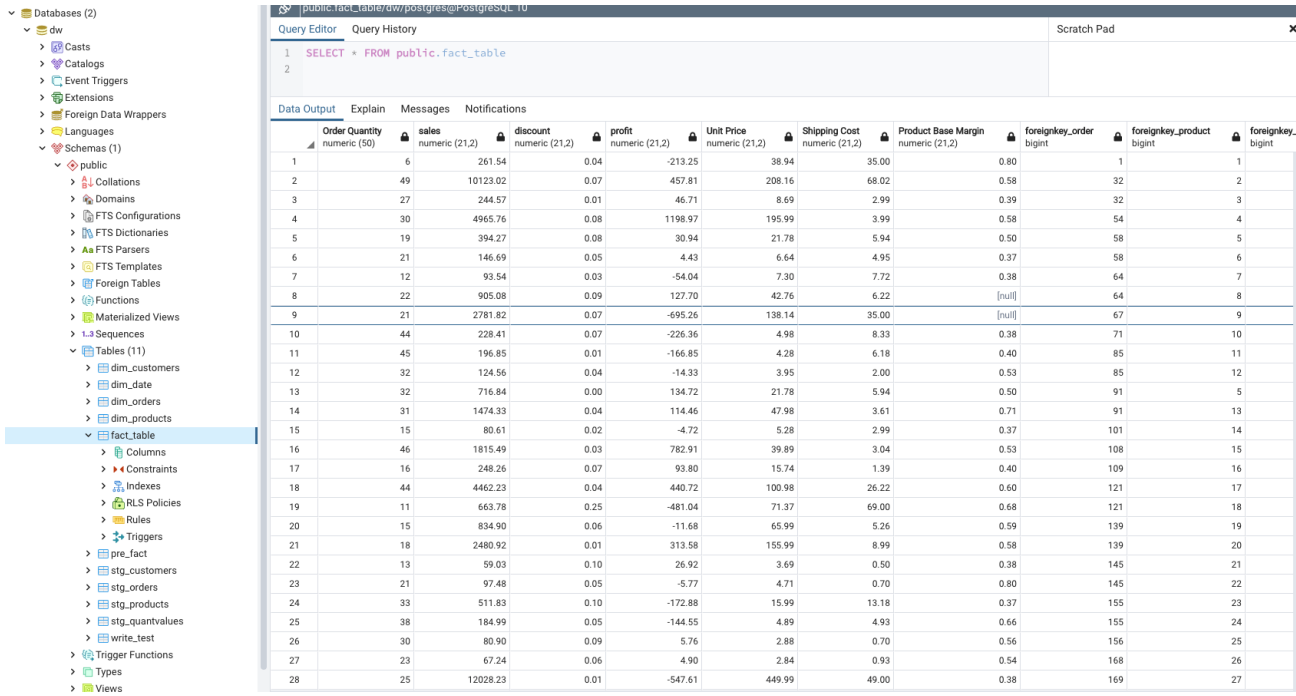
fact_table.ktr

Clear the **cache of our system**

cda_clear_cache.ktr

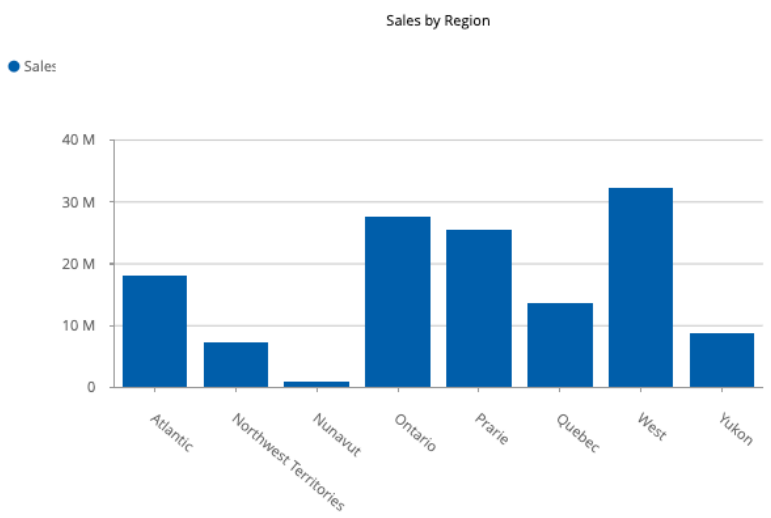
Example create a job – creating a fact table





	Order Quantity numeric (50)	sales numeric (21,2)	discount numeric (21,2)	profit numeric (21,2)	Unit Price numeric (21,2)	Shipping Cost numeric (21,2)	Product Base Margin numeric (21,2)	foreignkey_order bigint	foreignkey_product bigint	foreignkey_ bigint
1	6	261.54	0.04	-213.25	38.94	35.00	0.80	1	1	
2	49	10123.02	0.07	457.81	208.16	68.02	0.58	32	2	
3	27	244.57	0.01	46.71	8.69	2.99	0.39	32	3	
4	30	4965.76	0.08	1198.97	195.99	3.99	0.58	54	4	
5	19	394.27	0.08	30.94	21.78	5.94	0.50	58	5	
6	21	146.69	0.05	4.43	6.64	4.95	0.37	58	6	
7	12	93.54	0.03	-54.04	7.30	7.72	0.38	64	7	
8	22	905.08	0.09	127.70	42.76	6.22	[null]	64	8	
9	21	2781.82	0.07	-695.26	138.14	35.00	[null]	67	9	
10	44	228.41	0.07	-226.36	4.98	8.33	0.38	71	10	
11	45	196.85	0.01	-166.85	4.28	6.18	0.40	85	11	
12	32	124.56	0.04	-14.33	3.95	2.00	0.53	85	12	
13	32	716.84	0.00	134.72	21.78	5.94	0.50	91	5	
14	31	1474.33	0.04	114.46	47.98	3.61	0.71	91	13	
15	15	80.61	0.02	-4.72	5.28	2.99	0.37	101	14	
16	46	1815.49	0.03	782.91	39.89	3.04	0.53	108	15	
17	16	248.26	0.07	93.80	15.74	1.39	0.40	109	16	
18	44	4462.23	0.04	440.72	100.98	26.22	0.60	121	17	
19	11	663.78	0.25	-481.04	71.37	69.00	0.68	121	18	
20	15	834.90	0.06	-11.68	65.99	5.26	0.59	139	19	
21	18	2480.92	0.01	313.58	155.99	8.99	0.58	139	20	
22	13	59.03	0.10	26.92	3.69	0.50	0.38	145	21	
23	21	97.48	0.05	-5.77	4.71	0.70	0.80	145	22	
24	33	511.83	0.10	-172.88	15.99	13.18	0.37	155	23	
25	38	184.99	0.05	-144.55	4.89	4.93	0.66	155	24	
26	30	80.90	0.09	5.76	2.88	0.70	0.56	156	25	
27	23	67.24	0.06	4.90	2.84	0.93	0.54	168	26	
28	25	12028.23	0.01	-547.61	449.99	49.00	0.38	169	27	

Mashboard example



Show 10 entries Search:

Products.Category	Sales	Discount	Profit
Office Machines	19518274.26	153.18	2769416.37
Telephones and Communication	17003825.28	381.06	2852564.58
Copiers and Fax	10173251.7	40.32	1506253.41

Show 10 entries Search:

Priority	Sales	Discount	Profit
Low	29445293.16	775.44	3401532.54
High	29400466.41	785.7	3663224.1
Medium	25762068.9	736.92	2943465.57
Not Specified	25005269.79	729.54	1967263.47
Critical	24627309.93	727.11	1720426.14

Showing 1 to 5 of 5 entries Previous 1 Next