Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

SUPER-X Data Warehousing Case Study

Procurement Department

Study Program

M.Sc. Business Intelligence and Process Management

Course

581091 Data Warehousing

Academic Supervisor

Prof. Dr. Roland Müller

Authors

Christian Kregelin 334983

s_kregelin@stud.hwr-berlin.de

Pranav Pandya 552590

s_pandya@stud.hwr-berlin.de

Philipp Krank 537528

s_krank@stud.hwr-berlin.de

Date of Submission

February 09, 2017

Table of Content

1. BUSINESS REQUIREMENTS & KPIS

The very first step in designing a meaningful data warehouse is to identify the business requirements. Super-X's main objective is to analyze the performance of its business processes in the procurement department. Therefore, our team gathered 10 business requirements which specifically address these subjects. Based on these business requirements, Key Performance Indicators (KPIs) were derived. Furthermore, a validation of the proposed data warehouse solution to answer these business requirements is given in chapter seven.

Business Requirement Gathering

No.	***Business Requirement (Question)***	***Importance***	***High Level Entities***	***Measures***
1	What is the quantity ordered for a particular period?	High	• Purchase_order_items	• Quantity
2	What is the amount ordered for a specific period?	High	• Purchase_order_items	• Amount
3	Add supplier rating (on time delivery, delivery reliability)	High	• Supplier • Deliveries • Purchase_order	• On-time Delivery (OTD) = Successful Delivery / Total Deliveries * 100
4	What's the delivery default rate of the suppliers?	High	• Supplier • Deliveries • Purchase_order	• Delivery? Yes / No

5	How many suppliers are unreliable / underperforming?	High	• Supplier
			• Deliveries
			• Purchase_order
				• OTD ABC Rating
				  ○ A = (Delivery Default Rate = 0) AND 95% OTD Rate)
				  ○ B = (Delivery Default Rate = 0) AND >= 90%
				  ○ C = (Delivery Default Rate 0 - 10) AND < 85%
				  ○ D = (Delivery Default Rate > 10%) OR < 85%

6	Material Cost Ranking	High	• Materials • Purchase_order_items	• Comparison of unit prices per currency
7	What is the delivery volatility? (Constant, seasonal, one-time, etc.?)	High	• Purchase_order_items (→ timestamp)	• Quantity of ordered materials over time
8	How flexible is the supplier (rate of accepted change requests)?	High	• Change_requests • Supplier	• Flexibility Rating = Accepted Change Request / Total Amount of Change Requests * 100
9	How did the prices for materials develop over time?	High	• Purchase_order_items • Material • Supplier • Purchase_Orders	• Compound Annual Growth Rate (CAGR)

10	What is the material price among our suppliers?	High	• Supplier • Purchase_order_items • Material • Purchase_Orders	• Comparison

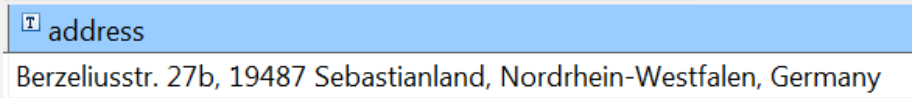
Table 1 - Business Requirements

2. RELEVANT DATA SOURCES & DATA QUALITY

Each table of the OLTP database was analyzed and classified to identify the relevance for the procurement department. First, the structure and logic of each table was analyzed and evaluated. Additionally, data quality issues were investigated.

2.1 OLTP Database Structure & Logic

In total, 31 structural and logical issues were identified (3 issues in the provided CSV files). Critical issues are marked in red (8) and non-critical issues in black (23). The following table shows the results.

Table	*Issue*	*Solution*	*Comment*
		Critical Issues	
Purchase_order_item	No status column	Create status column (useful for partial delivery)	
deliveries	Columns "supplier_id" and "employee_id" are redundant, as they already exist in the table "purchase_order". The link to "purchase_order" is missing.	Delete columns "supplier_id" and "employee_id" and create a FK to "purchase_order"	There might be more than one delivery per purchase order
general	The OLTP database is not fully normalized	Normalize the OLTP database	
	 *Figure 1 - OLTP Database Normalization Issue*		
suppliers	Column "category" is manually entered	Calculate category on-the-fly	Category might change because of changed contracts

suppliers	There is no supplier ranking	Implement supplier ranking (calculation on-the-fly)	Ranking probably changes many times during the year.
suppliers	There are a couple of suppliers mapped to POs which contain materials of the types semi-finished as well finished goods which who shall be declared as retailers	Declare those suppliers as retailers own or as both retailers and suppliers.	
contract_notes	Columns "purchase_contract_id", "material_id" and "supplier_id" are not needed for identification	Replace the three columns "purchase_contract_id", "material_id" and "supplier_id" with a FK to "purchase_contract_items id" Add FK to "delivery_item_id", in order to refer to a to-be-delivered item directly	
contract_notes	Contract notes are not linked to purchase_order and/or deliveries	Create a FK to "purchase_order" and/or "deliveries"	

Table 2 - OLTP Database Critical Issues

Table	*Issue*	*Solution*	*Comment*
		Non-Critical Issues	
bill_of_materials	Column „id" is useless/confusing	Delete "id" and make the columns "part_id" & parent_id a combined PK	If there is a new product or the quantity of single materials for a product changes, a new material id and parent id must be created.
Materials	Column "id" and "ean" are unique identifiers	Delete "id" and use "ean", only. This is also easier for the supplier and retailer communicatio n.	Even semi-finished products which are never publicly seen, need an EAN number
Materials	Typo line 30 > description > "delux(E)"	"Deluxe"	
Purchase_order_item	Delivery date is missing	Add column "delivery date"	
delivery_items	There is no status column	Optional: create status columns e.g. to track material quality	
delivery_items	Delivery date is missing	Add column "delivery date"	
change_request	"Material_id" is redundant, as there is "purchase_contract_ite m_id"	Delete "material_id"	
purchase_contract_ite ms	"state" and "contract type" are redundant as they are in the table	Delete columns "state" and	If one item of the contract is changed, a new contract (id) has to be created. In addition,

	"purchase_contracts" already	"contract type"	the valid from - to information is stored in the table “contract”, only.
contract_notes	Column “note” is free-text.	Replace the *varchar* column with column “issue_id” (*integer*). As a result, a new table “issues”, that describes the issue cases (e.g. 1 = quality issue, 2 = no delivery, 3 = partial delivery, 4 = delayed delivery etc.) needs to be created	
monthly_material_demands	Typo column "timestampe"	“timestamp”	
monthly_material_demands *[for OEM only]*	Timestamp is not filled out. There is no status column (e.g. to track, whether the forecast has been already adjusted via the sales figures/retailer orders for next month)	Create column “status” to indicate, whether the demand has already been updated Optional: delete timestamp	Alert (no update after 7th working day) and Dashboard (how much is updated already; what is the difference in quantity: annually forecast vs. monthly figures on the fly “forecast” vs. “monthly_material_demands”)
monthly_material_supplies	Typo column "timestampe"	“timestamp”	
monthly_material_supplies	Timestamp is not filled out. There is no status	Create column “status” to	Alert (no update after 15th working day) and

	column (e.g. to track, whether the forecast has been already adjusted via the sales figures/retailer orders for next month).	indicate, whether the demand has already been updated. Optional: delete timestamp	Dashboard (how much is updated already; what is the difference in quantity annually forecast vs. monthly figures on the fly "forecast" vs. "monthly_material_supplies")
suppliers	Column "contact_person" Some contact persons are missing	NOT NULL rule	
suppliers	Column "contact_person": names, titles and gender are mixed	Split up to "names", "title" and "gender"	
suppliers	Phone number is not in a specific format	Use standard format	
suppliers	Fax number is not in a specific format	Use standard format	
suppliers	Email address is not in a specific format	Use standard format	
suppliers	Duns_number is not in a specific format	Use standard format	
event_logs		Since this field is filled out manually, either drop it or atomize it	
CSV	Material description is used	Use EAN material number	New suppliers cannot identify the material by its description only. Therefore, an EAN number is more useful.
CSV	Delivery date is missing	Create column "delivery date"	
CSV	No reference to "contract_item_id"	Optional: add column "contract_item_id"	Helpful, as the new quantity might be in the tolerance of our contract.

change_request	There is a table for change requests and purchase orders.	Add column "change_request" to the table "purchase order" and drop table change_request.

Table 3 - OLTP Database Non-Critical Issues

2.2 OLTP Data Quality

To check for further data quality issues of each table, TALEND was used. The following issues were identified.

2.2.1 Email addresses

The email addresses did not match the standard pattern of an email in around 29% of all cases.

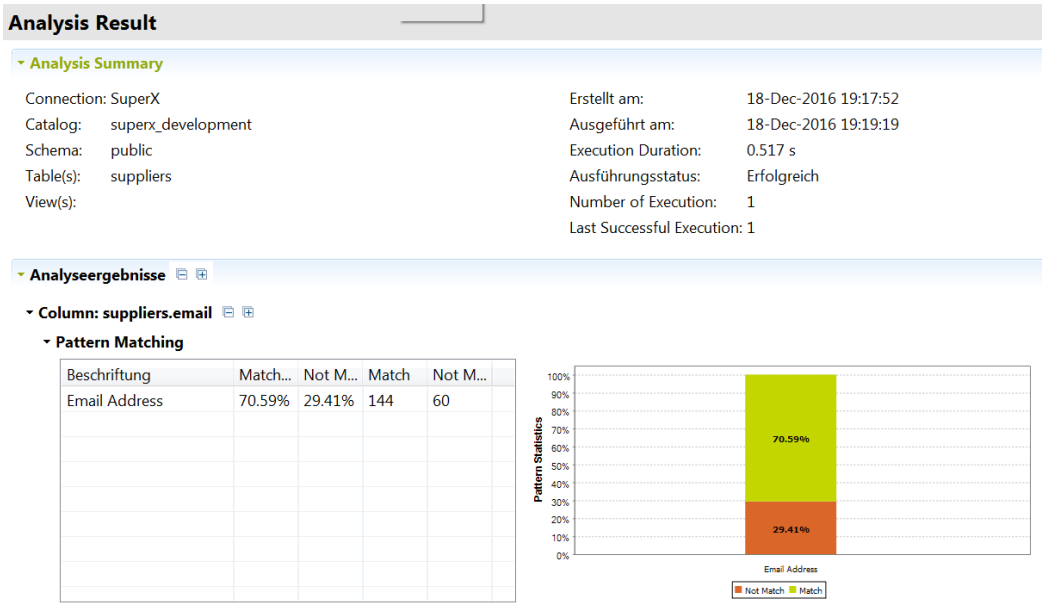


Figure 2 - Data Quality: Email Addresses

Below is an example of the variance of the email addresses.

email
palmieri_ausonio@barone. t
brian.s hmidt@schnherr.de
bre tenberg_vernie@braunmcdermott.ca
gabriela_valle_sra_cuellar@gr ego.info
m_renault_clara@le.inf
miranda.a.ochoa.cr stina.mar@escobedo.info
de_santis_clea@ben dettiroetti.it
eunice.i.l ettgen@langosh.net
tamino.ahren erg@brner.de
demetris_lo e@ko.us

Figure 3 - Data Quality: Email Addresses Variance

2.2.2 Missing Contact Persons

In 16 cases, a contact person is missing.

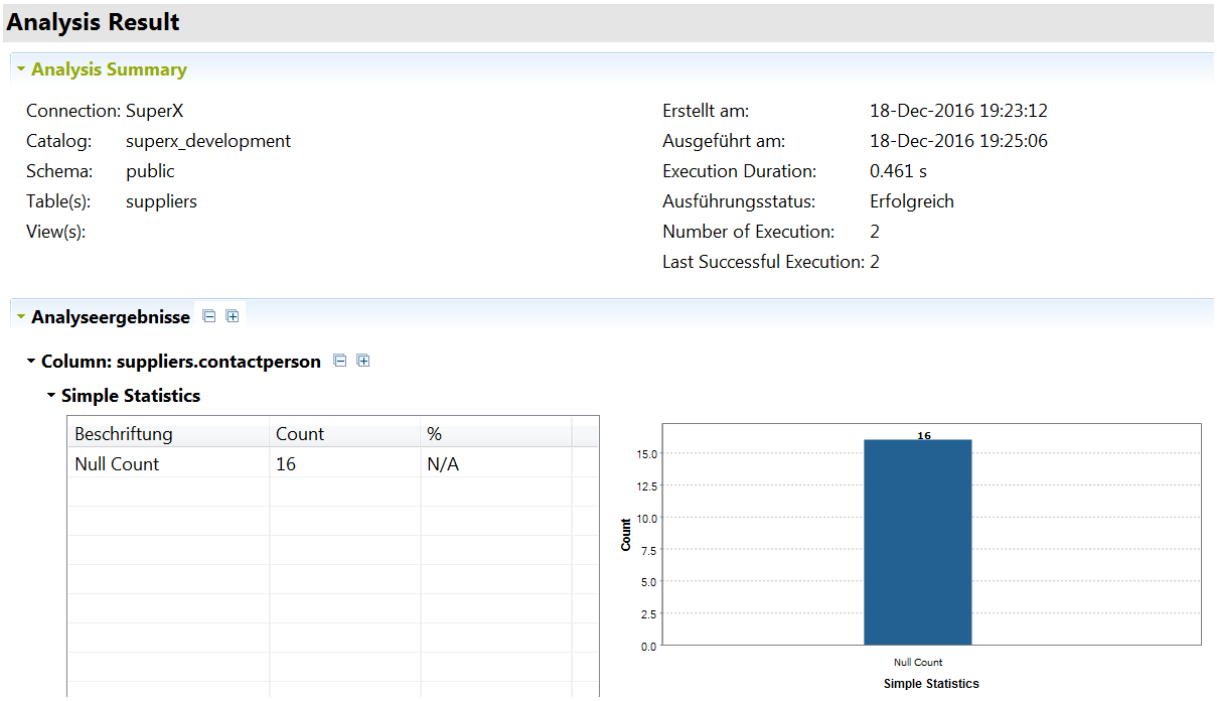


Figure 4 - Data Quality: Missing Contact Persons

2.2.3 Phone Number Patterns

The patterns of the phone numbers showed a high variance

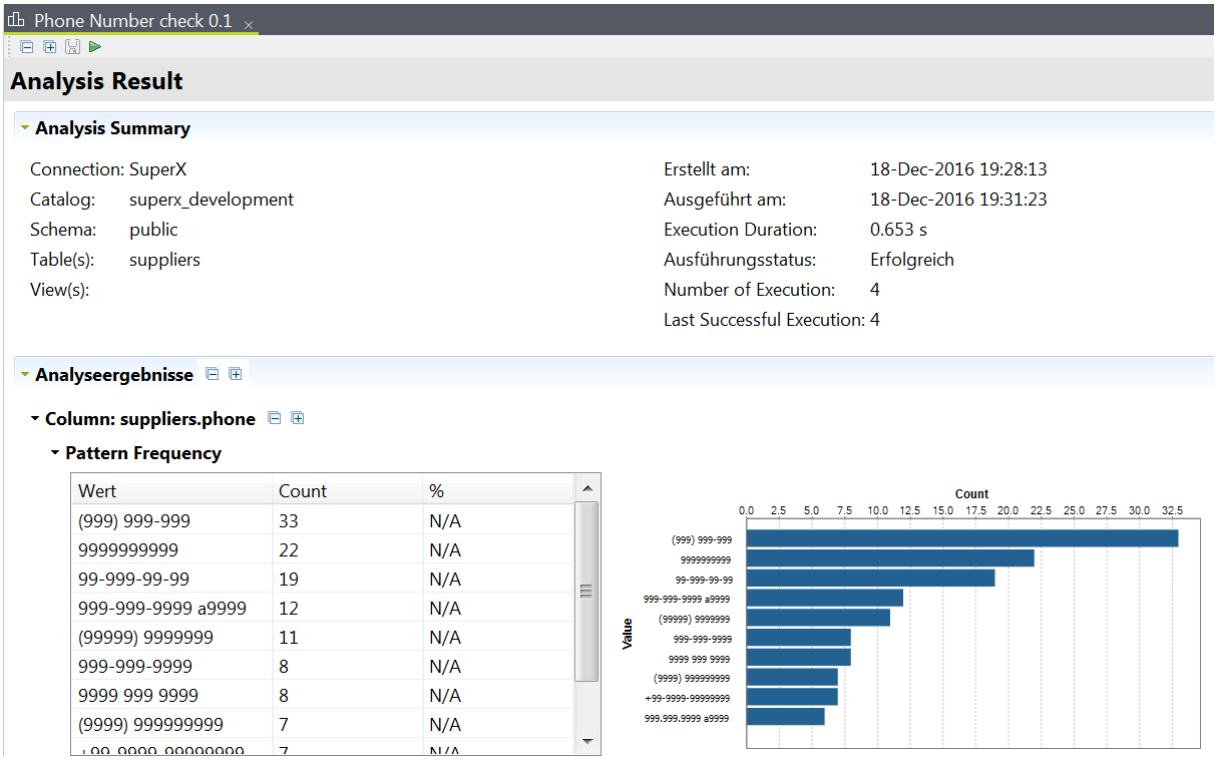


Figure 5 - Data Quality: Phone Number Variance

3. MULTI DIMENSIONAL DESIGN

Two cubes were designed to address the business requirements. In the following, the conceptual design of the cubes is outlined.

3.1 M/ER Diagram Purchase Orders

Our team developed the "Purchase Orders" cube, which answers the majority of business requirement questions. In order to address these requirements, the three dimensions "Time", "Supplier" and "Purchase Order Item" were added.

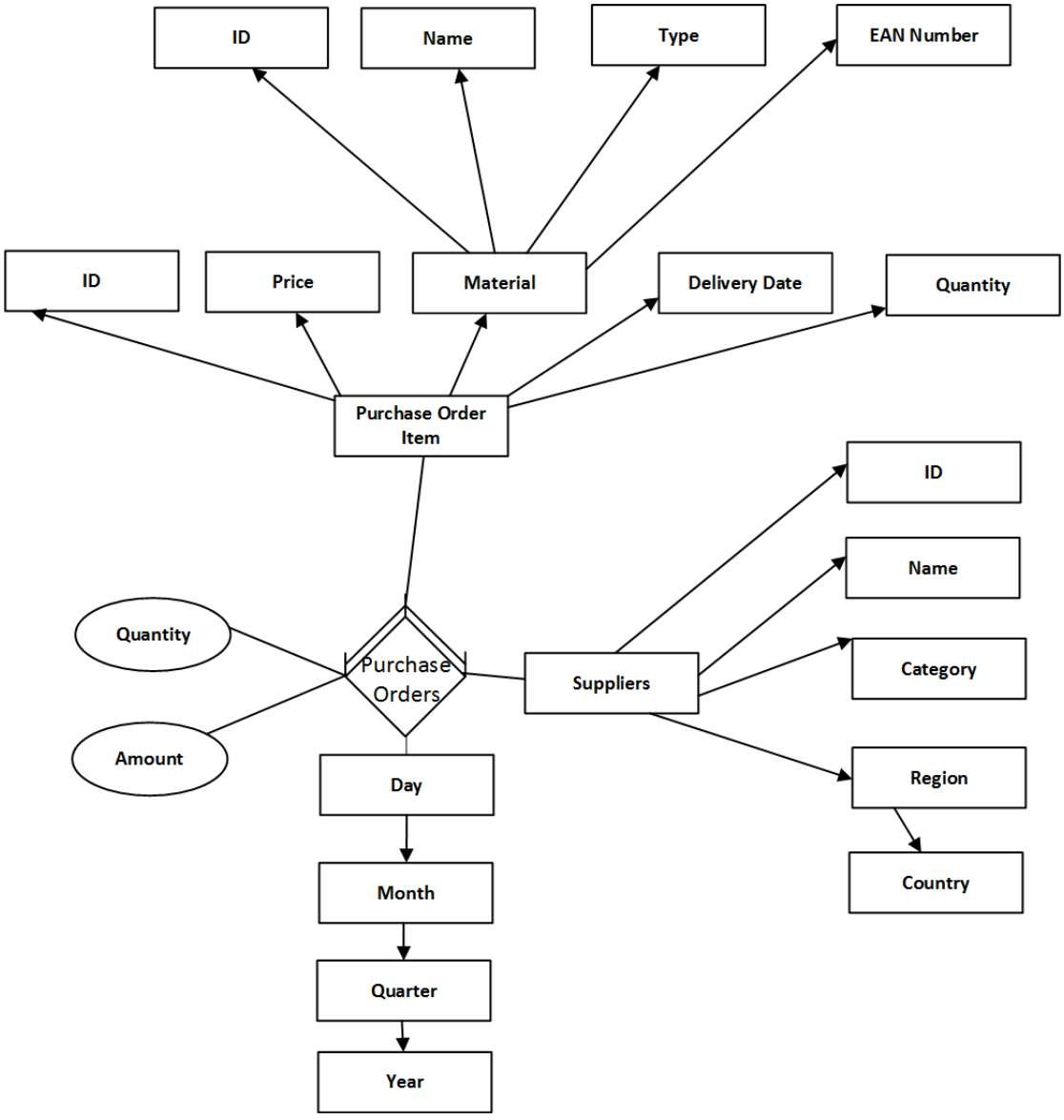


Figure 6 - M/ER Diagram: Purchase Orders

3.2 M/ER Diagram Contract Note

A second "Contract Note" cube was designed to specifically answer business requirement No. 5. In this case, the dimensions "Material", "Suppliers" and "Year" were created.

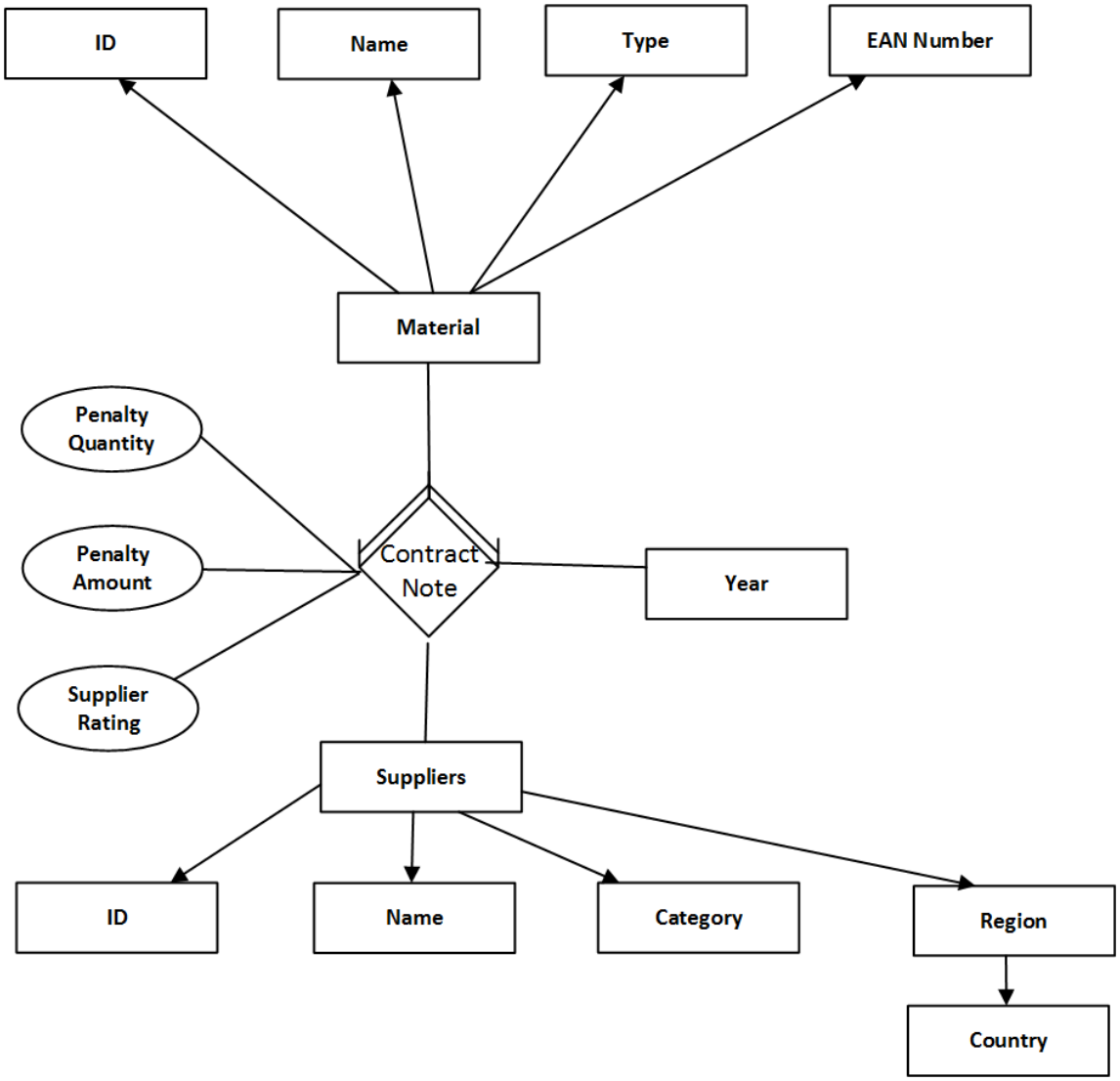


Figure 7 - M/ER Diagram: Contract Note

4. PROOF-OF-CONCEPT IMPLEMENTATIONS

4.A SAP HANA

Due to the design of SAP HANA (ETL Process before Multidimensional Implementation), we decided to switch the order of the following two chapters (Multidimensional Implementation & ETL Process).

4.A.1 ETL Process

In order to provide clean and standardized data for our cubes and dimensions, the very first step was to perform the extraction and transformation part of Extract, Transfom and Load (ETL) process in the SAP HANA Workbench Editor. Below is the step by step procedure for this process.

Remote Source

First, we connected the "Super_X_development" database from PostgreSQL to SAP HANA. Below the tables which are connected from the source to SAP HANA are illustrated.

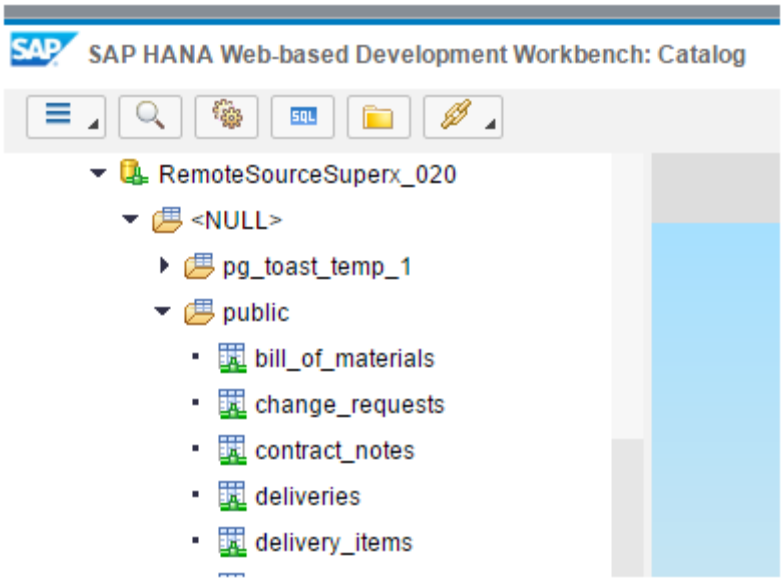


Figure 8 - Tables Extracted from the Remote Source

Flowgraphs

Within the flowgraph process, we have used virtual tables from the remote source for the initial transformation to detect anomalies, clean data and geocoding. The picture shown below illustrates steps and filters used for data cleaning and geocoding.

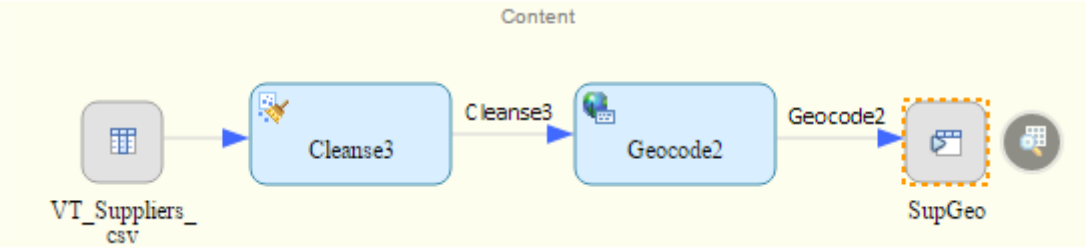


Figure 9 - Use of Data Cleaning and Geocoding Filters in the Flowgraph Process

Here we have used the "Cleanse" filter to deal with data quality issues. The filter identified, parsed and formatted any field that contains information, such as an address, person name (in our case, supplier contact person), organization name (in our case, supplier name), occupational title, phone number or email address.

Geocoding

We have split the supplier's address to extract the first line of the address, region and country. This helped us to geocode suppliers in order to visualize it in a map. Below is the output of a few columns in the suppliers table. The geocoding separates the region and country by latitude and longitude.

204 row(s)

STD_FIRM	STD_ADDR_R...	STD_ADDR_C...	STD_ADDR_C...
Döring, Bloch und Klem	Bayern	Deutschland	DE
Leeuwen V.O.F.	Noord-Holland	Netherlands	NL
Lombardi e Figli	Reggio Calabria	Italy	IT
Moulin SAS	Languedoc-Roussillon-	France	FR
Noel Scop	Pays de la Loire	France	FR
Pawlicki-Rosa	Świętokrzyskie	Poland	PL

Figure 10 - Output of Supplier's Address after Geocoding Filter

Instead of using a sink tables, we used template tables which allowed us to extract all the columns (i.e. geocoded columns from suppliers' address) which were transformed and could be useful for the front-end (Lumira).

Below is a list of all our flowgraphs, dimensions and cubes that we created in the SAP HANA Workbench Editor under the "020_SuperX_new" package. The first two cubes (CUBE_CON_PENALTIES and CUBE_PO_AmtQtyMatSup) were used for the front-end (SAP Lumira and SAP Expert Analytics).

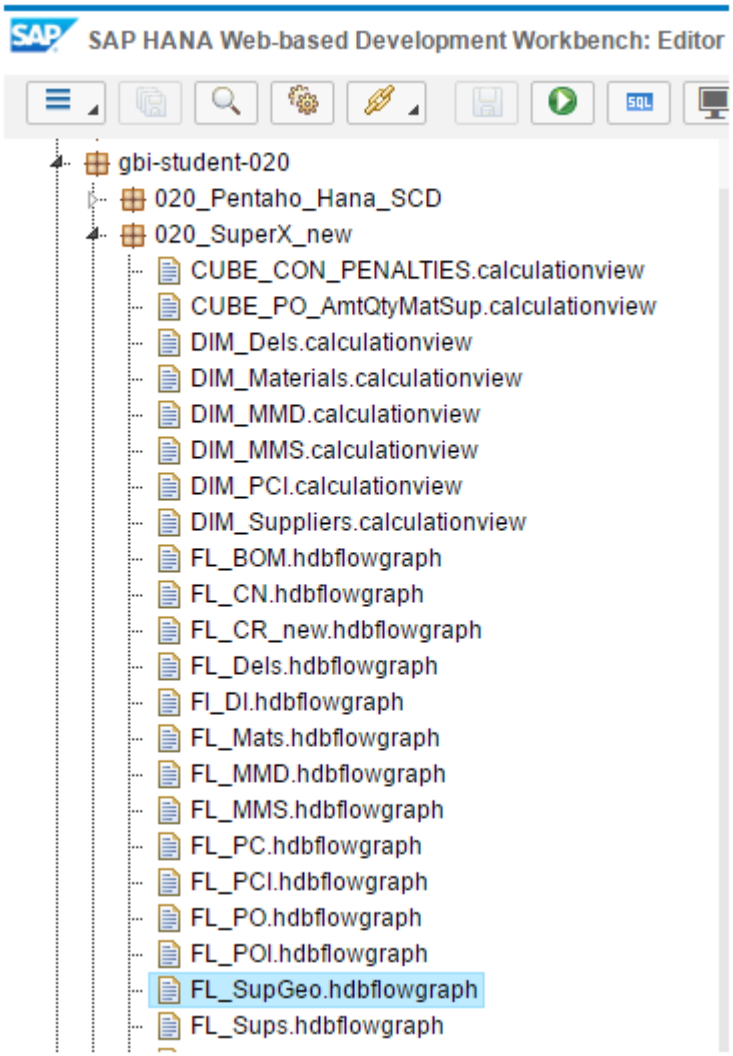


Figure 11 - List of Cubes, Dimensions and Flowgraphs in HANA Editor (under 020_SuperX_new package)

The illustration below shows the tables saved in the catalog perspective, after the transformation from our flowgraphs.

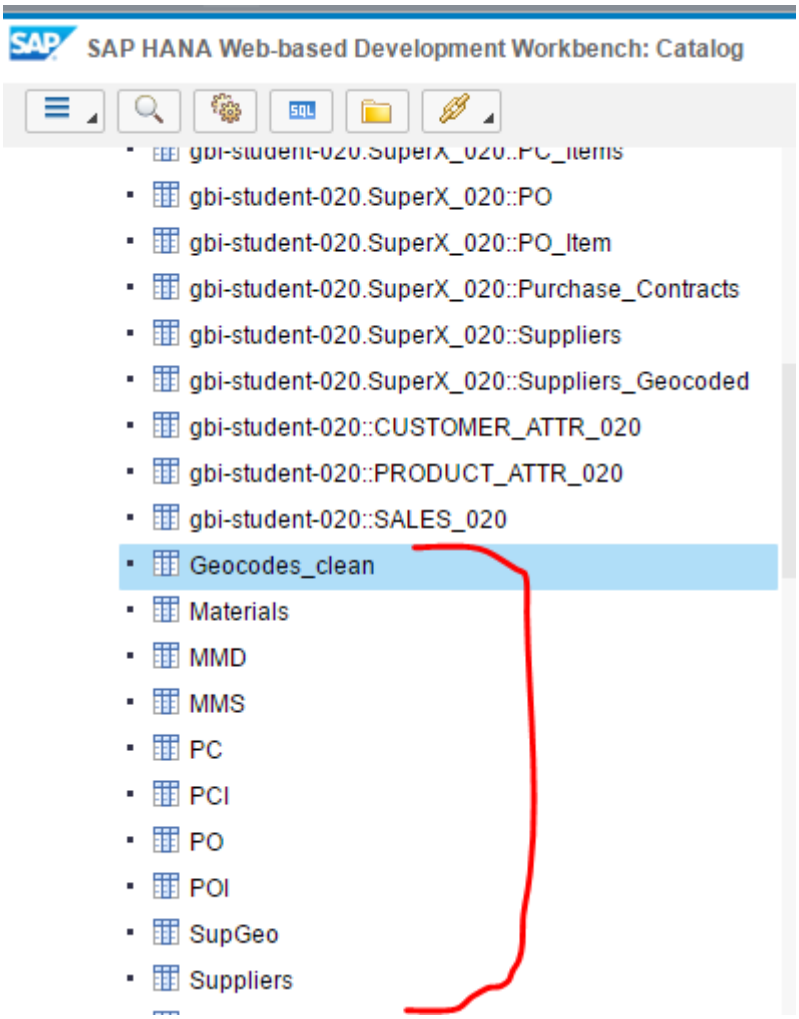


Figure 12 - Tables saved in the Catalog Perspective after the Transformation

Once the tables were ready after the flowgraph procedure, we identified relevant cubes and dimension tables which will be further explained in chapter 4.A.3 Multidimensional Implementation.

4.A.2 ELT Process

By using the Flowgraph Filter and Expressions Editors for the calculated columns, we had limited success for data cleaning. Therefore, we explored additional functionalities of SAP Lumira and Expert Analytics.

Tool: SAP Expert Analytics

After performing the ETL process in SAP HANA, we have used the ELT approach for further data cleaning in SAP Expert Analytics after loading the Cubes from SAP HANA. For example, the supplier category had typographical mistakes in the original data as well as some supplier addresses had Germany as a country and some of them had Deutschland.

The illustration below shows the data cleaning functions we have used in SAP Expert Analytics.

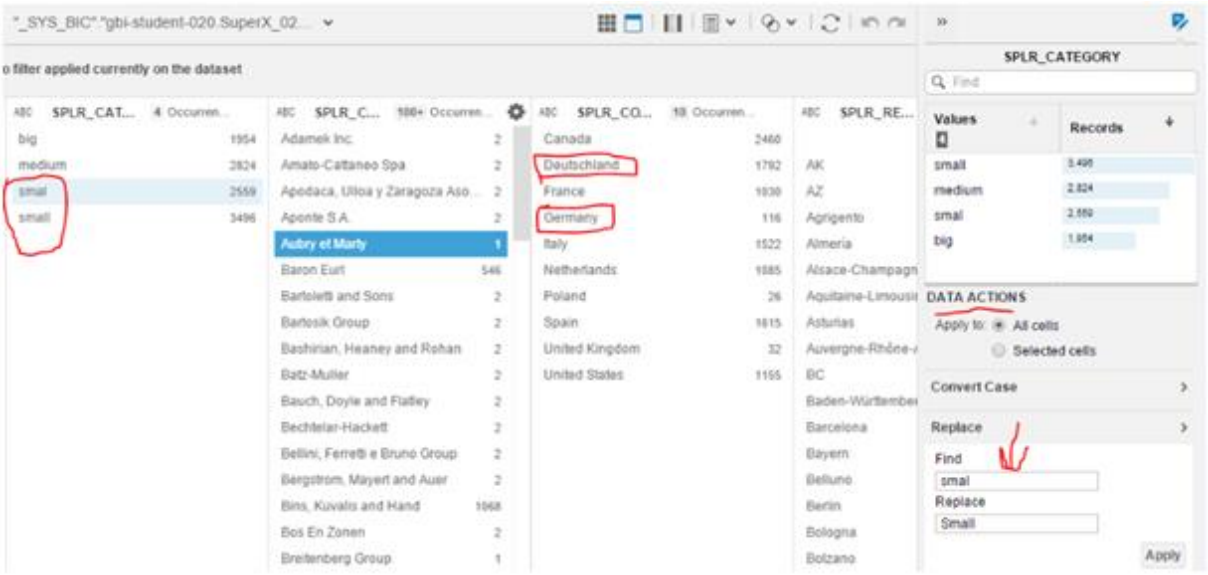


Figure 13 - Fixing Data Quality Issues in Expert Analytics (on the fly)

We found the ELT process relatively easy to use because we were able to identify anomalies in data when we tried to validate the results from the visualization of our cubes in the front-end. So, we were able to fix those issues in the "Prepare" tab in SAP Expert Analytics to visualize the correct and standardized data.

4.A.3 Multidimensional Implementation

Dimensions

Based on our conceptual design, we have implemented our multidimensional models in SAP HANA. We identified "Suppliers" and "Material" as our dimensions and "Purchase Order" and "Contract Penalty" as our main cubes. Afterwards, we joined the "supplier" table with geocoded elements to obtain the latitudes and longitudes which can be plotted on map in front-end (Lumira).

dim_supplier

In order to create the "Suppliers" dimension, we joined the suppliers table with "SupGeo" table (geocoded columns) which were created during the flowgraph process. This join was then used to create the projection.

The illustration below represents the procedure to create the "Suppliers" dimension.

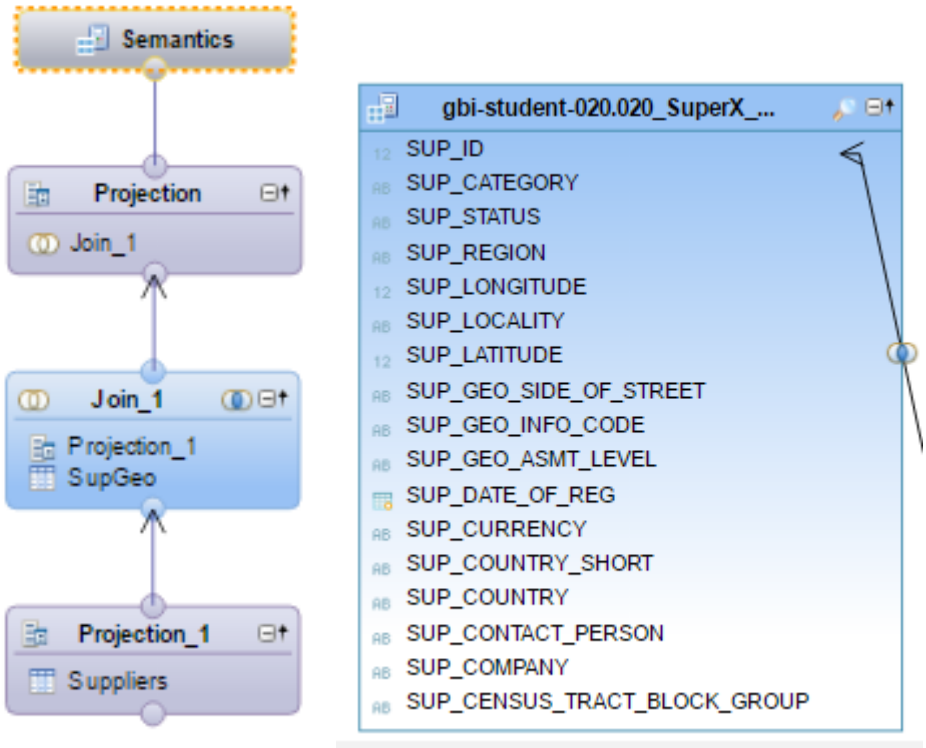


Figure 14 - Suppliers Dimension

dim_material

In order to create the "Material" dimension we simply used the virtual table and connected it to the template table as an output. The figure below shows the process.

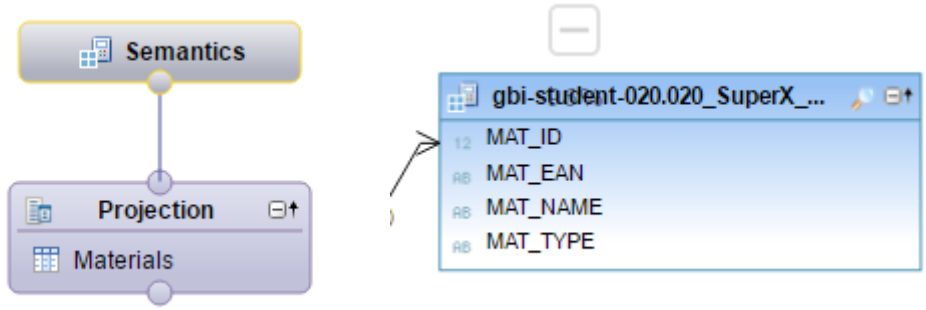


Figure 15 - Material Dimension

Cube 1: CUBE_PO_AmtQtyMatSup

In order to create this cube, first we joined the purchase_order and purchase_order_item tables so that we can connect our supplier and material dimensions. The illustration below shows the first join.

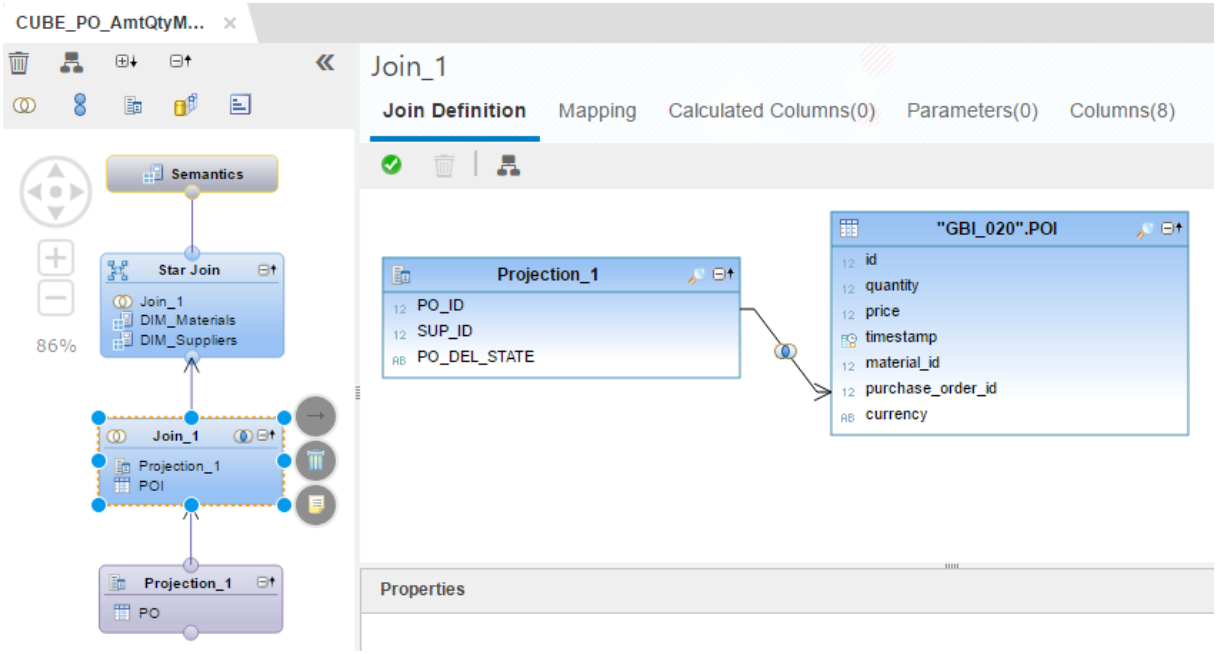


Figure 16 - 1st Cube: CUBE_PO_AmtQtyMatSup

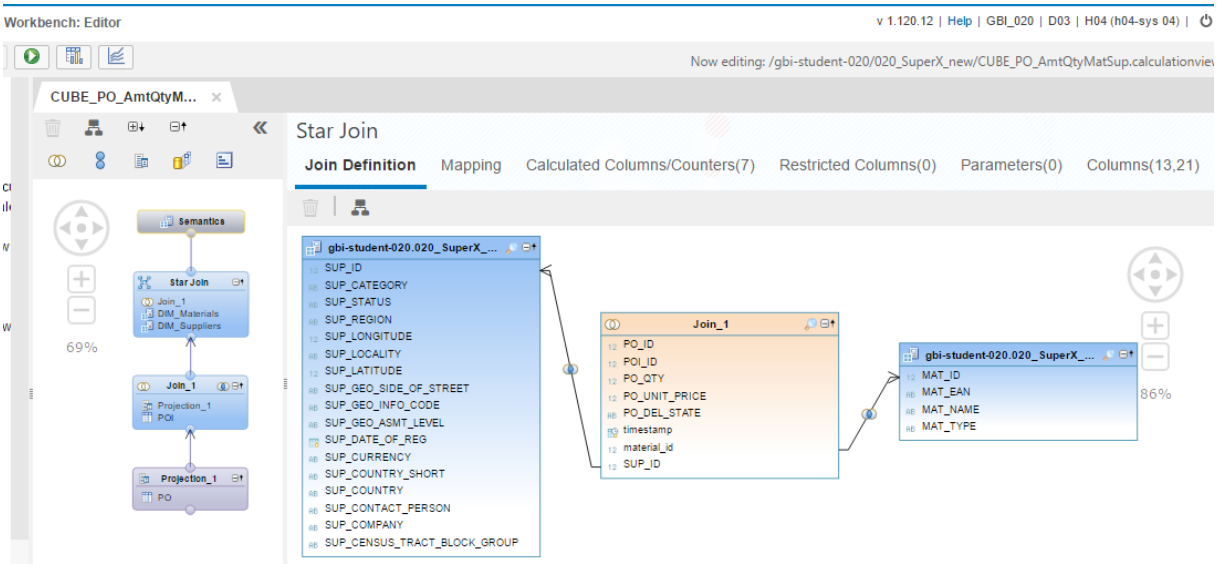


Figure 17 - Star Join for CUBE_PO_AmtQtyMatSup

Our cube uses a transaction type fact table and the granularity is purchase order id, material, quantity, unit price, order amount and supplier.

The illustration below shows the calculated columns/counters used in this cube where we used the unit price as a stock type measure, the quantity as a flow type measure and the order amount as a calculated measure (flow type).

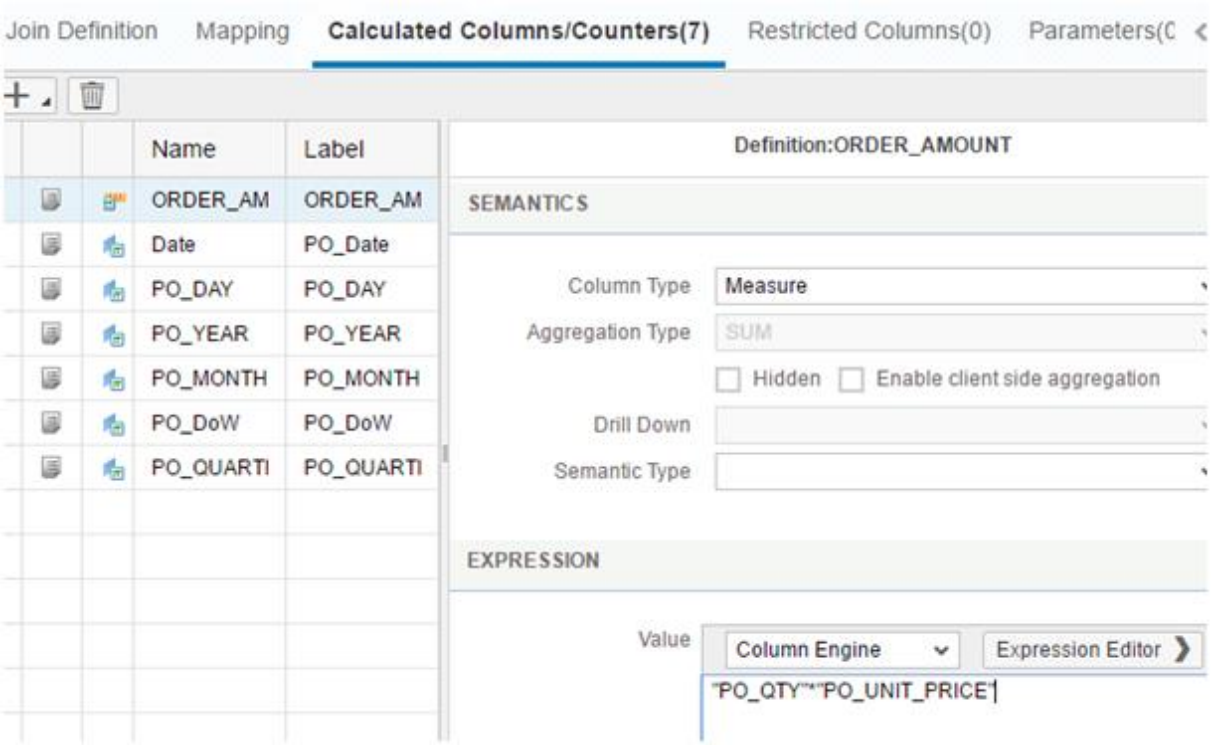


Figure 18 - Calculated Columns/Derived Measures for the 1st Cube

Cube 2: CUBE_CON_PENALTIES

This cube is related to contract notes where we used penalties as a measure. The picture below shows the star schema where we joined the "Contract Notes" table with the "Suppliers" and "Material" dimensions in order to evaluate the supplier performance by name of the material, material type and the supplier's geographical location.

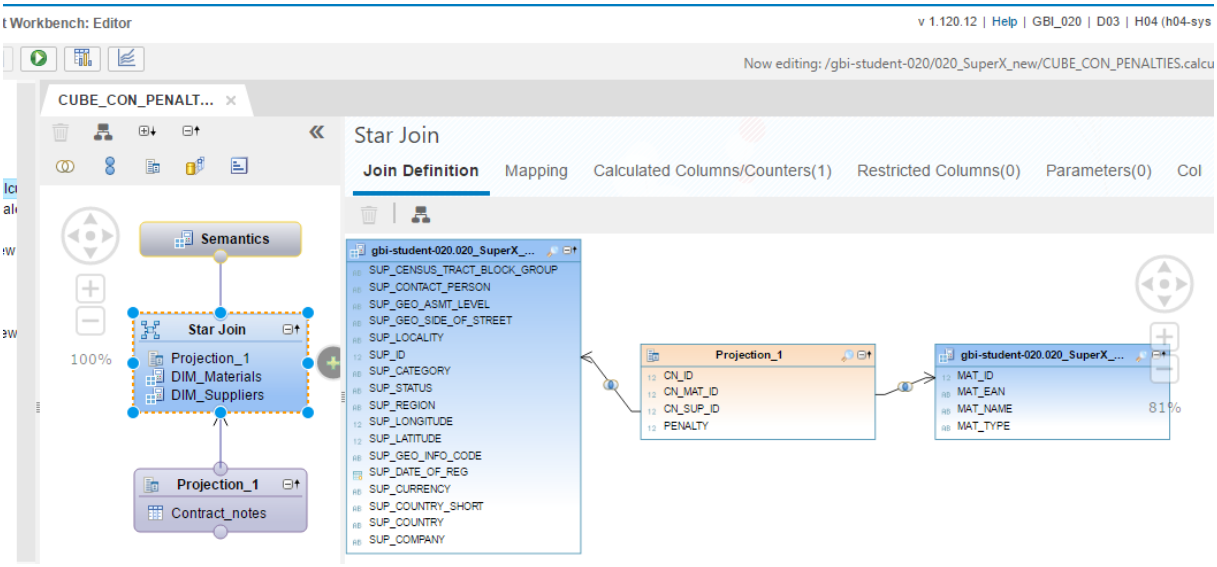


Figure 19 - Star Join for the 2nd Cube: CUBE_CON_PENALTIES

We used the Expression Editor to define penalties as a measure derived from the contract notes.

4.A.4 Implementation of a Front-End for our Data Mart

Tools: SAP Lumira & Expert Analytics

We used SAP Lumira and SAP Expert Analytics for our front-end implementation. Both cubes can be accessed from SAP Lumira and Expert Analytics under gbi-student-020.020_SuperX_new(2) user as shown in the illustration below.

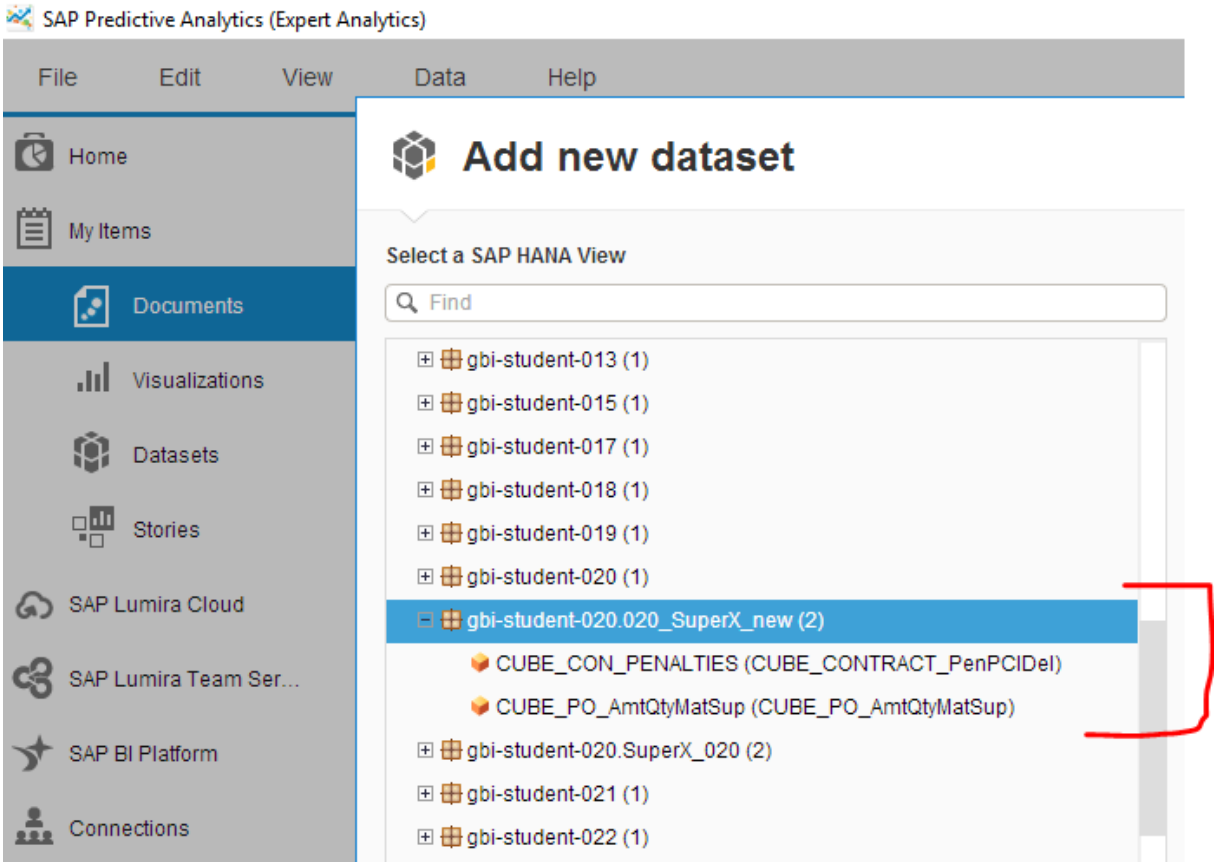


Figure 20 - Illustration of How and Where to Access our HANA Cubes in Lumira or Expert Analytics

Measures & Hierarchies

We used "Order Amount", "Quantity" and "Unit Price" as our measures and created a geographical hierarchy based on our geocoded latitude and longitude columns during the flowgraph process and time hierarchy to generate quarters as shown in the illustration below (in Expert Analytics).

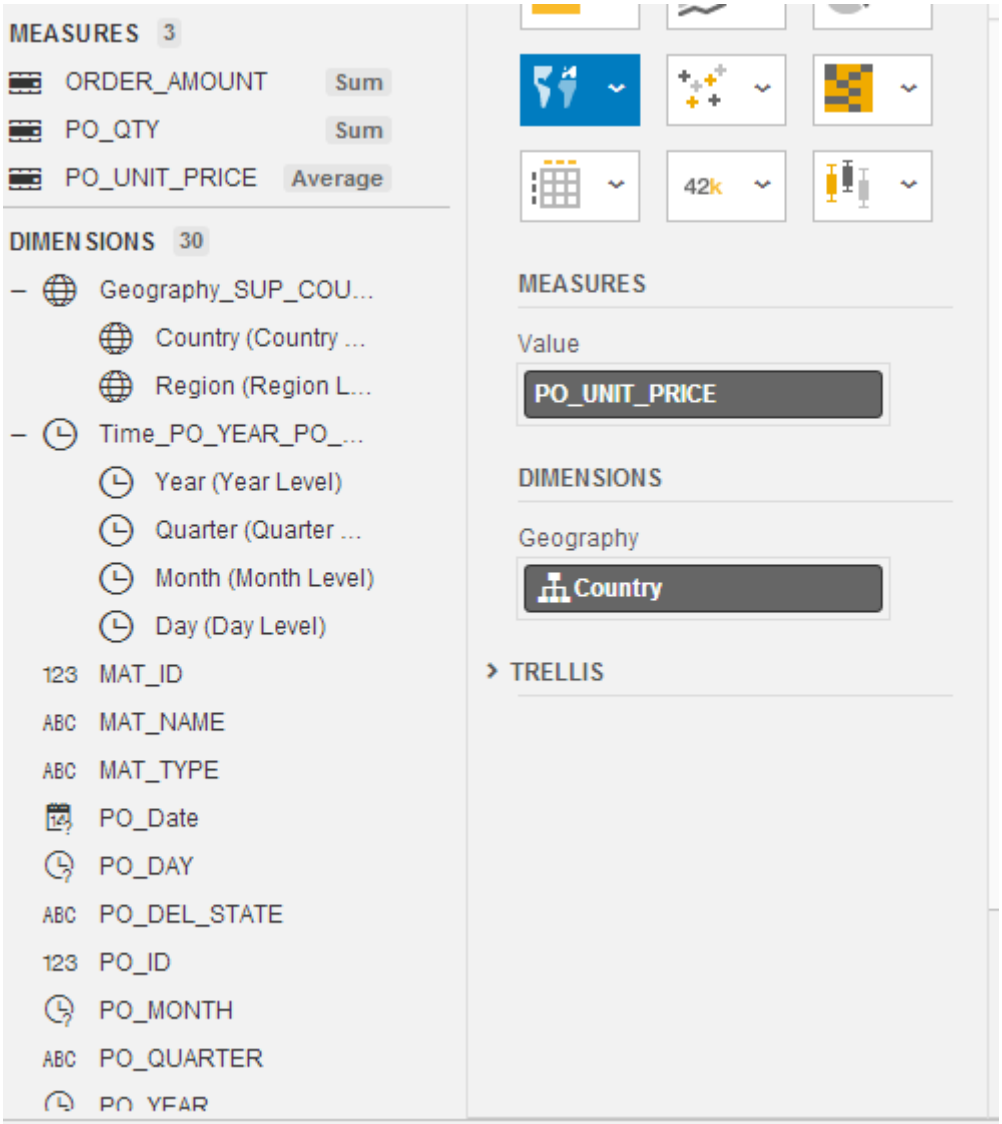


Figure 21 - Measures and Hierarchies created in Expert Analytics

Below is the illustration of our front-end implementation for the second cube (Contract Penalties) in SAP Lumira.

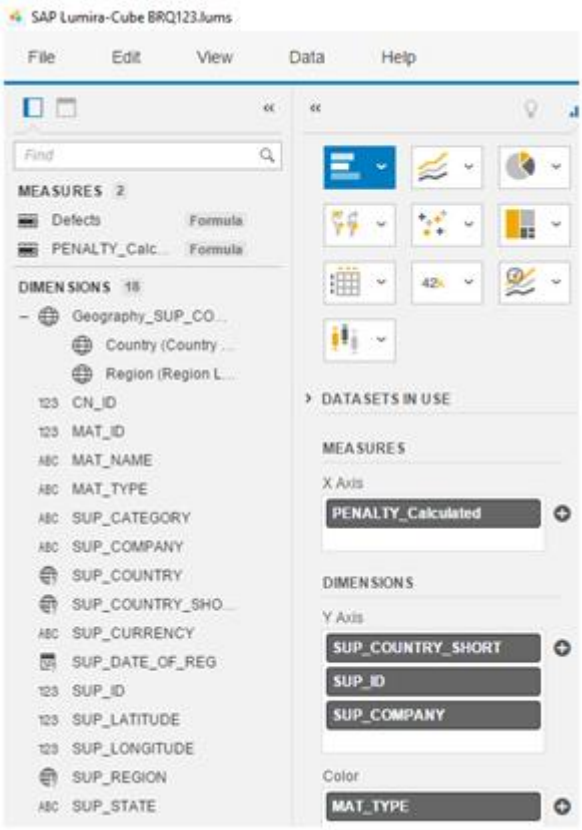


Figure 22 - Measures and Hierarchies for our Second Cube (Contract Penalty)

Visualizations

1. Variation in order quantities by years and European countries for all materials

   Countries in red indicate less ordered quantities and countries in green indicate a high amount of ordered quantities. It can be observed that France, the Netherlands and Germany had a moderate variation in ordered quantities, whereas Spain and Italy had a noticeable fluctuation between 2010-14.

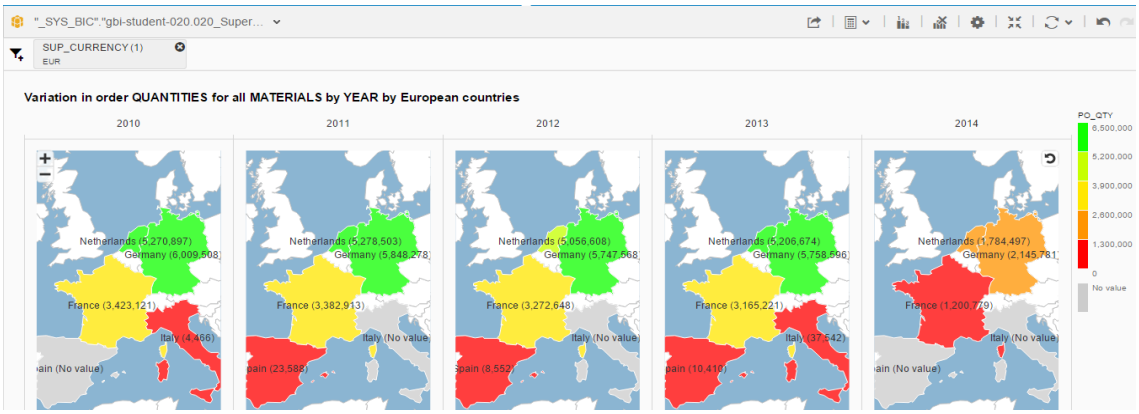


Figure 23 - Visualization in Expert Analytics (Variation in Order Quantities by Years and European Countries for all Materials)

2. Bubble Chart

The bubble chart below represents the fluctuation in order amount and quantities for raw materials by quarters and European countries. It can be observed that the plastic planes (blue bubble) have a much higher fluctuation compared to other raw materials.

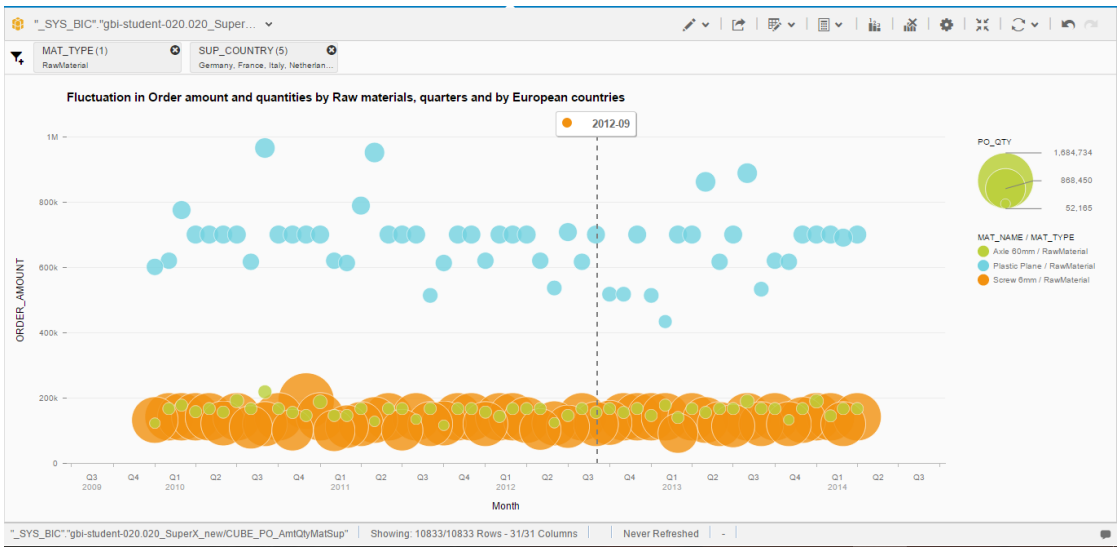


Figure 24 - Fluctuation in Order Amount for Raw Materials

The next bubble chart shows the fluctuation in order amount and quantities for OEM products by quarters and European countries. In this chart it can be seen that the Remote Controller 2-Channel 2MHz and Receiver 2-Channel 2MHz OEM products show a high fluctuation compared to other OEM products.

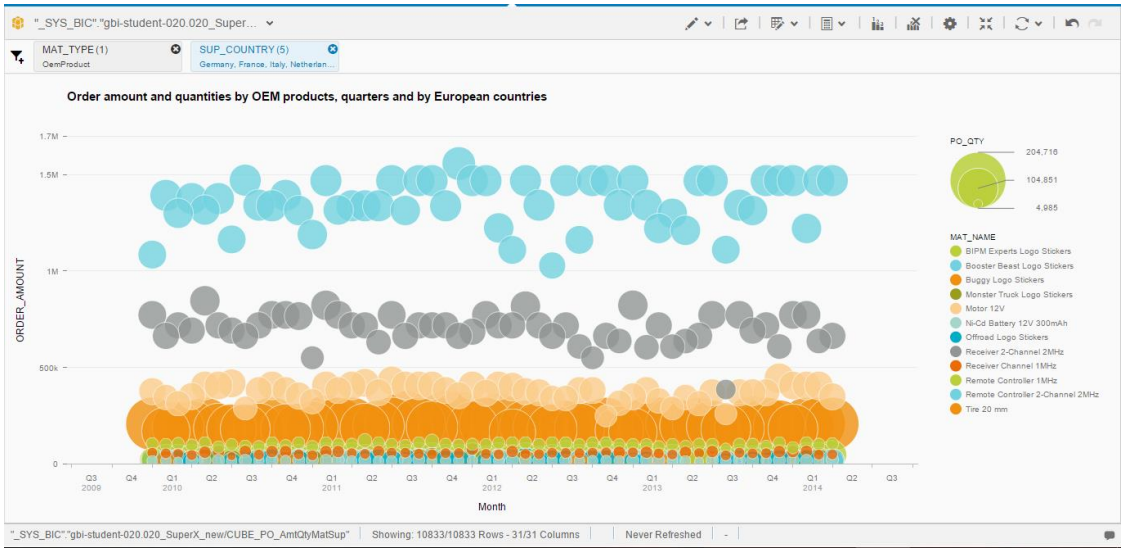


Figure 25 - Fluctuation in Order Amount for OEM Products

3. The next illustration provides high level overview of defects (penalties) received by the supplier till date by their name, country and categorized by material type. Dark green color means high penalties.

Figure 26 - Visualization on the Front-End (Supplier Evaluation)

We have provided more visualization to validate the business requirement questions (see chapter 7). Also, the cube can be accessed from SAP Lumira or Expert Analytics.

4.B Pentaho

4.B.1 Multidimensional Implementation

In order to visualize our intended multidimensional implementation, we created a Snowflake Schema, which displays the dimension tables “dim_supplier”, “dim_purchaseOrder”, “dim_material” and the fact table “fact_purchaseOrderItem” as well as their attributes. In addition, the cardinality is specified via Crow’s Foot notation.

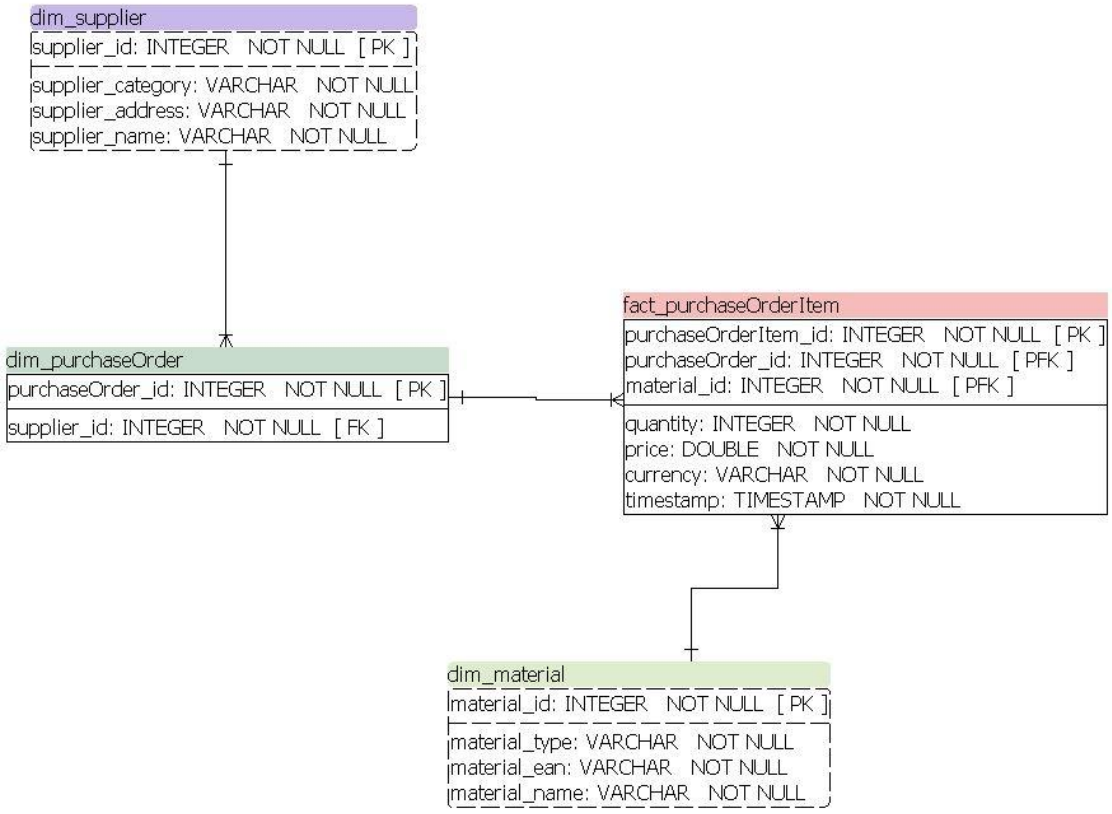


Figure 27 - Star Schema

4.B.2 ETL Process

In the following sections, the ETL process is described in detail.

Dimensions

dim_supplier

The following screenshots show the Pentaho Data Integration transformation step for dim_supplier:

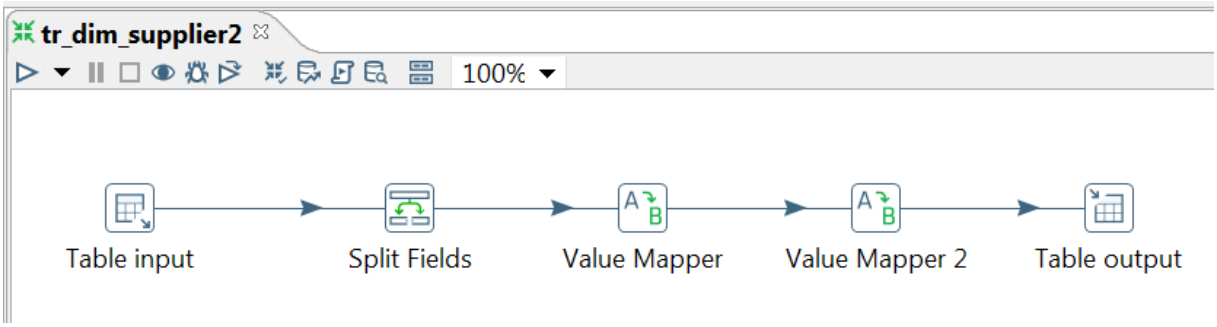


Figure 28 - Transformation for dim_supplier

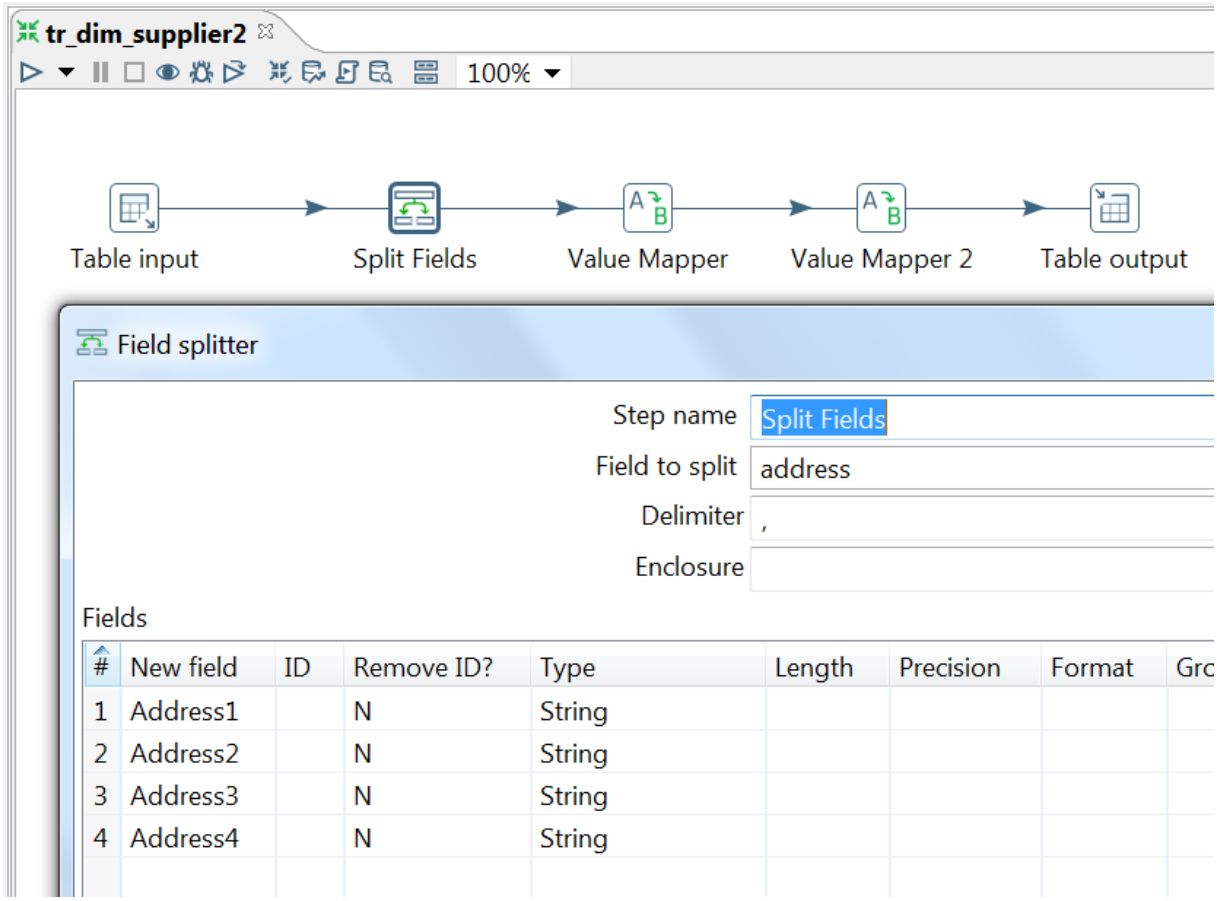


Figure 29 - Transformation for dim_supplier -> Split Fields

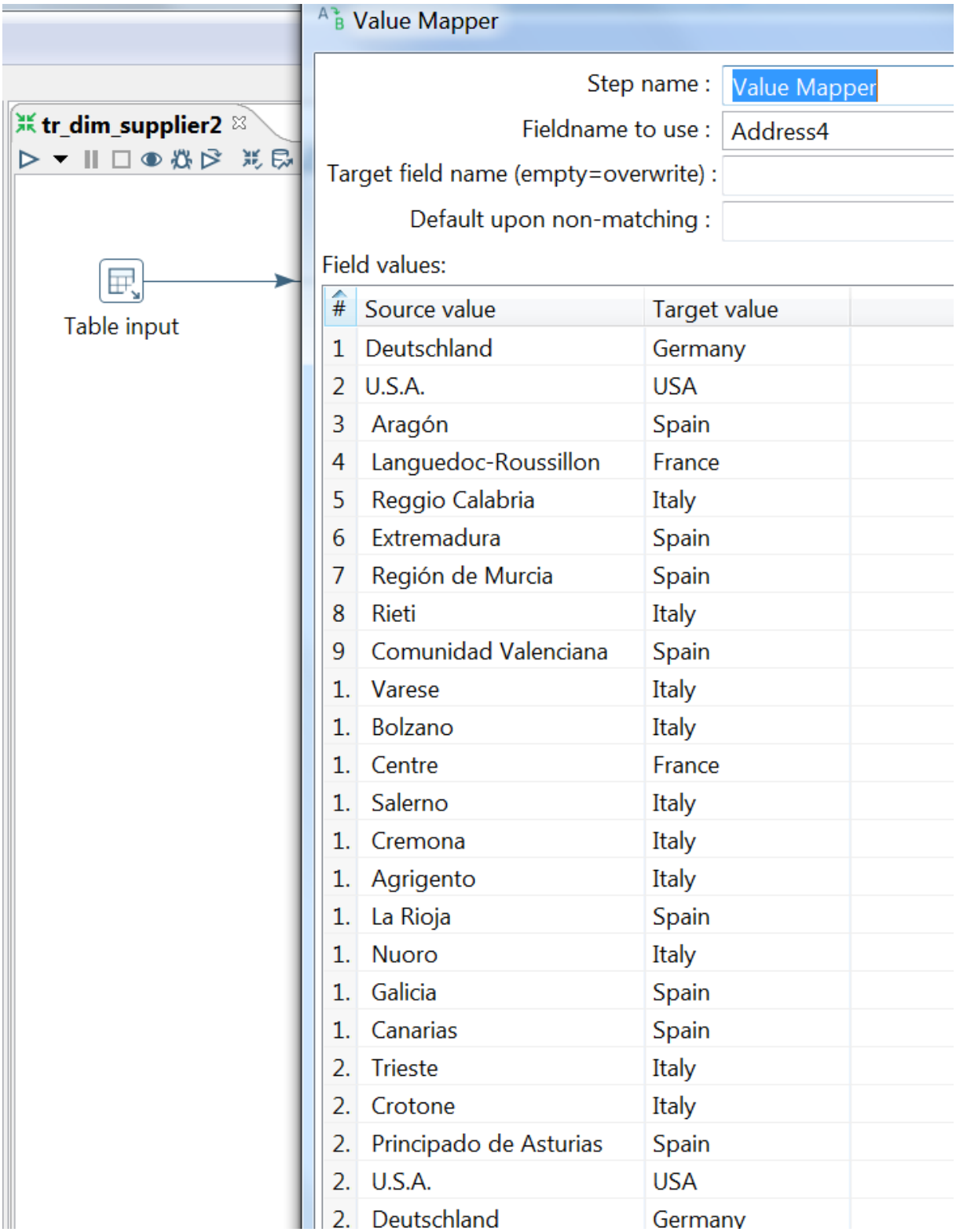


Figure 30 - Transformation for dim_supplier -> Value Mapper (1/2)

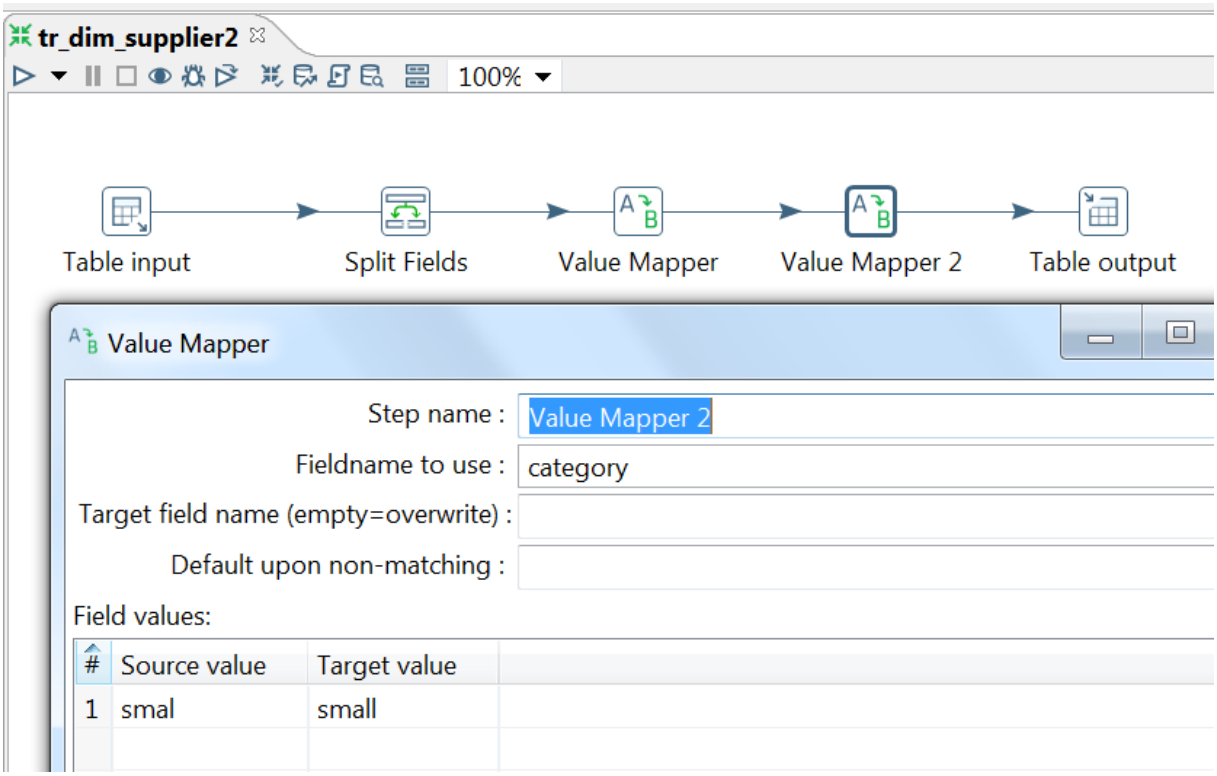


Figure 31 - Transformation for dim_supplier -> Value Mapper (2/2)

The result of the successful ETL process for dim_supplier are valid rows by design in the data warehouse:

*<superx_pcp> Script-6 | dim_material | dim_supplier | dim_purchaseorder | fact_pu

Properties | Data | Diagram | superx_

dim_supplier | Enter a SQL expression to filter results (use Ctrl+Space)

	supplier_id	supplier_category	supplier_name	country
182	182	medium	Jong, Peters and Groot	Netherlands
183	183	big	VonRueden-Jacobs	USA
184	184	big	Zemlak Inc	USA
185	185	small	Moretti Group	Italy
186	186	big	Wit-Ven	Netherlands
187	187	medium	Słowiński and Sons	Poland
188	188	small	Mehlhorn-Fleischmann	Germany
189	189	big	Bos en Zonen	Netherlands
190	190	small	Hilll Group	United Kingdom
191	191	small	Ghosh-Köhrbrück	Germany
192	192	big	Saflanis GmbH & Co. KG	Germany
193	193	medium	Guzik-Adamczak	Poland
194	194	medium	Zieme-Beatty	USA
195	195	small	Moll, Harter und Kutzner	Germany
196	196	medium	Jast-Wolff	Canada
197	197	big	Osinski Inc	United Kingdom
198	198	small	Walker-Lind	USA
199	199	big	Williamson Group	USA
200	200	big	Vieweg, Thränhardt und Friedek	Germany
201	201	small	Rohan-Emmerich	United Kingdom
202	202	medium	Ortiz y Abeyta	Spain
203	203	medium	Tillman, Kling and Moore	Canada
204	204	medium	Drees, Radtke und Deuschle	Germany

Figure 32 - Result: Successful ETL Process for dim_supplier

dim_material

The following screenshots show the Pentaho Data Integration transformation step for dim_material.

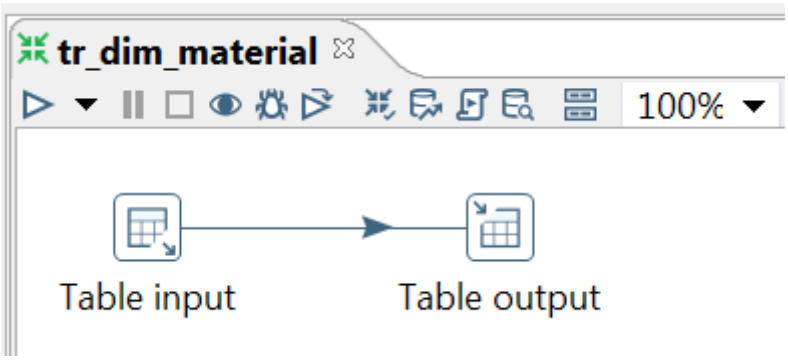
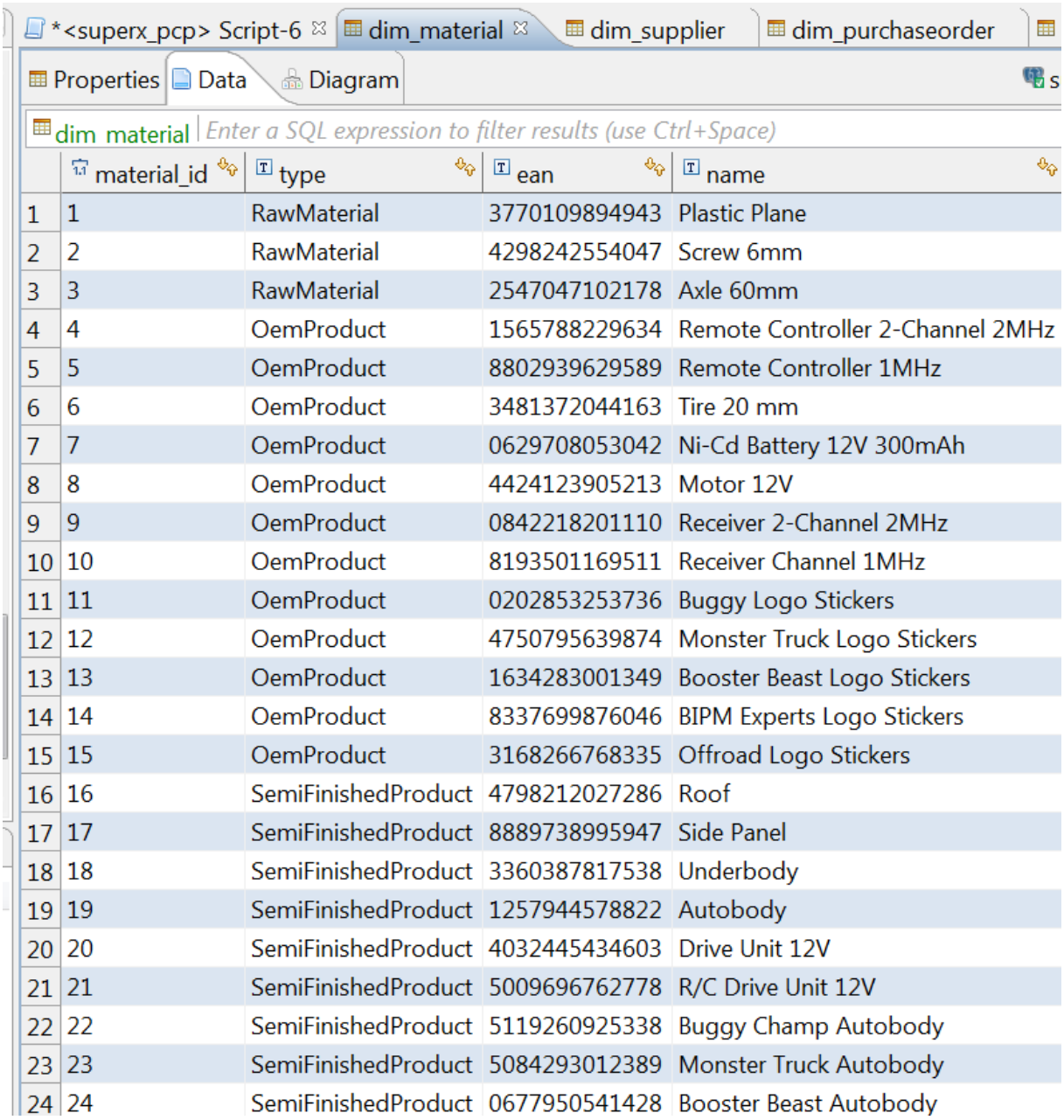


Figure 33 - Transformation for dim_material

The result of the successful ETL process for dim_material are valid rows by design in the data warehouse:



*<superx_pcp> Script-6 | dim_material | dim_supplier | dim_purchaseorder

Properties | Data | Diagram

dim_material | Enter a SQL expression to filter results (use Ctrl+Space)

	material_id	type	ean	name
1	1	RawMaterial	3770109894943	Plastic Plane
2	2	RawMaterial	4298242554047	Screw 6mm
3	3	RawMaterial	2547047102178	Axle 60mm
4	4	OemProduct	1565788229634	Remote Controller 2-Channel 2MHz
5	5	OemProduct	8802939629589	Remote Controller 1MHz
6	6	OemProduct	3481372044163	Tire 20 mm
7	7	OemProduct	0629708053042	Ni-Cd Battery 12V 300mAh
8	8	OemProduct	4424123905213	Motor 12V
9	9	OemProduct	0842218201110	Receiver 2-Channel 2MHz
10	10	OemProduct	8193501169511	Receiver Channel 1MHz
11	11	OemProduct	0202853253736	Buggy Logo Stickers
12	12	OemProduct	4750795639874	Monster Truck Logo Stickers
13	13	OemProduct	1634283001349	Booster Beast Logo Stickers
14	14	OemProduct	8337699876046	BIPM Experts Logo Stickers
15	15	OemProduct	3168266768335	Offroad Logo Stickers
16	16	SemiFinishedProduct	4798212027286	Roof
17	17	SemiFinishedProduct	8889738995947	Side Panel
18	18	SemiFinishedProduct	3360387817538	Underbody
19	19	SemiFinishedProduct	1257944578822	Autobody
20	20	SemiFinishedProduct	4032445434603	Drive Unit 12V
21	21	SemiFinishedProduct	5009696762778	R/C Drive Unit 12V
22	22	SemiFinishedProduct	5119260925338	Buggy Champ Autobody
23	23	SemiFinishedProduct	5084293012389	Monster Truck Autobody
24	24	SemiFinishedProduct	0677950541428	Booster Beast Autobody

Figure 34 - Result: Successful ETL Process for dim_material

dim_purchaseOrder

The following screenshots show the Pentaho Data Integration transformation steps for dim_purchaseOrder:

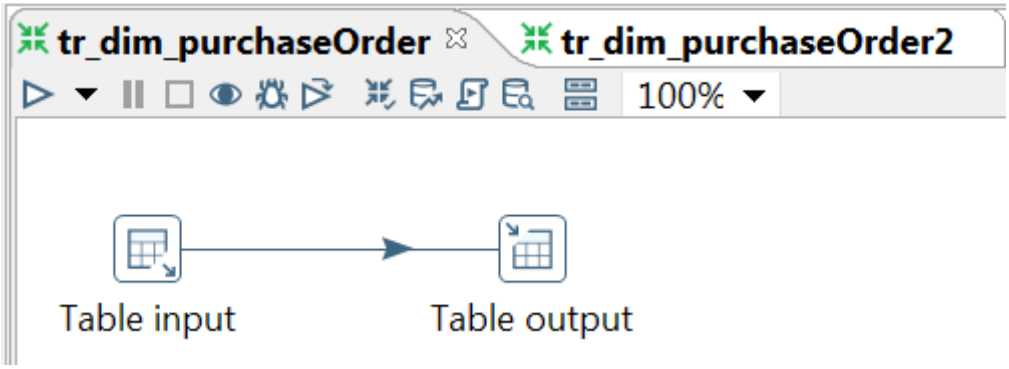


Figure 35 Transformation for dim_purchaseOrder (1/2)

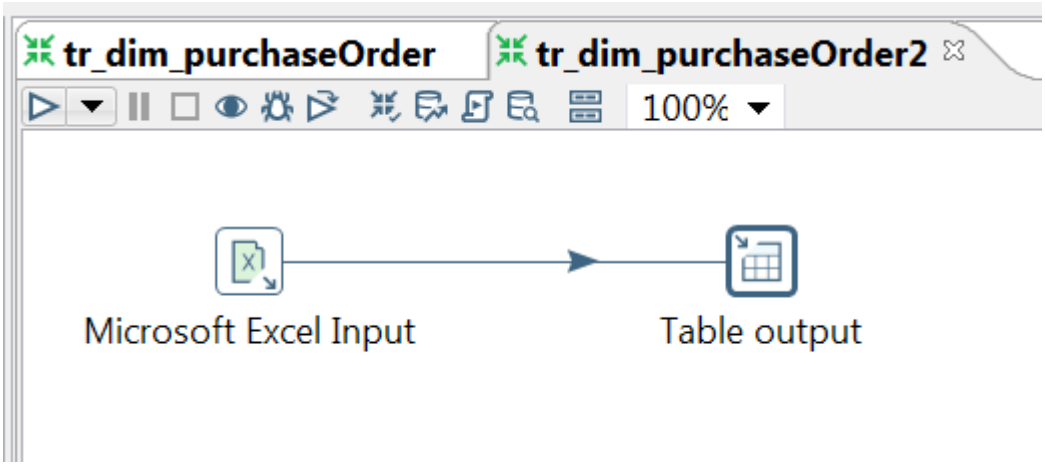


Figure 36 - Transformation for dim_purchaseOrder (2/2)

The result of the successful ETL process for dim_pruchaseOrder are valid rows by design in the data warehouse:

*<superx_pcp> Script-6 | dim_material | dim_supplier | dim_purchaseorder

Properties | Data | Diagram

dim_purchaseorder | Enter a SQL expression to filter results (use Ctrl+Space)

	purchaseorder_id	supplier_id	employee_id
1	13	13	60
2	10	10	25
3	7	7	89
4	40	13	113
5	37	10	84
6	34	7	44
7	27	27	25
8	25	25	66
9	24	24	71
10	23	23	36
11	22	22	54
12	21	21	116
13	20	20	12
14	18	18	36
15	17	17	104
16	16	16	116
17	15	15	60
18	14	14	71
19	12	12	104
20	11	11	57
21	6	6	94
22	5	5	6
23	4	4	35

Figure 37 - Result: Successful ETL Process for dim_purchaseOrder

Additional Purchase Order Excel Files

In addition to the operational database, separate Excel files (purchase orders for the years 2010, 2011, 2012, 2013 and 2014) were provided by the client:

	A	B	C	D	E	F
1	Material	Supplier	Quantity	Price	Currency	Timestamp,
2	oemote ContrRller 2-Channel 2MHz	Santoro-Barbieri s.r.l.	7179	21.21	EUR	2010-01-11 23:06:00 UTC,
3	Remote Controller 1MHz	Leeuwen V.O.F.	426	11.52	EUR	2010-01-11 23:06:00 UTC,
4	Screw 6mm	Baron EURL	83086	0.12	EUR	2010-01-11 23:06:00 UTC,
5	Ni-Cd Battery 12V 300mAh	Gorczany Inc	513	3.29	CAD	2010-01-11 23:06:00 UTC,
6	Btoster Beast Logo Soickers	GoldkÃ¼hle, Ne und Schedler	4680	0.46	EUR	2010-01-12 09:00:00 UTC
7	Axle 60mm	Santoro-Barbieri s.r.l.	10345	2.28	EUR	2010-01-12 09:00:00 UTC,
8	Sercw 6mm	Crona, Huels and Koelpin	115751	0.15	CAD	2010-01-12 09:00:00 UTC
9	Screw 6mm	Bins, Kuvalis and Hand	113533	0.17	CAD	2010-01-12 09:00:00 UTC
10	Remote Controller 2-Channel 2MHz	Leeuwen V.O.F.	3517	21.71	EUR	2010-01-12 09:00:00 UTC,
11	BIPM Experts Logo Stickers	Grasso Group	4815	0.98	EUR	2010-01-12 09:00:00 UTC,
12	BIPM Experts Logo Stickers	Smith-Vandervort	3279	1.01	USD	2010-01-12 09:00:00 UTC,
13	Offroad Logo Stickers	Crona, Huels and Koelpin	4565	0.64	CAD	2010-02-09 23:02:14 UTC
14	Tire 20 mm	Gamez, Viera y RamÃ³n Asociados	9504	1.09	EUR	2010-02-09 23:02:14 UTC
15	Remote Controller 2-Channel 2MHz	Leeuwen V.O.F.	3517	21.71	EUR	2010-02-09 23:02:14 UTC,
16	Buggy Logo Stickers	Brouwer BV	5672	0.49	EUR	2010-02-10 09:00:00 UTC,
17	Plastic Plane	Henkel-Ullrich	18771	4.27	EUR	2010-02-10 09:00:00 UTC,
18	Motor 12V	Henkel-Ullrich	3149	7.78	EUR	2010-02-10 09:00:00 UTC,

Figure 38 - Additional Purchase Orders (Excel Files)

The challenge was the deviating format of the Excel files, which we adjusted to the data warehouse format mainly by Excel formulas:

	A	B	C	D	E	F	G	H	I	J
1	PO_ID	POI_ID	Material_ID	Material Description	Supplier_ID	Supplier Name	Quantity	Price	Currency	Timestamp
2	990001	990001	4	Remote Controller 2-Channel 2MHz	3	Santoro-Barbieri s.r.l.	7179	21.21	EUR	11-01-2010 23:06:00
3	990002	990002	5	Remote Controller 1MHz	8	Leeuwen V.O.F.	426	11.52	EUR	11-01-2010 23:06:00
4	990003	990003	2	Screw 6mm	12	Baron EURL	83086	0.12	EUR	11-01-2010 23:06:00
5	990004	990004	7	Ni-Cd Battery 12V 300mAh	16	Gorczany Inc	513	3.29	CAD	11-01-2010 23:06:00
6	990005	990005	13	Booster Beast Logo Stickers	2	Goldkühle, Ne und Schedler	4680	0.46	EUR	12-01-2010 09:00:00
7	990006	990006	3	Axle 60mm	3	Santoro-Barbieri s.r.l.	10345	2.28	EUR	12-01-2010 09:00:00
8	990007	990007	2	Screw 6mm	5	Crona, Huels and Koelpin	115751	0.15	CAD	12-01-2010 09:00:00
9	990008	990008	2	Screw 6mm	6	Bins, Kuvalis and Hand	113533	0.17	CAD	12-01-2010 09:00:00
10	990009	990009	4	Remote Controller 2-Channel 2MHz	8	Leeuwen V.O.F.	3517	21.71	EUR	12-01-2010 09:00:00
11	990010	990010	14	BIPM Experts Logo Stickers	13	Grasso Group	4815	0.98	EUR	12-01-2010 09:00:00
12	990011	990011	14	BIPM Experts Logo Stickers	14	Smith-Vandervort	3279	1.01	USD	12-01-2010 09:00:00
13	990012	990012	15	Offroad Logo Stickers	5	Crona, Huels and Koelpin	4565	0.64	CAD	09-02-2010 23:02:14
14	990013	990013	6	Tire 20 mm	7	Gamez, Viera y Ramón Asociados	9504	1.09	EUR	09-02-2010 23:02:14
15	990014	990014	4	Remote Controller 2-Channel 2MHz	8	Leeuwen V.O.F.	3517	21.71	EUR	09-02-2010 23:02:14
16	990015	990015	11	Buggy Logo Stickers	1	Brouwer BV	5672	0.49	EUR	10-02-2010 09:00:00
17	990016	990016	1	Plastic Plane	4	Henkel-Ullrich	18771	4.27	EUR	10-02-2010 09:00:00
18	990017	990017	8	Motor 12V	4	Henkel-Ullrich	3149	7.78	EUR	10-02-2010 09:00:00
19	990018	990018	10	Receiver Channel 1MHz	4	Henkel-Ullrich	420	8.06	EUR	10-02-2010 09:00:00

Figure 39 - Adjusted Excel Format

Finally, we loaded the Excel data into the warehouse via Pentaho Data Integration:

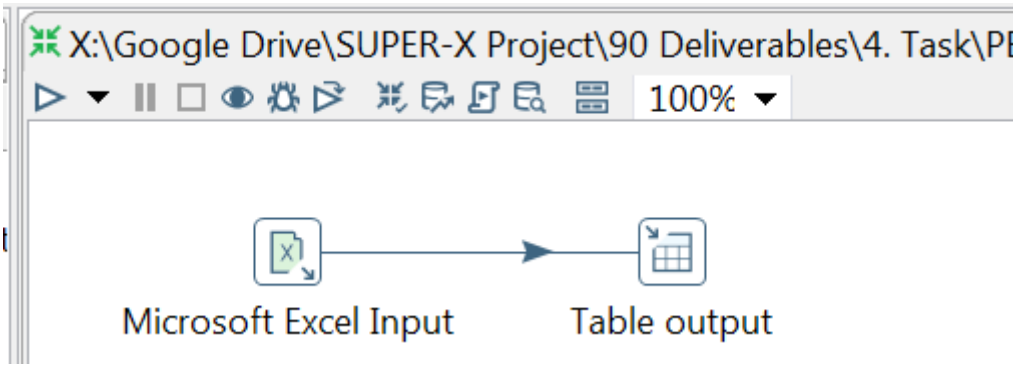


Figure 40 - Transformation for Additional Purchase Orders Excel Files

Fact Table

fact_purchaseOrderItem

The following screenshots show the Pentaho Data Integration transformation steps for fact_purchaseOrderItem. Data sources are the operational database table for purchase order items as well as the additional purchase order Excel files (→ chapter 4.B.2.4 Additional Purchase Order Excel Files).

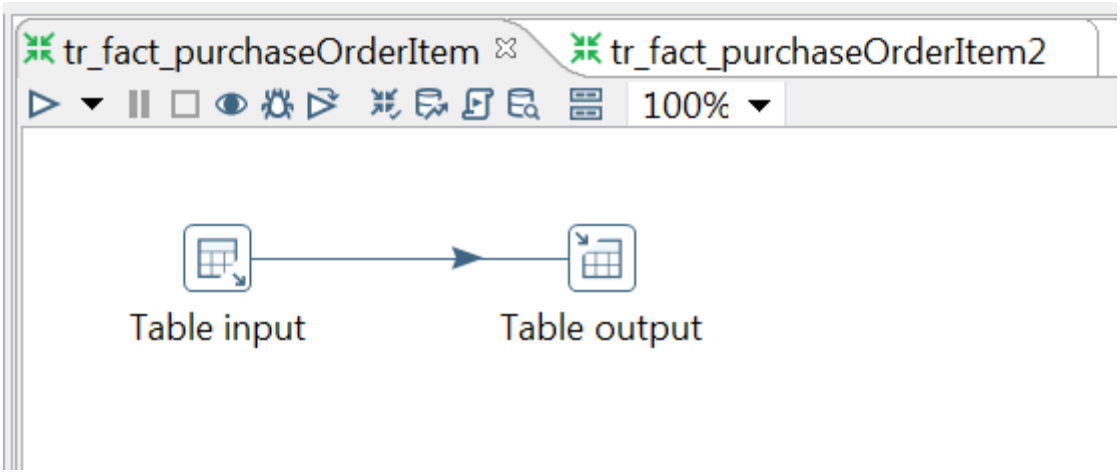


Figure 41 - Transformation for fact_purchaseOrderItem (1/2)

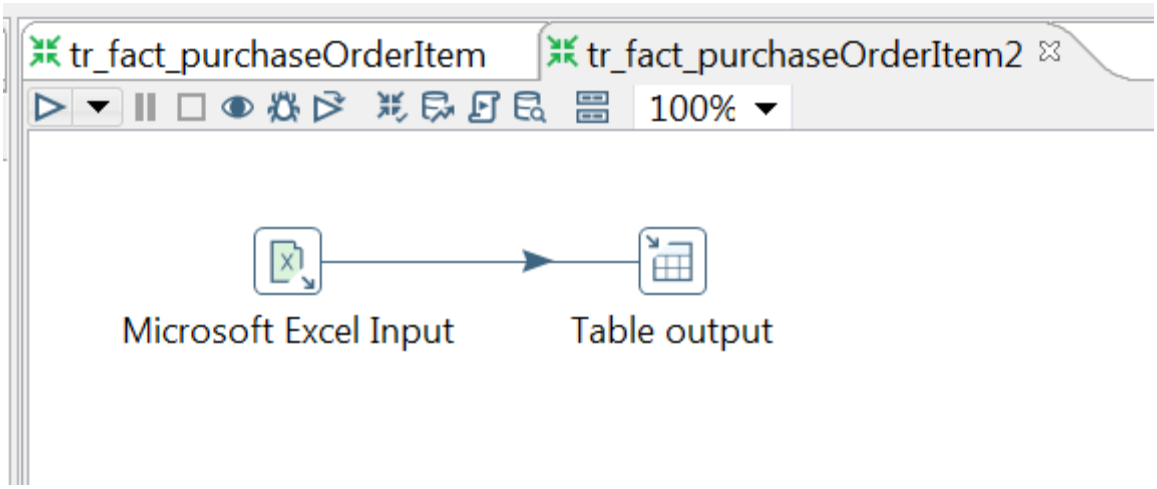


Figure 42 - Transformation for fact_purchaseOrderItem (2/2)

The result of the successful ETL process for fact_purchaseOrderItem are valid rows by design in the data warehouse:

*<superx_pcp> Script-6 | dim_material | dim_supplier | dim_purchaseorder | fact_purchaseorderitem

Properties | Data | Diagram — superx_pcp | bipm20 | Schemas | superx_pcp

fact_purchaseorderitem — Enter a SQL expression to filter results (use Ctrl+Space)

	purchaseorderitem_id	purchaseorder_id	material_id	quantity	price	currency	time_stamp
1	1	1	7	778	2,49	EUR	2010-01-11 23:06:00
2	2	1	9	7.247	15,11	EUR	2010-01-11 23:06:00
3	3	1	11	5.672	0,49	EUR	2010-01-11 23:06:00
4	4	1	13	6.005	0,48	EUR	2010-01-11 23:06:00
5	5	2	2	168.751	0,12	EUR	2010-01-11 23:06:00
6	6	2	3	4.973	2,12	EUR	2010-01-11 23:06:00
7	7	2	4	4.095	22,72	EUR	2010-01-11 23:06:00
8	8	2	5	1.089	11,34	EUR	2010-01-11 23:06:00
9	9	2	8	6.807	8,32	EUR	2010-01-11 23:06:00
10	10	3	3	10.345	2,28	EUR	2010-01-11 23:06:00
11	11	3	5	853	11,53	EUR	2010-01-11 23:06:00
12	12	3	6	14.615	1,08	EUR	2010-01-11 23:06:00
13	13	3	10	507	8	EUR	2010-01-11 23:06:00
14	14	3	11	2.397	0,47	EUR	2010-01-11 23:06:00
15	15	3	12	5.582	0,49	EUR	2010-01-11 23:06:00
16	16	3	14	6.025	0,96	EUR	2010-01-11 23:06:00
17	17	3	15	2.262	0,46	EUR	2010-01-11 23:06:00
18	18	4	1	18.771	4,27	EUR	2010-01-11 23:06:00
19	19	4	3	5.205	2,23	EUR	2010-01-11 23:06:00
20	20	4	4	5.517	23,05	EUR	2010-01-11 23:06:00
21	21	4	5	943	10,96	EUR	2010-01-11 23:06:00
22	22	4	8	3.149	7,78	EUR	2010-01-11 23:06:00
23	23	4	9	5.388	15,06	EUR	2010-01-11 23:06:00

Figure 43 - Result: Successful ETL process for fact_purchaseOrderItem

Update SQL Statements

After the transformation, there were 195 rows with missing currencies. Therefore, we enriched them with the standard currency of the respective suppliers via the following SQL statement (pseudo code):

update
superx_pcp.fact_purchaseorderitem
set
currency = 'Currency Name (EUR, USD, CAD, PLN)'
where
purchaseorder_id = PO_id;

Here, PO_id refers to the affected purchase order IDs.

In detail, we updated 41 fields where the currency “CAD” was blank.

```
update
    superx_pcp.fact_purchaseorderitem
set
    currency = 'CAD'
where
purchaseorder_id = 5 or
purchaseorder_id = 33 or
purchaseorder_id = 115 or
purchaseorder_id = 356 or
purchaseorder_id = 384 or
purchaseorder_id = 496 or
purchaseorder_id = 520 or
purchaseorder_id = 554 or
purchaseorder_id = 714 or
purchaseorder_id = 793 or
purchaseorder_id = 902 or
purchaseorder_id = 913 or
purchaseorder_id = 931 or
purchaseorder_id = 932 or
purchaseorder_id = 957 or
purchaseorder_id = 981 or
purchaseorder_id = 1071 or
purchaseorder_id = 1085 or
purchaseorder_id = 1086 or
purchaseorder_id = 1142 or
purchaseorder_id = 1192 or
purchaseorder_id = 1202 or
purchaseorder_id = 1281 or
purchaseorder_id = 1298 or
purchaseorder_id = 1431 or
purchaseorder_id = 1599 or
```

Figure 44 - SQL UPDATE Statement for the missing Currency "CAD"

Furthermore, we updated 132 fields where the currency “EUR” was blank.

```
-- Update blank EUR currencies
update
    superx_pcp.fact_purchaseorderitem
set
    currency = 'EUR'
where
purchaseorder_id = 7 or
purchaseorder_id = 28 or
purchaseorder_id = 62 or
purchaseorder_id = 83 or
purchaseorder_id = 85 or
purchaseorder_id = 121 or
purchaseorder_id = 140 or
purchaseorder_id = 1165 or
purchaseorder_id = 176 or
purchaseorder_id = 218 or
purchaseorder_id = 229 or
purchaseorder_id = 244 or
purchaseorder_id = 269 or
purchaseorder_id = 273 or
purchaseorder_id = 274 or
purchaseorder_id = 336 or
purchaseorder_id = 353 or
purchaseorder_id = 362 or
purchaseorder_id = 418 or
purchaseorder_id = 419 or
purchaseorder_id = 435 or
purchaseorder_id = 437 or
```

Figure 45 - SQL UPDATE Statement for the Missing Currency “EUR”

In addition, we updated 1 field where the currency “PLN” was blank.

```
-- Update blank PLN currencies
update
    superx_pcp.fact_purchaseorderitem
set
    currency = 'PLN'
where
    purchaseorder_id = 706;
```

Figure 46 - SQL UPDATE Statement for the Missing Currency “PLN”

Finally, we updated 21 fields where the currency “USD” was blank.

```
-- Update blank USD currencies
update
    superx_pcp.fact_purchaseorderitem
set
    currency = 'USD'
where
purchaseorder_id = 41 or
purchaseorder_id = 110 or
purchaseorder_id = 225 or
purchaseorder_id = 311 or
purchaseorder_id = 338 or
purchaseorder_id = 1660 or
purchaseorder_id = 476 or
purchaseorder_id = 621 or
purchaseorder_id = 877 or
purchaseorder_id = 1030 or
purchaseorder_id = 1150 or
```

Figure 47 - SQL UPDATE Statement for the Missing Currency “USD"

As a result, all rows in our fact table are mapped to a valid currency.

Schema Workbench

We used the Pentaho Schema Workbench as design interface to create and test our OLAP cube schema.

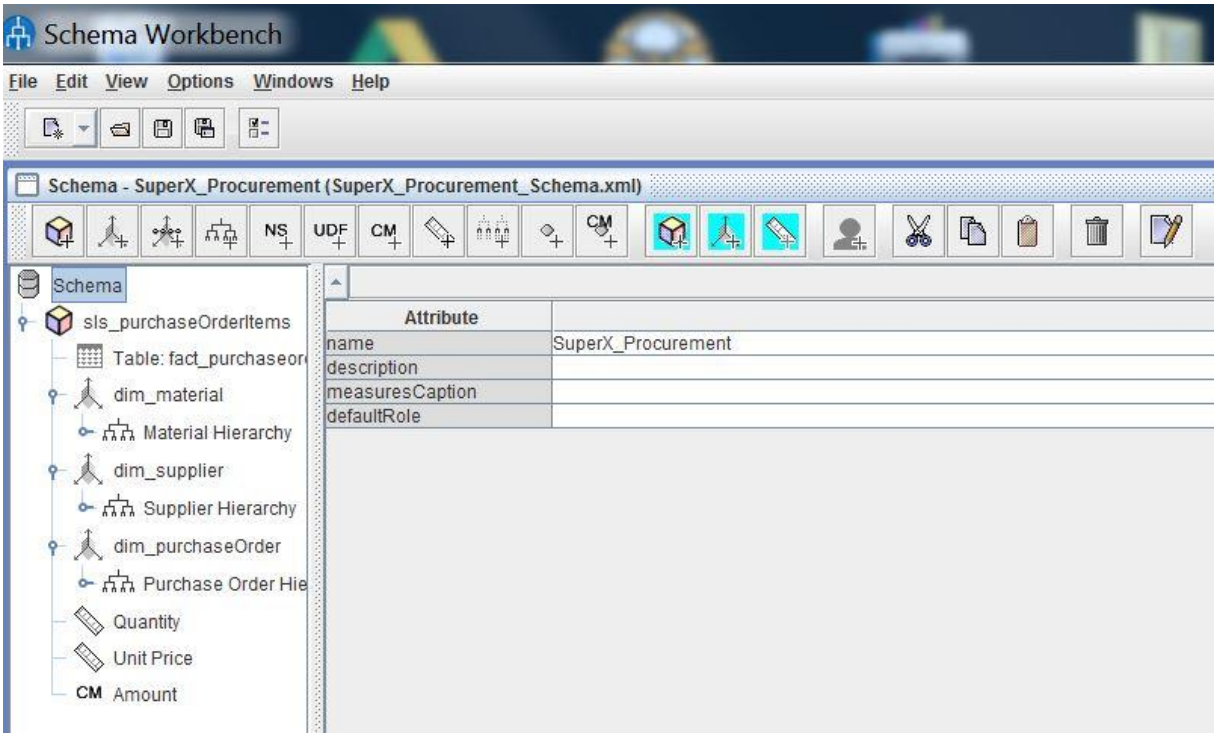


Figure 48 - Pentaho Schema Workbench

4.B.3 Implementation of Front-End for our Data Mart

After we finalized our OLAP cube, we used a Pentaho Pivot4JAnalytics View to implement a front-end in order to visualize:

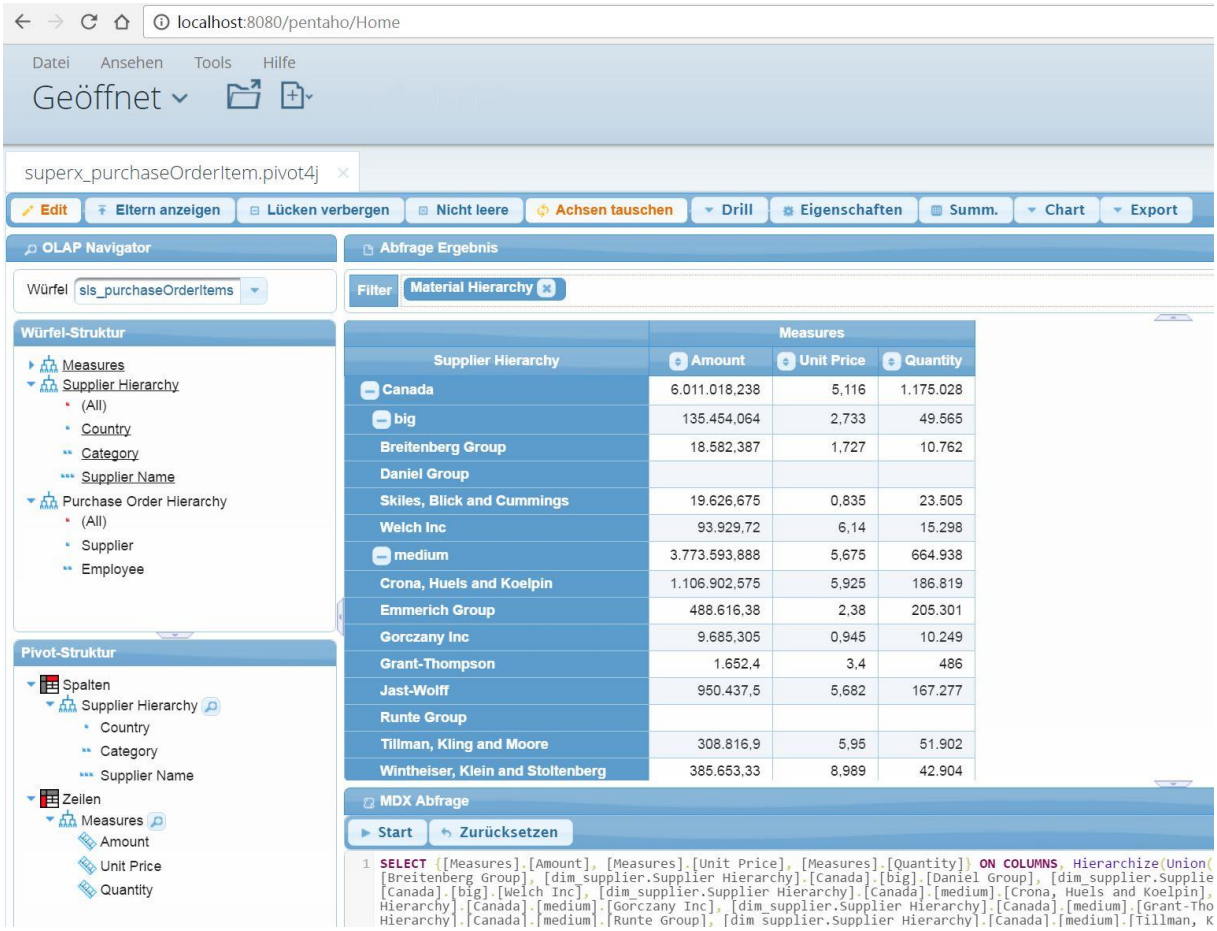


Figure 49 - Pentaho Pivot4JAnalytics

4.1 Evaluation & Comparison of both DW Technologies

4.1.1 Pentaho vs. SAP Lumira - Feature Overview

Pentaho	***SAP Lumira***
• *Data integration* • *Business Analytics* • *Big Data Analytics* • *Embedded Analytics* • *Cloud Analytics* • *Ad Hoc Analysis* • *Online Analytical Processing (OLAP)* • *Predictive Analysis* • *User-Friendly Interface* • *Ad Hoc Reporting* • *Customizable Features* • *Performance Measurements* • *Intuitive dashboards*	• Self Service Data Access • Data Transformation • Data Visualization • SAP HANA Data Discovery • Secure Sharing • Web/Mobile Support

Table 4 - DW Technology Benefits. Data from Finance Online (2016)

Features wise both tools have number of options to perform ETL process but compared to Pentaho, SAP HANA has better graphical user interface and less complex functions for ETL process. For reporting functionality, SAP Lumira doesn't have predictive analytics features but those features are available in Expert analytics where integration with R is possible with HANA as well as predictive analysis library is available.

The illustration below shows how HANA XS application (SAP Lumira) works:

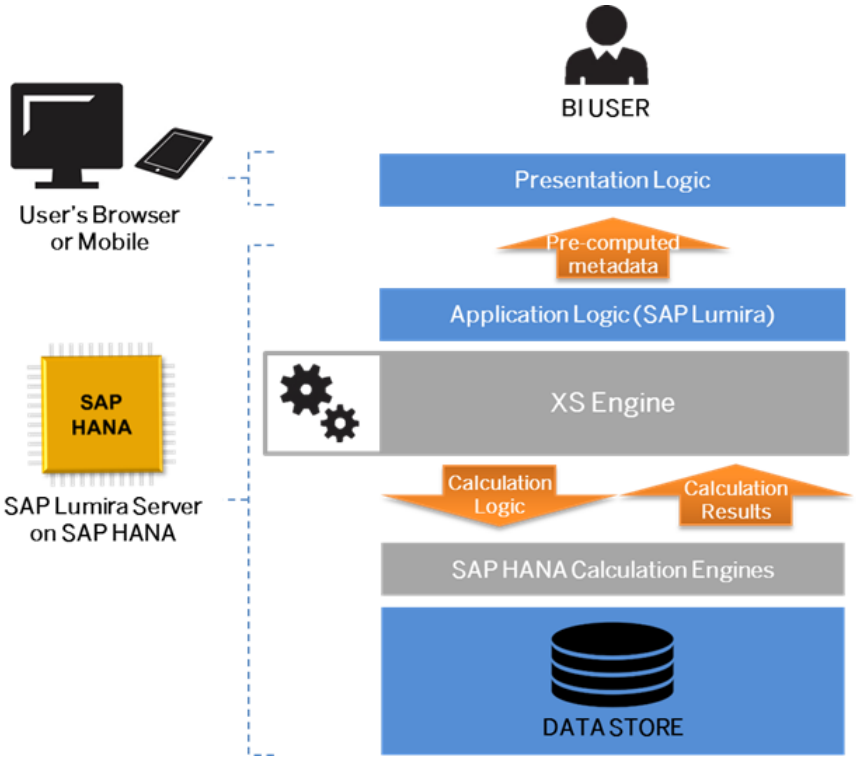


Figure 50 - Illustration of how HANA XS Application (SAP Lumira) Works. Data from Morzaria (2014)

4.1.2 Pentaho vs. SAP Lumira - Benefits

Pentaho	**SAP Lumira**
• Pentaho is a powerful data and analytics platform that offers a myriad of advantages for businesses that are keen on getting more from their data. • Powerful and thorough visualizations enable users to see data clearly, zoom in on information and other important details beyond statistical figures.	• The power of SAP Business Objects Lumira is that it pulls data directly from a variety of personal sources, blending it together in a seamless and repetitive way. Connection with HANA cube is as simple as clicking few buttons. • It's easy to enrich the traditional information framework with geographic information, to merge

• Get real-time analysis of information as the system derives information through in-memory data caching. • Total control with interactive and customizable web-based drag and drop dashboards and a library full filter functions. • Data integration software allows users to blend information sourced from other pools of information, including NoSQL, Hadoop, relational databases, and analytical databases.	data based on particular attributes, to format it with your own measures, and to apply sophisticated formulas without any coding expertise. • Visualizations are not simply accurate, but engaging too. Radars, tree maps, pies and bubble charts are also available. • Automating calculations and transformations is now possible without coding knowledge, and there are no predefined functions for data discovery which you must follow.

Table 5 - Benefit Wise Comparison of both DW Technologies. Data from Finance Online (2016)

4.1.3 Personal Reflection

SAP HANA

- tables, flowgraphs and calculation views (dimensions and cubes) were easy to create in SAP HANA editor
- within flowgraph process, we found many options to clean the dataset including geocoding which was were very useful. Template table output was also helpful in order to get all output columns after transformation instead of using "hdb" (sink) tables to predefine/limit the columns.
- difficulty level was higher in order to use standard SQL queries to fix data quality issues however Cleanse filter and expression editors were helpful.
- joining tables and creating star schema was also easy task because of user friendly graphical interface in HANA editor.
- one of the most exciting filter in HANA editor was to use R codes in order to clean the dataset but online help was much limited i.e. reference material such tutorials were very few and outdated.

SAP Lumira & Expert Analytics

- Lumira and Expert analytics both are reporting functionalities of SAP HANA which allows on the fly calculations to business requirement questions.
- it was possible to create running calculations, change measure type, create new calculated measure during visualization phase without need of making changes in HANA flowgraph
- data manipulation or fixing data quality issues was one of the unique feature and pretty much easy to use.
- rich graphical user interface, high end features, interactive visualizations including animations for time series data. Compared to Pentaho, it was much more sophisticated.
- using this front-end of HANA was much more intuitive compared to Pentaho

- option to save, publish or share the report on cloud or on local system
- "Prediction" is an advanced feature only available in Expert Analytics which has simple user interface and pre-installed machine learning algorithms. We found this feature very exciting and as future work for this project

Pentaho

- publishing the cube was not intuitive, minor bugs
- Kettle as an ETL tool is very powerful and was relatively easy to use, (still we relied on Excel for some minor tweaks as well as custom written SQL UPDATE statements)
- filling the fact table was pretty slow (for loading ca. 11500 fact table rows, it took 8 mins)
- creating the schema in Workbench and publishing it in Pentaho was tricky due to rather hidden data source connection parameters
- generating the pivot table in Pivot4J is not very user-friendly as it takes a lot of time for the data to load, furthermore the design is very plain compared to HANA
- Data visualization with the Pivot4J View is rather rudimentary and can be compared with an Excel Pivot
- For larger queries, Pivot4J reacts slowly and freezes sometimes

5. PROCESS MINING

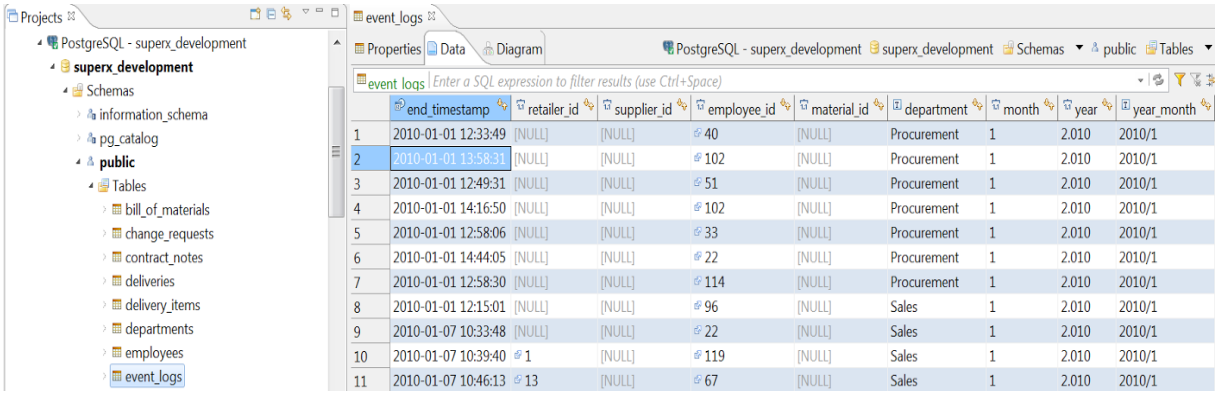
5.1 Tool - DISCO

The tool of our choice was DISCO as we already had the chance to get familiar with it in the DWH lecture. In addition, the handling of our data set was easy with DISCO to handle and the results were very entirely satisfactory.

Figure 51 - DISCO Logo

5.2 Data Source (Event Log)

The data source was the table "event logs" of the database "superx_development":



	end_timestamp	retailer_id	supplier_id	employee_id	material_id	department	month	year	year_month
1	2010-01-01 12:33:49	[NULL]	[NULL]	40	[NULL]	Procurement	1	2.010	2010/1
2	2010-01-01 13:58:31	[NULL]	[NULL]	102	[NULL]	Procurement	1	2.010	2010/1
3	2010-01-01 12:49:31	[NULL]	[NULL]	51	[NULL]	Procurement	1	2.010	2010/1
4	2010-01-01 14:16:50	[NULL]	[NULL]	102	[NULL]	Procurement	1	2.010	2010/1
5	2010-01-01 12:58:06	[NULL]	[NULL]	33	[NULL]	Procurement	1	2.010	2010/1
6	2010-01-01 14:44:05	[NULL]	[NULL]	22	[NULL]	Procurement	1	2.010	2010/1
7	2010-01-01 12:58:30	[NULL]	[NULL]	114	[NULL]	Procurement	1	2.010	2010/1
8	2010-01-01 12:15:01	[NULL]	[NULL]	96	[NULL]	Sales	1	2.010	2010/1
9	2010-01-07 10:33:48	[NULL]	[NULL]	22	[NULL]	Sales	1	2.010	2010/1
10	2010-01-07 10:39:40	1	[NULL]	119	[NULL]	Sales	1	2.010	2010/1
11	2010-01-07 10:46:13	13	[NULL]	67	[NULL]	Sales	1	2.010	2010/1

Figure 52 - Table "event logs" of the Database "superx_development"

As disco only accepts flat files, we extracted the data to an xls file and filtered the content on Procurement (column "department" = "Procurement").

5.3 Import Event Log to DISCO

The data columns in DISCO determine the analysis possibilities. The minimum requirements of an event log for Process Mining are the elements "Case ID", "Activity", and "Timestamp".

Case-ID:

Identifies an instance of the process. As we investigate the monthly Procurement Planning process, one instance is a specific month and hence we selected the column "year_month" as a proxy for the Case-ID.

Activity:

Refers to one process step and determines its level of detail. Consequently, we selected the column "activity".

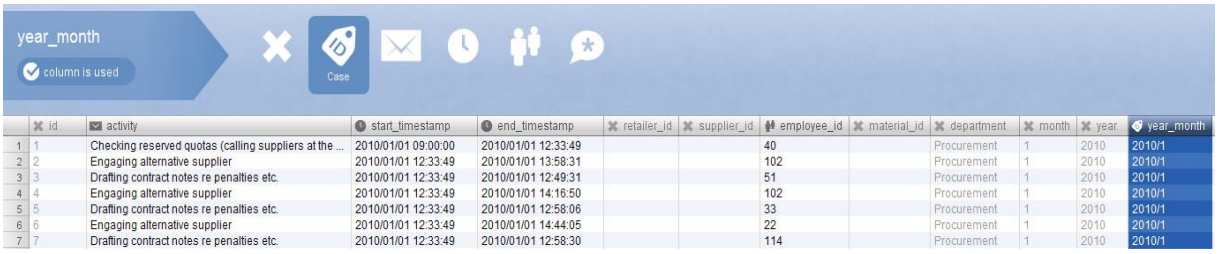
Timestamp:

We assigned the columns "start_timestamp" and "end_timestamp" for this element category, as they indicate the point in time the activity took place. Both times are also used for deriving the activity order in your process.

Resource:

We assigned "employee" for this element type to map our employees to the respective processes or rather process steps in order to measure their performance.

Furthermore, we removed the columns "id" (→ n/a), "month" & "year" (→ not needed as we declared "year_month" as Case-ID) as well as "retailer_id", "supplier_id" and "material_id" (→ empty cells).



	id	activity	start_timestamp	end_timestamp	retailer_id	supplier_id	employee_id	material_id	department	month	year	year_month
1	1	Checking reserved quotas (calling suppliers at the ...	2010/01/01 09:00:00	2010/01/01 12:33:49			40		Procurement	1	2010	2010/1
2	2	Engaging alternative supplier	2010/01/01 12:33:49	2010/01/01 13:58:31			102		Procurement	1	2010	2010/1
3	3	Drafting contract notes re penalties etc.	2010/01/01 12:33:49	2010/01/01 12:49:31			51		Procurement	1	2010	2010/1
4	4	Engaging alternative supplier	2010/01/01 12:33:49	2010/01/01 14:16:50			102		Procurement	1	2010	2010/1
5	5	Drafting contract notes re penalties etc.	2010/01/01 12:33:49	2010/01/01 12:58:06			33		Procurement	1	2010	2010/1
6	6	Engaging alternative supplier	2010/01/01 12:33:49	2010/01/01 14:44:05			22		Procurement	1	2010	2010/1
7	7	Drafting contract notes re penalties etc.	2010/01/01 12:33:49	2010/01/01 12:58:30			114		Procurement	1	2010	2010/1

Figure 53 - Case Resources

5.4 Results

In the following sections, the Process Mining results of DISCO based on the event log are listed.

5.4.1 Overview

The following picture shows all possible activity and path variants:

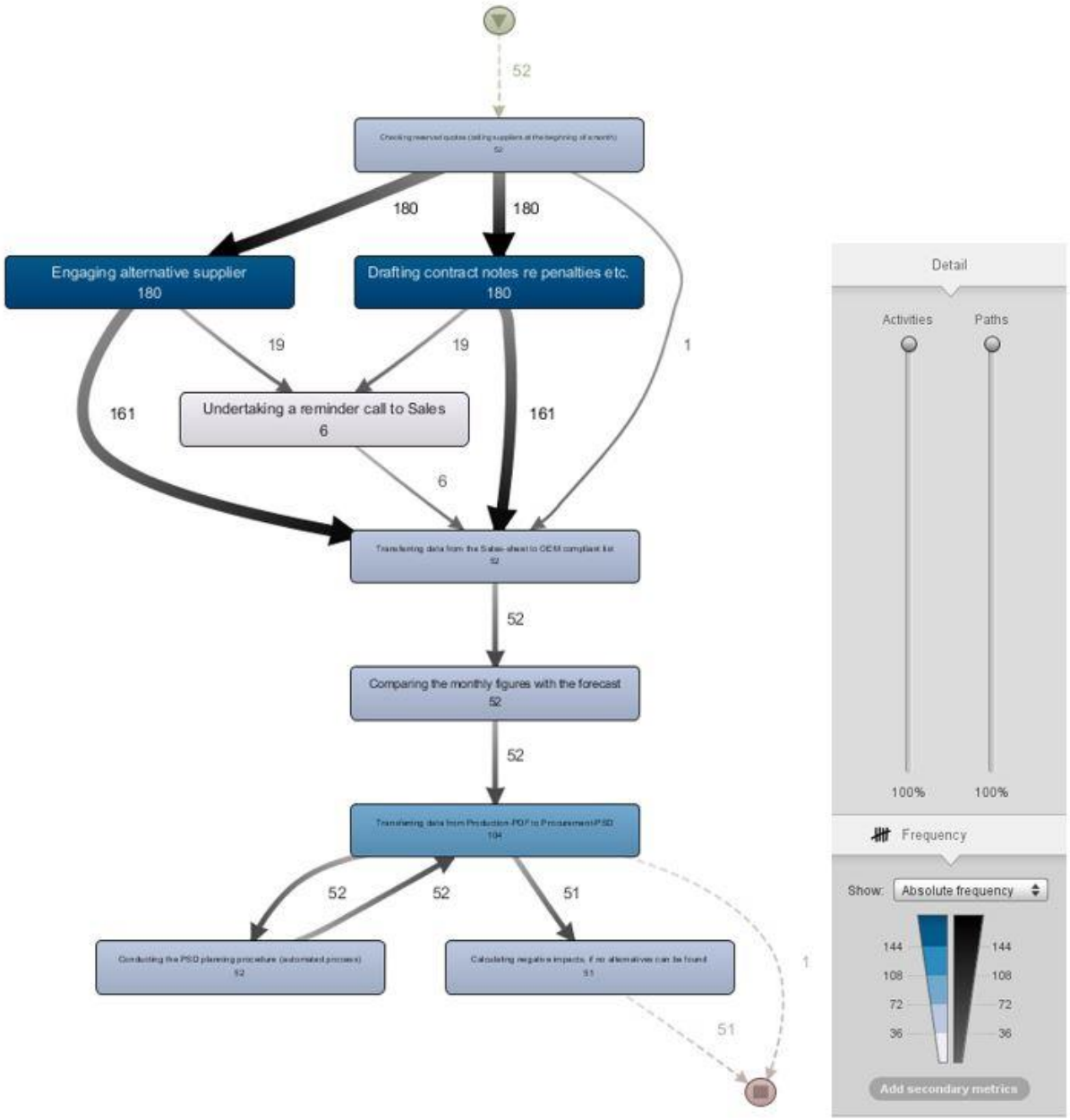


Figure 54 - All Activity and Path Variants

In total, there are 52 cases (with 11 variants) with 729 events and 28 resources assigned. A case has a mean duration of 23.1 days.

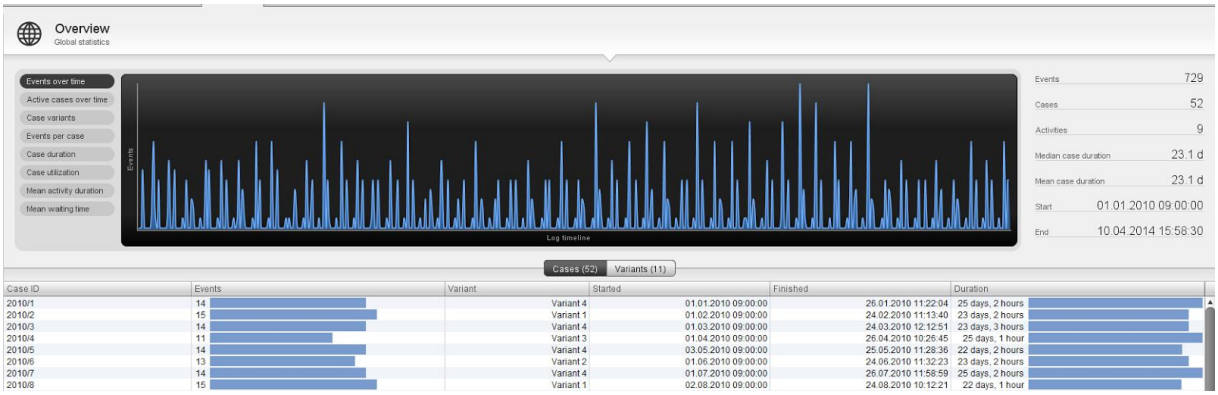


Figure 55 - Results Overview

There are 9 unique activities, whereas at least 6 activities occur for each case:

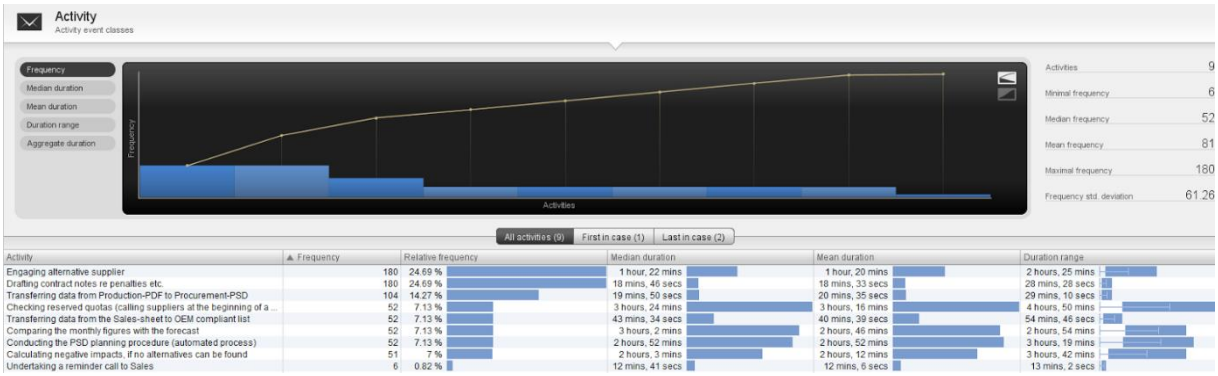


Figure 56 - Unique Activities

5.4.2 Case Variants

In total, there are 11 case variants. The maximum occurrence is 12 cases for one variant, whereas the minimum occurrence is 1 case per variant:

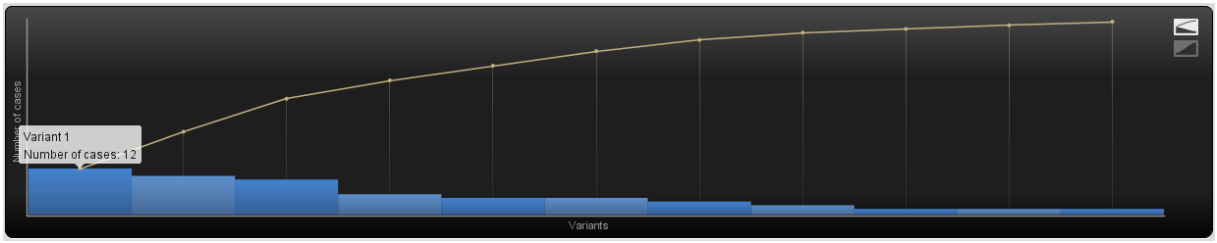


Figure 57 - Case Variants

5.4.3 Events per Case

In total, there are 729 events. The minimum number of events for a case is 7 events (occurs for one case), whereas the maximum number of events for a case is 21 events (occurs for two cases):

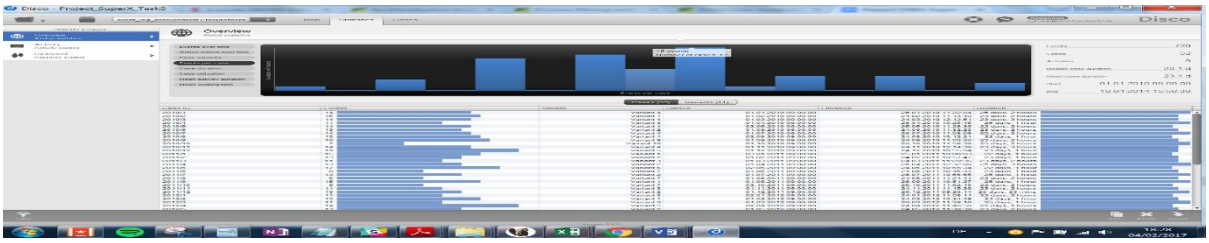


Figure 58 - Events per Case

5.4.4 Case Duration

The mean case duration is 23.1 days, whereas the minimum duration is 9 days and 10 hours (for one case) and the maximum duration is 25 days and 7 hours (for one case):

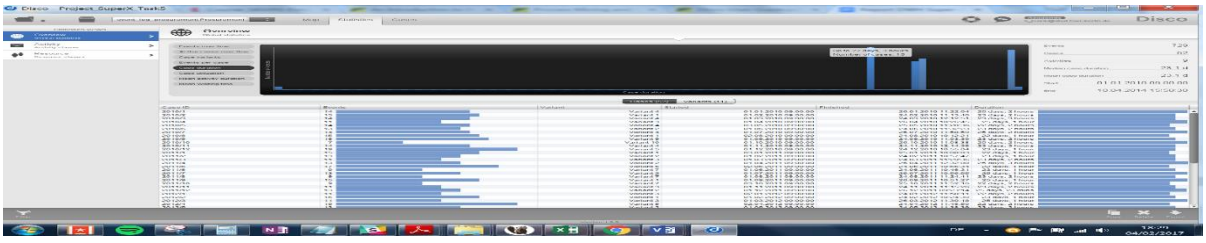


Figure 59 - Case Duration

5.4.5 Case Utilization

The lowest case utilization is 1.9% and the highest is 4.3%, whereas the highest value can be seen as an outlier:

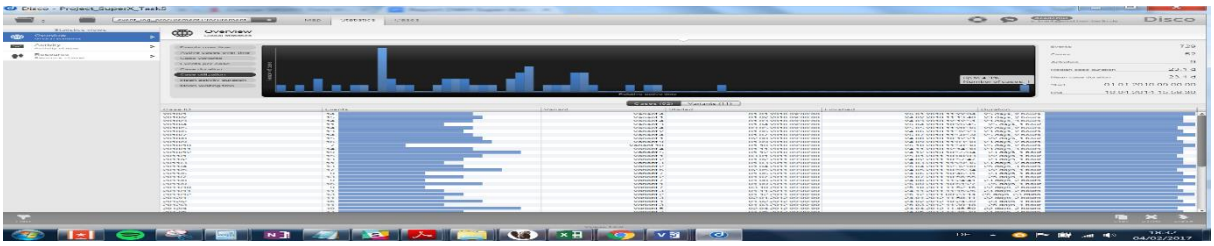


Figure 60 - Case Utilization

5.4.6 Mean Activity Duration

The minimum activity duration is 2:07 hours while the maximum is 3:44 hours:

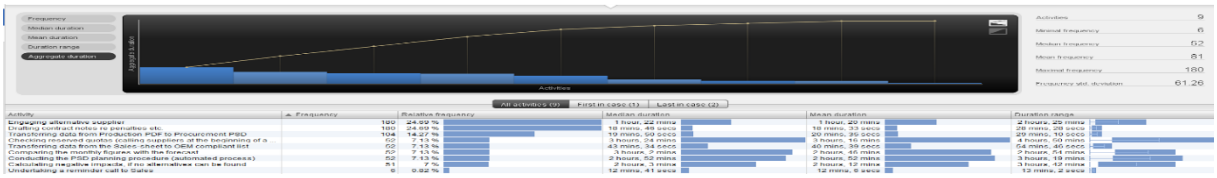


Figure 61 - Mean Activity Duration)

5.4.7 Mean Waiting Time

The minimum waiting time is 2 days and 23 hours, whereas the maximum is 6 days and 3 hours:

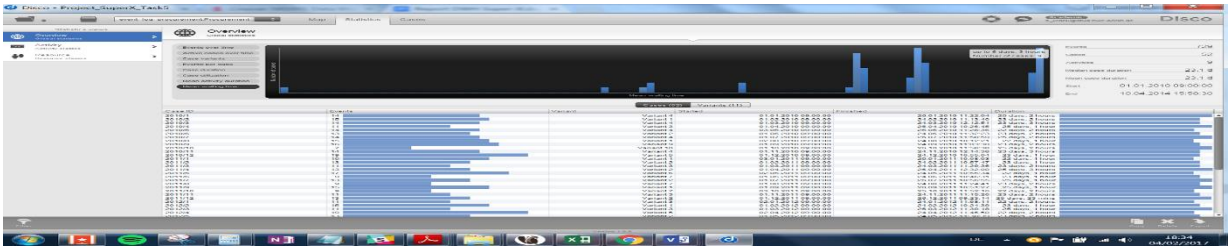


Figure 62 - Mean Waiting Time

6. RECOMMENDATIONS

This chapter features business and IT recommendations to increase the performance of Super-X's procurement department. The section is divided into business and IT improvements.

6.1 Business Improvements

6.1.1 Supplier Reliability

The overall supplier reliability is inconsistent and could be improved, as Super-X needs to search for alternative suppliers and draft penalties for existing suppliers 3 times per instance/month in average and in 99.73% of the cases the reserved quotas are not available.

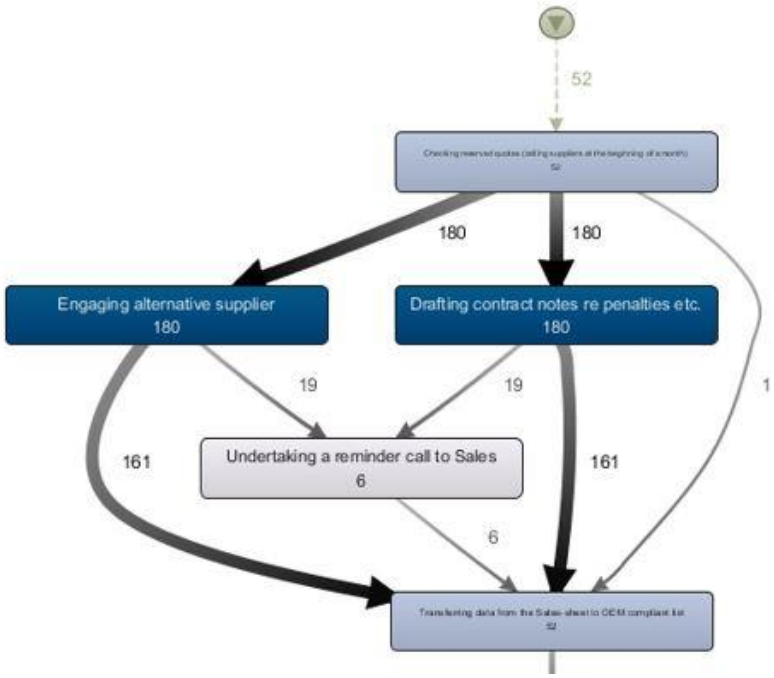


Figure 63 - Supplier Reliability

As a next step, investigate whether this is due to "black sheep" or common phenomenon and initiate corresponding steps (dismiss unreliable suppliers).

6.1.2 Procurement Process Duration

The mean duration of the whole procurement planning process is 23.1 days, which is too long (→ should be reduced to 15 days):

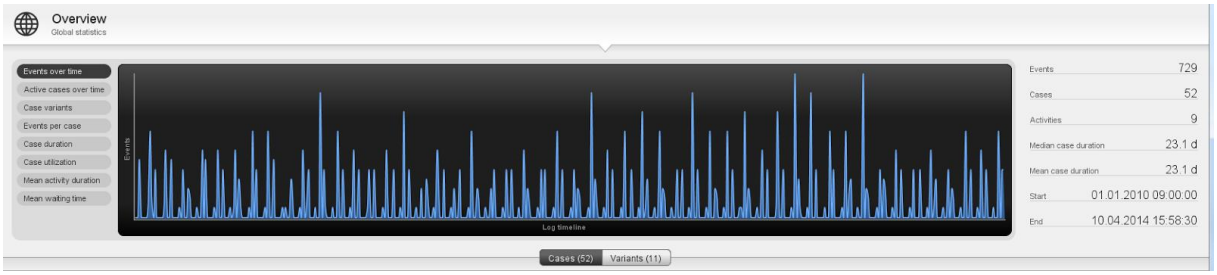
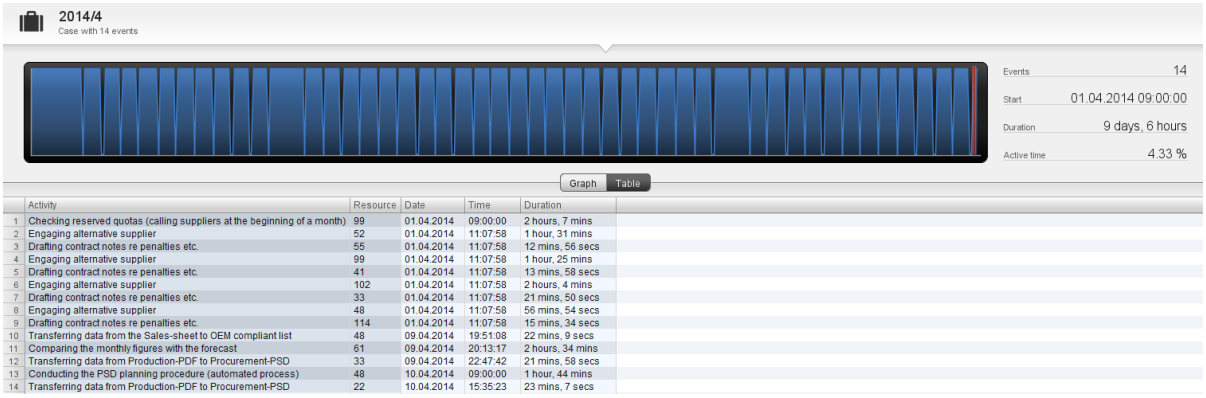


Figure 64 - Process Duration Overview

For the month 2014/04, the total process duration is 13 days and 19 hours faster than the average (9 days and 16 hours). Therefore, Super-X should further investigate this occurence (e.g. interview respective employees) in order to find out why the processing time was significantly faster in this month in order to speed up future planning processes:



	Activity	Resource	Date	Time	Duration
1	Checking reserved quotas (calling suppliers at the beginning of a month)	99	01.04.2014	09:00:00	2 hours, 7 mins
2	Engaging alternative supplier	52	01.04.2014	11:07:58	1 hour, 31 mins
3	Drafting contract notes re penalties etc.	55	01.04.2014	11:07:58	12 mins, 56 secs
4	Engaging alternative supplier	99	01.04.2014	11:07:58	1 hour, 25 mins
5	Drafting contract notes re penalties etc.	41	01.04.2014	11:07:58	13 mins, 58 secs
6	Engaging alternative supplier	102	01.04.2014	11:07:58	2 hours, 4 mins
7	Drafting contract notes re penalties etc.	33	01.04.2014	11:07:58	21 mins, 50 secs
8	Engaging alternative supplier	48	01.04.2014	11:07:58	56 mins, 54 secs
9	Drafting contract notes re penalties etc.	114	01.04.2014	11:07:58	15 mins, 34 secs
10	Transferring data from the Sales-sheet to OEM compliant list	48	09.04.2014	19:51:08	22 mins, 9 secs
11	Comparing the monthly figures with the forecast	61	09.04.2014	20:13:17	2 hours, 34 mins
12	Transferring data from Production-PDF to Procurement-PSD	33	09.04.2014	22:47:42	21 mins, 58 secs
13	Conducting the PSD planning procedure (automated process)	48	10.04.2014	09:00:00	1 hour, 44 mins
14	Transferring data from Production-PDF to Procurement-PSD	22	10.04.2014	15:35:23	23 mins, 7 secs

Figure 65 - Process Performance for the Month 2014/04

6.1.3 Process Utilization Time

The active time (→ case utilization) of all case variants has a range of 2.18 – 4.33 %, only. Hence, there is a very long waiting time associated with the Procurement Planning processes, which could be shortened to reduce the whole process duration.

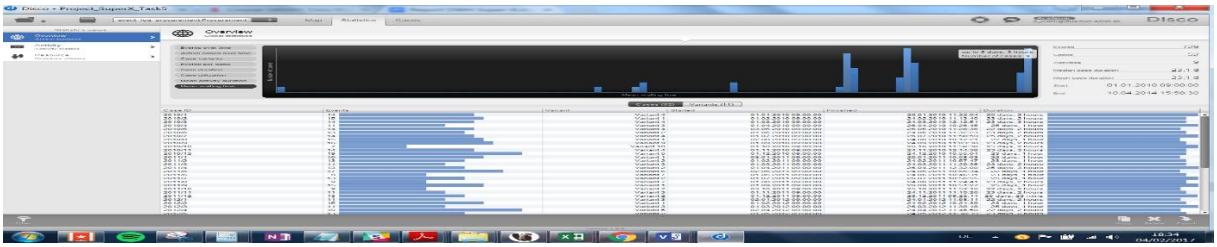


Figure 66 - Mean Waiting Time

6.1.4 Identified Rework

In all case variants, the task "Transferring data from Production-PDF to Procurement-PSD" occurs twice, even though this should be a task which should be done only once per instance. Therefore, we should further investigate this phenomenon (e.g. interview respective employees) in order to get rid of this unnecessary doublet.

6.1.5 Ordered Quantities

The ordered quantities for material 2 and 6 differ a lot by country over time. Material_id 2 (Screw 6mm) is by far the most bought item in terms of quantity.

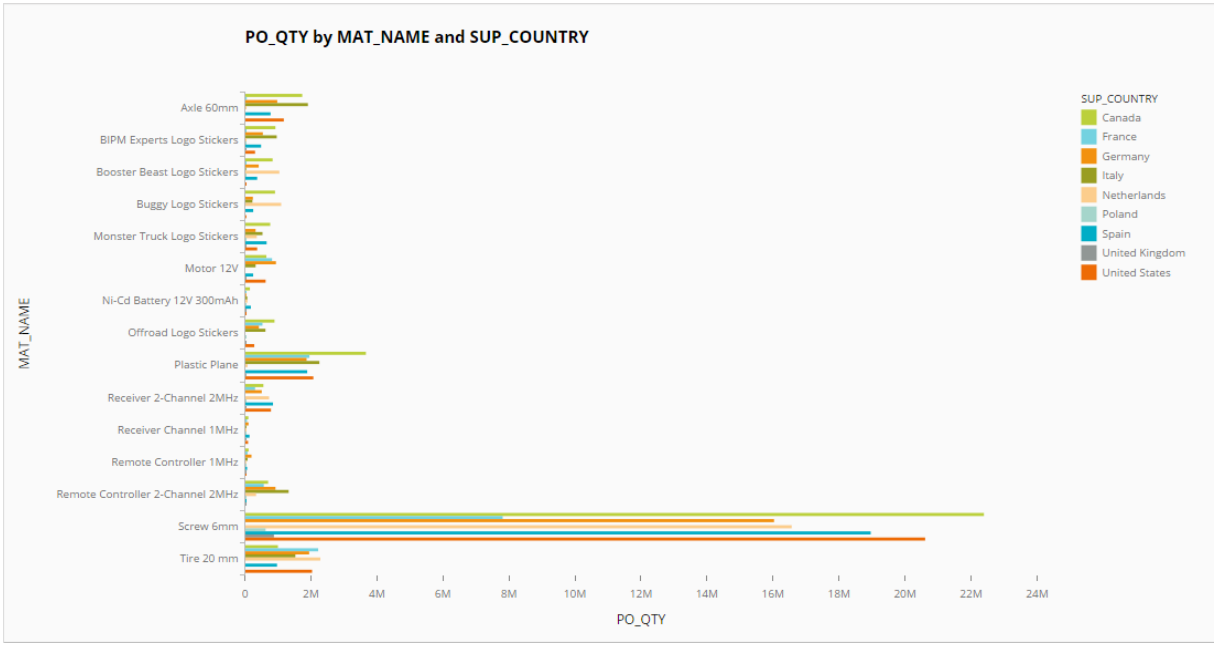


Figure 67 - Order Quantity per Material and Country

More precisely, the ordered quantities for material_id 2 differ extremely by country:

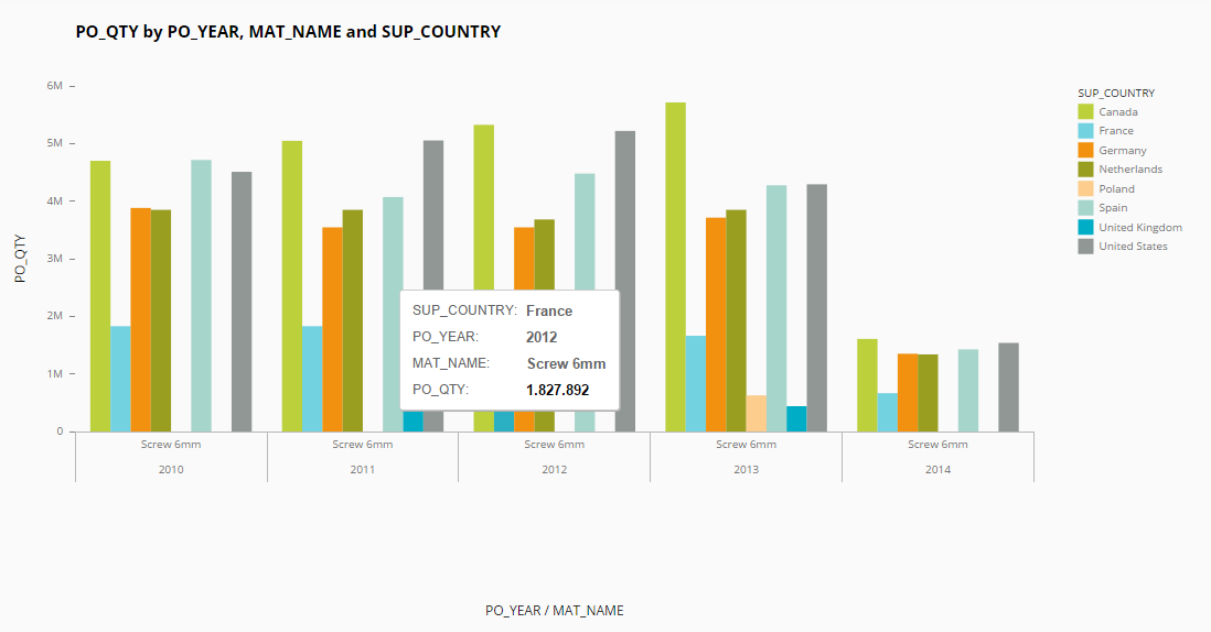


Figure 68 - Ordered Quantity for Material_id 2 by Country and Time

Due to the large ordered quantities and the differences by country, it is highly recommended to consolidate the purchases of this material by a minimum number of suppliers in order to achieve discounts and centralize distribution.

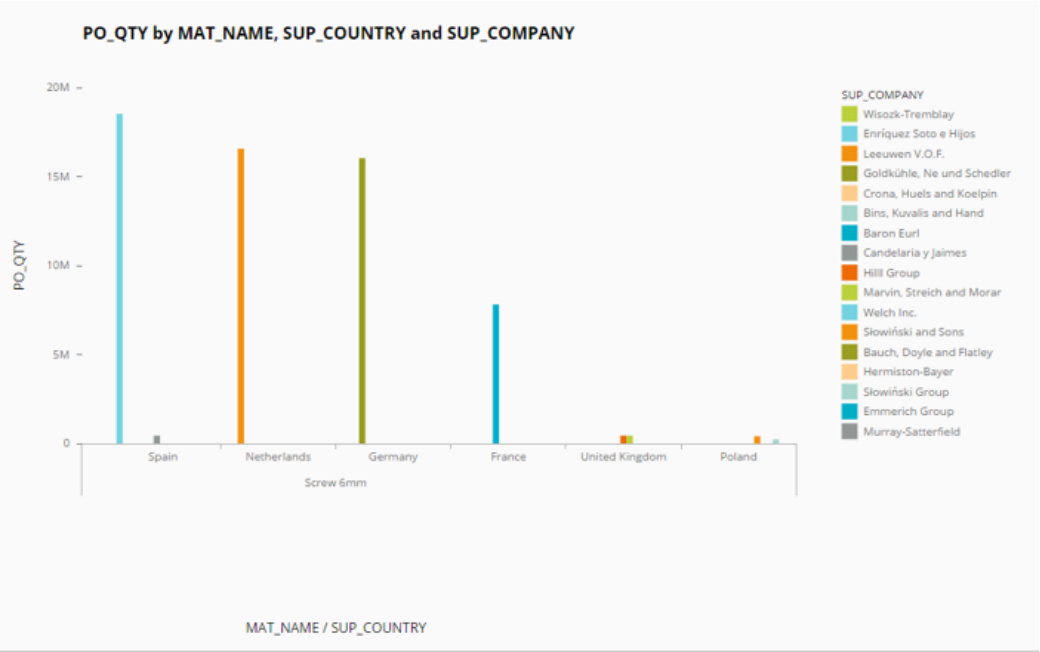


Figure 69 - Supplier and Corresponding Ordered Quantities for Material_id 2

6.1.6 Disparity of Unit Prices Among Suppliers

Prices among suppliers for material_id 4 (Remote Controller 2-Channel) and material_id 9 (Receiver 2-Channel) vary by roughly 10%.

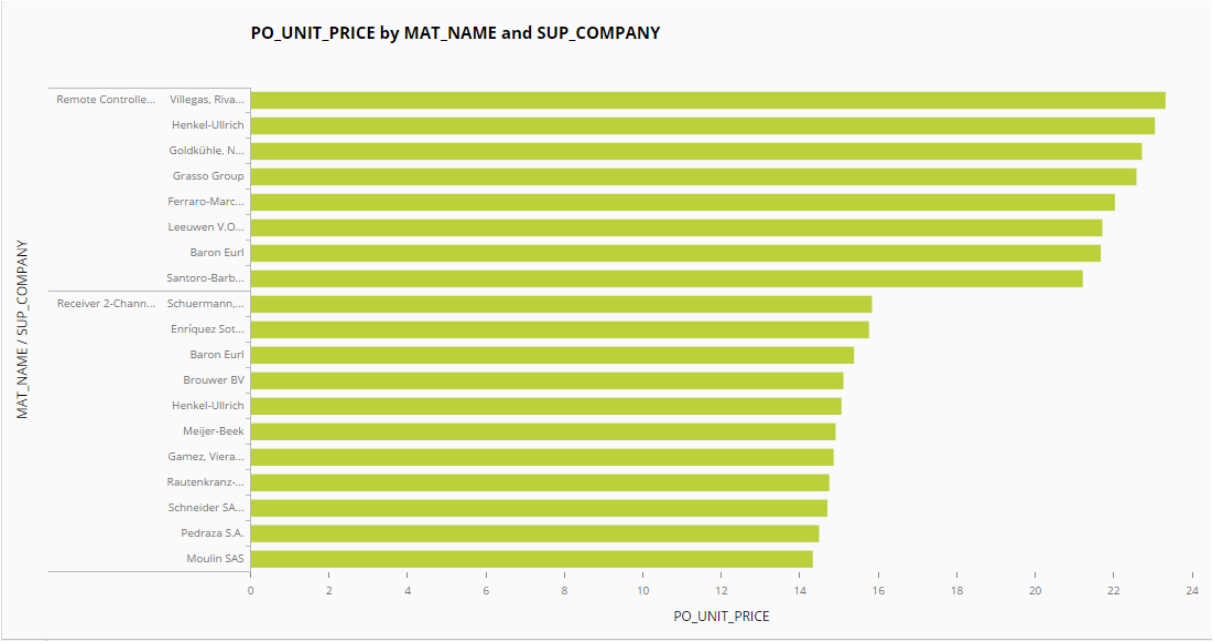


Figure 70 - Disparity of Unit Prices Among Suppliers

Investigate whether the differences in qualities could explain the price variations. If there is no difference in quality, renegotiate better price conditions.

6.1.7 Average Material Prices

Even though the unit prices differ among suppliers the overall average unit price is very stable over time and countries (e.g. Axle 60mm, European countries).

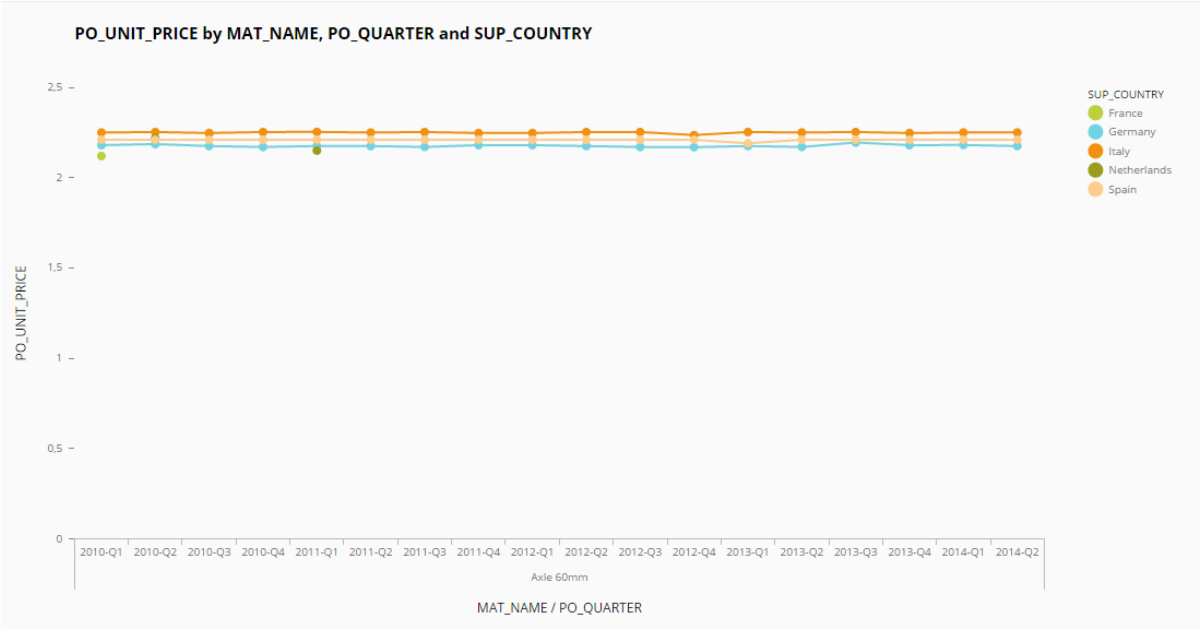


Figure 71 - PO Unit Price by Quarter and European Countries for Axle 60mm

At a first glance those figures look terrific. However, those figures could be misleading as we have seen in the previous chapters (e.g. the unit prices differ significantly among supplier and countries).

6.2 IT Improvements

6.2.1 OLTP Database

The data quality of the OLTP database needs to be improved (31 improvements → see chapter 2 / Solution column). Focus on the eight critical issues first.

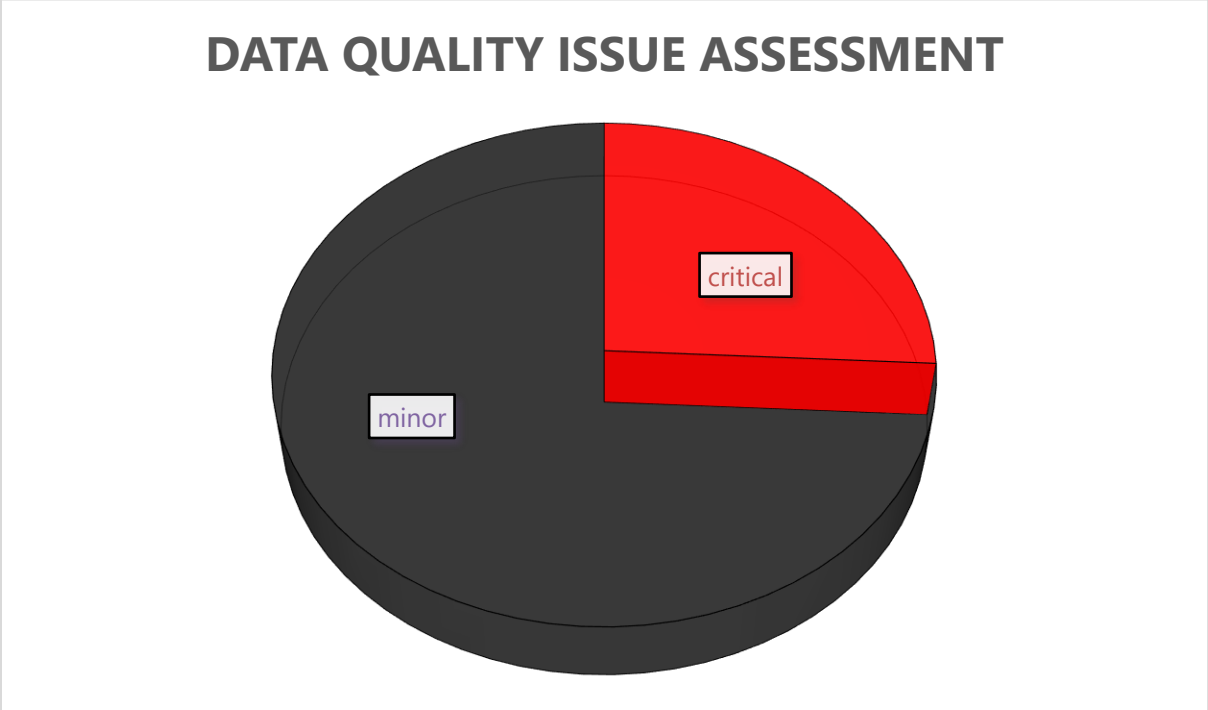


Figure 72 - Data Quality Issue Assessment

6.2.2 OLAP Architecture

After the testing our small data mart you might decide to expand your DWH possibilities and consequently implement a more suitable DWH architecture. In the following table, we summarized the most appropriate architectures which tend to fit your needs the most.

	Alternative A	***Alternative B***
Architecture	Hub and Spoke	Data Mart Bus
Scope	Several subject areas (building a scalable and maintainable infrastructure that includes a centralized DW and several dependent data marts)	One subject area (individual marts are linked to each other via conformed dimensions, better chance of maintain data consistency across the enterprise)
Development Time	Long-term	Short- to medium-term
Development Cost	More than $1,000,000	Up to $100,000
Sources	Many operational and external systems	Only some operation and external systems
Size	Up to several petabytes	Up to several gigabytes

Figure 73 - Overview Alternative. Data from Turban, Sharda and Delen (2015)

Due to the fact that we only know the data side for procurement and not the company in total, we cannot provide a clear recommendation. In addition, we are not familiar with the financial figures (e.g. IT budget) of Super-X, as well as further key figures such as potential system users.

Furthermore, we are not aware of the in-house capabilities for a potential DWH implementation support and Hypercare. Moreover, we suggest to that you evaluate our data mart first in order to estimate the business value as well as the user acceptance of such DWH technology.

7. BUSINESS REQUIREMENT VALIDATION

Out of 10 business requirement questions, our SAP Hana cubes are able to answer six completely. Unfortunately, one can only be answered partially and three not at all. This is due to missing data and structures in the OLTP database (contract notes are not linked to purchase_order and/or deliveries). In the following tables are listing the answers to each business requirement in detail.

No.	Business Requirement (Question)	Importance	High Level Entities	Measures	Achieved?
1	What is the quantity ordered for a particular period?	High	Purchase_order_items	Quantity	YES

Table 6 - Business Requirement Validation (1/10)

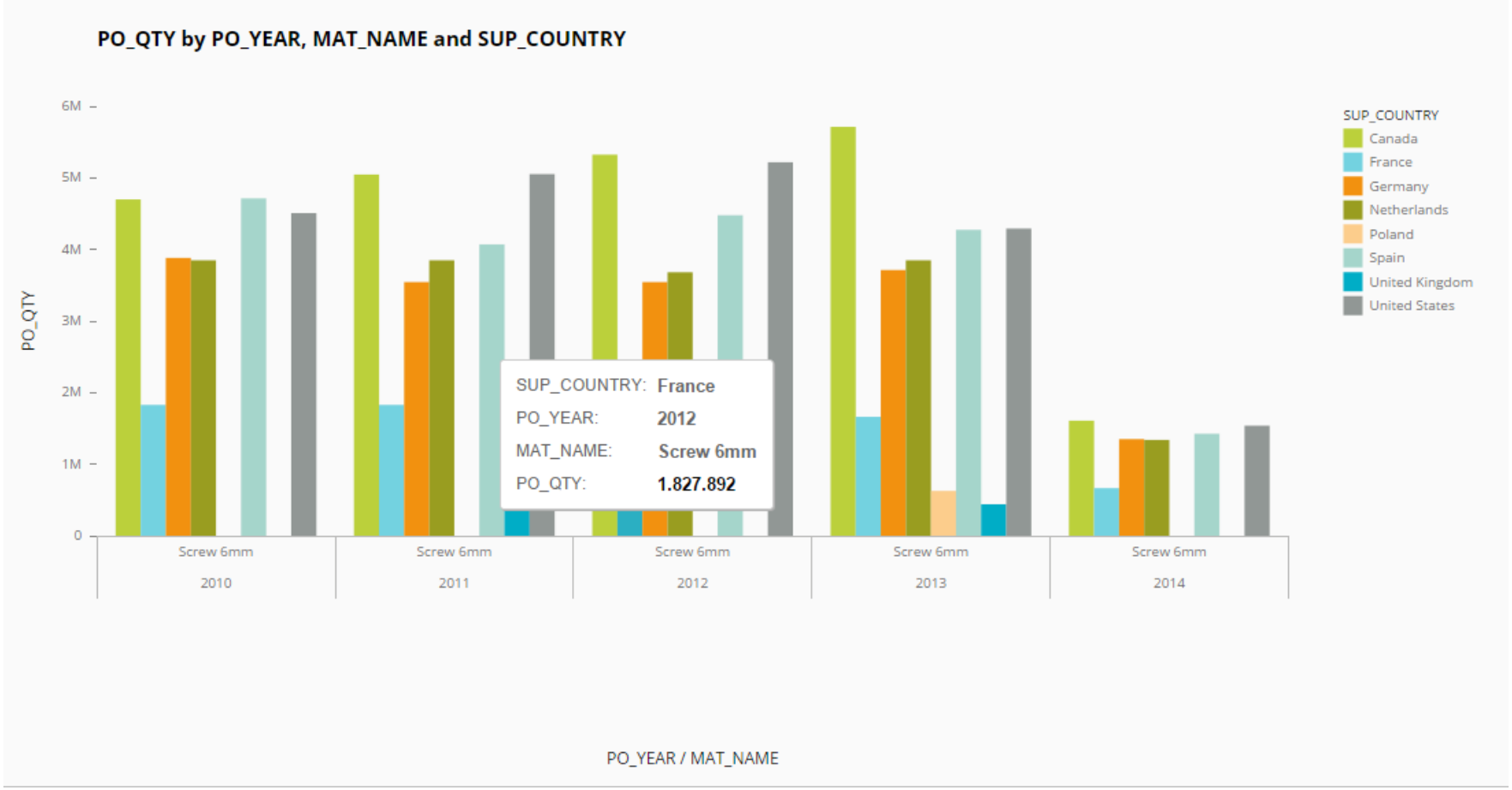


Figure 74 - Business Requirement Answer (1/10)

No.	Business Requirement (Question)	Importance	High Level Entities	Measures	Achieved?
2	What is the amount ordered for a specific period?	High	Purchase_order_items	Amount	YES

Table 7 - Business Requirement Validation (2/10)

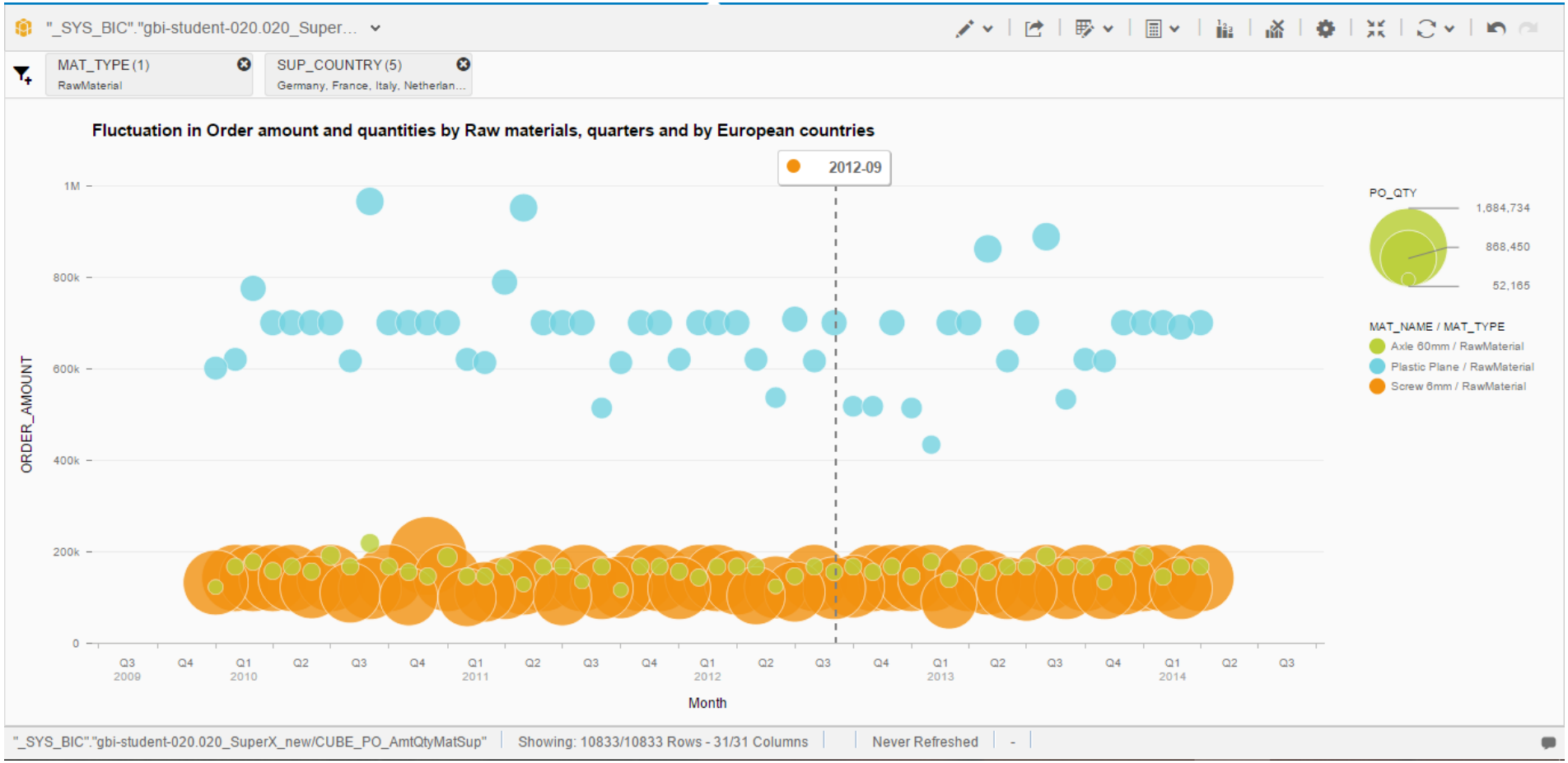


Figure 75 - Business Requirement Answer (2/10)

No.	Business Requirement (Question)	Importance	High Level Entities	Measures	Achieved?
3	Add supplier rating (on time delivery, delivery reliability)	High	• Supplier • Deliveries • Purchase_order	• On-time Delivery (OTD) = Successful Delivery / Total Deliveries * 100	**NO**

Table 8 - Business Requirement Validation (3/10)

No.	Business Requirement (Question)	Importance	High Level Entities	Measures	Achieved?
4	What's the delivery default rate of the suppliers?	High	• Supplier • Deliveries • Purchase_order	• Delivery? Yes / No	**NO**

Table 9 - Business Requirement Validation (4/10)

No.	*Business Requirement (Question)*	*Importance*	*High Level Entities*	*Measures*	*Achieved?*

5	How many suppliers are unreliable / underperforming?	High	• Supplier
			• Deliveries
			• Purchase_order
				• OTD ABC Rating
				  ○ A = (Delivery Default Rate = 0) AND 95% OTD Rate)
				  ○ B = (Delivery Default Rate = 0) AND >= 90%
				  ○ C = (Delivery Default Rate 0 - 10) AND < 85%
				  ○ D = (Delivery Default Rate > 10%) OR < 85%
					Partial (Penalty volume per supplier can be analyzed

Table 10 - Business Requirement Validation (5/10)

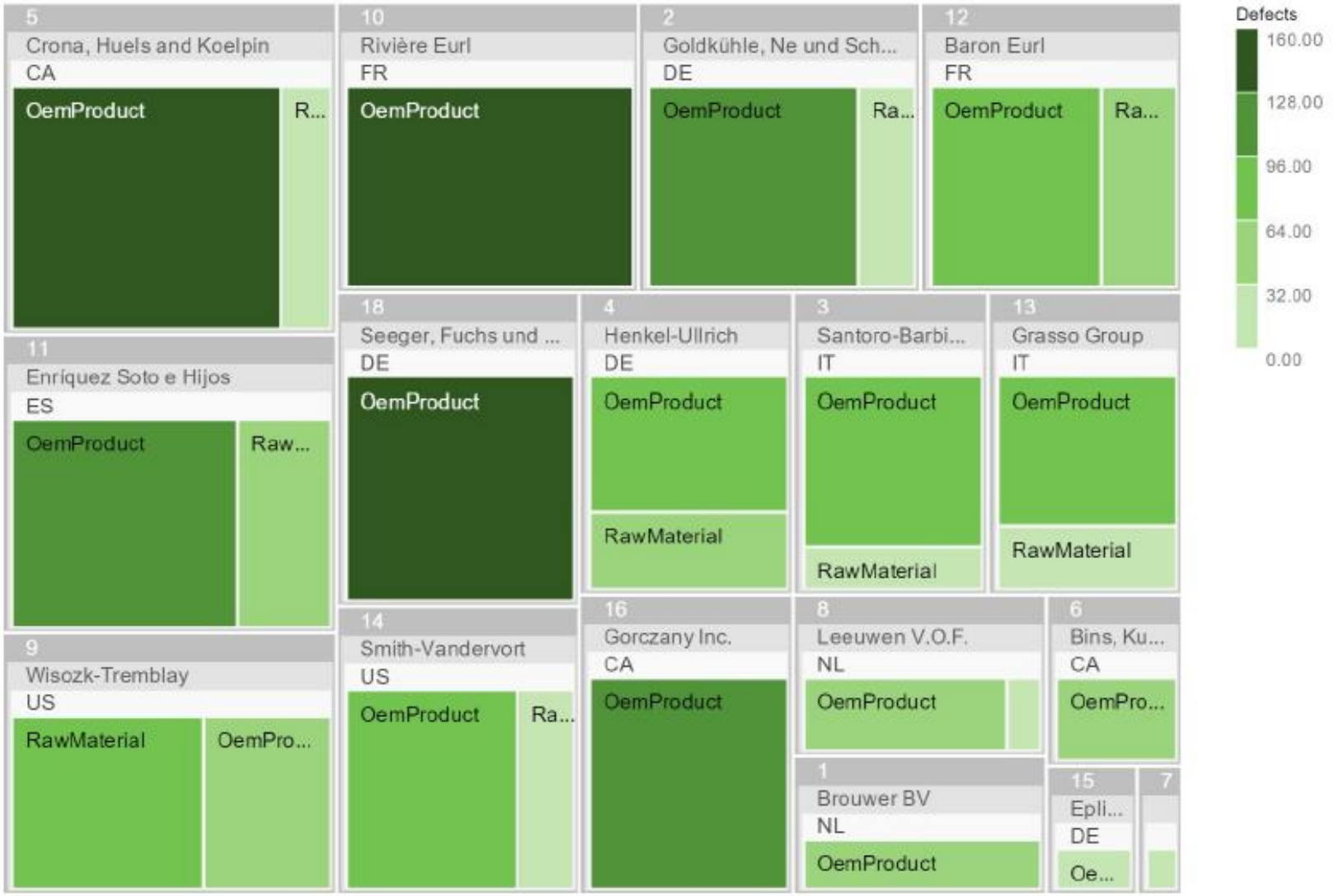


Figure 76 - Business Requirement Answer (5/10)

No.	Business Requirement (Question)	Importance	High Level Entities	Measures	Achieve
6	Material Cost Ranking	High	• Materials • Purchase_order_items	• Comparison of unit prices per currency	YES

Table 11 - Business Requirement Validation (6/10)

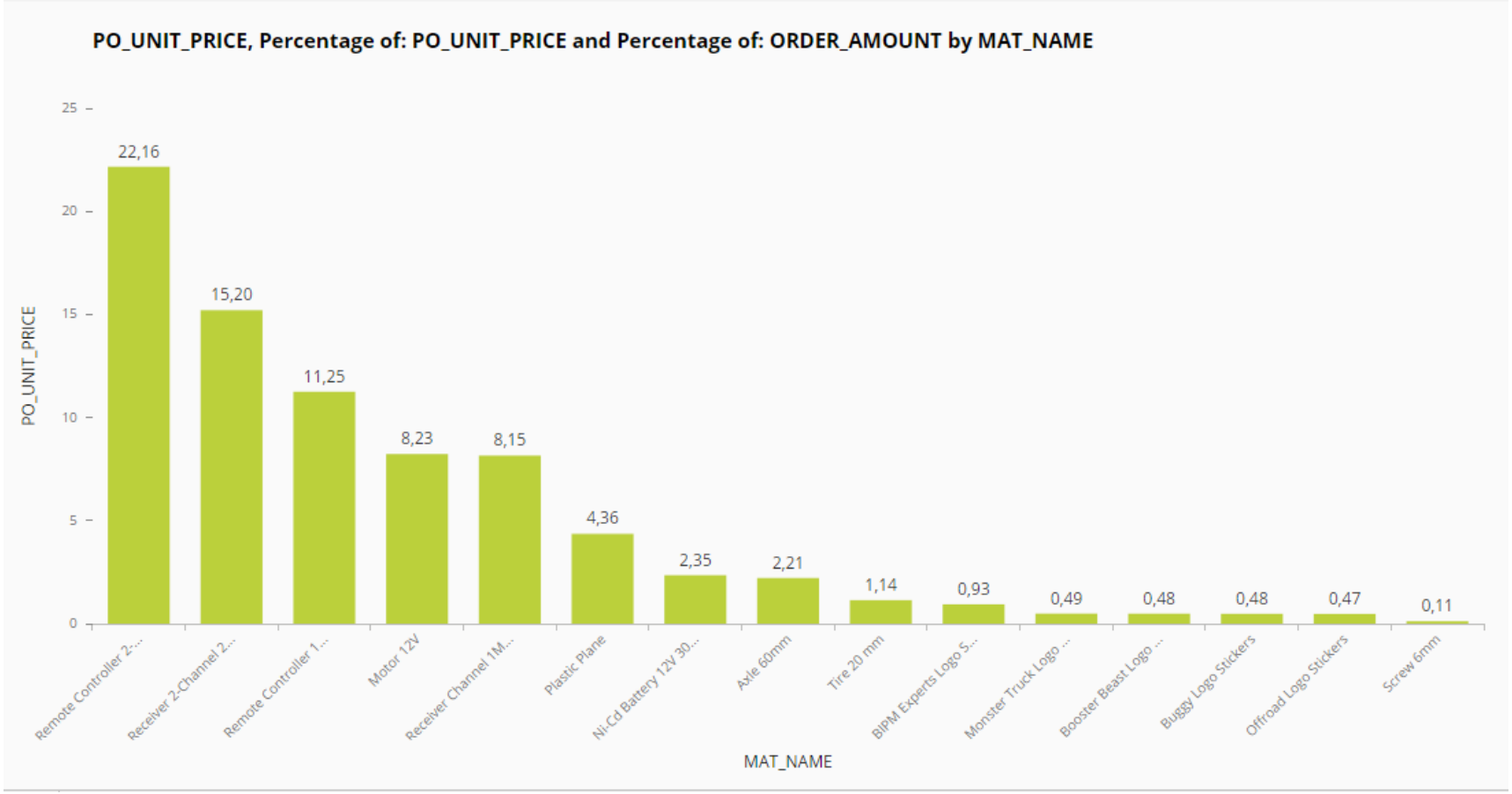


Figure 77 – Business Requirement Answer (6/10)

No.	Business Requirement (Question)	Importance	High Level Entities	Measures	Achieved?
7	What is the delivery volatility? (Constant, seasonal, one-time, etc.?)	High	Purchase_order_items (→ timestamp)	Quantity of ordered materials over time	YES

Table 12 - Business Requirement Validation (7/10)

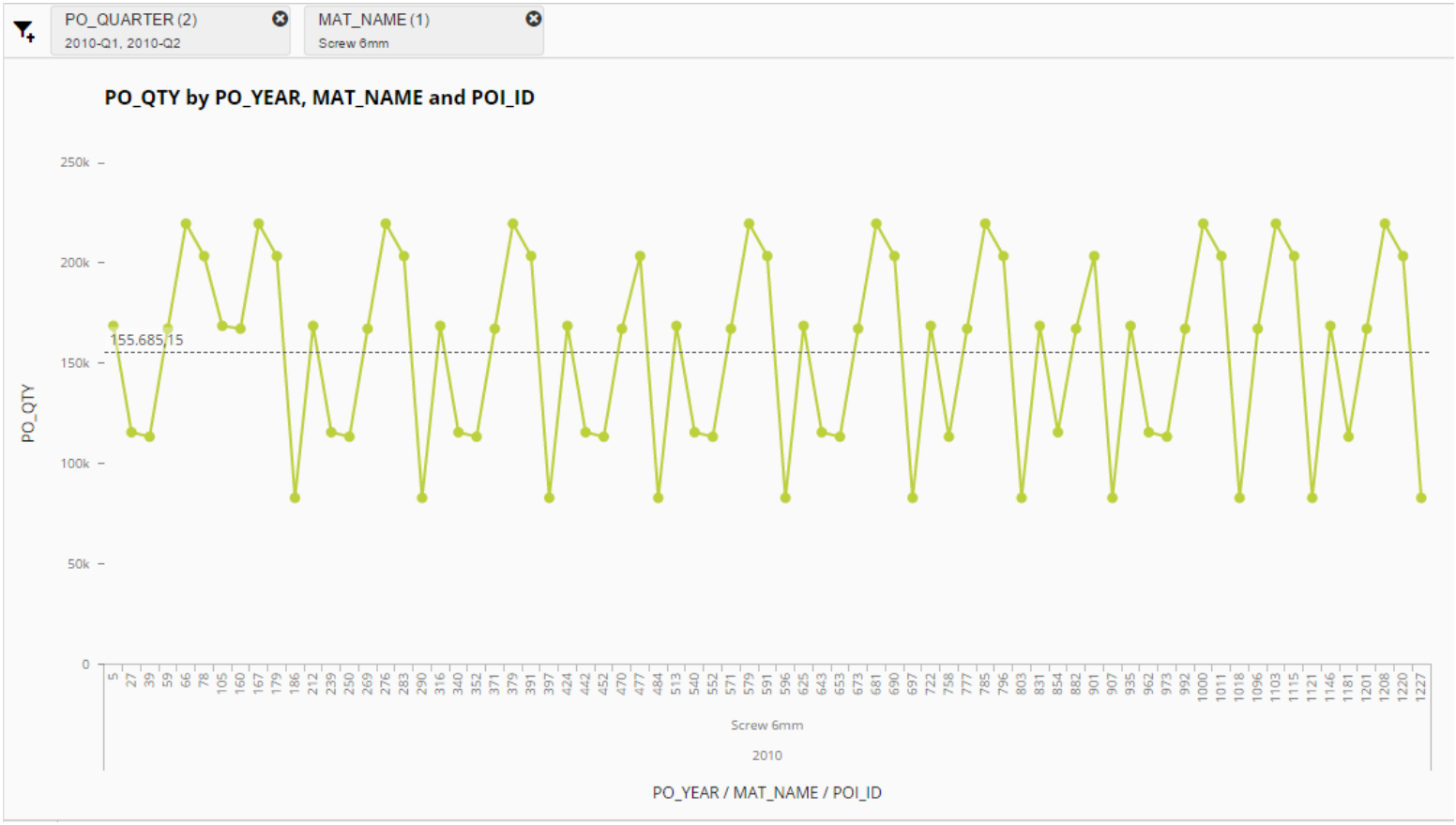


Figure 78 – Business Requirement Answer (7/10)

No.	***Business Requirement (Question)***	***Importance***	***High Level Entities***	***Measures***	***Achieved?***
8	How flexible is the supplier (rate of accepted change requests)?	High	• Change_requests • Supplier	• Flexibility Rating = Accepted Change Request / Total Amount of Change Requests * 100	**NO** (Data not available in the OLTP database table change_requests)

Table 13 - Business Requirement Validation (8/10)

No.	Business Requirement (Question)	Importance	High Level Entities	Measures	Achieved?
9	How did the prices for materials develop over time?	High	• Purchase_order_items • Material • Supplier • Purchase_Orders	• Compound Annual Growth Rate (CAGR)	**YES**

Table 14 - Business Requirement Validation (9/10)

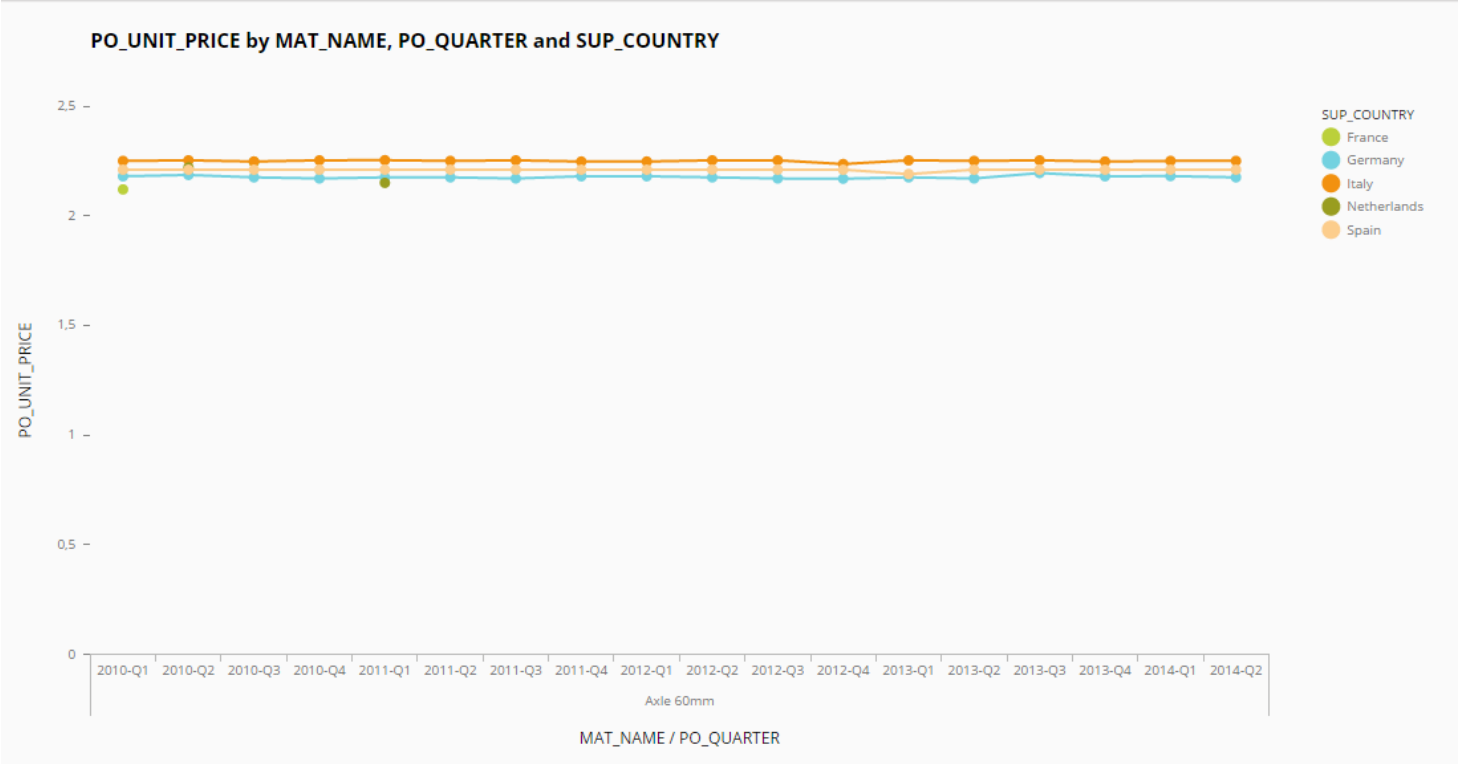


Figure 79 - Business Requirement Answer (9/10)

No.	Business Requirement (Question)	Importance	High Level Entities	Measures	Achieved?
10	What is the material price among our suppliers?	High	• Supplier • Purchase_order_items • Material • Purchase_Orders	• Comparison	YES

Table 15 - Business Requirement Validation (10/10)

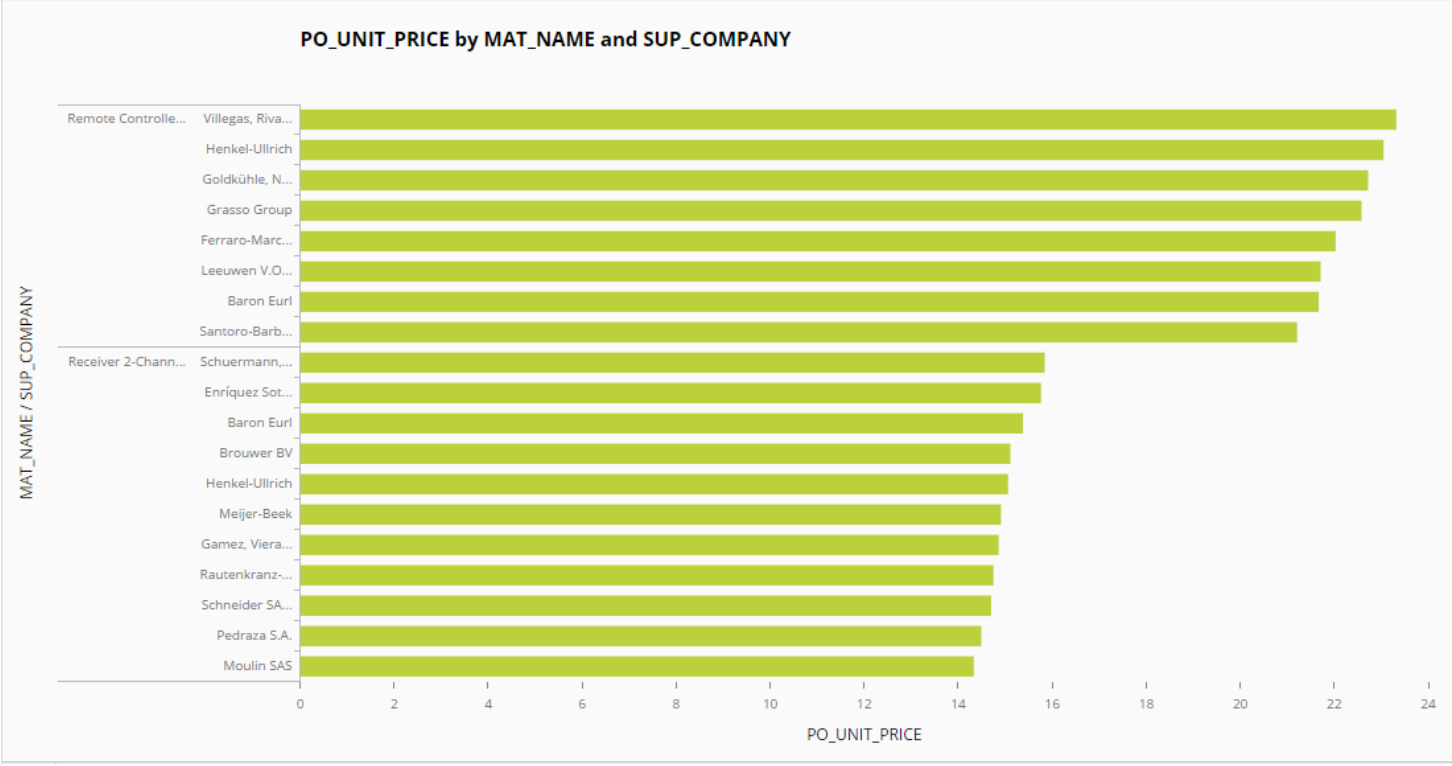


Figure 80 - Business Requirement Answer (10/10)

8. REFERENCES

Finances Online. (December 13, 2016). *Pentaho vs SAP BusinessObjects Lumira 2017 Comparison*. Retrieved February 03, 2017, from https://comparisons.financesonline.com/pentaho-vs-sap-businessobjects-lumira

Morzaria, A. (September 15, 2014). *How SAP Lumira Server Runs on SAP HANA*. Retrieved February 03, 2017, from https://blogs.sap.com/2014/09/15/how-sap-lumira-server-runs-on-sap-hana/

Turban, E., Sharda, R., & Delen, D. (2015). *Business Intelligence and Analytics: Systems for Decision Support (10th ed.)*. Upper Saddle River, N.J.: Pearson Education Limited.

9. APPENDIX

Main Task	Responsible Person
1. BUSINESS REQUIREMENTS & KPIS	All
2. RELEVANT DATA SOURCES & DATA QUALITY	
2.1 OLTP Database Structure & Logic	All
2.2 OLTP Data Quality	All
2.2.1 Email addresses	All
2.2.2 Missing Contact Persons	All
2.2.3 Phone Number Patterns	All
3. MULTI DIMENSIONAL DESIGN	
3.1 M/ER Diagram Purchase Orders	All
3.2 M/ER Diagram Contract Note	All
4. PROOF-OF-CONCEPT IMPLEMENTATIONS	
4.A SAP HANA	
4.A.1 ETL Process	Pranav
Remote Source	Pranav
Flowgraphs	Pranav
Geocoding	Pranav
4.A.2 ELT Process	Pranav
Tool: SAP Expert Analytics	Pranav
4.A.3 Multidimensional Implementation	Pranav
Dimensions	Pranav
Cube 1: CUBE_PO_AmtQtyMatSup	Pranav
Cube 2: CUBE_CON_PENALTIES	Pranav
4.A.4 Implementation of a Front-End for our Data Mart	Christian
Tools: SAP Lumira & Expert Analytics	Christian

Measures & Hierarchies	Christian
Visualizations	Christian
4.B Pentaho	
4.B.1 Multidimensional Implementation	Philipp
4.B.2 ETL Process	Philipp
Dimensions	Philipp
Additional Purchase Order Excel Files	Christian
Fact Table	Philipp
Update SQL Statements	Christian
Schema Workbench	Philipp
4.B.3 Implementation of Front-End for our Data Mart	Pranav
4.1 Evaluation & Comparison of the DW Technologies	Pranav
4.1.1 Pentaho vs. SAP Lumira - Feature Overview	Pranav
4.1.2 Pentaho vs. SAP Lumira - Benefits	Pranav
4.1.3 Personal Reflection	Philipp
SAP HANA	Pranav
SAP Lumira & Expert Analytics	Christian
Pentaho	Philipp
5. PROCESS MINING	
5.1 Tool - DISCO	Christian
5.2 Data Source (Event Log)	Christian
5.3 Import Event Log to DISCO	Christian
5.4 Results	Philipp
5.4.1 Overview	Philipp
5.4.2 Case Variants	Philipp
5.4.3 Events per Case	Philipp

5.4.4 Case Duration	Philipp
5.4.5 Case Utilization	Philipp
5.4.6 Mean Activity Duration	Philipp
5.4.7 Mean Waiting Time	Philipp
6. RECOMMENDATIONS	
6.1 Business Improvements	
6.1.1 Supplier Reliability	Christian
6.1.2 Procurement Process Duration	Christian
6.1.3 Process Utilization Time	Christian
6.1.4 Identified Rework	Christian
6.1.5 Ordered Quantities	Christian
6.1.6 Disparity of Unit Prices Among Suppliers	Pranav
6.1.7 Average Material Prices	Pranav
6.2 IT Improvements	
6.2.1 OLTP Database	Philipp
6.2.2 OLAP Architecture	Philipp
7. BUSINESS REQUIREMENT VALIDATION	Pranav
8. REFERENCES	Christian
9. Appendix	Christian

Table 16 - Division of Responsibilities