

# DDVUG DWH Automation Challenge

TDWI München  
June 2023

# WhereScape Intro



**Sultan Shiffa**

Senior Solutions  
Architect, WhereScape



**Simon Spring**

Operations Director,  
WhereScape



**Thomas Heuer**

Sales Director EMEA,  
WhereScape

**WhereScape®**  
Data Automation

# WhereScape Applications

## WhereScape® 3D

- Discover source systems
- Profile source data
- Model Design
- Metadata Tagging
- Generate Models

## WhereScape® RED

- Generate physical structures
- Generate Integration/ETL/ELT code
- Schedule
- Document
- Deploy



# WhereScape Enablement Packs



## Your Modeling Styles

Data Vault / Data Lake / Data  
Mart / ODS / Dimensional /  
EDW / Normalised /  
Bi-Temporal / Custom



## Your Platforms

Cloud / On-premises /  
Hybrid / RDBMS / Hadoop /  
Single platform or  
Federated



## Your Patterns

Micro Batch / Real-time /  
Full load / Incremental  
load / History / Slowly  
Changing

MULTI PLATFORM  
Optimised, Native and  
Portable



# Willibald Solution

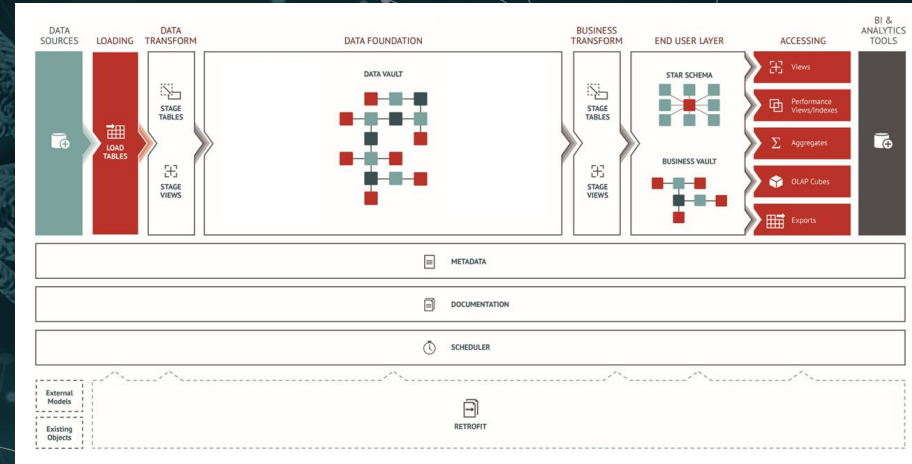


Samen- und Pflanzenhandel  
**WILLIBALD**

**WhereScape®**  
Data Automation

# High-level Walkthrough

- The end: Snowflake+SQL Server:
  - Raw vault
  - Business Vault
  - Information Mart
- The beginning:
  - Source Discovery & Profiling
  - Data Vault Design
- The middle:
  - Load and Stage
  - Raw Vault
  - Business Vault

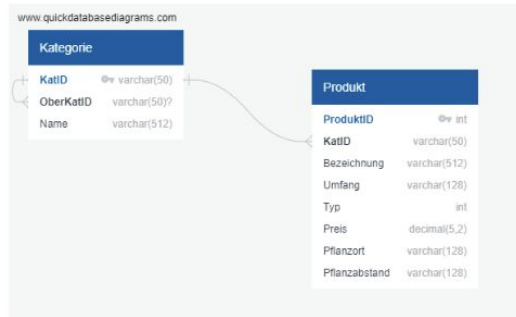


# Use cases 1 + 2

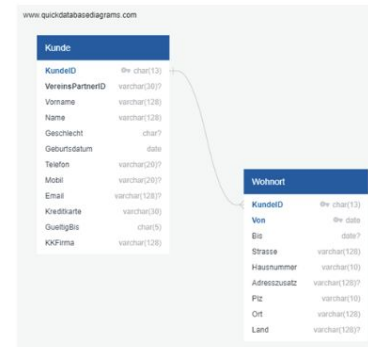
## Hierarchical Links & Multi-Active Satellites

Test cases in the data model  
hierarchical link

1



Test cases in the data model  
multi-active satellite



2

nur lösungs-  
beschreibung

# Use cases 3 + 4

## Relationships, Driving Keys and m:n Tables

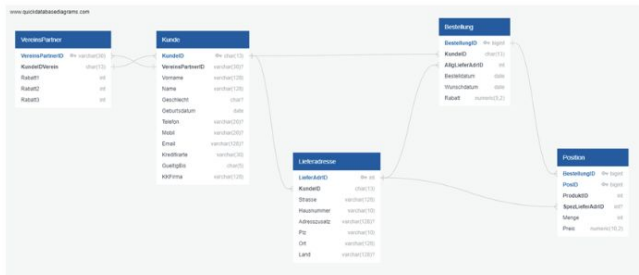
### Test cases in the data model Identifying Relationship and Driving Keys

3

The relationship between ORDER (Bestellung) and POSITION (Position) cannot change. The key situation makes every change a deletion and a new creation.

All other relationships can change. The test cases are all implemented on the foreign key in CUSTOMER (Kunde) to ASSOCIATION PARTNER (VereinsPartner). The following situations occur here:

- the foreign key is optional and therefore also NULL
- the foreign key changes between ASSOCIATION PARTNERS
- the foreign key changes from „valid“ to „invalid“ - and in some cases then even back to „valid“ again

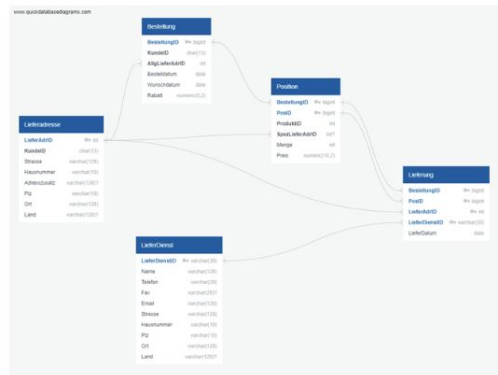


### Test cases in the data model m:n table without own key

4

The table of DELIVERIES (Lieferung) does not have its own primary key and usually only occurs once, as only successful deliveries are transmitted to the DWH.

This can be solved in several possible ways:  
keyed instance, transactional link, dependent child link, ...



# Use cases 5 + 6

## Multiple Sources and Historicised Ref. Tables

### Test cases in the data model Integration of the order

5

- We have 2 sources. These are to be integrated into ORDER (Bestellung) and POSITION (Position) Hubs.
- The primary key for order (BestellungID) of both systems are simply incremented. The Roadshow has an additional 'RS' in for safe differentiation.
- One could argue that integration should take place in the Business Vault. The point here is to present the capability for early integration.



6

### Test cases in the data model (Historicised) reference table

- There is one
  - Reference table
  - historicised reference table
- Both are available as csv file

## 5. Early Integration (of the order)

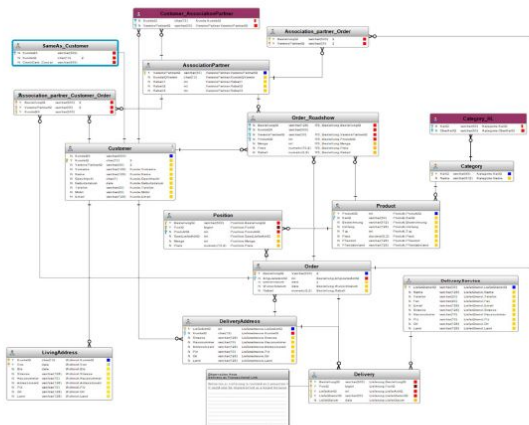
### Data Vault design

In order to integrate the different sources from the webshop and the roadshow to the order and position hubs, we do apply source mapping to both tables with the subset of the columns arriving from the different source tables. The Order\_Roadshow link keeps the history of the roadshow order.

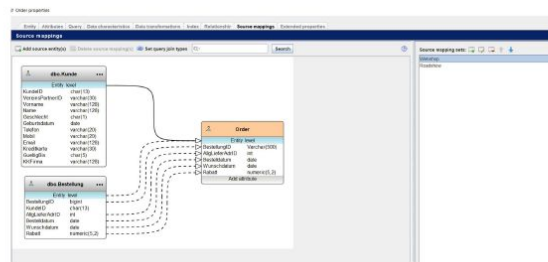
An additional KundeBK column is added as a derived column to support different customers from the two sources as same as link

KundenBK data transformation

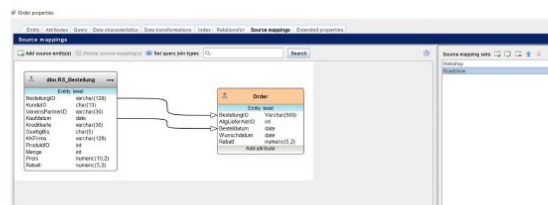
COALESCE(KundeID, CONCAT(KKFirma, '|', Kreditkarte, '|', GueltigBis))



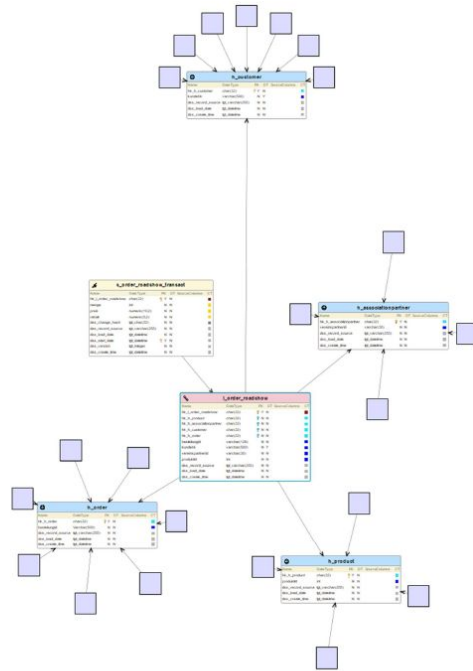
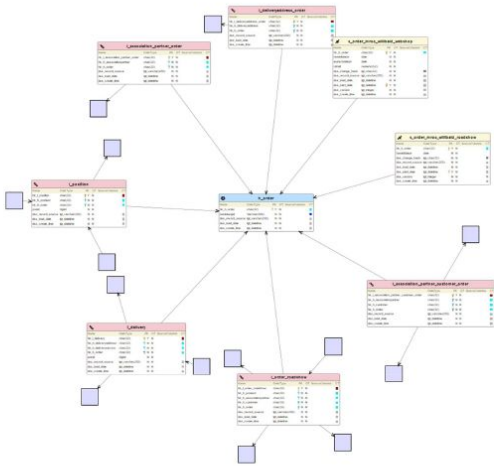
### Webshop source mapping for order



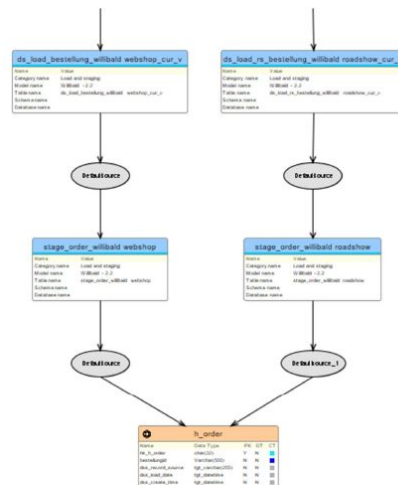
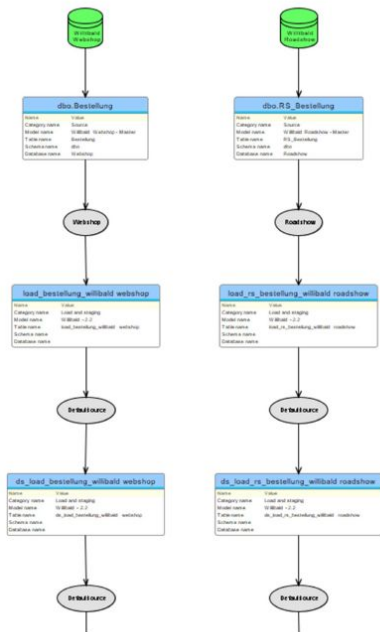
### Roadshow source mapping for order



Raw Vault generated automatically by 3D



Load and Staging generated automatically by 3D  
The order table is sourced from webshop and Roadshow.



### Data Lineage in WhereScape RED

Note: WhereScape Red uses the XML File to build the data warehouse objects and all code routines to ingest the load tables, ODS layers, views, stage tables and the raw vault or the business vault layers.



# Use cases 7-10

## Test cases in the data part 1 of 2

### Test cases in the data

- **Duplicates in the loading data**
  - There are 2 types of duplicates in PRODUCT (Produkt). In the first case, it is an actual duplicate (ProduktID 20), all attributes are the same. In the second case, the attributes contradict each other (ProduktID 21).
  - Here we just want to see how it is dealt with.
- **Rows without business key**
  - In DELIVERY SERVICE (LieferDienst), there are records with valid values without a key. Here, too, the only question is how to deal with this.
- **Changes in CUSTOMER (Kunde)**
  - A very simple test case, the data in the customer (KundeID 107) is changed to a value in delivery 2 and get the values from delivery 1 again in delivery 3.
- **Deletions in CUSTOMER (Kunde)**
  - There are cases in the customer where the customer (KundeID 70) was deleted in the second delivery and reappears in the third delivery.

7

8

9

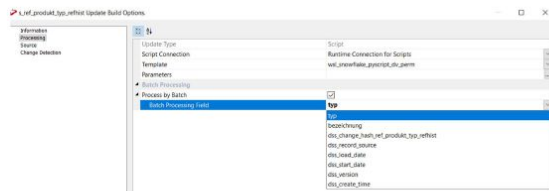
10

## 7. Duplicate Data

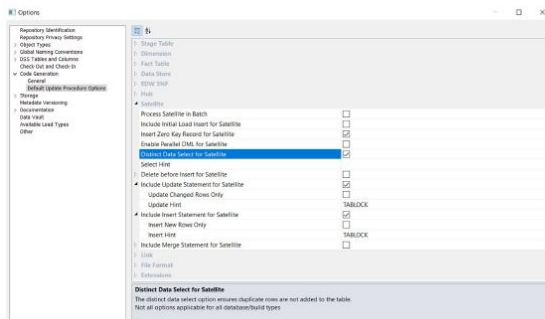
### Option 1

Technical DV duplicate data - two identical rows arriving at the same batch timestamp

- a) use a batch id different from the business key or add a version column using for example a SQ window function like `row_number()` to insert



- b) In WhereScape RED there is an option for each object to set a distinct property on an object.



### Option 2

Example two rows with the same business key but different data arrive on the same batch. Meaning the source is erroneous.

Solutions:

1. Change the satellite as a multi-active satellite, and load the received data as it is.
2. Communicate with the team delivering the source data and let them fix it.
3. Apply soft rules in Business Vault to fix the data

Technically, one can use an error handling to create an error reporting table for the team responsible for cleansing the data.

Add a stage entity with a `row_number` column.

```
ROW_NUMBER() OVER (PARTITION BY <business_key>
                    ORDER BY <business_key>)
```

Records with a value 1 from the stage go into the vault, and all other records go to an error reporting table.

## 8. Rows without business key

If there is a need to maintain an "unknown" business key in a hub, then a zero key hashkey can be inserted into the hub after creating the hub.

Example: deliveryService hub without BK

Column Name	Source Name	Table Type	Source Table	Source Column	Format	Null	Key	Alt	PK
id	id	Table	id	id	String	N	N		
name	name	Table	name	name	String	N	N		
description	description	Table	description	description	String	N	N		
created_date	created_date	Table	created_date	created_date	String	N	N		
last_date	last_date	Table	last_date	last_date	String	N	N		
total_date	total_date	Table	total_date	total_date	String	N	N		
data_date	data_date	Table	data_date	data_date	String	N	N		

-- Example: Generate the hash key for the business key "UNKNOWN"

SELECT

```
CONVERT(binary(16),HASHBYTES('MDS',
ISNULL(RTRIM(CONVERT(NVARCHAR(100),NULL)), 'UNKNOWN'+*')'),2)
```

Generated hash key: 0xD7F91BE8665F968B127DE83FB3777C0B

- OF -

SELECT

```
CONVERT(char(32),HASHBYTES('MDS',
ISNULL(RTRIM(CONVERT(NVARCHAR(100),NULL)), 'UNKNOWN'+*')'),2)
```

Generated hash key: D7F91BE8665F968B127DE83FB3777C0B

Insert the zero key record as a union or direct insert:

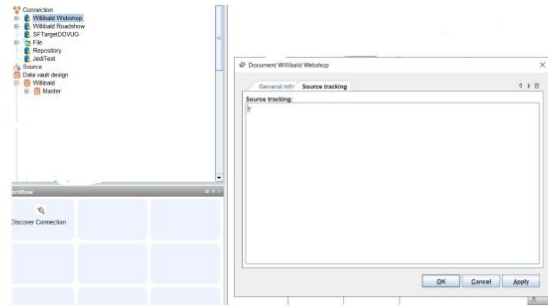
```
1 use WAREHOUSE;
2 use DATABASE;
3
4 INSERT INTO deliveryService (id, name, description, created_date, last_date, total_date, data_date)
5 VALUES (D7F91BE8665F968B127DE83FB3777C0B, 'UNKNOWN', 'SYSTEM', CAST('2023-05-16' AS TIMEZONE), CAST('2023-05-16' AS TIMEZONE), CAST('2023-05-16' AS TIMEZONE), CAST('2023-05-16' AS TIMEZONE));
```

Row	ID	NAME	DESCRIPTION	CREATED_DATE	LAST_DATE	TOTAL_DATE	DATA_DATE
48	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
70	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
71	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
72	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
73	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
74	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
75	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
76	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
77	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
78	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
79	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
80	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000
81	D7F91BE8665F968B127DE83FB3777C0B	UNKNOWN	SYSTEM	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000	2023-05-16 16:22:00.000

## 10. Deletion of Business Keys

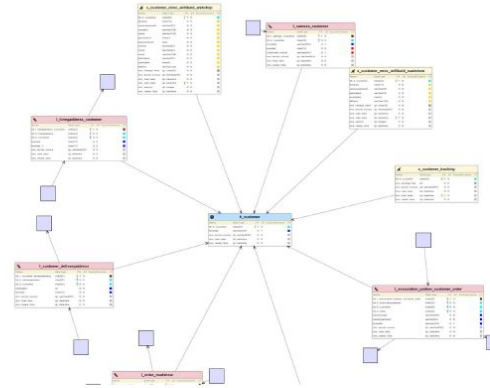
### Data Vault design

All connections which need source tracking must be configured for tracking and set as Y:



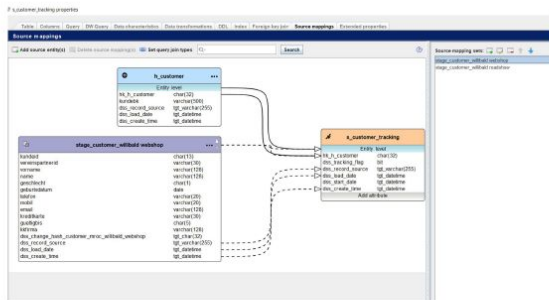
### Raw Vault generated automatically by 3D

Tracking satellites do not maintain dss\_change\_hash nor dss\_version columns. A new dss\_tracking\_flag attribute is added..

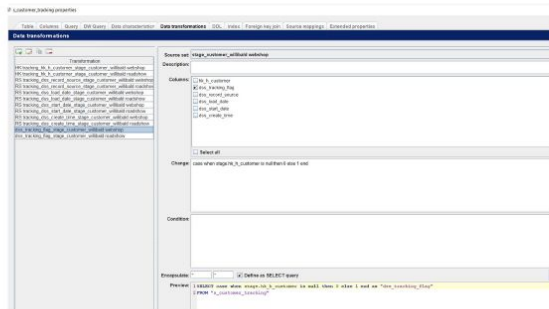


Load and Staging generated automatically by 3D

The load and stage mechanism is the same as other satellites.

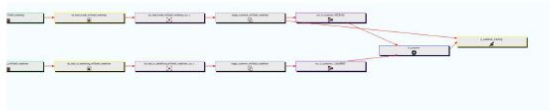


dss\_tracking transformation logic:



## Data Lineage in WhereScape RED

Note: WhereScape Red uses the XML File to build the data warehouse objects and all code routines to ingest the load tables, ODS layers, views, stage tables and the raw vault or the business vault layers.



### Sample Data of Satellite tracking for customer

[illegible]

# Use cases 11-13

## Test cases in the data part 2 of 2

### Test cases in the data II

- DELIVERY ADDRESSES (Lieferadresse) without CUSTOMER (Kunde)
  - The first delivery contains delivery addresses for which there is no record with the same customer ID (KundeID) in CUSTOMER (Kunde - KundeIDs 999, 998 and 997).
- Deletions of ORDERS (Bestellung)
  - The orders are relevant for counting and are deleted during the dates of deliveries.
  - Between period 1 and 2 the orderIDs (BestellungID) 99, 220 and 465.
  - Between periods 2 and 3 the orderIDs (BestellungID) 1470 and 1288.
- Changes in the dimensions
  - The hierarchy of the product CATEGORY (Kategorie) changes completely with both deliveries. So we have 3 different product hierarchies. These are to be displayed as as-what at the respective reporting time.

11

12

13

## Business Rules

15

1. Standardise ORDERS (Bestellung) of the ASSOCIATION PARTNERS (VereinsPartner)  
The ROADSHOW ORDERS (RS\_Bestellung) are directly linked to the ASSOCIATION PARTNERS (VereinsPartner). This must now be done for the orders of the association partner from the webshop. To do this, all orders of this customer are linked directly to the association partner using the association partner ID (VereinsPartnerID) of this CUSTOMER (Kunde).
2. Roadshow: assigning ORDERS (RS\_Bestellung) to CUSTOMERS (Kunde)  
If the customer ID (KundeID) is missing in the ROADSHOW ORDERS (RS\_Bestellung), the credit card, CC-Company (KKFirma) and the valid-to (GueltigBis) can be used to identify the customer.

## 15. Business Rules

Business rules and data transformations can be added at any time in WhereScape RED.

To standardize the Orders of the Association Partners from the webshop a join is created between the Customer table and the webshop Orders table and it is filtered to only include Association Partners.

A rule is implemented to identify customers without customer\_id in the (Roadshow order) RS\_Bestellung based on their credit card, cc-company and valid-to data.

The connection between Webshop Orders and Association Partner was also modeled in 3D to generate the proper Link tables.

We also modeled a Sames-As Link table to resolve the alternate Customer ID based on credit card information.

### Additional Business Vault

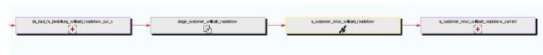
Satellites are insert-only and store historical changes. It is best practice to add the current and historical views on the top of the raw satellites. WhereScape is able to generate these two types of views for all the satellites automatically.

[illegible]

## Data Lineage in WhereScape RED

Note: WhereScape Red uses the XML File to build the data warehouse objects and all code routines to ingest the load tables, ODS layers, views, stage tables and the raw vault or the business vault layers.

**Current view**



### Historical view



### Sample Data

**Current satellite view**

[illegible]

### Historical satellite view

[illegible]

# Use cases 16-22

## Non-Data-Vault-Specific Requirements

16. Data  
Lineage

17. Error  
Handling /  
DQ Check

7

8

18.  
Orchestration

19.  
Deployment

20.  
Scheduling

21.  
supported  
dbs

22. installation  
requirements,  
etc.

# Get Hands On!



**90 minute Hands-on virtual lab July 12 17:00**

**90 minute Hands-on virtual lab July 19 17:00**



**Schedule a demo @ [www.wherescape.com](http://www.wherescape.com)**

Thank you



**WhereScape®**  
Data Automation