VAULTSPEED

# DDVUG
# Automation Challenge
# @TDWI Munich 2023

by VaultSpeed

VAULTSPEED

ABSTRACT ETL

# Introducing the next automation standard

## Model-driven

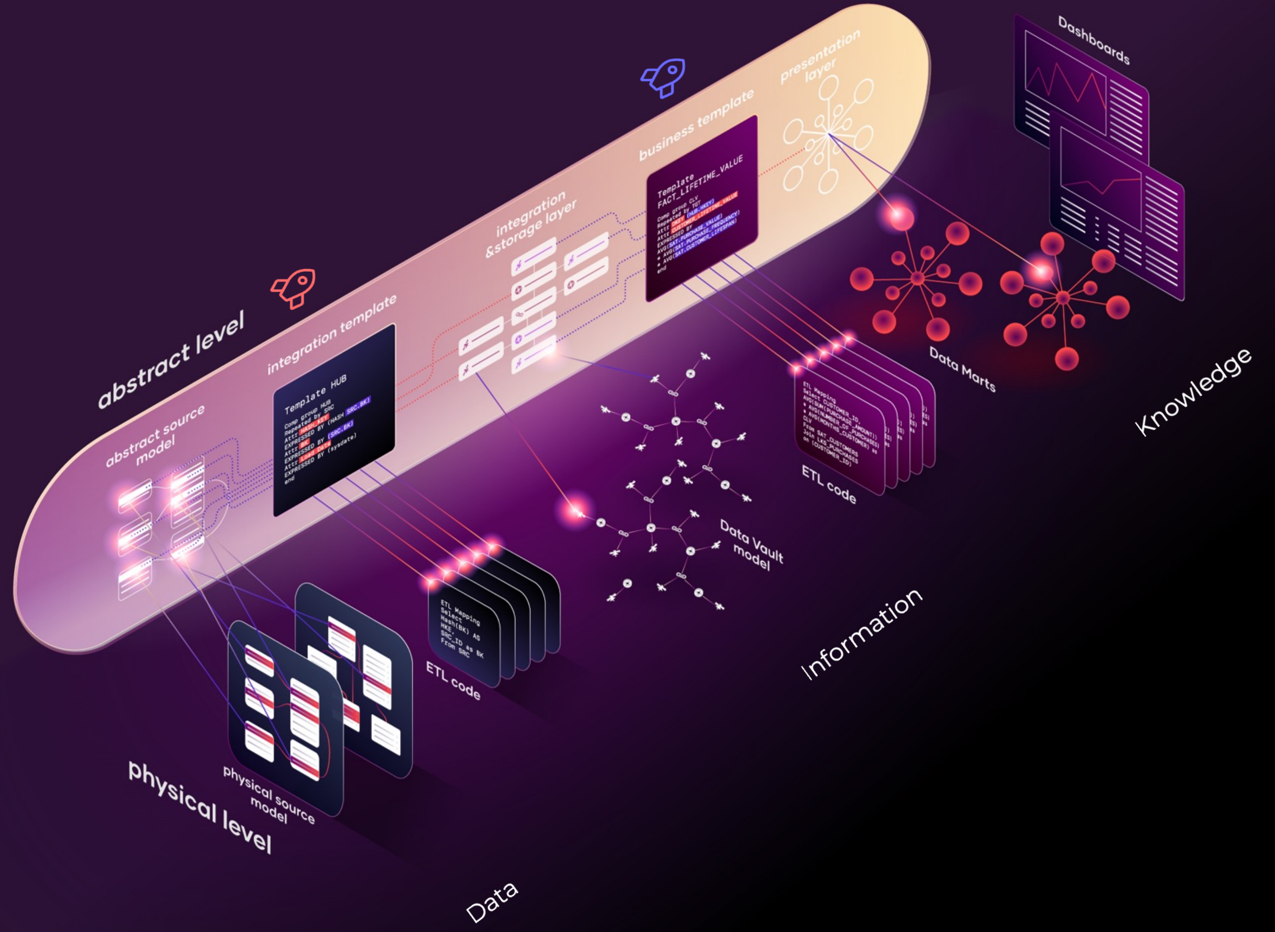
integrated data transformation patterns , not isolated data pipelines

## No code automation

Built-in integration templates to automate multi-source data integration, based on Data Vault 2.0 standards.

## Industry-specific metrics store

Customers write and share templates to automate custom business logic on top of the integration layer.



```
Template HUB
Comp group HUB
Repeated by SRC
Attr HASH_KEY
EXPRESSED BY (HASH SRC.BK)
Attr BK
EXPRESSED BY (SRC.BK)
Attr Load Date
EXPRESSED BY (sysdate)
end
```

```
Template
FACT_LIFETIME_VALUE
Comp group CLV
Repeated by TOT
Attr DKEY (HUB.HKEY)
Attr CUSTOMER_LIFETIME_VALUE
EXPRESSED BY
AVG(SAT.PURCHASE_VALUE)
* AVG(SAT.PURCHASE_FREQUENCY)
* AVG(SAT.CUSTOMER_LIFESPAN)
end
```

```
ETL Mapping
Select
Hash(BK) AS
HKE,
SRC_ID as BK
From SRC
```

```
ETL Mapping
Select CUSTOMER_ID,
AVG(SUM(PURCHASE_AMOUNT))
* AVG(NUMBER_OF_PURCHASES)
* AVG(MONTHS_CUSTOMER) as
CLV
From SAT_CUSTOMERS
Join LKS_PURCHASES
on (CUSTOMER_ID)
```

VAULTSPEED

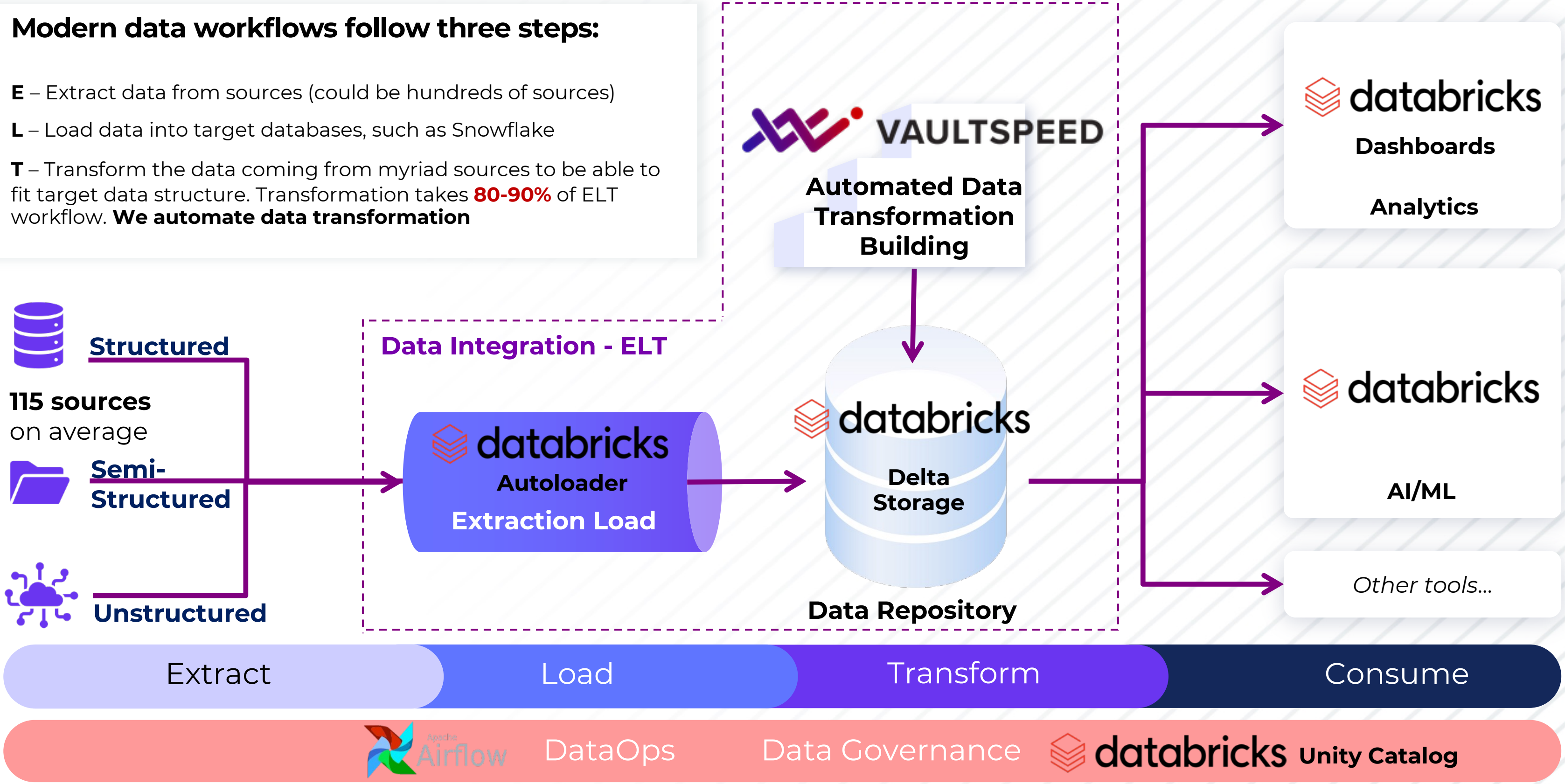
# Some decisions on tooling

**Modern data workflows follow three steps:**

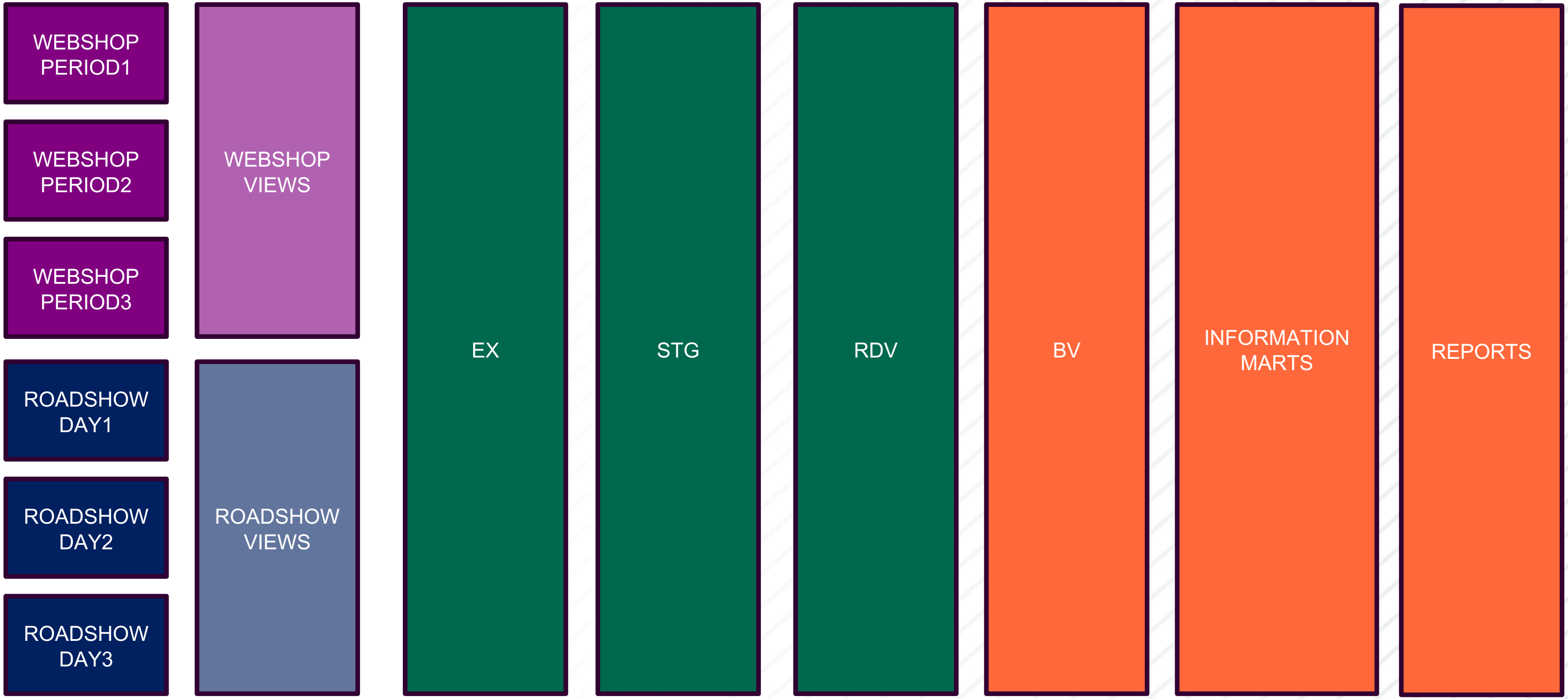
**E** – Extract data from sources (could be hundreds of sources)

**L** – Load data into target databases, such as Snowflake

**T** – Transform the data coming from myriad sources to be able to fit target data structure. Transformation takes **80-90%** of ELT workflow. **We automate data transformation**



# How the datasets were loaded



WEBSHOP PERIOD1

WEBSHOP PERIOD2

WEBSHOP PERIOD3

ROADSHOW DAY1

ROADSHOW DAY2

ROADSHOW DAY3

WEBSHOP VIEWS

ROADSHOW VIEWS

EX

STG

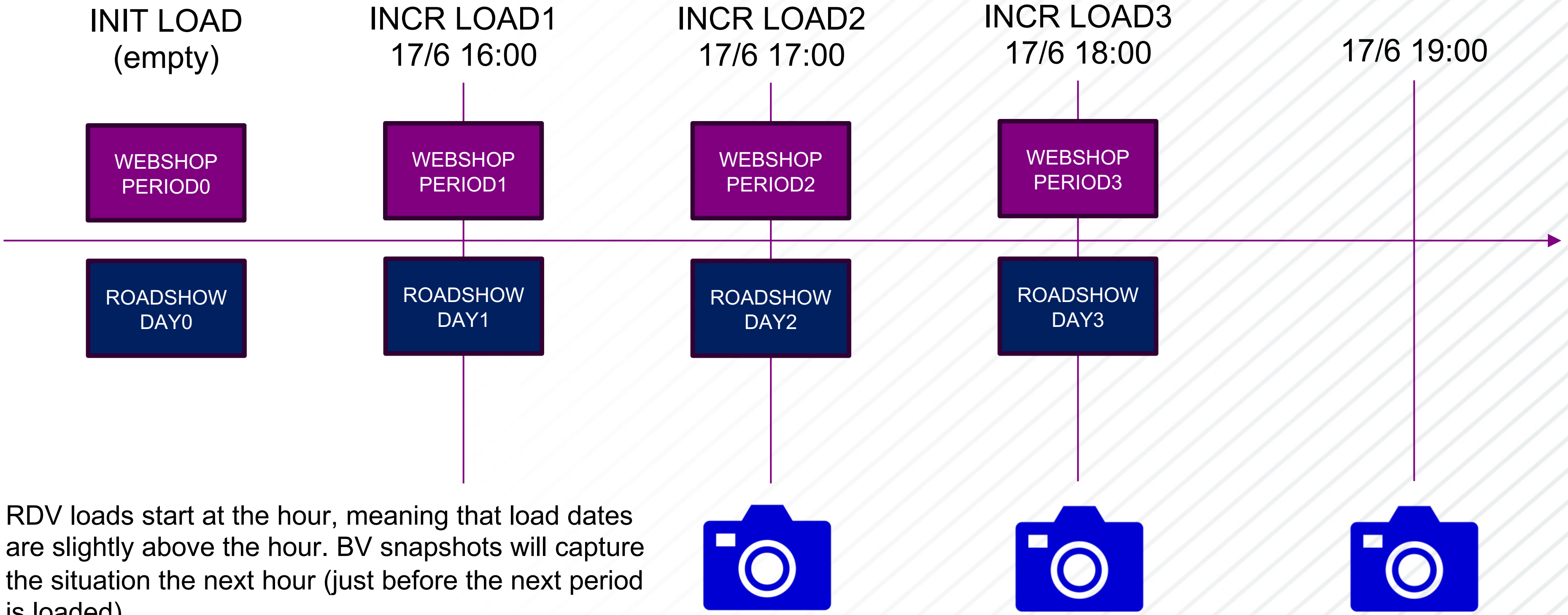
RDV

BV

INFORMATION MARTS

REPORTS

# How the datasets were loaded



INIT LOAD
(empty)

WEBSHOP
PERIOD0

ROADSHOW
DAY0

INCR LOAD1
17/6 16:00

WEBSHOP
PERIOD1

ROADSHOW
DAY1

INCR LOAD2
17/6 17:00

WEBSHOP
PERIOD2

ROADSHOW
DAY2

INCR LOAD3
17/6 18:00

WEBSHOP
PERIOD3

ROADSHOW
DAY3

17/6 19:00

RDV loads start at the hour, meaning that load dates are slightly above the hour. BV snapshots will capture the situation the next hour (just before the next period is loaded)

# Business view and reality don’t always match up

**Source data model**

Data
Facts

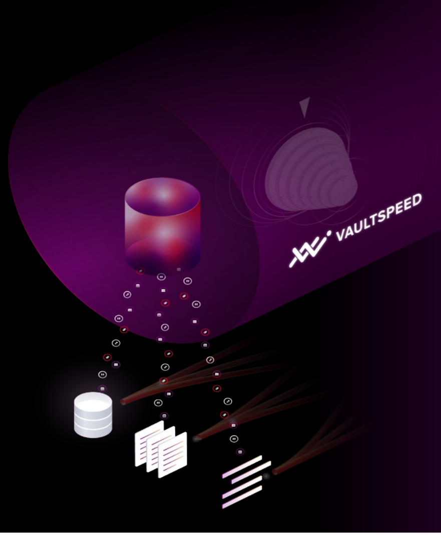
**Conceptual data model**

Views
Concepts

# Data-driven meets Model-driven

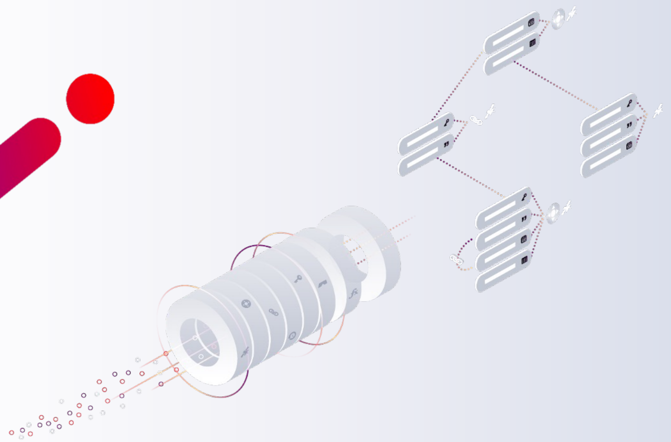
## Data-driven automation

Speed up building the physical data model using metadata harvested from your data sources



## Model-driven automation

Speed up building the physical data model by importing conceptual model metadata

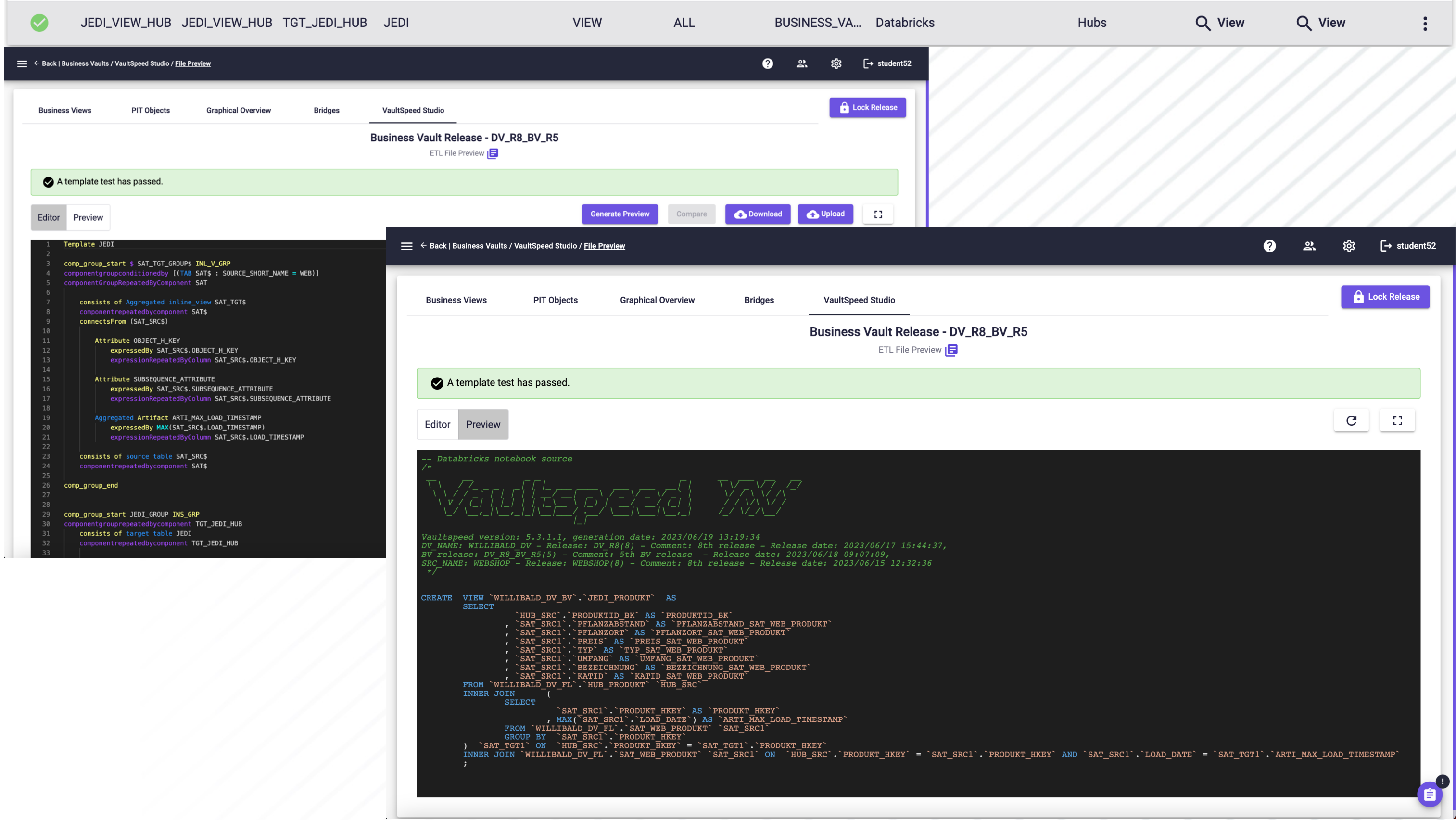


# 0. Jedi test

## Recreate the source from the DV

Using VaultSpeed studio, you can create anything on top of the Raw Vault.

For this use case, we built JEDI views that recreate the source.



JEDI_VIEW_HUB JEDI_VIEW_HUB TGT_JEDI_HUB JEDI VIEW ALL BUSINESS_VA... Databricks Hubs View View

Business Vault Release - DV_R8_BV_R5

ETL File Preview

A template test has passed.

```
Template JEDI

comp_group_start $ SAT_TGT_GROUP$ INL_V_GRP
componentgroupconditionedby [(TAB SAT$ : SOURCE_SHORT_NAME = WEB)]
componentGroupRepeatedByComponent SAT

    consists of Aggregated inline_view SAT_TGT$
    componentrepeatedbycomponent SAT$
    connectsFrom (SAT_SRC$)

        Attribute OBJECT_H_KEY
            expressedBy SAT_SRC$.OBJECT_H_KEY
            expressionRepeatedByColumn SAT_SRC$.OBJECT_H_KEY

        Attribute SUBSEQUENCE_ATTRIBUTE
            expressedBy SAT_SRC$.SUBSEQUENCE_ATTRIBUTE
            expressionRepeatedByColumn SAT_SRC$.SUBSEQUENCE_ATTRIBUTE

        Aggregated Artifact ARTI_MAX_LOAD_TIMESTAMP
            expressedBy MAX(SAT_SRC$.LOAD_TIMESTAMP)
            expressionRepeatedByColumn SAT_SRC$.LOAD_TIMESTAMP

    consists of source table SAT_SRC$
    componentrepeatedbycomponent SAT$

comp_group_end


comp_group_start JEDI_GROUP INS_GRP
componentgrouprepeatedbycomponent TGT_JEDI_HUB
    consists of target table JEDI
    componentrepeatedbycomponent TGT_JEDI_HUB

```

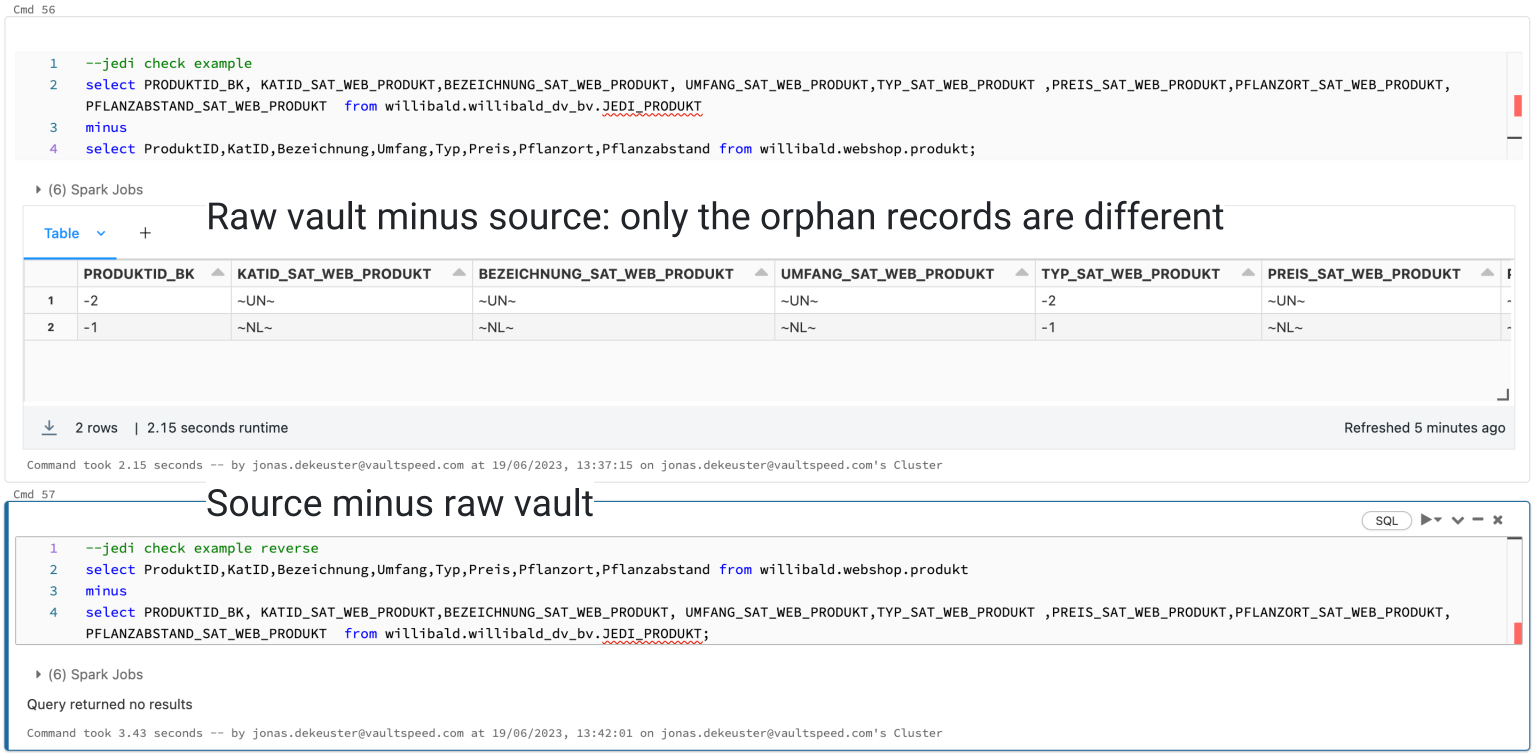
```
-- Databricks notebook source
/*

Vaultspeed version: 5.3.1.1, generation date: 2023/06/19 13:19:34
DV_NAME: WILLIBALD_DV - Release: DV_R8(8) - Comment: 8th release - Release date: 2023/06/17 15:44:37,
BV release: DV_R8_BV_R5(5) - Comment: 5th BV release  - Release date: 2023/06/18 09:07:09,
SRC_NAME: WEBSHOP - Release: WEBSHOP(8) - Comment: 8th release - Release date: 2023/06/15 12:32:36
 */


CREATE  VIEW `WILLIBALD_DV_BV`.`JEDI_PRODUKT`  AS
        SELECT
                `HUB_SRC`.`PRODUKTID_BK` AS `PRODUKTID_BK`
              , `SAT_SRC1`.`PFLANZABSTAND` AS `PFLANZABSTAND_SAT_WEB_PRODUKT`
              , `SAT_SRC1`.`PFLANZORT` AS `PFLANZORT_SAT_WEB_PRODUKT`
              , `SAT_SRC1`.`PREIS` AS `PREIS_SAT_WEB_PRODUKT`
              , `SAT_SRC1`.`TYP` AS `TYP_SAT_WEB_PRODUKT`
              , `SAT_SRC1`.`UMFANG` AS `UMFANG_SAT_WEB_PRODUKT`
              , `SAT_SRC1`.`BEZEICHNUNG` AS `BEZEICHNUNG_SAT_WEB_PRODUKT`
              , `SAT_SRC1`.`KATID` AS `KATID_SAT_WEB_PRODUKT`
        FROM `WILLIBALD_DV_FL`.`HUB_PRODUKT` `HUB_SRC`
        INNER JOIN      (
              SELECT
                        `SAT_SRC1`.`PRODUKT_HKEY` AS `PRODUKT_HKEY`
                      , MAX(`SAT_SRC1`.`LOAD_DATE`) AS `ARTI_MAX_LOAD_TIMESTAMP`
              FROM `WILLIBALD_DV_FL`.`SAT_WEB_PRODUKT` `SAT_SRC1`
              GROUP BY  `SAT_SRC1`.`PRODUKT_HKEY`
        )  `SAT_TGT1` ON  `HUB_SRC`.`PRODUKT_HKEY` = `SAT_TGT1`.`PRODUKT_HKEY`
        INNER JOIN `WILLIBALD_DV_FL`.`SAT_WEB_PRODUKT` `SAT_SRC1` ON  `HUB_SRC`.`PRODUKT_HKEY` = `SAT_SRC1`.`PRODUKT_HKEY` AND `SAT_SRC1`.`LOAD_DATE` = `SAT_TGT1`.`ARTI_MAX_LOAD_TIMESTAMP`
        ;
```

# 0. Jedi test

## Recreate the source from the DV

Using VaultSpeed studio, you can create anything on top of the Raw Vault.

For this use case, we built JEDI views that recreate the source.



Cmd 56

```
--jedi check example
select PRODUKTID_BK, KATID_SAT_WEB_PRODUKT,BEZEICHNUNG_SAT_WEB_PRODUKT, UMFANG_SAT_WEB_PRODUKT,TYP_SAT_WEB_PRODUKT ,PREIS_SAT_WEB_PRODUKT,PFLANZORT_SAT_WEB_PRODUKT,
PFLANZABSTAND_SAT_WEB_PRODUKT  from willibald.willibald_dv_bv.JEDI_PRODUKT
minus
select ProduktID,KatID,Bezeichnung,Umfang,Typ,Preis,Pflanzort,Pflanzabstand from willibald.webshop.produkt;
```

Raw vault minus source: only the orphan records are different

| | PRODUKTID_BK | KATID_SAT_WEB_PRODUKT | BEZEICHNUNG_SAT_WEB_PRODUKT | UMFANG_SAT_WEB_PRODUKT | TYP_SAT_WEB_PRODUKT | PREIS_SAT_WEB_PRODUKT | P |
|---|---|---|---|---|---|---|---|
| 1 | -2 | ~UN~ | ~UN~ | ~UN~ | -2 | ~UN~ | - |
| 2 | -1 | ~NL~ | ~NL~ | ~NL~ | -1 | ~NL~ | - |

2 rows | 2.15 seconds runtime

Refreshed 5 minutes ago

```
Command took 2.15 seconds -- by jonas.dekeuster@vaultspeed.com at 19/06/2023, 13:37:15 on jonas.dekeuster@vaultspeed.com's Cluster
```

Cmd 57

Source minus raw vault

```
--jedi check example reverse
select ProduktID,KatID,Bezeichnung,Umfang,Typ,Preis,Pflanzort,Pflanzabstand from willibald.webshop.produkt
minus
select PRODUKTID_BK, KATID_SAT_WEB_PRODUKT,BEZEICHNUNG_SAT_WEB_PRODUKT, UMFANG_SAT_WEB_PRODUKT,TYP_SAT_WEB_PRODUKT ,PREIS_SAT_WEB_PRODUKT,PFLANZORT_SAT_WEB_PRODUKT,
PFLANZABSTAND_SAT_WEB_PRODUKT  from willibald.willibald_dv_bv.JEDI_PRODUKT;
```

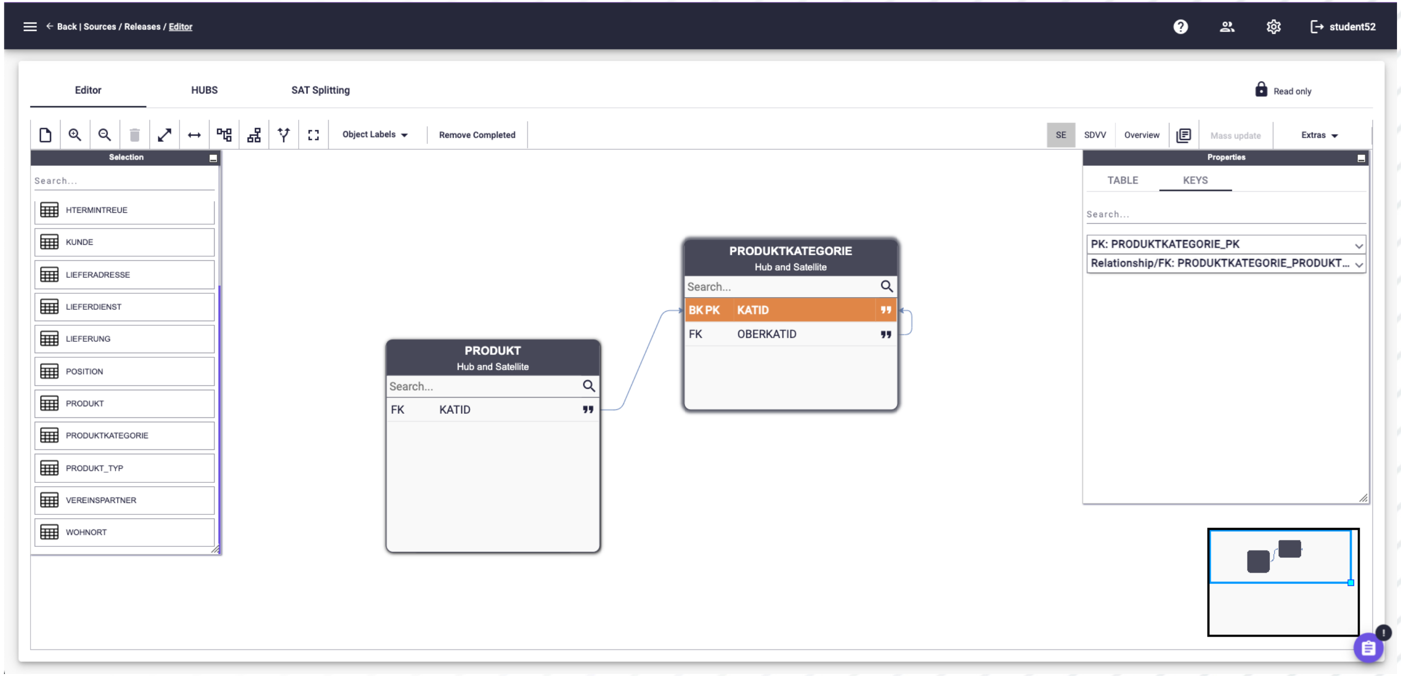
Query returned no results

```
Command took 3.43 seconds -- by jonas.dekeuster@vaultspeed.com at 19/06/2023, 13:42:01 on jonas.dekeuster@vaultspeed.com's Cluster
```

# 1. Hierarchical link

**Raw Vault modeling**

Self reference on Product category table



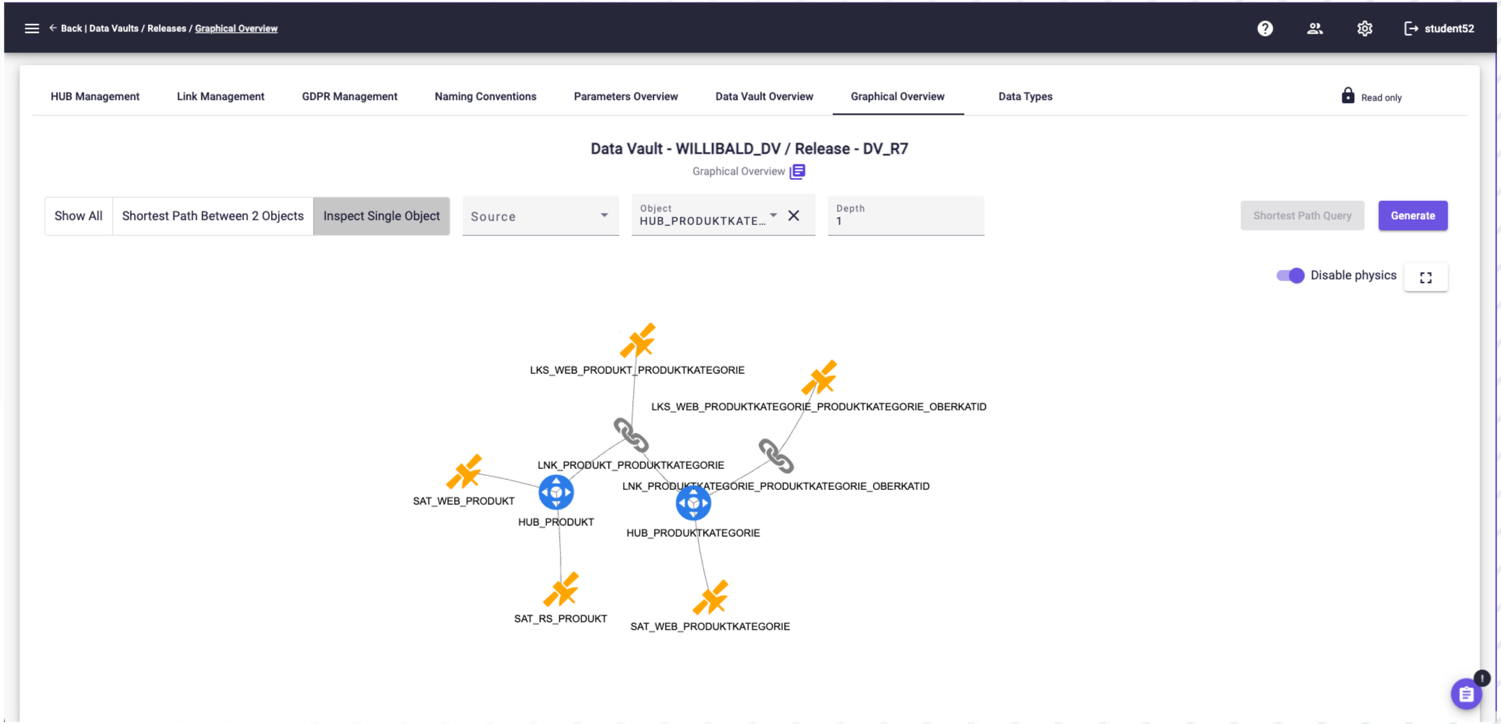
PK: PRODUKTKATEGORIE_PK

Relationship/FK: PRODUKTKATEGORIE_PRODUKT...

# 1. Hierarchical link

## Raw Vault modeling

The result in the raw vault is a hub with a link that contains the hierarchical data



Data Vault - WILLIBALD_DV / Release - DV_R7

Object: HUB_PRODUKTKATE…

Depth: 1

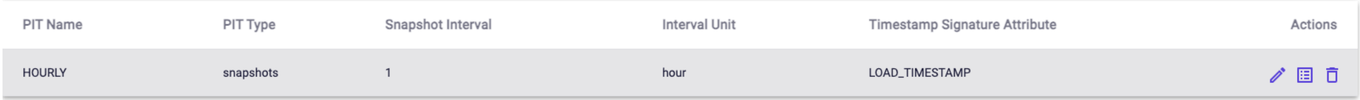
Disable physics

# 1. Hierarchical link

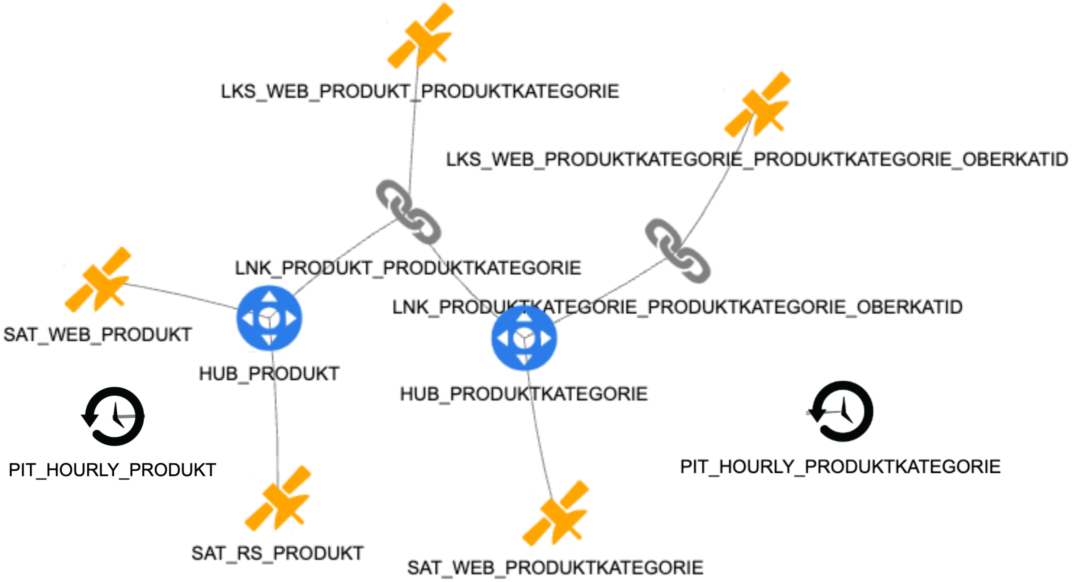
## Business Vault

Automatic creation of 2 snapshot PITs on each Hub.

Using the PITs we build 2 Dimensions (Product and Product Category) That we can combine in a view a bridge combining the 2 hubs and links.



| PIT Name | PIT Type | Snapshot Interval | Interval Unit | Timestamp Signature Attribute | Actions |
|---|---|---|---|---|---|
| HOURLY | snapshots | 1 | hour | LOAD_TIMESTAMP | ✎ ▤ 🗑 |


LKS_WEB_PRODUKT_PRODUKTKATEGORIE
LKS_WEB_PRODUKTKATEGORIE_PRODUKTKATEGORIE_OBERKATID
LNK_PRODUKT_PRODUKTKATEGORIE
LNK_PRODUKTKATEGORIE_PRODUKTKATEGORIE_OBERKATID
SAT_WEB_PRODUKT
HUB_PRODUKT
HUB_PRODUKTKATEGORIE
PIT_HOURLY_PRODUKT
PIT_HOURLY_PRODUKTKATEGORIE
SAT_RS_PRODUKT
SAT_WEB_PRODUKTKATEGORIE

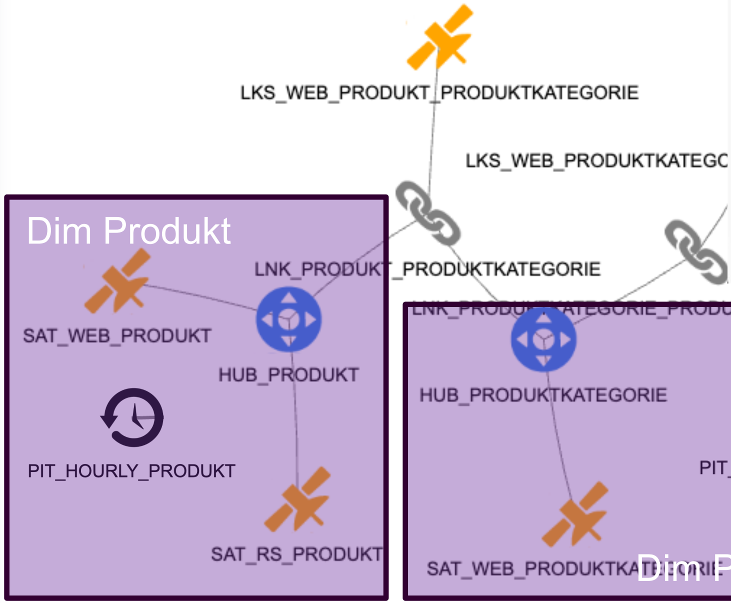
VAULTSPEED

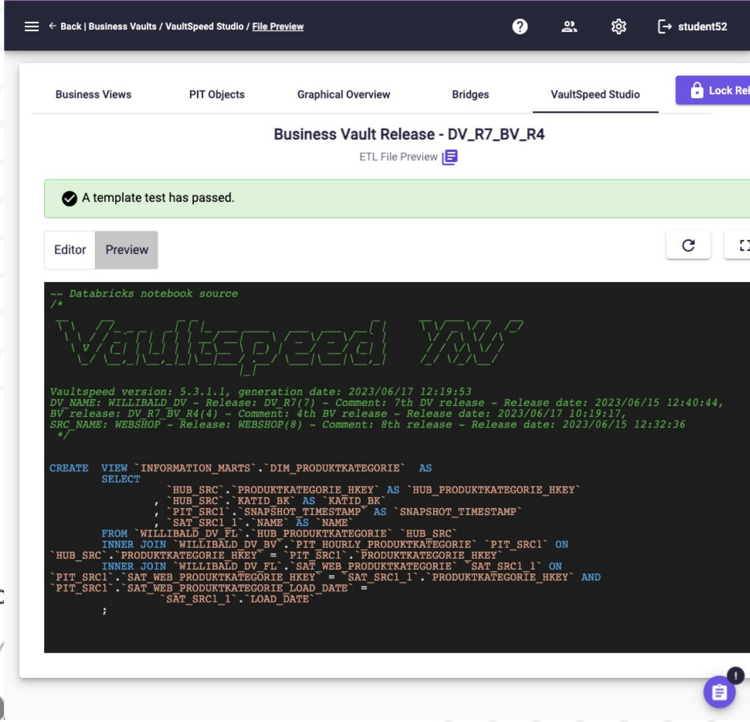
# 1. Hierarchical link

**Presentation Layer?**
VaultSpeed Studio template for dimension views based on snapshot pits

The Snapshot Pits are perfect to watch the data at a certain point in time.

We'll show how that is convenient to track changes in use case 13.





Business Vault Release - DV_R7_BV_R4

A template test has passed.

```
-- Databricks notebook source
/*

Vaultspeed version: 5.3.1.1, generation date: 2023/06/17 12:19:53
DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7th DV release - Release date: 2023/06/15 12:40:44,
BV release: DV_R7_BV_R4(4) - Comment: 4th BV release - Release date: 2023/06/17 10:19:17,
SRC_NAME: WEBSHOP - Release: WEBSHOP(8) - Comment: 8th release - Release date: 2023/06/15 12:32:36
 */


CREATE  VIEW `INFORMATION_MARTS`.`DIM_PRODUKTKATEGORIE`  AS
        SELECT
                `HUB_SRC`.`PRODUKTKATEGORIE_HKEY` AS `HUB_PRODUKTKATEGORIE_HKEY`
               , `HUB_SRC`.`KATID_BK` AS `KATID_BK`
               , `PIT_SRC1`.`SNAPSHOT_TIMESTAMP` AS `SNAPSHOT_TIMESTAMP`
               , `SAT_SRC1_1`.`NAME` AS `NAME`
        FROM `WILLIBALD_DV_FL`.`HUB_PRODUKTKATEGORIE` `HUB_SRC`
        INNER JOIN `WILLIBALD_DV_BV`.`PIT_HOURLY_PRODUKTKATEGORIE` `PIT_SRC1` ON
`HUB_SRC`.`PRODUKTKATEGORIE_HKEY` = `PIT_SRC1`.`PRODUKTKATEGORIE_HKEY`
        INNER JOIN `WILLIBALD_DV_FL`.`SAT_WEB_PRODUKTKATEGORIE` `SAT_SRC1_1` ON
`PIT_SRC1`.`SAT_WEB_PRODUKTKATEGORIE_HKEY` = `SAT_SRC1_1`.`PRODUKTKATEGORIE_HKEY` AND
`PIT_SRC1`.`SAT_WEB_PRODUKTKATEGORIE_LOAD_DATE` =
                `SAT_SRC1_1`.`LOAD_DATE`
        ;
```

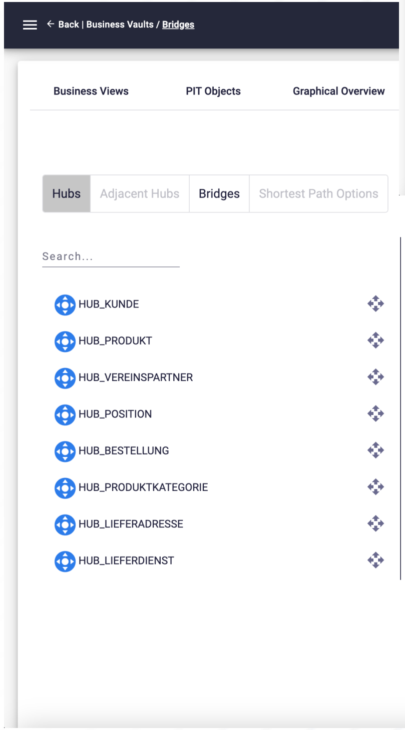
VAULTSPEED

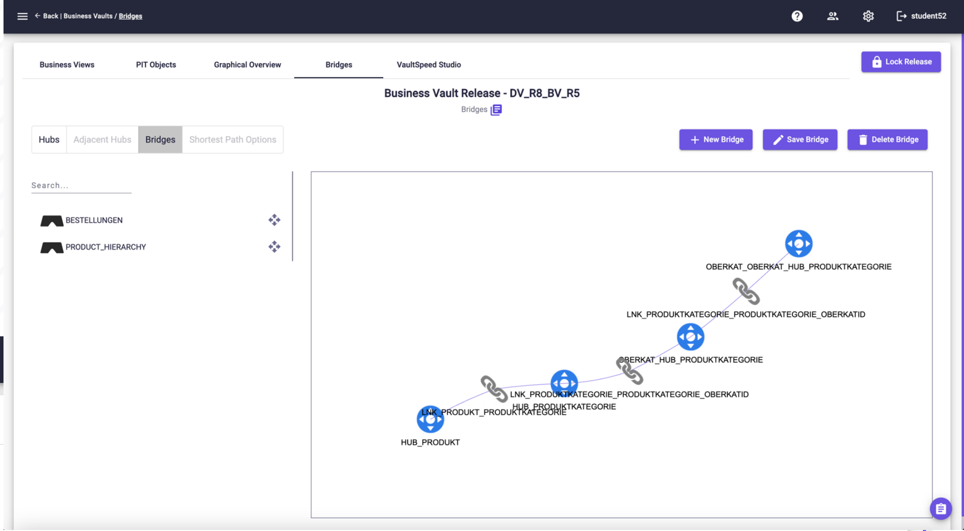
# 1. Hierarchical link

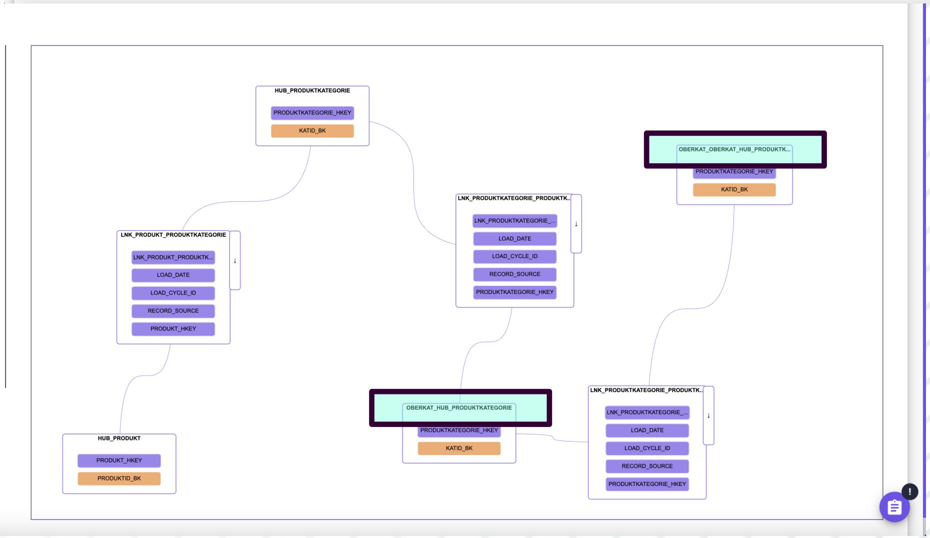
## Business Vault

A bridge combining the 2 hubs and links we need to build the product hierarchy

In the detailed view, notice how we prefixed multiple instances of the same hub for them to be able to exist in the same bridge more then once.



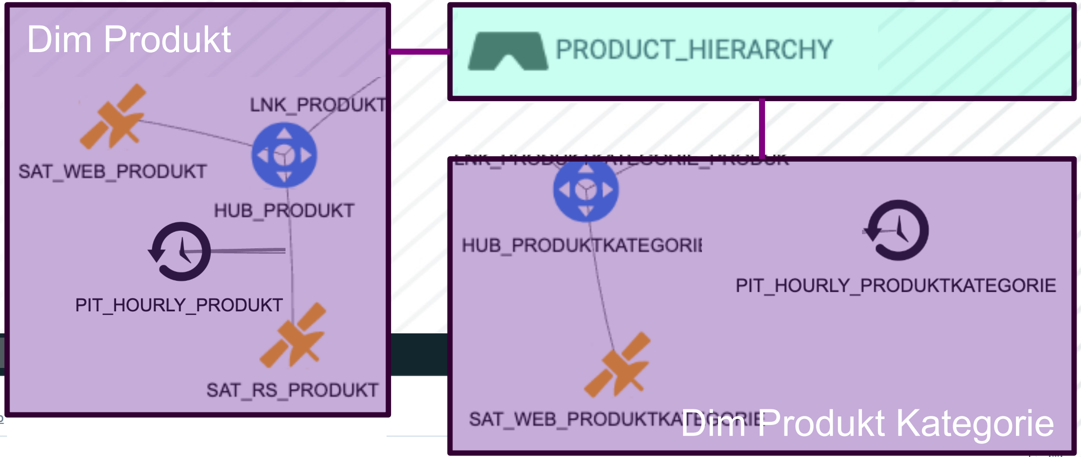

Back | Business Vaults / Bridges
student52
Business Views
PIT Objects
Graphical Overview
Bridges
VaultSpeed Studio
Lock Release
Business Vault Release - DV_R8_BV_R5
Bridges
Hubs
Adjacent Hubs
Bridges
Shortest Path Options
New Bridge
Save Bridge
Delete Bridge
Search...
BESTELLUNGEN
PRODUCT_HIERARCHY
OBERKAT_OBERKAT_HUB_PRODUKTKATEGORIE
LNK_PRODUKTKATEGORIE_PRODUKTKATEGORIE_OBERKATID
OBERKAT_HUB_PRODUKTKATEGORIE
LNK_PRODUKTKATEGORIE_PRODUKTKATEGORIE_OBERKATID
HUB_PRODUKTKATEGORIE
LNK_PRODUKT_PRODUKTKATEGORIE
HUB_PRODUKT

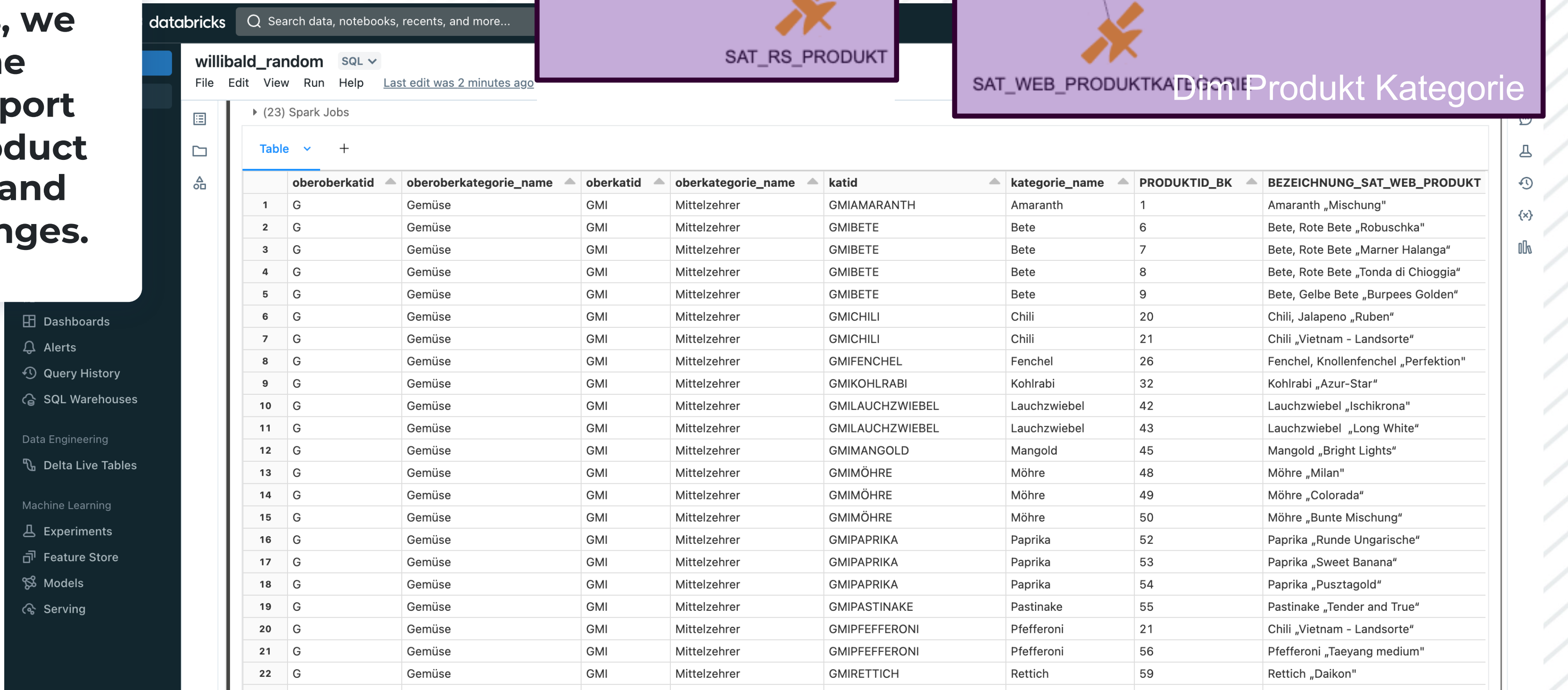


VAULTSPEED

# 1. Hierarchical link

Using these structures, we have all the tools to report on the product hierarchy and on its changes.





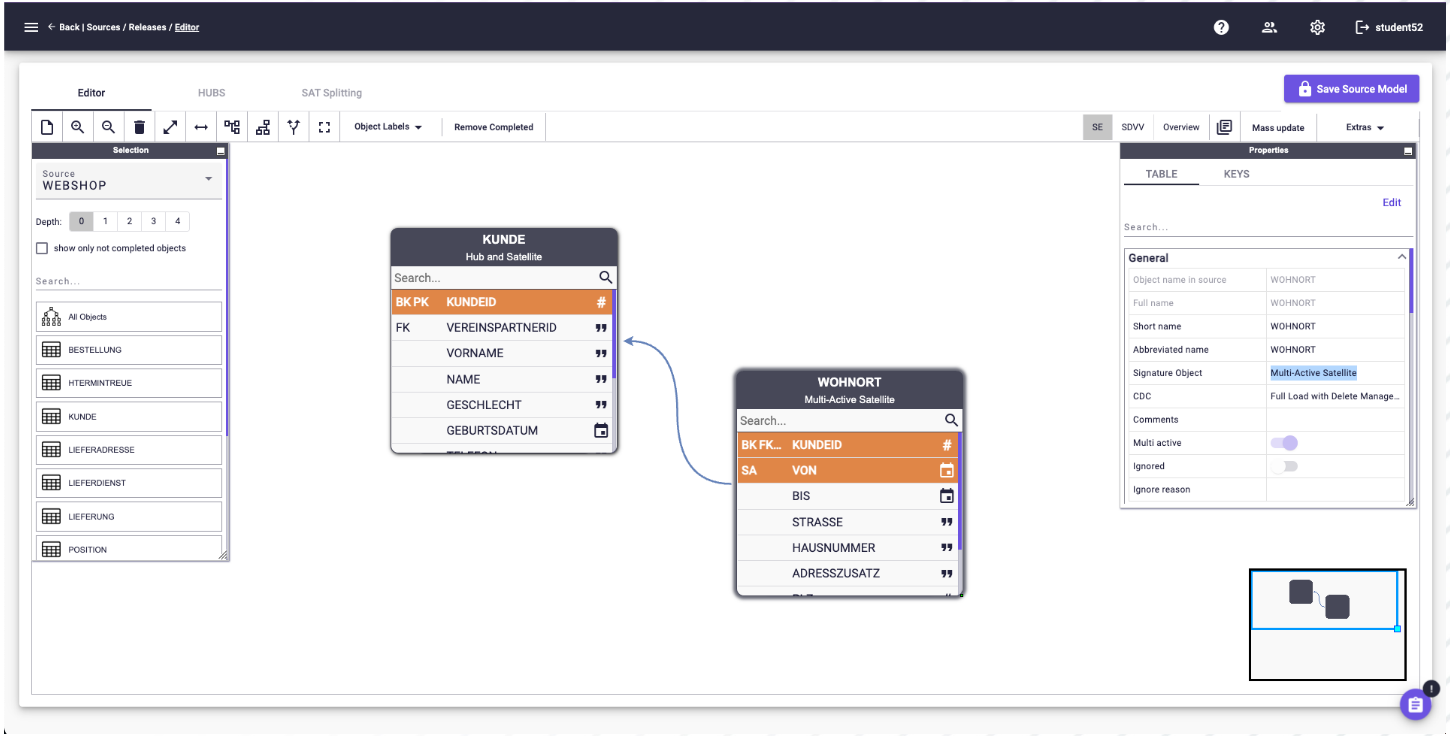
(23) Spark Jobs

| | oberoberkatid | oberoberkategorie_name | oberkatid | oberkategorie_name | katid | kategorie_name | PRODUKTID_BK | BEZEICHNUNG_SAT_WEB_PRODUKT |
|---|---|---|---|---|---|---|---|---|
| 1 | G | Gemüse | GMI | Mittelzehrer | GMIAMARANTH | Amaranth | 1 | Amaranth „Mischung" |
| 2 | G | Gemüse | GMI | Mittelzehrer | GMIBETE | Bete | 6 | Bete, Rote Bete „Robuschka" |
| 3 | G | Gemüse | GMI | Mittelzehrer | GMIBETE | Bete | 7 | Bete, Rote Bete „Marner Halanga“ |
| 4 | G | Gemüse | GMI | Mittelzehrer | GMIBETE | Bete | 8 | Bete, Rote Bete „Tonda di Chioggia“ |
| 5 | G | Gemüse | GMI | Mittelzehrer | GMIBETE | Bete | 9 | Bete, Gelbe Bete „Burpees Golden“ |
| 6 | G | Gemüse | GMI | Mittelzehrer | GMICHILI | Chili | 20 | Chili, Jalapeno „Ruben“ |
| 7 | G | Gemüse | GMI | Mittelzehrer | GMICHILI | Chili | 21 | Chili „Vietnam - Landsorte“ |
| 8 | G | Gemüse | GMI | Mittelzehrer | GMIFENCHEL | Fenchel | 26 | Fenchel, Knollenfenchel „Perfektion" |
| 9 | G | Gemüse | GMI | Mittelzehrer | GMIKOHLRABI | Kohlrabi | 32 | Kohlrabi „Azur-Star“ |
| 10 | G | Gemüse | GMI | Mittelzehrer | GMILAUCHZWIEBEL | Lauchzwiebel | 42 | Lauchzwiebel „Ischikrona" |
| 11 | G | Gemüse | GMI | Mittelzehrer | GMILAUCHZWIEBEL | Lauchzwiebel | 43 | Lauchzwiebel  „Long White“ |
| 12 | G | Gemüse | GMI | Mittelzehrer | GMIMANGOLD | Mangold | 45 | Mangold „Bright Lights“ |
| 13 | G | Gemüse | GMI | Mittelzehrer | GMIMÖHRE | Möhre | 48 | Möhre „Milan" |
| 14 | G | Gemüse | GMI | Mittelzehrer | GMIMÖHRE | Möhre | 49 | Möhre „Colorada“ |
| 15 | G | Gemüse | GMI | Mittelzehrer | GMIMÖHRE | Möhre | 50 | Möhre „Bunte Mischung“ |
| 16 | G | Gemüse | GMI | Mittelzehrer | GMIPAPRIKA | Paprika | 52 | Paprika „Runde Ungarische“ |
| 17 | G | Gemüse | GMI | Mittelzehrer | GMIPAPRIKA | Paprika | 53 | Paprika „Sweet Banana“ |
| 18 | G | Gemüse | GMI | Mittelzehrer | GMIPAPRIKA | Paprika | 54 | Paprika „Pusztagold“ |
| 19 | G | Gemüse | GMI | Mittelzehrer | GMIPASTINAKE | Pastinake | 55 | Pastinake „Tender and True“ |
| 20 | G | Gemüse | GMI | Mittelzehrer | GMIPFEFFERONI | Pfefferoni | 21 | Chili „Vietnam - Landsorte“ |
| 21 | G | Gemüse | GMI | Mittelzehrer | GMIPFEFFERONI | Pfefferoni | 56 | Pfefferoni „Taeyang medium" |
| 22 | G | Gemüse | GMI | Mittelzehrer | GMIRETTICH | Rettich | 59 | Rettich „Daikon" |

# 2. Multi-active satellite

## Build wohnort as a multi-active satellite

In the source modeler, tag the wohnort object with the satellite signature and make it multi-active.



Source
WEBSHOP

Depth: 0 1 2 3 4

□ show only not completed objects

- All Objects
- BESTELLUNG
- HTERMINTREUE
- KUNDE
- LIEFERADRESSE
- LIEFERDIENST
- LIEFERUNG
- POSITION

**KUNDE**
Hub and Satellite

| | | |
|---|---|---|
| BK PK | KUNDEID | # |
| FK | VEREINSPARTNERID | ❞ |
| | VORNAME | ❞ |
| | NAME | ❞ |
| | GESCHLECHT | ❞ |
| | GEBURTSDATUM | 📅 |

**WOHNORT**
Multi-Active Satellite

| | | |
|---|---|---|
| BK FK... | KUNDEID | # |
| SA | VON | 📅 |
| | BIS | 📅 |
| | STRASSE | ❞ |
| | HAUSNUMMER | ❞ |
| | ADRESSZUSATZ | ❞ |

**General**

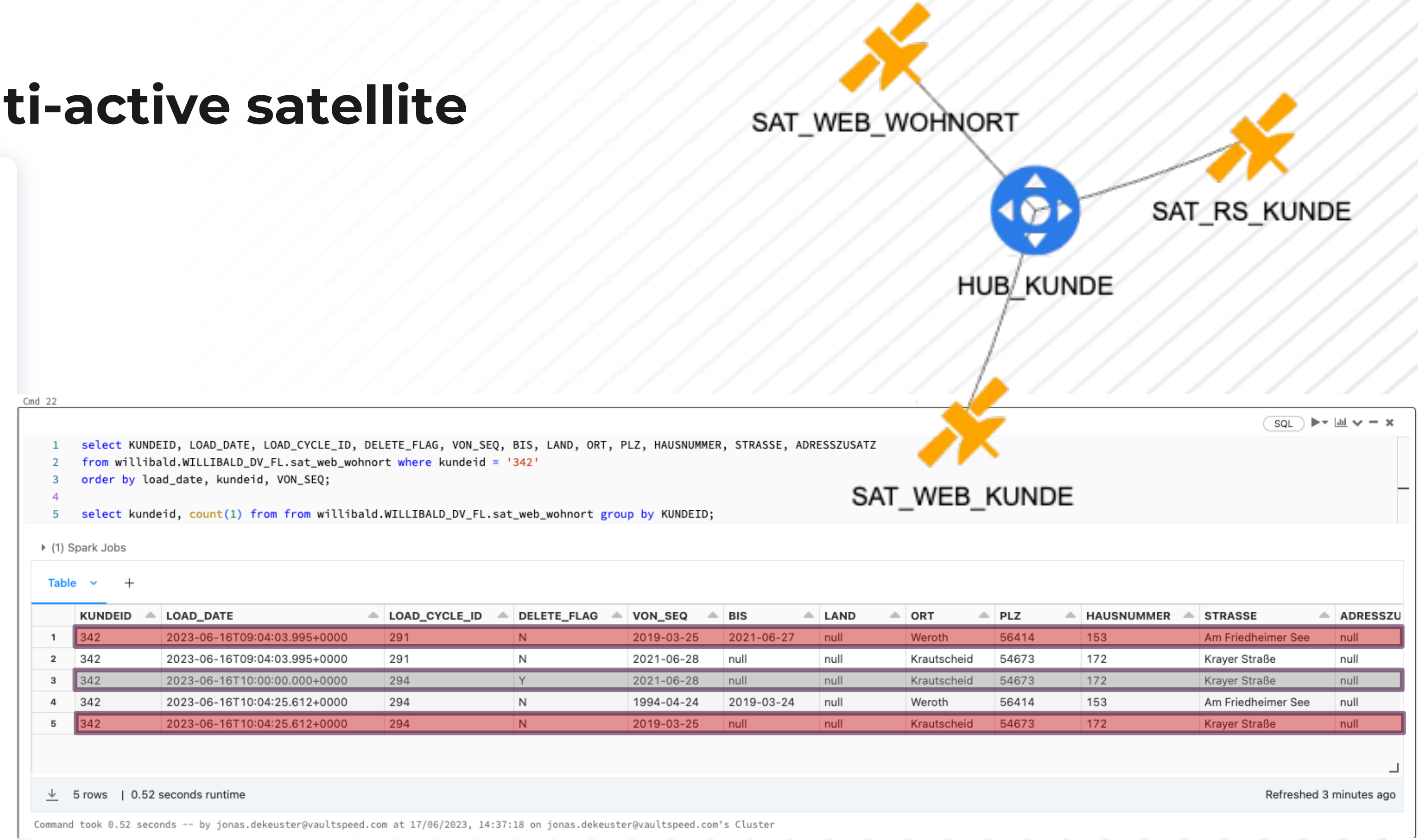
| | |
|---|---|
| Object name in source | WOHNORT |
| Full name | WOHNORT |
| Short name | WOHNORT |
| Abbreviated name | WOHNORT |
| Signature Object | Multi-Active Satellite |
| CDC | Full Load with Delete Manage... |
| Comments | |
| Multi active | ● on |
| Ignored | ○ off |
| Ignore reason | |

# 2. Multi-active satellite

## Build wohnort as a multi-active satellite

We can indeed find some real multi-active examples in the data with overlapping addresses.

We could also have modelled this by isolating the relation between address and kundeid and model that as a link between Kunde and Adress



```
select KUNDEID, LOAD_DATE, LOAD_CYCLE_ID, DELETE_FLAG, VON_SEQ, BIS, LAND, ORT, PLZ, HAUSNUMMER, STRASSE, ADRESSZUSATZ
from willibald.WILLIBALD_DV_FL.sat_web_wohnort where kundeid = '342'
order by load_date, kundeid, VON_SEQ;

select kundeid, count(1) from from willibald.WILLIBALD_DV_FL.sat_web_wohnort group by KUNDEID;
```

| | KUNDEID | LOAD_DATE | LOAD_CYCLE_ID | DELETE_FLAG | VON_SEQ | BIS | LAND | ORT | PLZ | HAUSNUMMER | STRASSE | ADRESSZU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 342 | 2023-06-16T09:04:03.995+0000 | 291 | N | 2019-03-25 | 2021-06-27 | null | Weroth | 56414 | 153 | Am Friedheimer See | null |
| 2 | 342 | 2023-06-16T09:04:03.995+0000 | 291 | N | 2021-06-28 | null | null | Krautscheid | 54673 | 172 | Krayer Straße | null |
| 3 | 342 | 2023-06-16T10:00:00.000+0000 | 294 | Y | 2021-06-28 | null | null | Krautscheid | 54673 | 172 | Krayer Straße | null |
| 4 | 342 | 2023-06-16T10:04:25.612+0000 | 294 | N | 1994-04-24 | 2019-03-24 | null | Weroth | 56414 | 153 | Am Friedheimer See | null |
| 5 | 342 | 2023-06-16T10:04:25.612+0000 | 294 | N | 2019-03-25 | null | null | Krautscheid | 54673 | 172 | Krayer Straße | null |

5 rows | 0.52 seconds runtime

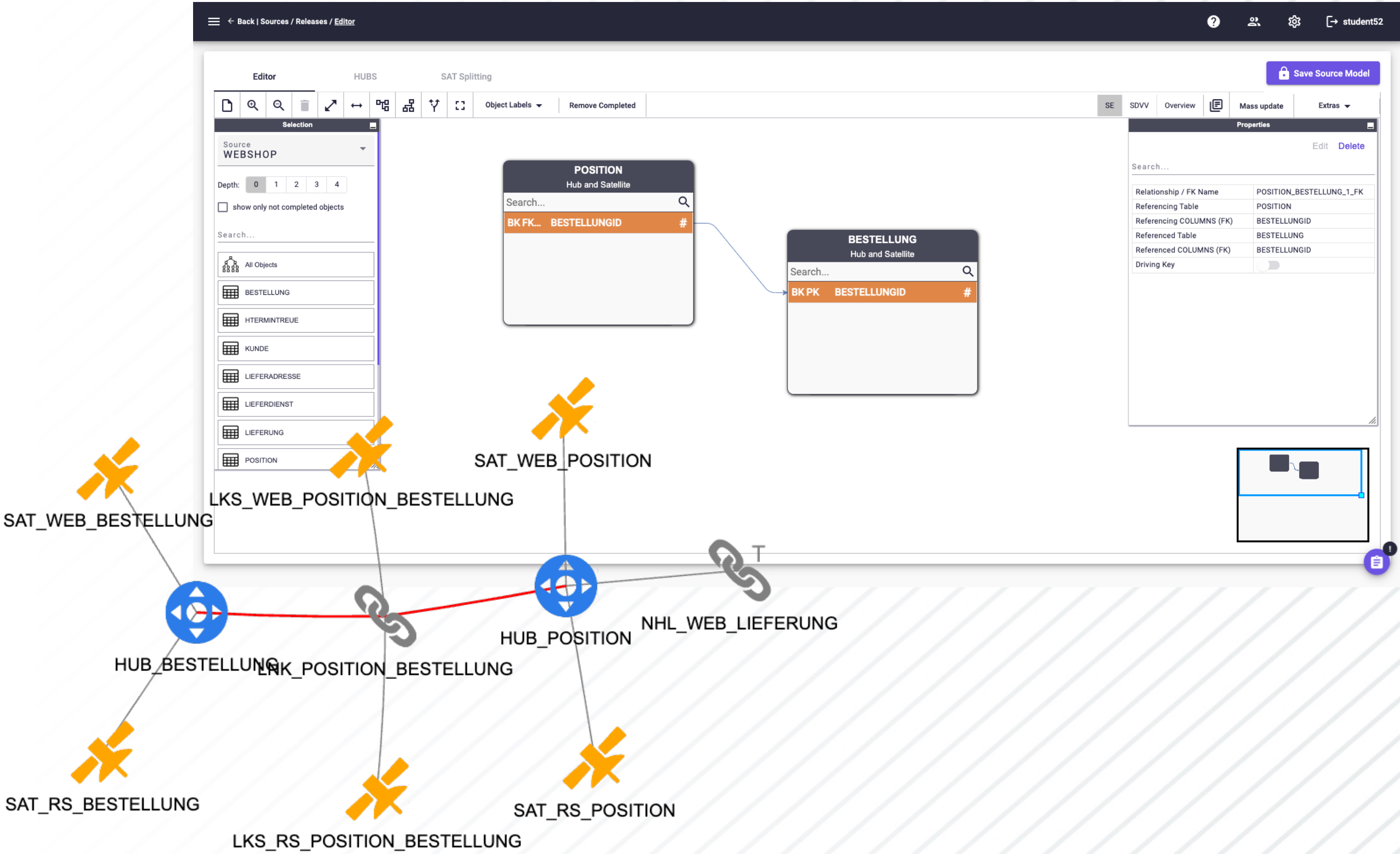
Refreshed 3 minutes ago

Command took 0.52 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 14:37:18 on jonas.dekeuster@vaultspeed.com's Cluster

# 3. Identifying relationship and driving keys

## Bestellungen and Positions are both modeled as hubs in this case

They are connected trough a link: **LNK_POSITION_BESTELLUNG**
In case of M:N links, VaultSpeed can set a driving key, but here it is not needed as the standard link with a satlink between both hubs will capture the relation correctly.

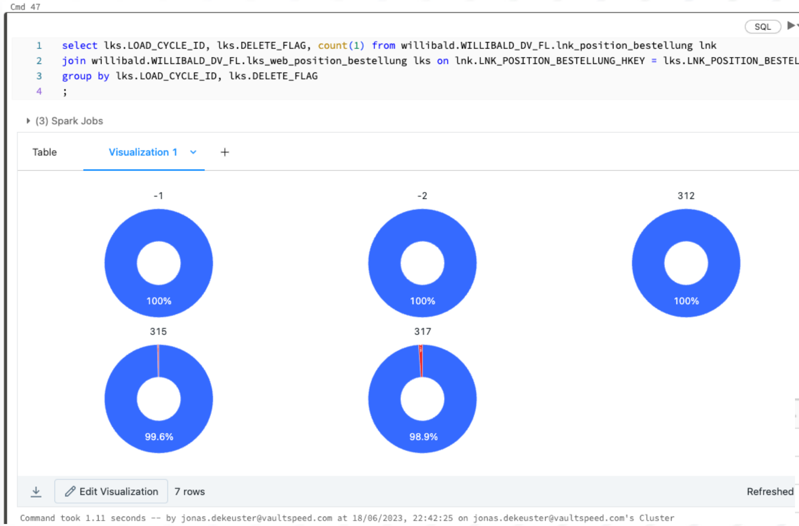


| | |
|---|---|
| Relationship / FK Name | POSITION_BESTELLUNG_1_FK |
| Referencing Table | POSITION |
| Referencing COLUMNS (FK) | BESTELLUNGID |
| Referenced Table | BESTELLUNG |
| Referenced COLUMNS (FK) | BESTELLUNGID |
| Driving Key | |

# 3. Identifying relationship and driving keys

## Bestellungen and Positions are both modeled as hubs in this case

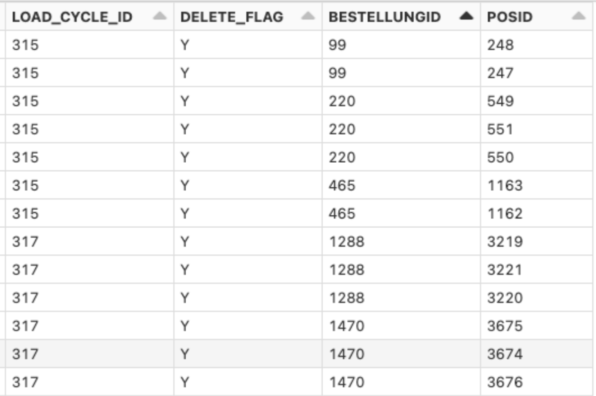
They are connected trough a link: **LNK_POSITION_BESTELLUNG** In case of M:N links, VaultSpeed can set a driving key, but here it is not needed as the standard link with a satlink between both hubs will capture the relation correctly.



Cmd 47

```
select lks.LOAD_CYCLE_ID, lks.DELETE_FLAG, count(1) from willibald.WILLIBALD_DV_FL.lnk_position_bestellung lnk
join willibald.WILLIBALD_DV_FL.lks_web_position_bestellung lks on lnk.LNK_POSITION_BESTELLUNG_HKEY = lks.LNK_POSITION_BESTELLUNG_HKEY
group by lks.LOAD_CYCLE_ID, lks.DELETE_FLAG
;
```

Command took 1.11 seconds -- by jonas.dekeuster@vaultspeed.com at 18/06/2023, 22:42:25 on jonas.dekeuster@vaultspeed.com's Cluster



| LOAD_CYCLE_ID | DELETE_FLAG | BESTELLUNGID | POSID |
|---|---|---|---|
| 315 | Y | 99 | 248 |
| 315 | Y | 99 | 247 |
| 315 | Y | 220 | 549 |
| 315 | Y | 220 | 551 |
| 315 | Y | 220 | 550 |
| 315 | Y | 465 | 1163 |
| 315 | Y | 465 | 1162 |
| 317 | Y | 1288 | 3219 |
| 317 | Y | 1288 | 3221 |
| 317 | Y | 1288 | 3220 |
| 317 | Y | 1470 | 3675 |
| 317 | Y | 1470 | 3674 |
| 317 | Y | 1470 | 3676 |

7 relations deleted between period 1 and 2,
6 relations deleted between period 2 and 3

(the ones that correspond to the 5 deletions of bestellung in test case 12 )

VAULTSPEED

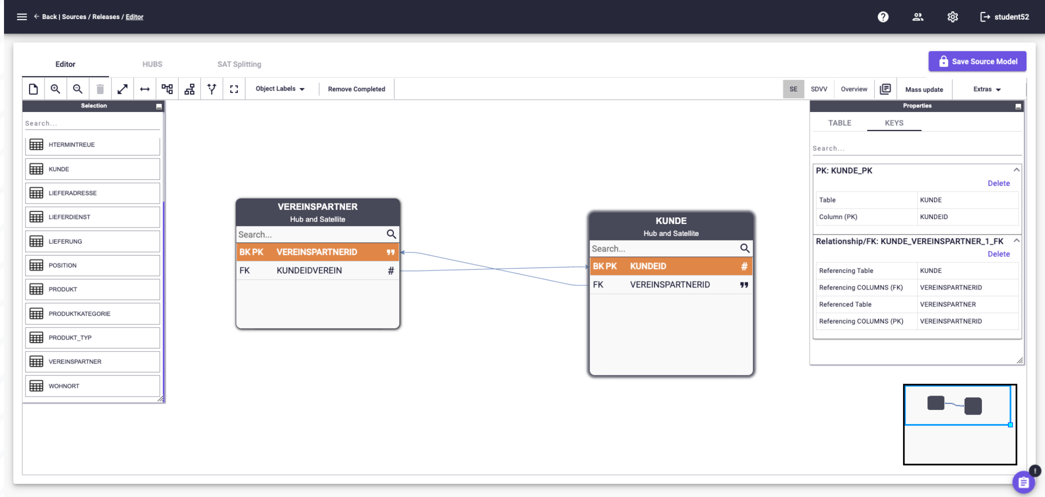
# 3. Identifying relationships

## Changes between Vereinspartner and Kunde

These changes are being tracked trough links: **LNK_KUNDE_VEREINSTPARTNER** and **LNK_VEREINSTPARTNER_ KUNDE**

We modelled a link in both directions. This is easy to do using VaultSpeed's source editor.

Note that have a tab in the source editor that shows the data vault model based on what we know from this source (so before any integration with other sources)



Back | Sources / Releases / Editor

Editor · HUBS · SAT Splitting · Save Source Model

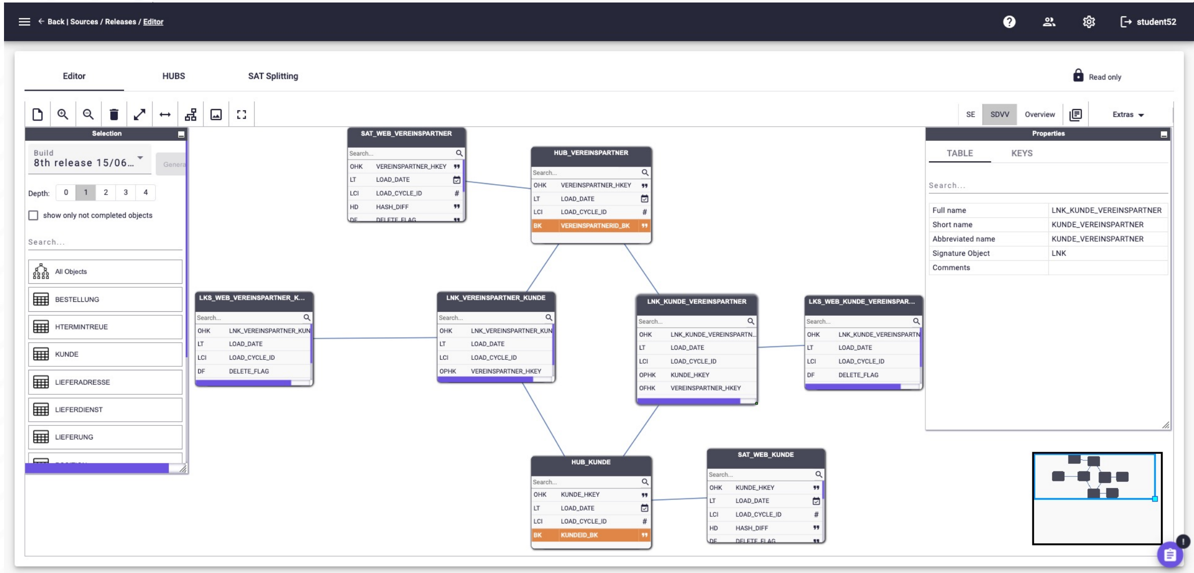
Properties — TABLE · KEYS

PK: KUNDE_PK — Delete

| | |
|---|---|
| Table | KUNDE |
| Column (PK) | KUNDEID |

Relationship/FK: KUNDE_VEREINSPARTNER_1_FK — Delete

| | |
|---|---|
| Referencing Table | KUNDE |
| Referencing COLUMNS (FK) | VEREINSPARTNERID |
| Referenced Table | VEREINSPARTNER |
| Referencing COLUMNS (PK) | VEREINSPARTNERID |



Back | Sources / Releases / Editor

Editor · HUBS · SAT Splitting · Read only

Properties — TABLE · KEYS

| | |
|---|---|
| Full name | LNK_KUNDE_VEREINSPARTNER |
| Short name | KUNDE_VEREINSPARTNER |
| Abbreviated name | KUNDE_VEREINSPARTNER |
| Signature Object | LNK |
| Comments | |

VAULTSPEED

# 3. Identifying relationships

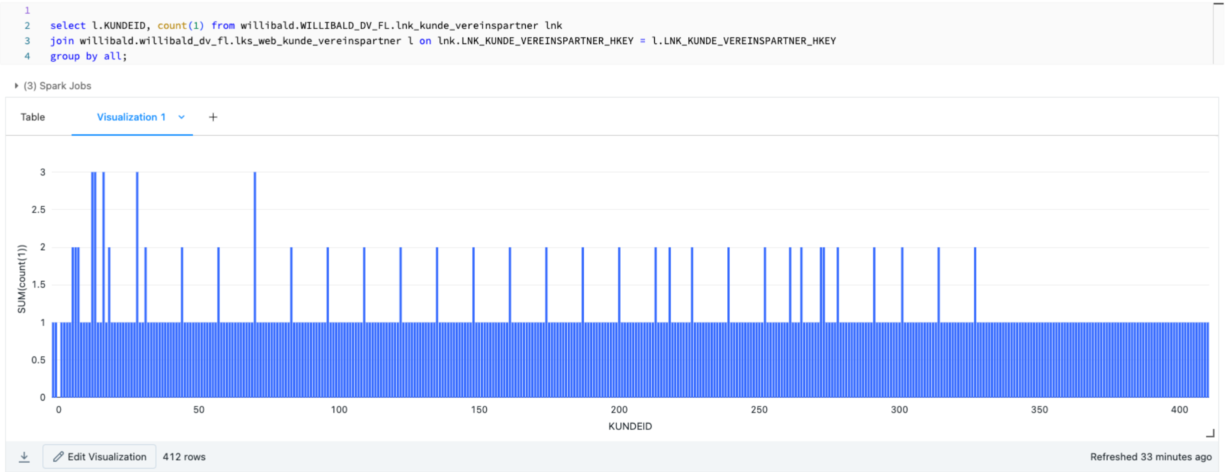
## Changes between Vereinspartner and Kunde

These changes are being tracked trough links:
**LNK_KUNDE_VEREINSTPARTNER**
and
**LNK_VEREINSTPARTNER_ KUNDE**

We modelled a link in both directions. This is easy to do using VaultSpeed's source editor.

Note that have a tab in the source editor that shows the data vault model based on what we know from this source (so before any integration with other sources)

Overview of how many satlinks per customer id (so how many times a customer changed association partner)



```
select l.KUNDEID, count(1) from willibald.WILLIBALD_DV_FL.lnk_kunde_vereinspartner lnk
join willibald.willibald_dv_fl.lks_web_kunde_vereinspartner l on lnk.LNK_KUNDE_VEREINSPARTNER_HKEY = l.LNK_KUNDE_VEREINSPARTNER_HKEY
group by all;
```

# 3. Identifying relationships

## Changes between Vereinspartner and Kunde

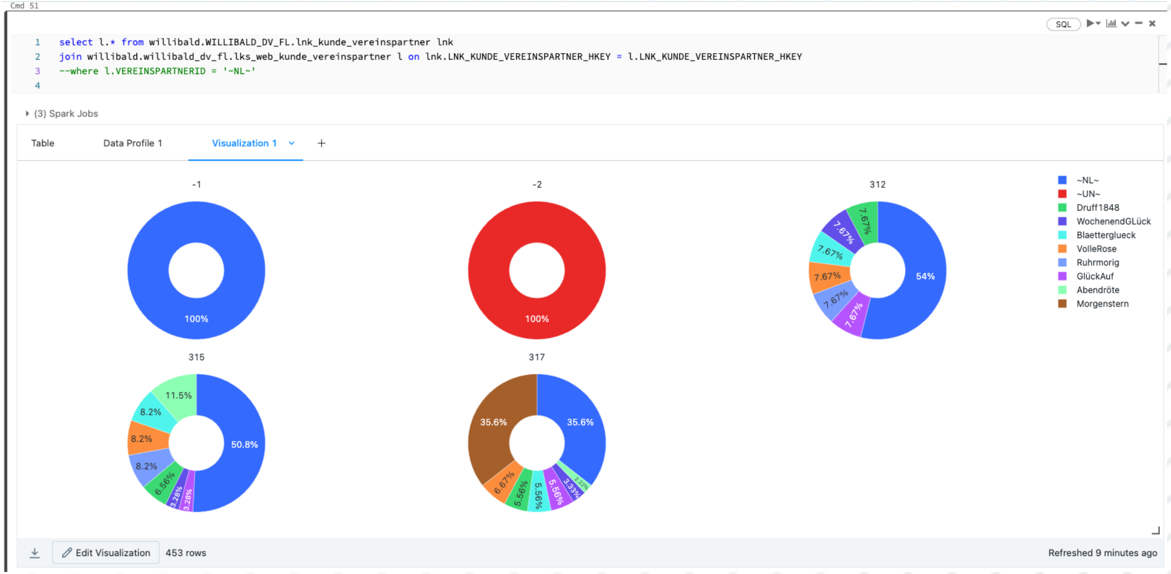
These changes are being tracked trough links: **LNK_KUNDE_VEREINSTPARTNER** and **LNK_VEREINSTPARTNER_ KUNDE**

We modelled a link in both directions. This is easy to do using VaultSpeed's source editor.

Note that have a tab in the source editor that shows the data vault model based on what we know from this source (so before any integration with other sources)

Overview of customers without vereinspartner (~NL~) across loads

(detail: the value that is give the the unknown and orphan records can be configured)



Cmd 51

```sql
select l.* from willibald.WILLIBALD_DV_FL.lnk_kunde_vereinspartner lnk
join willibald.willibald_dv_fl.lks_web_kunde_vereinspartner l on lnk.LNK_KUNDE_VEREINSPARTNER_HKEY = l.LNK_KUNDE_VEREINSPARTNER_HKEY
--where l.VEREINSPARTNERID = '~NL~'
```

# 3. Identifying relationships

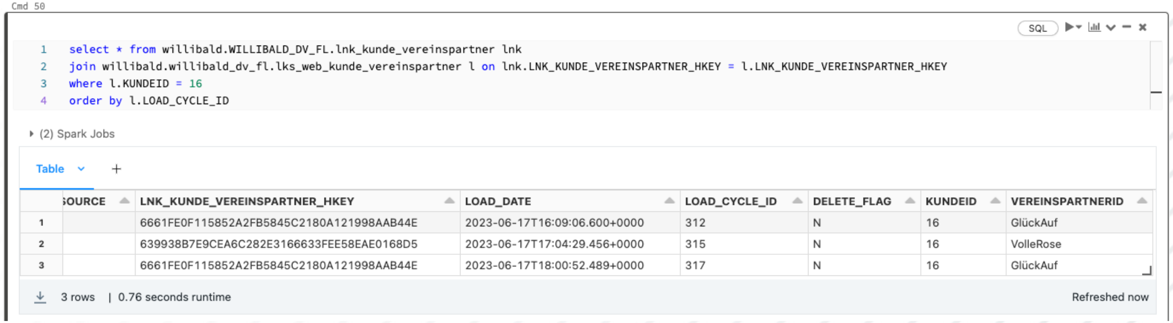
## Changes between Vereinspartner and Kunde

These changes are being tracked trough links:
**LNK_KUNDE_VEREINSTPARTNER**
and
**LNK_VEREINSTPARTNER_ KUNDE**

We modelled a link in both directions. This is easy to do using VaultSpeed's source editor.

Note that have a tab in the source editor that shows the data vault model based on what we know from this source (so before any integration with other sources)

Case where a relationship changes and also changes back to the first version



```sql
select * from willibald.WILLIBALD_DV_FL.lnk_kunde_vereinspartner lnk
join willibald.willibald_dv_fl.lks_web_kunde_vereinspartner l on lnk.LNK_KUNDE_VEREINSPARTNER_HKEY = l.LNK_KUNDE_VEREINSPARTNER_HKEY
where l.KUNDEID = 16
order by l.LOAD_CYCLE_ID
```

| | ;OURCE | LNK_KUNDE_VEREINSPARTNER_HKEY | LOAD_DATE | LOAD_CYCLE_ID | DELETE_FLAG | KUNDEID | VEREINSPARTNERID |
|---|---|---|---|---|---|---|---|
| 1 | | 6661FE0F115852A2FB5845C2180A121998AAB44E | 2023-06-17T16:09:06.600+0000 | 312 | N | 16 | GlückAuf |
| 2 | | 639938B7E9CEA6C282E3166633FEE58EAE0168D5 | 2023-06-17T17:04:29.456+0000 | 315 | N | 16 | VolleRose |
| 3 | | 6661FE0F115852A2FB5845C2180A121998AAB44E | 2023-06-17T18:00:52.489+0000 | 317 | N | 16 | GlückAuf |

3 rows | 0.76 seconds runtime

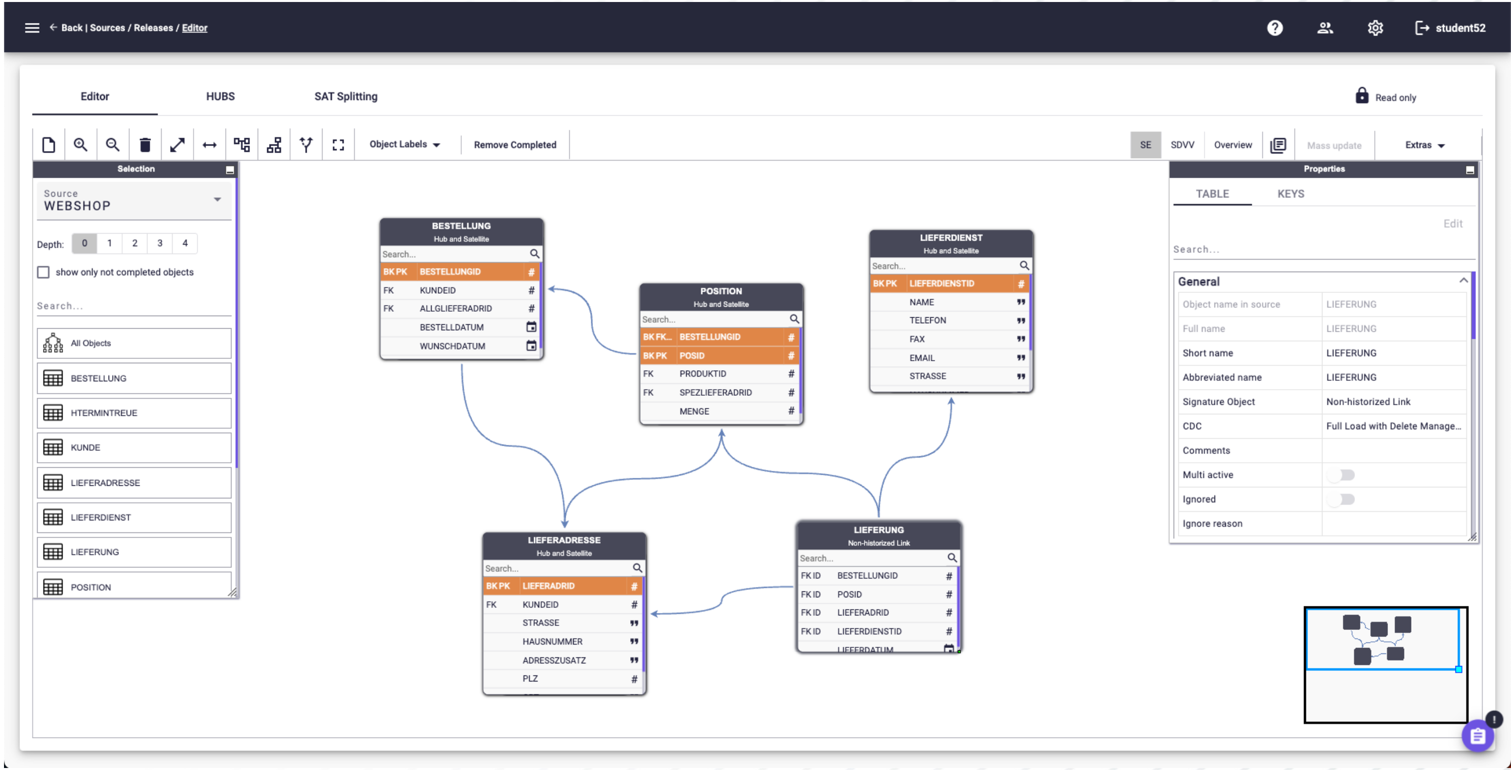
Refreshed now

VAULTSPEED

# 4. M:N relations without a key

**Transactional Links or Non-historized links**

Lieferung is modelled here as a transactional link. Each instance will be treated as a transaction.



Back | Sources / Releases / Editor

student52

Editor HUBS SAT Splitting

Read only

Object Labels Remove Completed

SE SDVV Overview Mass update Extras

**Selection**

Source
WEBSHOP

Depth: 0 1 2 3 4

□ show only not completed objects

Search...

- All Objects
- BESTELLUNG
- HTERMINTREUE
- KUNDE
- LIEFERADRESSE
- LIEFERDIENST
- LIEFERUNG
- POSITION

**BESTELLUNG**
Hub and Satellite

| | | |
|---|---|---|
| BK PK | BESTELLUNGID | # |
| FK | KUNDEID | # |
| FK | ALLGLIEFERADRID | # |
| | BESTELLDATUM | |
| | WUNSCHDATUM | |

**POSITION**
Hub and Satellite

| | | |
|---|---|---|
| BK FK... | BESTELLUNGID | # |
| BK PK | POSID | # |
| FK | PRODUKTID | # |
| FK | SPEZLIEFERADRID | # |
| | MENGE | # |

**LIEFERDIENST**
Hub and Satellite

| | | |
|---|---|---|
| BK PK | LIEFERDIENSTID | # |
| | NAME | ” |
| | TELEFON | ” |
| | FAX | ” |
| | EMAIL | ” |
| | STRASSE | ” |

**LIEFERADRESSE**
Hub and Satellite

| | | |
|---|---|---|
| BK PK | LIEFERADRID | # |
| FK | KUNDEID | # |
| | STRASSE | ” |
| | HAUSNUMMER | ” |
| | ADRESSZUSATZ | ” |
| | PLZ | # |

**LIEFERUNG**
Non-historized Link

| | | |
|---|---|---|
| FK ID | BESTELLUNGID | # |
| FK ID | POSID | # |
| FK ID | LIEFERADRID | # |
| FK ID | LIEFERDIENSTID | # |
| | LIEFERDATUM | |

**Properties**

TABLE KEYS

Edit

Search...

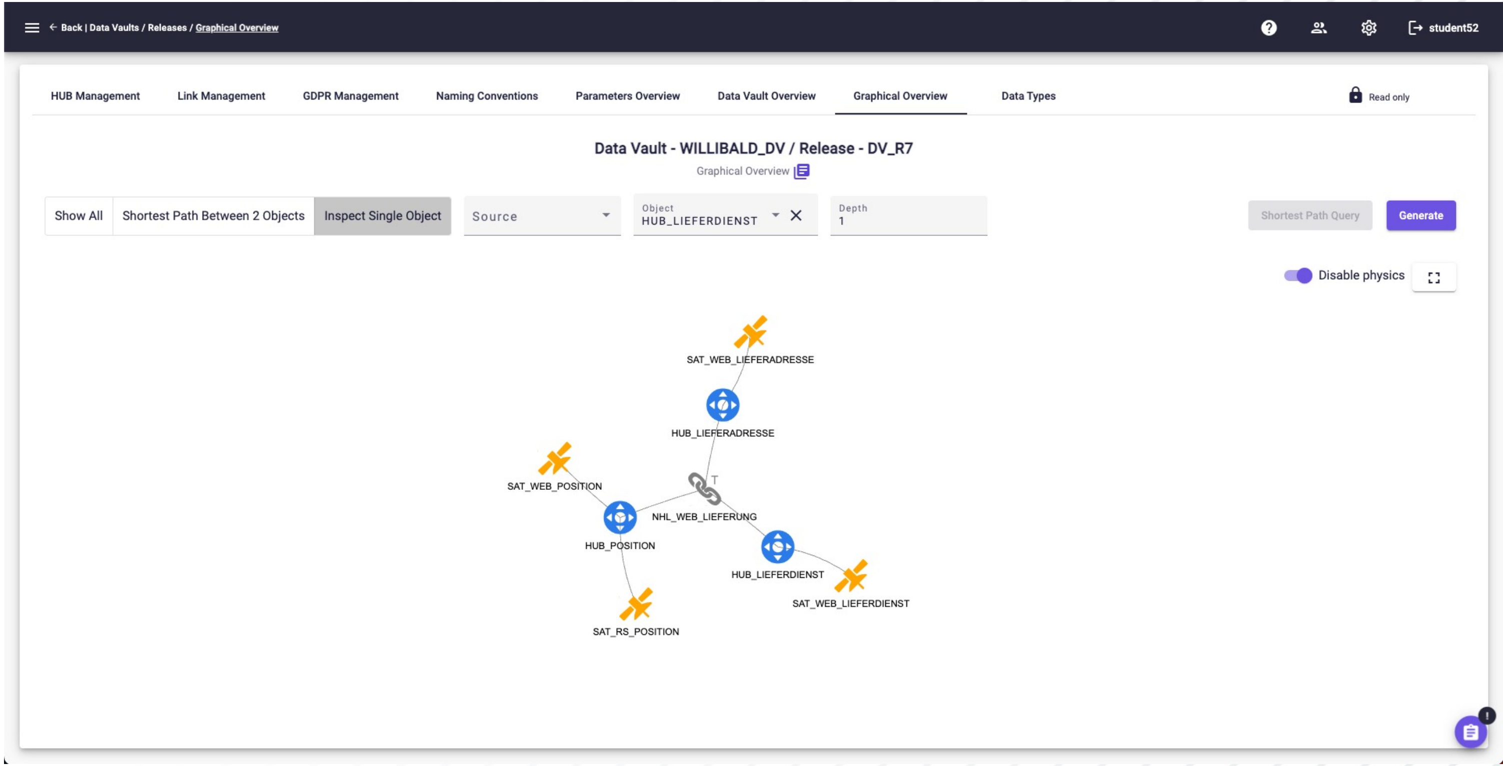
**General**

| | |
|---|---|
| Object name in source | LIEFERUNG |
| Full name | LIEFERUNG |
| Short name | LIEFERUNG |
| Abbreviated name | LIEFERUNG |
| Signature Object | Non-historized Link |
| CDC | Full Load with Delete Manage... |
| Comments | |
| Multi active | |
| Ignored | |
| Ignore reason | |

# 4. M:N relations without a key

**Transactional link**

Lieferung is modelled here as a transactional link. Each instance will be treated as a transaction.



Back | Data Vaults / Releases / Graphical Overview

student52

HUB Management | Link Management | GDPR Management | Naming Conventions | Parameters Overview | Data Vault Overview | Graphical Overview | Data Types

Read only

Data Vault - WILLIBALD_DV / Release - DV_R7

Graphical Overview

Show All | Shortest Path Between 2 Objects | Inspect Single Object

Source

Object: HUB_LIEFERDIENST

Depth: 1

Shortest Path Query | Generate

Disable physics

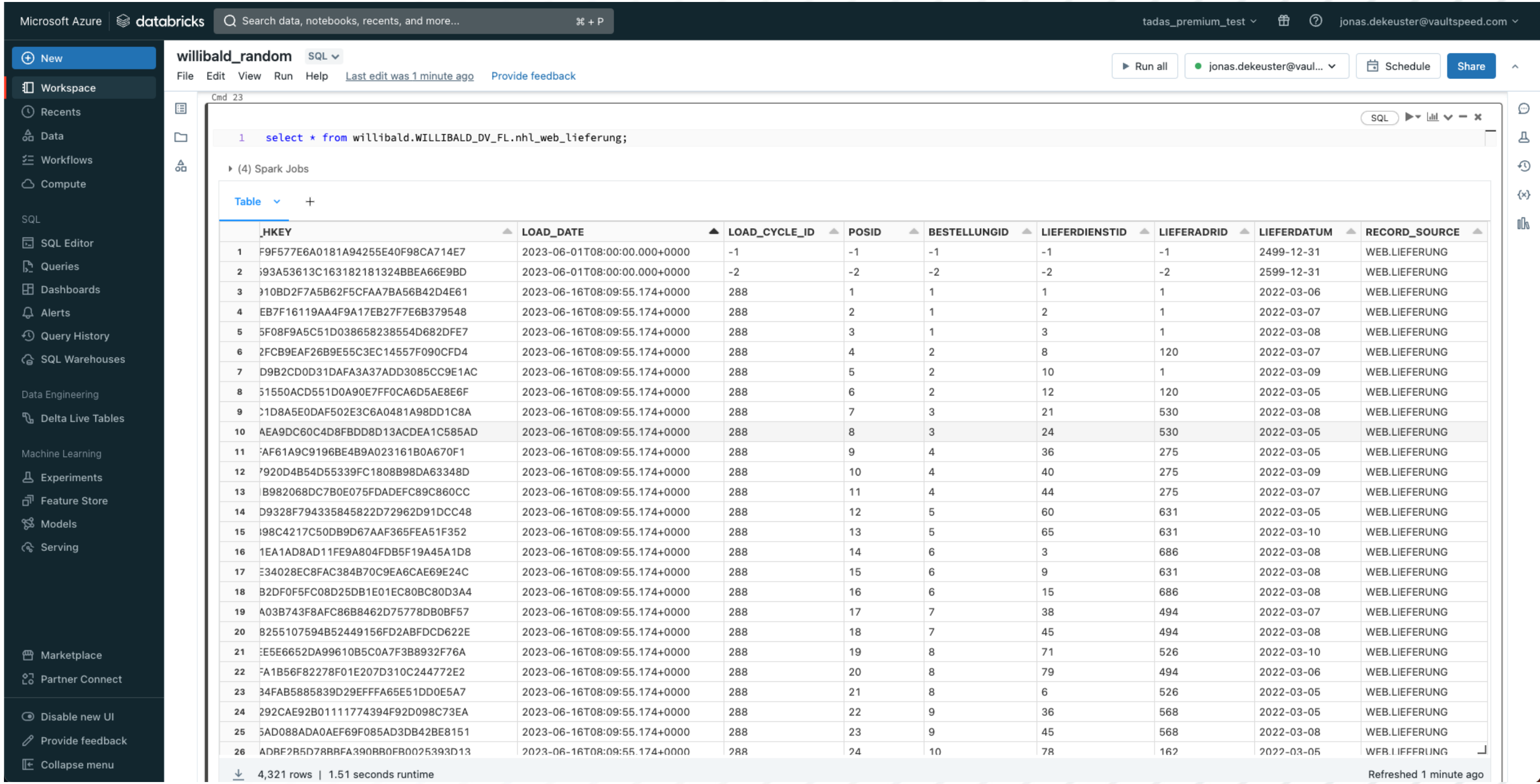
VAULTSPEED

# 4. M:N relations without a key

## Transactional link

Lieferung is modelled here as a transactional link. Each instance will be treated as a transaction.

**<u>Extra option:</u>**
in case you have transactional links with deletes, we offer the option to handle this by inserting the negative of the delete transaction.



```
select * from willibald.WILLIBALD_DV_FL.nhl_web_lieferung;
```

(4) Spark Jobs

| | _HKEY | LOAD_DATE | LOAD_CYCLE_ID | POSID | BESTELLUNGID | LIEFERDIENSTID | LIEFERADRID | LIEFERDATUM | RECORD_SOURCE |
|---|---|---|---|---|---|---|---|---|---|
| 1 | F9F577E6A0181A94255E40F98CA714E7 | 2023-06-01T08:00:00.000+0000 | -1 | -1 | -1 | -1 | -1 | 2499-12-31 | WEB.LIEFERUNG |
| 2 | 593A53613C163182181324BBEA66E9BD | 2023-06-01T08:00:00.000+0000 | -2 | -2 | -2 | -2 | -2 | 2599-12-31 | WEB.LIEFERUNG |
| 3 | 910BD2F7A5B62F5CFAA7BA56B42D4E61 | 2023-06-16T08:09:55.174+0000 | 288 | 1 | 1 | 1 | 1 | 2022-03-06 | WEB.LIEFERUNG |
| 4 | EB7F16119AA4F9A17EB27F7E6B379548 | 2023-06-16T08:09:55.174+0000 | 288 | 2 | 1 | 2 | 1 | 2022-03-07 | WEB.LIEFERUNG |
| 5 | 5F08F9A5C51D038658238554D682DFE7 | 2023-06-16T08:09:55.174+0000 | 288 | 3 | 1 | 3 | 1 | 2022-03-08 | WEB.LIEFERUNG |
| 6 | 2FCB9EAF26B9E55C3EC14557F090CFD4 | 2023-06-16T08:09:55.174+0000 | 288 | 4 | 2 | 8 | 120 | 2022-03-07 | WEB.LIEFERUNG |
| 7 | D9B2CD0D31DAFA3A37ADD3085CC9E1AC | 2023-06-16T08:09:55.174+0000 | 288 | 5 | 2 | 10 | 1 | 2022-03-09 | WEB.LIEFERUNG |
| 8 | 51550ACD551D0A90E7FF0CA6D5AE8E6F | 2023-06-16T08:09:55.174+0000 | 288 | 6 | 2 | 12 | 120 | 2022-03-05 | WEB.LIEFERUNG |
| 9 | C1D8A5E0DAF502E3C6A0481A98DD1C8A | 2023-06-16T08:09:55.174+0000 | 288 | 7 | 3 | 21 | 530 | 2022-03-08 | WEB.LIEFERUNG |
| 10 | AEA9DC60C4D8FBDD8D13ACDEA1C585AD | 2023-06-16T08:09:55.174+0000 | 288 | 8 | 3 | 24 | 530 | 2022-03-05 | WEB.LIEFERUNG |
| 11 | FAF61A9C9196BE4B9A023161B0A670F1 | 2023-06-16T08:09:55.174+0000 | 288 | 9 | 4 | 36 | 275 | 2022-03-05 | WEB.LIEFERUNG |
| 12 | 7920D4B54D55339FC1808B98DA63348D | 2023-06-16T08:09:55.174+0000 | 288 | 10 | 4 | 40 | 275 | 2022-03-09 | WEB.LIEFERUNG |
| 13 | B982068DC7B0E075FDADEFC89C860CC | 2023-06-16T08:09:55.174+0000 | 288 | 11 | 4 | 44 | 275 | 2022-03-07 | WEB.LIEFERUNG |
| 14 | D9328F794335845822D72962D91DCC48 | 2023-06-16T08:09:55.174+0000 | 288 | 12 | 5 | 60 | 631 | 2022-03-05 | WEB.LIEFERUNG |
| 15 | 98C4217C50DB9D67AAF365FEA51F352 | 2023-06-16T08:09:55.174+0000 | 288 | 13 | 5 | 65 | 631 | 2022-03-10 | WEB.LIEFERUNG |
| 16 | 1EA1AD8AD11FE9A804FDB5F19A45A1D8 | 2023-06-16T08:09:55.174+0000 | 288 | 14 | 6 | 3 | 686 | 2022-03-08 | WEB.LIEFERUNG |
| 17 | E34028EC8FAC384B70C9EA6CAE69E24C | 2023-06-16T08:09:55.174+0000 | 288 | 15 | 6 | 9 | 631 | 2022-03-08 | WEB.LIEFERUNG |
| 18 | B2DF0F5FC08D25DB1E01EC80BC80D3A4 | 2023-06-16T08:09:55.174+0000 | 288 | 16 | 6 | 15 | 686 | 2022-03-08 | WEB.LIEFERUNG |
| 19 | A03B743F8AFC86B8462D75778DB0BF57 | 2023-06-16T08:09:55.174+0000 | 288 | 17 | 7 | 38 | 494 | 2022-03-07 | WEB.LIEFERUNG |
| 20 | B255107594B52449156FD2ABFDCD622E | 2023-06-16T08:09:55.174+0000 | 288 | 18 | 7 | 45 | 494 | 2022-03-08 | WEB.LIEFERUNG |
| 21 | E5E6652DA99610B5C0A7F3B8932F76A | 2023-06-16T08:09:55.174+0000 | 288 | 19 | 8 | 71 | 526 | 2022-03-10 | WEB.LIEFERUNG |
| 22 | FA1B56F82278F01E207D310C244772E2 | 2023-06-16T08:09:55.174+0000 | 288 | 20 | 8 | 79 | 494 | 2022-03-06 | WEB.LIEFERUNG |
| 23 | 34FAB5885839D29EFFFA65E51DD0E5A7 | 2023-06-16T08:09:55.174+0000 | 288 | 21 | 8 | 6 | 526 | 2022-03-05 | WEB.LIEFERUNG |
| 24 | 292CAE92B01111774394F92D098C73EA | 2023-06-16T08:09:55.174+0000 | 288 | 22 | 9 | 36 | 568 | 2022-03-05 | WEB.LIEFERUNG |
| 25 | 5AD088ADA0AEF69F085AD3DB42BE8151 | 2023-06-16T08:09:55.174+0000 | 288 | 23 | 9 | 45 | 568 | 2022-03-08 | WEB.LIEFERUNG |
| 26 | ADBE2B5D78BBFA390BB0FB0025393D13 | 2023-06-16T08:09:55.174+0000 | 288 | 24 | 10 | 78 | 162 | 2022-03-05 | WEB.LIEFERUNG |

4,321 rows | 1.51 seconds runtime

Refreshed 1 minute ago

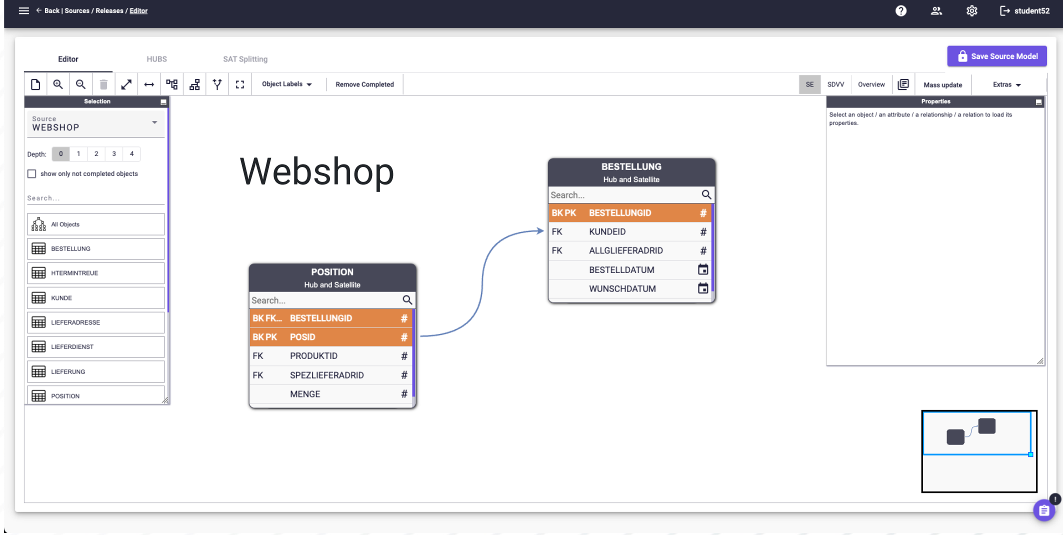
VAULTSPEED

# 5. Early integration

## Grouping and splitting bestellung and position objects

VaultSpeed offers the capabilities to integrate data early in the RDV. In this case we use our splitter and the hub grouping.

Bestellung and position are tangled up in 1 entity in RS source. Therefore we split them into the business keys we want to model: Bestellung and Position



Back | Sources / Releases / Editor

Editor HUBS SAT Splitting

Object Labels · Remove Completed · SE · SDVV · Overview · Mass update · Extras · Save Source Model

Selection

Source WEBSHOP

Depth: 0 1 2 3 4

□ show only not completed objects

- All Objects
- BESTELLUNG
- HTERMINTREUE
- KUNDE
- LIEFERADRESSE
- LIEFERDIENST
- LIEFERUNG
- POSITION

Properties

Select an object / an attribute / a relationship / a relation to load its properties.

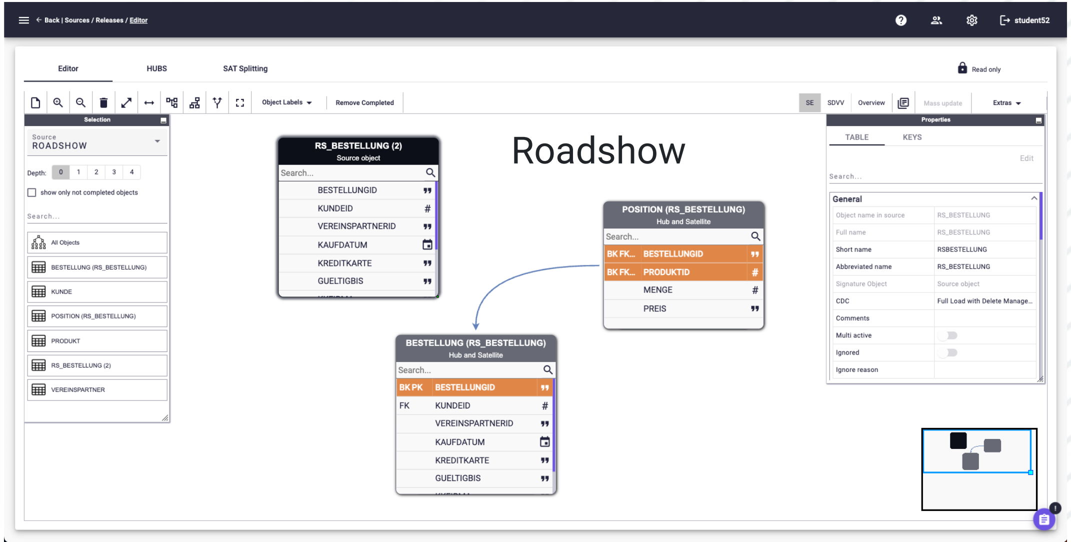
Webshop

POSITION
Hub and Satellite

| | | |
|---|---|---|
| BK FK... | BESTELLUNGID | # |
| BK PK | POSID | # |
| FK | PRODUKTID | # |
| FK | SPEZLIEFERADRID | # |
| | MENGE | # |

BESTELLUNG
Hub and Satellite

| | | |
|---|---|---|
| BK PK | BESTELLUNGID | # |
| FK | KUNDEID | # |
| FK | ALLGLIEFERADRID | # |
| | BESTELLDATUM | 📅 |
| | WUNSCHDATUM | 📅 |



Back | Sources / Releases / Editor

student52

Editor HUBS SAT Splitting

Read only

Object Labels · Remove Completed · SE · SDVV · Overview · Mass update · Extras

Selection

Source ROADSHOW

Depth: 0 1 2 3 4

□ show only not completed objects

- All Objects
- BESTELLUNG (RS_BESTELLUNG)
- KUNDE
- POSITION (RS_BESTELLUNG)
- PRODUKT
- RS_BESTELLUNG (2)
- VEREINSPARTNER

Roadshow

RS_BESTELLUNG (2)
Source object

| | |
|---|---|
| BESTELLUNGID | ❞ |
| KUNDEID | # |
| VEREINSPARTNERID | ❞ |
| KAUFDATUM | 📅 |
| KREDITKARTE | ❞ |
| GUELTIGBIS | ❞ |

POSITION (RS_BESTELLUNG)
Hub and Satellite

| | | |
|---|---|---|
| BK FK... | BESTELLUNGID | ❞ |
| BK FK... | PRODUKTID | # |
| | MENGE | # |
| | PREIS | ❞ |

BESTELLUNG (RS_BESTELLUNG)
Hub and Satellite

| | | |
|---|---|---|
| BK PK | BESTELLUNGID | ❞ |
| FK | KUNDEID | # |
| | VEREINSPARTNERID | ❞ |
| | KAUFDATUM | 📅 |
| | KREDITKARTE | ❞ |
| | GUELTIGBIS | ❞ |

Properties

TABLE KEYS

Edit

General

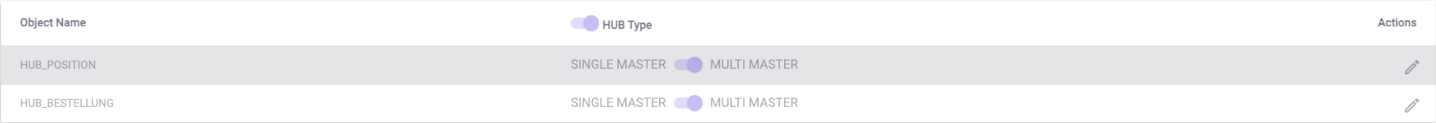
| | |
|---|---|
| Object name in source | RS_BESTELLUNG |
| Full name | RS_BESTELLUNG |
| Short name | RSBESTELLUNG |
| Abbreviated name | RS_BESTELLUNG |
| Signature Object | Source object |
| CDC | Full Load with Delete Manage... |
| Comments | |
| Multi active | ○ |
| Ignored | ○ |
| Ignore reason | |

# 5. Early integration

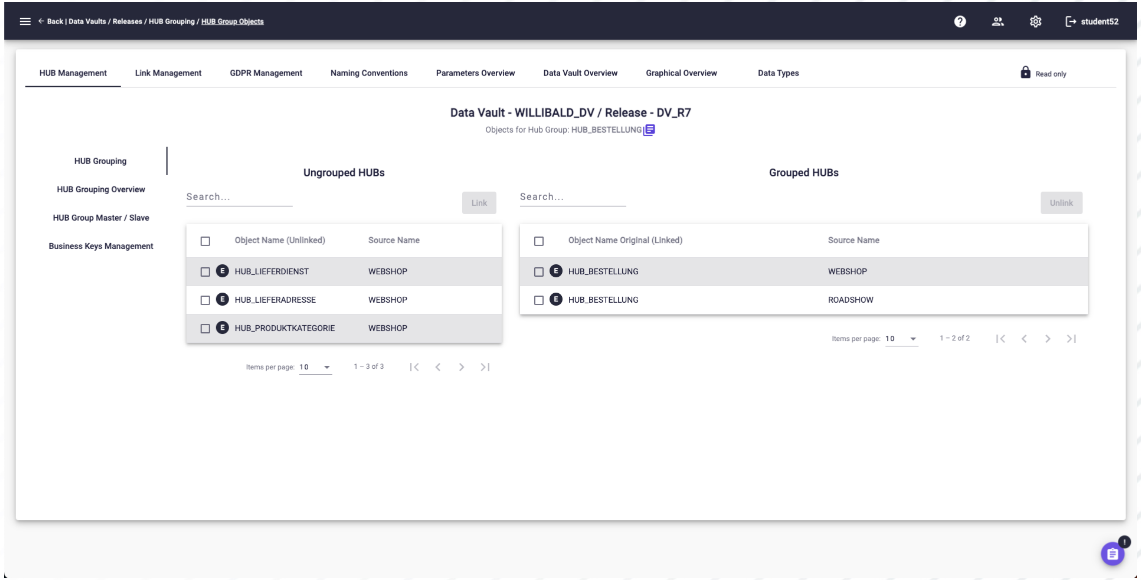
## Grouping and splitting bestellung and position objects

Once we prepare to build the RDV model. We integrate both source on their common business keys.

We modelled them as multi-master objects: Objects in the raw vault will get a SRC_BK as the business key collision code. This will make sure that potential overlapping business keys will be loaded correctly.



| Object Name | HUB Type | Actions |
|---|---|---|
| HUB_POSITION | SINGLE MASTER MULTI MASTER | |
| HUB_BESTELLUNG | SINGLE MASTER MULTI MASTER | |



Data Vault - WILLIBALD_DV / Release - DV_R7

Objects for Hub Group: HUB_BESTELLUNG

Ungrouped HUBs

| □ | Object Name (Unlinked) | Source Name |
|---|---|---|
| □ | HUB_LIEFERDIENST | WEBSHOP |
| □ | HUB_LIEFERADRESSE | WEBSHOP |
| □ | HUB_PRODUKTKATEGORIE | WEBSHOP |

Grouped HUBs

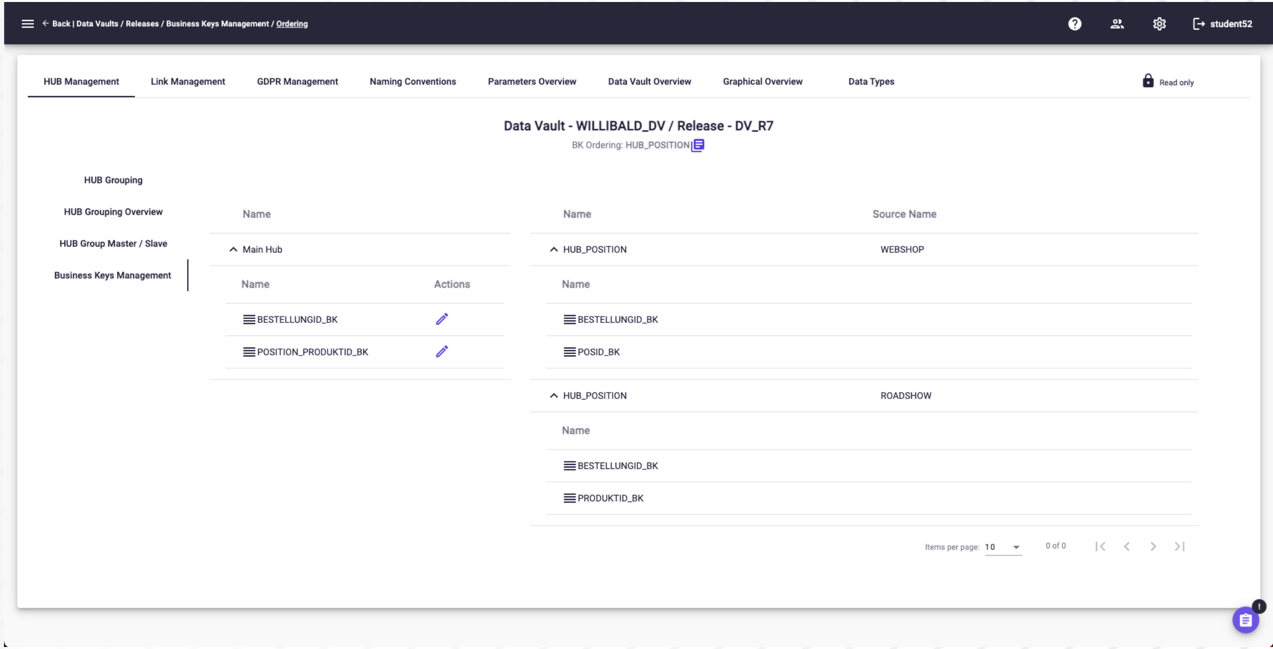
| □ | Object Name Original (Linked) | Source Name |
|---|---|---|
| □ | HUB_BESTELLUNG | WEBSHOP |
| □ | HUB_BESTELLUNG | ROADSHOW |

# 5. Early integration

## Grouping and splitting bestellung and position objects

We decided that the position key for the roadshow source should be bestellungID and ProductID, that makes it the most similar to the webshow source, where a position is influenced by having different products in 1 order, or having to deliver them to different addresses.

Therefore we make sure the business keys are ordered correctly, we can also rename the BK of the grouped hub.



Data Vault - WILLIBALD_DV / Release - DV_R7

BK Ordering: HUB_POSITION

| Name |
|---|
| Main Hub |

| Name | Actions |
|---|---|
| BESTELLUNGID_BK | ✎ |
| POSITION_PRODUKTID_BK | ✎ |

| Name | Source Name |
|---|---|
| HUB_POSITION | WEBSHOP |

| Name |
|---|
| BESTELLUNGID_BK |
| POSID_BK |

| Name | Source Name |
|---|---|
| HUB_POSITION | ROADSHOW |

| Name |
|---|
| BESTELLUNGID_BK |
| PRODUKTID_BK |

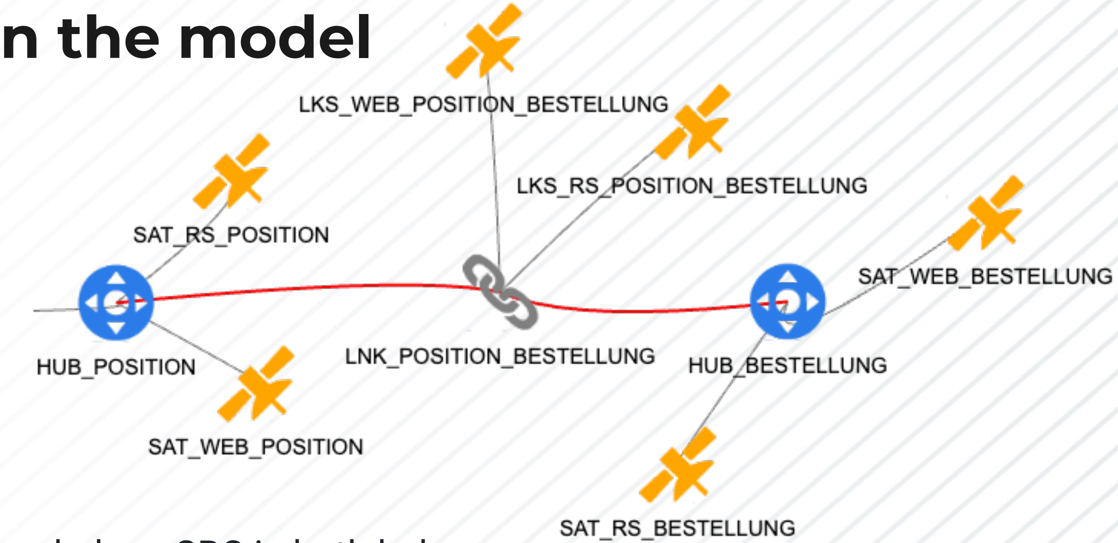
VAULTSPEED

# 5. Early integration in the model

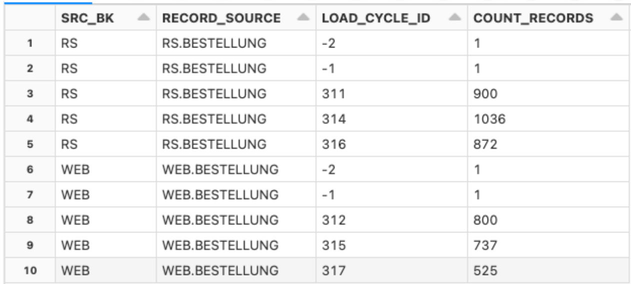
## Grouping and splitting bestellung and position objects

The positions and bestellungen hubs are both grouped together in 1 physical hub table. Each source has a satellite joined to this hub.

Note that also the link will be grouped together with two source specific satlinks attached

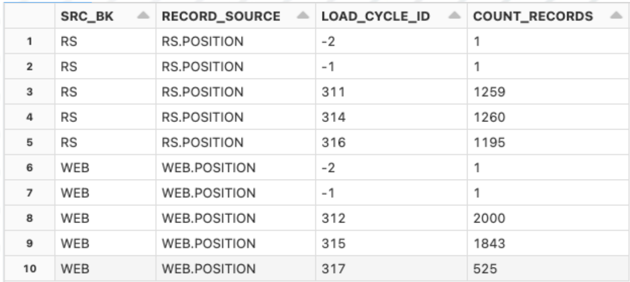


Number of records loaded per SRC in both hubs



|  | SRC_BK | RECORD_SOURCE | LOAD_CYCLE_ID | COUNT_RECORDS |
|---|---|---|---|---|
| 1 | RS | RS.BESTELLUNG | -2 | 1 |
| 2 | RS | RS.BESTELLUNG | -1 | 1 |
| 3 | RS | RS.BESTELLUNG | 311 | 900 |
| 4 | RS | RS.BESTELLUNG | 314 | 1036 |
| 5 | RS | RS.BESTELLUNG | 316 | 872 |
| 6 | WEB | WEB.BESTELLUNG | -2 | 1 |
| 7 | WEB | WEB.BESTELLUNG | -1 | 1 |
| 8 | WEB | WEB.BESTELLUNG | 312 | 800 |
| 9 | WEB | WEB.BESTELLUNG | 315 | 737 |
| 10 | WEB | WEB.BESTELLUNG | 317 | 525 |

hub_bestellungen



|  | SRC_BK | RECORD_SOURCE | LOAD_CYCLE_ID | COUNT_RECORDS |
|---|---|---|---|---|
| 1 | RS | RS.POSITION | -2 | 1 |
| 2 | RS | RS.POSITION | -1 | 1 |
| 3 | RS | RS.POSITION | 311 | 1259 |
| 4 | RS | RS.POSITION | 314 | 1260 |
| 5 | RS | RS.POSITION | 316 | 1195 |
| 6 | WEB | WEB.POSITION | -2 | 1 |
| 7 | WEB | WEB.POSITION | -1 | 1 |
| 8 | WEB | WEB.POSITION | 312 | 2000 |
| 9 | WEB | WEB.POSITION | 315 | 1843 |
| 10 | WEB | WEB.POSITION | 317 | 525 |

hub_positions

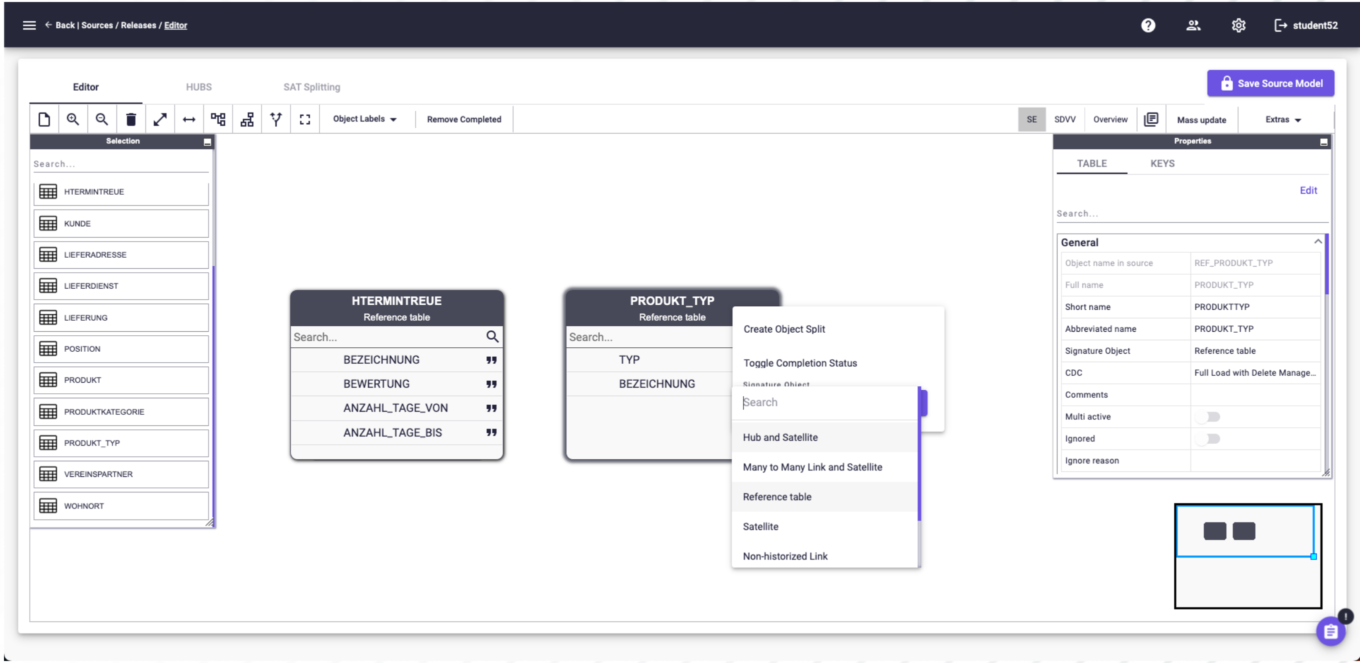
# 6. Reference tables

## Ref table

Just select the correct signature object and it will we treated as a ref table. A standard REF table does truncate insert

You can choose to load new records in there using the load date for historical comparison.

Easy to join to the snapshot PITs



| | |
|---|---|
| Object name in source | REF_PRODUKT_TYP |
| Full name | PRODUKT_TYP |
| Short name | PRODUKTTYP |
| Abbreviated name | PRODUKT_TYP |
| Signature Object | Reference table |
| CDC | Full Load with Delete Manage... |
| Comments | |
| Multi active | |
| Ignored | |
| Ignore reason | |

VAULTSPEED

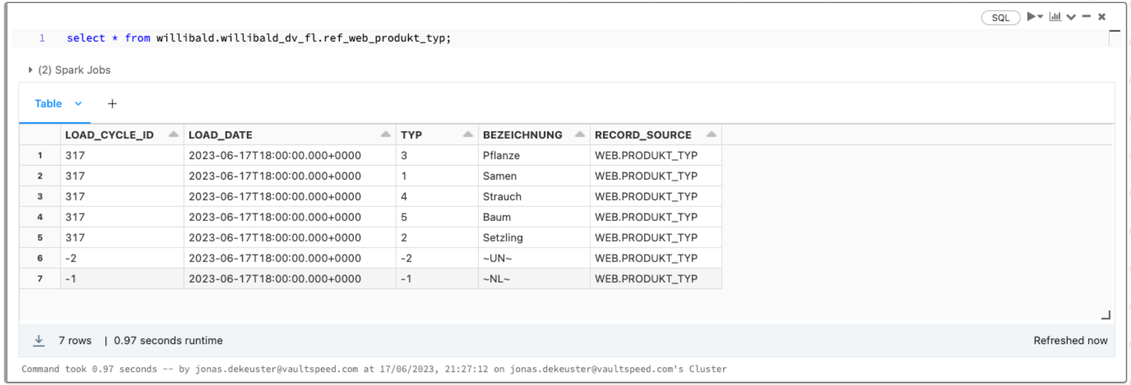
# 6. Reference tables

## Ref table

Just select the correct signature object and it will we treated as a ref table. A standard REF table does truncate insert

You can choose to load new records in there using the load date for historical comparison.

Easy to join to the snapshot PITs



```sql
select * from willibald.willibald_dv_fl.ref_web_produkt_typ;
```

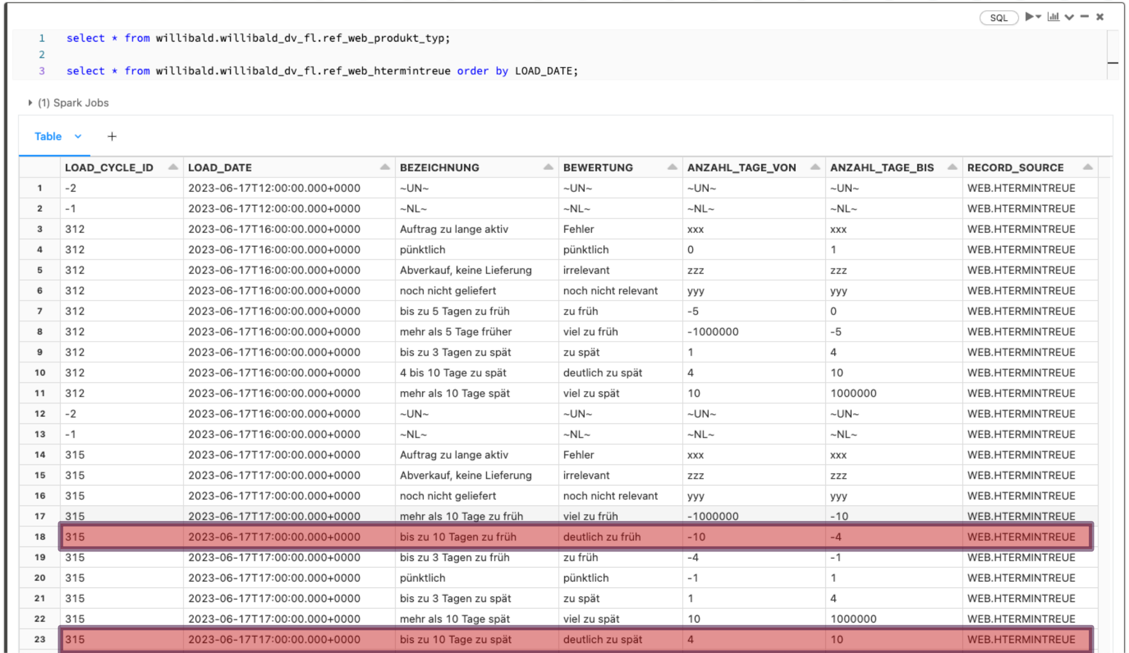
(2) Spark Jobs

| | LOAD_CYCLE_ID | LOAD_DATE | TYP | BEZEICHNUNG | RECORD_SOURCE |
|---|---|---|---|---|---|
| 1 | 317 | 2023-06-17T18:00:00.000+0000 | 3 | Pflanze | WEB.PRODUKT_TYP |
| 2 | 317 | 2023-06-17T18:00:00.000+0000 | 1 | Samen | WEB.PRODUKT_TYP |
| 3 | 317 | 2023-06-17T18:00:00.000+0000 | 4 | Strauch | WEB.PRODUKT_TYP |
| 4 | 317 | 2023-06-17T18:00:00.000+0000 | 5 | Baum | WEB.PRODUKT_TYP |
| 5 | 317 | 2023-06-17T18:00:00.000+0000 | 2 | Setzling | WEB.PRODUKT_TYP |
| 6 | -2 | 2023-06-17T18:00:00.000+0000 | -2 | ~UN~ | WEB.PRODUKT_TYP |
| 7 | -1 | 2023-06-17T18:00:00.000+0000 | -1 | ~NL~ | WEB.PRODUKT_TYP |

7 rows | 0.97 seconds runtime

Refreshed now

```
Command took 0.97 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 21:27:12 on jonas.dekeuster@vaultspeed.com's Cluster
```



```sql
select * from willibald.willibald_dv_fl.ref_web_produkt_typ;

select * from willibald.willibald_dv_fl.ref_web_htermintreue order by LOAD_DATE;
```

(1) Spark Jobs

| | LOAD_CYCLE_ID | LOAD_DATE | BEZEICHNUNG | BEWERTUNG | ANZAHL_TAGE_VON | ANZAHL_TAGE_BIS | RECORD_SOURCE |
|---|---|---|---|---|---|---|---|
| 1 | -2 | 2023-06-17T12:00:00.000+0000 | ~UN~ | ~UN~ | ~UN~ | ~UN~ | WEB.HTERMINTREUE |
| 2 | -1 | 2023-06-17T12:00:00.000+0000 | ~NL~ | ~NL~ | ~NL~ | ~NL~ | WEB.HTERMINTREUE |
| 3 | 312 | 2023-06-17T16:00:00.000+0000 | Auftrag zu lange aktiv | Fehler | xxx | xxx | WEB.HTERMINTREUE |
| 4 | 312 | 2023-06-17T16:00:00.000+0000 | pünktlich | pünktlich | 0 | 1 | WEB.HTERMINTREUE |
| 5 | 312 | 2023-06-17T16:00:00.000+0000 | Abverkauf, keine Lieferung | irrelevant | zzz | zzz | WEB.HTERMINTREUE |
| 6 | 312 | 2023-06-17T16:00:00.000+0000 | noch nicht geliefert | noch nicht relevant | yyy | yyy | WEB.HTERMINTREUE |
| 7 | 312 | 2023-06-17T16:00:00.000+0000 | bis zu 5 Tagen zu früh | zu früh | -5 | 0 | WEB.HTERMINTREUE |
| 8 | 312 | 2023-06-17T16:00:00.000+0000 | mehr als 5 Tage früher | viel zu früh | -1000000 | -5 | WEB.HTERMINTREUE |
| 9 | 312 | 2023-06-17T16:00:00.000+0000 | bis zu 3 Tagen zu spät | zu spät | 1 | 4 | WEB.HTERMINTREUE |
| 10 | 312 | 2023-06-17T16:00:00.000+0000 | 4 bis 10 Tage zu spät | deutlich zu spät | 4 | 10 | WEB.HTERMINTREUE |
| 11 | 312 | 2023-06-17T16:00:00.000+0000 | mehr als 10 Tage spät | viel zu spät | 10 | 1000000 | WEB.HTERMINTREUE |
| 12 | -2 | 2023-06-17T16:00:00.000+0000 | ~UN~ | ~UN~ | ~UN~ | ~UN~ | WEB.HTERMINTREUE |
| 13 | -1 | 2023-06-17T16:00:00.000+0000 | ~NL~ | ~NL~ | ~NL~ | ~NL~ | WEB.HTERMINTREUE |
| 14 | 315 | 2023-06-17T17:00:00.000+0000 | Auftrag zu lange aktiv | Fehler | xxx | xxx | WEB.HTERMINTREUE |
| 15 | 315 | 2023-06-17T17:00:00.000+0000 | Abverkauf, keine Lieferung | irrelevant | zzz | zzz | WEB.HTERMINTREUE |
| 16 | 315 | 2023-06-17T17:00:00.000+0000 | noch nicht geliefert | noch nicht relevant | yyy | yyy | WEB.HTERMINTREUE |
| 17 | 315 | 2023-06-17T17:00:00.000+0000 | mehr als 10 Tage zu früh | viel zu früh | -1000000 | -10 | WEB.HTERMINTREUE |
| 18 | 315 | 2023-06-17T17:00:00.000+0000 | bis zu 10 Tagen zu früh | deutlich zu früh | -10 | -4 | WEB.HTERMINTREUE |
| 19 | 315 | 2023-06-17T17:00:00.000+0000 | bis zu 3 Tagen zu früh | zu früh | -4 | -1 | WEB.HTERMINTREUE |
| 20 | 315 | 2023-06-17T17:00:00.000+0000 | pünktlich | pünktlich | -1 | 1 | WEB.HTERMINTREUE |
| 21 | 315 | 2023-06-17T17:00:00.000+0000 | bis zu 3 Tagen zu spät | zu spät | 1 | 4 | WEB.HTERMINTREUE |
| 22 | 315 | 2023-06-17T17:00:00.000+0000 | mehr als 10 Tage spät | viel zu spät | 10 | 1000000 | WEB.HTERMINTREUE |
| 23 | 315 | 2023-06-17T17:00:00.000+0000 | bis zu 10 Tage zu spät | deutlich zu spät | 4 | 10 | WEB.HTERMINTREUE |

# 7. Duplicates loading the data

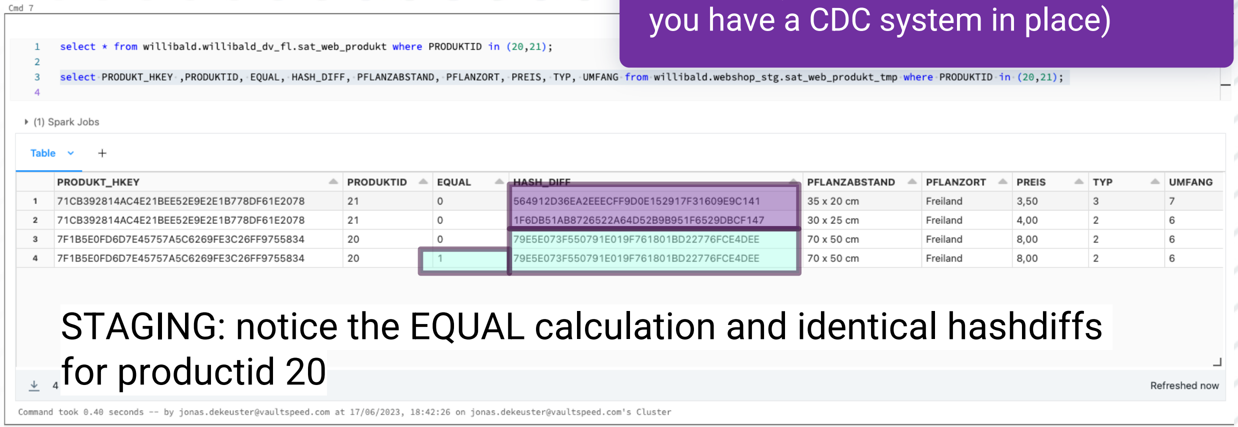
**EXTRA INFO**
VaultSpeed also supports having an attribute that contains the sequence of the data (for example the logposition if you have a CDC system in place)

## Duplicates are solved in RDV load

Using hash diff to tell the difference and intra load-cycle changes activated in parameters

A duplicate BK that is exactly the same will generate the same HASHDIFF and will be loaded once in the SAT.

A duplicate BK with different values will have a different hashdiff and will be loaded to the SAT with both instances. Otherwise, you lose data. The load date of the second record is augmented by +2 microseconds. Order of the SATs is in this case not exact as we do not know it.



Cmd 7

```
select * from willibald.willibald_dv_fl.sat_web_produkt where PRODUKTID in (20,21);

select PRODUKT_HKEY ,PRODUKTID, EQUAL, HASH_DIFF, PFLANZABSTAND, PFLANZORT, PREIS, TYP, UMFANG from willibald.webshop_stg.sat_web_produkt_tmp where PRODUKTID in (20,21);
```

(1) Spark Jobs

Table

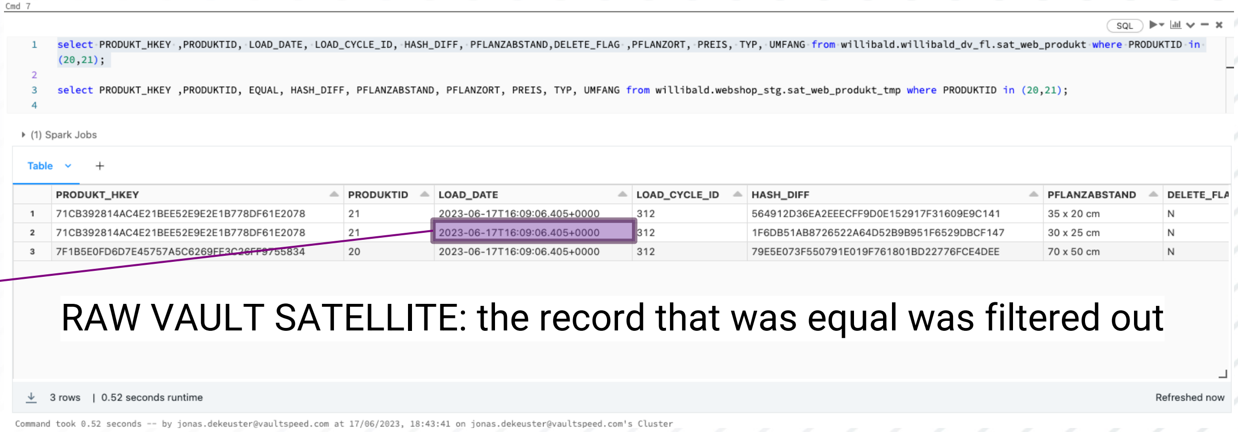
| | PRODUKT_HKEY | PRODUKTID | EQUAL | HASH_DIFF | PFLANZABSTAND | PFLANZORT | PREIS | TYP | UMFANG |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 71CB392814AC4E21BEE52E9E2E1B778DF61E2078 | 21 | 0 | 564912D36EA2EEECFF9D0E152917F31609E9C141 | 35 x 20 cm | Freiland | 3,50 | 3 | 7 |
| 2 | 71CB392814AC4E21BEE52E9E2E1B778DF61E2078 | 21 | 0 | 1F6DB51AB8726522A64D52B9B951F6529DBCF147 | 30 x 25 cm | Freiland | 4,00 | 2 | 6 |
| 3 | 7F1B5E0FD6D7E45757A5C6269FE3C26FF9755834 | 20 | 0 | 79E5E073F550791E019F761801BD22776FCE4DEE | 70 x 50 cm | Freiland | 8,00 | 2 | 6 |
| 4 | 7F1B5E0FD6D7E45757A5C6269FE3C26FF9755834 | 20 | 1 | 79E5E073F550791E019F761801BD22776FCE4DEE | 70 x 50 cm | Freiland | 8,00 | 2 | 6 |

STAGING: notice the EQUAL calculation and identical hashdiffs for productid 20

4

Refreshed now

Command took 0.40 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 18:42:26 on jonas.dekeuster@vaultspeed.com's Cluster



Cmd 7

```
select PRODUKT_HKEY ,PRODUKTID, LOAD_DATE, LOAD_CYCLE_ID, HASH_DIFF, PFLANZABSTAND,DELETE_FLAG ,PFLANZORT, PREIS, TYP, UMFANG from willibald.willibald_dv_fl.sat_web_produkt where PRODUKTID in (20,21);

select PRODUKT_HKEY ,PRODUKTID, EQUAL, HASH_DIFF, PFLANZABSTAND, PFLANZORT, PREIS, TYP, UMFANG from willibald.webshop_stg.sat_web_produkt_tmp where PRODUKTID in (20,21);
```

(1) Spark Jobs

Table

| | PRODUKT_HKEY | PRODUKTID | LOAD_DATE | LOAD_CYCLE_ID | HASH_DIFF | PFLANZABSTAND | DELETE_FLA |
|---|---|---|---|---|---|---|---|
| 1 | 71CB392814AC4E21BEE52E9E2E1B778DF61E2078 | 21 | 2023-06-17T16:09:06.405+0000 | 312 | 564912D36EA2EEECFF9D0E152917F31609E9C141 | 35 x 20 cm | N |
| 2 | 71CB392814AC4E21BEE52E9E2E1B778DF61E2078 | 21 | 2023-06-17T16:09:06.405+0000 | 312 | 1F6DB51AB8726522A64D52B9B951F6529DBCF147 | 30 x 25 cm | N |
| 3 | 7F1B5E0FD6D7E45757A5C6269FE3C26FF9755834 | 20 | 2023-06-17T16:09:06.405+0000 | 312 | 79E5E073F550791E019F761801BD22776FCE4DEE | 70 x 50 cm | N |

RAW VAULT SATELLITE: the record that was equal was filtered out

3 rows | 0.52 seconds runtime

Refreshed now

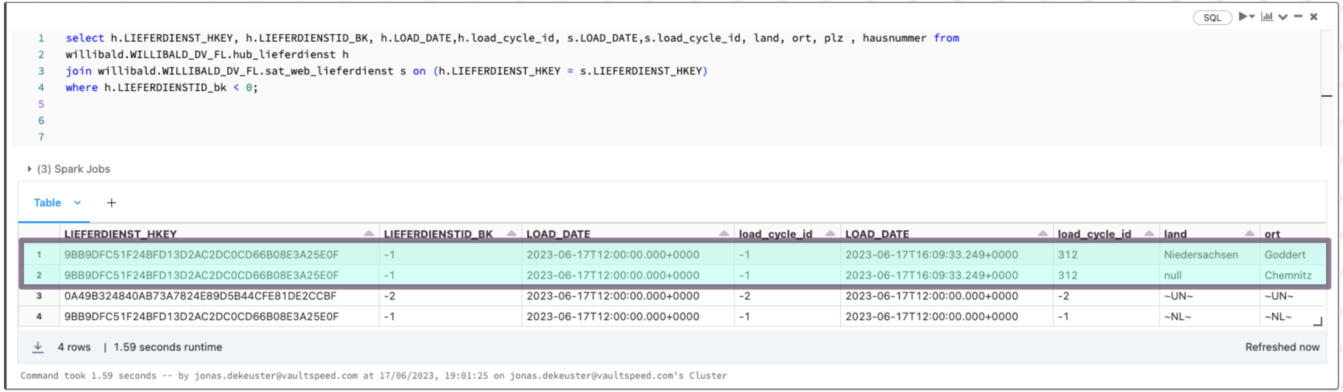
Command took 0.52 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 18:43:41 on jonas.dekeuster@vaultspeed.com's Cluster

# 8. Rows without business key

## Missing BK's are solved in the Raw Vault

Since the business key is not provided, we cannot calculate the hash key. That is why these records are loaded into the sat using the unknown record in the hub.

If the records are later on corrected, they will be loaded in the hub as well, and the old record will get delete flag = Y because it no longer exists.



```sql
select h.LIEFERDIENST_HKEY, h.LIEFERDIENSTID_BK, h.LOAD_DATE,h.load_cycle_id, s.LOAD_DATE,s.load_cycle_id, land, ort, plz , hausnummer from
willibald.WILLIBALD_DV_FL.hub_lieferdienst h
join willibald.WILLIBALD_DV_FL.sat_web_lieferdienst s on (h.LIEFERDIENST_HKEY = s.LIEFERDIENST_HKEY)
where h.LIEFERDIENSTID_bk < 0;
```

| | LIEFERDIENST_HKEY | LIEFERDIENSTID_BK | LOAD_DATE | load_cycle_id | LOAD_DATE | load_cycle_id | land | ort |
|---|---|---|---|---|---|---|---|---|
| 1 | 9BB9DFC51F24BFD13D2AC2DC0CD66B08E3A25E0F | -1 | 2023-06-17T12:00:00.000+0000 | -1 | 2023-06-17T16:09:33.249+0000 | 312 | Niedersachsen | Goddert |
| 2 | 9BB9DFC51F24BFD13D2AC2DC0CD66B08E3A25E0F | -1 | 2023-06-17T12:00:00.000+0000 | -1 | 2023-06-17T16:09:33.249+0000 | 312 | null | Chemnitz |
| 3 | 0A49B324840AB73A7824E89D5B44CFE81DE2CCBF | -2 | 2023-06-17T12:00:00.000+0000 | -2 | 2023-06-17T12:00:00.000+0000 | -2 | ~UN~ | ~UN~ |
| 4 | 9BB9DFC51F24BFD13D2AC2DC0CD66B08E3A25E0F | -1 | 2023-06-17T12:00:00.000+0000 | -1 | 2023-06-17T12:00:00.000+0000 | -1 | ~NL~ | ~NL~ |

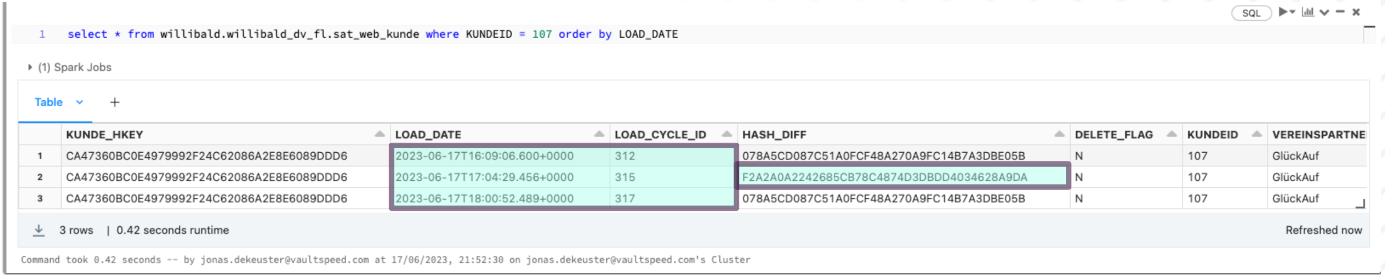
4 rows | 1.59 seconds runtime

Refreshed now

```
Command took 1.59 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 19:01:25 on jonas.dekeuster@vaultspeed.com's Cluster
```

VAULTSPEED

# 9. Changes in the customer



```
select * from willibald.willibald_dv_fl.sat_web_kunde where KUNDEID = 107 order by LOAD_DATE
```

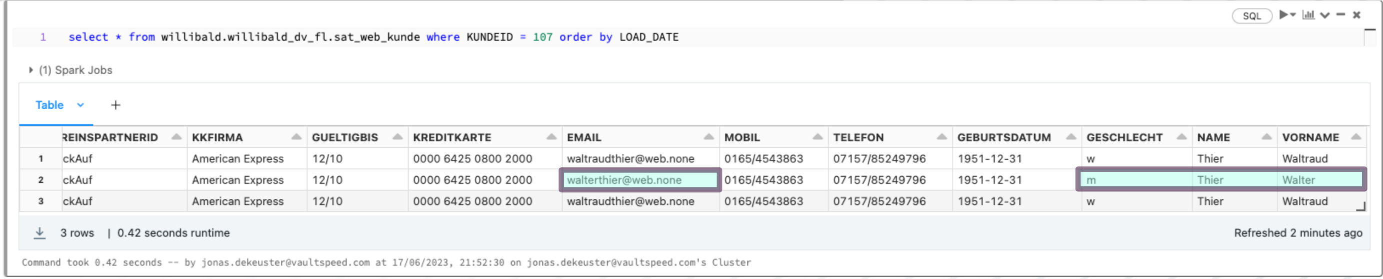
(1) Spark Jobs

| | KUNDE_HKEY | LOAD_DATE | LOAD_CYCLE_ID | HASH_DIFF | DELETE_FLAG | KUNDEID | VEREINSPARTNE |
|---|---|---|---|---|---|---|---|
| 1 | CA47360BC0E4979992F24C62086A2E8E6089DDD6 | 2023-06-17T16:09:06.600+0000 | 312 | 078A5CD087C51A0FCF48A270A9FC14B7A3DBE05B | N | 107 | GlückAuf |
| 2 | CA47360BC0E4979992F24C62086A2E8E6089DDD6 | 2023-06-17T17:04:29.456+0000 | 315 | F2A2A0A2242685CB78C4874D3DBDD4034628A9DA | N | 107 | GlückAuf |
| 3 | CA47360BC0E4979992F24C62086A2E8E6089DDD6 | 2023-06-17T18:00:52.489+0000 | 317 | 078A5CD087C51A0FCF48A270A9FC14B7A3DBE05B | N | 107 | GlückAuf |

3 rows | 0.42 seconds runtime

Refreshed now

Command took 0.42 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 21:52:30 on jonas.dekeuster@vaultspeed.com's Cluster



```
select * from willibald.willibald_dv_fl.sat_web_kunde where KUNDEID = 107 order by LOAD_DATE
```

(1) Spark Jobs

| | REINSPARTNERID | KKFIRMA | GUELTIGBIS | KREDITKARTE | EMAIL | MOBIL | TELEFON | GEBURTSDATUM | GESCHLECHT | NAME | VORNAME |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ckAuf | American Express | 12/10 | 0000 6425 0800 2000 | waltraudthier@web.none | 0165/4543863 | 07157/85249796 | 1951-12-31 | w | Thier | Waltraud |
| 2 | ckAuf | American Express | 12/10 | 0000 6425 0800 2000 | walterthier@web.none | 0165/4543863 | 07157/85249796 | 1951-12-31 | m | Thier | Walter |
| 3 | ckAuf | American Express | 12/10 | 0000 6425 0800 2000 | waltraudthier@web.none | 0165/4543863 | 07157/85249796 | 1951-12-31 | w | Thier | Waltraud |

3 rows | 0.42 seconds runtime

Refreshed 2 minutes ago

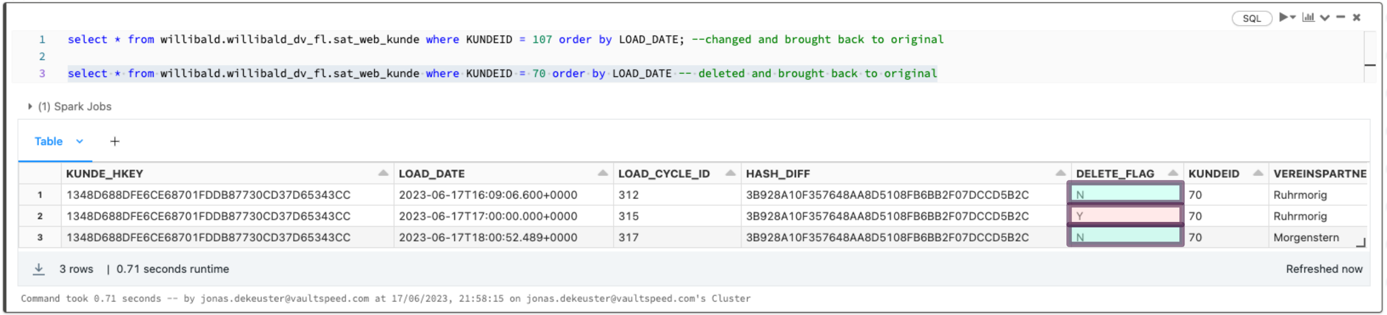
Command took 0.42 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 21:52:30 on jonas.dekeuster@vaultspeed.com's Cluster

## Changes in source records

These are simply changed by the standard DV2 approach. Using the hashdiff to identify changes. In this case the satellite will have 3 instances of this customer.

VAULTSPEED

# 10. Deletions in customer



```sql
select * from willibald.willibald_dv_fl.sat_web_kunde where KUNDEID = 107 order by LOAD_DATE; --changed and brought back to original

select * from willibald.willibald_dv_fl.sat_web_kunde where KUNDEID = 70 order by LOAD_DATE -- deleted and brought back to original
```

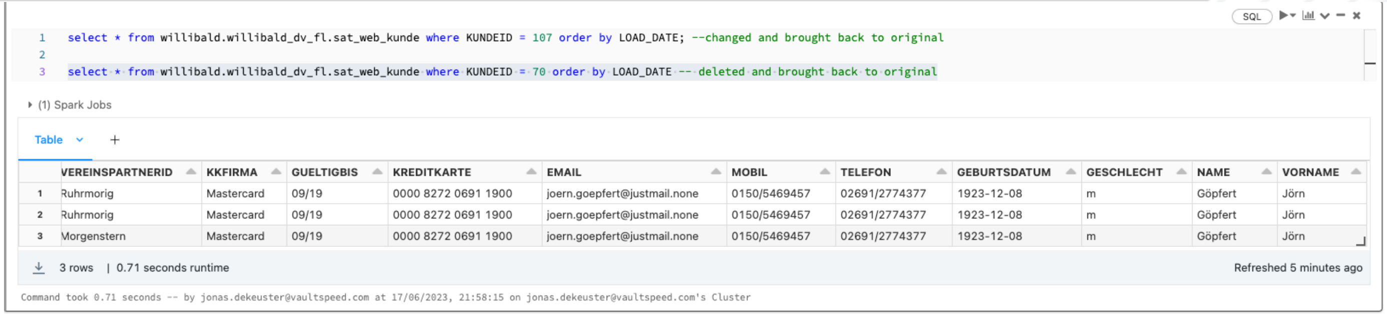
(1) Spark Jobs

| | KUNDE_HKEY | LOAD_DATE | LOAD_CYCLE_ID | HASH_DIFF | DELETE_FLAG | KUNDEID | VEREINSPARTNE |
|---|---|---|---|---|---|---|---|
| 1 | 1348D688DFE6CE68701FDDB87730CD37D65343CC | 2023-06-17T16:09:06.600+0000 | 312 | 3B928A10F357648AA8D5108FB6BB2F07DCCD5B2C | N | 70 | Ruhrmorig |
| 2 | 1348D688DFE6CE68701FDDB87730CD37D65343CC | 2023-06-17T17:00:00.000+0000 | 315 | 3B928A10F357648AA8D5108FB6BB2F07DCCD5B2C | Y | 70 | Ruhrmorig |
| 3 | 1348D688DFE6CE68701FDDB87730CD37D65343CC | 2023-06-17T18:00:52.489+0000 | 317 | 3B928A10F357648AA8D5108FB6BB2F07DCCD5B2C | N | 70 | Morgenstern |

3 rows | 0.71 seconds runtime

Refreshed now

Command took 0.71 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 21:58:15 on jonas.dekeuster@vaultspeed.com's Cluster



```sql
select * from willibald.willibald_dv_fl.sat_web_kunde where KUNDEID = 107 order by LOAD_DATE; --changed and brought back to original

select * from willibald.willibald_dv_fl.sat_web_kunde where KUNDEID = 70 order by LOAD_DATE -- deleted and brought back to original
```

(1) Spark Jobs

| | VEREINSPARTNERID | KKFIRMA | GUELTIGBIS | KREDITKARTE | EMAIL | MOBIL | TELEFON | GEBURTSDATUM | GESCHLECHT | NAME | VORNAME |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Ruhrmorig | Mastercard | 09/19 | 0000 8272 0691 1900 | joern.goepfert@justmail.none | 0150/5469457 | 02691/2774377 | 1923-12-08 | m | Göpfert | Jörn |
| 2 | Ruhrmorig | Mastercard | 09/19 | 0000 8272 0691 1900 | joern.goepfert@justmail.none | 0150/5469457 | 02691/2774377 | 1923-12-08 | m | Göpfert | Jörn |
| 3 | Morgenstern | Mastercard | 09/19 | 0000 8272 0691 1900 | joern.goepfert@justmail.none | 0150/5469457 | 02691/2774377 | 1923-12-08 | m | Göpfert | Jörn |

3 rows | 0.71 seconds runtime

Refreshed 5 minutes ago

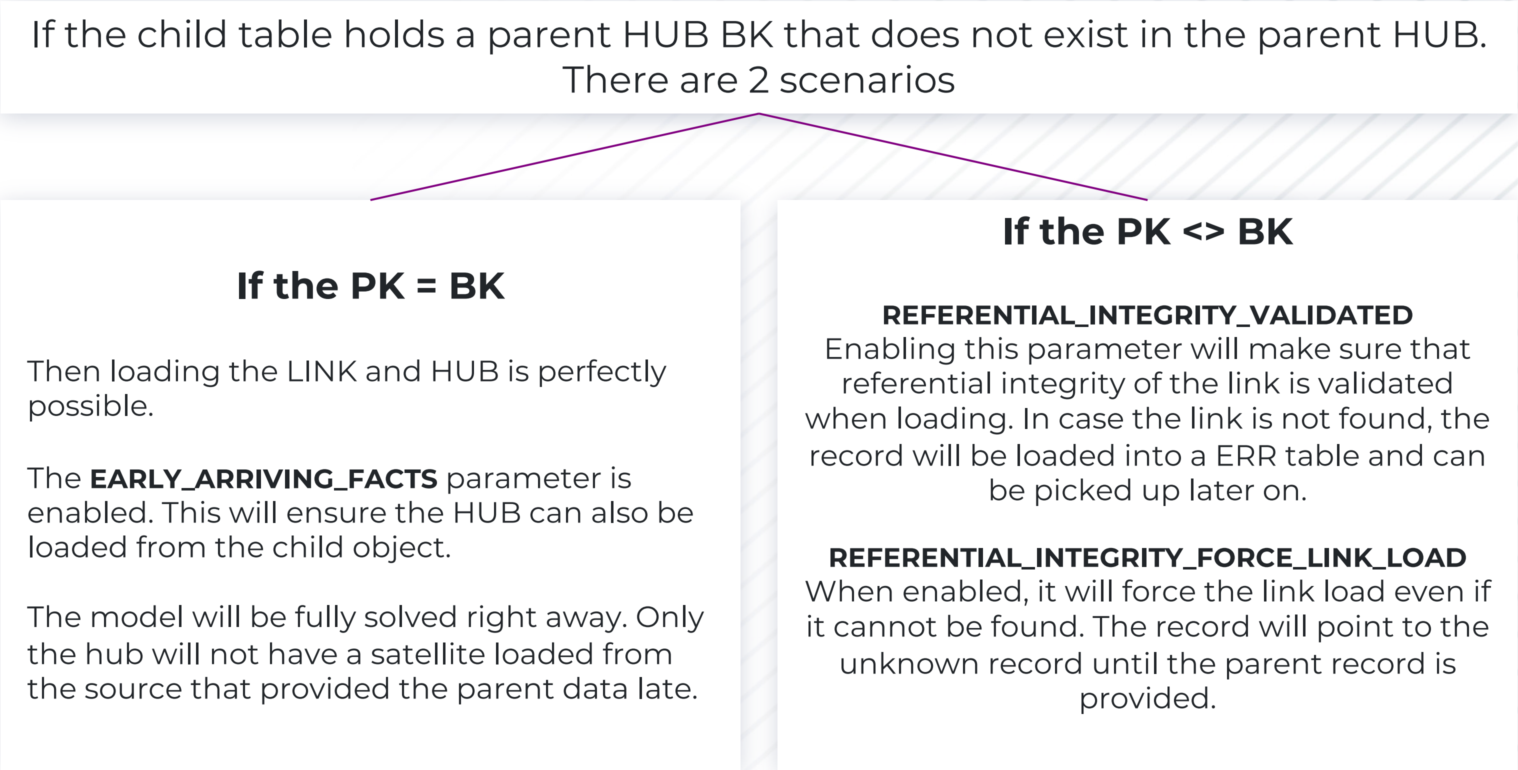
Command took 0.71 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 21:58:15 on jonas.dekeuster@vaultspeed.com's Cluster

## Deletes in source records

These are simply changed by the standard DV2 approach. Using the delete flag to indicate a record was deleted. In this case the satellite will have 3 instances of this customer. 1 with delete flag = Y .

Note that VaultSpeed can also do end dating logic. In that case you would only 2 SAT records. The first one would be end dated on the delete.

# 11. Adresses without customers



If the child table holds a parent HUB BK that does not exist in the parent HUB.
There are 2 scenarios

## If the PK = BK

Then loading the LINK and HUB is perfectly possible.

The **EARLY_ARRIVING_FACTS** parameter is enabled. This will ensure the HUB can also be loaded from the child object.

The model will be fully solved right away. Only the hub will not have a satellite loaded from the source that provided the parent data late.

## If the PK <> BK

**REFERENTIAL_INTEGRITY_VALIDATED**
Enabling this parameter will make sure that referential integrity of the link is validated when loading. In case the link is not found, the record will be loaded into a ERR table and can be picked up later on.

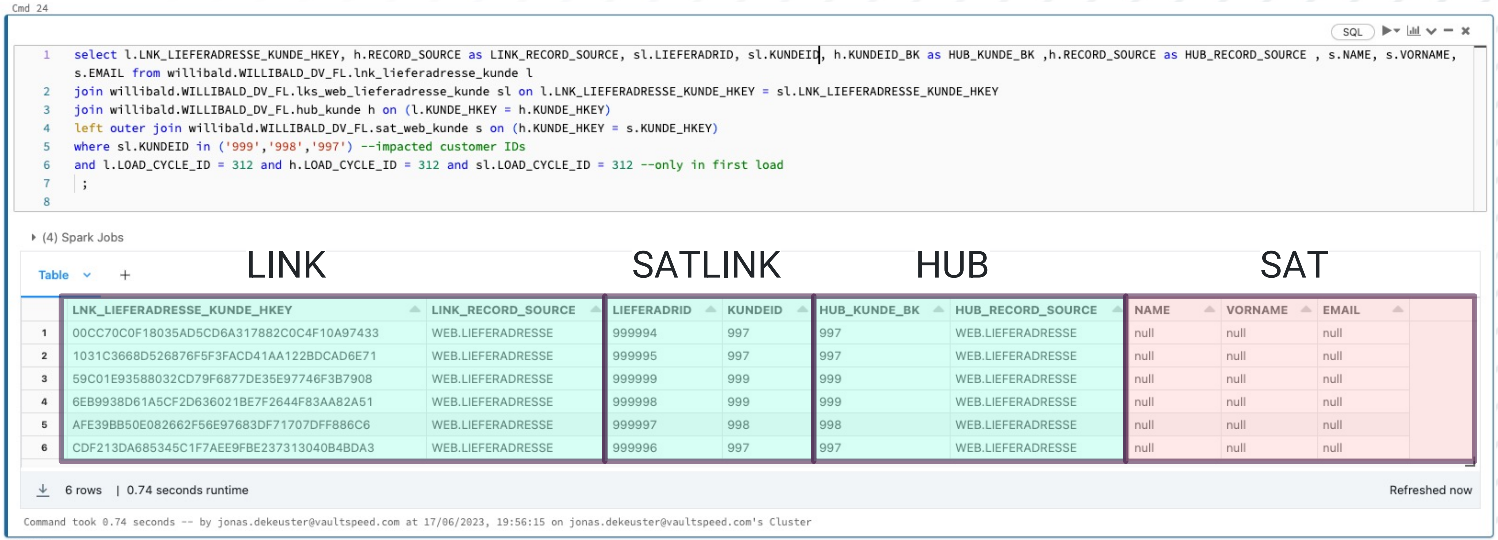
**REFERENTIAL_INTEGRITY_FORCE_LINK_LOAD**
When enabled, it will force the link load even if it cannot be found. The record will point to the unknown record until the parent record is provided.

VAULTSPEED

# 11. Adresses without customers

## Show test cases in report

The exercise only had examples where PK = BK. There are  no error tables needed here, VaultSpeed and the DV2 standard fixes this for us



```
select l.LNK_LIEFERADRESSE_KUNDE_HKEY, h.RECORD_SOURCE as LINK_RECORD_SOURCE, sl.LIEFERADRID, sl.KUNDEID, h.KUNDEID_BK as HUB_KUNDE_BK ,h.RECORD_SOURCE as HUB_RECORD_SOURCE , s.NAME, s.VORNAME,
s.EMAIL from willibald.WILLIBALD_DV_FL.lnk_lieferadresse_kunde l
join willibald.WILLIBALD_DV_FL.lks_web_lieferadresse_kunde sl on l.LNK_LIEFERADRESSE_KUNDE_HKEY = sl.LNK_LIEFERADRESSE_KUNDE_HKEY
join willibald.WILLIBALD_DV_FL.hub_kunde h on (l.KUNDE_HKEY = h.KUNDE_HKEY)
left outer join willibald.WILLIBALD_DV_FL.sat_web_kunde s on (h.KUNDE_HKEY = s.KUNDE_HKEY)
where sl.KUNDEID in ('999','998','997') --impacted customer IDs
and l.LOAD_CYCLE_ID = 312 and h.LOAD_CYCLE_ID = 312 and sl.LOAD_CYCLE_ID = 312 --only in first load
 ;
```

| | LINK | | SATLINK | | HUB | | SAT | | |
|---|---|---|---|---|---|---|---|---|---|
| | LNK_LIEFERADRESSE_KUNDE_HKEY | LINK_RECORD_SOURCE | LIEFERADRID | KUNDEID | HUB_KUNDE_BK | HUB_RECORD_SOURCE | NAME | VORNAME | EMAIL |
| 1 | 00CC70C0F18035AD5CD6A317882C0C4F10A97433 | WEB.LIEFERADRESSE | 999994 | 997 | 997 | WEB.LIEFERADRESSE | null | null | null |
| 2 | 1031C3668D526876F5F3FACD41AA122BDCAD6E71 | WEB.LIEFERADRESSE | 999995 | 997 | 997 | WEB.LIEFERADRESSE | null | null | null |
| 3 | 59C01E93588032CD79F6877DE35E97746F3B7908 | WEB.LIEFERADRESSE | 999999 | 999 | 999 | WEB.LIEFERADRESSE | null | null | null |
| 4 | 6EB9938D61A5CF2D636021BE7F2644F83AA82A51 | WEB.LIEFERADRESSE | 999998 | 999 | 999 | WEB.LIEFERADRESSE | null | null | null |
| 5 | AFE39BB50E082662F56E97683DF71707DFF886C6 | WEB.LIEFERADRESSE | 999997 | 998 | 998 | WEB.LIEFERADRESSE | null | null | null |
| 6 | CDF213DA685345C1F7AEE9FBE237313040B4BDA3 | WEB.LIEFERADRESSE | 999996 | 997 | 997 | WEB.LIEFERADRESSE | null | null | null |

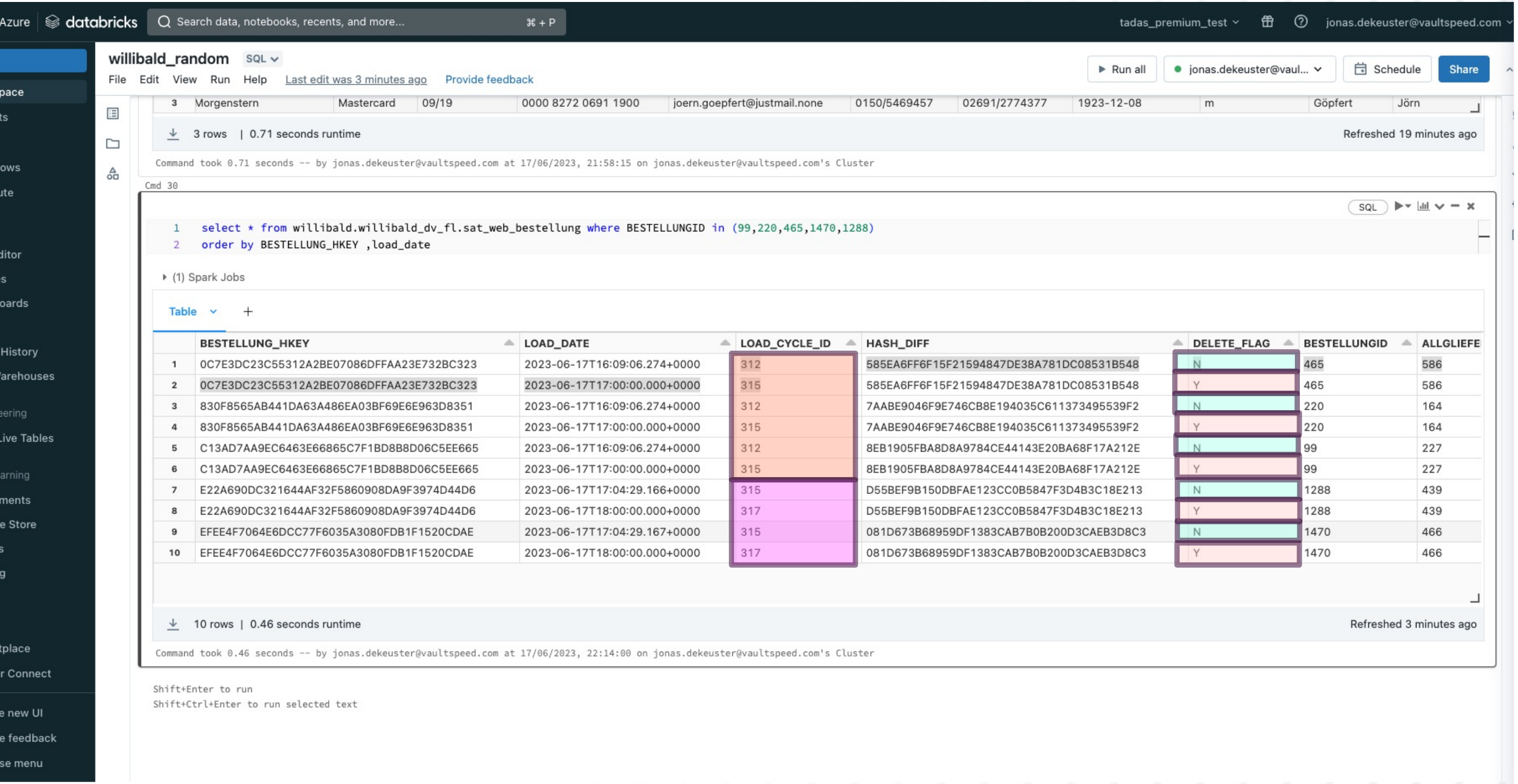
6 rows | 0.74 seconds runtime

Command took 0.74 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 19:56:15 on jonas.dekeuster@vaultspeed.com's Cluster

Note that the hub will be loaded from the link (you can see this in the record source) **EARLY_ARRIVING_FACTS** parameter

Note that the sat will of course not be populated for these records, as there is no valid SAT data for them

# 12. Deletions of orders



| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | Morgenstern | Mastercard | 09/19 | 0000 8272 0691 1900 | joern.goepfert@justmail.none | 0150/5469457 | 02691/2774377 | 1923-12-08 | m | Göpfert | Jörn |

3 rows | 0.71 seconds runtime — Refreshed 19 minutes ago

```
Command took 0.71 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 21:58:15 on jonas.dekeuster@vaultspeed.com's Cluster
```

```sql
select * from willibald.willibald_dv_fl.sat_web_bestellung where BESTELLUNGID in (99,220,465,1470,1288)
order by BESTELLUNG_HKEY ,load_date
```

| | BESTELLUNG_HKEY | LOAD_DATE | LOAD_CYCLE_ID | HASH_DIFF | DELETE_FLAG | BESTELLUNGID | ALLGLIEFE |
|---|---|---|---|---|---|---|---|
| 1 | 0C7E3DC23C55312A2BE07086DFFAA23E732BC323 | 2023-06-17T16:09:06.274+0000 | 312 | 585EA6FF6F15F21594847DE38A781DC08531B548 | N | 465 | 586 |
| 2 | 0C7E3DC23C55312A2BE07086DFFAA23E732BC323 | 2023-06-17T17:00:00.000+0000 | 315 | 585EA6FF6F15F21594847DE38A781DC08531B548 | Y | 465 | 586 |
| 3 | 830F8565AB441DA63A486EA03BF69E6E963D8351 | 2023-06-17T16:09:06.274+0000 | 312 | 7AABE9046F9E746CB8E194035C611373495539F2 | N | 220 | 164 |
| 4 | 830F8565AB441DA63A486EA03BF69E6E963D8351 | 2023-06-17T17:00:00.000+0000 | 315 | 7AABE9046F9E746CB8E194035C611373495539F2 | Y | 220 | 164 |
| 5 | C13AD7AA9EC6463E66865C7F1BD8B8D06C5EE665 | 2023-06-17T16:09:06.274+0000 | 312 | 8EB1905FBA8D8A9784CE44143E20BA68F17A212E | N | 99 | 227 |
| 6 | C13AD7AA9EC6463E66865C7F1BD8B8D06C5EE665 | 2023-06-17T17:00:00.000+0000 | 315 | 8EB1905FBA8D8A9784CE44143E20BA68F17A212E | Y | 99 | 227 |
| 7 | E22A690DC321644AF32F5860908DA9F3974D44D6 | 2023-06-17T17:04:29.166+0000 | 315 | D55BEF9B150DBFAE123CC0B5847F3D4B3C18E213 | N | 1288 | 439 |
| 8 | E22A690DC321644AF32F5860908DA9F3974D44D6 | 2023-06-17T18:00:00.000+0000 | 317 | D55BEF9B150DBFAE123CC0B5847F3D4B3C18E213 | Y | 1288 | 439 |
| 9 | EFEE4F7064E6DCC77F6035A3080FDB1F1520CDAE | 2023-06-17T17:04:29.167+0000 | 315 | 081D673B68959DF1383CAB7B0B200D3CAEB3D8C3 | N | 1470 | 466 |
| 10 | EFEE4F7064E6DCC77F6035A3080FDB1F1520CDAE | 2023-06-17T18:00:00.000+0000 | 317 | 081D673B68959DF1383CAB7B0B200D3CAEB3D8C3 | Y | 1470 | 466 |

10 rows | 0.46 seconds runtime — Refreshed 3 minutes ago

```
Command took 0.46 seconds -- by jonas.dekeuster@vaultspeed.com at 17/06/2023, 22:14:00 on jonas.dekeuster@vaultspeed.com's Cluster

Shift+Enter to run
Shift+Ctrl+Enter to run selected text
```

## Deletes in source records

These are simply changed by the standard DV2 approach and this logic comes OOTB with VaultSpeed.

Using the delete flag to indicate a record was deleted. The satellite for orders will have delete flag = Y for the orders that were deleted.

Note that VaultSpeed can also do end dating logic. In that case the SAT record would be end dated on the delete timestamp.

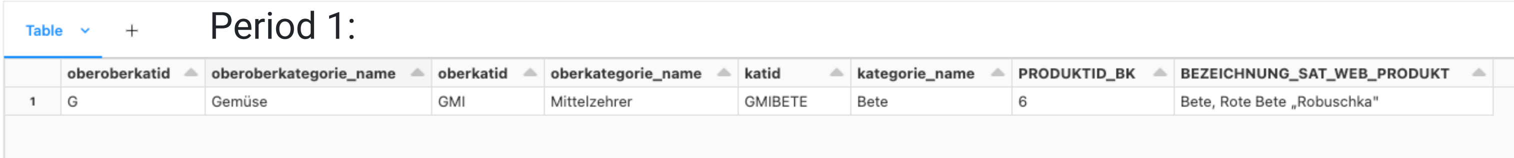
VAULTSPEED

# 13. Changes in the dimensions (product hierarchy)

## Changes in hierarchy

The product hierarchy changes can be captured by using the setup described in use case 1

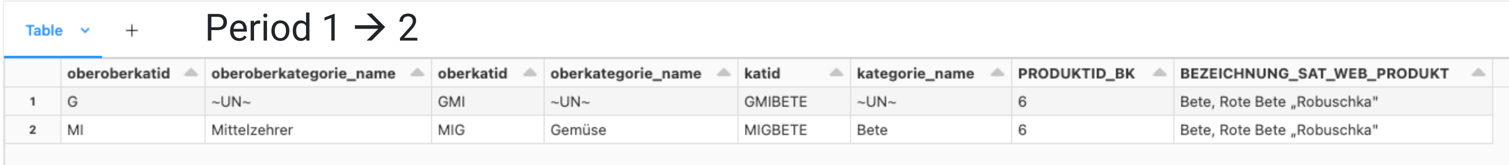
For example consider product id 6



Period 1:

Table ⌄ +

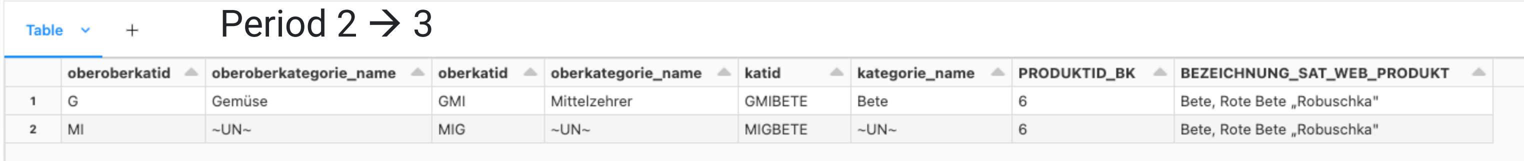
| | oberoberkatid | oberoberkategorie_name | oberkatid | oberkategorie_name | katid | kategorie_name | PRODUKTID_BK | BEZEICHNUNG_SAT_WEB_PRODUKT |
|---|---|---|---|---|---|---|---|---|
| 1 | G | Gemüse | GMI | Mittelzehrer | GMIBETE | Bete | 6 | Bete, Rote Bete „Robuschka" |



Period 1 → 2

Table ⌄ +

| | oberoberkatid | oberoberkategorie_name | oberkatid | oberkategorie_name | katid | kategorie_name | PRODUKTID_BK | BEZEICHNUNG_SAT_WEB_PRODUKT |
|---|---|---|---|---|---|---|---|---|
| 1 | G | ~UN~ | GMI | ~UN~ | GMIBETE | ~UN~ | 6 | Bete, Rote Bete „Robuschka" |
| 2 | MI | Mittelzehrer | MIG | Gemüse | MIGBETE | Bete | 6 | Bete, Rote Bete „Robuschka" |

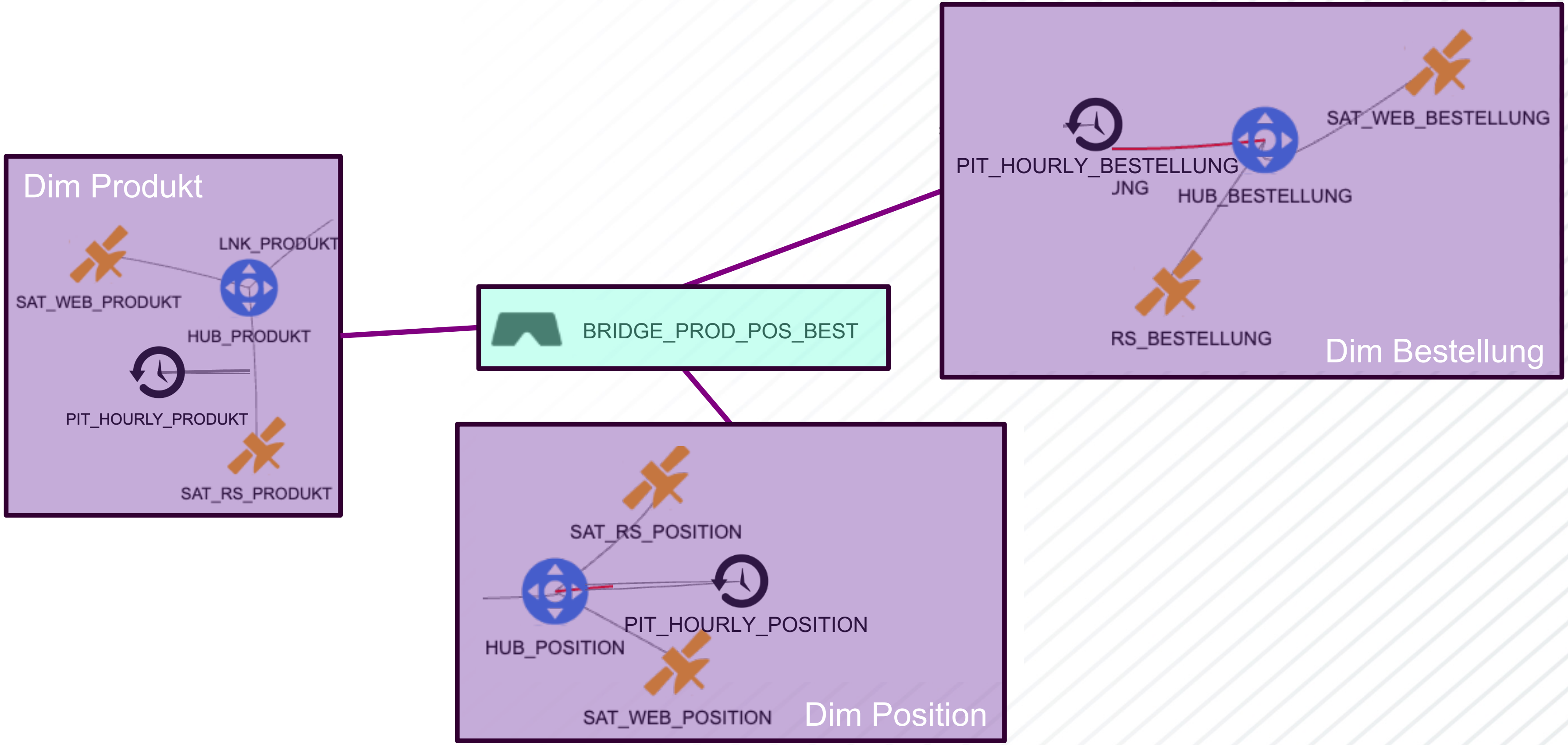


Period 2 → 3

Table ⌄ +

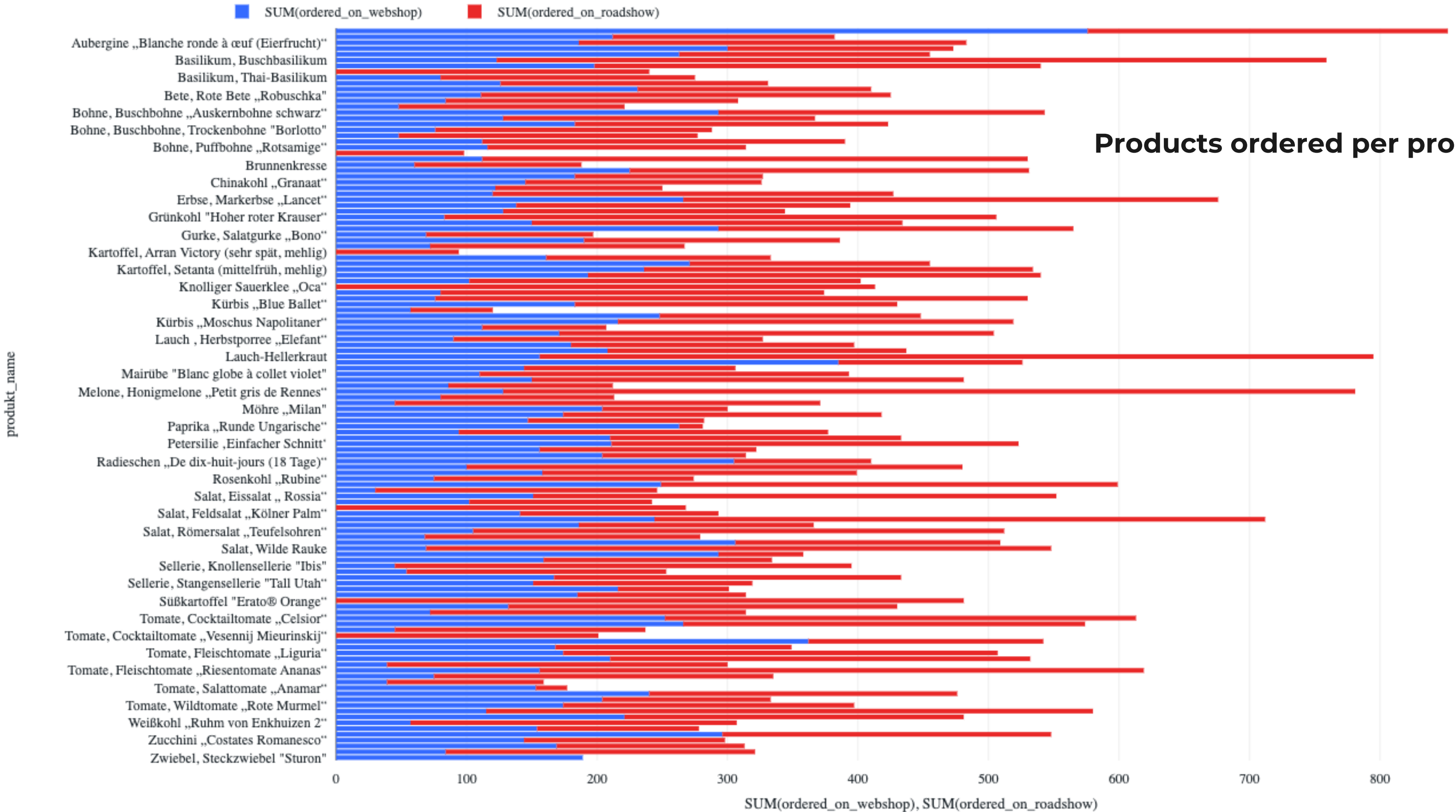
| | oberoberkatid | oberoberkategorie_name | oberkatid | oberkategorie_name | katid | kategorie_name | PRODUKTID_BK | BEZEICHNUNG_SAT_WEB_PRODUKT |
|---|---|---|---|---|---|---|---|---|
| 1 | G | Gemüse | GMI | Mittelzehrer | GMIBETE | Bete | 6 | Bete, Rote Bete „Robuschka" |
| 2 | MI | ~UN~ | MIG | ~UN~ | MIGBETE | ~UN~ | 6 | Bete, Rote Bete „Robuschka" |

# 14. Key figures as result control



VAULTSPEED

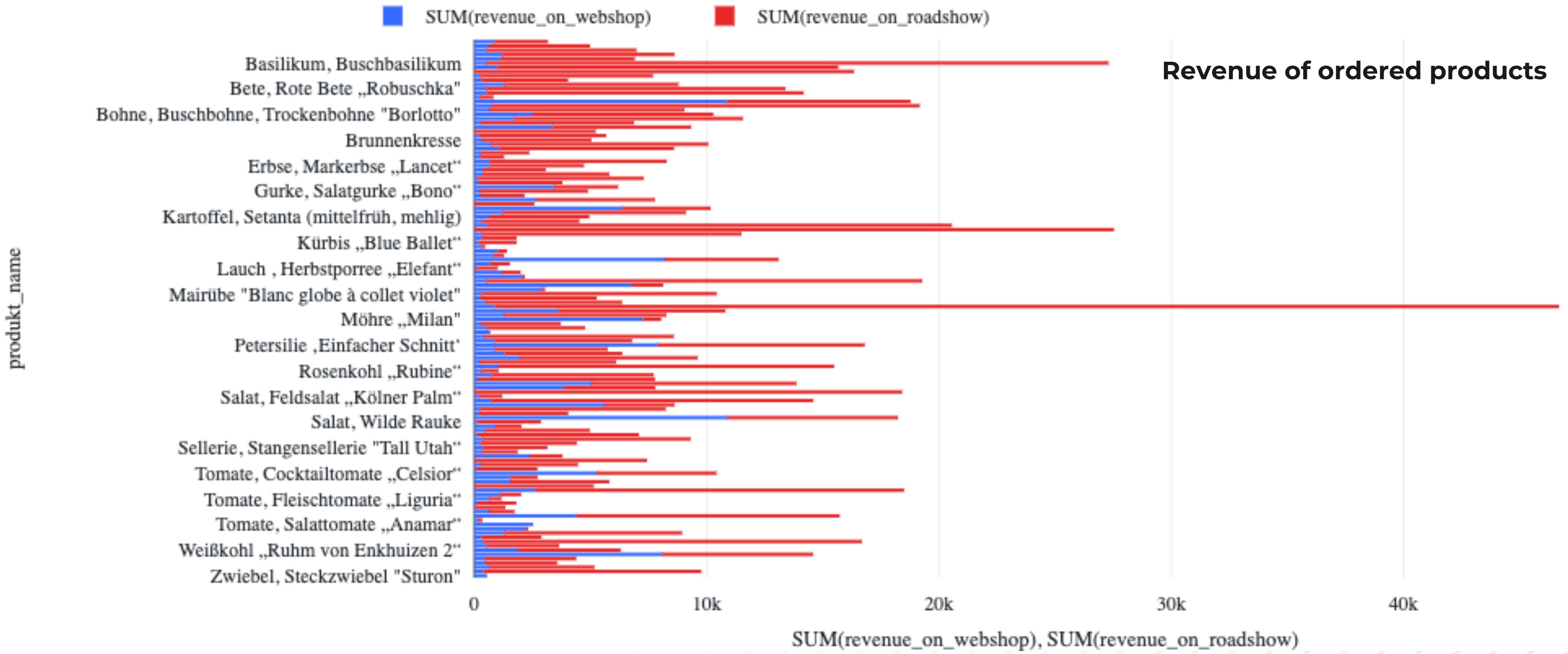
# 14. Key figures as result control



Products ordered per product

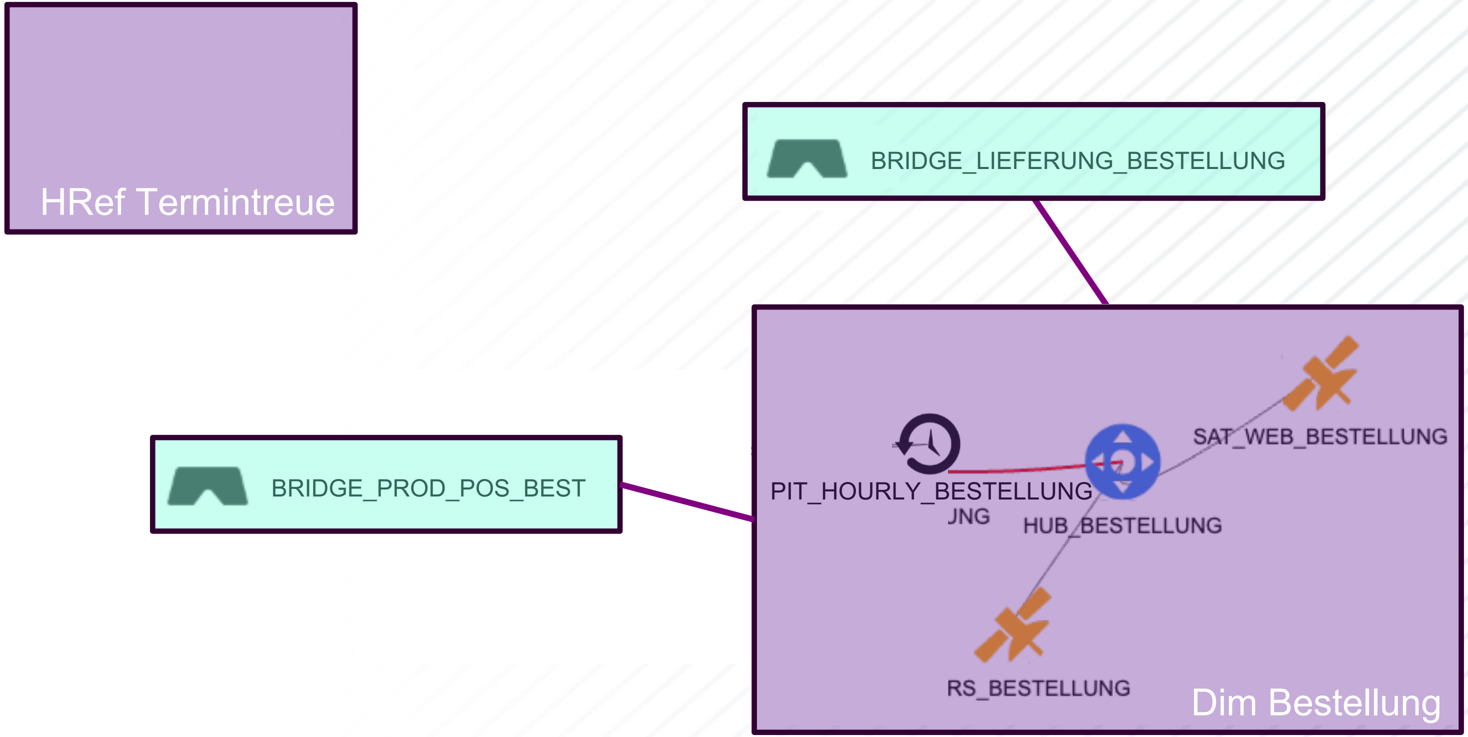
VAULTSPEED

# 14. Key figures as result control



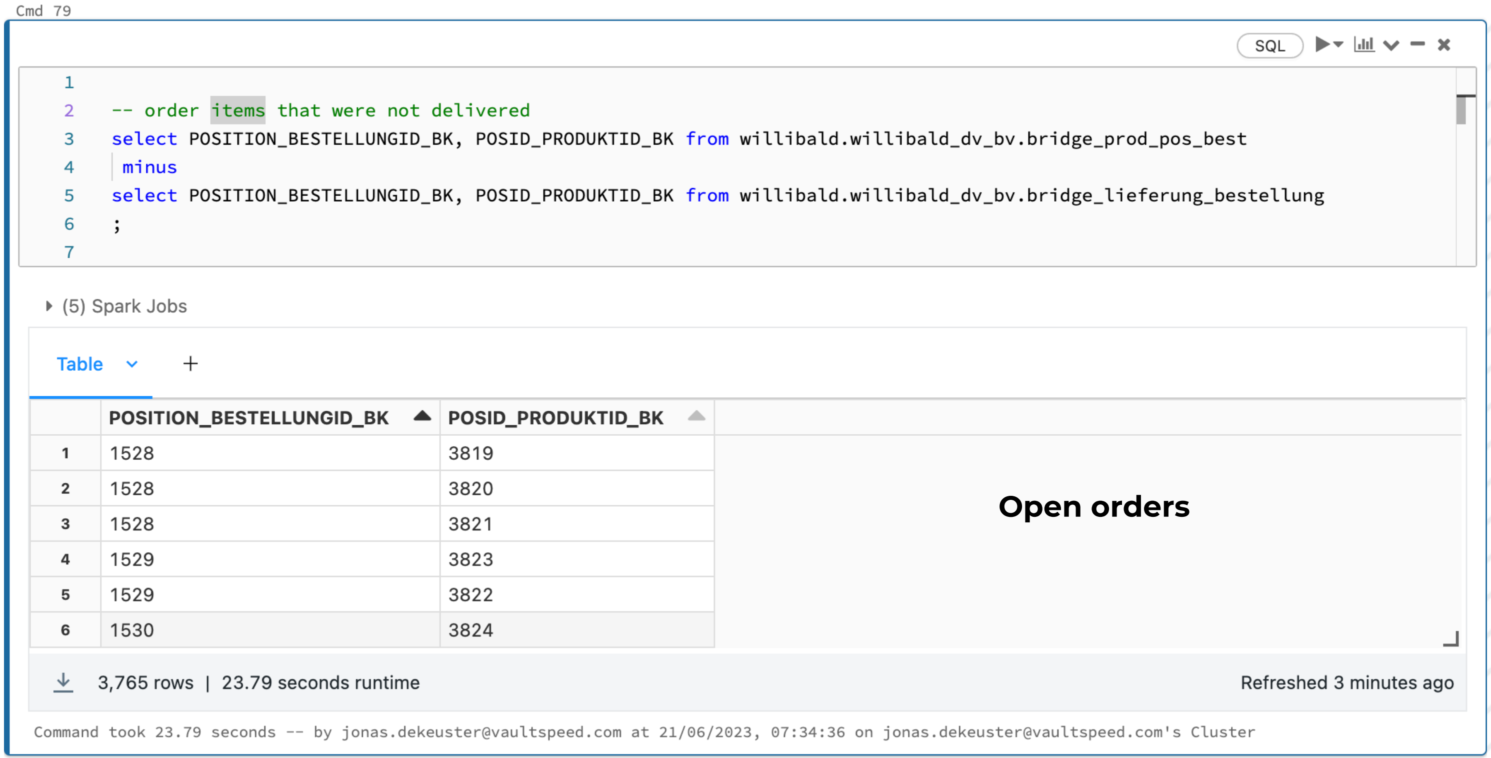
Revenue of ordered products

# 14. Key figures as result control



VAULTSPEED

# 14. Key figures as result control



Cmd 79

```
-- order items that were not delivered
select POSITION_BESTELLUNGID_BK, POSID_PRODUKTID_BK from willibald.willibald_dv_bv.bridge_prod_pos_best
 minus
select POSITION_BESTELLUNGID_BK, POSID_PRODUKTID_BK from willibald.willibald_dv_bv.bridge_lieferung_bestellung
;
```

(5) Spark Jobs

| | POSITION_BESTELLUNGID_BK | POSID_PRODUKTID_BK |
|---|---|---|
| 1 | 1528 | 3819 |
| 2 | 1528 | 3820 |
| 3 | 1528 | 3821 |
| 4 | 1529 | 3823 |
| 5 | 1529 | 3822 |
| 6 | 1530 | 3824 |

**Open orders**

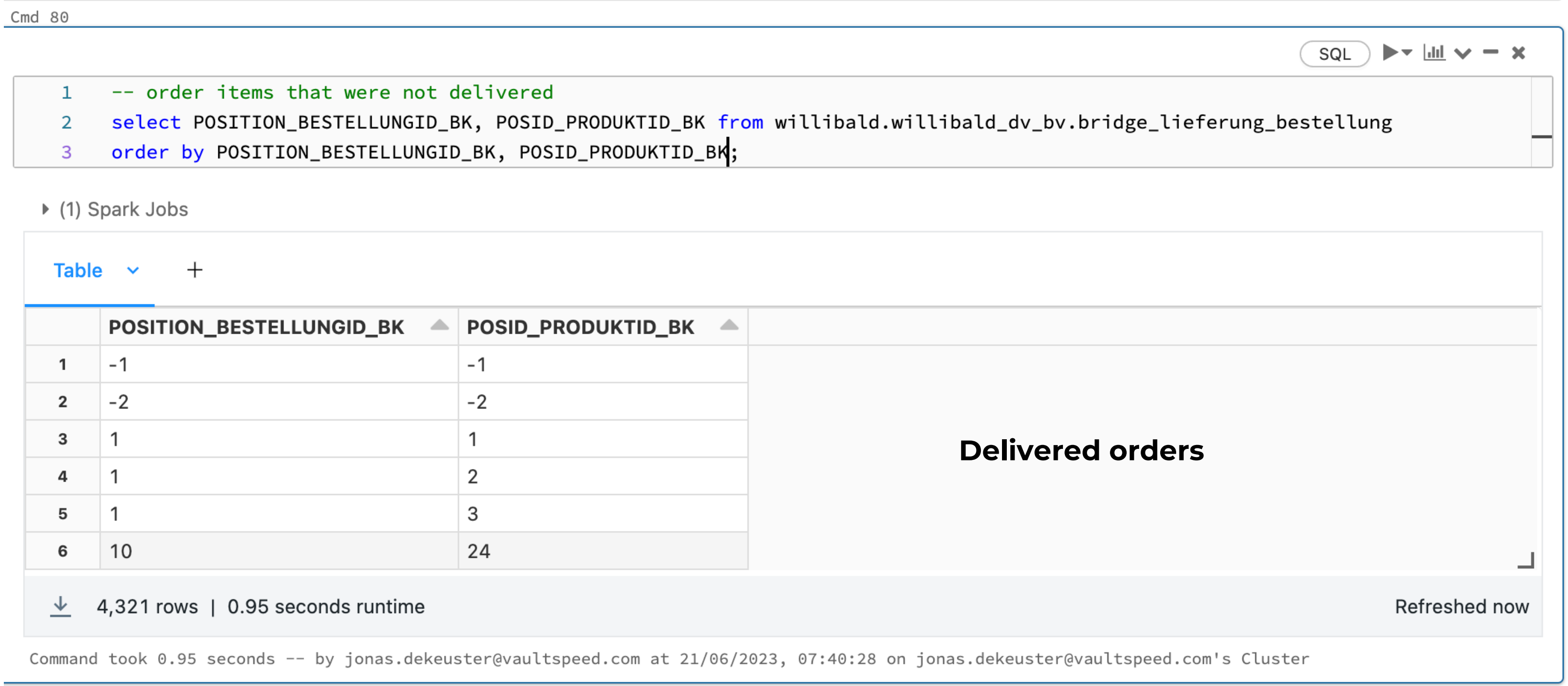
3,765 rows | 23.79 seconds runtime

Refreshed 3 minutes ago

```
Command took 23.79 seconds -- by jonas.dekeuster@vaultspeed.com at 21/06/2023, 07:34:36 on jonas.dekeuster@vaultspeed.com's Cluster
```

VAULTSPEED

# 14. Key figures as result control



Cmd 80

```
-- order items that were not delivered
select POSITION_BESTELLUNGID_BK, POSID_PRODUKTID_BK from willibald.willibald_dv_bv.bridge_lieferung_bestellung
order by POSITION_BESTELLUNGID_BK, POSID_PRODUKTID_BK;
```

| | POSITION_BESTELLUNGID_BK | POSID_PRODUKTID_BK |
|---|---|---|
| 1 | -1 | -1 |
| 2 | -2 | -2 |
| 3 | 1 | 1 |
| 4 | 1 | 2 |
| 5 | 1 | 3 |
| 6 | 10 | 24 |

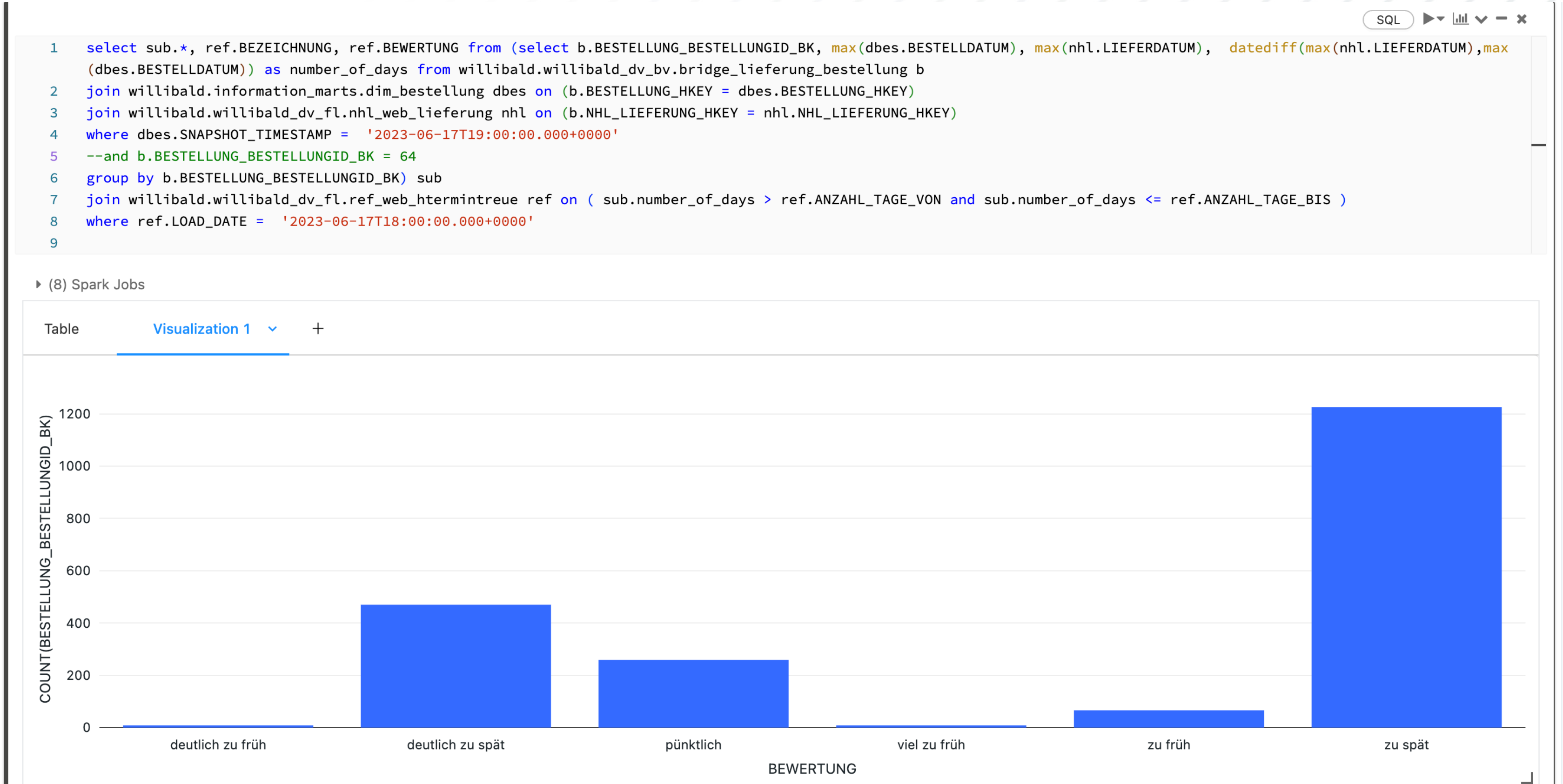
**Delivered orders**

4,321 rows | 0.95 seconds runtime

Refreshed now

Command took 0.95 seconds -- by jonas.dekeuster@vaultspeed.com at 21/06/2023, 07:40:28 on jonas.dekeuster@vaultspeed.com's Cluster

# 14. Key figures as result control



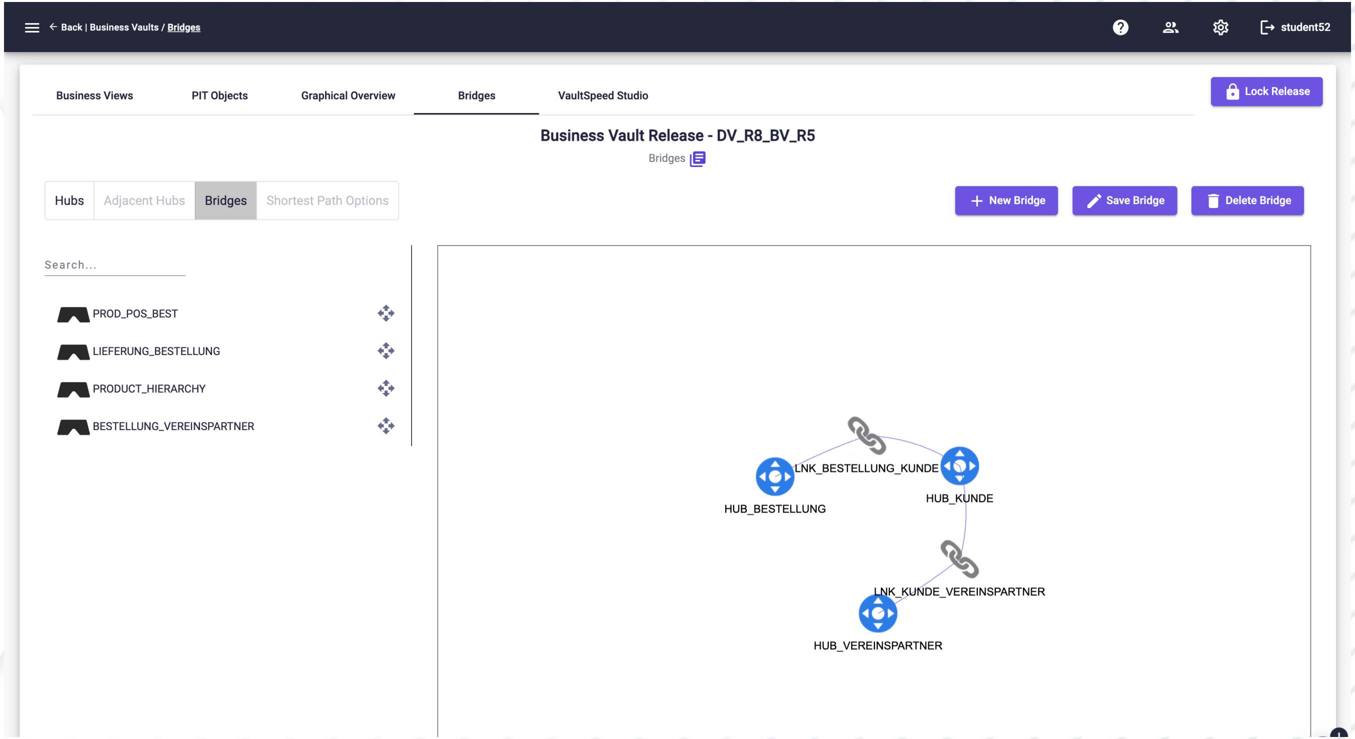
```sql
select sub.*, ref.BEZEICHNUNG, ref.BEWERTUNG from (select b.BESTELLUNG_BESTELLUNGID_BK, max(dbes.BESTELLDATUM), max(nhl.LIEFERDATUM),  datediff(max(nhl.LIEFERDATUM),max(dbes.BESTELLDATUM)) as number_of_days from willibald.willibald_dv_bv.bridge_lieferung_bestellung b
join willibald.information_marts.dim_bestellung dbes on (b.BESTELLUNG_HKEY = dbes.BESTELLUNG_HKEY)
join willibald.willibald_dv_fl.nhl_web_lieferung nhl on (b.NHL_LIEFERUNG_HKEY = nhl.NHL_LIEFERUNG_HKEY)
where dbes.SNAPSHOT_TIMESTAMP =  '2023-06-17T19:00:00.000+0000'
--and b.BESTELLUNG_BESTELLUNGID_BK = 64
group by b.BESTELLUNG_BESTELLUNGID_BK) sub
join willibald.willibald_dv_fl.ref_web_htermintreue ref on ( sub.number_of_days > ref.ANZAHL_TAGE_VON and sub.number_of_days <= ref.ANZAHL_TAGE_BIS )
where ref.LOAD_DATE =  '2023-06-17T18:00:00.000+0000'

```

# 15. Business rules

**Identify the customer if the name is not given.**

Create a bridge to “bridge” the gap between order and association partner.

On top of the bridge, you can build custom SQL logic.

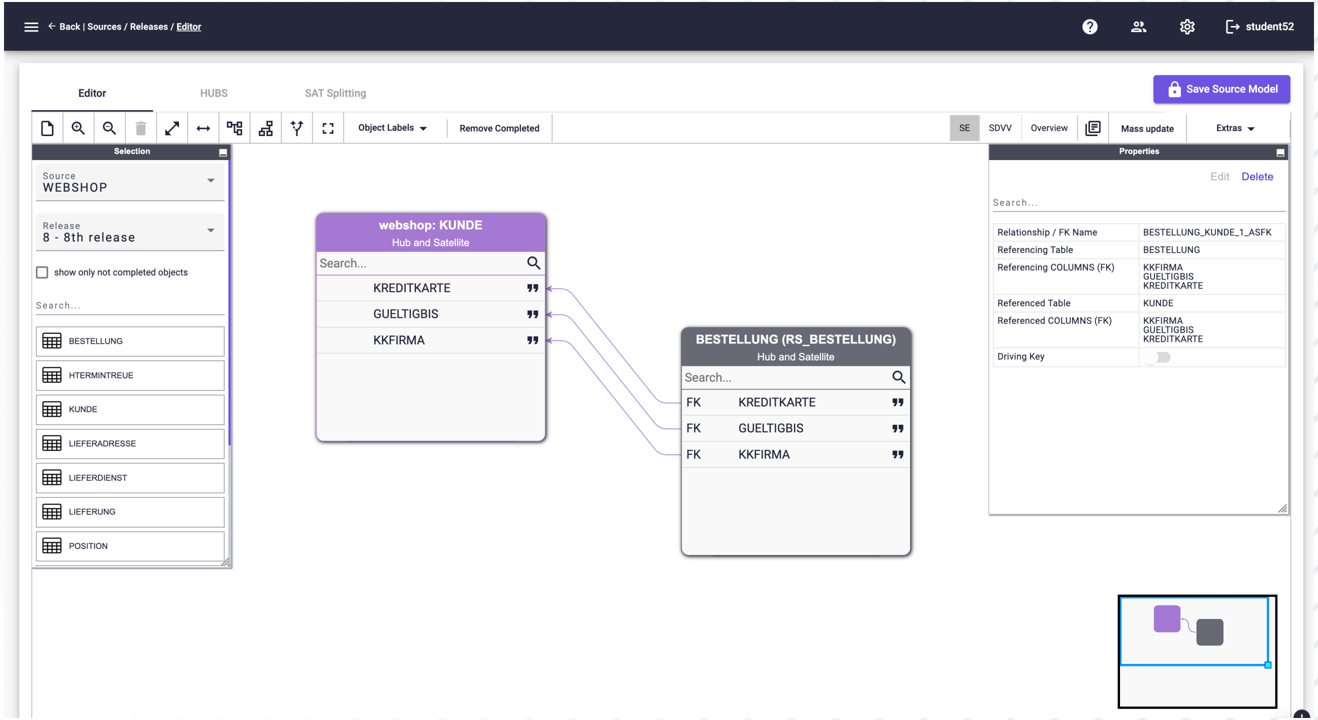


# 15. Business rules

**Identify the customer if the name is not given.**

Capture the relation in a ”link across sources”

These LAS will be deployed in the business vault



Source: WEBSHOP

Release: 8 - 8th release

□ show only not completed objects

- BESTELLUNG
- HTERMINTREUE
- KUNDE
- LIEFERADRESSE
- LIEFERDIENST
- LIEFERUNG
- POSITION

| | |
|---|---|
| Relationship / FK Name | BESTELLUNG_KUNDE_1_ASFK |
| Referencing Table | BESTELLUNG |
| Referencing COLUMNS (FK) | KKFIRMA<br>GUELTIGBIS<br>KREDITKARTE |
| Referenced Table | KUNDE |
| Referenced COLUMNS (FK) | KKFIRMA<br>GUELTIGBIS<br>KREDITKARTE |
| Driving Key | ○ |

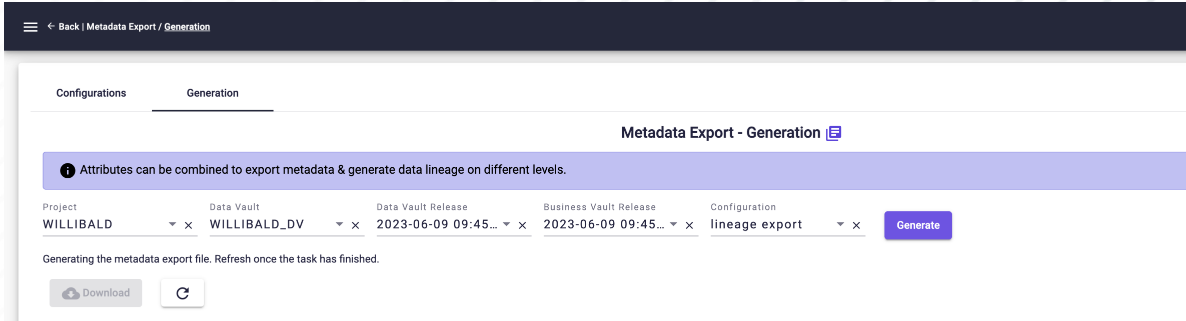
VAULTSPEED

# 16. Data Lineage

**Lineage provided by data VaultSpeed**

Full metadata export available providing input for technical impact analysis.

Can be exported using GUI or REST API



Back | Metadata Export / Generation

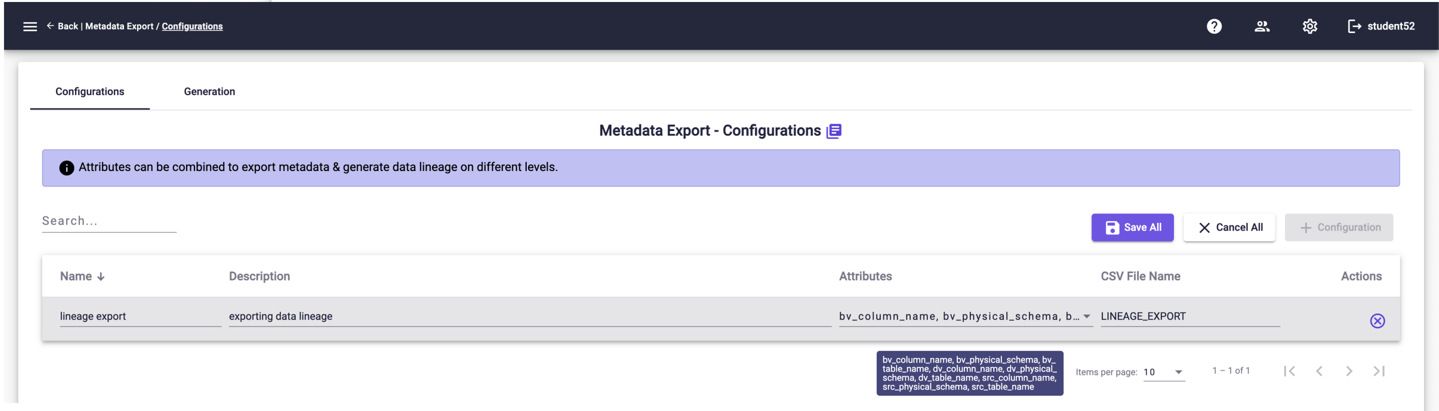
Configurations | Generation

**Metadata Export - Generation**

Attributes can be combined to export metadata & generate data lineage on different levels.

| Project | Data Vault | Data Vault Release | Business Vault Release | Configuration |
|---|---|---|---|---|
| WILLIBALD | WILLIBALD_DV | 2023-06-09 09:45… | 2023-06-09 09:45… | lineage export |

Generating the metadata export file. Refresh once the task has finished.



Back | Metadata Export / Configurations

Configurations | Generation

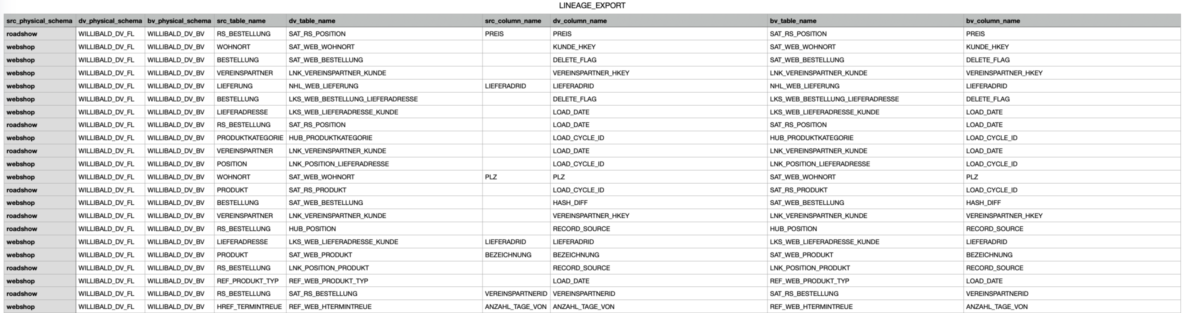
**Metadata Export - Configurations**

Attributes can be combined to export metadata & generate data lineage on different levels.

| Name ↓ | Description | Attributes | CSV File Name | Actions |
|---|---|---|---|---|
| lineage export | exporting data lineage | bv_column_name, bv_physical_schema, b… | LINEAGE_EXPORT | ⊗ |

bv_column_name, bv_physical_schema, bv_table_name, dv_column_name, dv_physical_schema, dv_table_name, src_column_name, src_physical_schema, src_table_name

Items per page: 10   1 – 1 of 1



LINEAGE_EXPORT

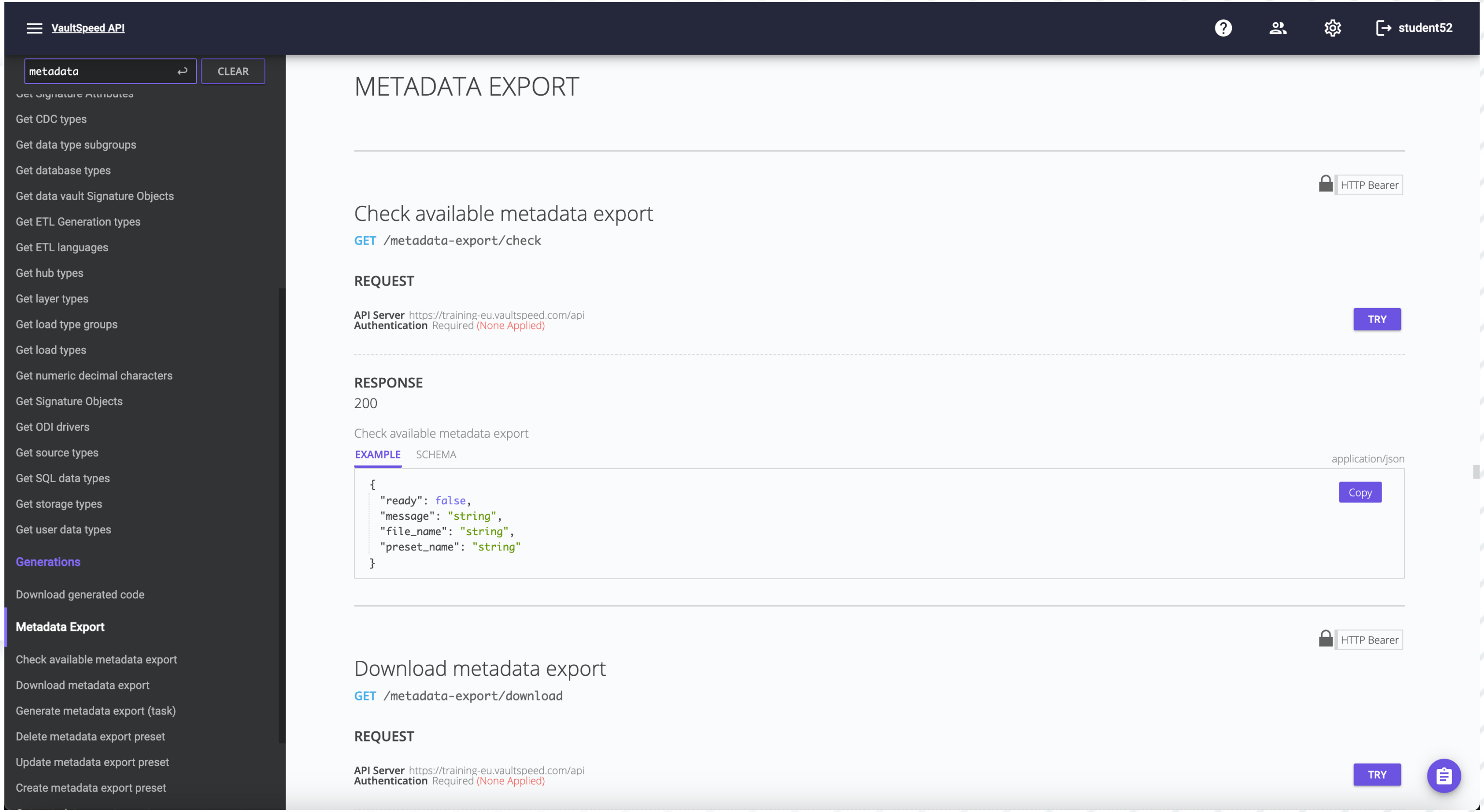
| src_physical_schema | dv_physical_schema | bv_physical_schema | src_table_name | dv_table_name | src_column_name | dv_column_name | bv_table_name | bv_column_name |
|---|---|---|---|---|---|---|---|---|
| roadshow | WILLIBALD_DV_FL | WILLIBALD_DV_BV | RS_BESTELLUNG | SAT_RS_POSITION | PREIS | PREIS | SAT_RS_POSITION | PREIS |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | WOHNORT | SAT_WEB_WOHNORT | | KUNDE_HKEY | SAT_WEB_WOHNORT | KUNDE_HKEY |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | BESTELLUNG | SAT_WEB_BESTELLUNG | | DELETE_FLAG | SAT_WEB_BESTELLUNG | DELETE_FLAG |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | VEREINSPARTNER | LNK_VEREINSPARTNER_KUNDE | | VEREINSPARTNER_HKEY | LNK_VEREINSPARTNER_KUNDE | VEREINSPARTNER_HKEY |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | LIEFERUNG | NHL_WEB_LIEFERUNG | LIEFERADRID | LIEFERADRID | NHL_WEB_LIEFERUNG | LIEFERADRID |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | BESTELLUNG | LKS_WEB_BESTELLUNG_LIEFERADRESSE | | DELETE_FLAG | LKS_WEB_BESTELLUNG_LIEFERADRESSE | DELETE_FLAG |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | LIEFERADRESSE | LKS_WEB_LIEFERADRESSE_KUNDE | | LOAD_DATE | LKS_WEB_LIEFERADRESSE_KUNDE | LOAD_DATE |
| roadshow | WILLIBALD_DV_FL | WILLIBALD_DV_BV | RS_BESTELLUNG | SAT_RS_POSITION | | LOAD_DATE | SAT_RS_POSITION | LOAD_DATE |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | PRODUKTKATEGORIE | HUB_PRODUKTKATEGORIE | | LOAD_CYCLE_ID | HUB_PRODUKTKATEGORIE | LOAD_CYCLE_ID |
| roadshow | WILLIBALD_DV_FL | WILLIBALD_DV_BV | VEREINSPARTNER | LNK_VEREINSPARTNER_KUNDE | | LOAD_DATE | LNK_VEREINSPARTNER_KUNDE | LOAD_DATE |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | POSITION | LNK_POSITION_LIEFERADRESSE | | LOAD_CYCLE_ID | LNK_POSITION_LIEFERADRESSE | LOAD_CYCLE_ID |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | WOHNORT | SAT_WEB_WOHNORT | PLZ | PLZ | SAT_WEB_WOHNORT | PLZ |
| roadshow | WILLIBALD_DV_FL | WILLIBALD_DV_BV | PRODUKT | SAT_RS_PRODUKT | | LOAD_CYCLE_ID | SAT_RS_PRODUKT | LOAD_CYCLE_ID |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | BESTELLUNG | SAT_WEB_BESTELLUNG | | HASH_DIFF | SAT_WEB_BESTELLUNG | HASH_DIFF |
| roadshow | WILLIBALD_DV_FL | WILLIBALD_DV_BV | VEREINSPARTNER | LNK_VEREINSPARTNER_KUNDE | | VEREINSPARTNER_HKEY | LNK_VEREINSPARTNER_KUNDE | VEREINSPARTNER_HKEY |
| roadshow | WILLIBALD_DV_FL | WILLIBALD_DV_BV | RS_BESTELLUNG | HUB_POSITION | | RECORD_SOURCE | HUB_POSITION | RECORD_SOURCE |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | LIEFERADRESSE | LKS_WEB_LIEFERADRESSE_KUNDE | LIEFERADRID | LIEFERADRID | LKS_WEB_LIEFERADRESSE_KUNDE | LIEFERADRID |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | PRODUKT | SAT_WEB_PRODUKT | BEZEICHNUNG | BEZEICHNUNG | SAT_WEB_PRODUKT | BEZEICHNUNG |
| roadshow | WILLIBALD_DV_FL | WILLIBALD_DV_BV | RS_BESTELLUNG | LNK_POSITION_PRODUKT | | RECORD_SOURCE | LNK_POSITION_PRODUKT | RECORD_SOURCE |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | REF_PRODUKT_TYP | REF_WEB_PRODUKT_TYP | | LOAD_DATE | REF_WEB_PRODUKT_TYP | LOAD_DATE |
| roadshow | WILLIBALD_DV_FL | WILLIBALD_DV_BV | RS_BESTELLUNG | SAT_RS_BESTELLUNG | VEREINSPARTNERID | VEREINSPARTNERID | SAT_RS_BESTELLUNG | VEREINSPARTNERID |
| webshop | WILLIBALD_DV_FL | WILLIBALD_DV_BV | HREF_TERMINTREUE | REF_WEB_HTERMINTREUE | ANZAHL_TAGE_VON | ANZAHL_TAGE_VON | REF_WEB_HTERMINTREUE | ANZAHL_TAGE_VON |

# 16. Data Lineage

**Lineage provided by data VaultSpeed**

Full metadata export available providing input for technical impact analysis.

Can be exported using GUI or REST API



## METADATA EXPORT

HTTP Bearer

### Check available metadata export

GET /metadata-export/check

**REQUEST**

**API Server** https://training-eu.vaultspeed.com/api
**Authentication** Required (None Applied)

TRY

**RESPONSE**

200

Check available metadata export

EXAMPLE SCHEMA application/json

```
{
  "ready": false,
  "message": "string",
  "file_name": "string",
  "preset_name": "string"
}
```

Copy

HTTP Bearer

### Download metadata export

GET /metadata-export/download

**REQUEST**

**API Server** https://training-eu.vaultspeed.com/api
**Authentication** Required (None Applied)

TRY

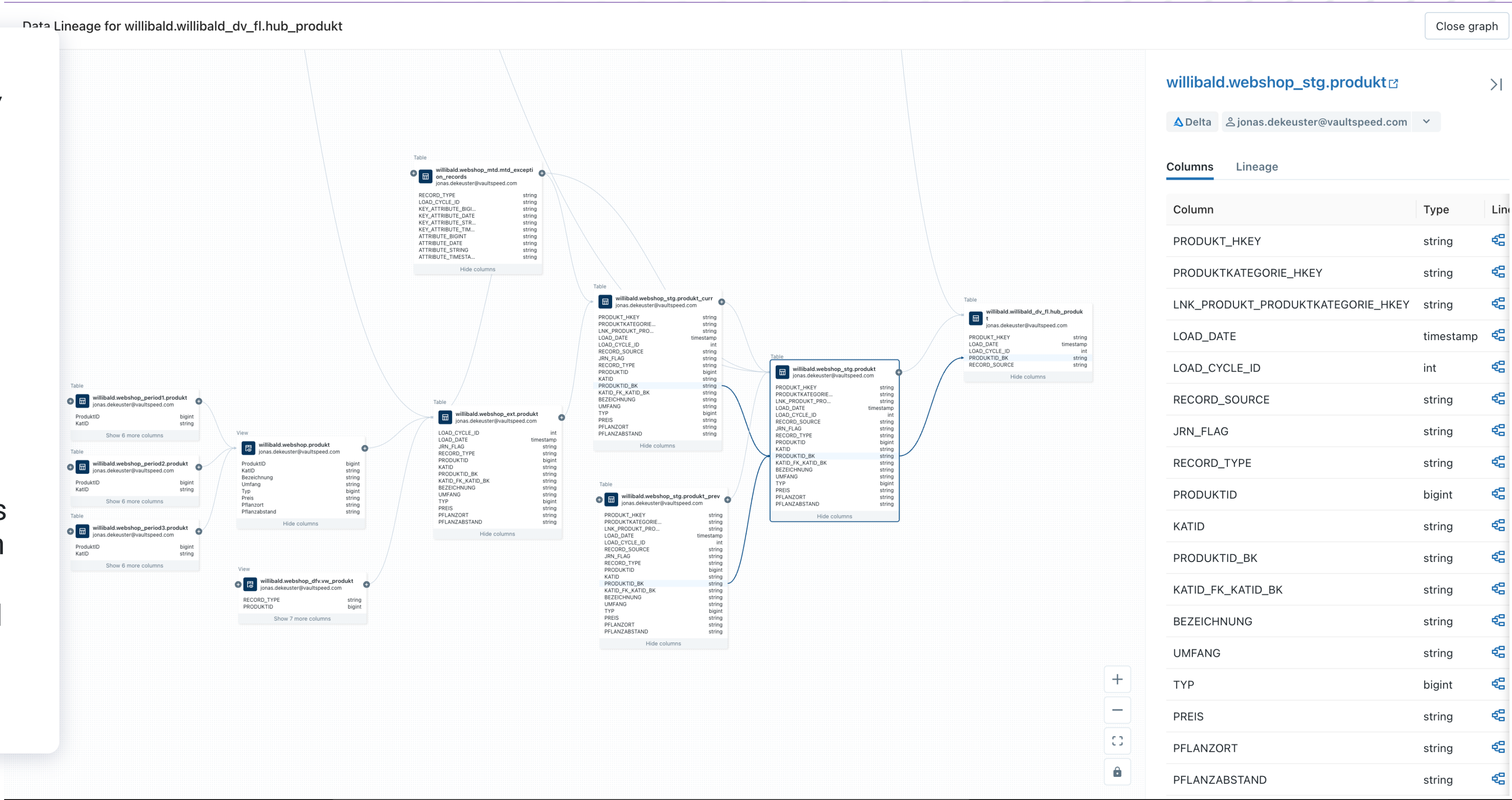
VAULTSPEED

# 16. Data Lineage

## Lineage provided by data platform

More and more modern data platforms introduce this in their standard offering.

Databricks Unity Catalog provides end2end column level lineage for end users based on actual data flows of the last 30 days.



Data Lineage for willibald.willibald_dv_fl.hub_produkt

Table
**willibald.webshop_period1.produkt**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| ProduktID | bigint |
| KatID | string |

Table
**willibald.webshop_period2.produkt**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| ProduktID | bigint |
| KatID | string |

Table
**willibald.webshop_period3.produkt**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| ProduktID | bigint |
| KatID | string |

View
**willibald.webshop.produkt**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| ProduktID | bigint |
| KatID | string |
| Bezeichnung | string |
| Umfang | string |
| Typ | bigint |
| Preis | string |
| Pflanzort | string |
| Pflanzabstand | string |

View
**willibald.webshop_dfv.vw_produkt**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| RECORD_TYPE | string |
| PRODUKTID | bigint |

Table
**willibald.webshop_mtd.mtd_exception_records**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| RECORD_TYPE | string |
| LOAD_CYCLE_ID | string |
| KEY_ATTRIBUTE_BIGI… | string |
| KEY_ATTRIBUTE_DATE | string |
| KEY_ATTRIBUTE_STR… | string |
| KEY_ATTRIBUTE_TIM… | string |
| ATTRIBUTE_BIGINT | string |
| ATTRIBUTE_DATE | string |
| ATTRIBUTE_STRING | string |
| ATTRIBUTE_TIMESTA… | string |

Table
**willibald.webshop_ext.produkt**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| LOAD_CYCLE_ID | int |
| LOAD_DATE | timestamp |
| JRN_FLAG | string |
| RECORD_TYPE | string |
| PRODUKTID | bigint |
| KATID | string |
| PRODUKTID_BK | string |
| KATID_FK_KATID_BK | string |
| BEZEICHNUNG | string |
| UMFANG | string |
| TYP | bigint |
| PREIS | string |
| PFLANZORT | string |
| PFLANZABSTAND | string |

Table
**willibald.webshop_stg.produkt_curr**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| PRODUKT_HKEY | string |
| PRODUKTKATEGORIE… | string |
| LNK_PRODUKT_PRO… | string |
| LOAD_DATE | timestamp |
| LOAD_CYCLE_ID | int |
| RECORD_SOURCE | string |
| JRN_FLAG | string |
| RECORD_TYPE | string |
| PRODUKTID | bigint |
| KATID | string |
| PRODUKTID_BK | string |
| KATID_FK_KATID_BK | string |
| BEZEICHNUNG | string |
| UMFANG | string |
| TYP | bigint |
| PREIS | string |
| PFLANZORT | string |
| PFLANZABSTAND | string |

Table
**willibald.webshop_stg.produkt_prev**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| PRODUKT_HKEY | string |
| PRODUKTKATEGORIE… | string |
| LNK_PRODUKT_PRO… | string |
| LOAD_DATE | timestamp |
| LOAD_CYCLE_ID | int |
| RECORD_SOURCE | string |
| JRN_FLAG | string |
| RECORD_TYPE | string |
| PRODUKTID | bigint |
| KATID | string |
| PRODUKTID_BK | string |
| KATID_FK_KATID_BK | string |
| BEZEICHNUNG | string |
| UMFANG | string |
| TYP | bigint |
| PREIS | string |
| PFLANZORT | string |
| PFLANZABSTAND | string |

Table
**willibald.webshop_stg.produkt**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| PRODUKT_HKEY | string |
| PRODUKTKATEGORIE… | string |
| LNK_PRODUKT_PRO… | string |
| LOAD_DATE | timestamp |
| LOAD_CYCLE_ID | int |
| RECORD_SOURCE | string |
| JRN_FLAG | string |
| RECORD_TYPE | string |
| PRODUKTID | bigint |
| KATID | string |
| PRODUKTID_BK | string |
| KATID_FK_KATID_BK | string |
| BEZEICHNUNG | string |
| UMFANG | string |
| TYP | bigint |
| PREIS | string |
| PFLANZORT | string |
| PFLANZABSTAND | string |

Table
**willibald.willibald_dv_fl.hub_produkt**
jonas.dekeuster@vaultspeed.com

| | |
|---|---|
| PRODUKT_HKEY | string |
| LOAD_DATE | timestamp |
| LOAD_CYCLE_ID | int |
| PRODUKTID_BK | string |
| RECORD_SOURCE | string |

**willibald.webshop_stg.produkt**

Delta
jonas.dekeuster@vaultspeed.com

| Column | Type | Lin |
|---|---|---|
| PRODUKT_HKEY | string | |
| PRODUKTKATEGORIE_HKEY | string | |
| LNK_PRODUKT_PRODUKTKATEGORIE_HKEY | string | |
| LOAD_DATE | timestamp | |
| LOAD_CYCLE_ID | int | |
| RECORD_SOURCE | string | |
| JRN_FLAG | string | |
| RECORD_TYPE | string | |
| PRODUKTID | bigint | |
| KATID | string | |
| PRODUKTID_BK | string | |
| KATID_FK_KATID_BK | string | |
| BEZEICHNUNG | string | |
| UMFANG | string | |
| TYP | bigint | |
| PREIS | string | |
| PFLANZORT | string | |
| PFLANZABSTAND | string | |

VAULTSPEED

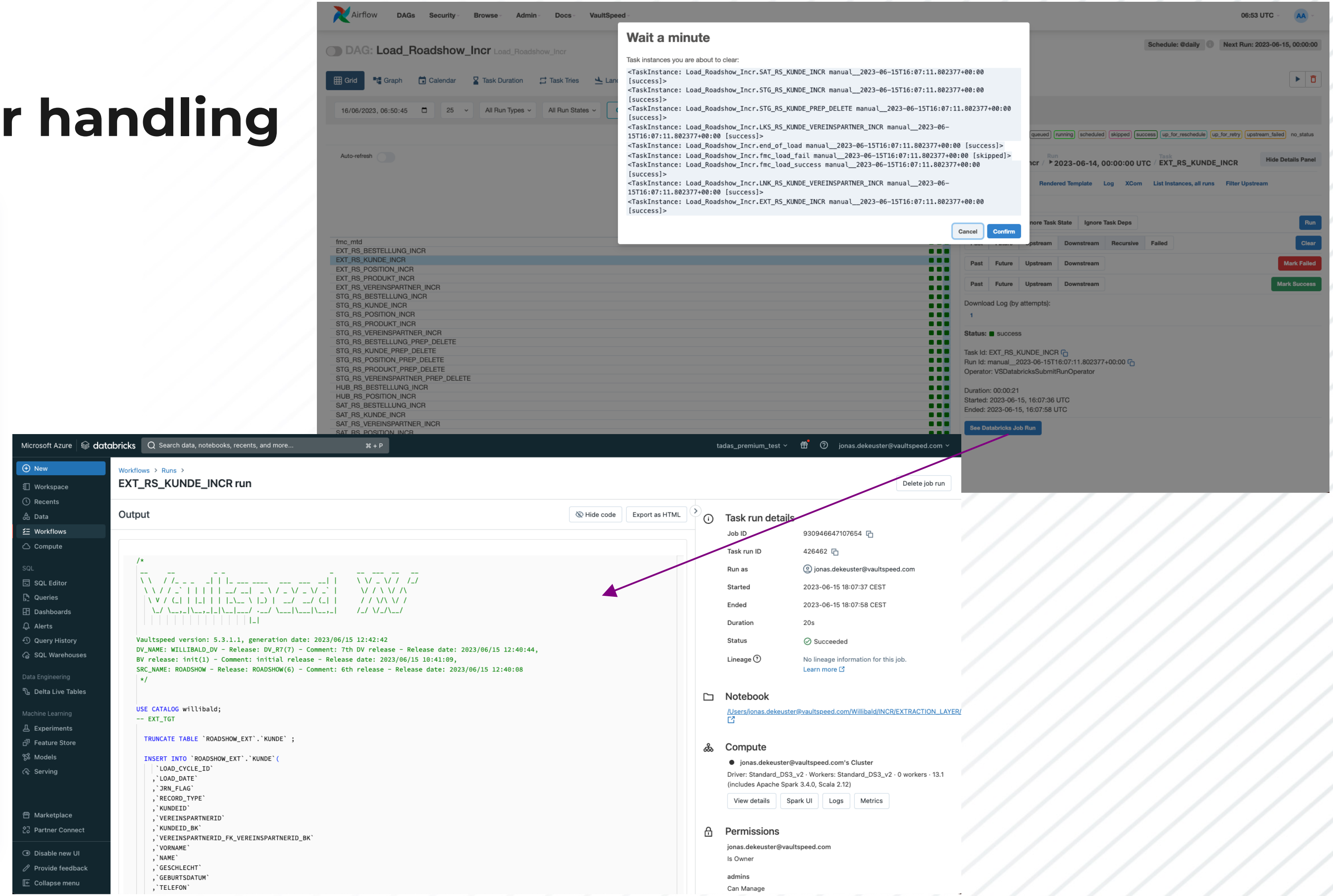
# 17. Error handling

**All workflows generated by VaultSpeed are fully restartable**

No loss of data or inconsistencies will arise.

Workflows can also partially be restarted.

Link to job log details in target platform



DAG: Load_Roadshow_Incr Load_Roadshow_Incr

Schedule: @daily

Next Run: 2023-06-15, 00:00:00

16/06/2023, 06:50:45 · 25 · All Run Types · All Run States

fmc_mtd
EXT_RS_BESTELLUNG_INCR
EXT_RS_KUNDE_INCR
EXT_RS_POSITION_INCR
EXT_RS_PRODUKT_INCR
EXT_RS_VEREINSPARTNER_INCR
STG_RS_BESTELLUNG_INCR
STG_RS_KUNDE_INCR
STG_RS_POSITION_INCR
STG_RS_PRODUKT_INCR
STG_RS_VEREINSPARTNER_INCR
STG_RS_BESTELLUNG_PREP_DELETE
STG_RS_KUNDE_PREP_DELETE
STG_RS_POSITION_PREP_DELETE
STG_RS_PRODUKT_PREP_DELETE
STG_RS_VEREINSPARTNER_PREP_DELETE
HUB_RS_BESTELLUNG_INCR
HUB_RS_POSITION_INCR
SAT_RS_BESTELLUNG_INCR
SAT_RS_KUNDE_INCR
SAT_RS_VEREINSPARTNER_INCR
SAT_RS_POSITION_INCR

**Wait a minute**

Task instances you are about to clear:

```
<TaskInstance: Load_Roadshow_Incr.SAT_RS_KUNDE_INCR manual__2023-06-15T16:07:11.802377+00:00
[success]>
<TaskInstance: Load_Roadshow_Incr.STG_RS_KUNDE_INCR manual__2023-06-15T16:07:11.802377+00:00
[success]>
<TaskInstance: Load_Roadshow_Incr.STG_RS_KUNDE_PREP_DELETE manual__2023-06-15T16:07:11.802377+00:00
[success]>
<TaskInstance: Load_Roadshow_Incr.LKS_RS_KUNDE_VEREINSPARTNER_INCR manual__2023-06-
15T16:07:11.802377+00:00 [success]>
<TaskInstance: Load_Roadshow_Incr.end_of_load manual__2023-06-15T16:07:11.802377+00:00 [success]>
<TaskInstance: Load_Roadshow_Incr.fmc_load_fail manual__2023-06-15T16:07:11.802377+00:00 [skipped]>
<TaskInstance: Load_Roadshow_Incr.fmc_load_success manual__2023-06-15T16:07:11.802377+00:00
[success]>
<TaskInstance: Load_Roadshow_Incr.LNK_RS_KUNDE_VEREINSPARTNER_INCR manual__2023-06-
15T16:07:11.802377+00:00 [success]>
<TaskInstance: Load_Roadshow_Incr.EXT_RS_KUNDE_INCR manual__2023-06-15T16:07:11.802377+00:00
[success]>
```

Run ▸2023-06-14, 00:00:00 UTC / Task EXT_RS_KUNDE_INCR

Download Log (by attempts):
1

Status: success

Task Id: EXT_RS_KUNDE_INCR
Run Id: manual__2023-06-15T16:07:11.802377+00:00
Operator: VSDatabricksSubmitRunOperator

Duration: 00:00:21
Started: 2023-06-15, 16:07:36 UTC
Ended: 2023-06-15, 16:07:58 UTC

Workflows › Runs ›

EXT_RS_KUNDE_INCR run

Output

```
/*
 __      __          _ _                       _       __  ___  __   __
 \ \    / /_ _ _   _| | |_ ___ _ __   ___  ___  __| |     \ \/ _ \/ /  /_/
  \ \  / / _` | | | | | __/ __| '_ \ / _ \/ _ \/ _` |      \/ / \ \/ /\
   \ V / (_| | |_| | | |_\__ \ |_) |  __/  __/ (_| |      / / \/\ \/ /
    \_/ \__,_|\__,_|_|\__|___/ .__/ \___|\___|\__,_|     /_/ \/_/\__/
                             |_|
Vaultspeed version: 5.3.1.1, generation date: 2023/06/15 12:42:42
DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7th DV release - Release date: 2023/06/15 12:40:44,
BV release: init(1) - Comment: initial release - Release date: 2023/06/15 10:41:09,
SRC_NAME: ROADSHOW - Release: ROADSHOW(6) - Comment: 6th release - Release date: 2023/06/15 12:40:08
 */


USE CATALOG willibald;
-- EXT_TGT

  TRUNCATE TABLE `ROADSHOW_EXT`.`KUNDE` ;

  INSERT INTO `ROADSHOW_EXT`.`KUNDE`(
     `LOAD_CYCLE_ID`
    ,`LOAD_DATE`
    ,`JRN_FLAG`
    ,`RECORD_TYPE`
    ,`KUNDEID`
    ,`VEREINSPARTNERID`
    ,`KUNDEID_BK`
    ,`VEREINSPARTNERID_FK_VEREINSPARTNERID_BK`
    ,`VORNAME`
    ,`NAME`
    ,`GESCHLECHT`
    ,`GEBURTSDATUM`
    ,`TELEFON`
```

**Task run details**

Job ID 930946647107654
Task run ID 426462
Run as jonas.dekeuster@vaultspeed.com
Started 2023-06-15 18:07:37 CEST
Ended 2023-06-15 18:07:58 CEST
Duration 20s
Status ✓ Succeeded
Lineage No lineage information for this job. Learn more

**Notebook**

/Users/jonas.dekeuster@vaultspeed.com/Willibald/INCR/EXTRACTION_LAYER/

**Compute**

● jonas.dekeuster@vaultspeed.com's Cluster
Driver: Standard_DS3_v2 · Workers: Standard_DS3_v2 · 0 workers · 13.1 (includes Apache Spark 3.4.0, Scala 2.12)

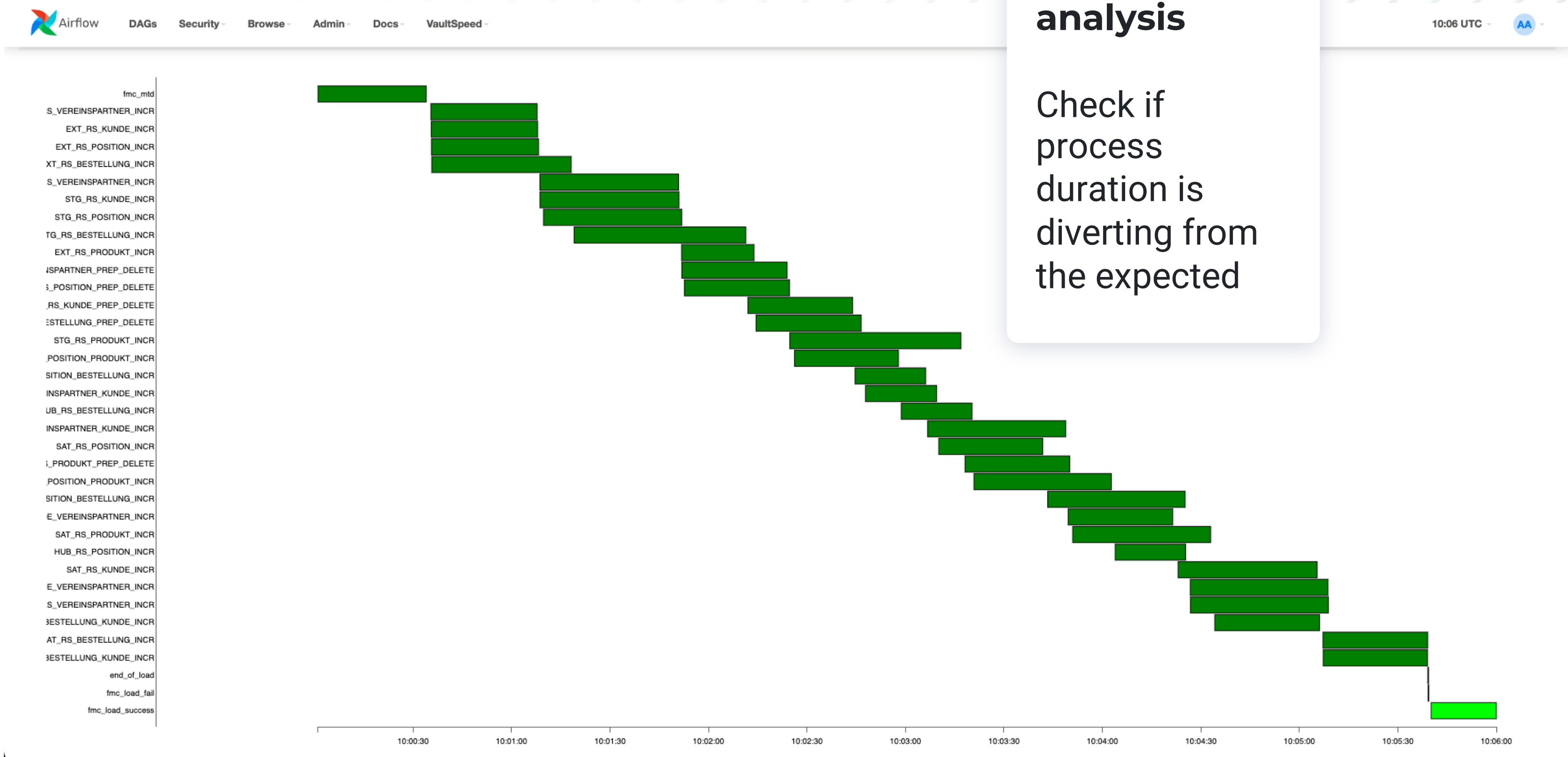
**Permissions**

jonas.dekeuster@vaultspeed.com
Is Owner

admins
Can Manage

VAULTSPEED

# 17. Error handling



## Load duration analysis

Check if process duration is diverting from the expected

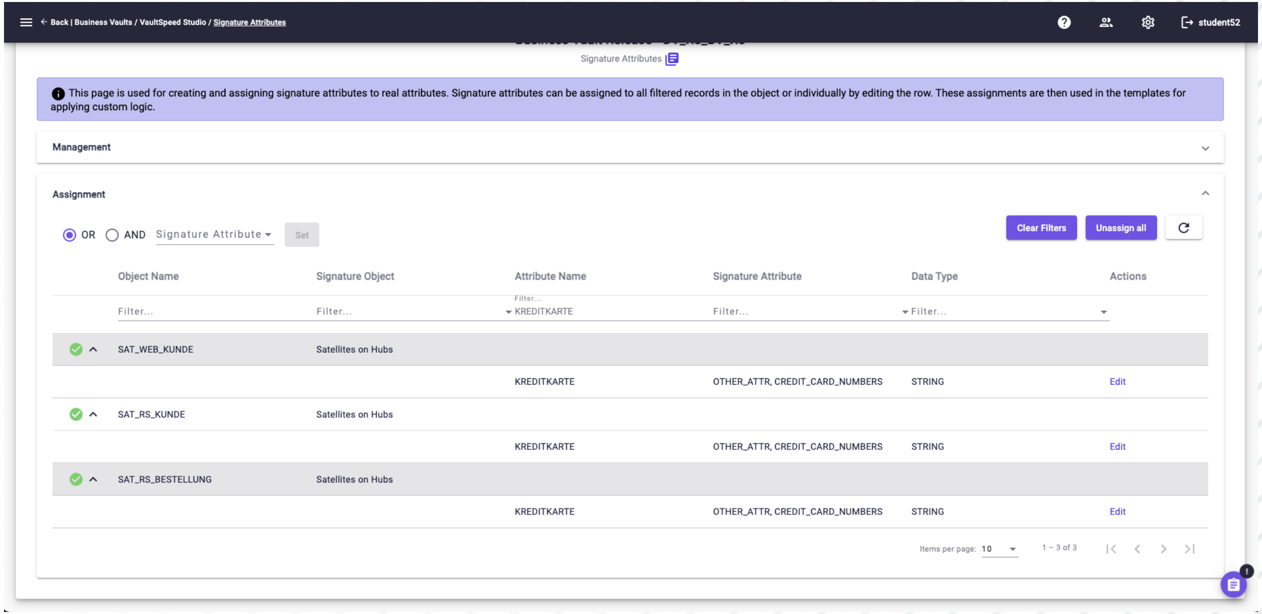
# 17. DQ check

## Build data quality templates with VaultSpeed studio

Example: build a template that checks if a certain field contains valid credit card data

Databricks has a function to do ah Luhn check on Credit Card no.

The results are alarming for our roadshow ;-)



Back | Business Vaults / VaultSpeed Studio / Signature Attributes
student52
Signature Attributes

This page is used for creating and assigning signature attributes to real attributes. Signature attributes can be assigned to all filtered records in the object or individually by editing the row. These assignments are then used in the templates for applying custom logic.

Management
Assignment
● OR ○ AND Signature Attribute
Set
Clear Filters
Unassign all

| | Object Name | Signature Object | Attribute Name | Signature Attribute | Data Type | Actions |
|---|---|---|---|---|---|---|
| | Filter... | Filter... | KREDITKARTE | Filter... | Filter... | |
| ✓ | SAT_WEB_KUNDE | Satellites on Hubs | | | | |
| | | | KREDITKARTE | OTHER_ATTR, CREDIT_CARD_NUMBERS | STRING | Edit |
| ✓ | SAT_RS_KUNDE | Satellites on Hubs | | | | |
| | | | KREDITKARTE | OTHER_ATTR, CREDIT_CARD_NUMBERS | STRING | Edit |
| ✓ | SAT_RS_BESTELLUNG | Satellites on Hubs | | | | |
| | | | KREDITKARTE | OTHER_ATTR, CREDIT_CARD_NUMBERS | STRING | Edit |

Easy to build in VaultSpeed Studio and to apply on all attributes across my RDV that contain credit card numbers.

Start by defining a business signature “CREDIT_CARD_NUMBERS” and applying that to the RDV fields we want to test.

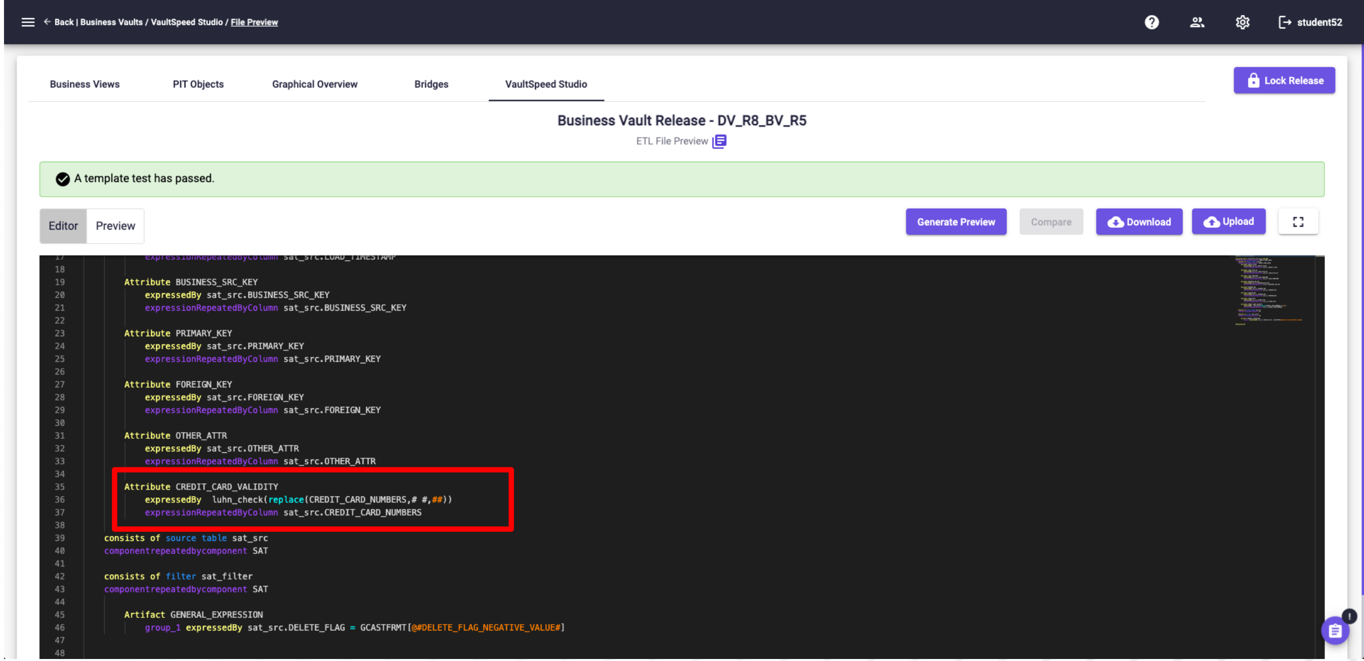
# 17. DQ check

## Build data quality templates with VaultSpeed studio

Example: build a template that checks if a certain field contains valid credit card data

Databricks has a function to do ah Luhn check on Credit Card no.

The results are alarming for our roadshow ;-)



Business Vault Release - DV_R8_BV_R5

A template test has passed.

```
            expressionRepeatedByColumn sat_src.LOAD_TIMESTAMP

        Attribute BUSINESS_SRC_KEY
            expressedBy sat_src.BUSINESS_SRC_KEY
            expressionRepeatedByColumn sat_src.BUSINESS_SRC_KEY

        Attribute PRIMARY_KEY
            expressedBy sat_src.PRIMARY_KEY
            expressionRepeatedByColumn sat_src.PRIMARY_KEY

        Attribute FOREIGN_KEY
            expressedBy sat_src.FOREIGN_KEY
            expressionRepeatedByColumn sat_src.FOREIGN_KEY

        Attribute OTHER_ATTR
            expressedBy sat_src.OTHER_ATTR
            expressionRepeatedByColumn sat_src.OTHER_ATTR

        Attribute CREDIT_CARD_VALIDITY
            expressedBy  luhn_check(replace(CREDIT_CARD_NUMBERS,# #,##))
            expressionRepeatedByColumn sat_src.CREDIT_CARD_NUMBERS

    consists of source table sat_src
    componentrepeatedbycomponent SAT

    consists of filter sat_filter
    componentrepeatedbycomponent SAT

        Artifact GENERAL_EXPRESSION
            group_1 expressedBy sat_src.DELETE_FLAG = GCASTFRMT[@#DELETE_FLAG_NEGATIVE_VALUE#]
```

Build a template using the signature we just defined and tagged. This will generate a DQ view in our designated schema (we chose to deploy it along with the rest in the information_marts schema, you could also define your own DQ shema…

VAULTSPEED

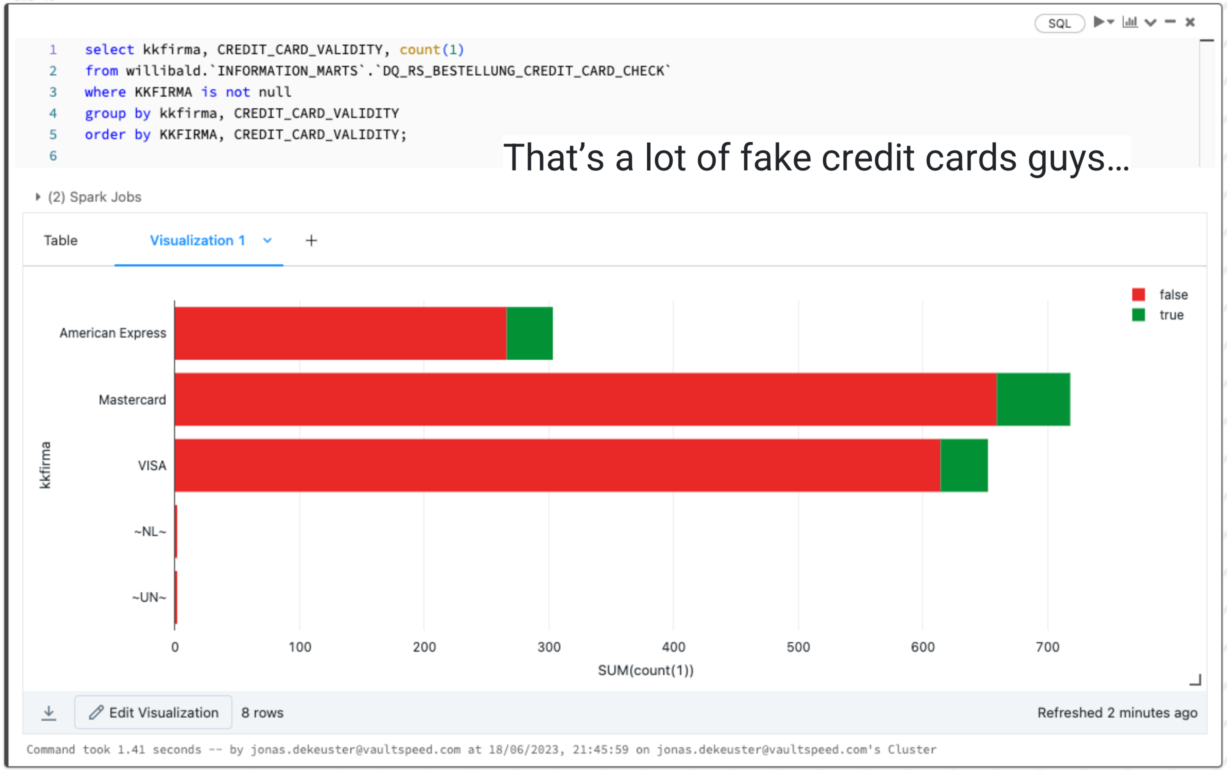
# 17. DQ check

## Build data quality templates with VaultSpeed studio

Example: build a template that checks if a certain field contains valid credit card data

Databricks has a function to do a Luhn check on Credit Card number strings.

The results are alarming for our roadshow ;-)



```sql
select kkfirma, CREDIT_CARD_VALIDITY, count(1)
from willibald.`INFORMATION_MARTS`.`DQ_RS_BESTELLUNG_CREDIT_CARD_CHECK`
where KKFIRMA is not null
group by kkfirma, CREDIT_CARD_VALIDITY
order by KKFIRMA, CREDIT_CARD_VALIDITY;
```

That’s a lot of fake credit cards guys…

```
Command took 1.41 seconds -- by jonas.dekeuster@vaultspeed.com at 18/06/2023, 21:45:59 on jonas.dekeuster@vaultspeed.com's Cluster
```

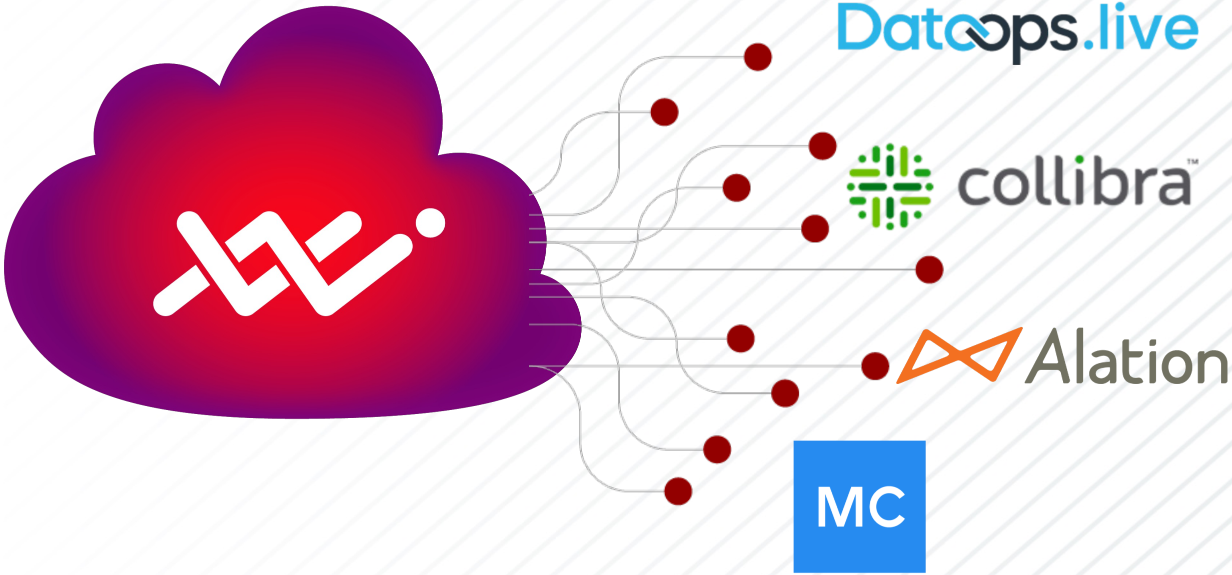
VAULTSPEED

# 17. DQ check

## Integrate with best of breed DQ solution providers

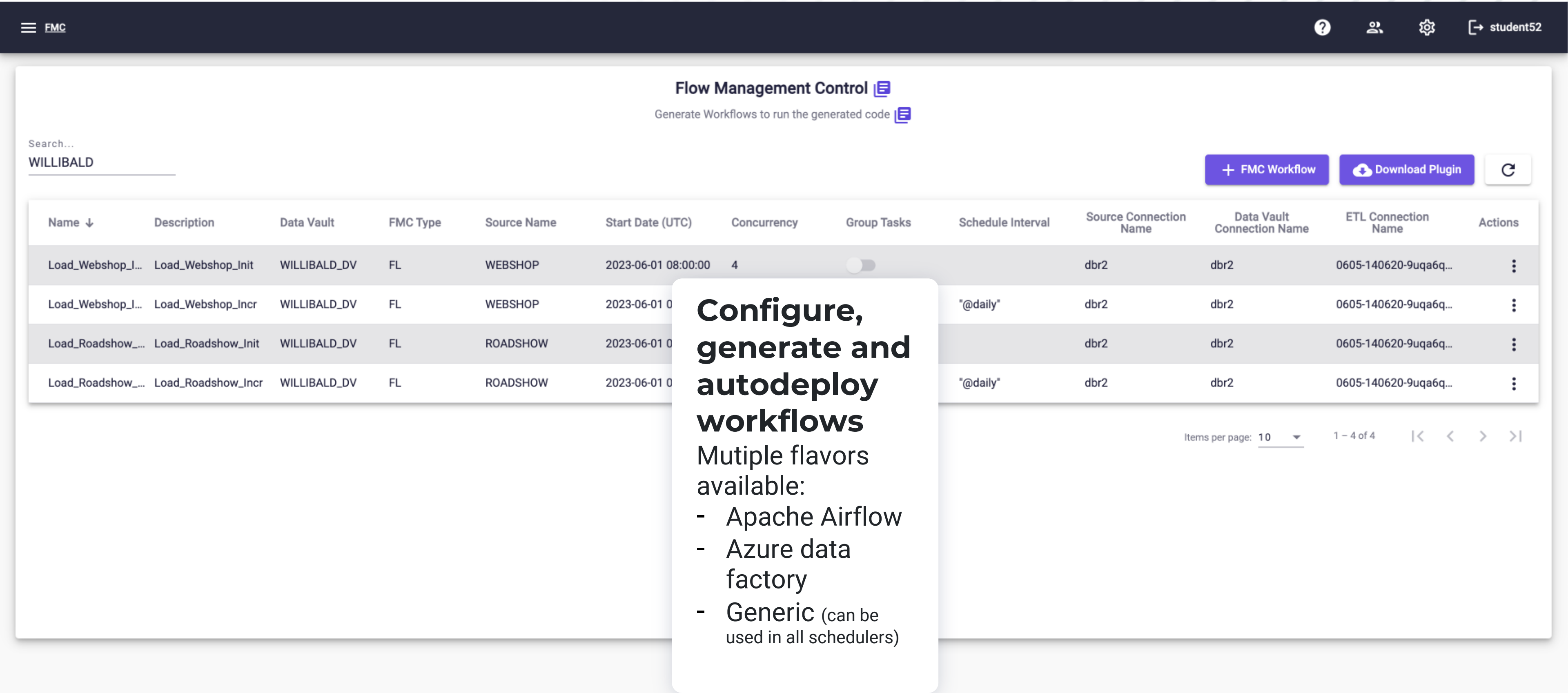
VaultSpeed fully unlocks its metadata trough its REST API

You can build REST API integrations with tools like Collibra, DataOps.live, Monte Carlo,…



VAULTSPEED

# 18. Orchestration



## Flow Management Control

Generate Workflows to run the generated code

Search...
WILLIBALD

| Name ↓ | Description | Data Vault | FMC Type | Source Name | Start Date (UTC) | Concurrency | Group Tasks | Schedule Interval | Source Connection Name | Data Vault Connection Name | ETL Connection Name | Actions |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Load_Webshop_I... | Load_Webshop_Init | WILLIBALD_DV | FL | WEBSHOP | 2023-06-01 08:00:00 | 4 | ○ | | dbr2 | dbr2 | 0605-140620-9uqa6q... | ⋮ |
| Load_Webshop_I... | Load_Webshop_Incr | WILLIBALD_DV | FL | WEBSHOP | 2023-06-01 0 | | | "@daily" | dbr2 | dbr2 | 0605-140620-9uqa6q... | ⋮ |
| Load_Roadshow_... | Load_Roadshow_Init | WILLIBALD_DV | FL | ROADSHOW | 2023-06-01 0 | | | | dbr2 | dbr2 | 0605-140620-9uqa6q... | ⋮ |
| Load_Roadshow_... | Load_Roadshow_Incr | WILLIBALD_DV | FL | ROADSHOW | 2023-06-01 0 | | | "@daily" | dbr2 | dbr2 | 0605-140620-9uqa6q... | ⋮ |

Items per page: 10    1 – 4 of 4

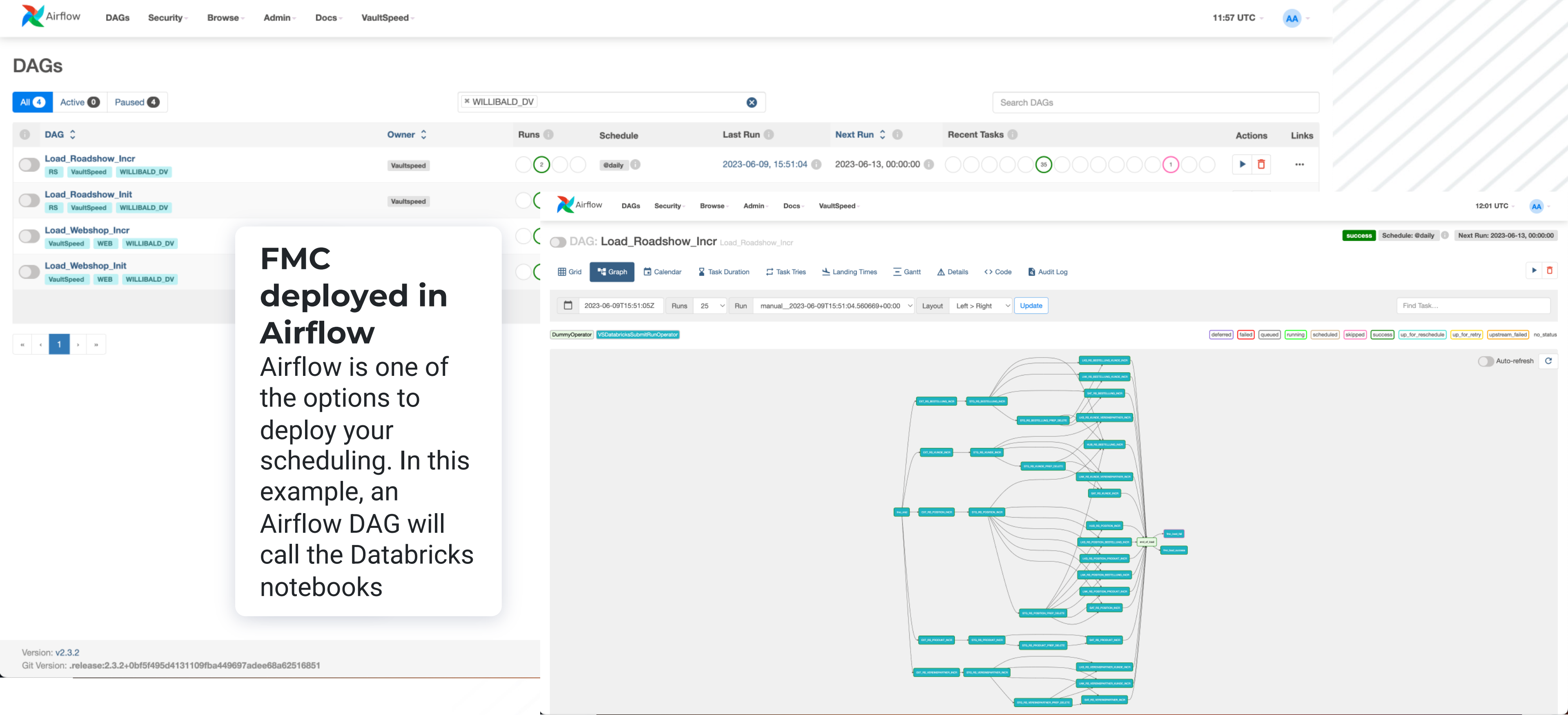
**Configure, generate and autodeploy workflows**

Mutiple flavors available:

- Apache Airflow
- Azure data factory
- Generic (can be used in all schedulers)

VAULTSPEED

# 18. Orchestration



Airflow DAGs Security Browse Admin Docs VaultSpeed 11:57 UTC AA

## DAGs

All 4 | Active 0 | Paused 4

× WILLIBALD_DV

Search DAGs

| | DAG | Owner | Runs | Schedule | Last Run | Next Run | Recent Tasks | Actions | Links |
|---|---|---|---|---|---|---|---|---|---|
| | Load_Roadshow_Incr<br>RS VaultSpeed WILLIBALD_DV | Vaultspeed | 2 | @daily | 2023-06-09, 15:51:04 | 2023-06-13, 00:00:00 | 35 1 | ▶ 🗑 | ... |
| | Load_Roadshow_Init<br>RS VaultSpeed WILLIBALD_DV | Vaultspeed | | | | | | | |
| | Load_Webshop_Incr<br>VaultSpeed WEB WILLIBALD_DV | | | | | | | | |
| | Load_Webshop_Init<br>VaultSpeed WEB WILLIBALD_DV | | | | | | | | |

« ‹ 1 › »

Version: v2.3.2
Git Version: .release:2.3.2+0bf5f495d4131109fba449697adee68a62516851

Airflow DAGs Security Browse Admin Docs VaultSpeed 12:01 UTC AA

DAG: Load_Roadshow_Incr Load_Roadshow_Incr

success Schedule: @daily Next Run: 2023-06-13, 00:00:00

Grid | Graph | Calendar | Task Duration | Task Tries | Landing Times | Gantt | Details | Code | Audit Log

2023-06-09T15:51:05Z Runs 25 Run manual__2023-06-09T15:51:04.560669+00:00 Layout Left > Right Update

Find Task...

DummyOperator VSDatabricksSubmitRunOperator

deferred failed queued running scheduled skipped success up_for_reschedule up_for_retry upstream_failed no_status

Auto-refresh

## FMC deployed in Airflow

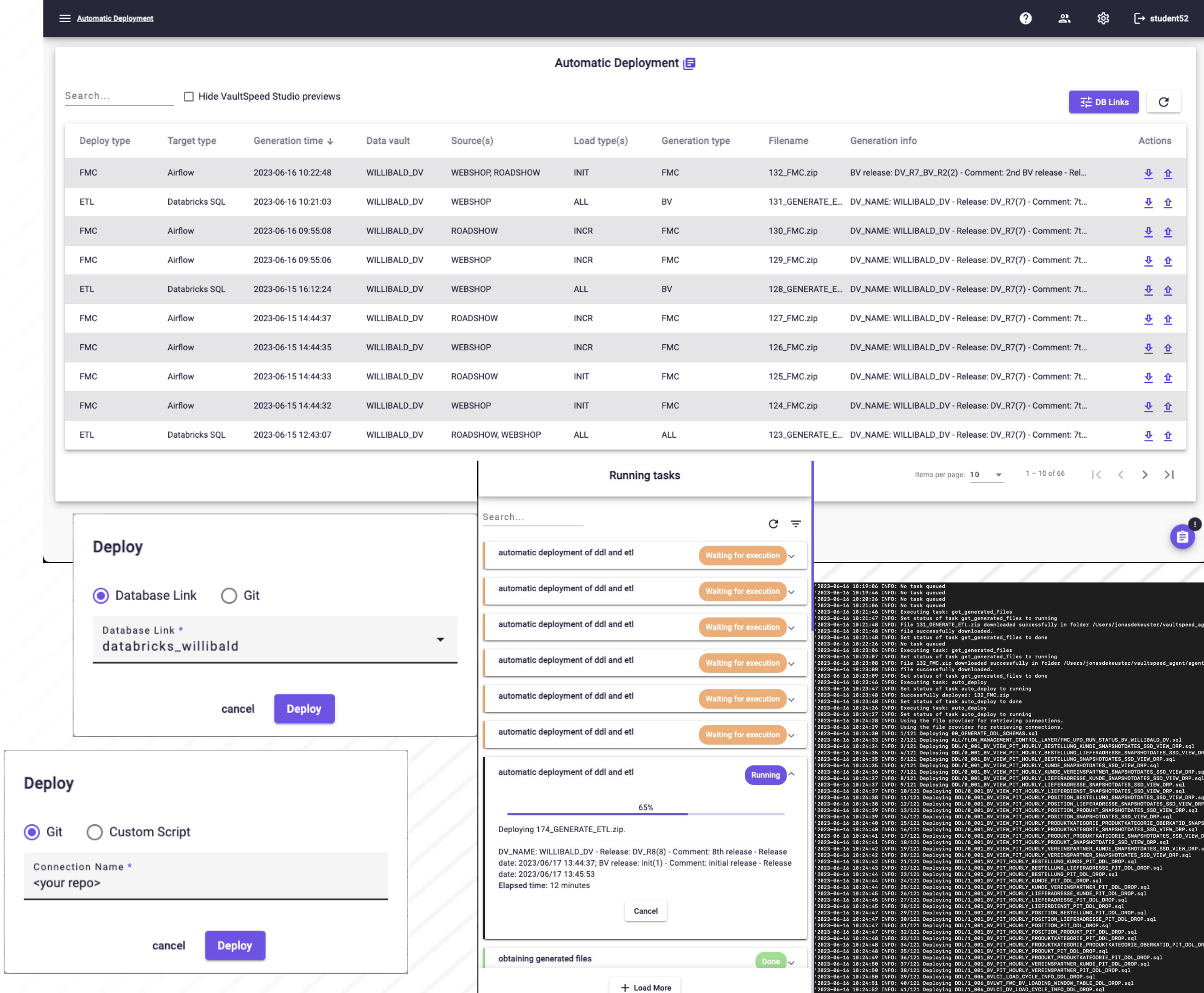
Airflow is one of the options to deploy your scheduling. In this example, an Airflow DAG will call the Databricks notebooks

VAULTSPEED

# 19. Deployment

## Deployment is done automatically

- Directly on the target platform into your data runtime environment (typically on development)
- Into your designated git-repo (typically to feed it into CI/CD processes)
- Using a custom deploy script (to tailor the deployment entirely to your needs)



Automatic Deployment

□ Hide VaultSpeed Studio previews

| Deploy type | Target type | Generation time ↓ | Data vault | Source(s) | Load type(s) | Generation type | Filename | Generation info | Actions |
|---|---|---|---|---|---|---|---|---|---|
| FMC | Airflow | 2023-06-16 10:22:48 | WILLIBALD_DV | WEBSHOP, ROADSHOW | INIT | FMC | 132_FMC.zip | BV release: DV_R7_BV_R2(2) - Comment: 2nd BV release - Rel... | |
| ETL | Databricks SQL | 2023-06-16 10:21:03 | WILLIBALD_DV | WEBSHOP | ALL | BV | 131_GENERATE_E... | DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7t... | |
| FMC | Airflow | 2023-06-16 09:55:08 | WILLIBALD_DV | ROADSHOW | INCR | FMC | 130_FMC.zip | DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7t... | |
| FMC | Airflow | 2023-06-16 09:55:06 | WILLIBALD_DV | WEBSHOP | INCR | FMC | 129_FMC.zip | DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7t... | |
| ETL | Databricks SQL | 2023-06-15 16:12:24 | WILLIBALD_DV | WEBSHOP | ALL | BV | 128_GENERATE_E... | DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7t... | |
| FMC | Airflow | 2023-06-15 14:44:37 | WILLIBALD_DV | ROADSHOW | INCR | FMC | 127_FMC.zip | DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7t... | |
| FMC | Airflow | 2023-06-15 14:44:35 | WILLIBALD_DV | WEBSHOP | INCR | FMC | 126_FMC.zip | DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7t... | |
| FMC | Airflow | 2023-06-15 14:44:33 | WILLIBALD_DV | ROADSHOW | INIT | FMC | 125_FMC.zip | DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7t... | |
| FMC | Airflow | 2023-06-15 14:44:32 | WILLIBALD_DV | WEBSHOP | INIT | FMC | 124_FMC.zip | DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7t... | |
| ETL | Databricks SQL | 2023-06-15 12:43:07 | WILLIBALD_DV | ROADSHOW, WEBSHOP | ALL | ALL | 123_GENERATE_E... | DV_NAME: WILLIBALD_DV - Release: DV_R7(7) - Comment: 7t... | |

**Deploy**

● Database Link ○ Git

Database Link *: databricks_willibald

cancel | Deploy

**Deploy**

● Git ○ Custom Script

Connection Name *: <your repo>

cancel | Deploy

**Running tasks**

- automatic deployment of ddl and etl — Waiting for execution
- automatic deployment of ddl and etl — Waiting for execution
- automatic deployment of ddl and etl — Waiting for execution
- automatic deployment of ddl and etl — Waiting for execution
- automatic deployment of ddl and etl — Waiting for execution
- automatic deployment of ddl and etl — Waiting for execution
- automatic deployment of ddl and etl — Running

  65%

  Deploying 174_GENERATE_ETL.zip.

  DV_NAME: WILLIBALD_DV - Release: DV_R8(8) - Comment: 8th release - Release date: 2023/06/17 13:44:37; BV release: init(1) - Comment: initial release - Release date: 2023/06/17 13:45:53
  Elapsed time: 12 minutes

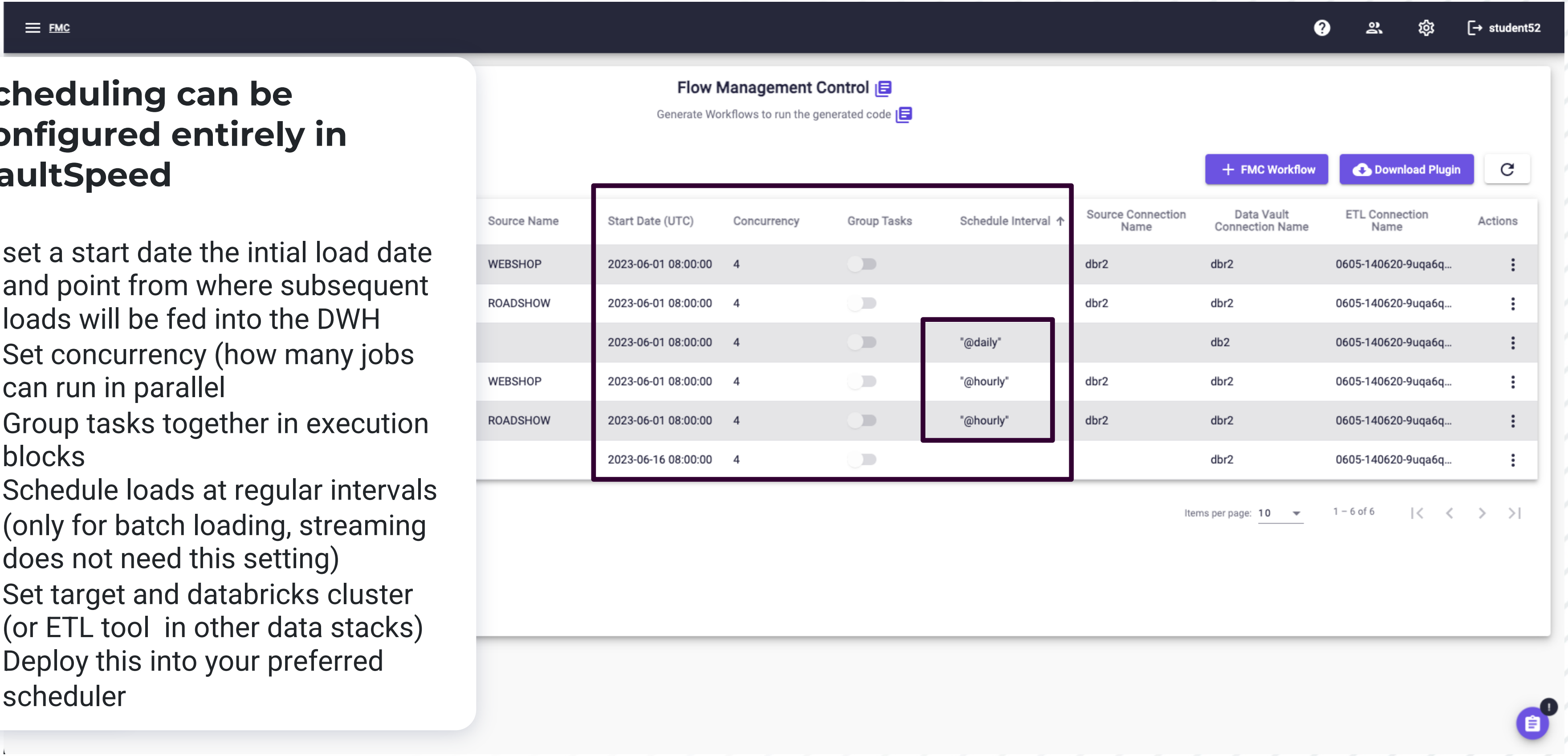
  Cancel
- obtaining generated files — Done

```
'2023-06-16 10:19:06 INFO: No task queued
'2023-06-16 10:19:46 INFO: No task queued
'2023-06-16 10:20:26 INFO: No task queued
'2023-06-16 10:21:06 INFO: No task queued
'2023-06-16 10:21:46 INFO: Executing task: get_generated_files
'2023-06-16 10:21:47 INFO: Set status of task get_generated_files to running
'2023-06-16 10:21:48 INFO: File 131_GENERATE_ETL.zip downloaded successfully in folder /Users/jonasdekeuster/vaultspeed_ag
'2023-06-16 10:21:48 INFO: file successfully downloaded.
'2023-06-16 10:21:48 INFO: Set status of task get_generated_files to done
'2023-06-16 10:22:26 INFO: No task queued
'2023-06-16 10:23:06 INFO: Executing task: get_generated_files
'2023-06-16 10:23:07 INFO: Set status of task get_generated_files to running
'2023-06-16 10:23:08 INFO: File 132_FMC.zip downloaded successfully in folder /Users/jonasdekeuster/vaultspeed_agent/agent
'2023-06-16 10:23:08 INFO: file successfully downloaded.
'2023-06-16 10:23:09 INFO: Set status of task get_generated_files to done
'2023-06-16 10:23:46 INFO: Executing task: auto_deploy
'2023-06-16 10:23:47 INFO: Set status of task auto_deploy to running
'2023-06-16 10:23:48 INFO: Successfully deployed: 132_FMC.zip
'2023-06-16 10:23:48 INFO: Set status of task auto_deploy to done
'2023-06-16 10:24:26 INFO: Executing task: auto_deploy
'2023-06-16 10:24:27 INFO: Set status of task auto_deploy to running
'2023-06-16 10:24:28 INFO: Using the file provider for retrieving connections.
'2023-06-16 10:24:29 INFO: Using the file provider for retrieving connections.
'2023-06-16 10:24:30 INFO: 1/121 Deploying 00_GENERATE_DDL_SCHEMAS.sql
'2023-06-16 10:24:33 INFO: 2/121 Deploying ALL/FLOW_MANAGEMENT_CONTROL_LAYER/FMC_UPD_RUN_STATUS_BV_WILLIBALD_DV.sql
'2023-06-16 10:24:34 INFO: 3/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_BESTELLUNG_KUNDE_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:35 INFO: 4/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_BESTELLUNG_LIEFERADRESSE_SNAPSHOTDATES_SSD_VIEW_DR
'2023-06-16 10:24:35 INFO: 5/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_BESTELLUNG_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:35 INFO: 6/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_KUNDE_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:36 INFO: 7/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_KUNDE_VEREINSPARTNER_SNAPSHOTDATES_SSD_VIEW_DRP.sq
'2023-06-16 10:24:37 INFO: 8/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_LIEFERADRESSE_KUNDE_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:37 INFO: 9/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_LIEFERADRESSE_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:37 INFO: 10/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_LIEFERDIENST_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:38 INFO: 11/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_POSITION_BESTELLUNG_SNAPSHOTDATES_SSD_VIEW_DRP.sq
'2023-06-16 10:24:38 INFO: 12/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_POSITION_LIEFERADRESSE_SNAPSHOTDATES_SSD_VIEW_DRP
'2023-06-16 10:24:39 INFO: 13/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_POSITION_PRODUKT_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:39 INFO: 14/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_POSITION_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:40 INFO: 15/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_PRODUKTKATEGORIE_PRODUKTKATEGORIE_OBERKATID_SNAPS
'2023-06-16 10:24:40 INFO: 16/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_PRODUKTKATEGORIE_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:41 INFO: 17/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_PRODUKT_PRODUKTKATEGORIE_SNAPSHOTDATES_SSD_VIEW_D
'2023-06-16 10:24:41 INFO: 18/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_PRODUKT_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:42 INFO: 19/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_VEREINSPARTNER_KUNDE_SNAPSHOTDATES_SSD_VIEW_DRP.s
'2023-06-16 10:24:42 INFO: 20/121 Deploying DDL/0_001_BV_VIEW_PIT_HOURLY_VEREINSPARTNER_SNAPSHOTDATES_SSD_VIEW_DRP.sql
'2023-06-16 10:24:42 INFO: 21/121 Deploying DDL/1_001_BV_PIT_HOURLY_BESTELLUNG_KUNDE_PIT_DDL_DROP.sql
'2023-06-16 10:24:43 INFO: 22/121 Deploying DDL/1_001_BV_PIT_HOURLY_BESTELLUNG_LIEFERADRESSE_PIT_DDL_DROP.sql
'2023-06-16 10:24:44 INFO: 23/121 Deploying DDL/1_001_BV_PIT_HOURLY_BESTELLUNG_PIT_DDL_DROP.sql
'2023-06-16 10:24:44 INFO: 24/121 Deploying DDL/1_001_BV_PIT_HOURLY_KUNDE_PIT_DDL_DROP.sql
'2023-06-16 10:24:44 INFO: 25/121 Deploying DDL/1_001_BV_PIT_HOURLY_KUNDE_VEREINSPARTNER_PIT_DDL_DROP.sql
'2023-06-16 10:24:45 INFO: 26/121 Deploying DDL/1_001_BV_PIT_HOURLY_LIEFERADRESSE_KUNDE_PIT_DDL_DROP.sql
'2023-06-16 10:24:45 INFO: 27/121 Deploying DDL/1_001_BV_PIT_HOURLY_LIEFERADRESSE_PIT_DDL_DROP.sql
'2023-06-16 10:24:45 INFO: 28/121 Deploying DDL/1_001_BV_PIT_HOURLY_LIEFERDIENST_PIT_DDL_DROP.sql
'2023-06-16 10:24:47 INFO: 29/121 Deploying DDL/1_001_BV_PIT_HOURLY_POSITION_BESTELLUNG_PIT_DDL_DROP.sql
'2023-06-16 10:24:47 INFO: 30/121 Deploying DDL/1_001_BV_PIT_HOURLY_POSITION_LIEFERADRESSE_PIT_DDL_DROP.sql
'2023-06-16 10:24:47 INFO: 31/121 Deploying DDL/1_001_BV_PIT_HOURLY_POSITION_PIT_DDL_DROP.sql
'2023-06-16 10:24:47 INFO: 32/121 Deploying DDL/1_001_BV_PIT_HOURLY_POSITION_PRODUKT_PIT_DDL_DROP.sql
'2023-06-16 10:24:48 INFO: 33/121 Deploying DDL/1_001_BV_PIT_HOURLY_PRODUKTKATEGORIE_PIT_DDL_DROP.sql
'2023-06-16 10:24:48 INFO: 34/121 Deploying DDL/1_001_BV_PIT_HOURLY_PRODUKTKATEGORIE_PRODUKTKATEGORIE_OBERKATID_PIT_DDL_DR
'2023-06-16 10:24:48 INFO: 35/121 Deploying DDL/1_001_BV_PIT_HOURLY_PRODUKT_PIT_DDL_DROP.sql
'2023-06-16 10:24:49 INFO: 36/121 Deploying DDL/1_001_BV_PIT_HOURLY_PRODUKT_PRODUKTKATEGORIE_PIT_DDL_DROP.sql
'2023-06-16 10:24:50 INFO: 37/121 Deploying DDL/1_001_BV_PIT_HOURLY_VEREINSPARTNER_KUNDE_PIT_DDL_DROP.sql
'2023-06-16 10:24:50 INFO: 38/121 Deploying DDL/1_001_BV_PIT_HOURLY_VEREINSPARTNER_PIT_DDL_DROP.sql
'2023-06-16 10:24:50 INFO: 39/121 Deploying DDL/1_006_BVLCI_LOAD_CYCLE_INFO_DDL_DROP.sql
'2023-06-16 10:24:51 INFO: 40/121 Deploying DDL/1_006_BVLWT_FMC_BV_LOADING_WINDOW_TABLE_DDL_DROP.sql
'2023-06-16 10:24:52 INFO: 41/121 Deploying DDL/1_006_DVLCI_DV_LOAD_CYCLE_INFO_DDL_DROP.sql
```

# 20. Scheduling

## Scheduling can be configured entirely in VaultSpeed

- set a start date the intial load date and point from where subsequent loads will be fed into the DWH
- Set concurrency (how many jobs can run in parallel
- Group tasks together in execution blocks
- Schedule loads at regular intervals (only for batch loading, streaming does not need this setting)
- Set target and databricks cluster (or ETL tool  in other data stacks)
- Deploy this into your preferred scheduler



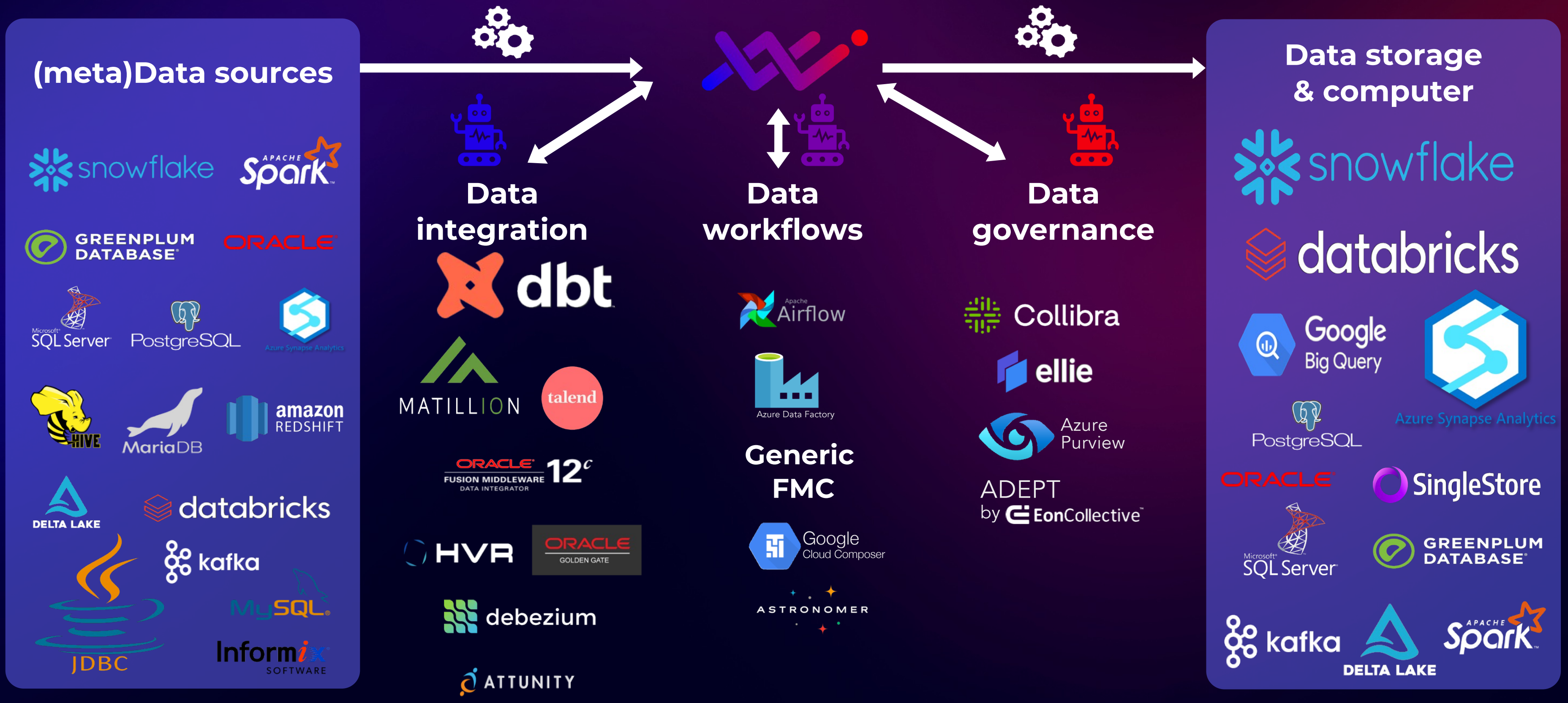
Flow Management Control

Generate Workflows to run the generated code

| Source Name | Start Date (UTC) | Concurrency | Group Tasks | Schedule Interval ↑ | Source Connection Name | Data Vault Connection Name | ETL Connection Name | Actions |
|---|---|---|---|---|---|---|---|---|
| WEBSHOP | 2023-06-01 08:00:00 | 4 | ○ |  | dbr2 | dbr2 | 0605-140620-9uqa6q... | ⋮ |
| ROADSHOW | 2023-06-01 08:00:00 | 4 | ○ |  | dbr2 | dbr2 | 0605-140620-9uqa6q... | ⋮ |
|  | 2023-06-01 08:00:00 | 4 | ○ | "@daily" |  | db2 | 0605-140620-9uqa6q... | ⋮ |
| WEBSHOP | 2023-06-01 08:00:00 | 4 | ○ | "@hourly" | dbr2 | dbr2 | 0605-140620-9uqa6q... | ⋮ |
| ROADSHOW | 2023-06-01 08:00:00 | 4 | ○ | "@hourly" | dbr2 | dbr2 | 0605-140620-9uqa6q... | ⋮ |
|  | 2023-06-16 08:00:00 | 4 | ○ |  |  | dbr2 | 0605-140620-9uqa6q... | ⋮ |

Items per page: 10 1 – 6 of 6

# 21. Supported DBs



## (meta)Data sources

- snowflake
- APACHE Spark
- GREENPLUM DATABASE
- ORACLE
- Microsoft SQL Server
- PostgreSQL
- Azure Synapse Analytics
- HIVE
- MariaDB
- amazon REDSHIFT
- DELTA LAKE
- databricks
- kafka
- JDBC
- MySQL
- Informix SOFTWARE

## Data integration

- dbt
- MATILLION
- talend
- ORACLE FUSION MIDDLEWARE DATA INTEGRATOR 12c
- HVR
- ORACLE GOLDEN GATE
- debezium
- ATTUNITY

## Data workflows

- Apache Airflow
- Azure Data Factory
- Generic FMC
- Google Cloud Composer
- ASTRONOMER

## Data governance

- Collibra
- ellie
- Azure Purview
- ADEPT by EonCollective

## Data storage & computer

- snowflake
- databricks
- Google Big Query
- Azure Synapse Analytics
- PostgreSQL
- ORACLE
- SingleStore
- Microsoft SQL Server
- GREENPLUM DATABASE
- kafka
- DELTA LAKE
- APACHE Spark

# 22. Installation requirements

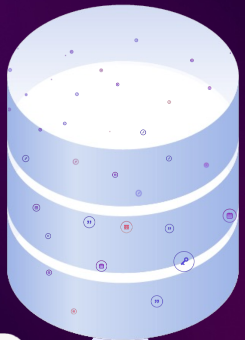
- VaultSpeed is cloud SaaS
- No virtual machines, no AMI's, no Dockers or Kubernetes needed
- Always running on the latest version
- Always using the latest DV2.0 support

- Lightweight Java Agent that sits in your network (hybrid approach)
- Agent handles source metadata harvest and code deployments
- Agent can be installed on any OS, using Java (oracle) JDK 8
- Install takes 10 minutes, installation scripts for Linux, Windows, MacOS available

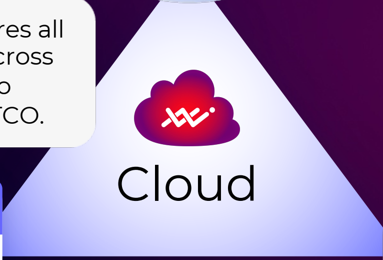
VAULTSPEED

# Core components to automate data transformation

## Metadata repository



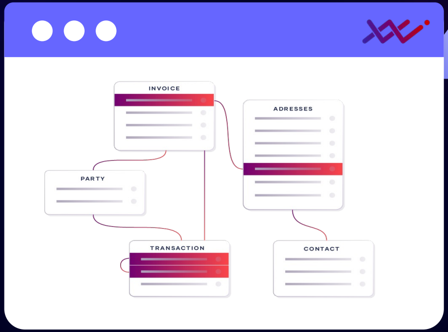
Store all types of metadata with zero effort.
Is capable of grouping metadata objects in signature groups.



## Cloud

Cloud-native Saas ensures all learnings are shared across all customers. It also drastically lowers the TCO.

## GUI



Powerful data modeling interface enables users to harvest and enrich source metadata, to model the target outcome, to create signature groups and tag objects accordingly. The data modeler proposes you a solution, you just need to confirm, enrich, correct

## Template engine

```
Template NPV
Comp Group NPV
Repeat by SAT
CF/(1.rate)^time
end group
```

Template engine that is capable of translating repeatable logic into data transformation code. Both integration and business logic can be applied.

# 23? Not included: Data model changes

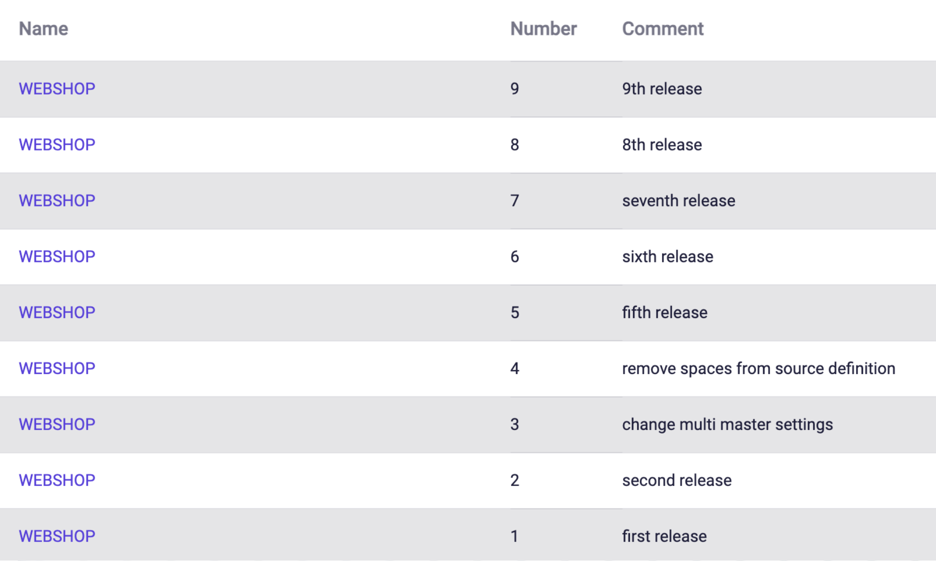
## Changes to the data model

VaultSpeed has over 600 rules in the back checking for change and automatically deriving what do to if you change the model.

- New attribute
- Deleted attribute
- New BK attribute
- New source
- New relationship
- Deleted relationship
- Dropped table
- Master/slave changes
- Single master to multi-master change
- …

Our delta generation will generate the delta between 2 releases and also deliver the **data migration scripts** in case they are needed.

For example: a hash recalculation when the business key changes.



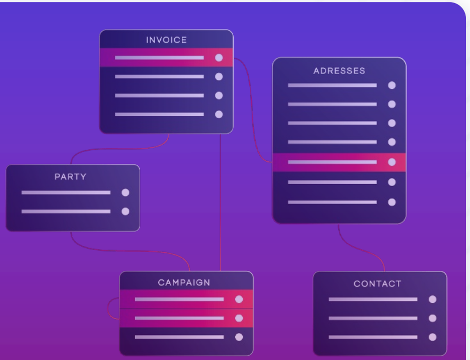
| Name | Number | Comment |
|---|---|---|
| WEBSHOP | 9 | 9th release |
| WEBSHOP | 8 | 8th release |
| WEBSHOP | 7 | seventh release |
| WEBSHOP | 6 | sixth release |
| WEBSHOP | 5 | fifth release |
| WEBSHOP | 4 | remove spaces from source definition |
| WEBSHOP | 3 | change multi master settings |
| WEBSHOP | 2 | second release |
| WEBSHOP | 1 | first release |

VAULTSPEED

# Disruptive Automation

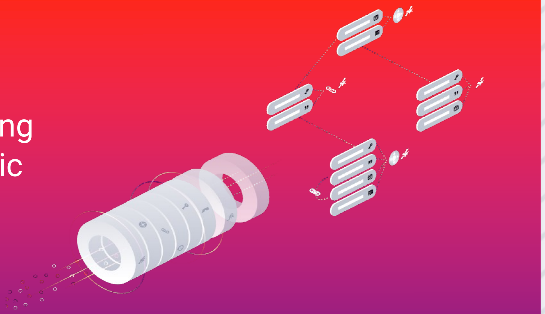
## Data model driven

develop integrated loading patterns instead of isolated pipelines



## GUI driven

object & attribute tagging to apply repeatable logic



## Pre-built integration templates for Data Vault

integration testing of DV2.0 template logic is our responsibility

## Abstract model templates

to apply industry-specific logic

LTV, CAC, NPV, PRR, Defect density...

```
Template NVP_CALC

comp_group_start MIV_GROUP
componentgrouprepeatedbycom

consists of aggregated inli
componentrepeatedbycomponen
Attribute FOREIGN_KEY
expressedBy NHL_SRC.FOREIGN
expressionRepeatedByColumn
NHL_SRC.FOREIGN_KEY
Aggregated Attribute NPV
expressedBy NHL_SRC.NET_CAS
NHL_SRC.DISCOUNT_RATE) ^
(NHL_SRC.NUMBER_OF_PERIODS)
expressionRepeatedByColumn
NHL_SRC.NET_CASHFLOW
```

VAULTSPEED

Want to learn more?

**Visit our website**
www.vaultspeed.com

**Read the docs**
docs.vaultspeed.com

**Join the community**
community.vaultspeed.com

**Talk to sales**
sales@vaultspeed.com

## Thank you

**Jonas De Keuster**
VP Marketing
VaultSpeed
Jonas.dekeuster@vaultspeed.com

**Florian Mayer**
Regional Director DACH & EE
VaultSpeed
Florian.mayer@vaultspeed.com