



DDVUG DWH Automation Challenge

TDWI München June 2023





WhereScape Intro



Sultan ShiffaSenior Solutions
Architect, WhereScape



Simon SpringOperations Director,
WhereScape



Thomas HeuerSales Director EMEA,
WhereScape



WhereScape Applications

WhereScape[®]



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

WhereScape®



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





WhereScape Enablement Packs



Your Modeling Styles

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



Your Platforms

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



Your Patterns

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





































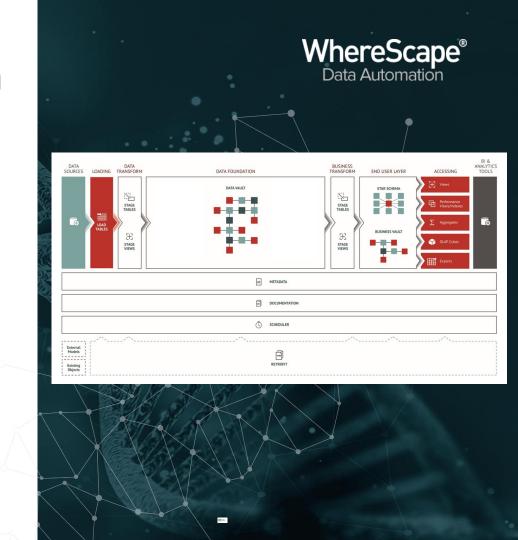






High-level Walkthrough

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



Use cases 1 + 2 Hierarchical Links & Multi-Active Satellites



Test cases in the data model hierarchical link





Test cases in the data model multi-active satellite



2



Use cases 3 + 4 Relationships, Driving Keys and m:n Tables



Test cases in the data model Identifying Relationship and Driving Keys



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

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

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

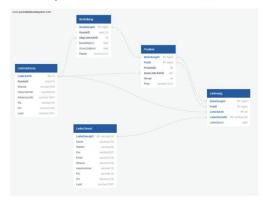


Test cases in the data model m:n table without own key

4

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

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



Use cases 5 + 6 Multiple Sources and Historicised Ref. Tables



Test cases in the data model Integration of the order



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



6

Test cases in the data model (Historicised) reference table

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



5. Early Integration (of the order)

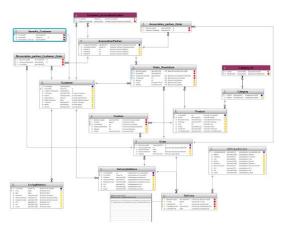
Data Vault design

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

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

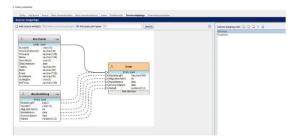
KundenBK data transformation

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



WhereScape® Data Automation

Webshop source mapping for order

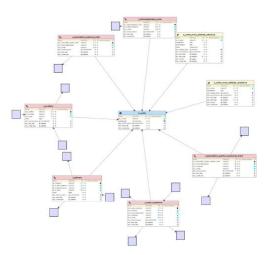


Roadshow source mapping for order

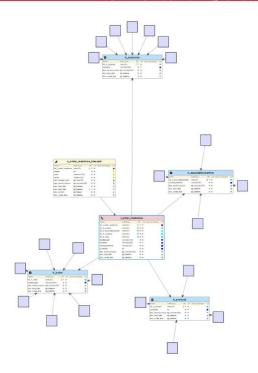




Raw Vault generated automatically by 3D



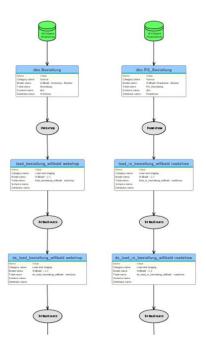
WhereScape® Data Automation



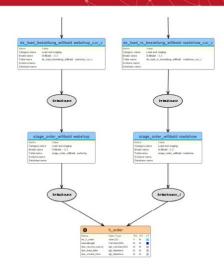


Load and Staging generated automatically by 3D

The order table is sourced from webshop and Roadshow.



WhereScape® Data Automation



Data Lineage in WhereScape RED

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



Use cases 7-10 Test cases in the data part 1 of 2



Test cases in the data

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

7

8

9

10



7. Duplicate Data

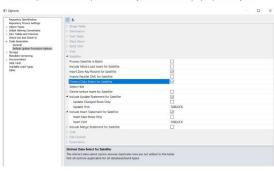
Option 1

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

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



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





Option 2

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

Solutions:

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

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

Add a stage entity with a row_number column.

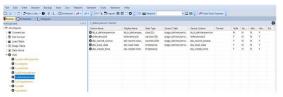
Records with a value 1 from the stage go into the vault, and all other records go to an error reporting table.



8. Rows without business key

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

Example: deliveryService hub without BK



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

SELECT

CONVERT(binary(16), HASHBYTES('MD5',

ISNULL(RTRIM(CONVERT(NVARCHAR(100),NULL)),'UNKNOWN')+'|'),2)

Generated hash key: 0xD7F91BEB665F968B127DE83FB3777C0B

- or -

SELECT

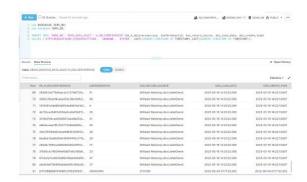
CONVERT(CHAR(32), HASHBYTES('MD5',

ISNULL(RTRIM(CONVERT(NVARCHAR(100),NULL)),'UNKNOWN')+'|'),2)

Generated hash key: D7F91BEB665F968B127DE83FB3777C0B



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

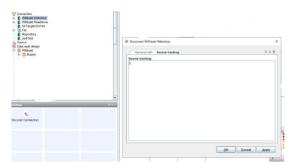




10. Deletion of Business Keys

Data Vault design

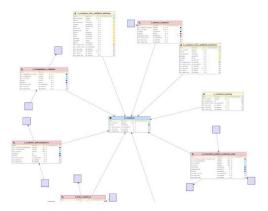
All connections which need source tracking must be configured for tracking and set as Y:



WhereScape® Data Automation

Raw Vault generated automatically by 3D

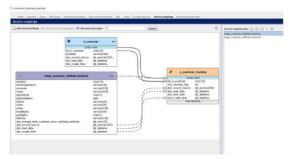
Tracking satellites do not maintain dss_change_hash nor dss_version columns. A new dss_tracking_flag attribute is added..



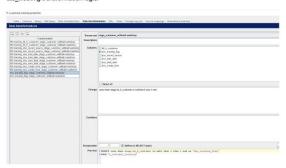


Load and Staging generated automatically by 3D

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



dss_tracking transformation logic:



WhereScape® Data Automation

Data Lineage in WhereScape RED

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



Sample Data of Satellite tracking for customer



Use cases 11-13 Test cases in the data part 2 of 2



Test cases in the data II

- DELIVERY ADDRESSES (Lieferadresse) without CUSTOMER (Kunde)
 - The first delivery contains delivery addresses for which there is no record with the same customer ID (KundeID) in CUSTOMER (Kunde KundeIDs 999, 998 and 997).

11

- Deletions of ORDERS (Bestellung)
 - The orders are relevant for counting and are deleted during the dates of deliveries.
 - Between period 1 and 2 the orderIDs (BestellungID) 99, 220 and 465.
 - Between periods 2 and 3 the orderIDs (BestellungID) 1470 and 1288.
- Changes in the dimensions
 - The hierarchy of the product CATEGORY (Kategorie) changes completely with both deliveries. So we have 3 different product hierarchies. These are to be displayed as aswhat at the respective reporting time.

13

12

Use case 15 Business Rules



Business Rules

15

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



15. Business Rules

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

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

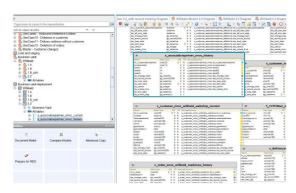
A rule is implemented to identify customers without customer_id in the (Roadshow order) RS_Bestellung based on their credit card, cc-company and valid-to data.

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

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

Additional Business Vault

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

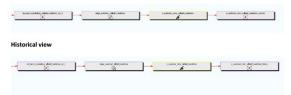


Data Lineage in WhereScape RED

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



Current view



Sample Data

Current satellite view

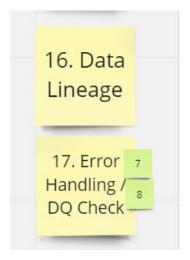
Builder Schoolser Buggers Sever Development Buggers	26 July 1 v COURS has della according white sector and											
Saroues -	N Form. Date	et renopero.	incline	SHROW	Minna .	to three.	des sarrot	des less, date	da.mr.dec	- frs. serson	do.out	
C Supera Study	*Des*** -	Goddy	0000 2925		Madestant			2021-06-02			2823-66-60.	
CiCities	79757 See. 15	Goodel	0000 1551	09/71	Markeyard	43007365	Withold F	2523-06-52	2825.06.02	11	2823.66.60	
25's promotive enter more named	120 x 25 - 120	Goodel	0000 9055	00.03	1856	014CFW5-	WRIDNIS F	3629-06-02	2629-06-02	11	2823-06-02	
F s granted imparter, many history	Changling.	COMMIT	0000 6325	06/98	Manhercard	11.1×9.00%	Without ft	2923-06-62	2623-06-02	1	2823-66-60	
35 s.category, M. offectivity current	Composid. IN	Gustaf	2517 5500	06/12	Mastercard	Style-Cookle	WHISHIR R.	2923-06-62	2825-06-62	1	2623-06-00	
S. s. rategory, N. offectivity history	3-30 (BL), V. 27	COOKE	0000 1768	09/15	WA	1010c.4792.i.	William R.	2823-09-62.	2925-09-02	1	2023-04-02	
	(A) (Frend W)., 170	Goodel	0000 0000	00/99	WSA	mindage.	Williams R.	2022-06-62	2623-06-02		2623 66-62	
(C) s, category, evoc, correct	Nove-161 L. 186	00004	0000 8757	0454	WSA	HIVEOG.	Withold FL.	2525-06-02.	2625-06-02		2623.66-60	
(S) IL category, treat, history	N 10m C. 130	GUSAN	0000-4114_	04/03	Maidencard	WASS-TWO.	Willhold R.	2629-06-62	2623-06-02	1	2823 66 60	
3 special contract and services of the street of the service of th	3/3940/- 40T	Godwar	6501 0106	10/00	Mastercard	(thisobal.		2823-06-62	2529-06-02		2823 66-62	
Experience association partner of factivity (Nation	NadOts	Gooke	0000 SZTE	09/18	VISA	O-MINTEL	Willbeld F	2923-06-62	2625-06-02	1	2623-06-02	
Countries year, will belief, to a few parties.	3-300 page 55	900Ad	0000 8621		VISA			39-99-635			2027-06-02	
Especial procedures and a second security and a second security and a second se	* DIRECT N 193	Glodukut	00004844	00/23	VISA			2525-06-62			2823-06-02	
Equations may will be in present a pursuant	Containe.	Global	0000 5720		964		Williams R.		3621-06-02		2023-06-00	
25 continue, moc without wearing history	90-750° L.	Godvint	0000 2813		American E.				2023-06-02		2623-06-02	
Espatome passing overt	€ MODEN'	Quaker	0000 1578	11/23	Maidmand	W50A-4075	Williams R.	2022-06-02	2623-06-62	1	2623 66-62	
25's nationer Various history	-N6004	godaler	0808 1765		VISA			2525 00 02.			20-90 1585	
Explores/resect cores	⊕w '->/ *- 65	Gustar	0000 1562		Madercard			2023-06-62			2023-66-62	
Estellandamentamen	23 "ps:(04)(**)	COMAN	0000 2730		American E.			3825-09-02.			30-60-035	
Es delivereddess mos curent	⊕31400—. 185	Quotair	0605.5806			11.1/15450					2823-66-62	
Statisfieredden nochhory	€0000××× 200	(Social)	0000 1750		984	+60-29-C-		2525-09-02-			-52 40 1555	
S. s. deliverservice, moc., current	€05515+m4., 401	Quality	2101 5500		Madercard			2823-06-02			2623-06-02	
	(A)300 bbytg. 307	Goodel	6613.3800		VISA			2029-04-02-			2023-06-00	
3 julieli veryservice ji moc ji lidory	-301-1EAL	GOSAN	0000 TE36_		Metercard	\$13A45e)		2522-06-62			20-00-02	
(E) s_fringaddwsc_ma_cument	3 AT RES 146	Gooder	0000 1674		956	PRECENTAL.		2623-04-62			5853-08-05-	
(E) x, Evingaddrass, FW, Natory	DOMPTH.	CONNE	0643.3806		American E.		Willbeld F.				203-00-02	
30 spederumocy/literal poeterwormers	(€)00×6×50A 190	GOORIE	0000 25TE		VISA	10.10	Williams R.		2623-06-62		2623-66-60	
30 code, no.c. will bald, code leve, bittory	30068-0-R, 221	GOOMIL	_ E159 6000		Minitercard	INTROOPS.	WRISHIT.		2525 09-02.		2823-69-62	
30 sunderunes cultilated, websing purent	(c) 12 (17	Goodel	0000 7784		Mastercard			2623-06-02			2833-66-02	

Historical satellite view

Ot Sever Development # # # # # # # # # # # # # # # # # # #	N.A.com.		WITTOWNER.	Switters.	outlists.	Minn		Are report -		du me des				
T bearing that	NA. 100.	A1	Gashi	2000 ATG)		American E.	Our SEE-A.	William R.	2523-06-62				9999-12-21	
vSC View	S street		COAR	DADO ENDO		WA.				2023-06-02			0999, 12-11	
X sussciptospaties moscoures	3. New York,	319	Good	0000 1268		VICA	5000511A			3503-86-00.			6999-12-31	
S cancertorpather rescured	Continu	100	GOAL	2000 1868		Maderart	ABHIROS.	MINIOR II.		2373-36-00			6009-12-21	
	30 Ta No.	799	manage	0000 K550		Motorcast	umeshs	Millshoot E.	7873 56 87	2323-96-52			1999.32.33.	
(i) s_coregory_V_effectivity_current	30 traine.		Guant	0000 1178 -		985A	iness SeAL	Wilheld IL.					6909-12-21	
(ii) s_coregory_N_effectivity_history	37.00996		Guahi	E700 2000		1056	William Street	Willand II.	2022-06-62	2022-06-05			6999-12-21.	
St congrey/weignmen	30°s/40.	293	COAM!	0000 1581		Makeraal	9-08S+-11U	Wilheld S.	1973 56 67	2523-96-02			1999 12:31.	
(ii) is considered, were failed as	976-HIS.		Guand	25% 3000		WSA	NIORY-	Wilself.		2023-96-02.			6969-12-21.	
 Container, seasof all importment of factivity, in. 	Systema.	155	Good 4	0000 0052	05/21	1956	tolymen.	Williams S.	2679-06-67	2523-96-52		2923-66-61	1999 12-31.	*
X sustainer personal impartment leftectivity in	W dealto	158	COAR	0000 5950		Markettand		Wilheld R.		2523-86-02		2013-06-01	9999-12-31	
(if a customer proc. will bold roadshow current	80	259	GUARA	0000 6152	09/18	Mutercard	-Dalnydo.	Willbald R.	2623-06-62.	2003-96-00	1	2823-66-82.	9999-12-21.	¥
C a contrast two sell bold southern his or	82.00	ins	manar	6933 D000	09/21	986	wedshark.	PRIDWING No.	2579-09-51.	2023-06-02	1	2573 09-67	9999-17-71.	*
Of a codomer rever will bold webshoo current	SCHAPM.		goods.	0000 1787	12/25	American L.	1043.264	Wilbeld IL.	2523-06-62.	2523-96-02	1	2523-06-82	9999-12-21.	
It's cotomer range will bald webshop below.	Exc1-05.	74	Gasour	0000 9050	01,03	WSA	fairus.	Wilbeld R.	2823-86-62	2023-36-02	1	2823-66-82.	9999-12-21	.7
35 supplement fracking corners	3-7mid-46.		DONAL	5660 6000	06/23	Minlestard	REWHOUL.	Williams R.	2829-09-62	20-96-02.	1	2823-66-82	9999-12-31.	
25 s. contamer, treating history	Parwellin.	373	09004	0000 6625	11/13	American C.	oftion no.	Wilbeld FL.	2823-06-61	2023-96-02	- 1	2023-66-62.	9999-12-31L	1
2 safetymy, hamsest commit	2500 AL		Gasker	6700 1000	00/78	Nedercard	názečeni-	Wilbeld R.	2023-06-02.	2023-36-02	- 1	2023-66-82	9909-12-21	7
C. Safety Special Distory	235-yat-		00004	0000 2465	09/94	Material	0.000000	Williams R.	JEG 09-62.	2023-99-02	1	2513-09-65.	5999-12-31.	*
2. satisfactions and county	Excident.	207	dissolut	0524 5100		Metercard	PI-ct VH	Wilbeld IL.		2023-96-62			999-U-JL	
X saniveyantes one name	336		Granulat	0000 Fr52		Mindercoid	N-276	Willbaid R.	2523-06-62	2523-36-55	1	2823-66-82.	9999-12-21	
	2.8/Aberto		00004	D000 1585		Amerian E.	#+C#131-			2023-06-02.			9999 12:51.	
25 Citeliveryservice/executament	* pobjects.		Goddulf	0000 3680		American L.		Wilbeld R.		2023-06-02			9999-12-31L	
(%) subeliverseevice/worchistory	Etroppie.		GUNNA	0000 6405		American E.				2503-96-02			9999-12-21.	
(%) s_thingscorest_ma_nument	Extint're.		DONAL	8624 5180		Modernat	opastich.	WRMT.					9999 32-33	
[5] s_fringeddress_ma_triebry	*2854W.		QUART	0000 8164	12/23	Metercard	AUTTIT/No.	Willbald R.		2023-06-02			9999-12-31L	
(K. KURRE, PROC, WEST BARRY, SOMEON, SURRES	31-85090		GUOVA	0000 7568		Nedercard	Measur.	Willoud R.					9999-52-25	
(i) context more without produces bitting	329544		DROOM	0000 8965	11/01	YOA	OF MANY	William K.		2025-96-02			9999-12-51	*
(i) carein, more, will be 4, sweeting, current	N TV PHIND.	278	CONNE	0000.7798	3931	Modernst	WITESAL.	Withwist R.	2011-06-07	3073-36-02		201-66-02	\$999.D.TL	3

Use cases 16-22 Non-Data-Vault-Specific Requirements











Get Hands On!



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

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





Schedule a demo @ www.wherescape.com



Thank you

