Banian

Banian

MODEL DRIVEN DWH MEETS DATA-OPS

DDVUG, 26.11.2020

AGENDA

- Einführung Model Driven DWH
- Beispiele aus Kundenprojekten zu DataOps
- Erfahrungen

ÜBER MICH

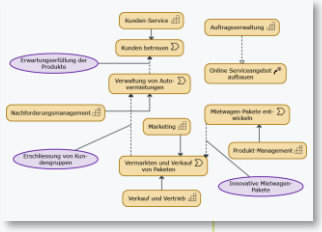
>16 Jahre Erfahrung rund um Daten

2019	Mitgründer der Banian AG
	BI Consultant
	BI & CDQ CC Manager
	DWH Engineer & Team Lead
2004	Business Analyst

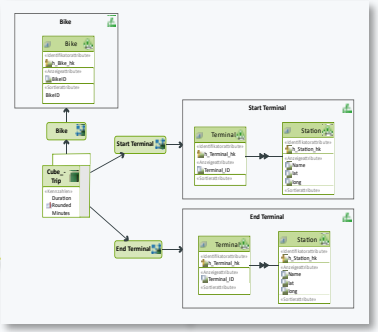
MODEL DRIVEN

MODELLIERUNG

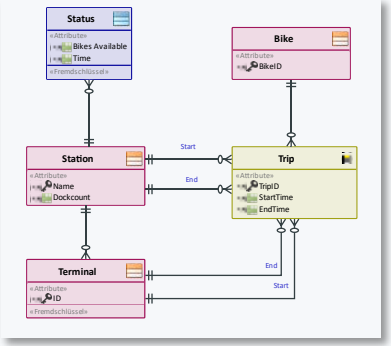
Motivation and Goal View (Archimate)

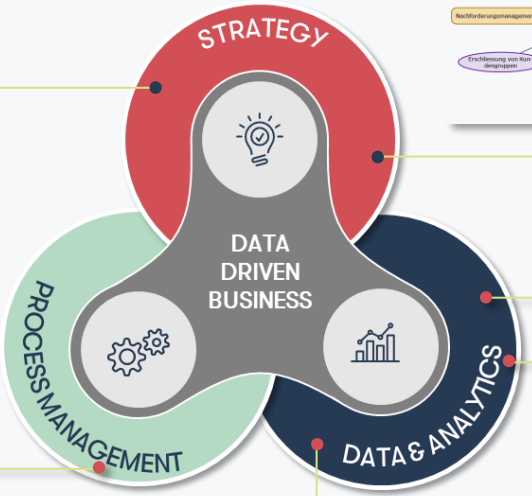


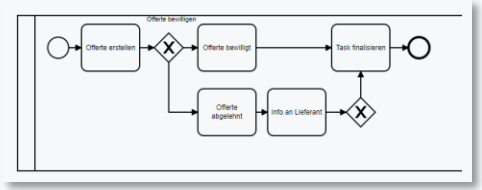
Analytical Model (ADAPT)



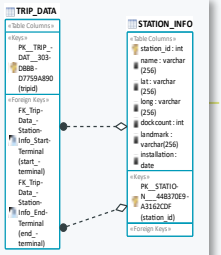
Business Object Model



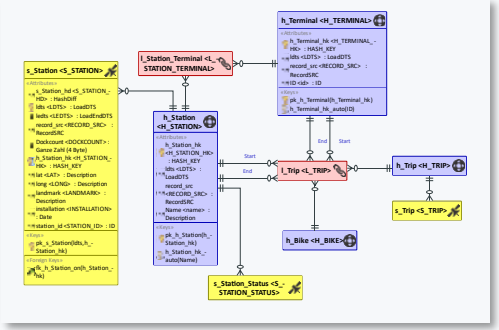




Process Model (BPMN)



Physical Data Model

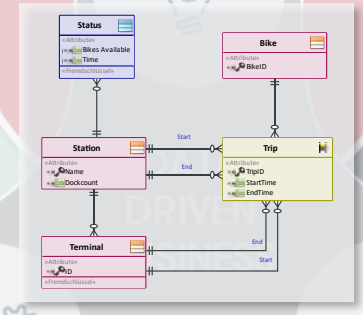


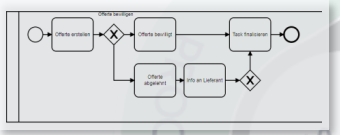
Logical Data Model (Data Vault)

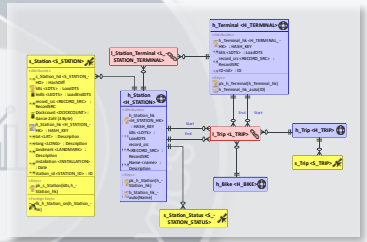
VOM MODELL ZUM SYSTEM / APPLIKATION

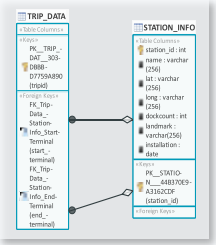
Transaktionelles System

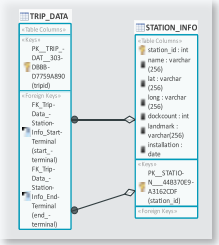
Analytisches System







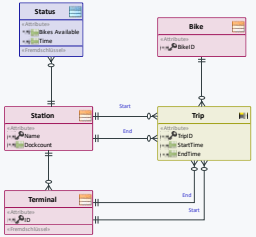




Automatisierung

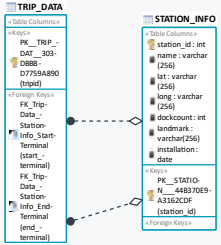
MODEL DRIVEN DWH AUTOMATION

01



GESCHÄFTSOBJEKT-MODELL

Modellierung der Geschäftsobjekte auf Basis der effektiven Informationsträger

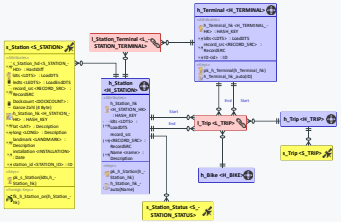


PHYSISCHES MODELL

Abbildung der Quellsysteme
Mapping der Quellsysteme auf die Geschäftsobjekte

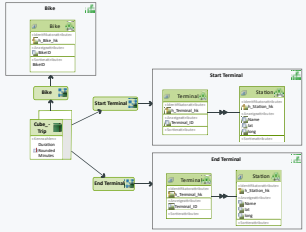
02

03



DWH MODELL (DATA VAULT)

Generierung des DWH Layers gemäss Standard Vorgaben und Templates



ANALYSE MODELL

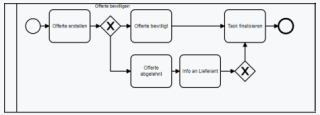
Dokumentation der Analysemodelle

04

VALUE ← TCO → VALUE

MODEL DRIVEN COMPANY

Verknüpfung der Prozesse / Applikationen mit dem Geschäftsobjektmodell

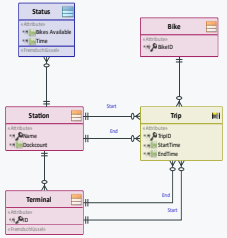


PROZESSMODELL

Abbildung der Geschäftsprozesse mit Geschäftsobjekten und Datenobjekten

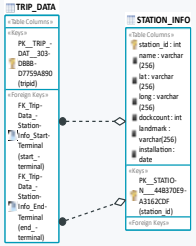
01

02



GESCHÄFTSOBJEKT-MODELL

Modellierung der Geschäftsobjekte auf Basis der effektiven Informationsträger

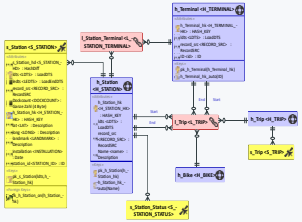


PHYSISCHES MODELL

Abbildung der Quellsysteme
Mapping der Quellsysteme auf die Geschäftsobjekte

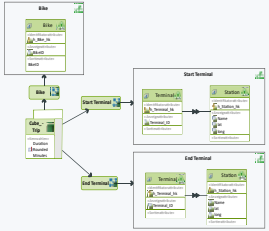
03

04



DWH MODELL (DATA VAULT)

Generierung des DWH Layers gemäss Standard Vorgaben und Templates



ANALYSE MODELL

Dokumentation der Analysemodelle

05

UMFRAGE

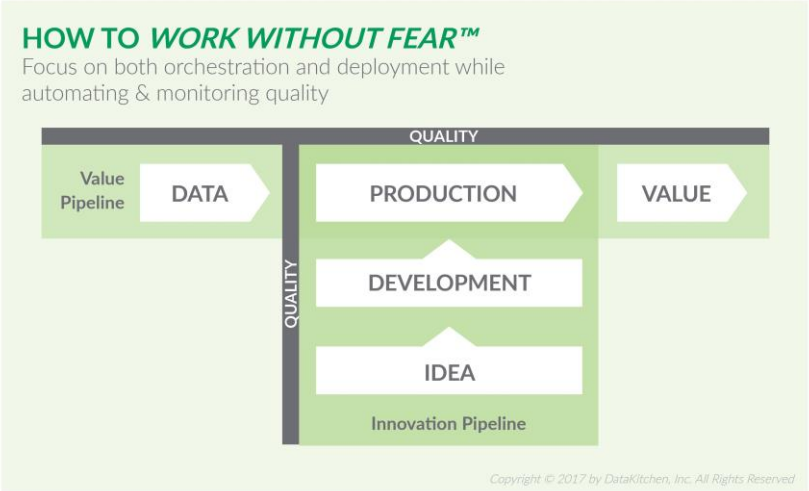
VERWENDEST DU EINEN “MODEL DRIVEN” ANSATZ?

a) JA
b) JA, ABER OHNE MODELLIERUNGS-TOOL
c) JA, ABER VON DEN QUELLEN ABGELEITET
d) NEIN

DATA OPS

-- *KUNDENBEISPIELE* --

DATAOPS



1. Add Data and Logic tests
2. Use a Version Control System
3. Branch and Merge
4. Use Multiple Environments
5. Reuse & Containerize
6. Parameterize Your Processing
7. Work Without Fear™

https://medium.com/data-ops/dataops-in-7-steps-f72ff2b37812
https://www.dataopsmanifesto.org/

ANFORDERUNGEN

- Persönliche Entwicklungsumgebung
- Standardisierte Entwicklung
- Continous Integration & Quality
- Continous Delivery

UMFRAGE

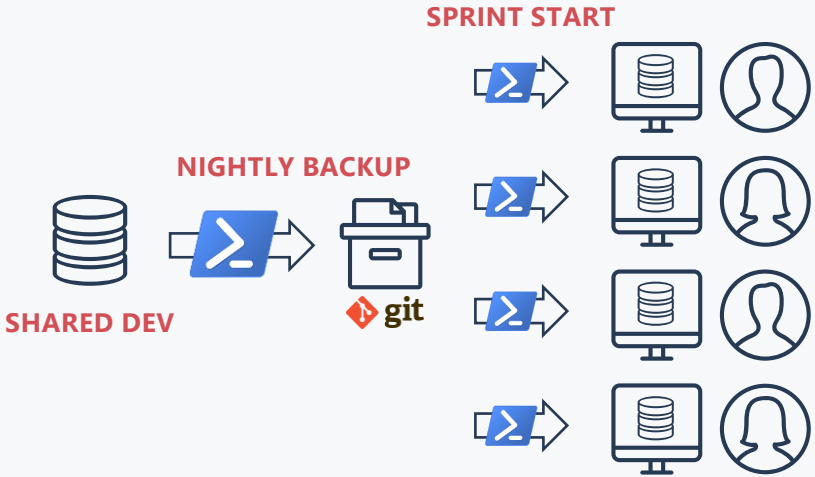
ERFÜLLT DEINE DWH AUTOMATISIERUNG DIE DATA OPS “KRITERIEN”

a) JA, VOLLSTÄNDIG
b) TEILWEISE; STANDARDISIERTE ENTWICKLUNG
c) TEILWEISE; PERSÖNLICHE ENTWICKLUNGSUMGEBUNGEN
d) NEIN

Beispiel 1

PERSÖNLICHE ENTWICKLUNGS-UMGEBUNG

SQL SERVER, WHERESCAPE, POWERSHELL, GIT



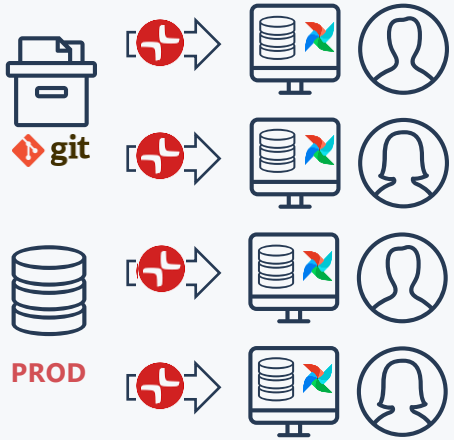
1. Aufsetzen einer frischen Umgebung
2. Deployment Application einspielen
3. SQL Agent Jobs einspielen
4. Clean Up

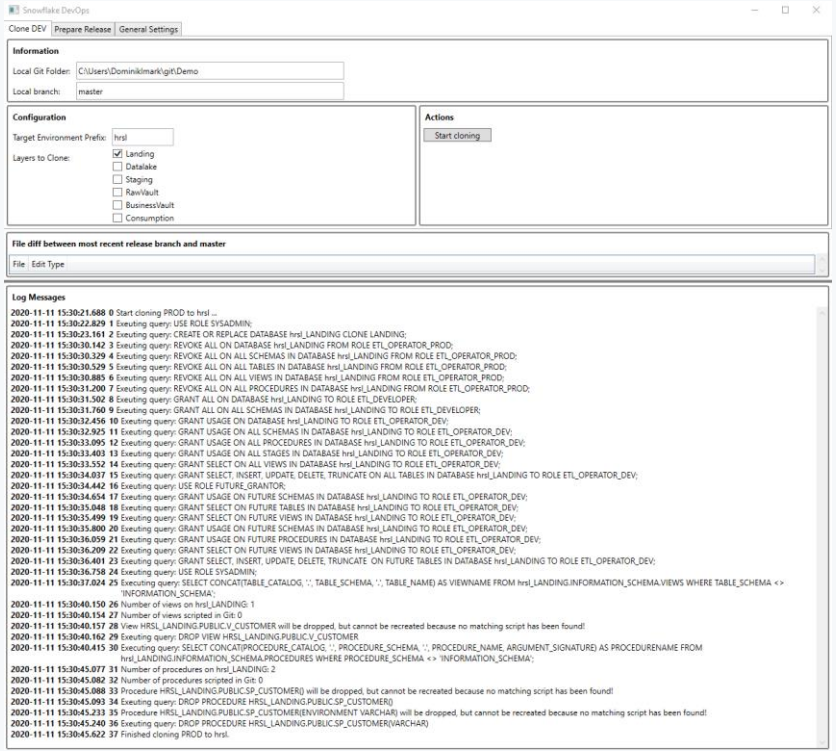
Beispiel 2

PERSÖNLICHE ENTWICKLUNGS-UMGEBUNG

SNOWFLAKE, MID INNOVATOR, C#, GIT, APACHE AIRFLOW

SPRINT START





1. Clone der Prod DB Umgebung
2. Erstellen einer Airflow Umgebung
3. Deployment der Jobs
4. Checkout des Code und Anpassung an die neue Umgebung

ANFORDERUNGEN

- Persönliche Entwicklungsumgebung
- Standardisierte Entwicklung
- Continous Integration & Quality
- Continous Delivery

Beispiel 1

STANDARDISIERTE ENTWICKLUNG (MODEL DRIVEN)

SQL SERVER, WHERESCAPE, GIT, AZURE DEVOPS, BIML

INTERFACE DESIGN

01

	A	B	C	D	E	F	G	H	I	J	K	L
1	Ordinal	Column	Rename	Type	Length	Precision	Scale	IsNullable	FastParse	Textqualifier	Column Separator	Conversion Rule
2	1	full_name	full_name	nvarchar	100			0			.	
3	2	age	age	smallint				0			.	
4	3	league	league	nvarchar	20			0			.	
5	4	season	season	nvarchar	9			0			.	
6	5	position	position	nvarchar	10			0			.	
7	6	Current Club	Current_Club	nvarchar	50			0			.	
8	7	minutes_played_overall	minutes_played_overall	smallint				0			.	
9	8	minutes_played_home	minutes_played_home	smallint				0			.	
10	9	minutes_played_away	minutes_played_away	smallint				0			.	
11	10	nationality	nationality	nvarchar	50			0			.	
12	11	appearances_overall	appearances_overall	smallint				0			.	
13	12	appearances_home	appearances_home	smallint				0			.	
14	13	appearances_away	appearances_away	smallint				0			.	
15	14	goals_overall	goals_overall	smallint				0			.	
16	15	goals_home	goals_home	smallint				0			.	
17	16	goals_away	goals_away	smallint				0			.	
18	17	assists_overall	assists_overall	smallint				0			.	
19	18	assists_home	assists_home	smallint		3	2	0			.	
20	19	assists_away	assists_away	smallint		3	2	0			.	
21	20	penalty_goals	penalty_goals	smallint		3	2	0			.	
22	21	penalty_misses	penalty_misses	smallint				0			.	
23	22	clean_sheets_overall	clean_sheets_overall	smallint				0			.	
24	23	clean_sheets_home	clean_sheets_home	smallint				0			.	
25	24	clean_sheets_away	clean_sheets_away	smallint				0			.	
26	25	conceded_overall	conceded_overall	smallint				0			.	
27	26	conceded_home	conceded_home	smallint				0			.	
28	27	conceded_away	conceded_away	smallint				0			.	
29	28	yellow_cards_overall	yellow_cards_overall	smallint				0			.	
30	29	red_cards_overall	red_cards_overall	smallint				0			.	
31	30	goals_involved_per_90_overall	goals_involved_per_90_overall	decimal		4	2	0			.	
32	31	assists_per_90_overall	assists_per_90_overall	decimal		4	2	0			.	
33	32	goals_per_90_overall	goals_per_90_overall	decimal		3	2	0			.	
34	33	goals_per_90_home	goals_per_90_home	decimal		3	2	0			.	
35	34	goals_per_90_away	goals_per_90_away	decimal		3	2	0			.	
36	35	min_per_goal_overall	min_per_goal_overall	smallint				0			.	
37	36	conceded_per_90_overall	conceded_per_90_overall	decimal		3	2	0			.	
38	37	min_per_conceded_overall	min_per_conceded_overall	smallint				0			.	
39	38	min_per_match	min_per_match	smallint				0			.	
40	39	min_per_card_overall	min_per_card_overall	smallint				0			.	
41	40	min_per_assist_overall	min_per_assist_overall	smallint				0			.	
42	41	cards_per_90_overall	cards_per_90_overall	decimal		3	2	0			.	
43	42	rank_in_league_top_attackers	rank_in_league_top_attackers	smallint				0			.	
44	43	rank_in_league_top_midfielders	rank_in_league_top_midfielders	smallint				0			.	
45	44	rank_in_league_top_defenders	rank_in_league_top_defenders	smallint				0			.	
46	45	rank_in_club_top_scorer	rank_in_club_top_scorer	smallint				0			LF	

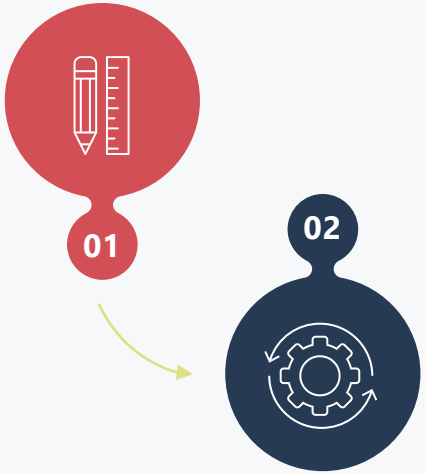
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Worksheet	FileNamePattern	FileExtension	HeaderRowsToSkip	ColumnNamesInFirstRow	IsUnicode	FlatFileType	TextQualifier	RowDelimiter	HeaderRowDelimiter	CodePage	RetainNulls	LoadingGroup	FileContentChronological	IsFullload	SortAttributeChronology
2	Demo_Teams	*-teams*.csv	.csv	0	1	0	Delimited		{LF}	Comma	65001	1		0	0	
3	Demo_Players	*-players*.csv	.csv	0	1	0	Delimited		{LF}	Comma	65001	1		0	0	
4	Demo_Matches	*-matches*.csv	.csv	0	1	0	Delimited	"	{LF}	Comma	65001	1		0	0	
5	Demo_League	*-league*.csv	.csv	0	1	0	Delimited		{LF}	Comma	65001	1		0	0	
6																

Beispiel 1

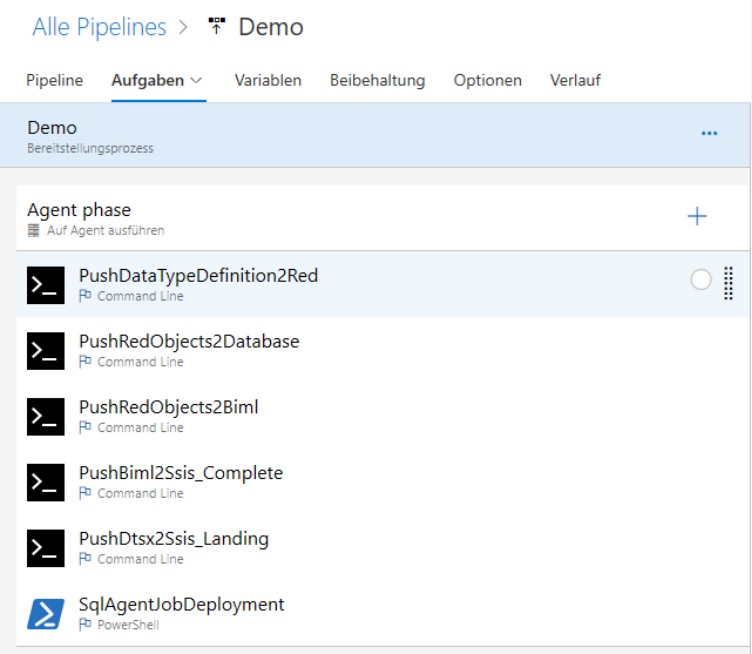
STANDARDISIERTE ENTWICKLUNG (MODEL DRIVEN)

SQL SERVER, WHERESCAPE, GIT, AZURE DEVOPS, BIML

INTERFACE DESIGN



BUILD & DEPLOY DEV



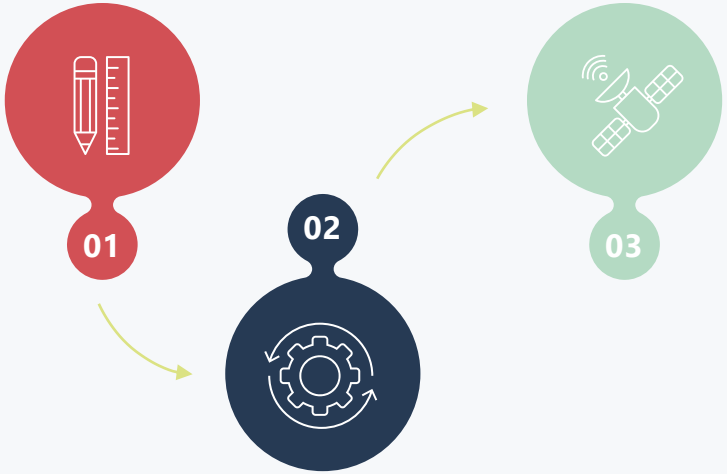
Beispiel 1

STANDARDISIERTE ENTWICKLUNG (MODEL DRIVEN)

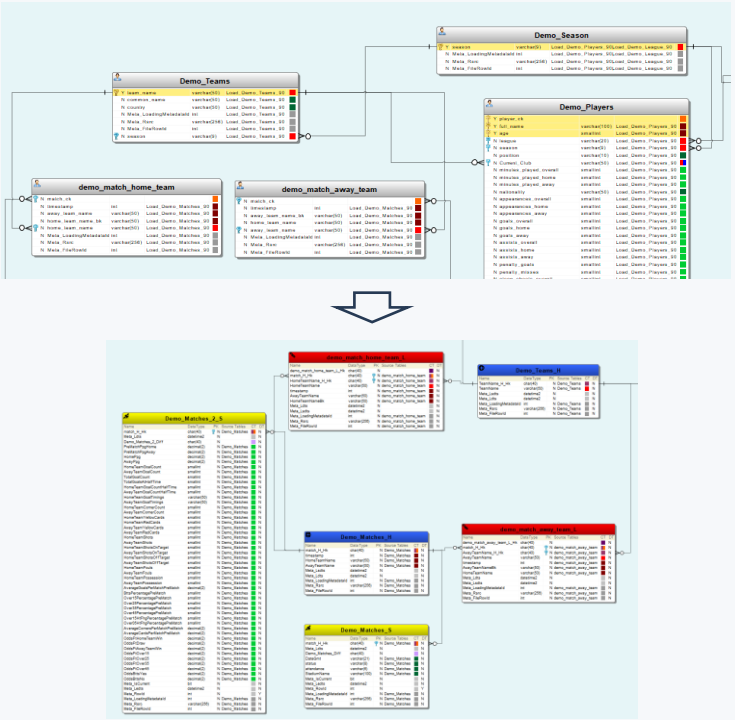
SQL SERVER, WHERESCAPE, GIT, AZURE DEVOPS, BIML

INTERFACE DESIGN

MODELLIERUNG



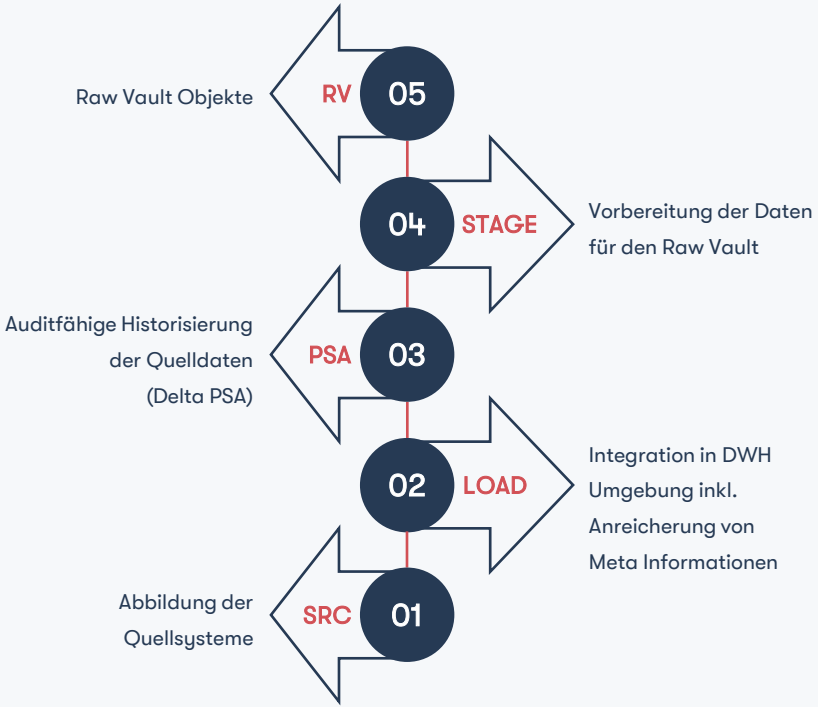
BUILD & DEPLOY DEV

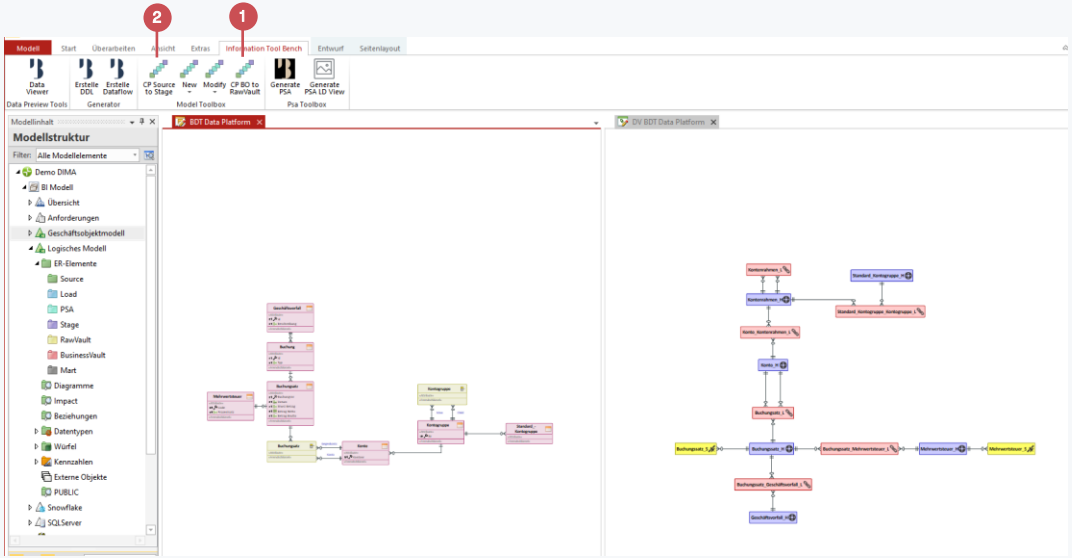


Beispiel 2

STANDARDISIERTE ENTWICKLUNG (MODEL DRIVEN)

SNOWFLAKE, MID INNOVATOR, C#, GIT, APACHE AIRFLOW

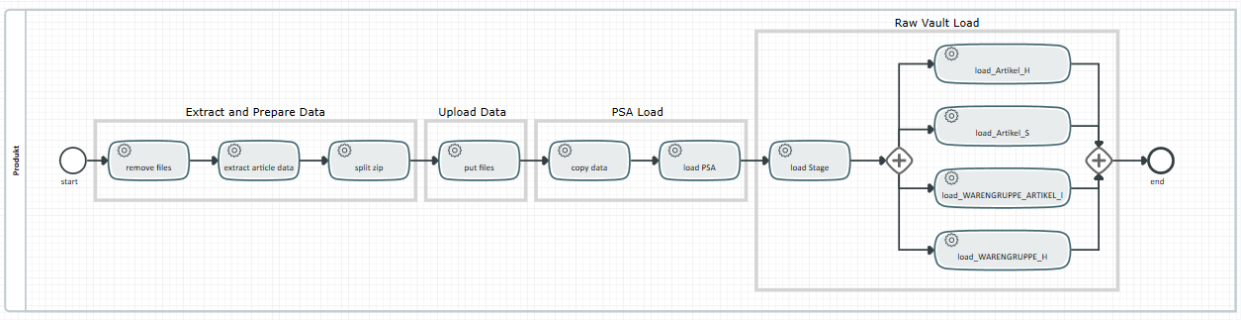


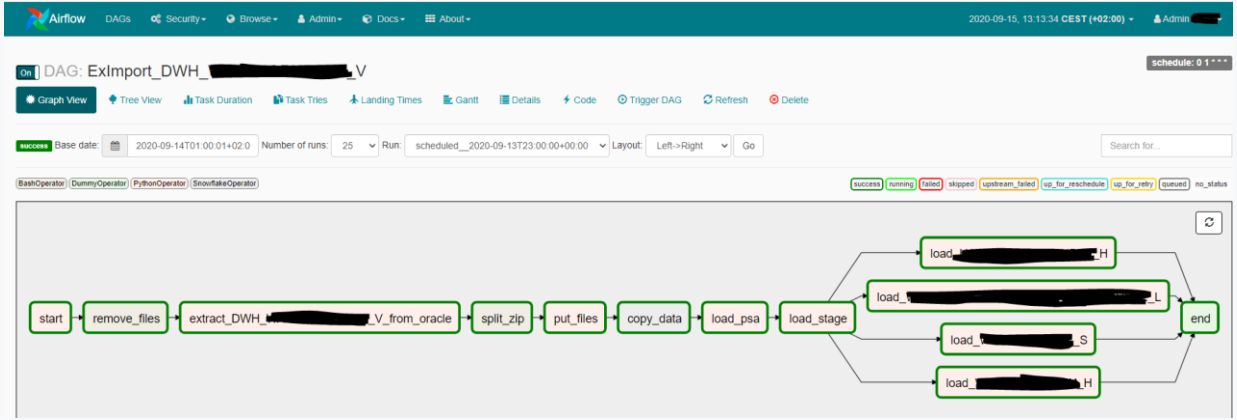


Beispiel 2

BESCHREIBUNG DER DATAFLOWS IM MODELL

SNOWFLAKE, MID INNOVATOR, C#, GIT, APACHE AIRFLOW





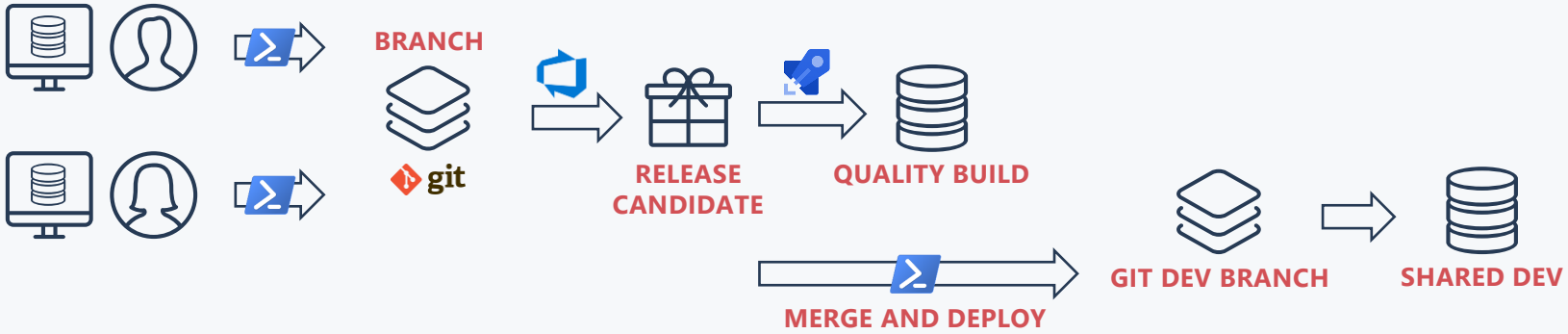
ANFORDERUNGEN

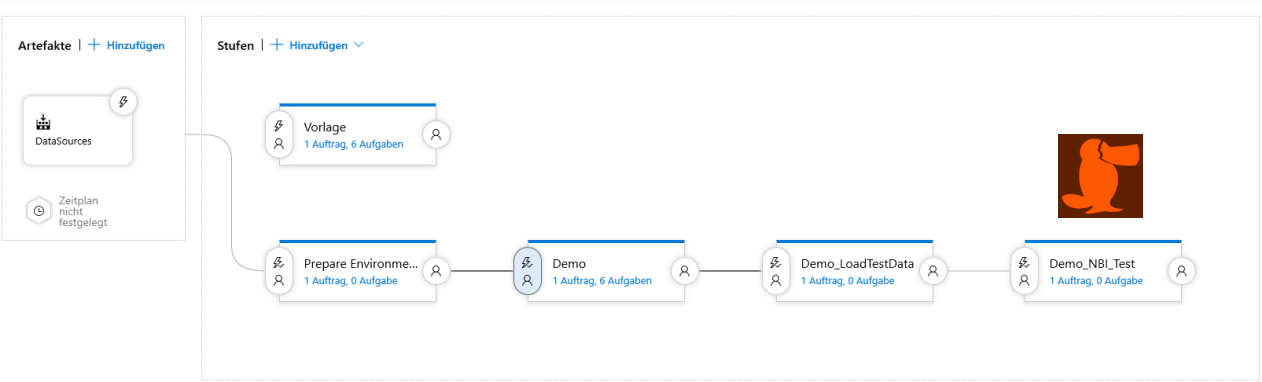
- Standardisierte Entwicklung
- Persönliche Entwicklungsumgebung
- Continous Integration & Quality
- Continous Delivery

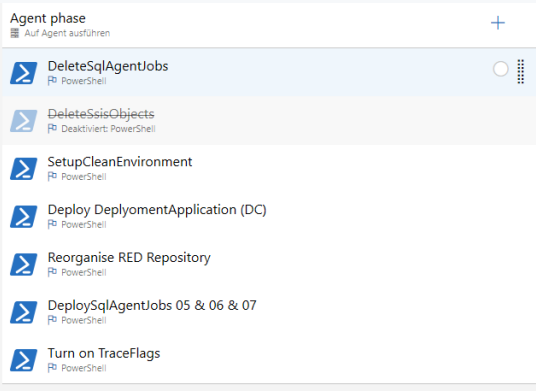
Beispiel 1

CONTINOUS INTEGRATION & QUALITY

SQL SERVER, WHERESCAPE, GIT, AZURE DEVOPS, BIML, POWERSHELL





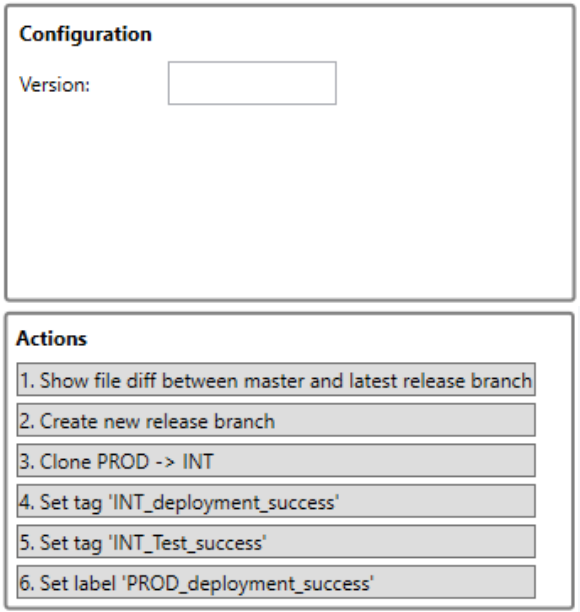


Beispiel 2

CONTINOUS INTEGRATION / DELIVERY

SNOWFLAKE, MID INNOVATOR, C#, GIT, APACHE AIRFLOW

1. Analyse der Änderungen –ggf. Delta Skripte erstellen
2. Neuen Branch erstellen mit den Änderungen
3. Neue Umgebung erstellen
   Deployment durchführen und testen
4. Strukturen laden und testen
5. Rollout auf Produktion



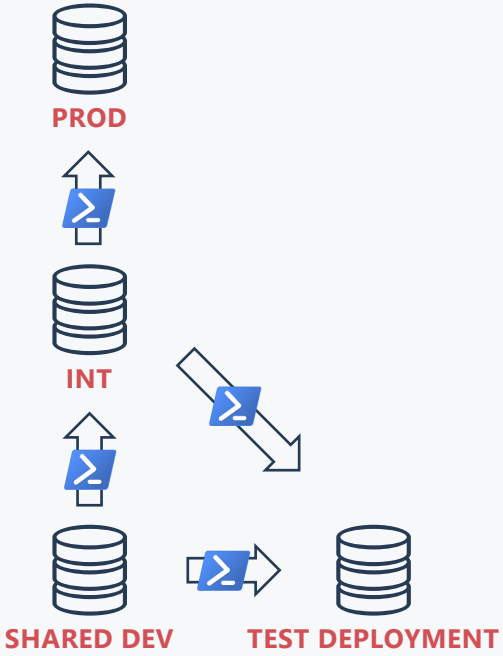
ANFORDERUNGEN

- ✔ Persönliche Entwicklungsumgebung
- ✔ Standardisierte Entwicklung
- ✔ Continous Integration & Quality
- ● Continous Delivery

Beispiel 1

CONTINOUS DELIVERY

SQL SERVER, WHERESCAPE, GIT, AZURE DEVOPS, POWERSHELL



1. Test Deployment gegen Clone der INT (optional)
2. Deployment gemäss Release Branch
3. Log prüfen
3. Lade- / Qualitätstest

ERFAHRUNGEN / FAZIT

ERFAHRUNGEN / FAZIT

1. Add Data and Logic tests
2. Use a Version Control System
3. Branch and Merge
4. Use Multiple Environments
5. Reuse & Containerize
6. Parameterize Your Processing
7. Work Without Fear™

ERFAHRUNGEN / FAZIT

- Entwicklungskompetenz
- Datenvolumen
- Regression Testing
- Compute <> Storage

GROW TOGETHER

Dominik Imark

+41 79 202 99 06

dominik.imark@banian.ch

@d_imark

linkedin.com/in/dominikimark

Banian AG
St. Jakobs-Strasse 3, CH-4052 Basel

banian.ch

+41 61 551 00 12

hey@banian.ch

@banianag

linkedin.com/company/banian-ag

Banian