



DATA VAULT, SELF SERVICE BI UND UNIFIED STAR SCHEMA

WILLKOMMEN

Petr Beles

Seit 2000 mit Daten unterwegs -
Hauptsächlich im Backend

Hat auch schon Reports entwickeln
müssen

Seit 2010 bei Datavault Builder

petr.beles@datavault-builder.com

Bitte kontaktiert mich auf LinkedIn

<https://datavault-builder.com/data-vault-blog/>

AGENDA

SELF SERVICE BI

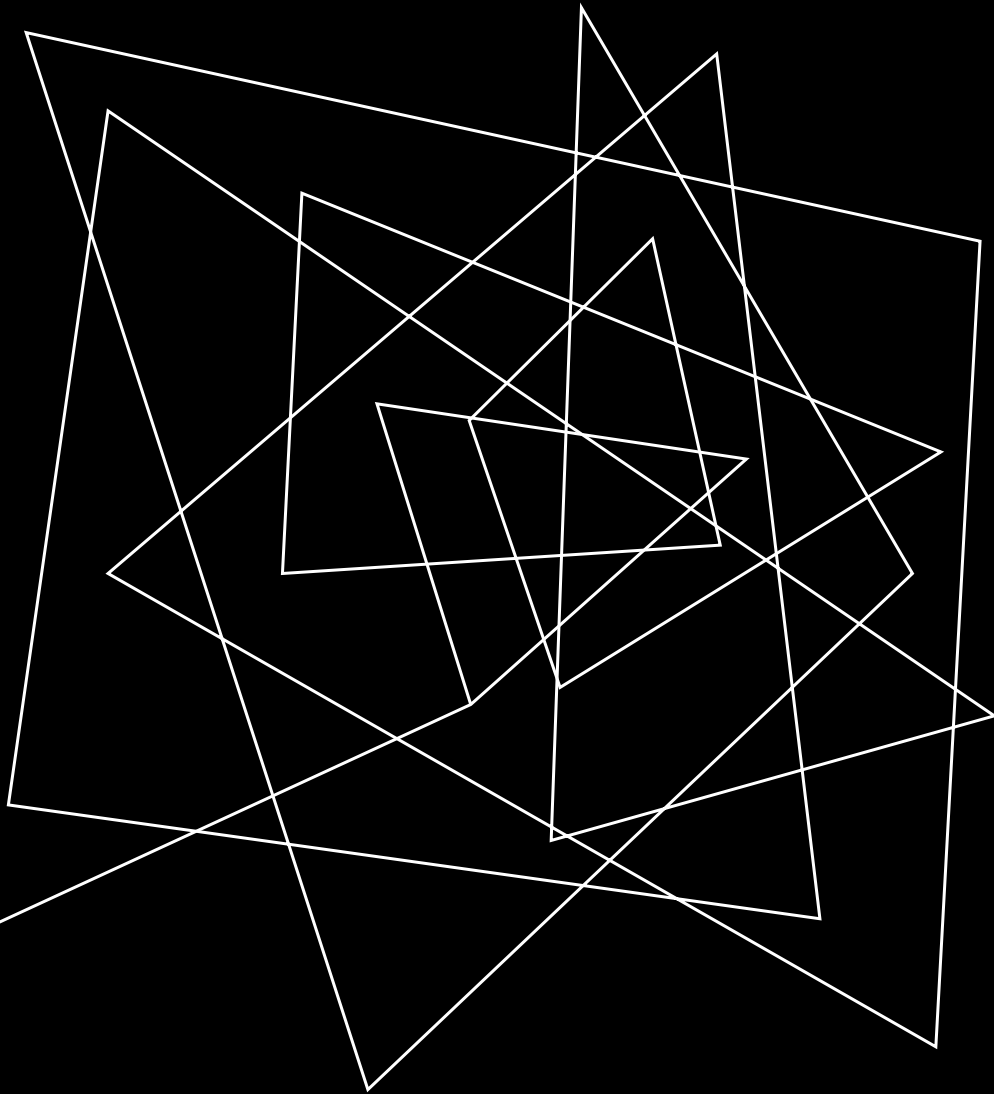
UNIVERSAL STAR SCHEMA (USS)

USS & DATA VAULT

BEISPIEL (DEMO)

ZUSAMMENFASSUNG &

DISKUSSION



WIESO SELF SERVICE BI?

WIR STEUERN PROZESSE / DECKEN RECHTLICHE VORGABEN AB

Daten Integration und Reporting ist kein Selbstzweck



WER STEUERT PROZESSE? FACHANWENDER

Fachanwender brauchen Zugang zu ihren Daten

Fachanwender wollen plötzlich neue Informationen
zusammenstellen

Fachanwender möchten /
können nicht auf die IT warten



DAS QLIKVIEW LADESKRIPT

Funktioniert sofort – aber wohl nicht für lange

Erzeugt falsche Daten / Informationen

Ist nicht wiederverwendbar





LÖSUNG 1

Integration durch Data Engineers

Zusammenstellung von Abfragen durch Fachanwender

Fachanwender müssen die IT Systeme verstehen





LÖSUNG 2

- Was braucht ihr?
- Alles!
- So was machen wir doch nicht!
- Aber wieso nicht?

Wir liefern ihnen einen Report, der alles enthält, was sie jemals brauchen werden.

IDEE

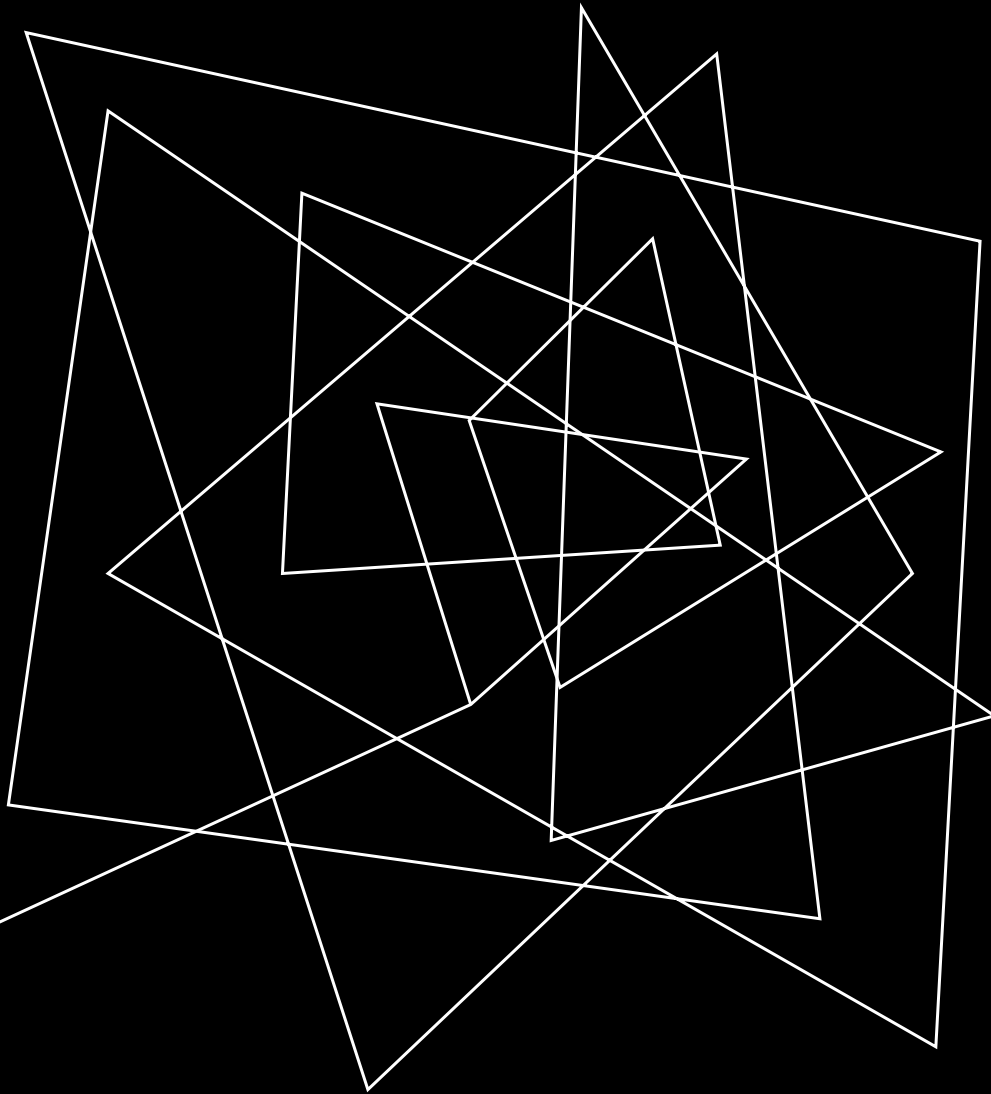
Der Fachanwender

(zumindest der fachnahe Report Entwickler)

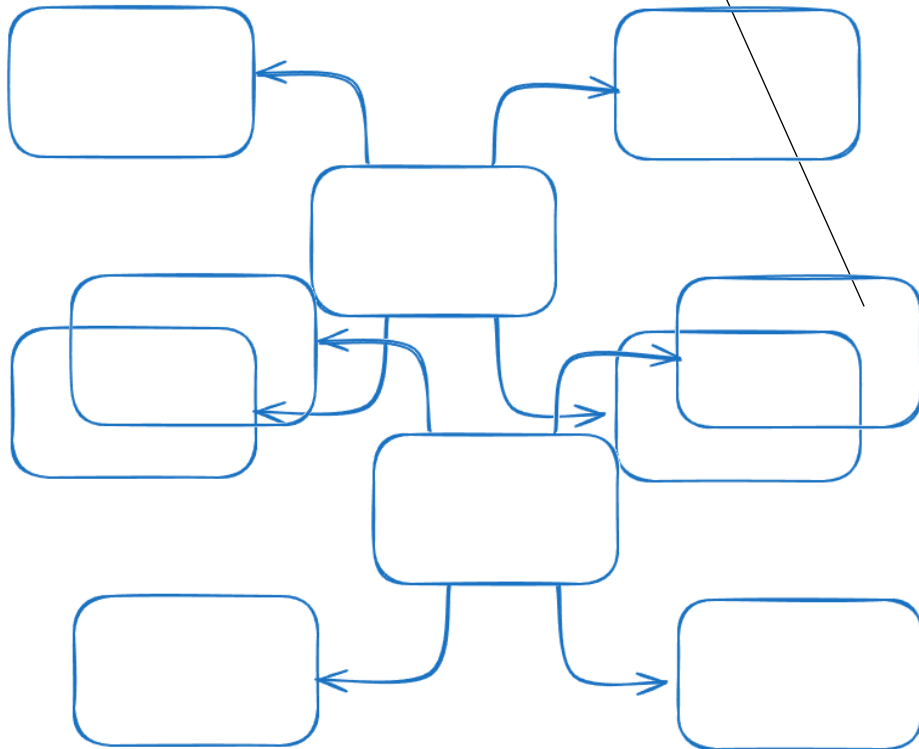
Kann in seinem Lieblingstool direkt die Daten auswerten, auch wenn er sie vorher nicht bestellt hat.

Das erreicht man durch ein universelles Präsentationsmodell:

das Universal Star Schema (USS)



UNIVERSAL STAR SCHEMA (USS)

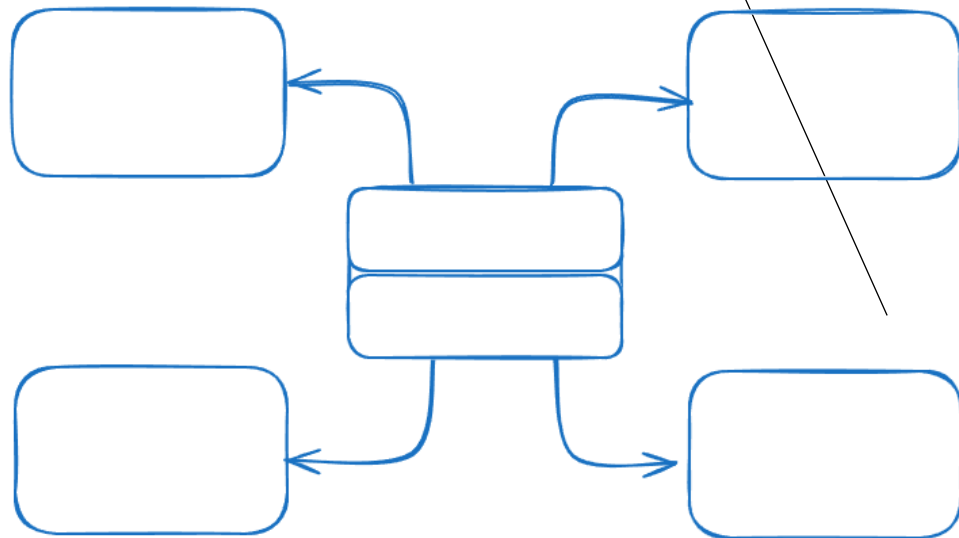


HISTORIE

Kimballs Dimensional Model hat eine Fakten-Tabelle und beliebig viele Dimensionen.

Kimball hat nie beschrieben, was mit mehrere Fakten-Tabellen geschieht.

Durch die Limitation in gewissen Tools, wie Qlikview, wurde ein Ansatz etabliert mit einer zentralen Link Tabelle, welche verschiedene Fakten aufnehmen kann.



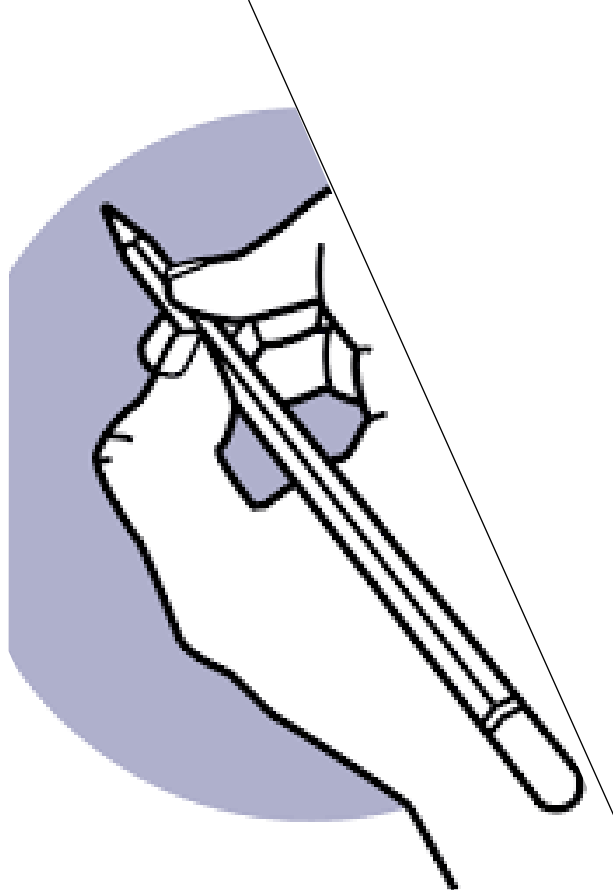
USS GRUNDIDEE

Wenn wir alle möglichen Prozesse bereits in einem Modell abbilden, müssen wir nachträglich nicht immer Änderungen machen, falls die Fachanwender etwas Neues möchten.

Es soll nicht 100% der Anforderungen abdecken, aber bereits 80% der Standardreports.

Es ist nicht Anforderungsgetrieben.

Aber auch nicht nur Datengetrieben



DIMENSIONEN BLEIBEN - FAST

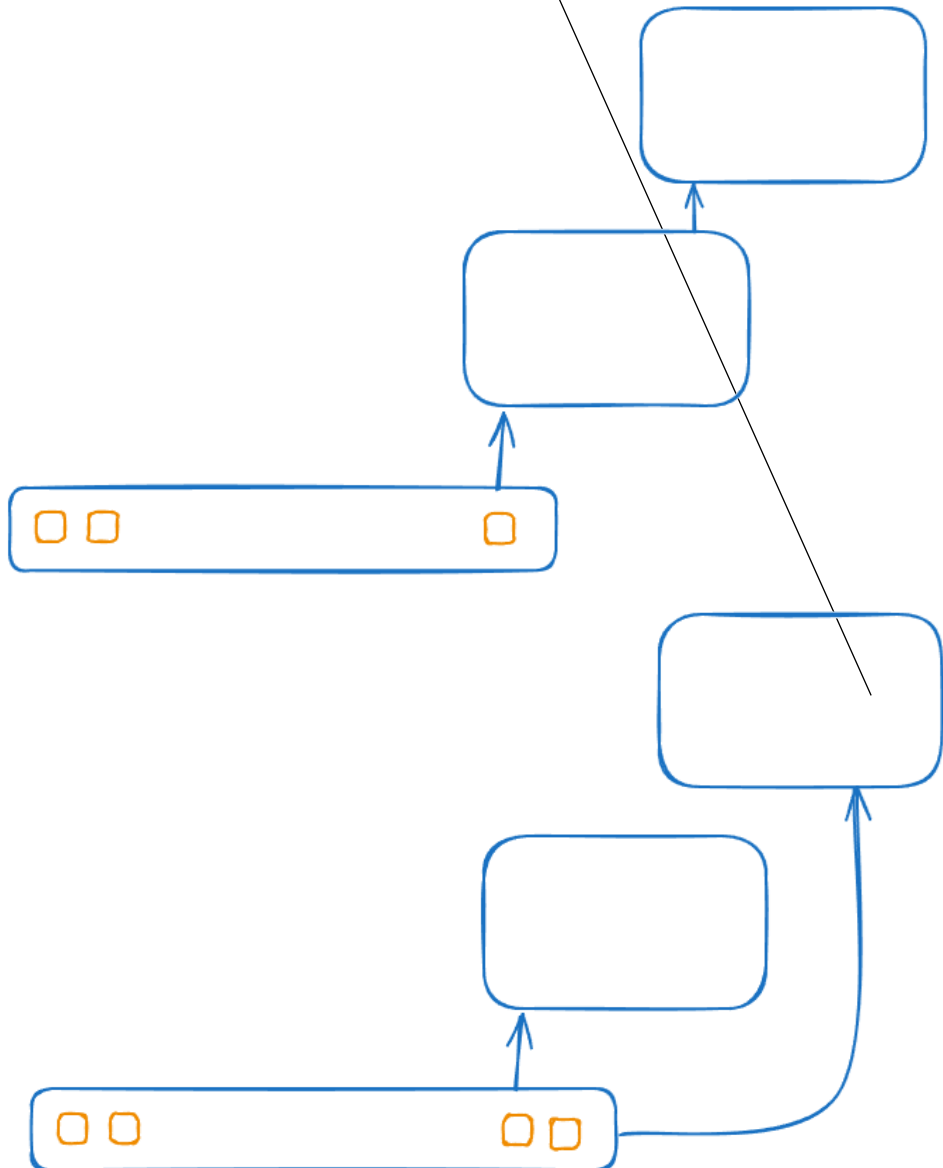
Fast alles, was ihr über Dimensionen wisst, könnt ihr weiterhin nutzen.

Solange es kein Schneeflocken-Modell ist.

Die Annahme ist: Schneeflocken
Gemeinde – Bundesland – Land ist bereits zu kompliziert.

Wieso denormalisieren wir nicht alles?

- Um richtig zu zählen
- Gewisse Daten haben evtl. nur Beziehung zum Land



FAKTEN

Fakten werden in einer Tabelle gesammelt.

Spalten in Fakten werden spärlich ausgefüllt, wenn nötig

Fakten haben direkte Beziehungen zu allen Dimensionen:

also direkt zu Gemeinde, Bundesland und Land)

!Datumsdimension!

UNION ALL

Alle Fakten werden vorbereitet in sogenannten "Stages"

Alle Fakten werden durch ein UNION ALL in eine zentrale Faktentabelle geschrieben. USS nennt diese Tabelle die Brücke (Bridge)



STAGES UND BRÜCKEN?

Wieso nicht einfach Fakten?

Weil Stages auch Sachen beschreiben, die man klassisch nicht als Fakten sieht:

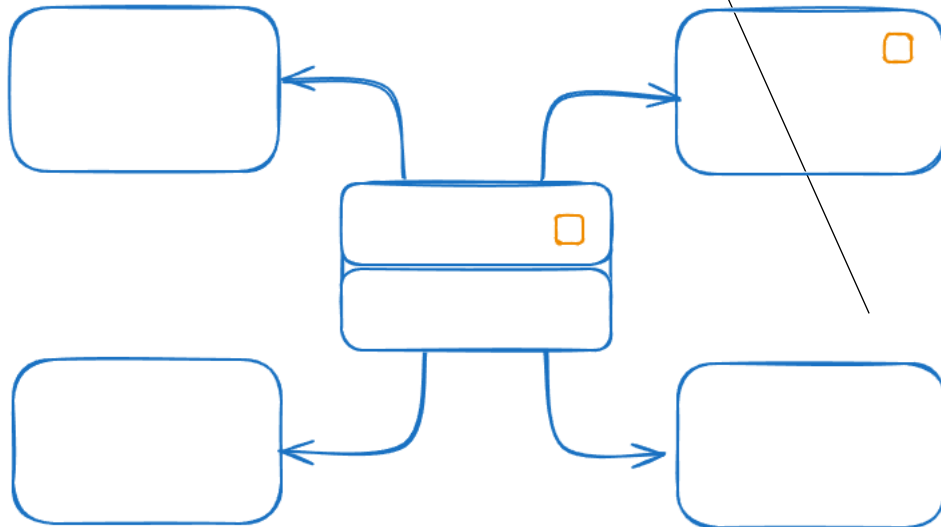
- Der Kunde
- Der Lieferdienst
- ...

Eine Stage ist ein Objekt, welches Beziehungen hat.

Bridges.....

Ein Objekt mit Beziehungen kann in einer Businessfrage zum Betrachtungsobjekt werden:

Wie viele Kunden haben wir in Deutschland, auch wenn sie nichts bestellt haben.



WOHIN MIT DEN MESSWERTEN?

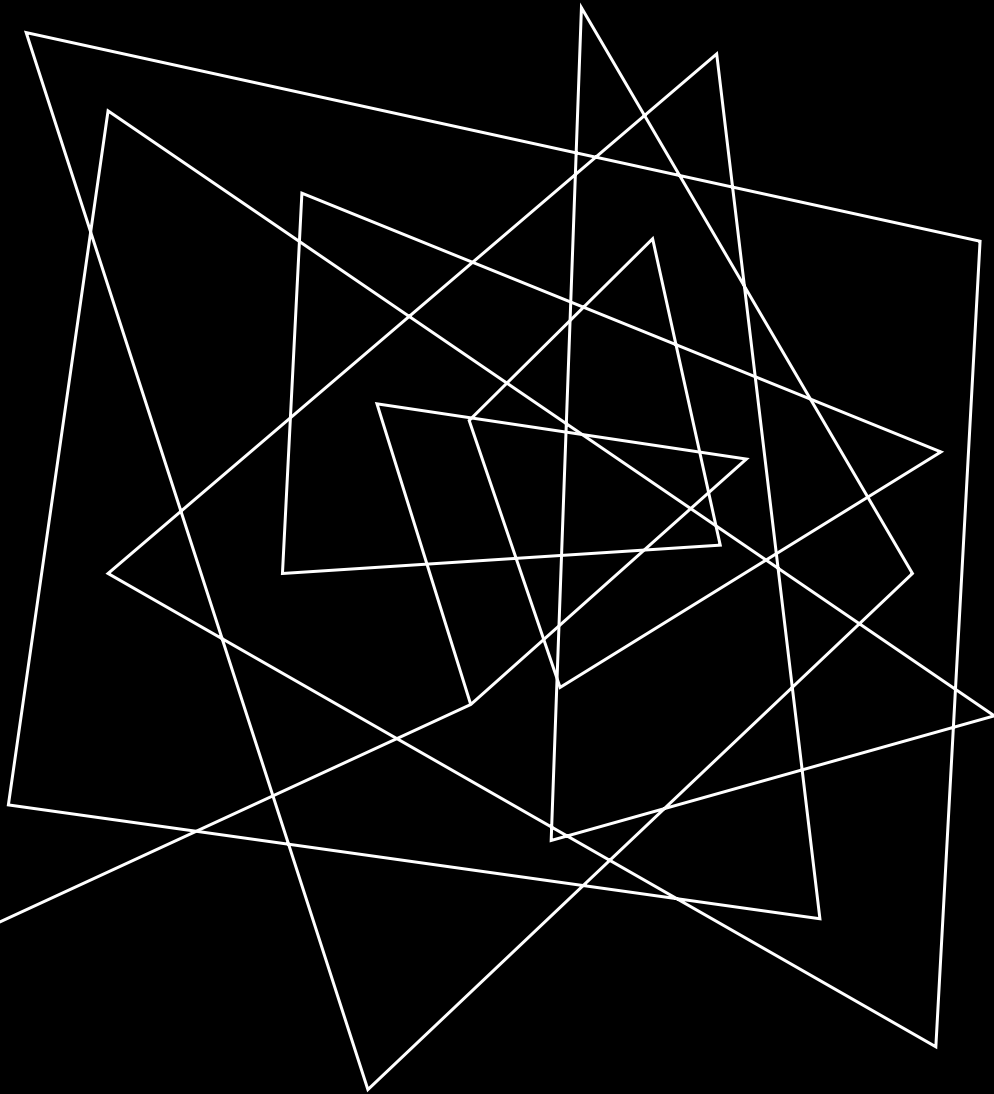
Das ist nicht so einfach.

Je nachdem, was für eine Frage wir beantworten an verschiedene Orte.

Beispiel:

Durchschnittsalter der Besteller eines Produktes

- Kevin ist 5 Jahre alt und hat 1000 Lollipops bestellt.
- David ist 60, hat keine Zähne mehr und hat einen Lollipop bestellt.



USS UND DATA VAULT

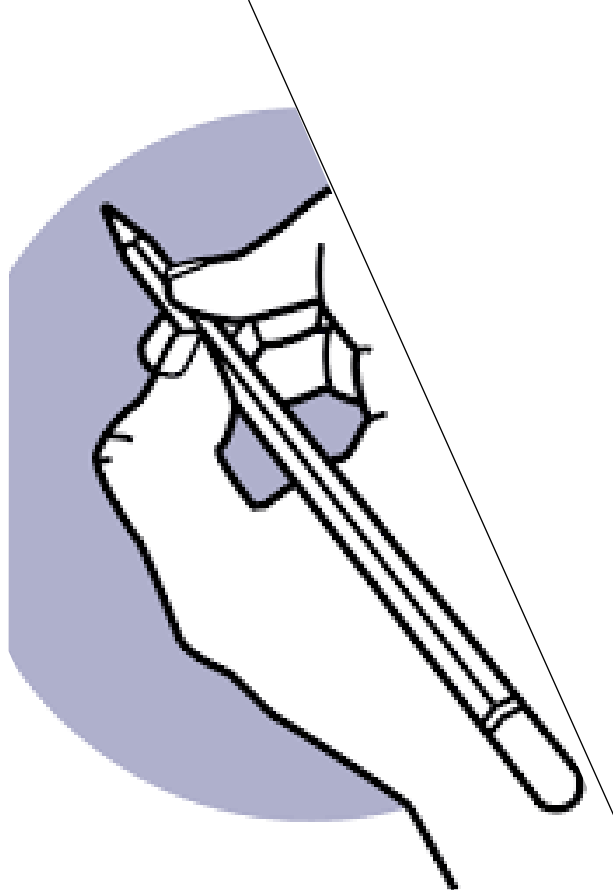
WIESO USS, WENN WIR DATA VAULT HABEN

Verschiedene Anforderungen:

- Erweiterbarkeit
- Wartbarkeit

Vs

- Verständlichkeit



WIESO DATA VAULT, WENN MAN USS WILL

Integration von verschiedenen Quellen

Historisierung

Refactoring

Auflösung der direkten Beziehungen der Fakten

ANPASSUNGEN IM DATA VAULT

Im Data Vault gibt es eigentlich keine Anpassungen.

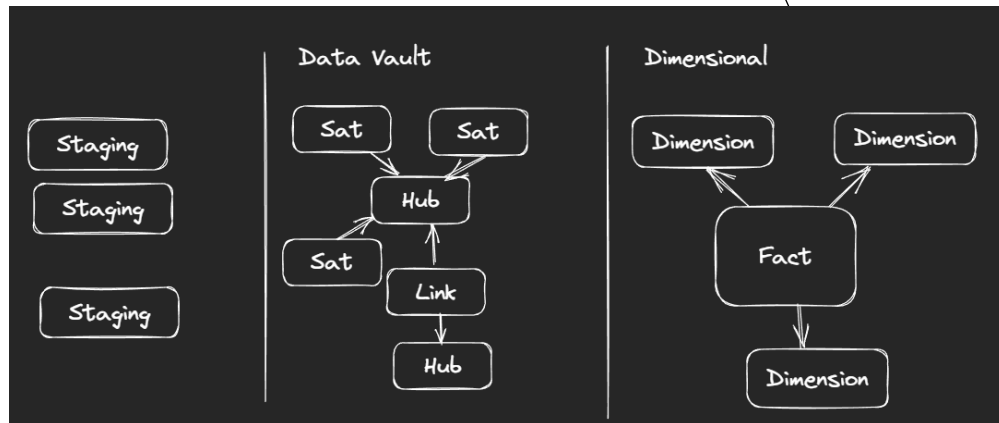
Eigentlich....

Also eine kleine...

Aber die hat nicht direkt was mit USS zu tun:

Modelliert doch bitte die erwarteten Kardinalitäten mit.

Also nur wenn ihr das Ganze auch automatisieren wollt.



DATA VAULT GRUNDANSATZ

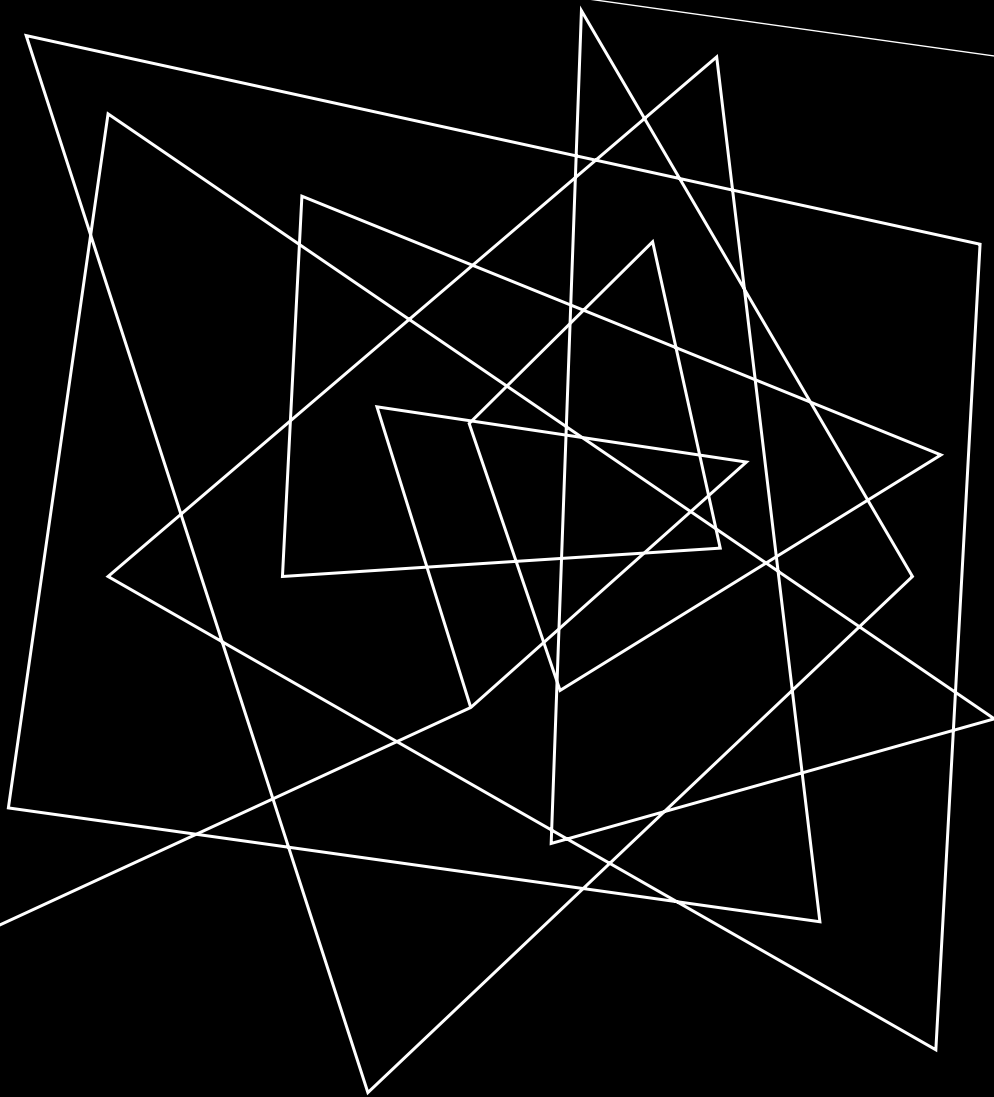
Trennung von Data Vault und Präsentation durch "Schnittstellen"

Die Schnittstelle fragt schon die richtige Zeitsicht ab

Das Schnittstelle denormalisiert wo nötig

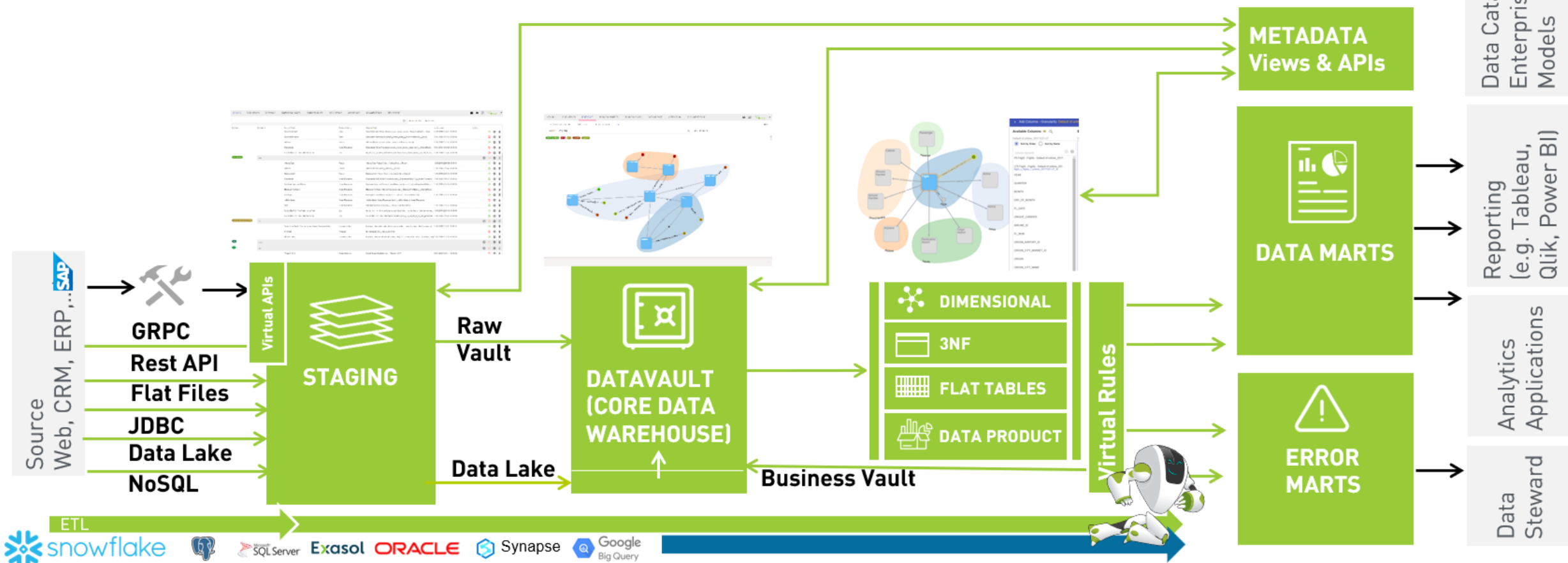
Das Schnittstelle harmonisiert Feldnamen

Stages und Dimensionen sind Schnittstelle auf dem Data Vault



BEISPIEL (DEMO)

FROM SOURCE TO DESTINATION



Connect to source systems & stage data

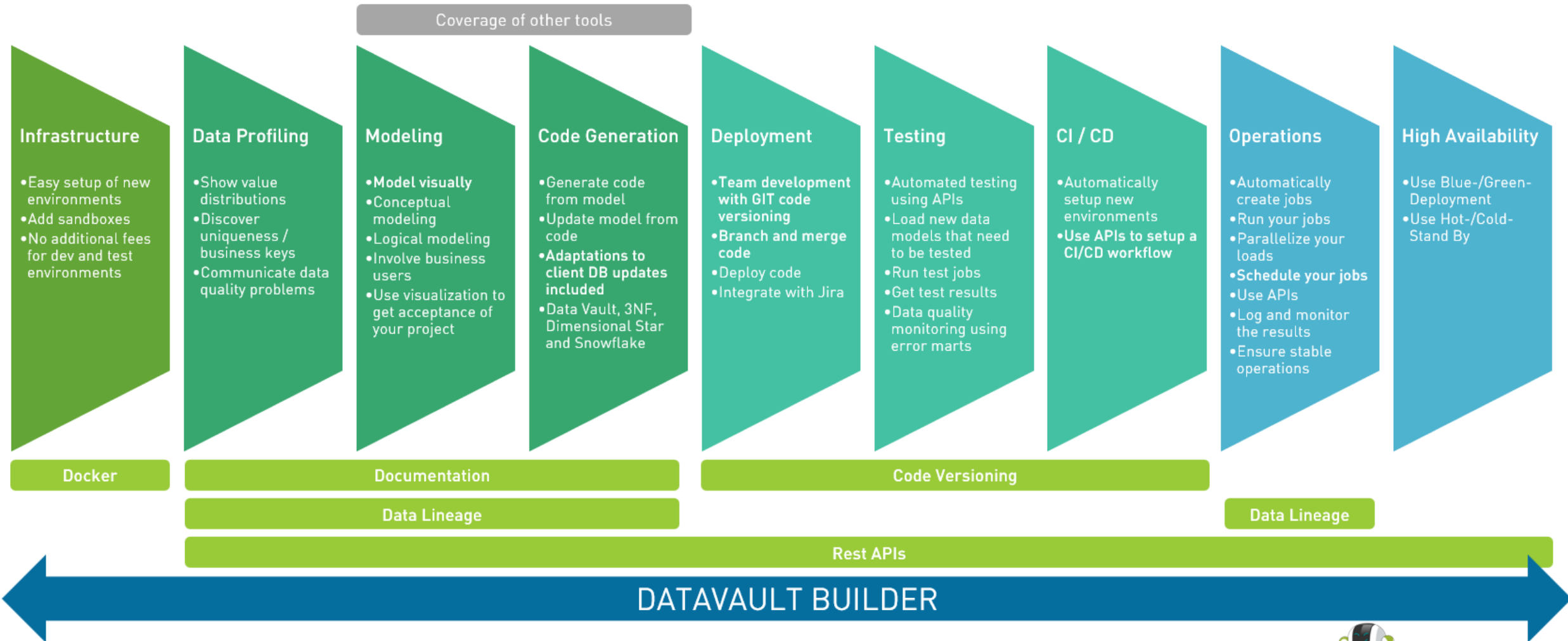
Historize and persist

Apply business logic

Deliver Data Marts, Data Products, AI & ML

Empower Reporting and other Applications

COVERING THE FULL DWH LIFECYCLE



STAGING BIS DATA VAULT

SCHNITTSTELLEN

Dimensionen und Stages



BRIDGE TABELLE

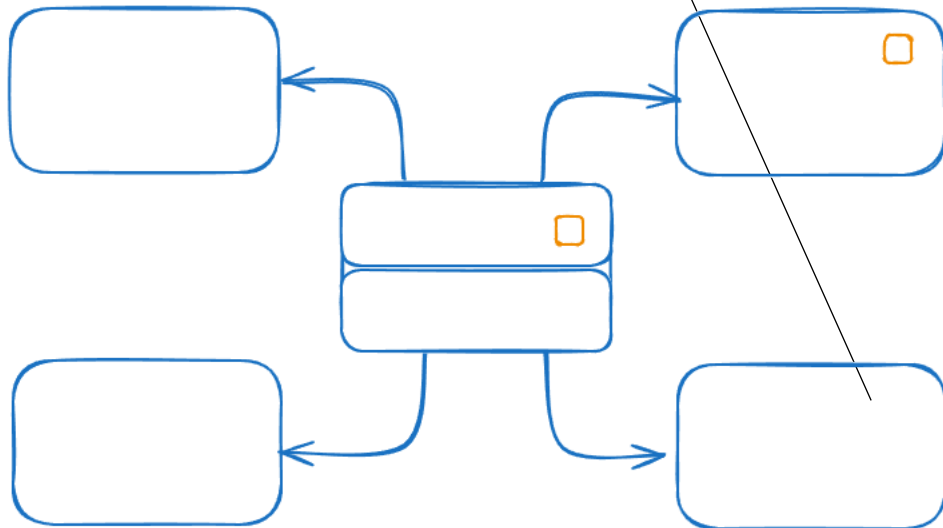
Automatisiertes UNION ALL

VIDEO

<https://datavault-builder.com/2023/11/20/automating-the-unified-star-schema-with-datavault-builder/>

BONUS

Wenn wir Zeit haben,
schauen wir das
Mühsame Thema der
Messwerte an



WO KOMMEN DIE KENNZAHLEN HIN?

In eine Stage?

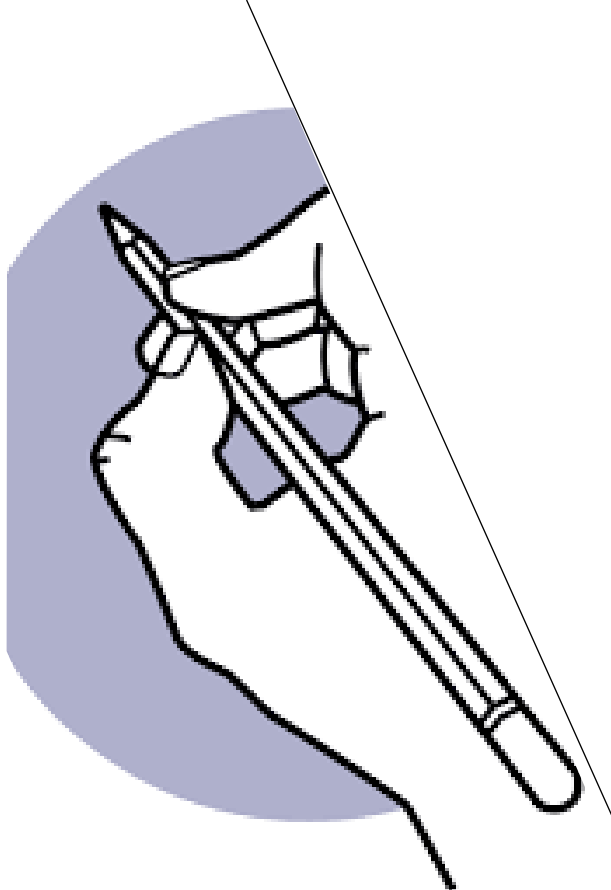
In eine Dimension?

In mehrere Stages?

Faktoren:

Was wollen wir wissen?

Welches Tool verwenden wir?

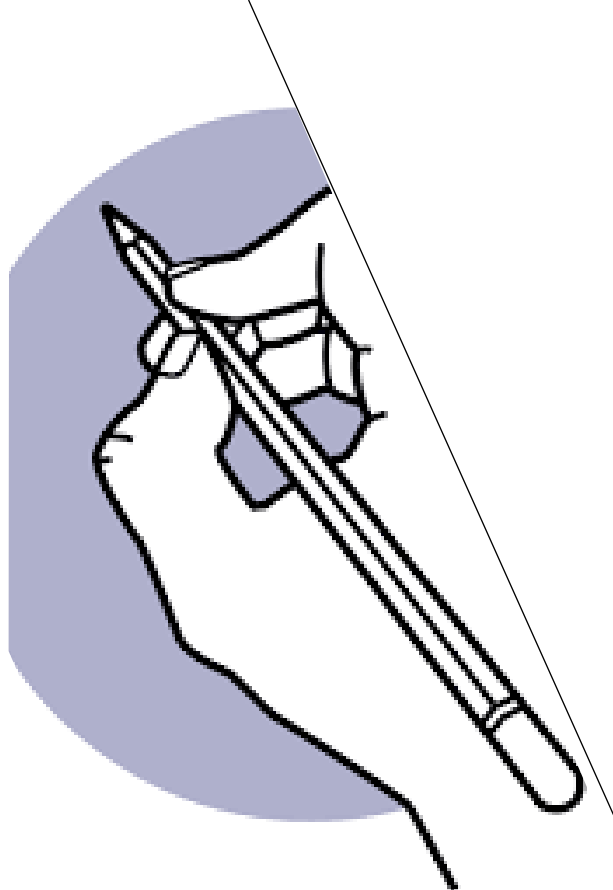


BEISPIEL VERSANDKOSTEN

Im Quellsystem auf dem Bestellkopf
(oder noch schlimmer auf einer Pseudo-Bestellposition)
Wie berechne ich den Deckungsbeitrag eines Produktes?

Ich kann die Versandkosten auf dem Bestellkopf belassen
In einem anderen Feld in der Stage erfasse ich die
verteilten Versandkosten

Der Fachanwender kann wählen, was er sehen möchte



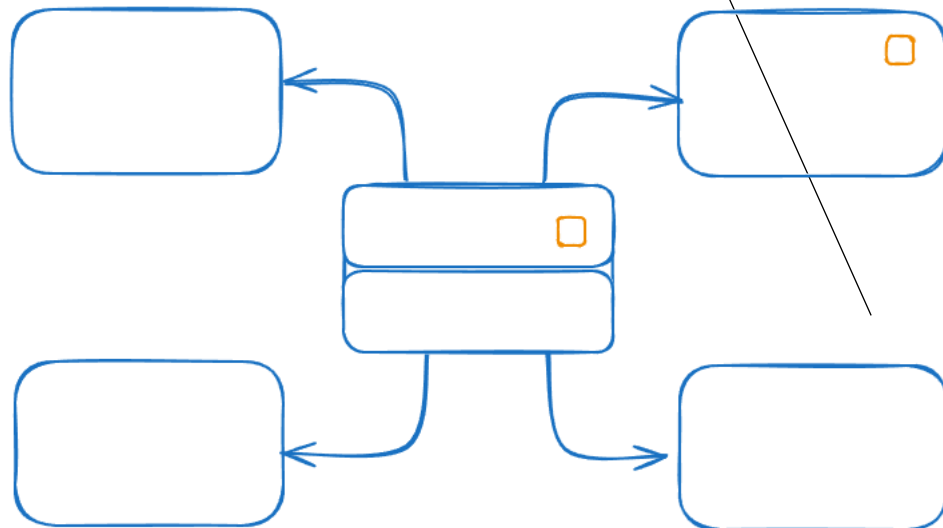
DURCHSCHNITTSALTER KUNDE

Kevin ist 5 Jahre alt und hat 1000 Lollipops bestellt.

David ist 60, hat keine Zähne mehr und hat einen Lollipop bestellt.

Petr, 45, hat keine Lollipops bestellt.

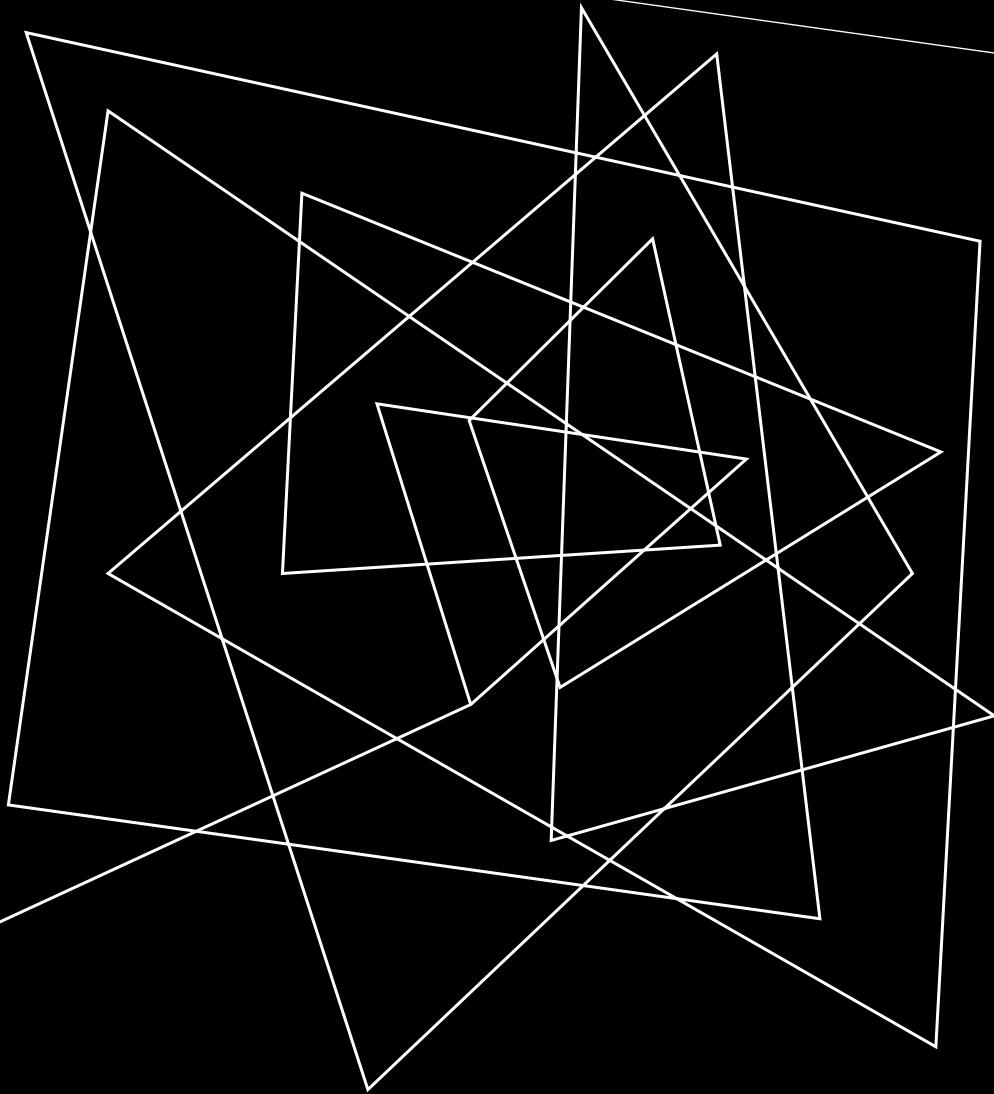
Ist das Durchschnittsalter 32.5, 35.6..., 5.039 Jahre?



KURZ ZUSAMMENGEFASST

Je nach Frage (und Tool), kann es sein, dass man eine Kennzahl denormalisiert auf ein feineres Korn, oder dort lässt, wo sie war....

Oder beides.



ZUSAMMENFASSUNG & DISKUSSION

ZUSAMMENFASSUNG

Ansatz

Data Vault und USS ergänzen sich

USS ist eine Evolution des Dimensionalen Modells wie Data Vault eine Evolution von 3NF ist.

Stages lösen die Beziehungen von einem Objekt auf direkt zu allen verbundenen Elementen

Stages und Dimensionen sind Interfaces auf dem Data Vault

Die Stages werden per UNION ALL kombiniert

Automatisierung

Data Vault sollte automatisiert sein

Interfaces sollten visuell zusammenstellbar sein

UNION ALL der Stages sollte automatisch generiert werden

Daten sollten in der Breite und nach Stages gefiltert werden können

FREUDE HERRSCHT!

Seid ihr jetzt alle glücklich?
und / oder gibt es noch Fragen und Einwände :)

THANK YOU

Petr Beles

petr.beles@datavault-builder.com

Bitte kontaktiert mich auf LinkedIn



SCAN ME