

Die Zukunft der Data Vault Automatisierung

Reinhard Mense

Deutschsprachige
Data Vault
User Group e.V.

Erfolgsmaßstab

$$\text{ROI} = \frac{\text{Gewinn}}{\text{eingesetztes Kapital}} \times 100$$


Nutzen
?
Investment

Der einzige Hebel





Mechanisierung

Das Ergebnis wird geprüft

Dieselbe Erfahrung ist nötig

Ein Mensch führt es aus



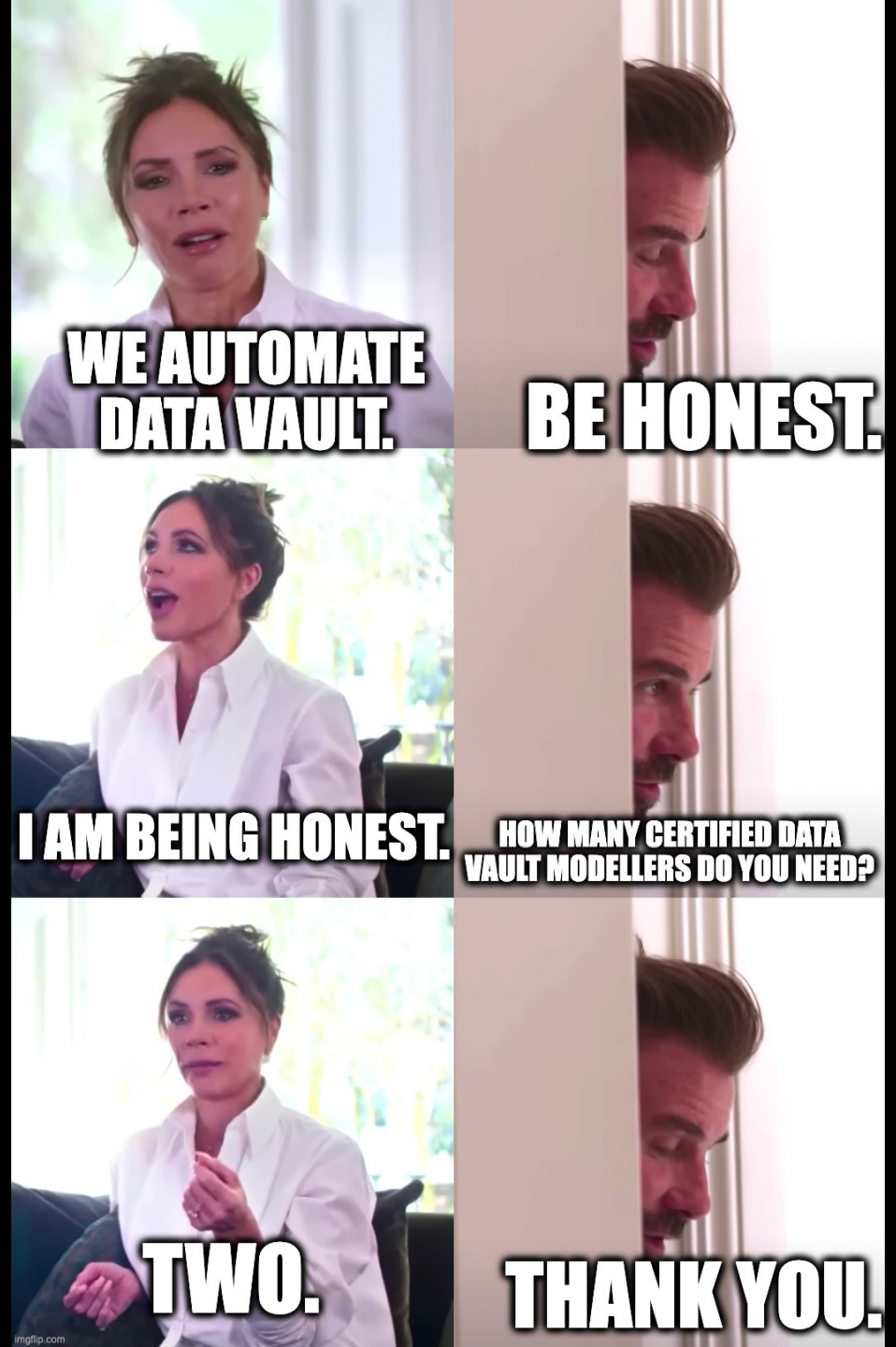
Automatisierung

Dem Ergebnis wird vertraut

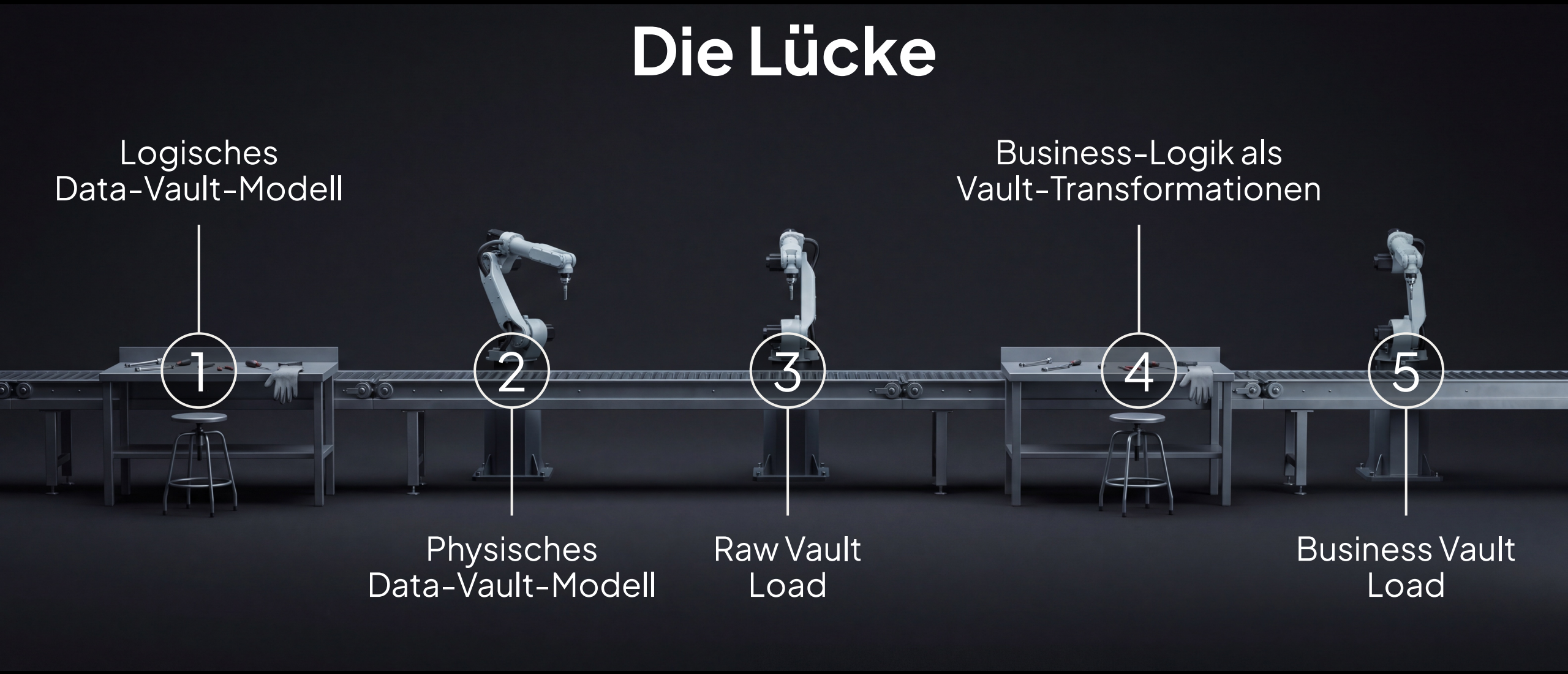
Die Erfahrung ist nicht mehr nötig

Es läuft von allein

KI generiert


WE AUTOMATE DATA VAULT.
BE HONEST.
I AM BEING HONEST.
HOW MANY CERTIFIED DATA VAULT MODELLERS DO YOU NEED?
TWO.
THANK YOU.
imgflip.com

Die Lücke


Logisches
Data-Vault-Modell
1
2
Physisches
Data-Vault-Modell
3
Raw Vault
Load
Business-Logik als
Vault-Transformationen
4
5
Business Vault
Load

Einfach die KI fragen?

Anthropic hat es ausprobiert

"How Anthropic enables self-service data analytics with Claude", claude.com/blog, Juni 2026

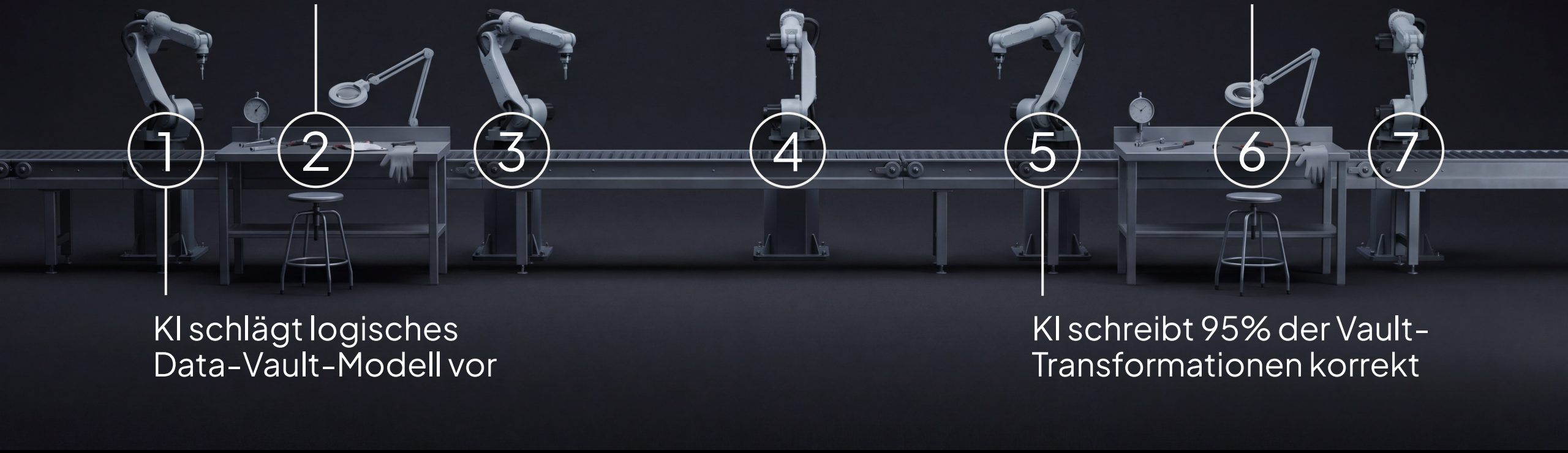
"Our bottleneck wasn't *access* to prior work, it was *structure*. "

"The fix is fewer, more heavily governed logical models: curate a small set of canonical, single source-of-truth datasets."

"Without skills, Claude's ability to answer analytics questions accurately didn't exceed 21% on our evals. Adding skills gets these numbers consistently above 95% in aggregate."*

* gemessen mit vorhandenem Data Warehouse, dimensionaler Modellierung, Tests, Metadaten und Semantic Layer

Schließt KI die Lücke?



Dieselbe Erfahrung ist nötig

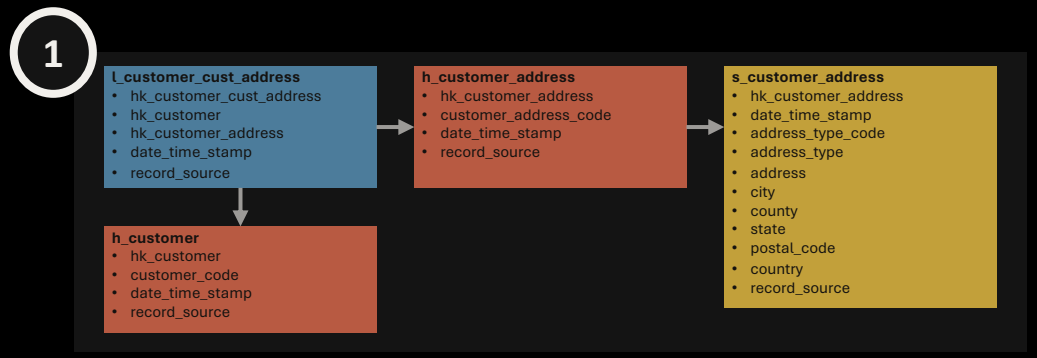
Das ist Mechanisierung.

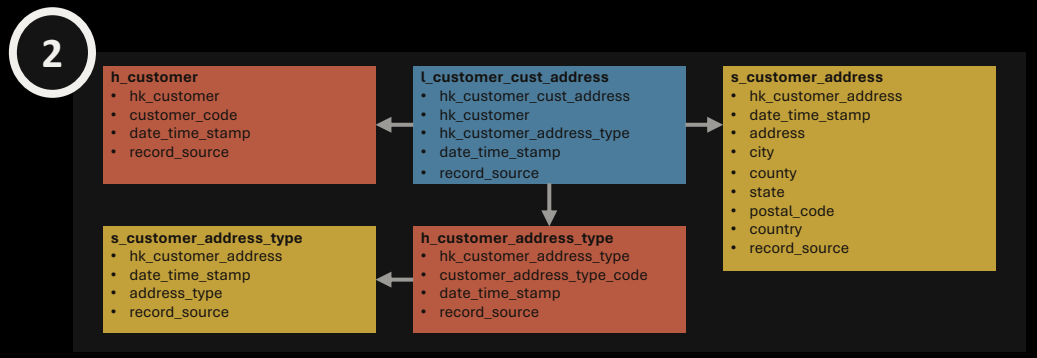
Bei einer Datenbank denken wir in Tabellen.
Wo ihre Zeilen physisch liegen, fragt niemand mehr.

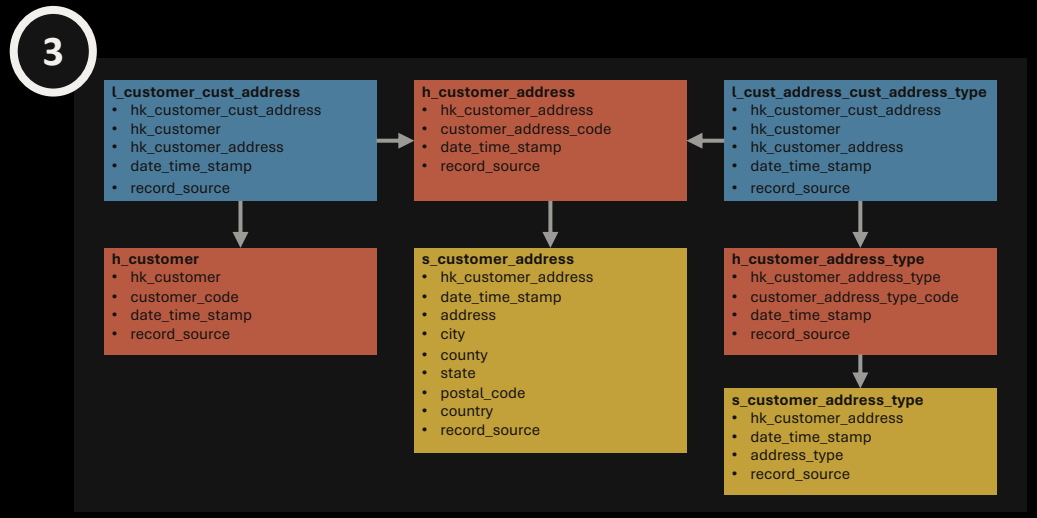
Beim Data Warehouse denken wir in Business Concepts.
In welchen Tabellen sie liegen, müssen wir jedes Mal wissen.

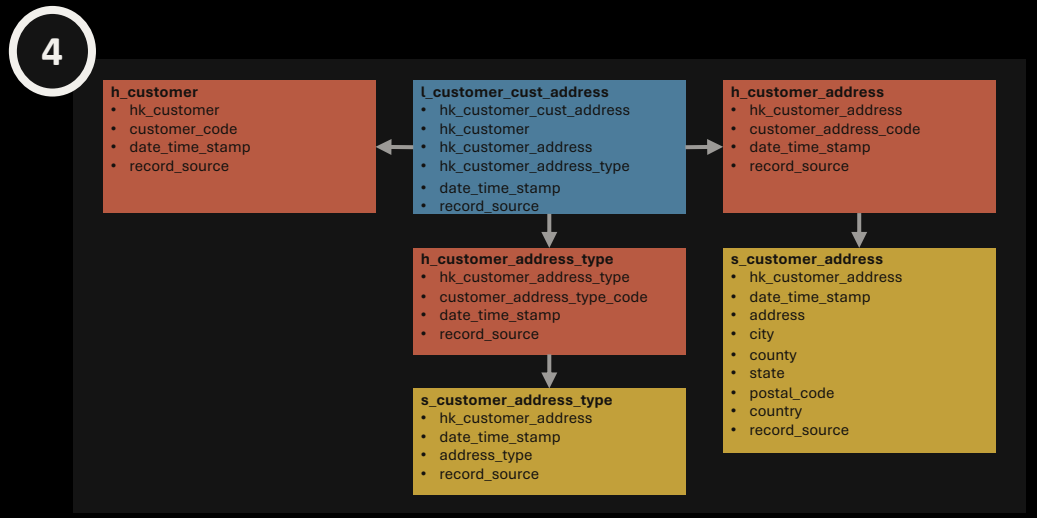


Dieselbe Anforderung, vier Modelle







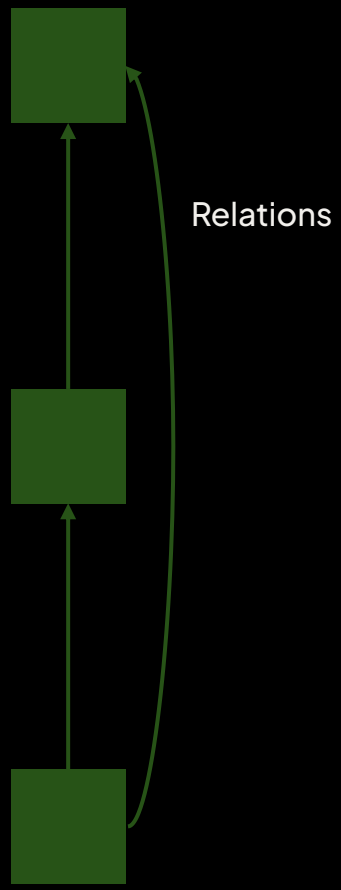


Wer mit dem Concept arbeitet muss jedes Mal ins Modell schauen

Eine Ebene höher beschreiben.
Data Vault wird abgeleitet.

Eine Ebene höher beschreiben.

Conceptual Layer




```yaml
# concepts/product/product.yaml (excerpt)
kind: concept
name: product

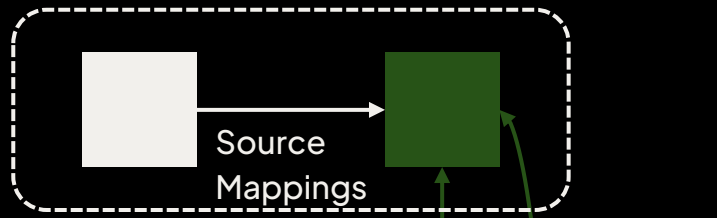
attributes:
  - name: product_number
    type: varchar(20)
  - name: list_price
    type: decimal(10, 2)
  - name: category_id
    type: integer

business_key:
  - product_number

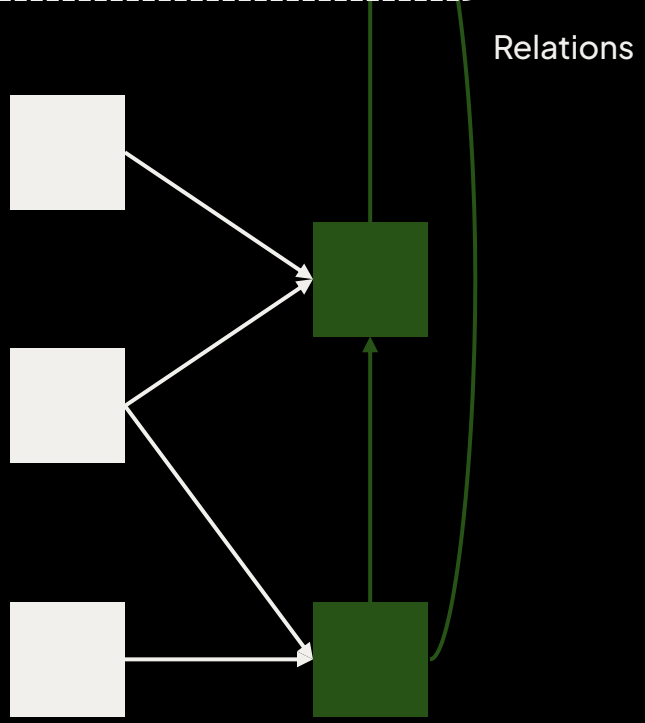
relations:
  - name: product_category
    parent: category
    foreign_key: [category_id]
    parent_key: [category_id]
```

Eine Ebene höher beschreiben.

Conceptual Layer ➡ Logical Data Vault Model



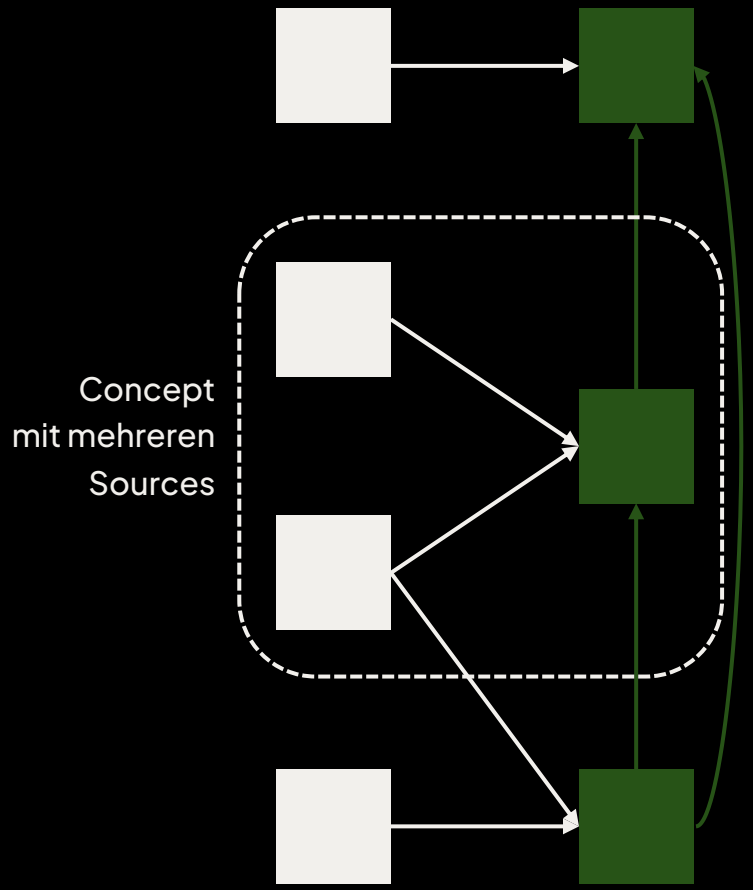


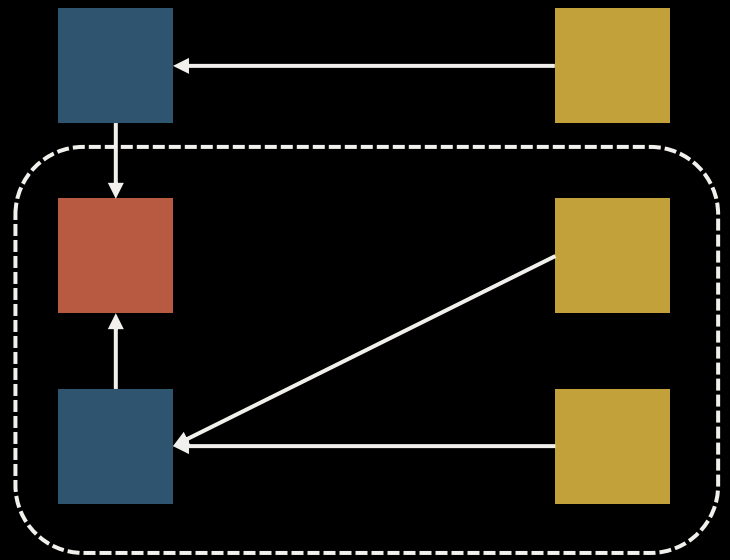


Source Concept Hub Satellite

Eine Ebene höher beschreiben.

Conceptual Layer ➡ Logical Data Vault Model

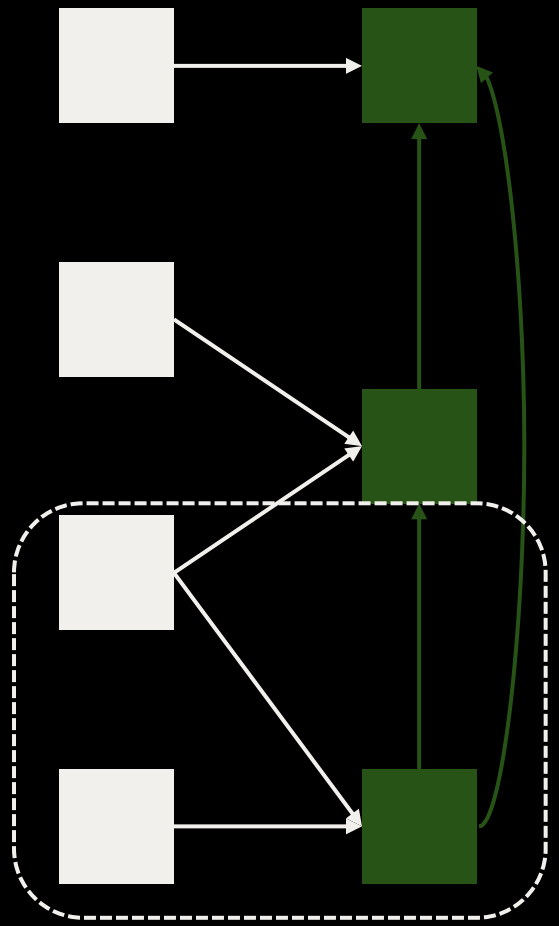




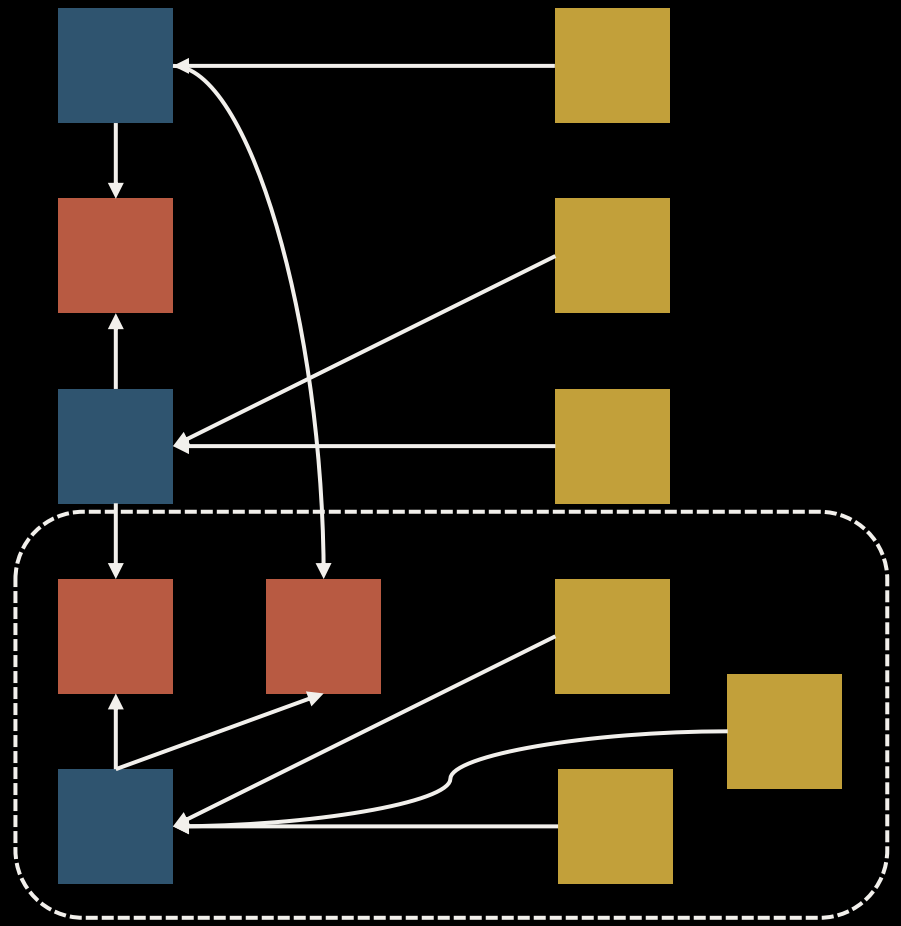
Source Concept Hub Satellite Link

Eine Ebene höher beschreiben.

Conceptual Layer ➡ Logical Data Vault Model



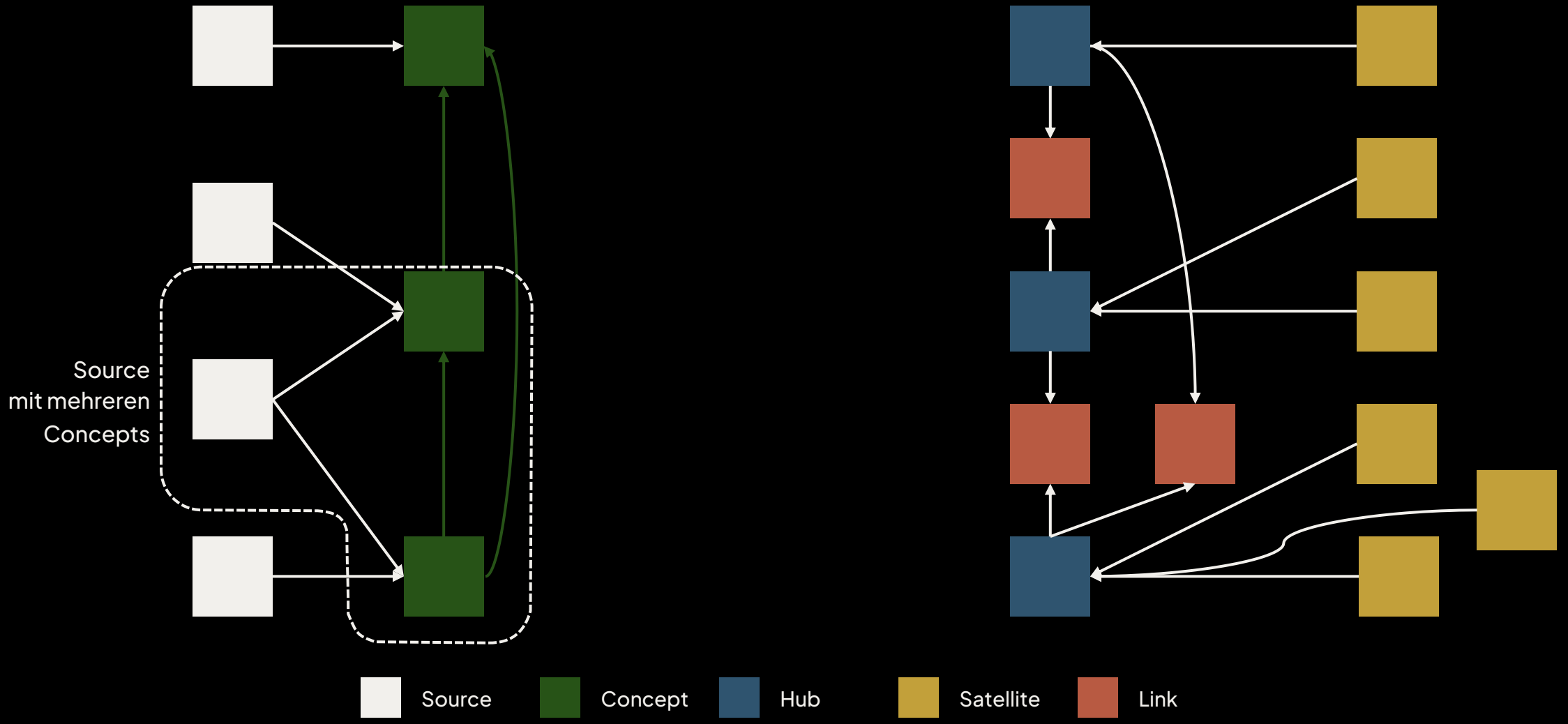
Concept
mit mehreren
Sources
und Attribut-
überschneidungen



Source Concept Hub Satellite Link

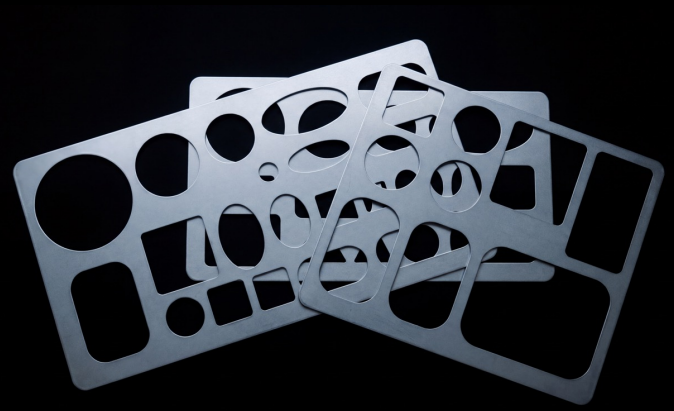
Eine Ebene höher beschreiben.

Conceptual Layer ➡ Logical Data Vault Model



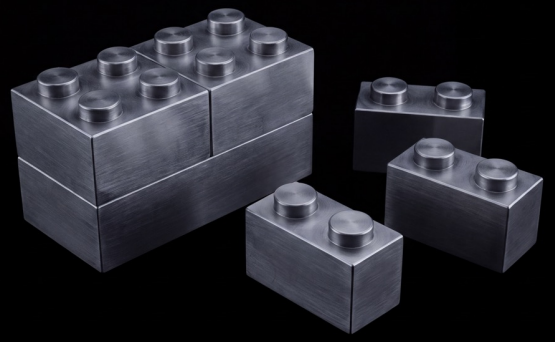
Weitere Muster im Conceptual Layer

Alternate Keys
Self References
Circular References
Business Timelines
Calculated Attributes
...



Weitere Bausteine für das Logical Data Vault Model

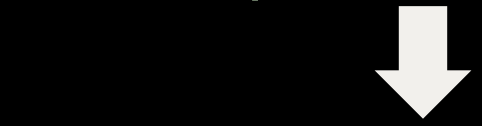
Same-as-Link
Effectivity Satellite
Multi-Active-Satellite
Point-in-Time Table
...



Eine Ebene höher beschreiben.

Calculated Attributes

```
1  SELECT <calculation with attributes>
2    FROM concept
```



```
1  SELECT <calculation with columns>
2    FROM pit_concept AS p
3    JOIN hub_concept AS h
4      ON p.hk = h.hk
5    JOIN sat_concept__1 AS s1
6      ON p.hk = s1.hk
7     AND p.s1_ldts = s1.ldts
8    JOIN sat_concept__2 AS s2
9      ON p.hk = s2.hk
10    AND p.s2_ldts = s2.ldts
```

Eine Ebene höher beschreiben.

Calculated Attributes mit Joins

```
SELECT <calculation with attributes>
  FROM concept_a
  JOIN concept_b
 USING (relation_a_b)
```

Calculated Attributes mit Joins und Timeline

```
SELECT <calculation with attributes>
  FROM concept_a
  JOIN concept_b
 USING (relation_a_b, business_timeline)
```

Änderungen sind auch nur Muster

Erweiterung

neues Concept · neues Attribut · neue Beziehung

Der Regelfall

Das Modell wächst

Die Daten bleiben

Umbau

der Business Key ändert sich

Der seltenere Fall

Ein Teil wird neu abgeleitet

Dieser Teil wird aus der PSA neu geladen

materialisiert · schnellere Abfragen
virtualisiert · weniger Storage, schnellere Beladung

1. Standard — es greift eine Voreinstellung
2. Vorgabe — über Concepts und Attribute
3. Statistik — nach Nutzung und Datenmengen

Materialisierung ändert das Modell nicht

Automatisieren heißt die Lücke schließen.
Nicht beschleunigen.

Data Vault wird vollständig abgeleitet.
Deterministisch.

KI hilft eine Ebene höher.

Womit verbringen wir die meiste Zeit?

conflux9

Ein Projekt, in dem ich an diesen Themen arbeite

conflux9.com — Blog und Newsletter

Reinhard Mense