

m[method] 2 data

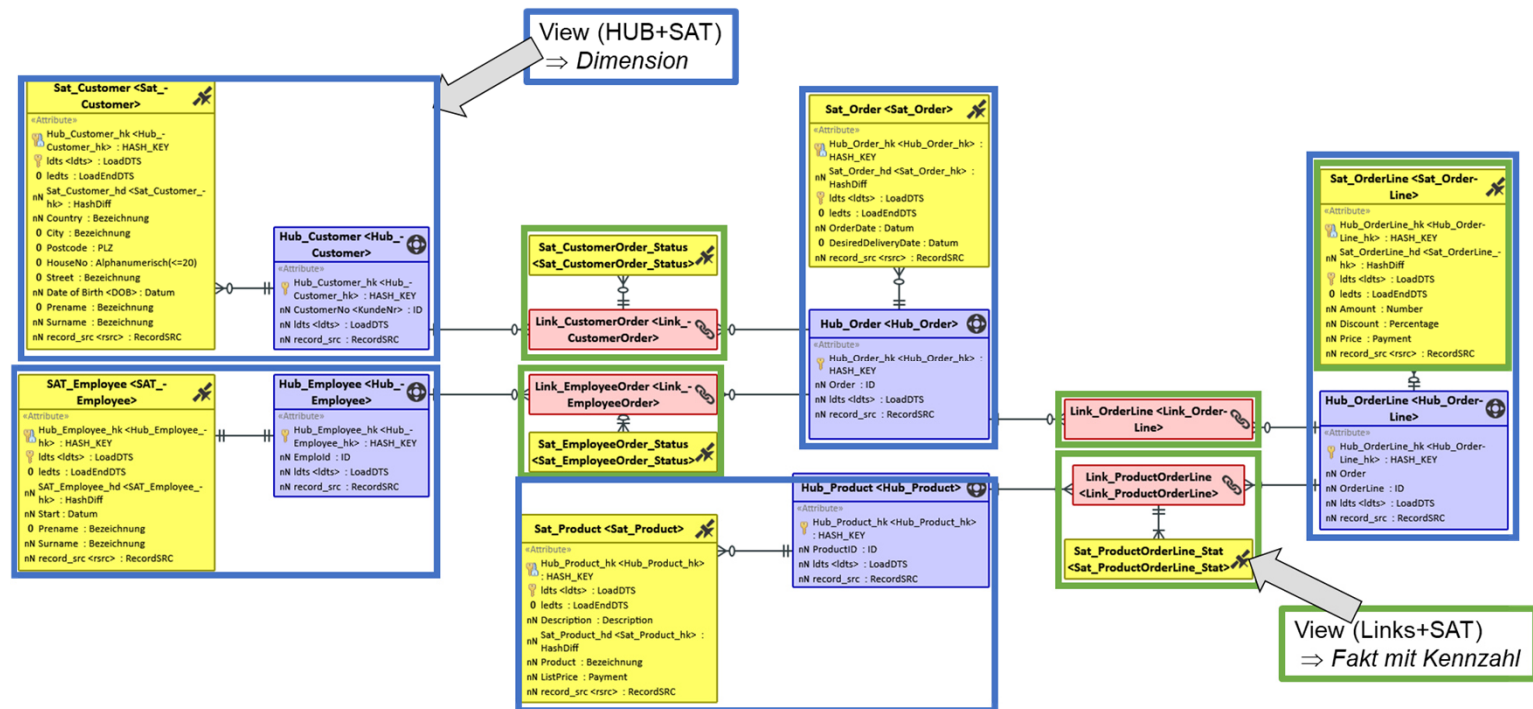
Und wenn wir doch mal auch  
den Data Mart automatisieren?

18. Tagung der DDVUG, 16. April 2026, Berlin

# Data Vault und Dimensional

*I have written extensively on the steps required to administer conformed dimensions and conformed facts. I have never seen a comparable set of specific guidelines for the normalized EDW approach.*

Kimball, Ross, "The Kimball Group Reader", Wiley 2015, S. 200



## Was bedeutet 'specific guidelines' ?

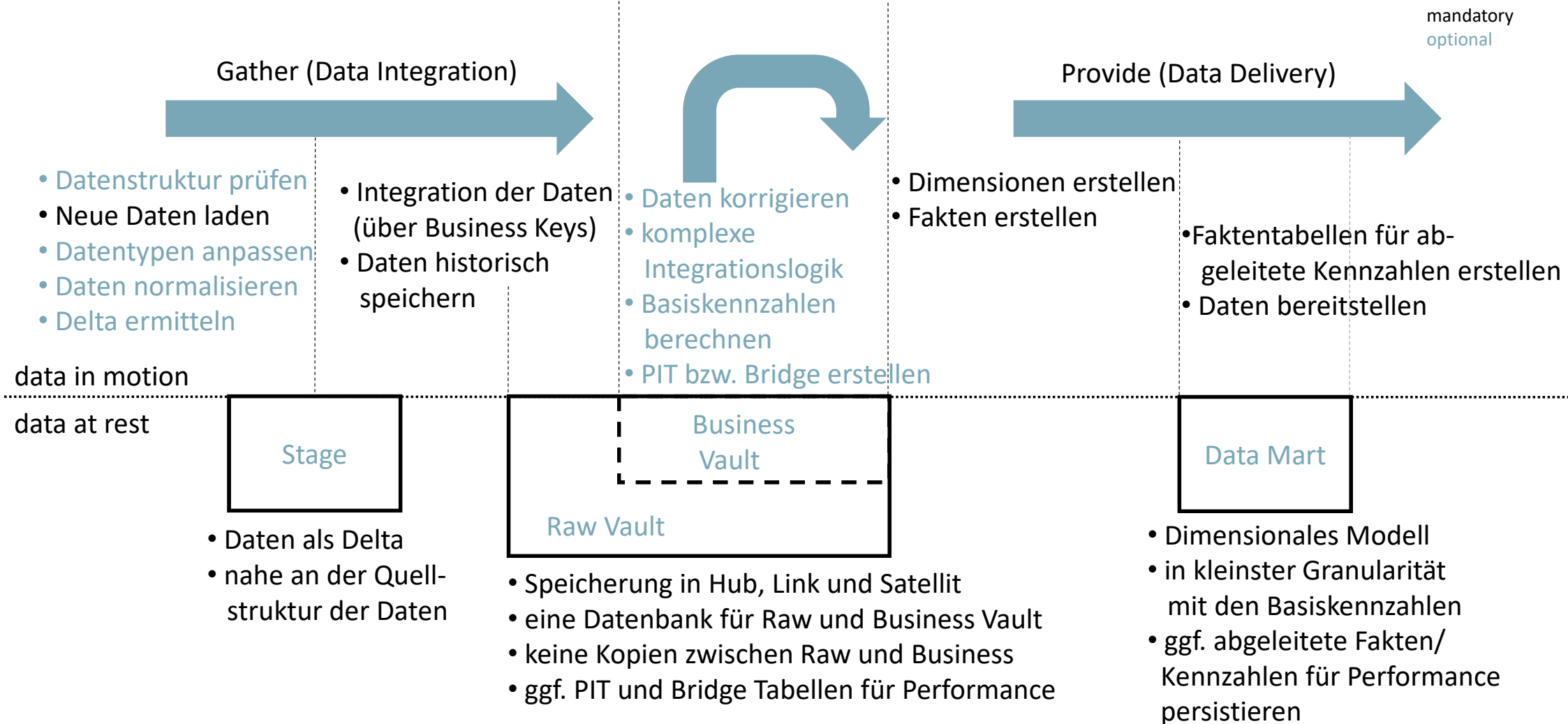
### Faktentypen

- ★ **Transaction Grain**  
fester Punkt in Raum und Zeit
- ★ **Periodic Snapshot Grain**  
regelmäßig wiederkehrende Messung
- ★ **Cumulating Snapshot Grain**  
Kombination existierender Fakten

SCD Type	Dimension Table Action	Impact on Fact Analysis
Type 0	No change to attribute value	Facts associated with attribute's original value
Type 1	Overwrite attribute value	Facts associated with attribute's current value
Type 2	Add new dimension row for profile with new attribute value	Facts associated with attribute value in effect when fact occurred
Type 3	Add new column to preserve attribute's current and prior values	Facts associated with both current and prior attribute alternative values
Type 4	Add mini-dimension table containing rapidly changing attributes	Facts associated with rapidly changing attributes in effect when fact occurred
Type 5	Add type 4 mini-dimension, along with overwritten type 1 mini-dimension key in base dimension	Facts associated with rapidly changing attributes in effect when fact occurred, plus current rapidly changing attribute values
Type 6	Add type 1 overwritten attributes to type 2 dimension row, and overwrite all prior dimension rows	Facts associated with attribute value in effect when fact occurred, plus current values
Type 7	Add type 2 dimension row with new attribute value, plus view limited to current rows and/or attribute values	Facts associated with attribute value in effect when fact occurred, plus current values

Quelle: <http://www.kimballgroup.com/2013/02/design-tip-152-slowly-changing-dimension-types-0-4-5-6-7/>

# Architektur Datenintegration komplett



mandatory  
optional

## Kennzahlen so einfach wie möglich machen.



- ★ Die Berechnung der Kennzahl möglichst nur an einer Stelle.
- ★ Kennzahlen zwischen den verschiedenen Data Marts gleich halten.
- ★ Entspricht  
    ,Schadensfälle nach Verträgen‘  
    aus dem Data Mart ,Schadensfälle‘  
wirklich  
    ,Verträge mit Schadensfall‘  
    aus dem Data Mart ,Verträge‘?

## Klassifizierung der Kennzahlen nach ihrer Entstehung

### ★ **Basis-Kennzahlen**

werden aus dem Core Warehouse (Raw und Business Vault) ausgewählt:

- ★ Einnahmen
- ★ Anzahl der Verträge

### ★ **Abgeleitete Kennzahlen**

werden im Data Mart aus Kennzahlen und Dimensionen berechnet

- ★ Einnahmen nach Zahlungseingang, Einnahmen nach Auftragseingang
- ★ Anzahl der Verträge mit Personen unter 25 Jahren, mit Rentnern

### ★ **Berechnete Kennzahlen**

können nur im Frontend berechnet werden, da die Berechnung erst nach der Auswahl der Dimensionen erfolgen kann

- ★ insbesondere Quoten
- ★ Schadenfreie Verträge in %

## Unterschiede zwischen Transactions- und periodischem Snapshot-Grain

### ★ Events

1. März: Auftragseingang

25. April: Änderung der  
Auftragspositionen

(zwei neue, eine wird  
gestrichen)

9. Juni: Bestellung wird geliefert

### ★ Transaction Grain (Auftragseingang)

Faktentabelle März: Bestellung

Faktentabelle April: Delta auf  
Bestellung

Faktentabelle Juni: Lieferung

Spezielle Aggregationsfunktionen für  
die Zeit:

- Bisheriges Jahr
- Erster Wert
- Letzter Wert

### ★ Periodic snapshot Grain (Auftragsbuch)

Faktentabelle März: Zufluss

Bestellung

Faktentabelle April: Bestellung

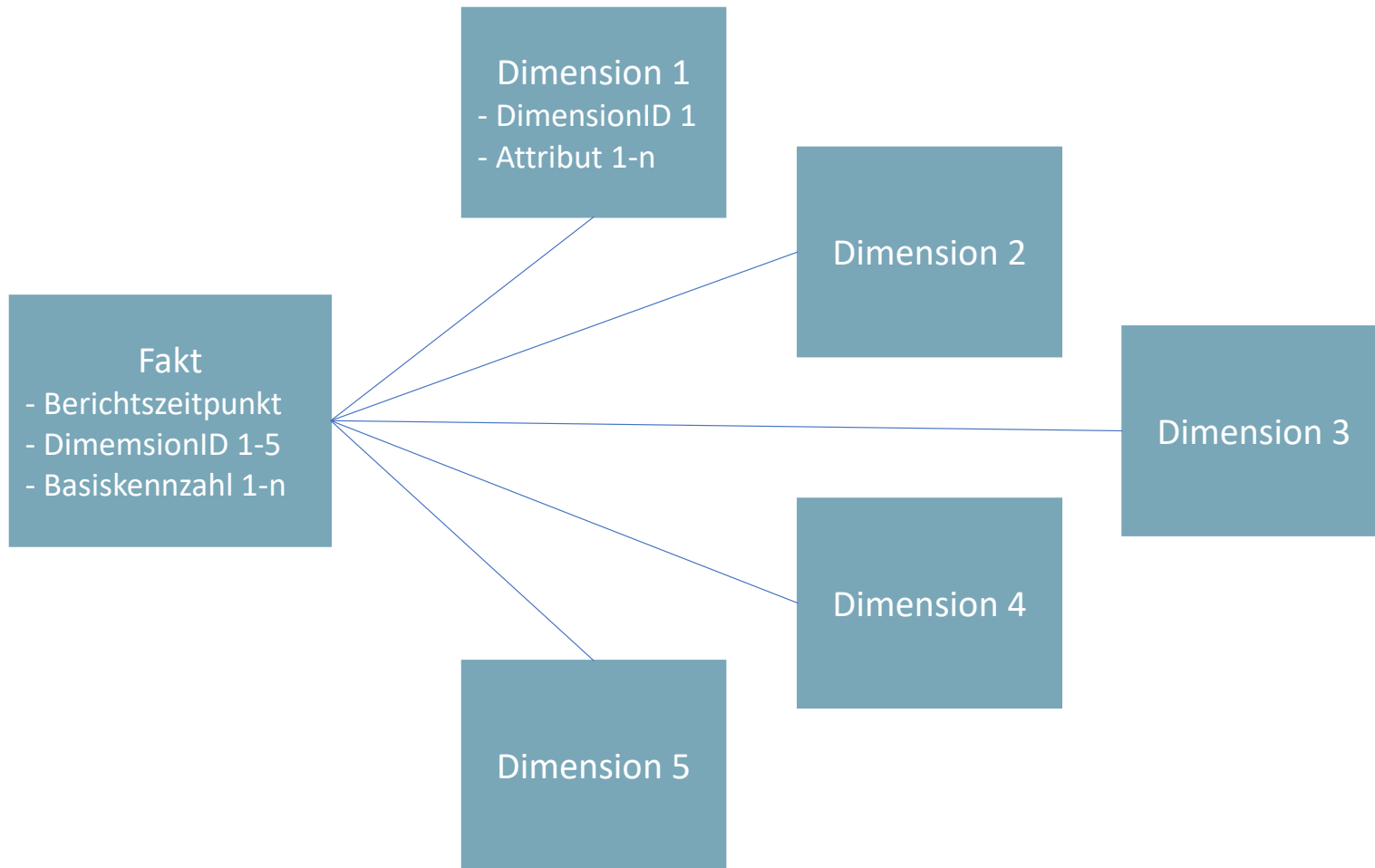
Faktentabelle Mai: Bestellung

Faktentabelle Juni: Abfluss

Nur die Zu- und Abflüsse können über  
die Zeit aggregiert werden.

Normalerweise werden alle Daten als  
Zeitreihen ohne zeitliche Aggregation  
dargestellt.

# Änderungen und Zeit im dimensionalen Modell



- ★ Der Berichtszeitpunkt als fixer Zeitpunkt für die Fakten
- ★ Fakt und Dimension sind fix verbunden
- ★ Attribute in Dimensionen können sich ändern
- ★ Die SCD-Typen 0-7 wurden geschaffen, um Änderungen in den Dimensionen zu behandeln

# Zeit in Dimensionen oder ,Wie man Historie in Dimensionen speichert'.

überhaupt keine Veränderung → nur den aktuellen Wert → Historische Speicherung von Werten

Vermeidung der Aktualisierung der Faktentabelle

SCD Type	Dimension Table Action	Impact on Fact Analysis
Type 0	No change to attribute value	Facts associated with attribute's original value
Type 1	Overwrite attribute value	Facts associated with attribute's current value
Type 2	Add new dimension row for profile with new attribute value	Facts associated with attribute value in effect when fact occurred
Type 3	Add new column to preserve attribute's current and prior values	Facts associated with both current and prior attribute alternative values
Type 4	Add mini-dimension table containing rapidly changing attributes	Facts associated with rapidly changing attributes in effect when fact occurred
Type 5	Add type 4 mini-dimension, along with overwritten type 1 mini-dimension key in base dimension	Facts associated with rapidly changing attributes in effect when fact occurred, plus current rapidly changing attribute values
Type 6	Add type 1 overwritten attributes to type 2 dimension row, and overwrite all prior dimension rows	Facts associated with attribute value in effect when fact occurred, plus current values
Type 7	Add type 2 dimension row with new attribute value, plus view limited to current rows and/or attribute values	Facts associated with attribute value in effect when fact occurred, plus current values

Data Vault ist immer Typ 2

Anpassung der Faktentabelle notwendig, da mehr als ein Wert pro ID in der Dimension gültig ist.

# Klassifizierung der Dimensionen nach ihrer Geschichte

## ★ **as-is:**

Alle Fakten beziehen sich auf den aktuellen Dimensionswert

*Basierend auf  
HubID*

## ★ **as-was:**

Alle Fakten beziehen sich auf den Dimensionswert, der zum Zeitpunkt der Erfassung des Fakts gültig war (Berichtszeitpunkt).

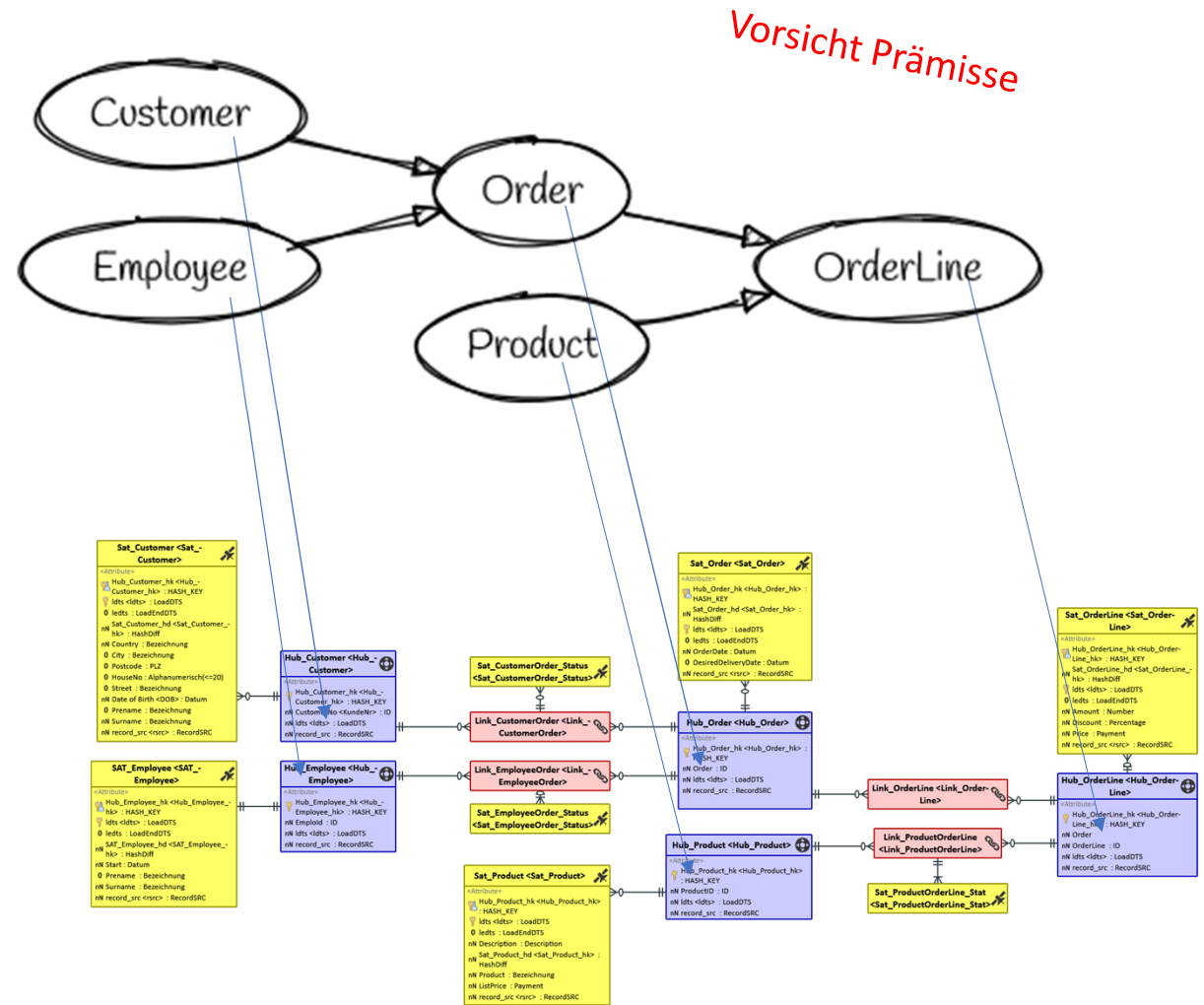
*Basierend auf  
HubID und LDTs*

## ★ **as-of:**

Alle Fakten beziehen sich auf den Dimensionswert, der zu einem bestimmten, frei wählbaren Zeitpunkt galt

*Basierend auf  
HubID und LDTs*

Data Vault integriert die Daten anhand von Geschäftsobjekten mit jeweils einem Hub pro Geschäftsobjekt



## Welche Integrationsaufgaben hat das zur Folge?

### ★ 1 Quellsystem

-> Daten einfach laden

### ★ 2-n Quellsysteme mit jeweils nur einer Quelle pro Hub/GO

-> Daten einfach laden

### ★ 2-n Quellsysteme:

#### ★ mit disjunkten Hub/GO

-> einfaches union

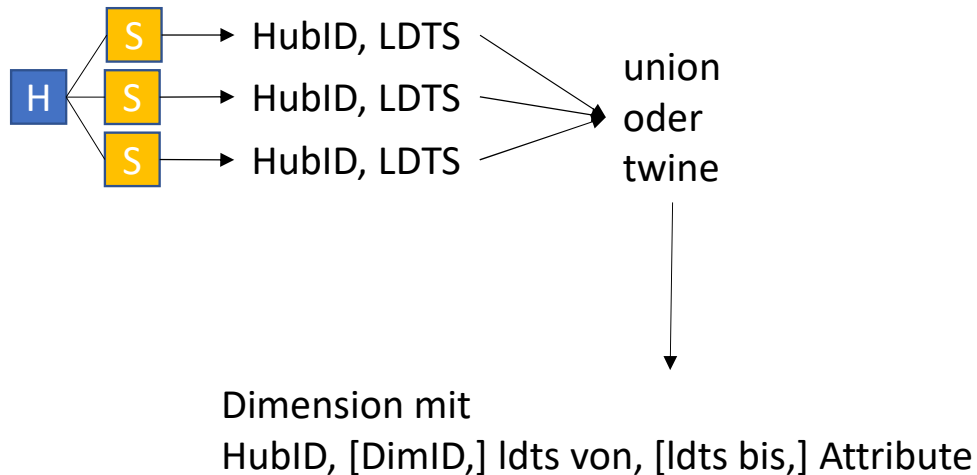
#### ★ mit überschneidenden Hub/GO

-> twine o.ä., coalesce() mit prio

### Davon unabhängig:

- ★ Kennzahl geliefert/zu berechnen
- ★ Same-as wg. Dubletten/Migration
- ★ Bitemp m/o rückwärtigen Änderungen

## Dimension, die verschiedene Attribute von mehreren Satelliten aufnimmt



- ★ Die Zahl der HubID bleibt immer gleich
- ★ Die LDTS variieren auf Grund von
  - ★ Attributauswahl
  - ★ Satellitenauswahl
  - ★ Zeitraumauswahl

### Damit haben wir keine stabile DimensionID (DimID)

- ★ Können uns aber eine rekonstruierbare errechnen:
  - ★ DimID mittels `row_number()` erzeugen partitioniert und sortiert nach Idts und HubID
  - ★ HubID stabil, DimID kann sich verändern
  - ★ Diese DimID ist deterministisch, d.h. immer gleich bei gleicher ausgewählter Anzahl Sätze

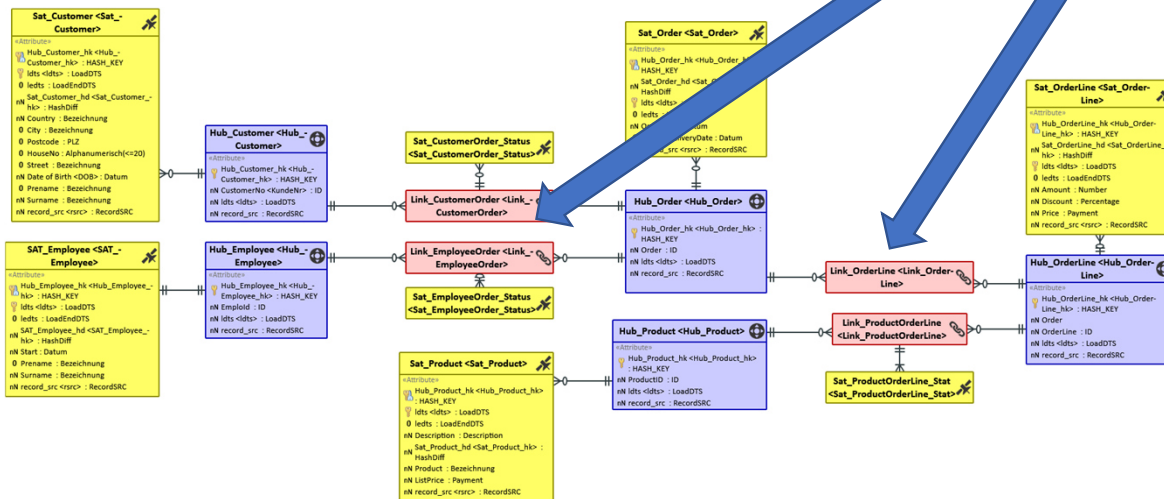
# Ergänzen der Faktentabelle um die HubIDs der Dimensionen

Hinzufügen von HubIDs durch Zugriff auf die Links und deren Status-Satelliten

- ★ normalerweise zum Berichtszeitpunkt
- ★ in seltenen Fällen kann dies auch mit einer anderen Zeit geschehen, um eine Dimension zu einem bestimmten Zeitpunkt zu erhalten, wie:

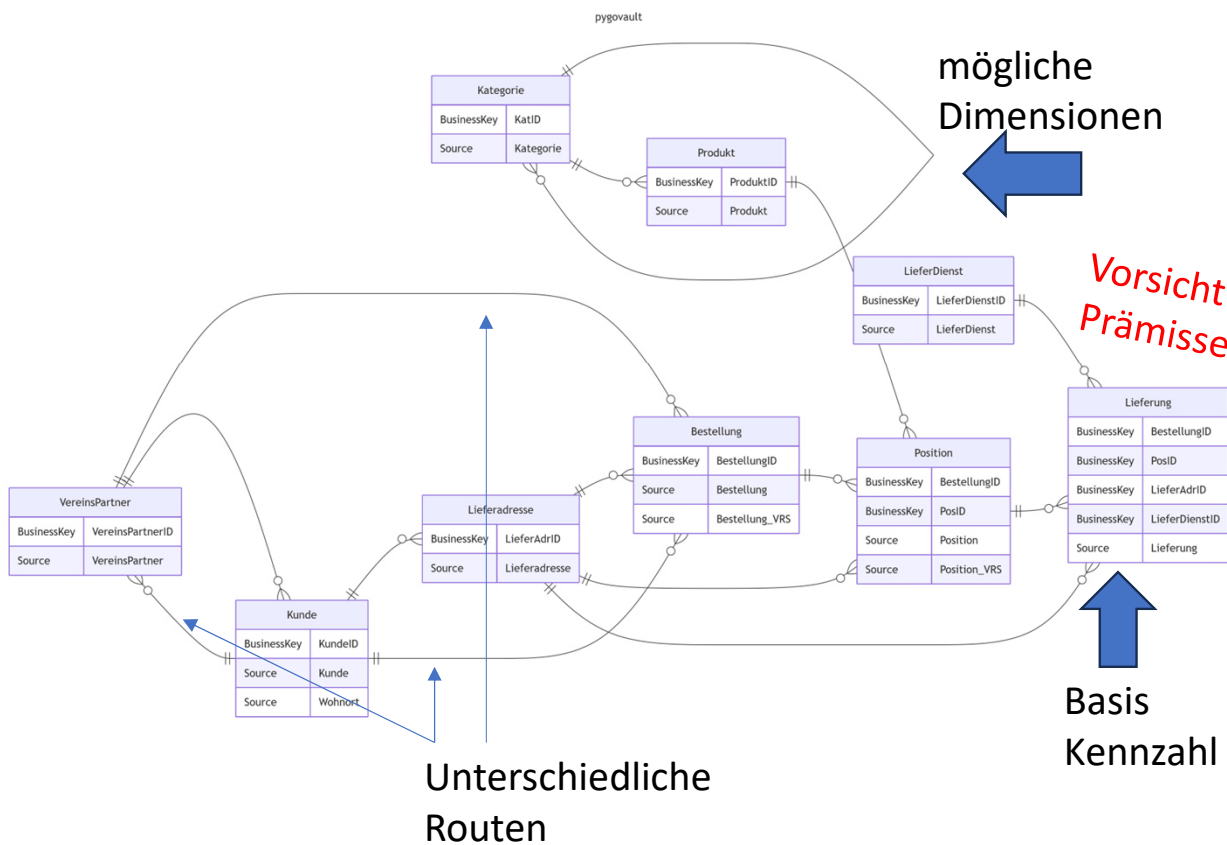
- ★ Wo ist die Person beschäftigt bei Beginn des Versicherungsvertrags
- ★ Familienstand bei der Registrierung auf der Website
- ★ ...

- ★ Mit oder ohne Auflösung eines möglichen same-as-link zu diesem Hub



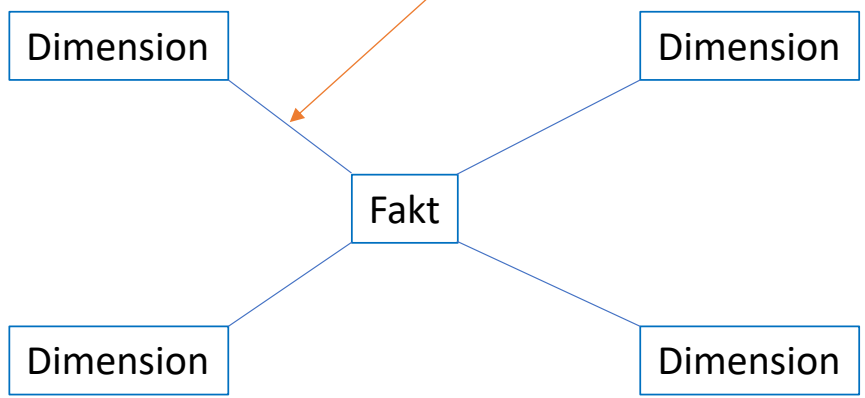
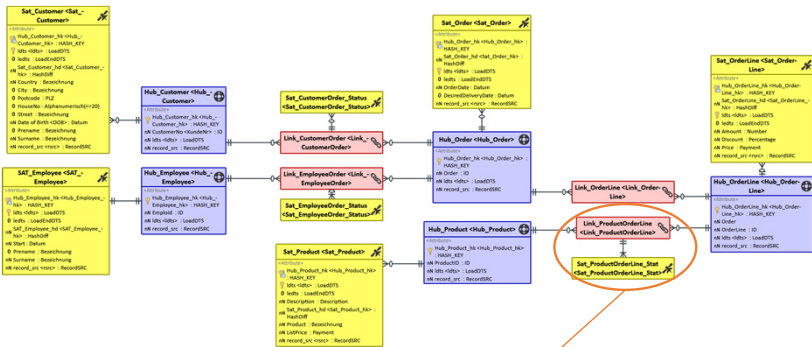
***Dies ist unabhängig von der Dimension Zeit (as-\*)***

# Erstellen eines Geschäftsobjektmodells aus Ihrem Data Vault-Modell



- ★ Für jeden Hub ein Business-Objekt
- ★ Zeichnen Sie für jeden binären Link einen Pfeil zum anderen Geschäftsobjekt (Richtung von 1:n, die Stage verrät die Kardinalität)
- ★ n-äre Links ist oder Links mit Satelliten (Status-Satelliten zählen hier nicht) -> ein Business-Objekt anlegen
- ★ Zeichnen Sie eine Linie zwischen allen beteiligten Business-Objekten (Hubs) und dem neuen Business-Objekt mit dem Pfeil in Richtung des neuen Business-Objekts
- ★ Organisieren Sie alle Geschäftsobjekte entlang der Pfeile in Leserichtung (von links nach rechts, oben-unten)

# Der richtige Dimensionsschlüssel (HubID) für eine Faktentabelle

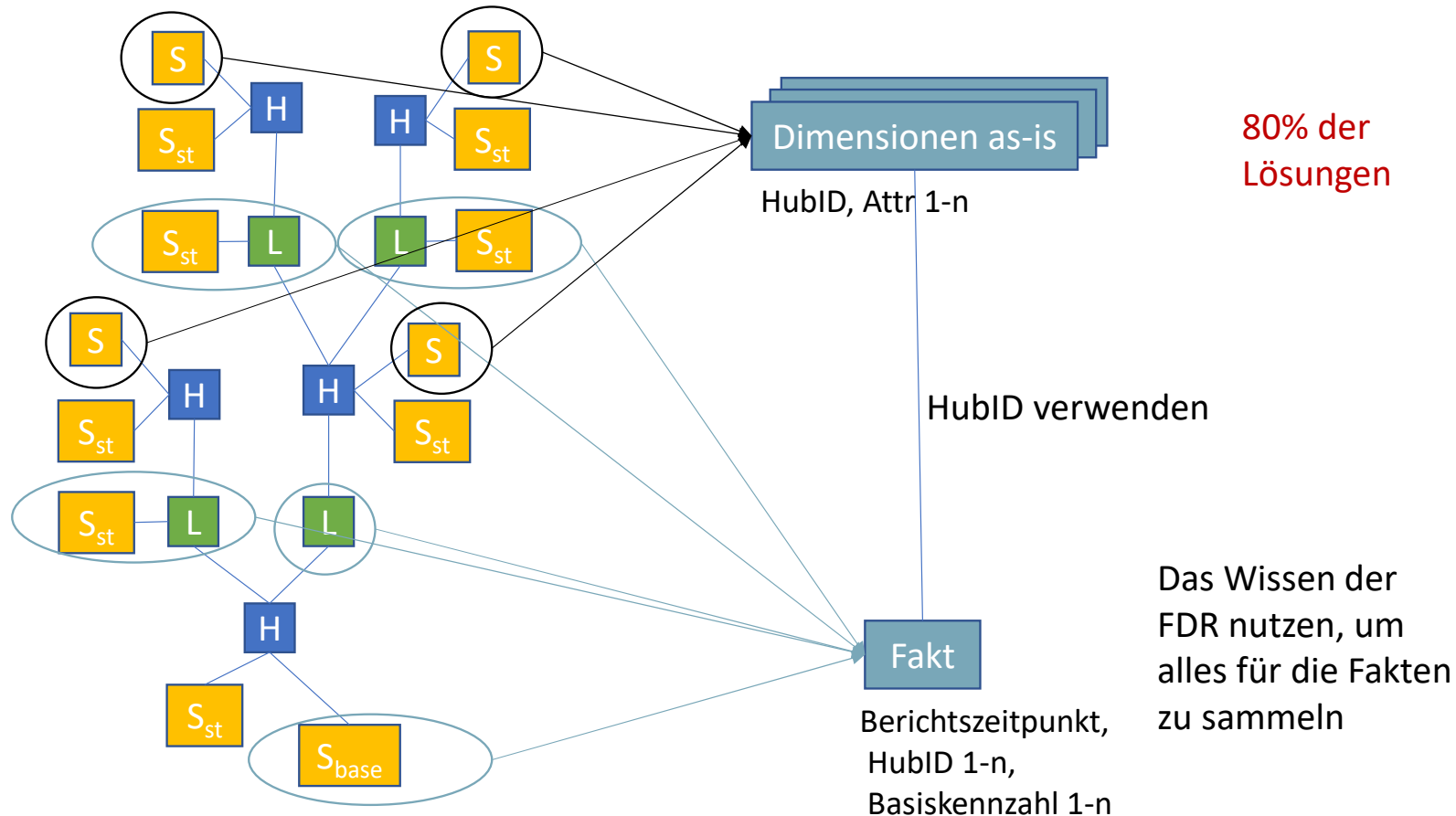


- ★ Hier geht es um die Frage, welches Objekt für die Faktentabelle ausgewählt werden soll.
- ★ Das ist etwas anderes als die Frage, welcher Datenstand der Dimension zu wählen ist.
- ★ Verwendung einer anderen Zeit als dem Berichtszeitpunkt, mit der durch die Links/Linksatelliten zwischen Faktentabelle und der Dimension navigiert wird.
- ★ Es kann mehr als einen Weg vom Satelliten mit der Basiskennzahl zu dieser HubID geben
- ★ Dies ist die Route zwischen Fakt und Dimension (FDR - Fakt-Dimension-Route) mit den 2 Spezialfällen:
  - ★ Zu einer anderen Zeit
  - ★ Roleplaying Dimension

## Alle Anforderungen für einen Data Mart

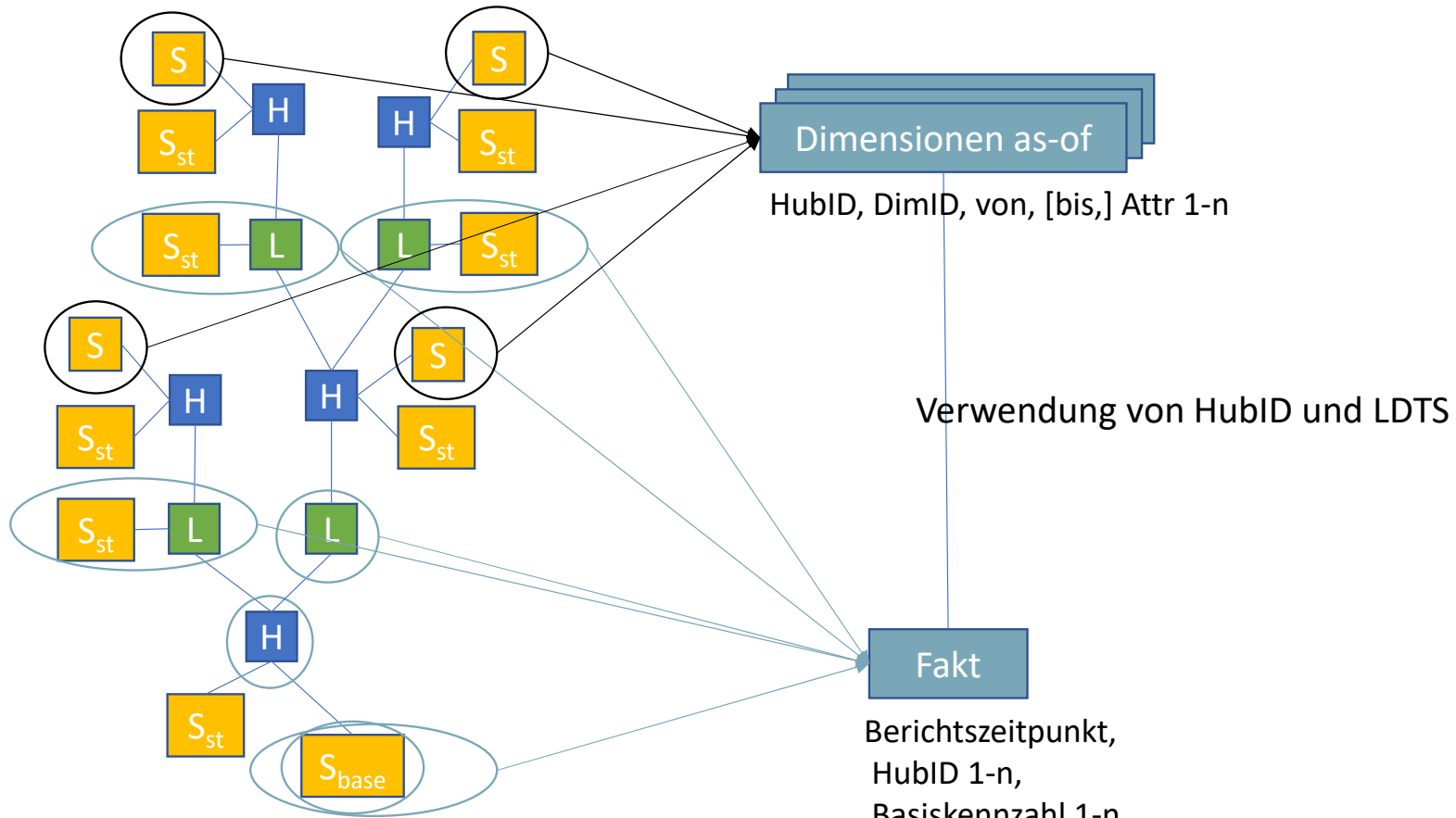
- ★ Kennzahlen
  - ★ Basiskennzahlen (mit jeweiligem Satellitenattribut)
  - ★ Abgeleitete Kennzahlen
  - ★ Berechnete Kennzahlen
- ★ Fakten
  - ★ Transaction Grain
  - ★ Periodic snapshot Grain
- ★ Berichtszeitpunkt
- ★ Dimensionen
  - ★ aus welchem Geschäftsobjekt
  - ★ as-is, as-was, as-of
- ★ Fakt-Dimension-Relation (FDR)
  - ★ Pfad
  - ★ Zeit

# Blick auf die Umsetzung – aktueller Zustand



- Legende
- H - Hub
  - L - Link
  - S - Satellit
  - $S_{base}$  - Satellit mit Basiskennzahl
  - $S_{st}$  - Status Satellite für Löschungen

# Betrachtung der Umsetzung as-of



## Legende

H - Hub

L - Link

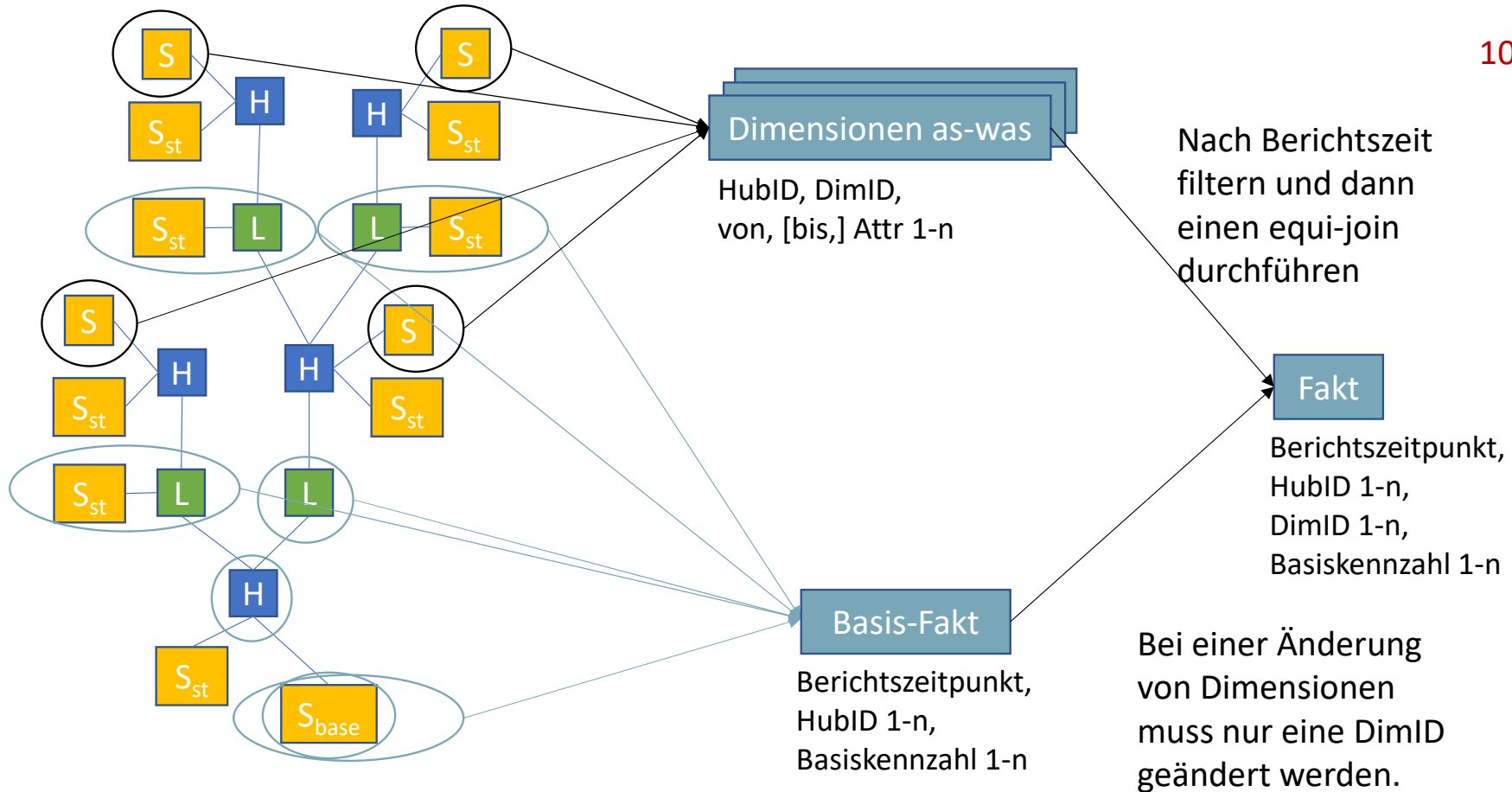
S - Satellit

S<sub>base</sub> - Satellit mit Basiskennzahl

S<sub>st</sub> - Status Satellite für Löschungen

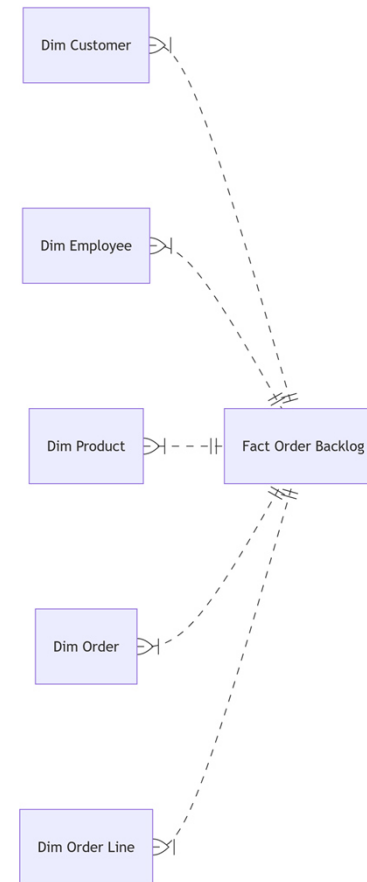
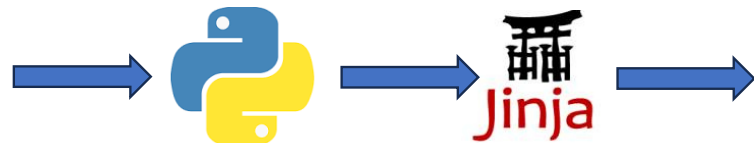
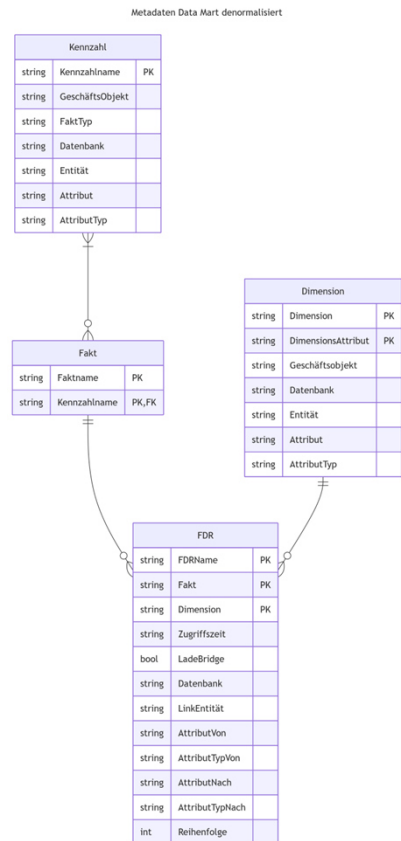
# Betrachtet man die Umsetzung - as-was

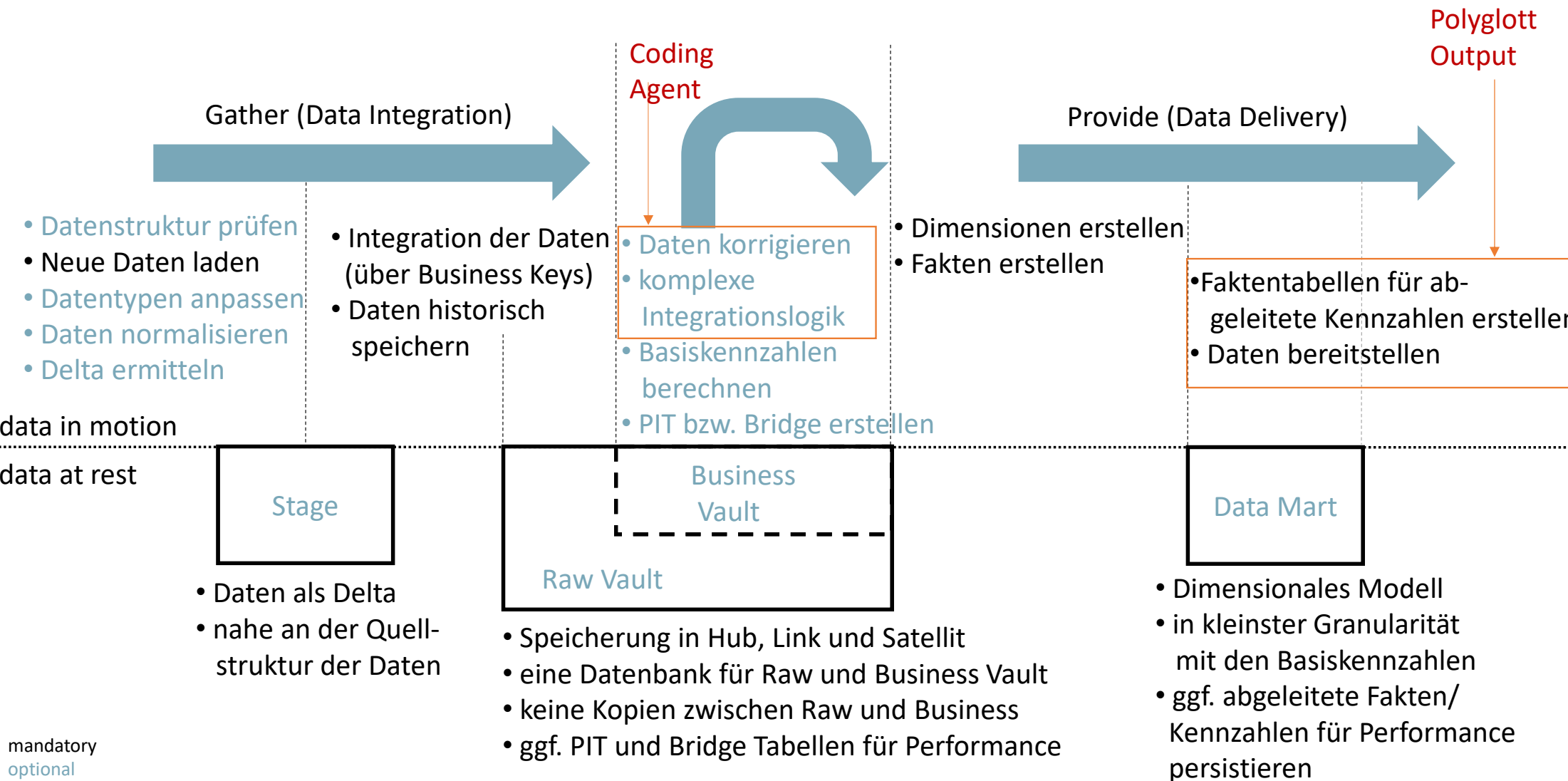
10%



- Legende  
 H - Hub  
 L - Link  
 S - Satellit  
 S<sub>base</sub> - Satellit mit Basiskennzahl  
 S<sub>st</sub> - Status Satellite für Löschungen

# Betrachtet man die Umsetzung - as-was / as-of DimID





Danke für Ihre Aufmerksamkeit!

m[method] 2 data