

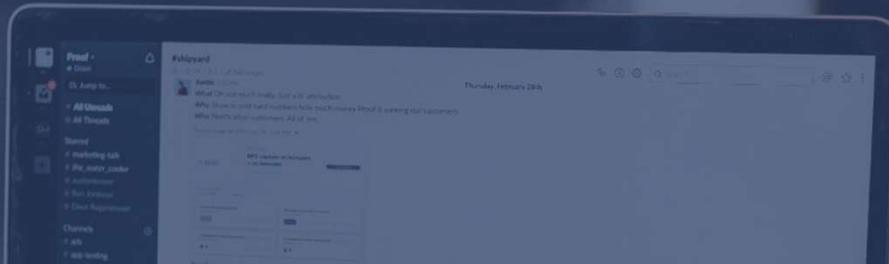


Data Vault on a diet

Pragmatismus vs. Perfektionismus

Eckhard Zemp

Hannover, 27. November 2025



Use Case

- Kunde hat Databricks, nur Standard-Tools der Plattform sollen genutzt werden
- Für Analysten, wenn kein Unternehmensdatenmodell vorhanden ist

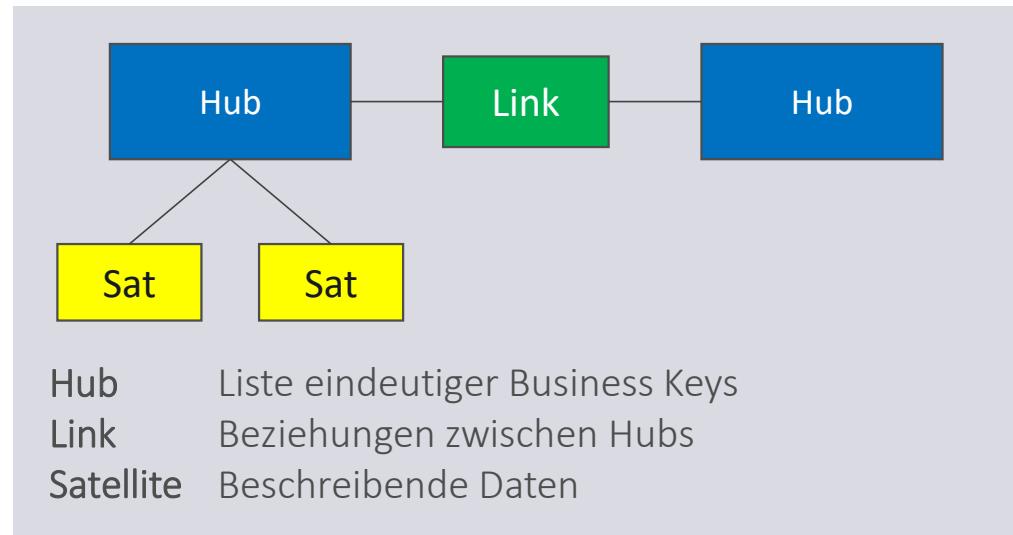


“ DISCLAIMER ”

Es sieht aus wie Data Vault,
ist es aber gemäss Lehrbücher nicht



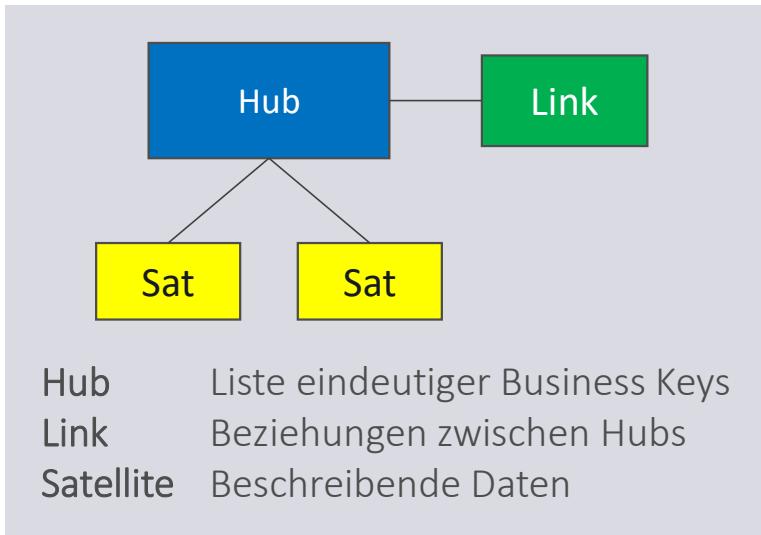
Was sind die Core Elemente eines Data Vault?



Grundregeln des Data Vault

- Hubs stehen für sich alleine und verweisen nie auf andere Objekte.
- Links verweisen auf zwei oder mehr Hubs, aber nie auf andere Links oder Satelliten.
- Satelliten verweisen auf genau einen Hub oder einen Link, nie auf andere Satelliten.

Was „erweitert“ das einfache Modell, um nun als Data Vault bezeichnet zu werden?



- Dan Linstedt:
 - Methodology
 - Architecture
 - Implementation
 - Modelling
- Zusätzliche Metadaten: Load Date und Record Source
- Zusätzlicher HashKey basierend auf dem Business Key
- Historisierung von Satelliten
- Auditierfähigkeit des Raw Vault
- Unterscheidung von Raw Vault und Business Vault
- Ermöglicht komplexe Link-Strukturen
- Effectivity Satellites bei Links
- Ghost Records (und Zero Keys)
- ...

► Ohne Automatisierung (fast) unmöglich zu bauen/bedienen/nutzen



Was Liebe ich an Data Vault

- Unique List of Business Keys

Grain wird enforced und stellt sicher, dass keine Duplikate vorhanden sind

- Die Freiheit, so viele Satelliten anzuhängen, wie man will

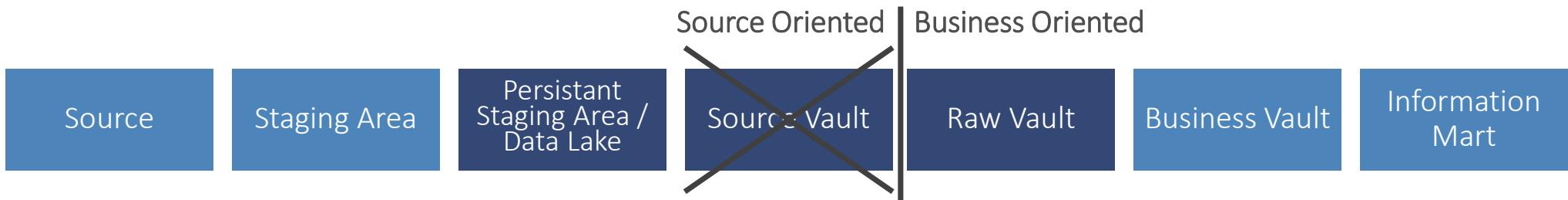
- Die Freiheit, logische Beziehungen zwischen den Hubs zu erstellen
(Ich mag Binary Links mit 2 Enden)

- Die Fähigkeit, Berechnungsschritte über viele Layer hinweg zu verarbeiten und den Content eines Hubs mit jedem Schritt zu erweitern

- Automatisierbar!

Hubs, Links und Satelliten werden immer identisch beladen, ohne Ausnahme!

Wo ist der Single Point of Truth der Unternehmensdaten?



Die Verlockung ist gross, die Quelldatenbank mit Data Vault nachzumodellieren. Damit hat die Fachabteilung nichts gewonnen. Jeder will es vermeiden, aber es ist schwer, sich davon zu distanzieren.

Standard Data Vault



Erfindung und Vorgehen von Dan Linstedt. Daten werden direkt mit dem Business Key der Fachabteilung beladen.
Schwierigkeit: 80-90% der Tabellen enthalten den eigentlichen Business Key nicht und muss bei jedem Load dazugeholt werden.

Verwendung einer Persisting Staging Area



Entkopplung von Daten Beschaffung und Daten Modellierung. Keine doppelte Datenhaltung mit einem zusätzlichen Raw Vault notwendig. Im Business Vault liegen nur die Daten, worüber wir ALLES wissen.

Was wäre ein schlanker Ansatz? (I)



➤ Trennung von Daten Beschaffung und Daten Modellierung

Wie weiss ich, wie mein Raw Vault aussehen soll, wenn ich keinen Zugriff auf die Quell-Daten habe? Vielleicht brauche ich mehr als 2 Quellobjekte (BK + Attribute), um einen sinnvollen Satelliten zu bauen (ohne das Modell aufzublähen)

➤ Jedes Quellobjekt wird als historisiertes unabhängiges Objekt der Quelle geladen (=PSA)

➤ Es braucht keinen Raw Vault, wenn es eine PSA gibt

Ein Satellit mit hunderten von Spalten, dessen Akronyme niemand versteht, braucht niemand (Ausnahmen bestätigen die Regel). Doppelte Datenhaltung kostet Storage

➤ Kein Unterschied mehr zwischen Raw und Business Vault → Es gibt nur noch Content/Wissen

➤ Die Modellierung orientiert sich nach wie vor am Bedarf der Fachabteilung und/oder ggf. dem Information Mart

Die Fachabteilung ist direkt eingebunden und kann sich damit identifizieren

➤ Die Business Keys sollen eindeutig mit der Fachabteilung identifiziert werden

... und kommen nicht schleichend als „Zwängerei“ der Quelle ins Data Vault

➤ Wir fokussieren uns alleine auf die Business Rules

Wir gestalten unseren Business Vault so, dass er zum Kunden und seinen Data Marts passt

➤ Beziehungen sind normalerweise 2-seitig (Natürliche 1:m Beziehungen)

Bei Performance-Problemen kann man später immer noch verdichten

➤ Logische Zwischenelemente können trotzdem unsichtbar für den Data Mart modelliert und gepflegt werden

Was wäre ein schlanker Ansatz? (II)



➤ Hub Hash Heys haben keinen Mehrwert

Data Vault "0" war ohne Hashing und ohne Surrogate Keys. Bei modernen Datenplattformen gibt es keinen nennenswerten Vorteil, den Hub Business Key zu hashen.
Hashing kostet Compute

➤ Satelliten dürfen auch ohne Historie beladen werden

Ist offiziell erlaubt, jedoch nicht üblich. Wähle SCD1, wenn SCD2 zu kompliziert oder nicht erforderlich ist.
Bedenke jedoch, üblicherweise kommt der Appetit für die Historie im Verlauf des Projekts.

➤ Links wie Satelliten bauen

„Netherland Style“. Datavault Builder baut seine Transactional Links genau so, benennt sie jedoch als Links.
Dadurch benötigt es keinen Effectivity Satellite um die Gültigkeit eines Links zu tracken.

➤ Keine Angst vor Veränderung!

Stellt sich heraus, dass etwas anders benannt oder modelliert werden soll, keine Panik. Die PSA hat alles für die Ewigkeit

Ist das Ergebnis noch ein Data Vault?

Die Core-Elemente werden wie vorgesehen, verwendet.

Viele Elemente, die gemäss Lehrbuch notwendig sind, können bei Bedarf trotzdem verwendet werden, um die damit verbundenen Vorteile zu nutzen.

→ Benutze diese Vorschläge, benenne es jedoch nicht als Data Vault



Ein Beispiel (I)

1

Satellite Sales Order Main

| Sales Order BK | Shop | Sales Order No | Order Date | Order Status | Payment Method |
|----------------|------|----------------|------------|--------------|----------------|
| DE 1234 | DE | 1234 | 01.04.2025 | ordered | creditcard |

Satellite Sales Order Shipment Address

| Sales Order BK | Shipment Address 1 | ... | ... | ... | ... |
|----------------|--------------------|-----|-----|-----|-----|
| | | | | | |

Link Sales Order → Customer

| Sales Order BK | Customer BK |
|----------------|-------------|
| | |

Hub Sales Order

Sales Order BK

- Jeder Satellite oder Link ist eine eigene Tabelle oder View
- Concatenated Business Keys als eigenständige System Column betrachten
- Der Hub ist ein Distinct View auf den vorhandenen Objekten

2



Ein Beispiel (II)

3

Zusammenführen der einzelnen Objekte zu einer Gesamt-View auf Ebene Hub

| Hub Sales Order | Satellite Sales Order Main | | | | | Satellite Sales Order Shipment Address | | |
|-----------------|----------------------------|----------------|------------|--------------|----------------|--|-----|-----|
| Sales Order BK | Shop | Sales Order No | Order Date | Order Status | Payment Method | Shipment Address 1 | ... | ... |

4

Bauen des Information Marts (Dimensionen, Fakten, Reports, etc.)

Nicht vergessen: Unique Check
Vor oder nach Beladung der Objekte



Es ist ganz einfach

- Unique List of Business Keys!
- Was wissen wir darüber?

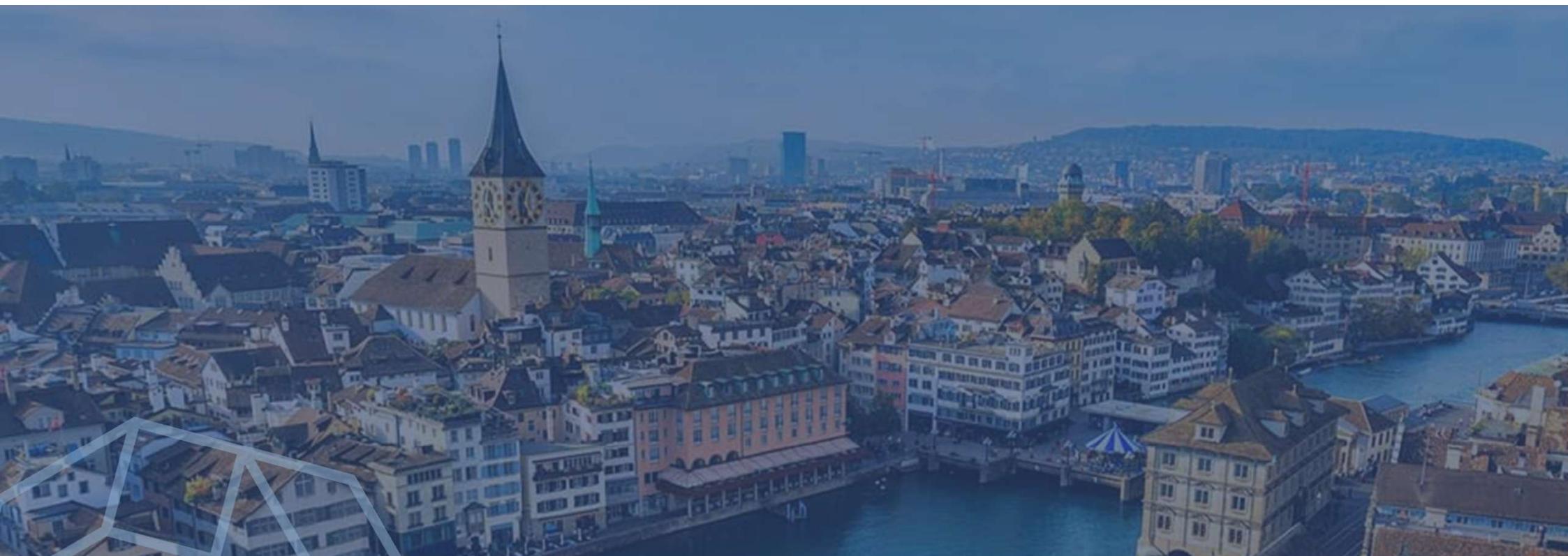




Vielen Dank!

Join us





Eckhard Zemp



b.telligent Schweiz GmbH
Kanzleistrasse 57
8004 Zürich
+41 44 585 39 80
www.btelligent.com