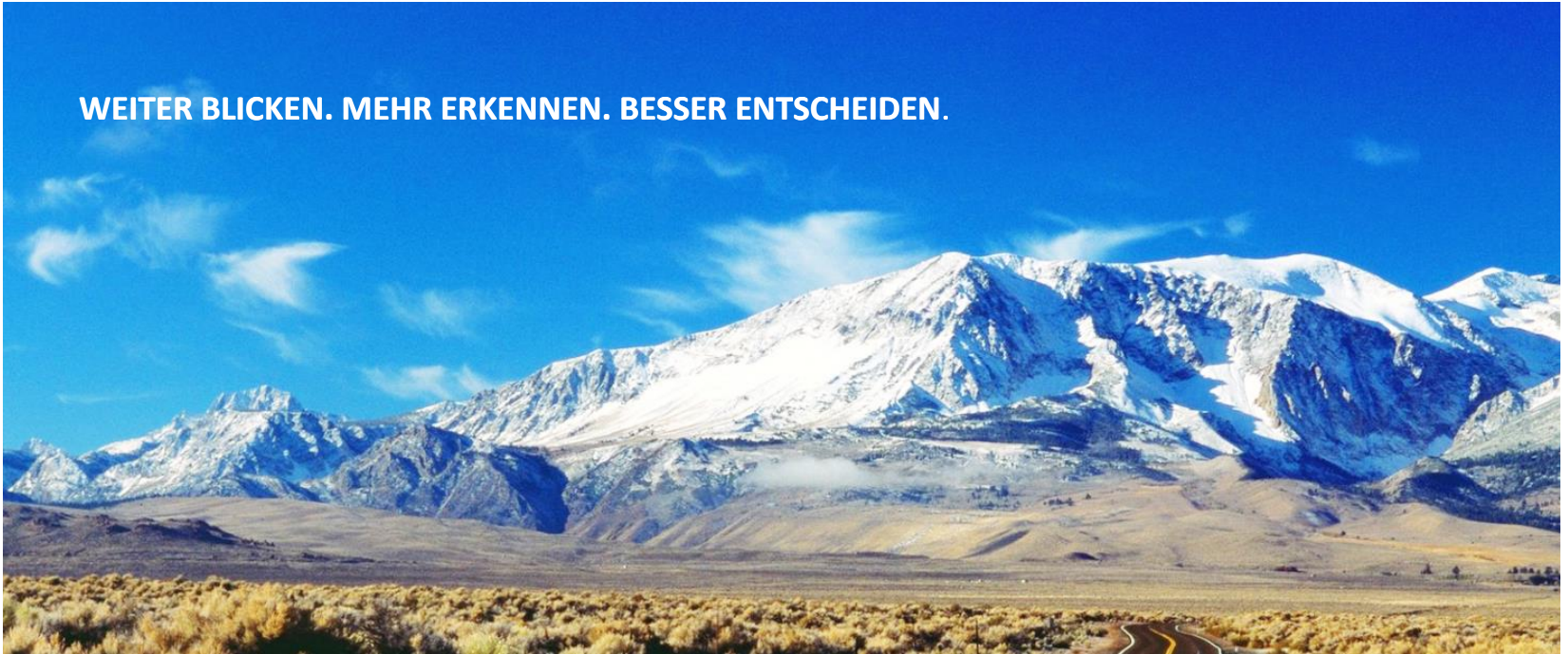



Ensemble Modeling am Beispiel von Data Vault

WEITER BLICKEN. MEHR ERKENNEN. BESSER ENTSCHIEDEN.



- › Was ist ein Ensemble?
- › Was ist Data Vault?
- › Warum will ich das?
- › Warum will ich das nicht?
- › Wie kann das überhaupt funktionieren?
(hier mogelt der Ben dann Biml rein)

- › DBAs?
- › Architekten?
- › Entwickler?
- › Controller?
- › Irgendwelche Data Vault Erfahrung?

- › Ben Weissman, Solisyon, Nürnberg
- ›  @bweissman
- › b.weissman@solisyon.de
- › SQL Server seit Version 6.5



Data Science



Big Data



Certified Data Vault Modeler

**BimlHero**
CERTIFIED EXPERT**Microsoft**
CERTIFIED

Solutions Expert

Business Intelligence

Microsoft
CERTIFIED

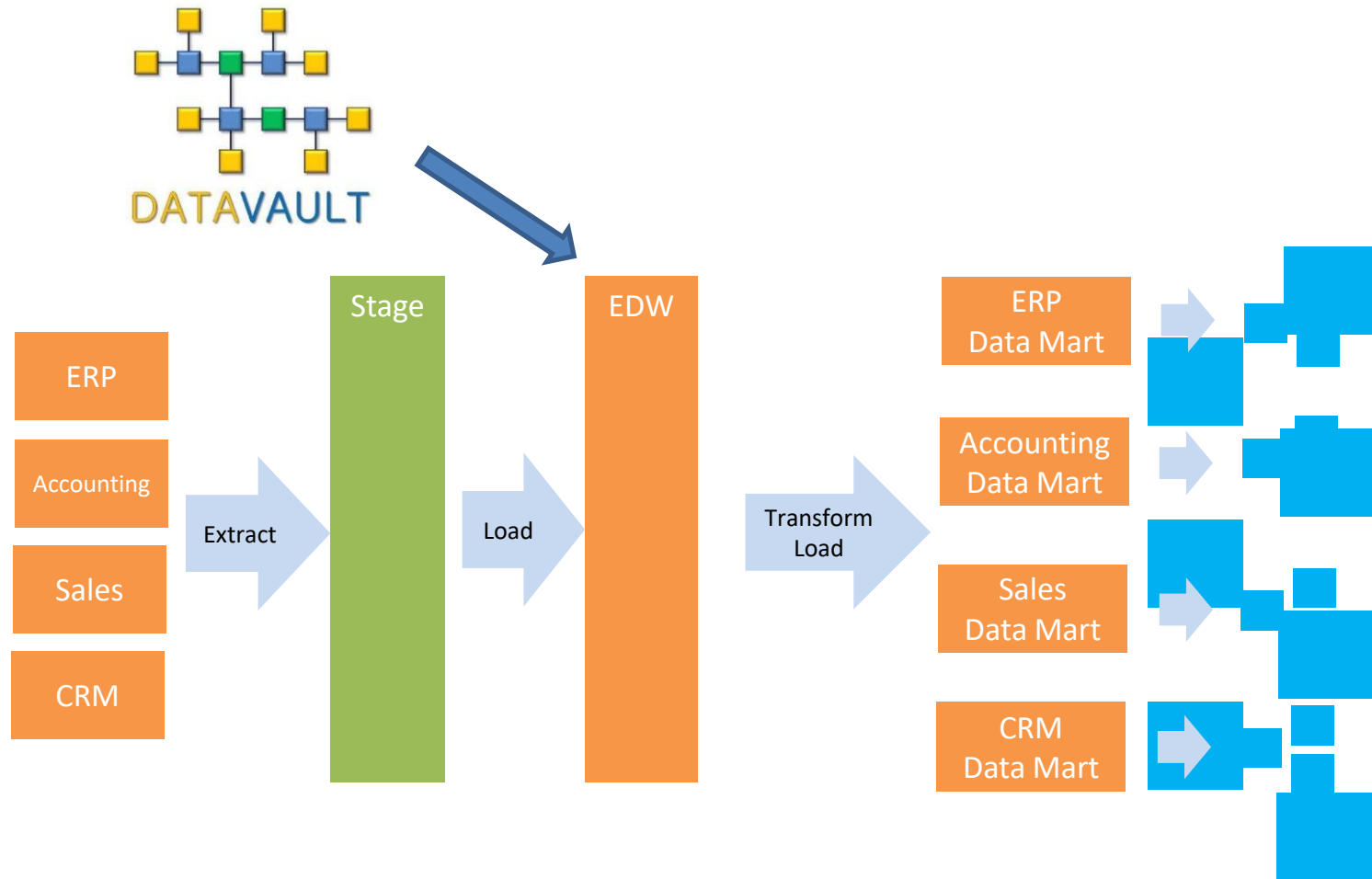
Solutions Expert

Data Platform

- › Ein Ensemble ist ein Hub mit allen zugehörigen Satelliten und Links
- › Hä?
- › Gleich!
- › Ensembles gibt es nicht nur bei Data Vault, sondern auch bei anderen Methoden wie:
 - › Focal Point
 - › Anchor
 - › Hyper Agility

- › Methodologie (KEINE Technologie!)
- › Vereinfacht gesagt: Alles wird Type 2 und sinnvoll gruppiert
- › Besteht aus:
 - › HUBs
 - › LINKs
 - › SAT_(elitten)S
- › HUBs + Links formen den Backbone, Satelliten bringen den Inhalt

- › Verschiedene Ansätze (z.B. Dan Linstedt vs. Hans Hultgren)
 - › „Immer“ ist also nicht immer „immer“
 - › Es gibt klar falsche aber nie einzig richtige Wege 😊
- › Modellierung nach Sprache und Logik des Business
- › Eine einmal angelegte Tabelle wird nicht mehr geändert
- › Es gibt nur Inserts (Ausnahmen bestätigen die Regel)



- › Formt den Business Key
- › Bestehen „immer“ aus 4 Spalten:
 - › Primary Key (Surrogate Key)
 - › Business Key
 - › Load Timestamp
 - › Source System
- › Wird immer als erstes befüllt

- › Formt Zusammenhänge / Relationships
- › Enthält 5-n Spalten:
 - › Surrogate Key
 - › Load Timestamp
 - › Source System
 - › 2-n Foreign Keys (Surrogate Keys der zugehörigen Hubs)

- › Formt Context, Beschreibung und Historie
- › Enthält 5-n Spalten:
 - › Surrogate Key
 - › Foreign Key (Surrogate Key des zugehörigen Hubs)
 - › Load Timestamp
 - › Source System
 - › 1-n Beschreibungsspalten
- › Ein Hub kann beliebig viele Satelliten haben!
- › Daten die nichts miteinander zu tun haben oder sich unterschiedlich häufig ändern gehören also meist in unterschiedliche Satelliten
- › Je Business Key ist i.d.R. nur ein Eintrag je Satellit und Zeitpunkt gültig
 - › Verschiedene Telefonnummern wären z.B. eher eigene Satelliten mit Typ-Link

- › Wozu passt die Überschrift am besten?
 - › Kunde
 - › Mitarbeiter
 - › Zuständiger Mitarbeiter
 - › Name des Mitarbeiters
 - › Verkaufsmenge in einer bestimmten Transaktion

- › Wie könnte bei den folgenden Kunden-Informationen ein Satelliten Design aussehen?
- › (Remember: Es gibt kein eindeutiges „Richtig“)
 - › Vorname
 - › Nachname
 - › Geschlecht
 - › Telefonnummer
 - › Geburtsdatum
 - › Familienstand
 - › Adresse
 - › Kinder (jeweils Name & Geburtsdatum)

- › Für die heutige Session schon 😊
- › Generell sind aber „Keyed Instances“, „Linked Sats“, „Valid From/To“, „Multiactive“ etc. durchaus denkbar und erlaubt
- › Da kommen dann auch UPDATES ins Spiel
- › (erlaubt ist eh alles, ist ja EUER Modell!)

- › Mit Post Its!
 - › Hubs: Blau
 - › Links: Rot
 - › Satelliten: Gelb
- › Was sind die „Core Business Concepts“ (→ Hubs)?
 - › Was beschäftigt das Business? Was sind verständliche Business Keys?
- › Wie hängen diese zusammen? (→ Links)
- › Was benötigte ich an Kontext? (→ Satelliten)
- › Wie sind die Hubs ggf. „aufzubrechen“ aus technischer Sicht?

Artikel

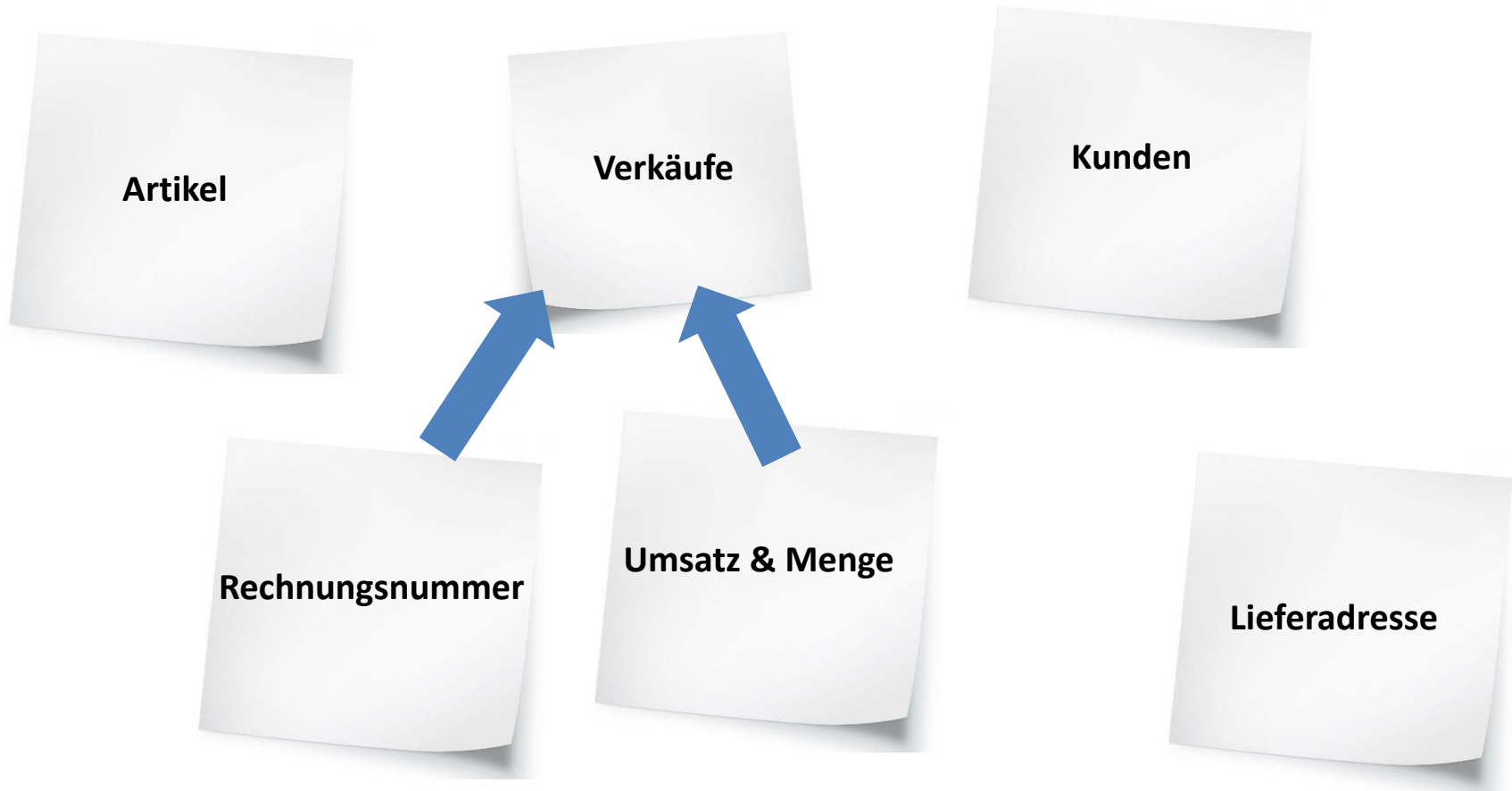
Verkäufe

Kunden

Rechnungsnummer

Umsatz & Menge

Lieferadresse



Artikel

Verkäufe

Umsatz & Menge
Rechnungsnummer

Kunden

Lieferadresse



Artikel

Kunden

Verkaufskopf

Rechnungsnummer

Verkaufszeile

Umsatz & Menge

Lieferadresse

Artikel

Kunden

Merkerliste

Umsatz & Menge

Rechnungsnummer

Verkaufskopf

Verkaufszeile

Lieferadresse

Artikel

Kunden

Merkerliste

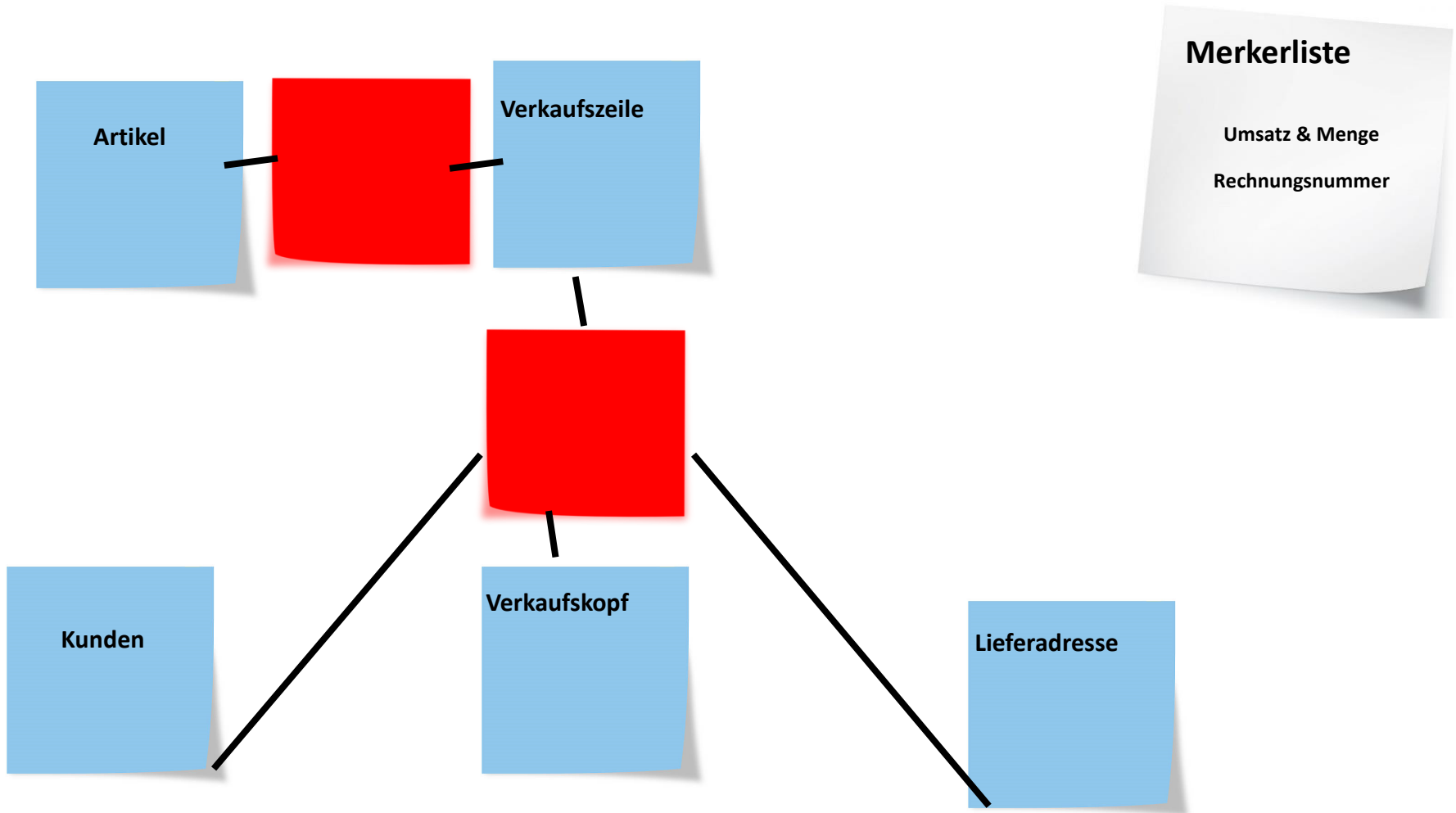
Umsatz & Menge

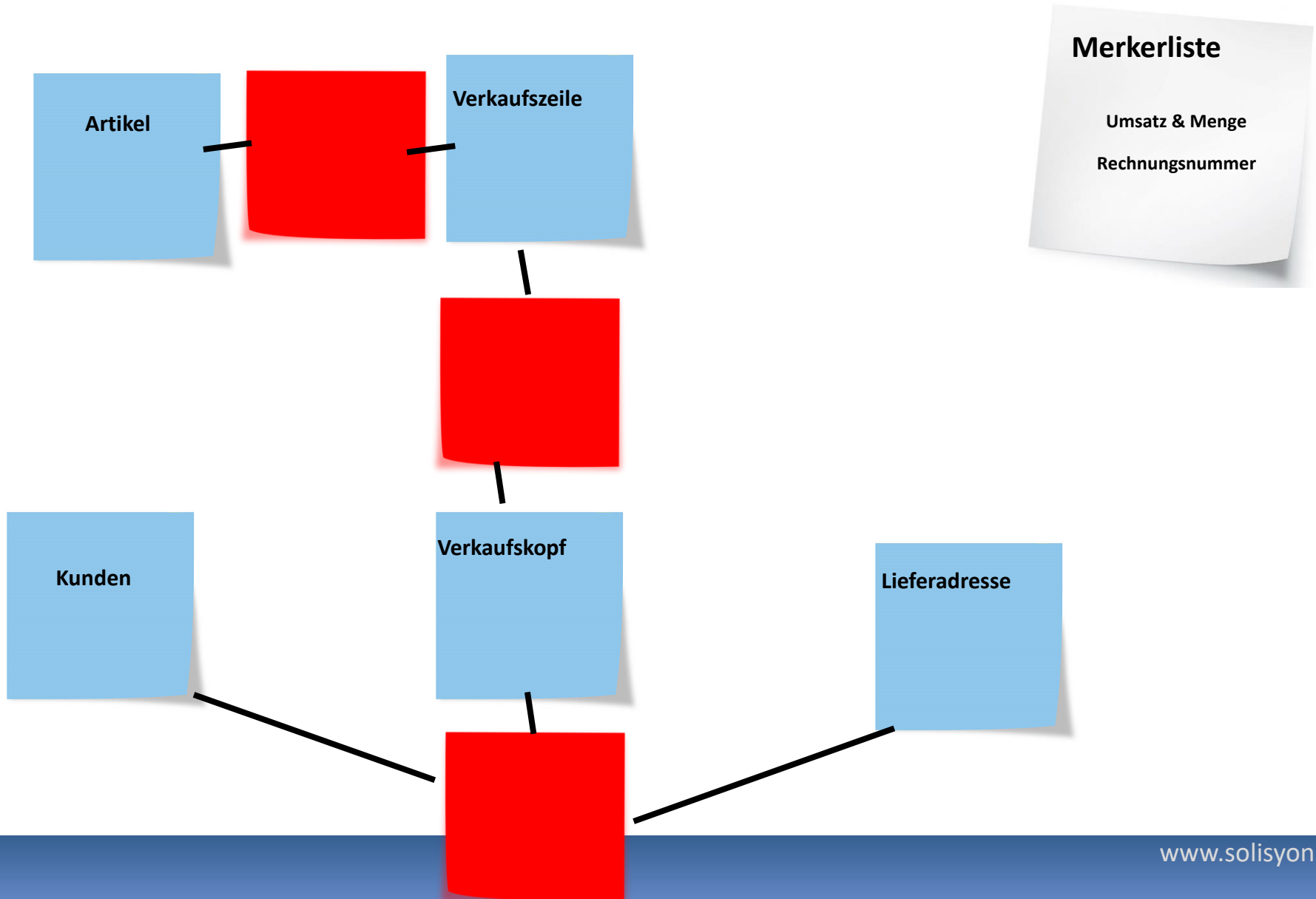
Rechnungsnummer

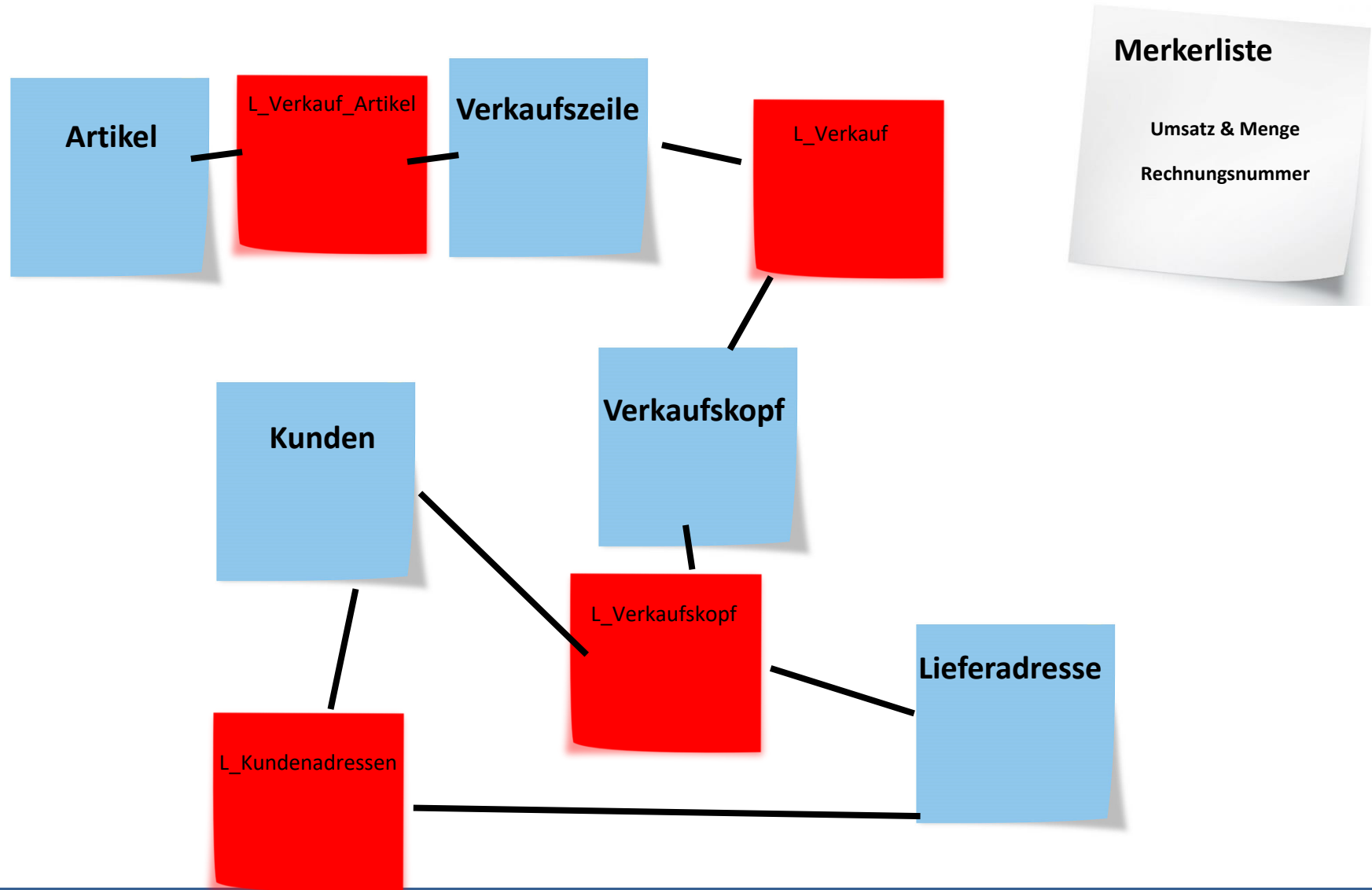
Verkaufskopf

Verkaufszeile

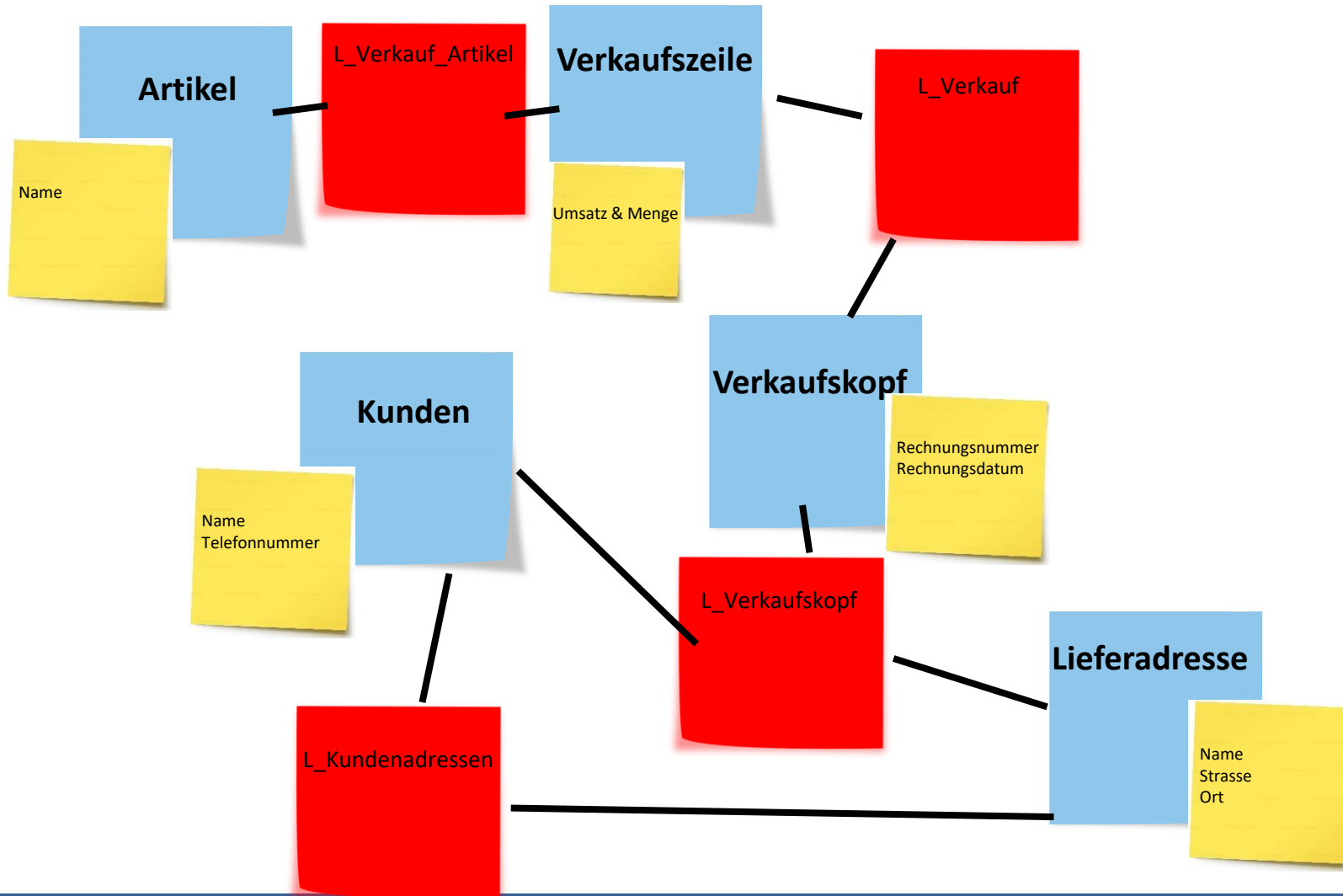
Lieferadresse

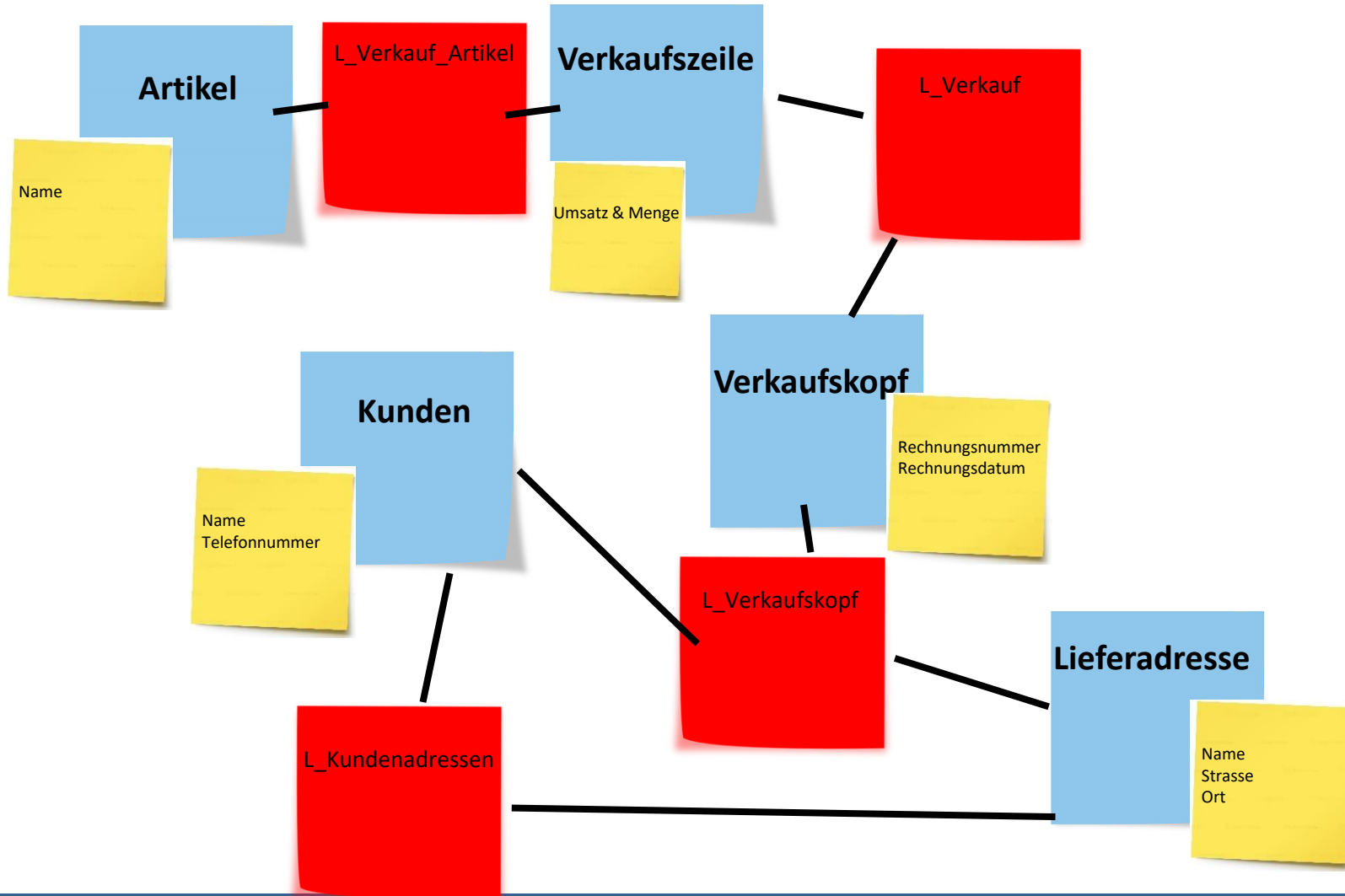


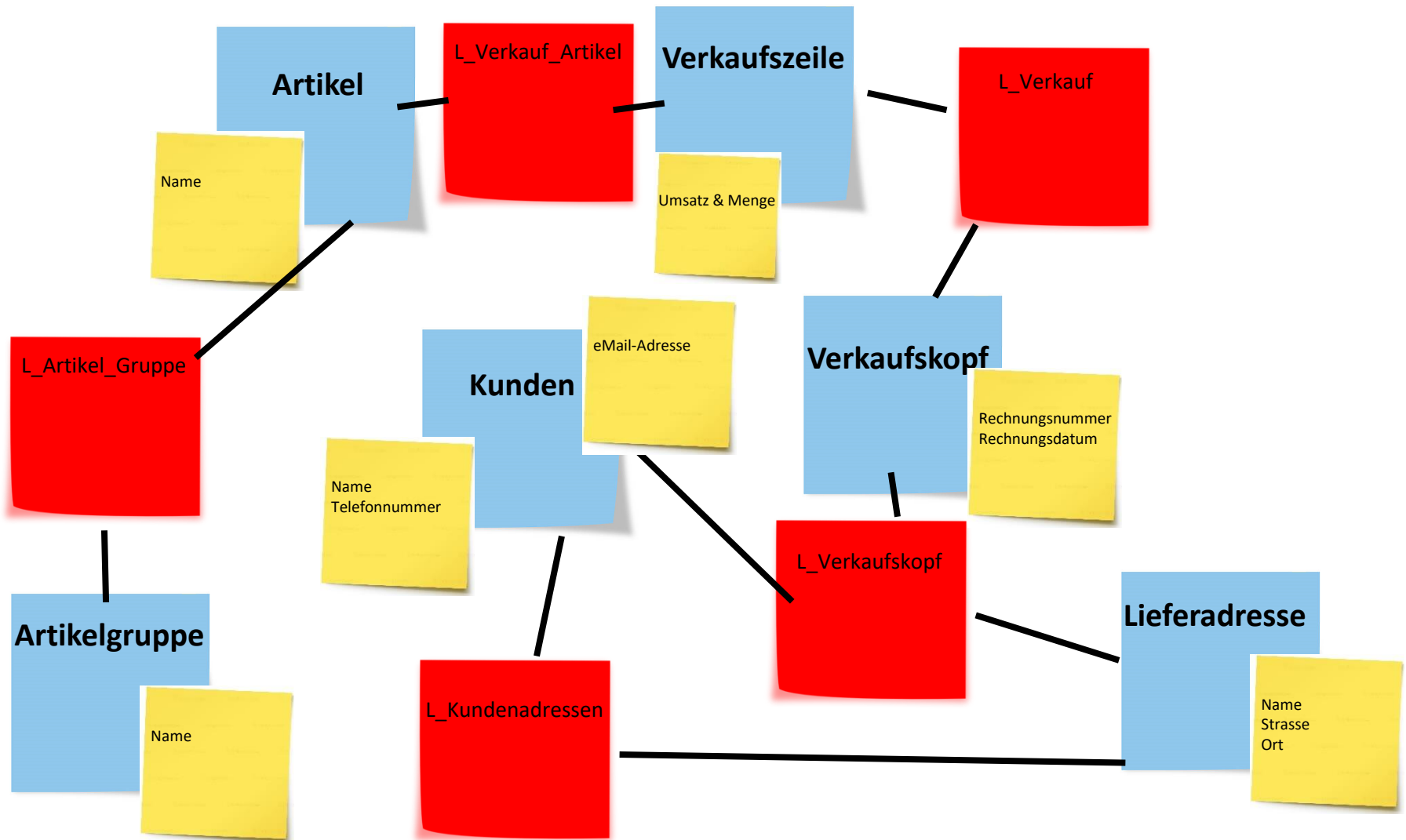








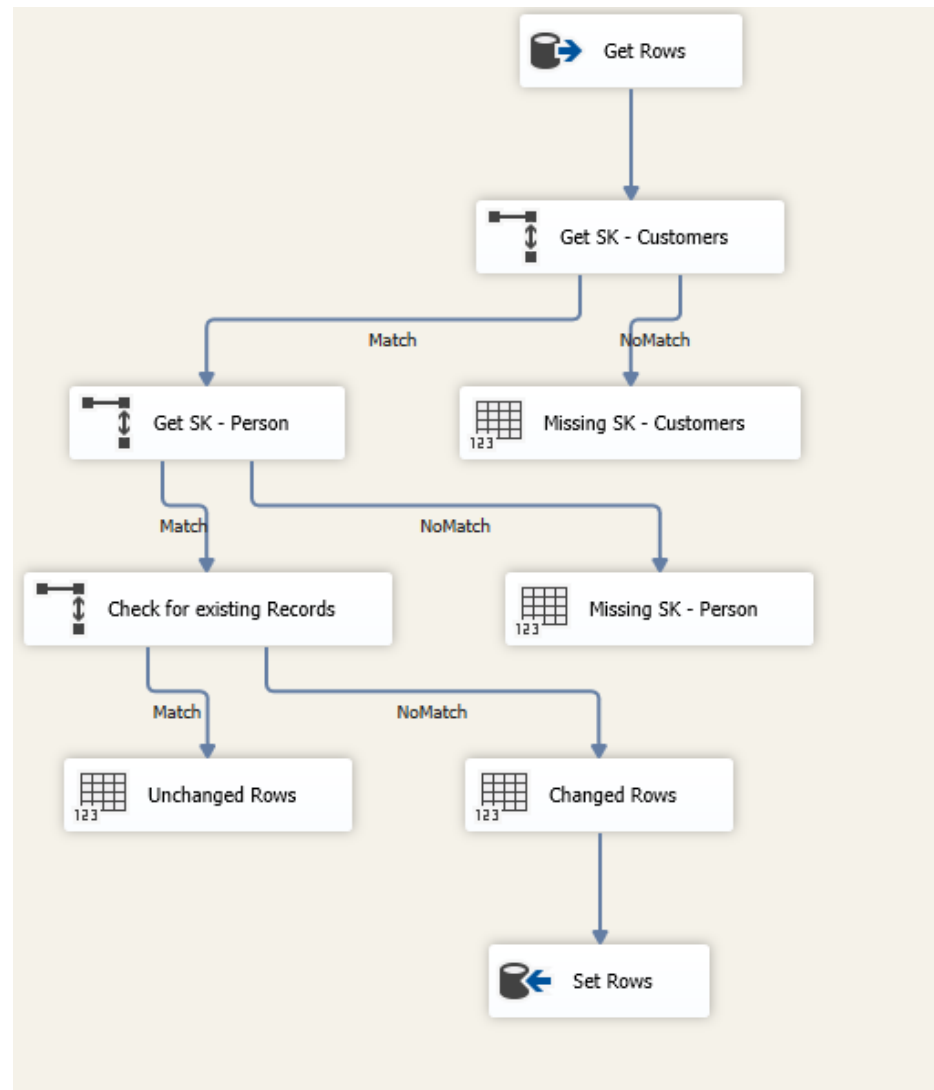




- › Nur INSERT Operationen, somit sehr gute Lade-Performance
- › Keine Anpassungen an bestehenden Tabellen, somit nicht nur incremental Load sondern auch incremental Build – in einem iterativen Prozess
- › Historie für alle Datensätze lösen Auditing und auch eventuelle spätere Point in Time Fragen
- › Geschicktes Satelliten Design hilft bei Datenschutz-Compliance

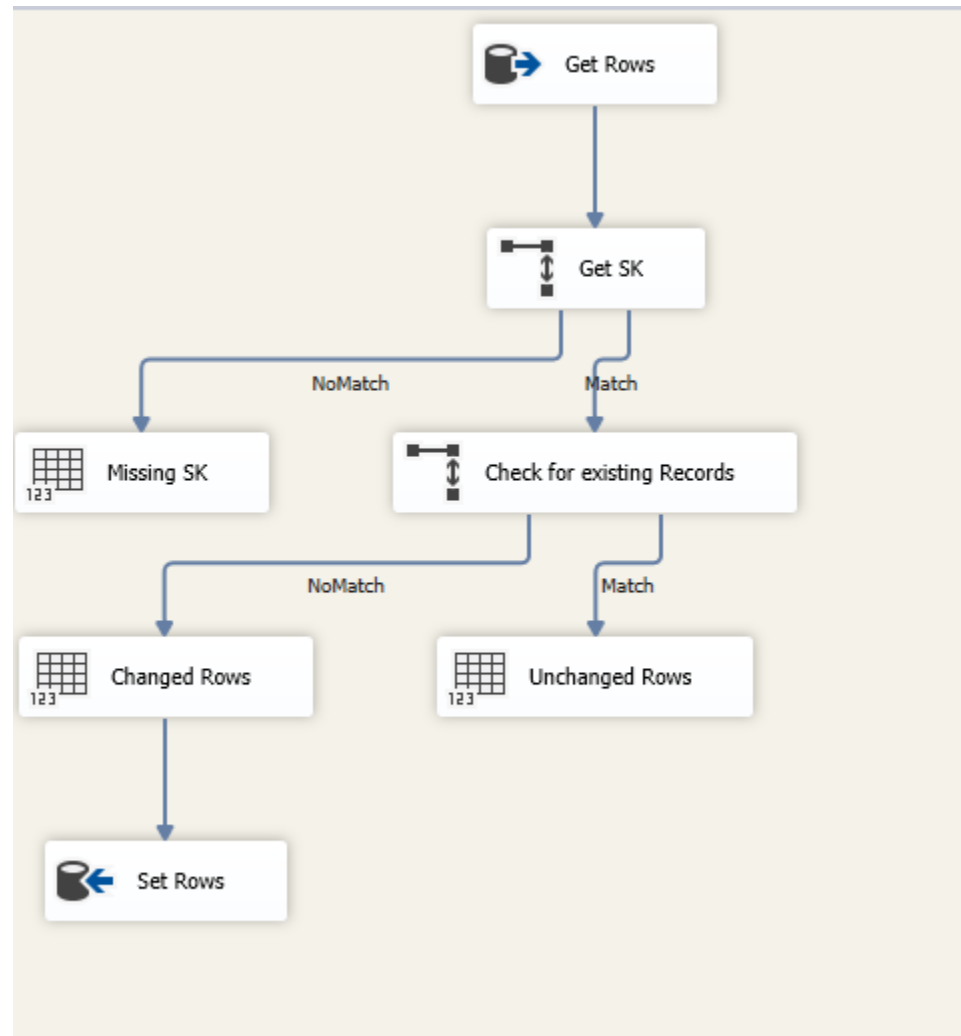
- › Viele Extra Joins bringen auch reichlich extra Abfragekosten
 - › (Im Gegenzug kann ich aber meine Datamarts sehr schön inkrementell beladen!)
- › Erfordert deutlich mehr Wissen über die Struktur des EDW um sinnvolle Abfragen zu erstellen, somit sind für ad hoc Abfragen häufiger Views erforderlich
- › Je nach (Vor-)Architektur habe ich sogar zwei Datawarehouses

```
INSERT INTO dv.[HUB_Customers] (H_Customers_BK,Source_Connection)  
SELECT BusinessKey,SRC.Source_Connection FROM stage.[AW_Sales_Customer] SRC  
LEFT JOIN dv.[HUB_Customers] HUB on SRC.BusinessKey = HUB.[H_Customers_BK]  
WHERE HUB.Load_TS IS NULL
```

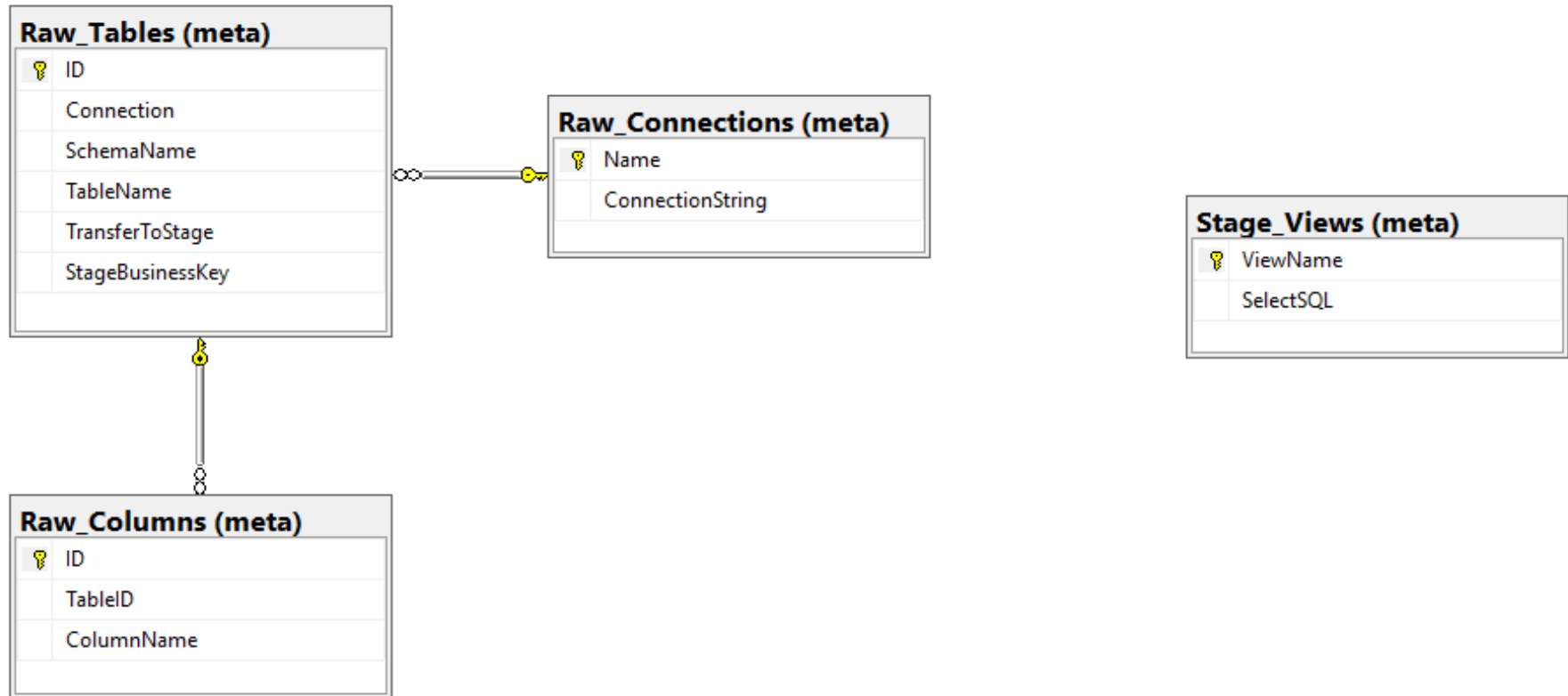



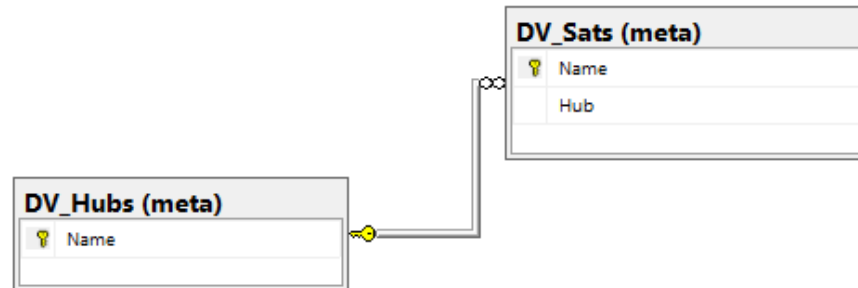


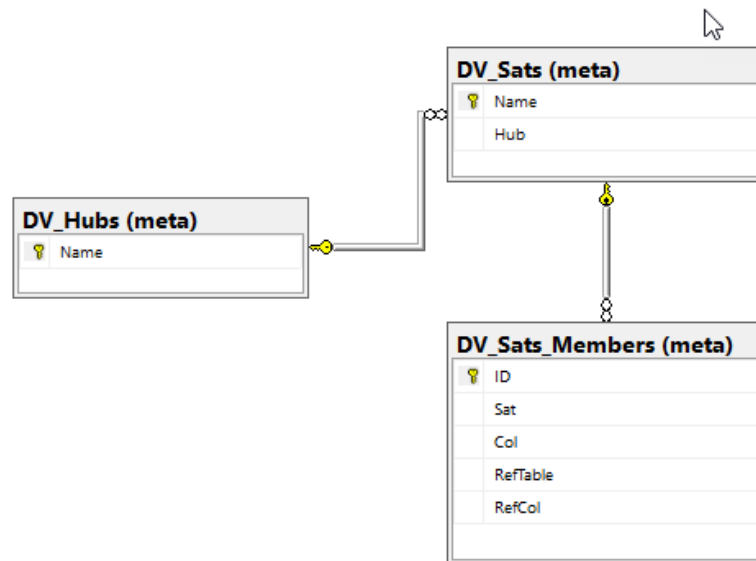
SOLISYON Wie kann das im Praxisfall aussehen? Laden von Satelliten

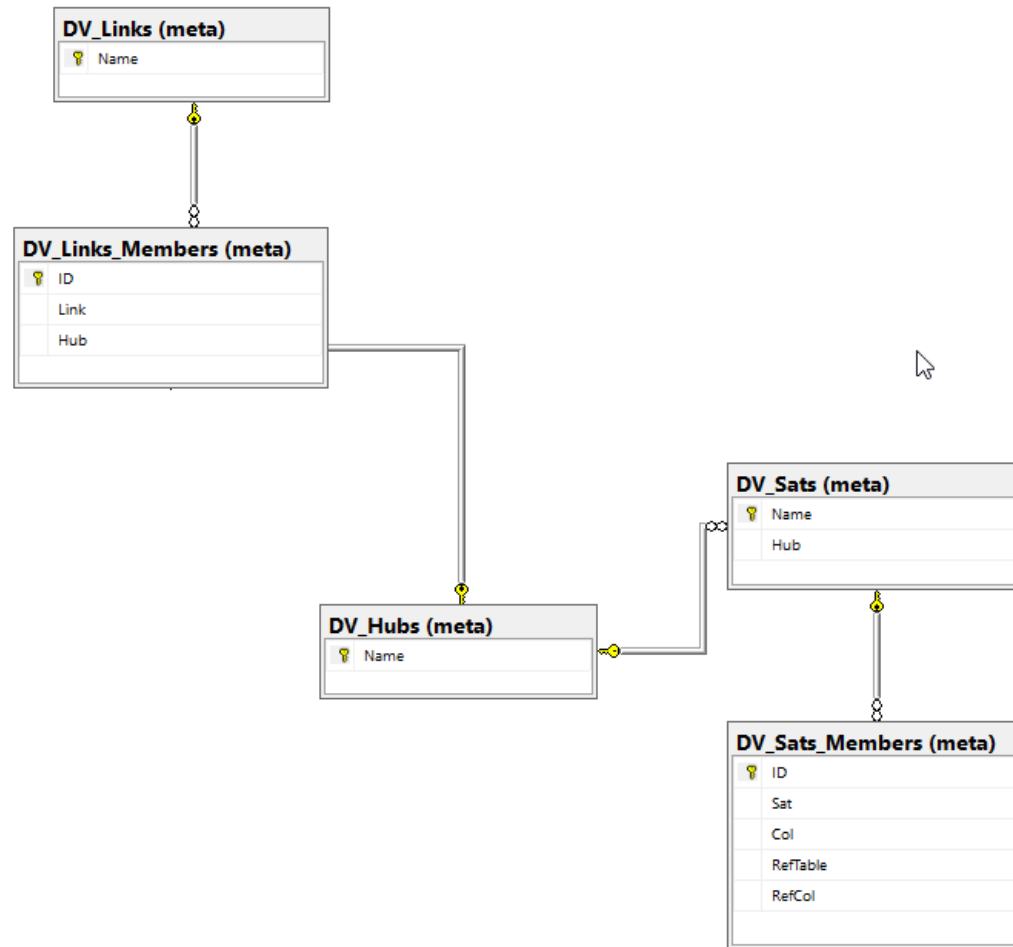


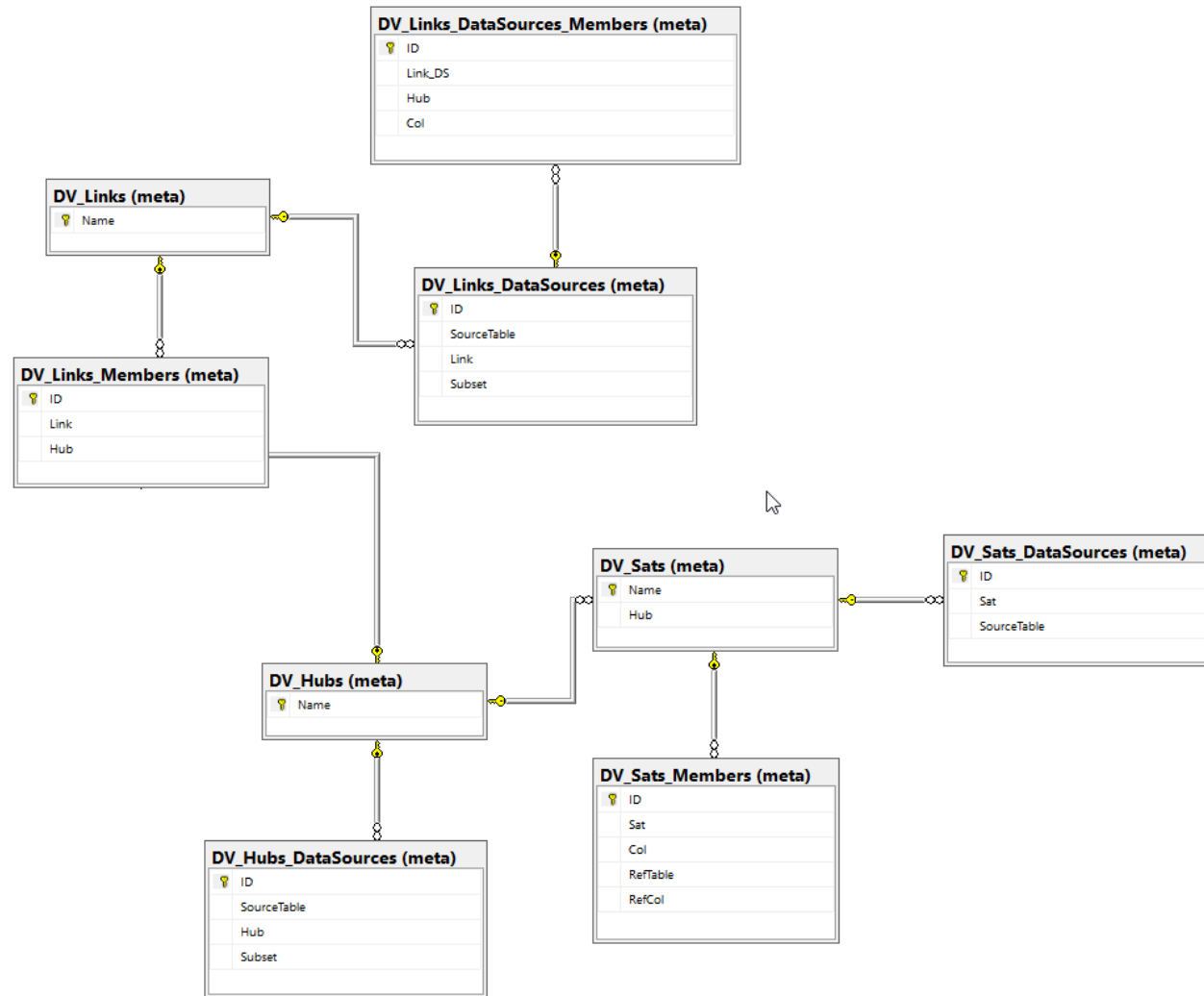
- › Nur mit Metadaten
- › Nur mit Automatisierung
- › Zum Beispiel mit: Biml 😊



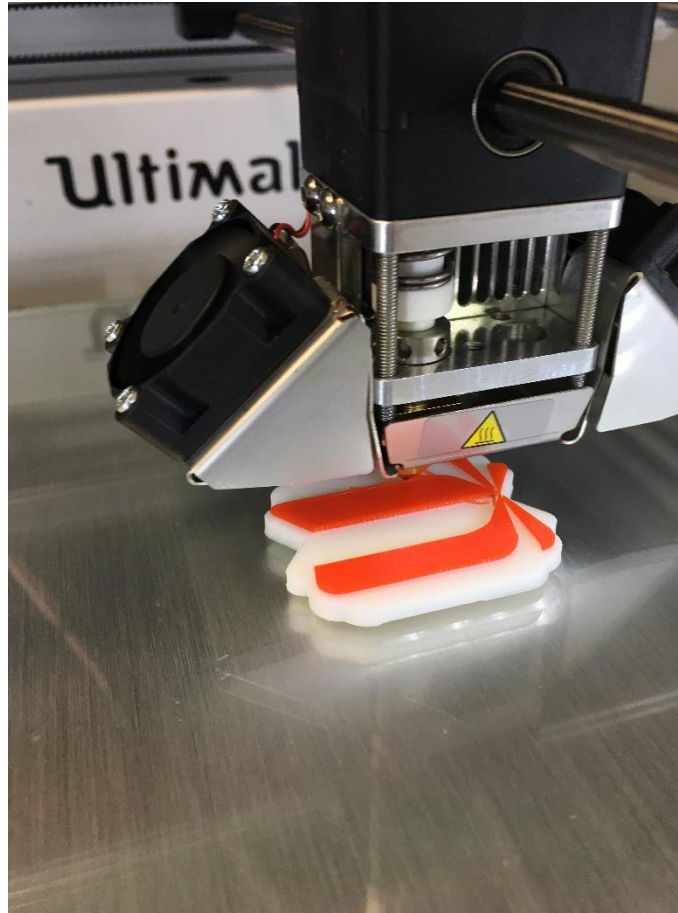








Kleines Intro in Biml



- › Biml ist eine Markup Sprache – also: XML
- › Erfunden/Entwickelt von Varigence
- › So sieht Biml aus:

```
<Biml xmlns="http://schemas.varigence.com/biml.xsd">  
  <Packages>  
    <Package Name="HelloBiml"/>  
  </Packages>  
</Biml>
```
- › Wird erst in Verbindung mit BimlScript/APIs mächtig
- › Kann neben SSIS auch für T-SQL und SSAS genutzt werden
- › Verschiedenste Frontends

 **BimlOnline**

 **BimlExpress**

 **BimlStudio**

Soweit, so gut...

Demo Time

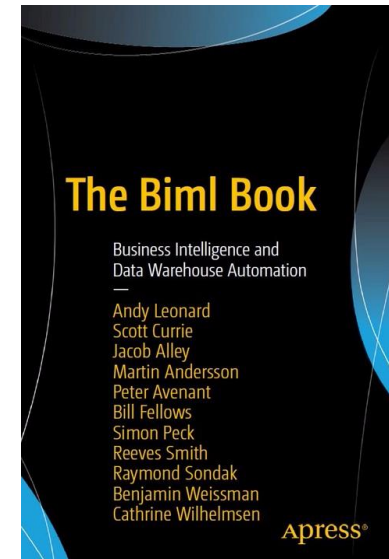
Oder auch: Biml Time!

› Data Vault:

- › <http://www.GeneseAcademy.com/>
- › <http://DataVaultBook.blogspot.com/>

› Biml:

- › <http://biml.blog/>
- › <http://www.bimlscript.com/>
- › http://sqlblog.com/blogs/andy_leonard/archive/2016/06/02/so-you-want-to-learn-more-about-biml.aspx
- › <http://www.cathrinewilhelmsen.net/biml/>



Fragen?

Gerne auch im Nachgang
per Mail (b.weissman@solisyon.de)

oder

Twitter (@bweissman)

Vielen Dank!

www.solisyon.de

