

Mach das beste daraus – Datenlieferungen an BI

Michael Müller, m2data,
24.10.2019

10. Tagung der DDVUG
e.V.

Die Gestaltung der Datenlieferung an BI



Hauptsache Daten



Opportunistisch



Der Bedarf an Arbeitsaufwand
wird nur im Nachhinein
verstanden



Anforderungen an eine Datenschnittstelle

- Vollständig
- Zuverlässig
- Robust



Typen von
Datenschnittstellen

Lieferung	Kopie	CDC	Export- programm	Message Queue
Komplett	x		x	
Delta	x	x	x	x
Event			x	x

Normalisierung: 3NF, denormalisiert, hierarchisch

Formate: JSON, Text, csv, XML, REST, GraphQL, ...

Komplett

Vollständig bis auf die eingangs erwähnten, nicht persistierten Daten. Gelöschte Daten werden über den Vergleich mit der letzten Lieferung ermittelt.

Es entsteht wenig Aufwand bei der Implementierung.

Robust, wenn rechtzeitig über Änderungen im Datenmodell informiert wird.

CDC

Vollständig bis auf die eingangs erwähnten, nicht persistierten Daten.

Es gibt gute Werkzeuge, die dies zuverlässig für das Quellsystem erledigen und keine negativen Auswirkungen auf die Performance des operativen Systems haben.

Robust, wenn rechtzeitig über Änderungen im Datenmodell informiert wird.

Programmiert

Die Verantwortung liegt komplett beim Ersteller des Programms.

Es werden vollständige und abgestimmte Ergebnisse geliefert.

Die Abstimmung erfolgt anhand eines logischen Modells. Die Erstellung kann sehr komplex werden.

Änderungen müssen rechtzeitig geplant werden, es geht Flexibilität und Geschwindigkeit verloren.

Event

Ideal für service-orientierte Architekturen (SOA).

Oft fehlt ein Push-Verfahren im Tool für den Enterprise Service Bus (ESB).

Nah an den Transaktionen und an der Ausführungszeit und somit schnell in der Auswertung.

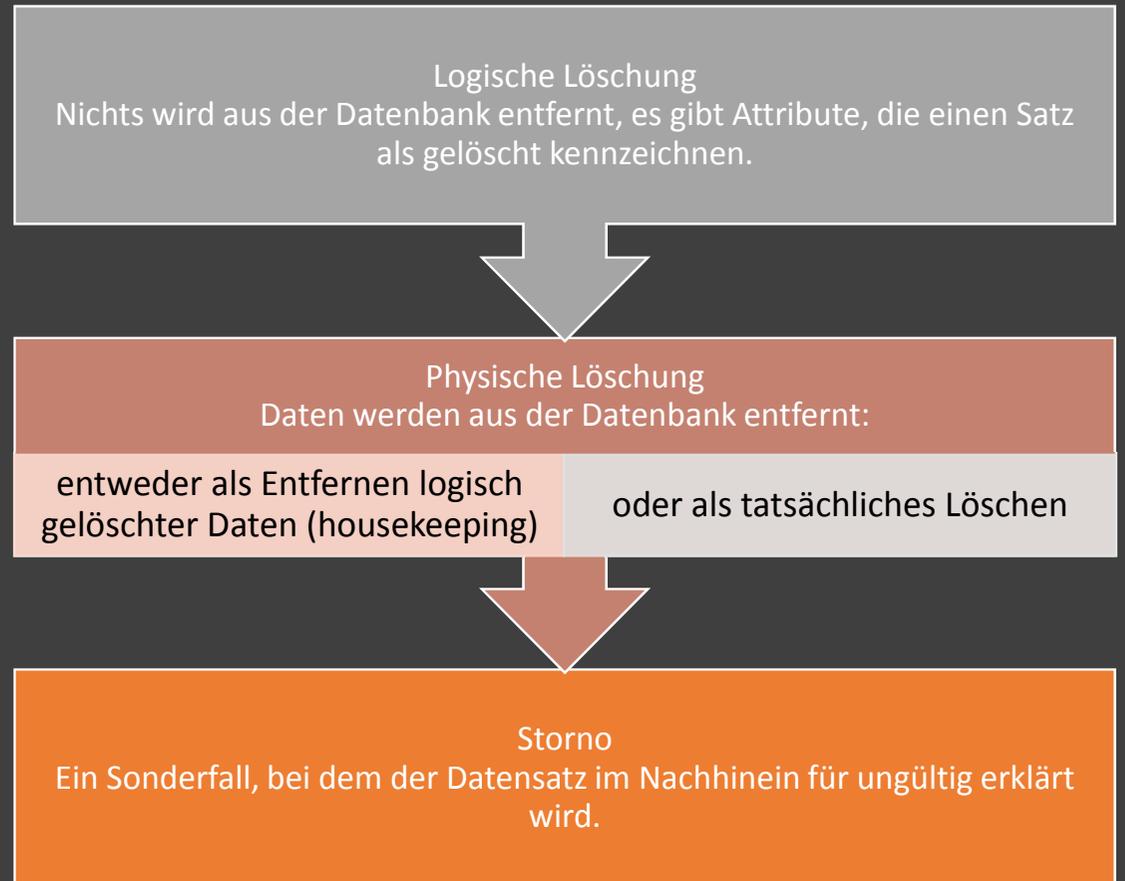
Das Sicherstellen der Vollständigkeit von den Services kann sehr komplex werden.



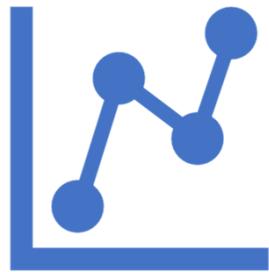
Was beinhaltet vollständig?

- Die benötigten Daten
- Korrekte Abbildung der Löschungen
- Lückenlos auch bei denormalisierten Daten

Arten von Löschungen



Warum sind Löschungen relevant?



Nicht gelöschte Daten werden
gezählt und ausgewertet.



Gilt also vor allem für
Bewegungsdaten.

Gefahr von denormalisierten Daten

Datum	CDC	Kundennr	Name	Geschlecht	Geburtsdatum	Adresse	Kundenkarte
2.1.2019	I	23457	Otto Malus	m	30.9.1978	Nürnberg	B27Z23
3.2.2019	U	23457	Otto Malus	d	30.9.1975	Nürnberg	X23B27
6.2.2019	U	23457	Otto Malus	m	30.9.1975	Fürth	X23B27



Logisches Datenmodell als Basis

Letztlich braucht man ein
komplettes Verständnis der
Daten und ihrer Beziehungen.

Schnittstellenvereinbarung



KLARE VEREINBARUNG,
SERVICE LEVEL AGREEMENT
(SLA)



UMFANG DER DATEN KLAR
BESCHRIEBEN



RECHTZEITIGE
KOMMUNIKATION DER
ÄNDERUNGEN AM
OPERATIVEN ALS TEIL DES SLA



NICHT-FUNKTIONALES
EBENFALLS VEREINBAREN:
PERFORMANCE, SICHERHEIT
UND SKALIERBARKEIT.



BESSER SIND EINFACHE
LÖSUNGEN.



STANDARDWEGE, DIE DIE
UMSETZUNG OPERATIV
EINFACH HALTEN.



Neue
zusätzliche
Anforderung

Mit mitgelieferten Metadaten
Daten automatisiert überprüfen.



Logische Schnittstelle



Trennung von operativ und dispositiv aufheben.



Ein gemeinsames logisches Datenmodell.



Die Bereitstellung erfolgt anhand dieses Datenmodells.



Das physische Modell im operativen System ist irrelevant.



Eine weit hergeholte Idee

- Anhand der Metadaten laden
- Änderungen automatisiert übernehmen
- Neue Attribute bis in den Data Mart ziehen



Word of Advice

Thou shall load
all your data
quickly and
pragmatically

Wünsche an Datenlieferungen

- Die beste Lösung für die vorhandene Umgebung.
- Schnelle und frühzeitige Reaktion auf Änderungen.
- Vollständig in Umfang und Form.
- Absicherung mit Metadaten
- Klare Kommunikationsvereinbarungen für Änderungen.
- Bewusste Entscheidungen.

