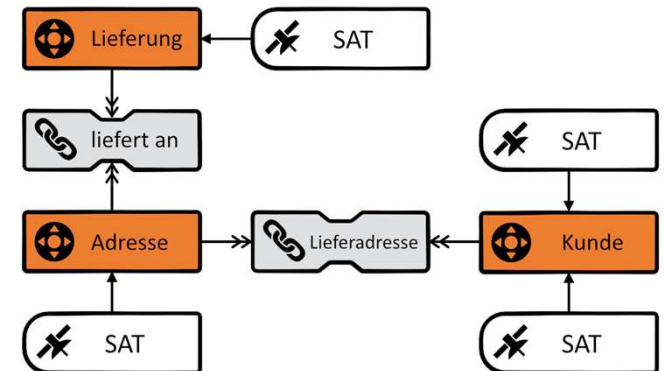


Ein Data Warehouse mit Data Vault

Toolauswahl und Einbettung ins Unternehmen



Deutschsprachige Data Vault User Group e.V.



- gegründet 2014
- im Schnitt 2 Konferenzen im Jahr
- Alles rund um Data Vault und Data Warehousing
- Beispieldatensatz für Data Vault unter CC-4.0-BY
- Vergleichsportal für Data Vault Automation



<https://ddvug.de/vortragsarchiv/>



<https://github.com/m2data/Willibald-Data>

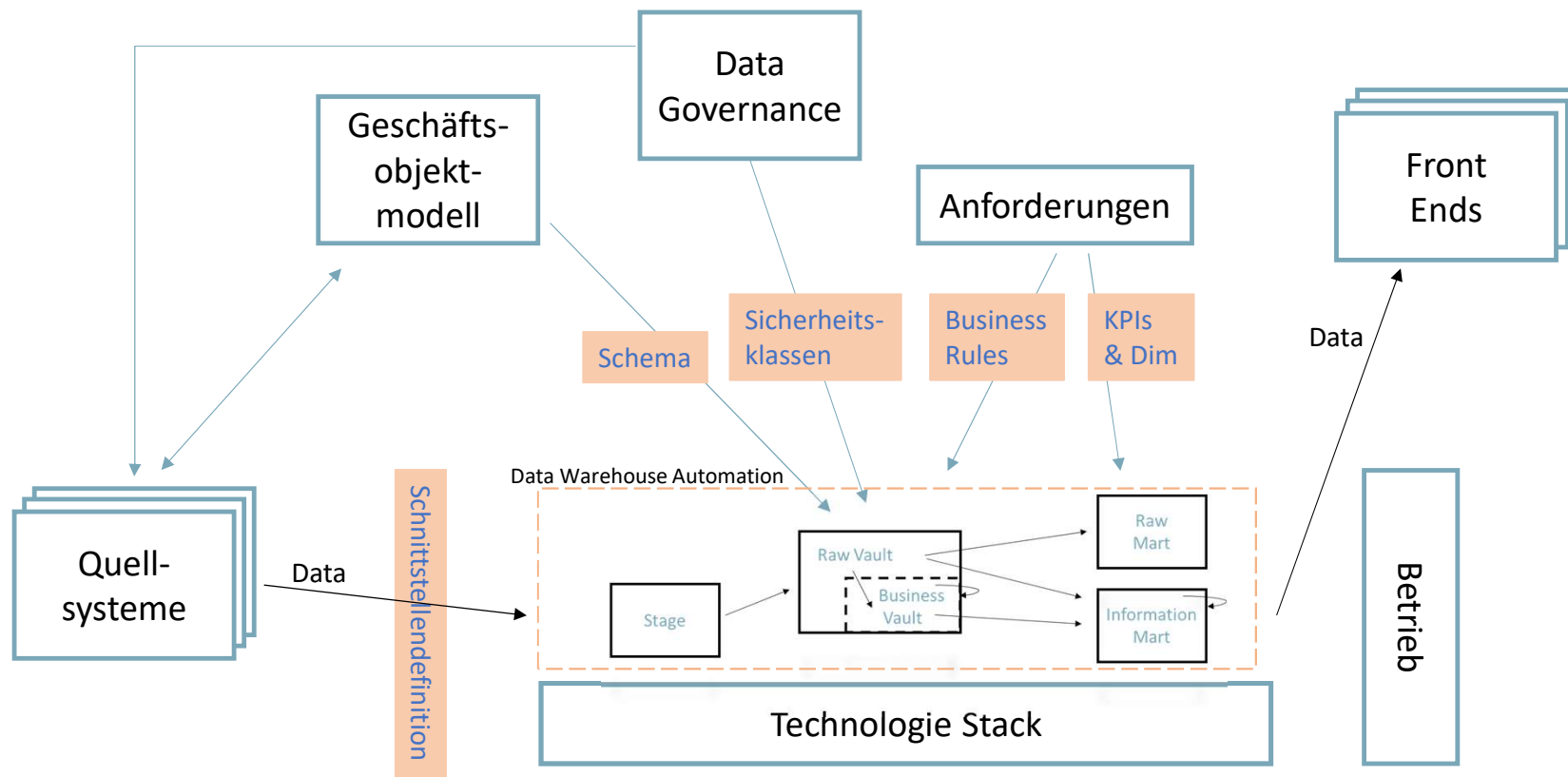


Samen- und Pflanzenhandel
WILLIBALD

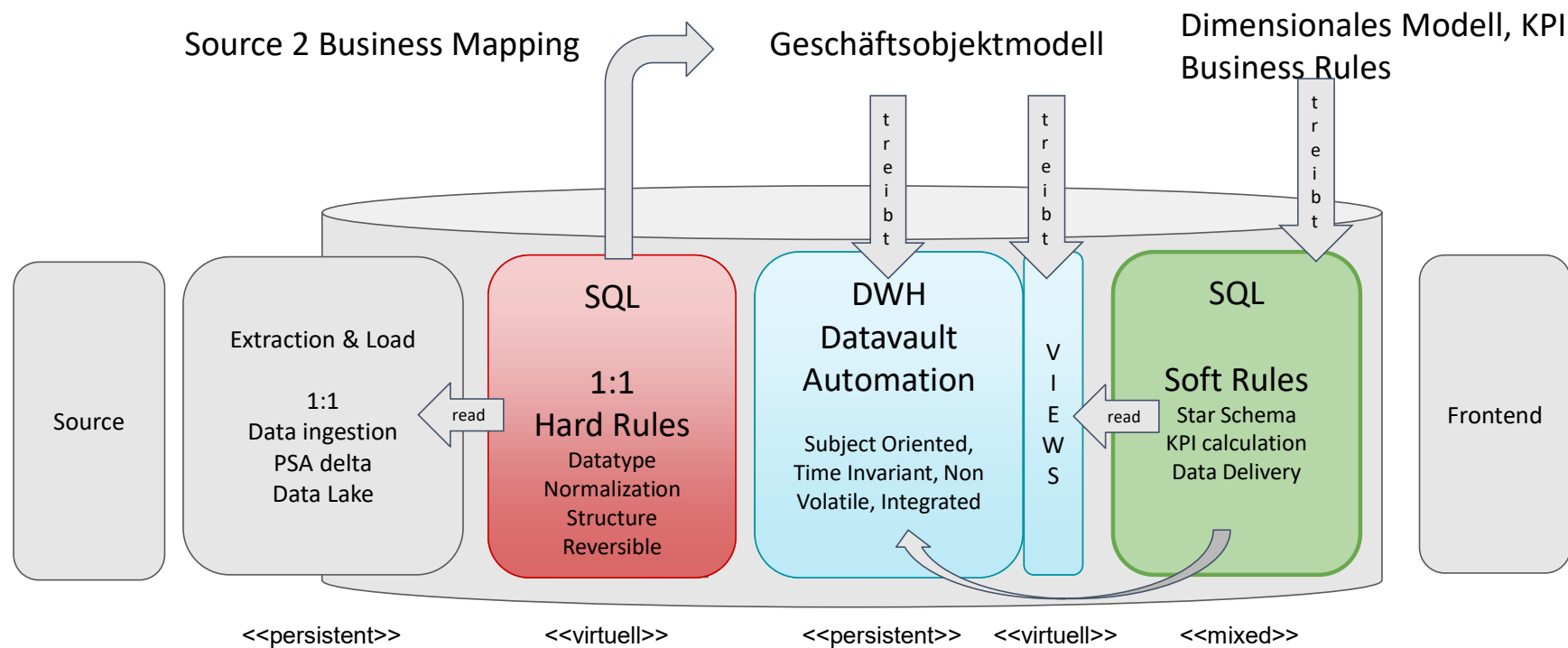


<https://dwa-compare.info/>

Einbettung Data Vault Warehouse



Datavault / DWA - Divide & Conquer in der Transformation





Toolauswahl beschleunigen

- Extrem vielfältig und schwer zu vergleichen
- Bewertung für Beschaffungsprozess
- Adaption des DWA-Tools
- Lernkurve
 - Automatisierung selbst
 - Data Vault
 - Neue Sprachen Python, SQL
 - Neue Verfahren CI/CD, Lineage, ...



Wie testet man ein DWA Tool?

- Nimm die typischen Fehler
- Integrier die typischen Probleme
- Baue ein Beispieldatenmodell und sei schlecht dabei
- Prüfe alles

“Wir liefern keine fertige Evaluation, sondern nur einen Satz von Lösungsbeschreibungen zu typischen Problemen und prüfen das Ergebnis. Bewerten muss der Nutzer selbst.”

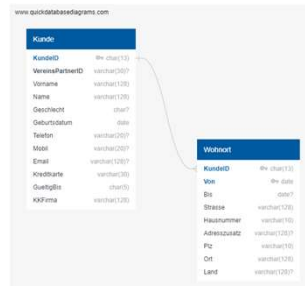
Data Vault Funktionalität (Testcases 1,2,3,4,6)



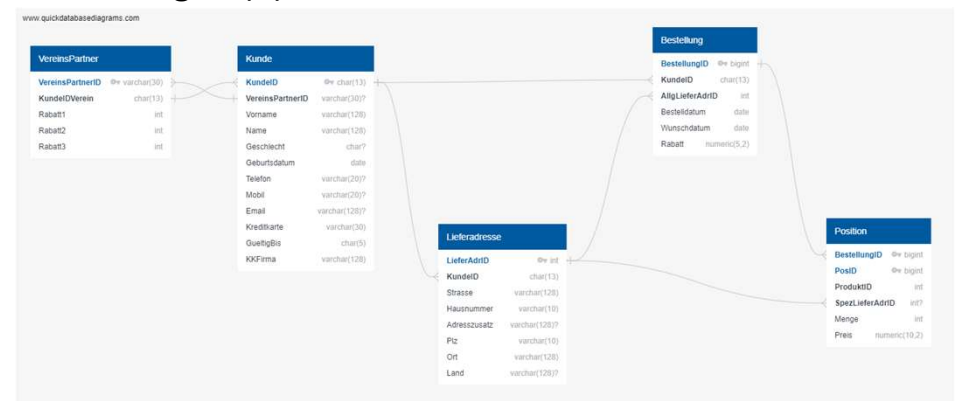
Hierarchical Link (1)



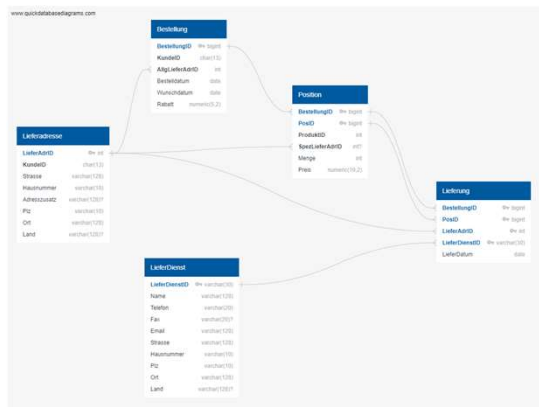
Multi-aktive Satelliten (2)



Beziehungen (3)

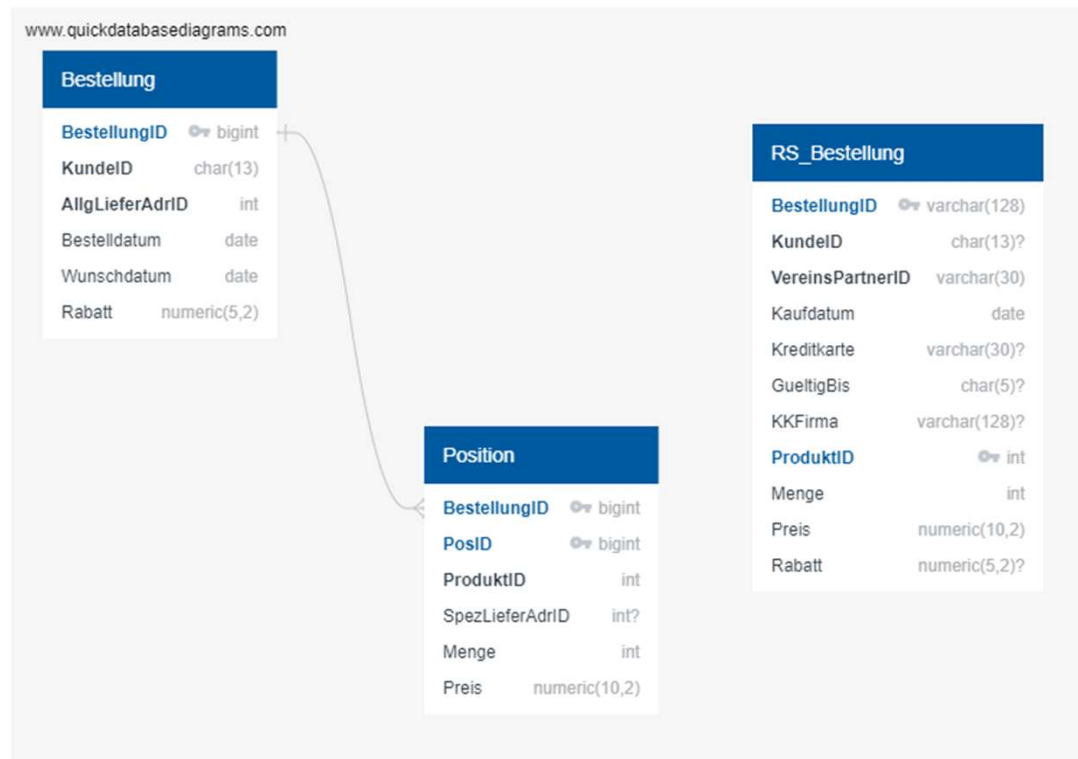


M:N Beziehung ohne ID und ist GO (4)



Referenztabellen und
historisierte Referenztabellen (6)

Frühe Integration (Testcase 5)



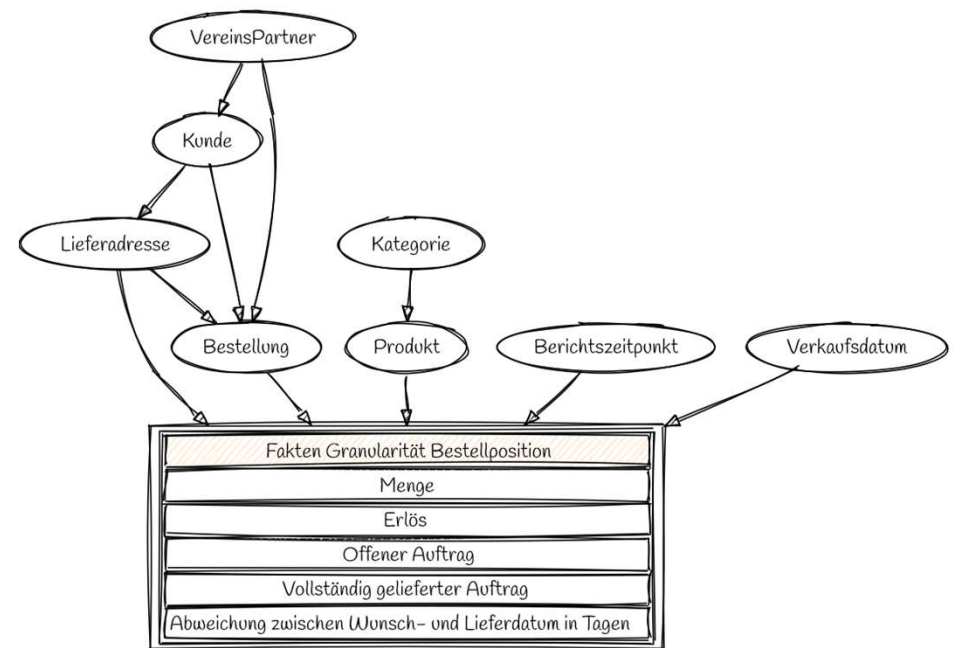
Business Rules & Data Mart (Testcase 14 und 15)



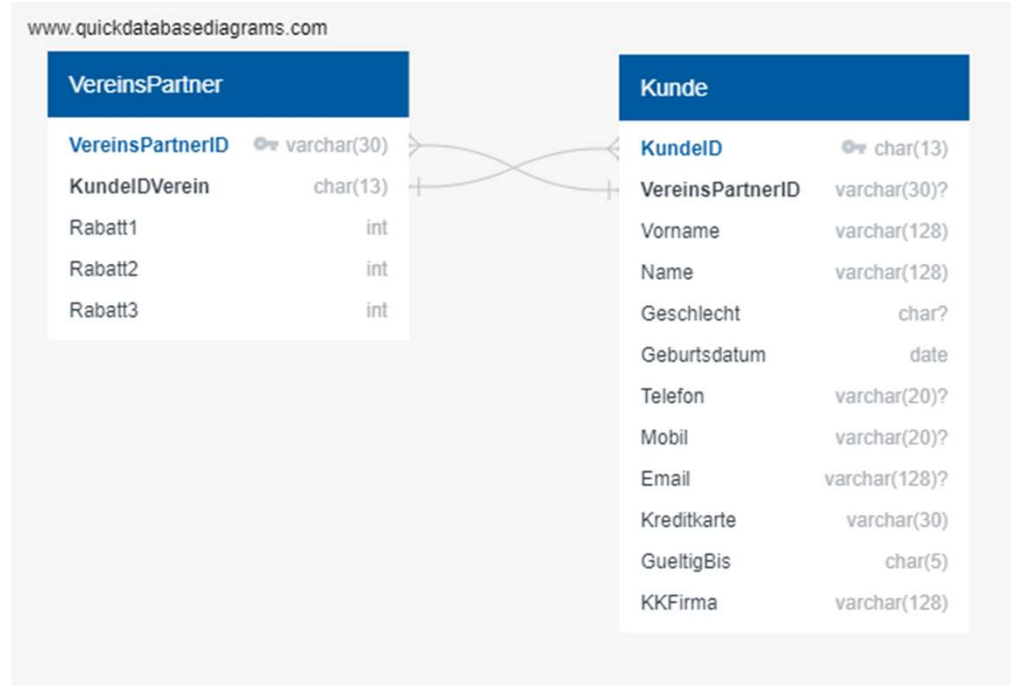
1. Die Bestellungen der Vereinspartner vereinheitlichen:

Die Bestellungen der Roadshow sind direkt mit dem Vereinspartner verbunden. Dies muss nun noch für die Bestellungen des Vereinspartners aus dem Webshop erfolgen. Hierzu werden anhand der VereinsPartner.KundeIDVerein alle Bestellungen zu diesem Kunden direkt mit dem VereinsPartner verknüpft.

2. Roadshow Bestellungen zu den Kunden zuordnen:
Wenn keine KundeID angegeben ist, wird anhand der Kreditkarte, der KKFirma und dem GueltigBis die Beziehungen von Bestellung zum Kunden ergänzt



Datenmodell



Datenqualität und –korrektheit (Testcase 7,8,9,10,11,12,13)



- Dubletten in den Ladedaten
- Sätze ohne Business Key
- Änderungen in den Kundendaten
- Löschungen im Kunden
- Lieferadressen ohne Kunde
- Löschungen von Bestellungen
- Änderungen in den Dimensionen



Querschnittsfunktionalität

- Lineage (16)
- Error Handling (17)
- Orchestration (18)
- Deployment (19)
- Scheduling (20)
- Supported Databases (21)
- Installation Requirements (22)



Was haben wir nicht getestet?

- bi-temporal
- DSGVO
- mehr Details zu CI/CD

Beschleunigung im Auswahlprozess



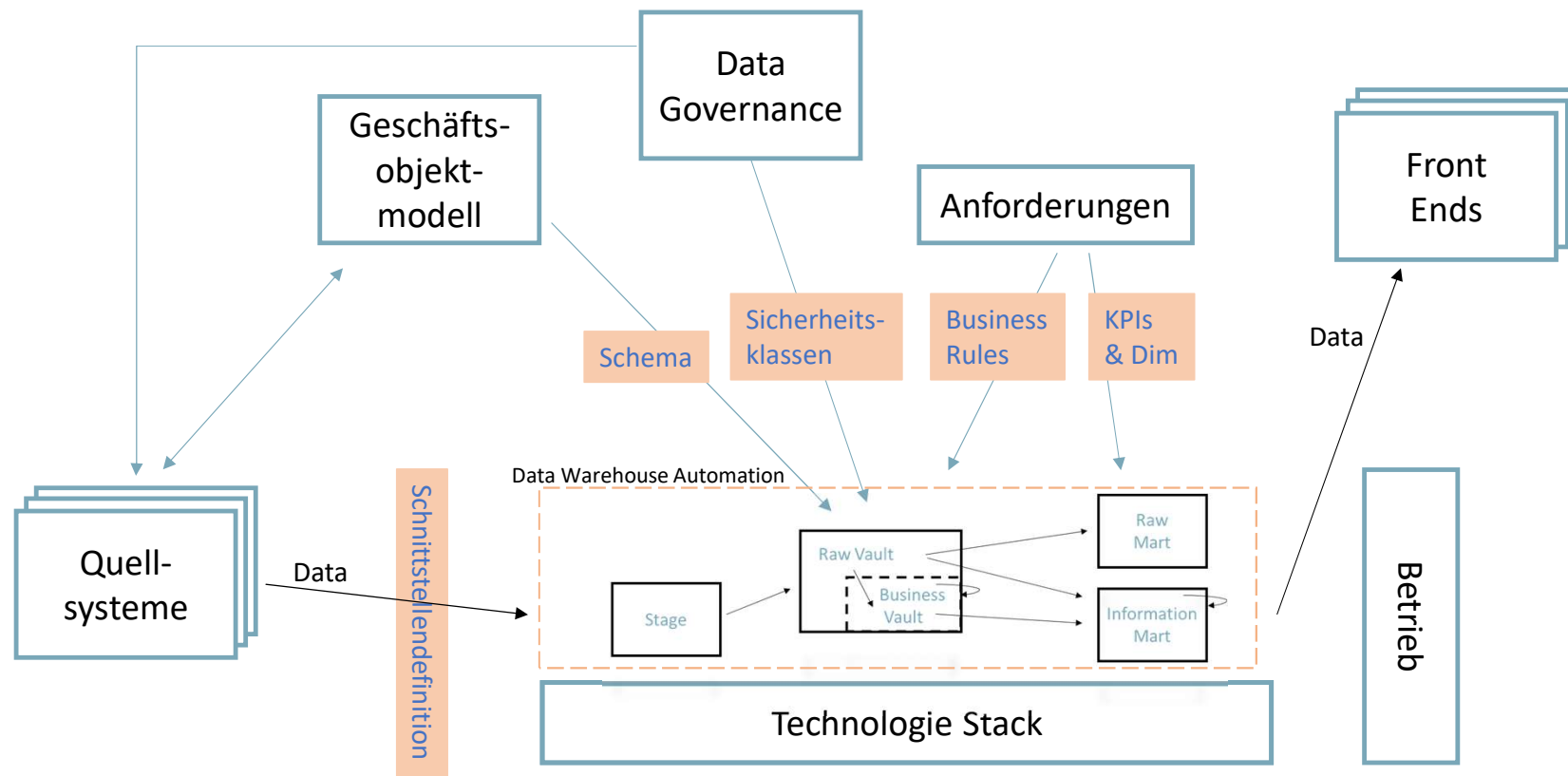
- Testcases spezifizieren
- Relevante Daten finden
- Kriterienkatalog aufsetzen
- Mögliche Werkzeuge finden
- Screening, Präsentationen
- 3 Tools auswählen
- Kriterienliste füllen



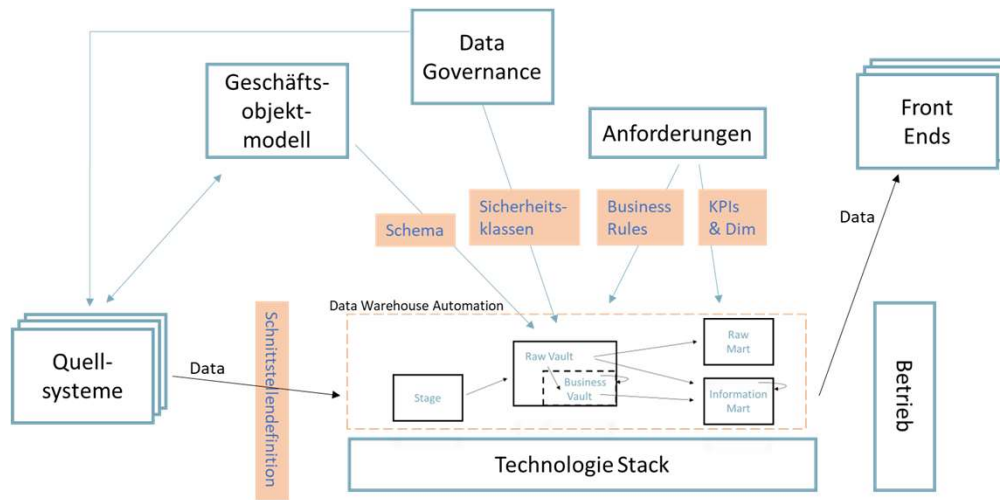
- Liste aus dwa-compare.info laden
- Liste um eigene Themen erweitern
- Veto-Kriterien identifizieren (was geht nicht?)
- Punkte aus der Liste gewichten
- Bewerten
- Auszählen
- Entscheiden
- POC mit eigenen Daten machen
- Einsetzen (kaufen)

➤ Zeit und Aufwand gespart

Und wenn sie nicht gestorben sind...

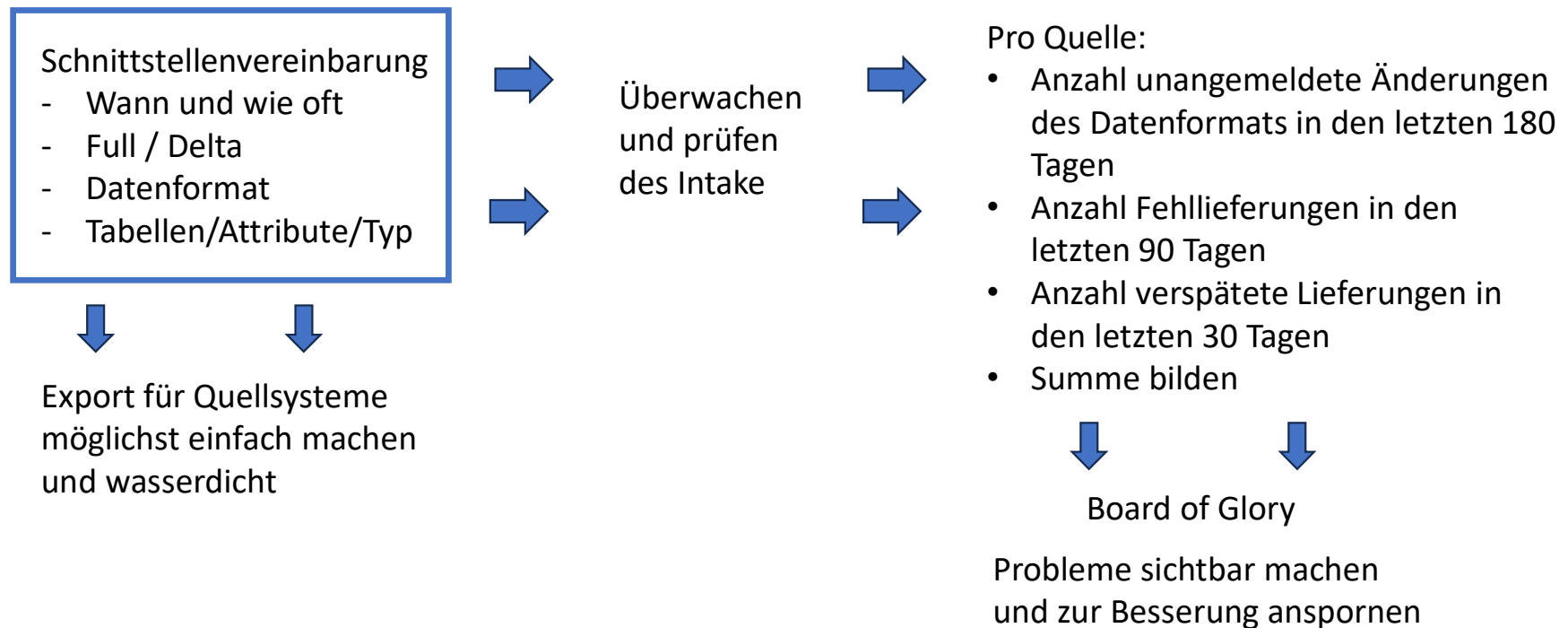


Gut abstimmen und immer anpassen



- Das zentrale Problem dieser Themen ist die geteilte Verantwortung
- Die Ergebnisse werden für das DWH gebraucht
- Der jeweils andere braucht hoffentlich das DWH
- Die Arbeitsteilung ist verhandelbar

Schnittstellen



Fachbereich einbinden – was ist die Motivation?



- Gemeinsame Verantwortung
 - Geschäftsobjektmodell als gemeinsame Basis
 - Zum Beispiel mit Ensemble Logical Model Workshops
 - Motivation Self Service
- Anforderungen
 - KPI und Dimensionen formulieren, wie geht das?
 - Geschäftsregeln beschreiben, klassisches Anforderungsmanagement für Software
- Data Governance
 - Definition der Basis für die Steuerung der Geschäftsprozesse
 - Sicherheitsklassen, Regulatorik
 - Schnittstellen (Data Contract), Quellsysteme
- Metadatenintegration
 - Strukturierte Erfassung aller Daten, Regeln, Modelle und Anforderungen
 - Aktive Metadaten durch Data Catalog oder Data Product Marketplace

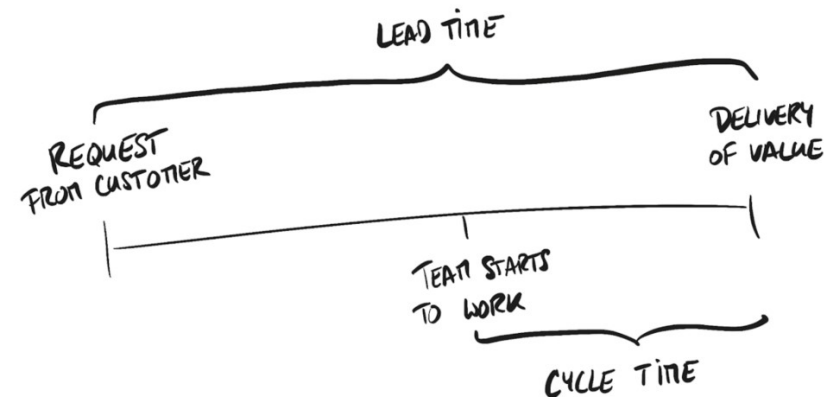


Samen- und Pflanzenhandel
WILLIBALD

Die Wartungsfalle



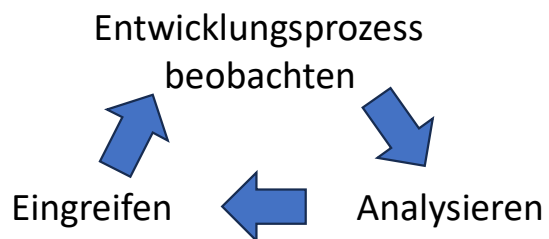
- Technische Schulden reduzieren
 - Data Vault an Geschäftsobjektmodell ausrichten
 - Transformation in Teilschritten (Hardrule, Data Vault, Softrule)
 - Inkrementelle Entwicklung / auf Erweiterbarkeit setzen / kleine Scopes
- ... und verwalten
 - Messen der Entwicklungsgeschwindigkeit
 - Gibt es ein Budget für technische Verbesserungen?
 - Erfassen der Ratio Wartung zu Weiterentwicklung



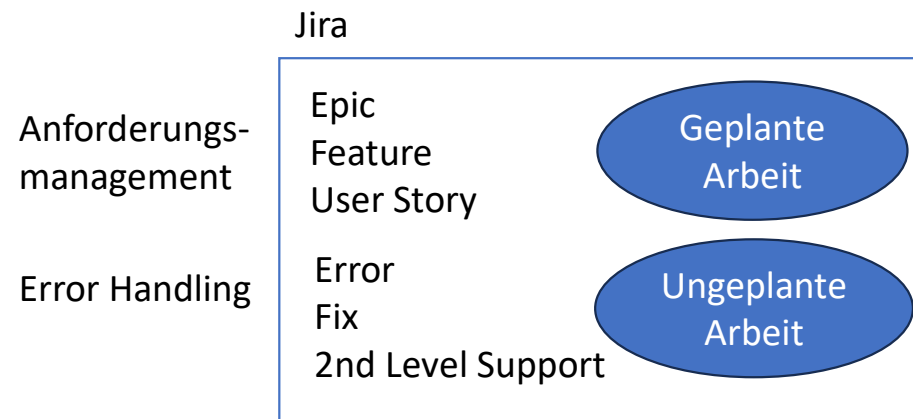
Ungeplante Arbeit tracken



- Ungeplante Arbeit ist ein Hindernis für schnelle Entwicklung



- Wo finden wir die Fakten?





Data Vault To Go

- ELM Modellierung – Integration des Fachbereichs
- Data Vault 2.0 Implementierung
- HandsOn Workshops
 - Datavault Builder, Carsten Schweiger
 - DVG, Joerg Stahnke
 - Vaultspeer, Jonas DeKeuster
 - Wherescape, Paul Watson-Gover
- - Paneldiskussion - Erfahrungsaustausch Use Cases



TDWI MÜNCHEN 2024

11. - 13. Juni 2024

Wie hat Ihnen die Session gefallen?

Geben Sie uns gerne Ihr Feedback



Mi 8.1

12.06.2024

09:00-10:00

Ein Data Warehouse mit Data Vault – Toolauswahl und Einbettung ins Unternehmen

