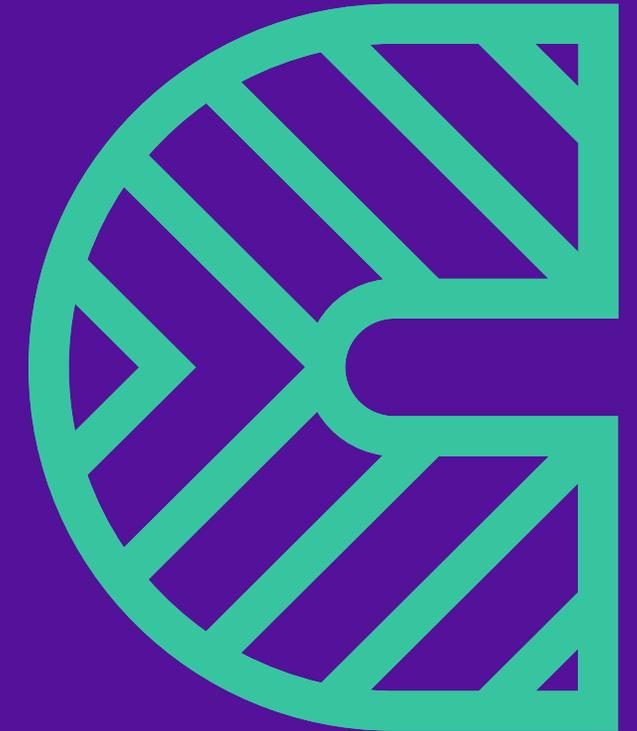




vollständig automatisierte Datenprojekte

DDVUG Frühjahrskonferenz 2025



Zielsetzung

Automatisierung ermöglicht Folgendes

- » Um 8:00 Uhr erhalten Sie eine Anforderung, die Berechnung einer Kennziffer für bestimmte Sonderfälle anzupassen
- » Um 16:00 Uhr ist diese Anforderung umgesetzt und im Produktionsbetrieb
- » es werden trotzdem keine Kompromisse gemacht
 - Codequalität
 - Dokumentationen
 - Data-Quality
 - DSGVO
 - revisionssicherer Test
 - Risikominimierung
 - u.a.

Vortragsthema:

- » Welche Projektschritte müssen dafür automatisiert werden?



Anforderung

Anforderung

Aufgaben

- » Anforderung formulieren

Beispiel:

- » Anpassung der Kennziffer für Nettoerträge
- » Kunden, die in den letzten 5 Jahren jeweils mehr als 1 Mio. Umsatz hatten erhalten zukünftig 2% Treuerabatt

Automatisierung



Analyse

Anforderung

Analyse

Aufgaben

Fragen beantworten:

- » Sind alle Informationen, die ich für die neue Kennzifferberechnung benötige, vorhanden?
- » In welchen Feldern finde ich diese Informationen?

Automatisierung

Bereitstellen Datenkatalog

- » grafische Modellansichten für verschiedene Detailgrade und Themen
 - Unternehmensmodell
 - grobe Bereichsübersichten
 - themenbezogene Detailübersichten
- » inklusive fachlicher Feldnamen + Beschreibungen der Feldinhalte
- » auswertbar mit BI-Tools
- » Versorgung Data Governance Tools



Impact

Anforderung

Analyse

Impact

Aufgaben

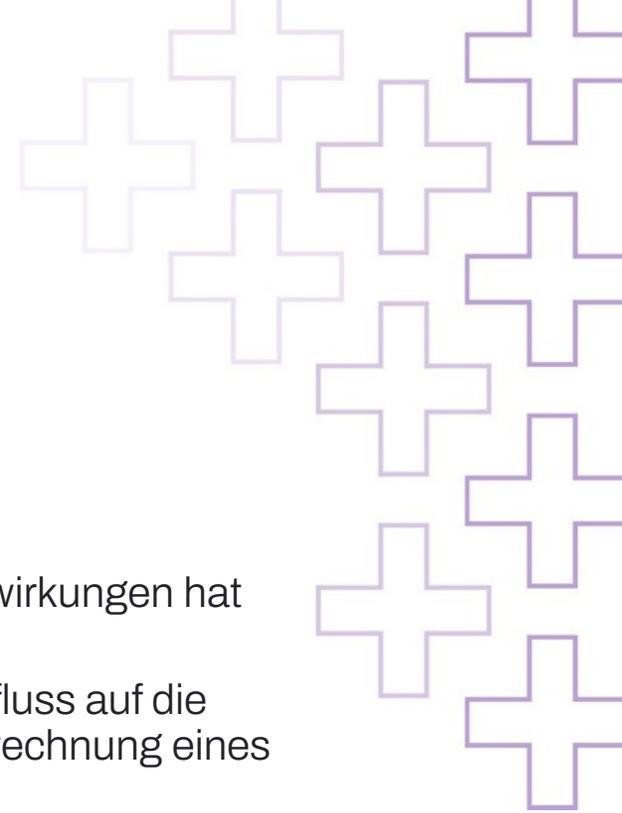
Fragen beantworten:

- » Welche direkten oder indirekten Auswirkungen hat die geplante Änderung auf andere Daten?
- » Welche Abteilungen sind betroffen?
- » Wer muss bei der Abnahme beteiligt werden?

Automatisierung

Bereitstellen Data- Lineage

- » auf Tabellen- und Feldebene
- » zwei Varianten:
 - Impactanalyse: Welche Auswirkungen hat eine Änderung?
 - Sourceanalyse: Was hat Einfluss auf die Erstellung einer Tabelle / Berechnung eines Feldes?
- » auswertbar mit BI-Tools



Konzeption

Anforderung

Analyse

Impact

Konzeption

Aufgaben

fachliche Vorgaben metadatenbasiert erstellen / anpassen

- » erfolgt durch Datenmodellierer mit fachlichem KnowHow Schwerpunkt
- » Geschäftsregel für Nettoertrag anpassen
- » neue Teilregel:
 - wenn Umsatz >1 Mio. in letzten 5 Jahren
 - dann Rabatt = Rabatt + 2%

**letzter Schritt mit
„menschlichem“ Aufwand**

Automatisierung

- » komplett metadatenbasiert
- » automatisches Einlesen von Quellsystemstrukturen
- » automatisches Erzeugen RawVault
- » Unterstützung bei Modellierung DataMarts, BusinessVault und Geschäftsregeln
 - sofortige Validierung bei Eingabe
 - grafische Anzeige der Datenmodelle
 - vorbereitete Templates
 - nur fachliche Daten-Aspekte modelliert (keine technische DataVault-Felder, Effektivitätssatelliten, bitemporale Historie, Bridgetabellen u.a.)
 - automatische Ergänzung aller technischen Code-Aspekte (z.B. Fehlerhandling, Protokollierung u.a.)
 - leicht ergänzbar



Programmierung

Anforderung

Analyse

Impact

Konzeption

~~Programmierung~~

Aufgaben

Umsetzung der Konzeption in Code

- » Anpassung des Datenmodells
- » Anpassung der Beladungsroutinen
- » Anpassung der fachlichen Logik

Automatisierung

- » komplettes Erzeugen sämtlicher Codeteile
 - technische Felder hinzufügen
 - SQL-Code
 - Programm Code für Beladungen
 - inklusive Fehlerhandling, Protokollierung u.a.
- » Entwickler werden im Projekt nicht mehr benötigt
- » Beschleunigung der Abläufe durch
 - weniger beteiligte Personen !
 - weniger Prozessschritte



Steuerung

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Aufgaben

Erstellen einer Jobsteuerung

- » fachlich notwendige Reihenfolge beachten
- » hoher Parallelisierungsgrad
- » schnelle Verarbeitung der Daten
- » keine Überlastung der Infrastruktur

Automatisierung

- » komplettes Erzeugen aller Jobabhängigkeiten
 - abgeleitet aus dem Data-Lineage
 - keinerlei manueller Konfigurationsaufwand
- » Bereitstellung aller Abhängigkeiten in Tabellenform
- » automatische Auswertung dieser Tabellen während der DWH-Beladung
- » Statushandling
 - jederzeit Überblick, wann welche Beladung lief
 - Fortschritt der aktuellen Beladung



Dokumentation

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Aufgaben

Aktualisieren sämtlicher Dokumentationen

- » Datenkatalog
- » Data-Lineage
- » Geschäftsregeln
- » u.a.

Automatisierung

- » komplettes Erzeugen der Dokumentationen
- » Bereitstellung in verschiedenen Formaten
 - grafische Übersichten in verschiedenen Detailgraden
 - Unternehmensmodell
 - grobe Bereichübersichten
 - Details zu speziellen Themen
 - in Tabellen zur Auswertung mit BI-Tools
 - Excel Sheets
 - evtl. Word oder Html
 - Versorgung bestehender Governance Tools
- » inklusive fachlicher Daten und Beschreibungen



bitemporale Historie

technische Zeit: Kenntnisnahme im DWH + fachliche Zeit: fachliche Gültigkeit

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Historie

Aufgaben

bitemporale Historie aller Daten

- » ermöglicht trotz fachlich rückwirkender Änderungen die konsistente Reproduktion jeglicher Datenstände
 - z.B. Unternehmensbilanzen
- » rückwirkende fachliche Korrekturen stehen jederzeit bereit
 - Korrektur Fehleingaben
 - Änderungen Logik der Kennzifferberechnung
 - rückwirkende Schadensmeldungen bei Versicherungen
- » KI-Modelle trainieren / kontrollieren
 - Was sagt mein Modell mit den vor 4 Wochen bekannten Daten für heute voraus? Wie gut trifft dies die Realität?

Automatisierung

- » komplettes Erzeugen aller für bitemporale Historie erforderlichen Komponenten
 - Datenmodell
 - Beladungsroutinen
 - korrekte Joins zwischen mehreren Tabellen mit jeweils unterschiedlicher bitemporaler Historie
- » einfache Abfragemöglichkeiten durch Anwender
 - Bereitstellung von Views mit eindimensionaler fachlich historisierter Sicht zu einem bestimmten Kenntniszeitpunkt
 - zielgruppenspezifische Zugriffsschicht mit verschiedenen (einfachen) Historisierungen



Migration

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Historie

Migration

Aufgaben

Migration bei Änderungen

» Datenmodellstrukturen

» Programmcode

» Dateninhalte in fachlich rückwirkender Historie neu berechnen

- bei angepasster Kennzifferberechnung
- neue Tabellen / Felder
- verschobene Felder

Automatisierung

» Erstellen und Ausführen „ALTER TABLE“ Befehle

» Austausch der geänderten Programme

» geänderte Programme aktualisieren automatisch komplette Fachhistorie

- nur beim ersten Start nach Änderung
- keine Konfiguration durch Anwender notwendig
- keine Änderung alter technischer Kenntnisstände in bitemporaler Historie



DSGVO

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Historie

Migration

DSGVO

Aufgaben

Umsetzung DSGVO

- » personenbezogene Daten schützen
- » sparsame Datenverwendung
- » Zugriff für berechtigte Personen ermöglichen
- » Recht auf Vergessen umsetzen

Automatisierung

- » einfaches Umsetzen ohne Aufwand
 - Markieren der DSGVO relevanten Felder reicht
 - Automatisierung kümmert sich um den „Rest“
- » Bereitstellen verschiedener Zugriffsschichten
 - nur berechtigte Personen sehen DSGVO Daten im Klartext
- » Entfernen bei Löschwunsch des Dateneigentümers
 - echte Löschung
 - auch berechtigte Personen sehen keine Daten mehr
 - alle Auswertungen von Nicht-DSGVO-Daten weiterhin mit gleichen Ergebnissen möglich



Datenqualität

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Historie

Migration

DSGVO

Datenqualität

Aufgaben

Datenqualität sichern

- » Qualitätsprobleme finden und überwachen
- » Datenqualität verbessern

Automatisierung

- » technische Qualitätskontrollen automatisch erstellen
 - Pflichtfeldprüfungen
 - Eindeutigkeit prüfen
 - Konsistenz prüfen (z.B. Konto mit unbekanntem Kunden)
 - u.a.
- » fachliche Qualitätsprüfungen ermöglichen
 - Hat ein Versicherungssachbearbeiter einen Schaden oberhalb seiner Befugnisgrenze genehmigt?
- » regelmäßig Status bereitstellen
 - Wann gab es wie viele und welche Qualitätsprobleme?
 - Welche Daten sind betroffen?
 - auswertbar im BI-Tool



Test

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Historie

Migration

DSGVO

Datenqualität

Test

Aufgaben

Test durchführen

» Entwicklertest

» Abnahmetest

» Performancetest

Automatisierung

» Entwicklertest

- jederzeit durch einen Aufruf startbar
- prüft, ob mit Testdaten die erwarteten Ergebnisdaten erzeugt werden

» Abnahmetest

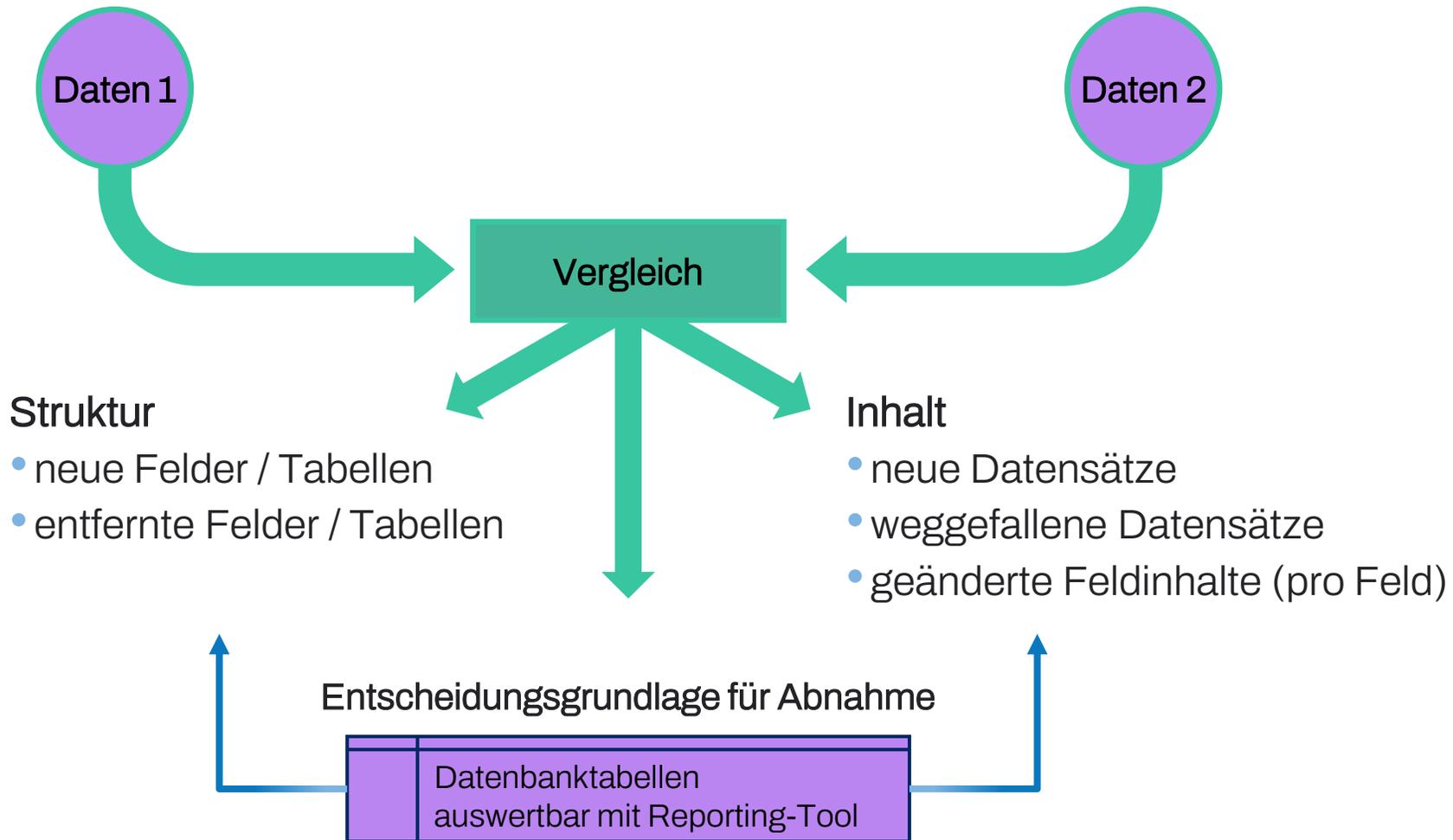
- vergleicht zwei Datenstände (vor und nach Änderung)
- vergleicht alle Tabellen und Felder
- ermöglicht zielgerichtete und schnelle Abnahmen

» Performancetest

- Kann vorhandene Datenstände **konsistent** um vorgegebene Faktoren vervielfältigen
- Laufzeitmessungen auf erhöhten Datenvolumen möglich



Abnahmetest



- » vollständiger Negativtest
- » effiziente Abnahme, da alle Änderungen sofort übersichtlich vorliegen
- » komplett automatisiert durchführbar
- » Ergebnisse liegen dauerhaft in Datenbank vor
- » Nachvollziehbarkeit der Abnahme gegeben



Produktivnahme

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Historie

Migration

DSGVO

Datenqualität

Test

Produktivnahme

Aufgaben

Anwendung produktiv setzen

» Deployment

» Installation

» Test

» Abnahme

» Änderung aktivieren

Anforderungen

» Risiko minimal (de fakto 0)

» 4 Augen Prinzip

» unterbrechungsfreier Produktionsbetrieb

Automatisierung

» Deployment Files bereitstellen

» einfache Installskripte bereitstellen

- Aufrufe sollten immer gleich sein
- unabhängig von der Art der fachlichen Änderung

» risikolose Produktivnahme ermöglichen

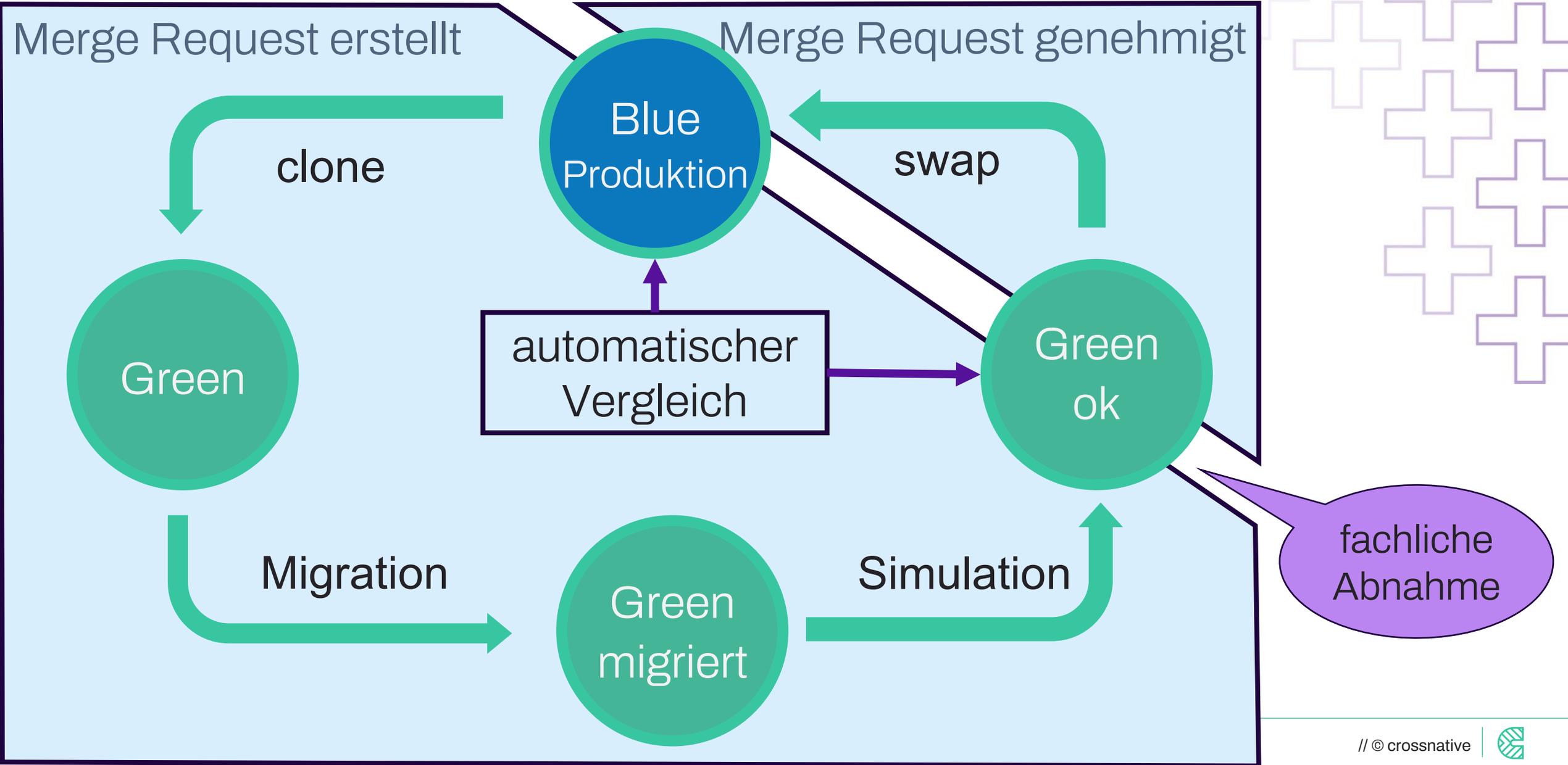
» Nutzung von Git Pipeline



Quelle Git Logo: <https://git-scm.com/downloads/logos>



Risikolose Produktivnahme (Blue Green Verfahren)



Produktionsbetrieb

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Historie

Migration

DSGVO

Datenqualität

Test

Produktivnahme

Produktivbetrieb

Aufgaben

1. regelmäßige Datenversorgung

» stabil

» nachvollziehbar

» risikolos

» einfach in Jobsteuerungssoftware integrierbar

» schnelle Ursachenermittlung bei Fehlern

2. Installation / Verwaltung neuer Versionen

Automatisierung

» Bereitstellung von Skripten

» aufzurufende Skripte ändern sich bei fachlich inhaltlichen Anpassungen der DWH-Logik nicht

- Beladungsskripte
- Installationskripte

» Bereitstellung von Statusinformationen

- Wann lief welche Beladung?
- Status der aktuellen Beladung
- Wann wurde welche Version installiert?
- Ursache von Abbrüchen

» einfaches Rollback

- leichtes Zurücksetzen der Datenstände für Notfälle

» Restartfähigkeit der Beladung bei Abbrüchen



Multienvironment

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Historie

Migration

DSGVO

Datenqualität

Test

Produktivnahme

Produktivbetrieb

Multienvironment

Aufgaben

gleichzeitige Umsetzung verschiedenster Anforderungen

- » Test letztes Release verzögert sich durch Erkrankung eines Testers (Fachabteilung)
- » Test neues Release für Controlling muss gestartet werden
- » BI Bereich braucht Übergangsphase aktuelles / neues Release in Produktion zur Anpassung von Auswertungen
- » Data Science möchte Spielwiese
- » Ad Hoc Kennziffer für Vorstand benötigt
- » Hotfix für Produktion nötig

Automatisierung

beliebig viele Umgebungen können erstellt und mit aktuellen Daten versorgt werden

- » eine oder mehrere Staging-Schichten möglich
- » unterschiedliche parallele DataVault / Datamart Modelle auf jeder Staging Schicht möglich
- » unterschiedliche Geschäftsregeln auf jeder Modellschicht möglich



Multienvironment

Produktion

Simulation

Test Hotfix

neues Release

weitere Projekte
(z.B. Data Science)

Rule
Engine

Rule
Engine

Rule
Engine

Rule
Engine

Rule
Engine

Rule
Engine

Datamarts

Datamarts

Datamarts

BusinessVault

BusinessVault

BusinessVault

RawVault

RawVault

RawVault

Staging Schicht

einfache Datenanalyse

Anforderung

Analyse

Impact

Konzeption

Programmierung

Steuerung

Dokumentation

Historie

Migration

DSGVO

Datenqualität

Test

Produktivnahme

Produktivbetrieb

Multienvironment

Datenanalyse

Aufgaben

Datenanalyse vereinfachen

- » Anwenderfrage: „Warum wird mir im Feld X ein Betrag von 1000,- Euro angezeigt? Ich hätte 1500,- Euro erwartet.“
- » fundierte Antwort auf diese Frage kann sehr aufwendig sein
- » systematische, faktenbasierte Analyse ermöglichen

Automatisierung

» Data Lineage

- Welche Quelldaten fließen in welche Zieldaten ein?
- Wo entsprechen die Daten auf der Verarbeitungsstrecke letztmalig den Erwartungen und ab welcher Stelle weichen Sie ab?

» permanente Speicherung Originaldaten in der Staging-Schicht

- Fehlerursache „falsche gelieferte Daten“ bzw. „Berechnungsfehler im DWH“ klar
- Zuständigkeit (Team des Quellsystems / DWH Team) immer ermittelbar

» Nachvollziehbarkeit Prozessabläufe

- Welcher Prozess hat wann welchen Datensatz erzeugt?
- Welche SQL-Befehle hat dieser Prozess ausgeführt?
- Nach welchem Befehl sind die Daten erstmal unerwartet?





Ausblick

- **Kontakt: joerg.stahnke@crossnative.com**
- **weitere Details**
- **Wie kann man dies umsetzen?**
**=> mein Buch zum Thema wird dieses Jahr
erscheinen**

