



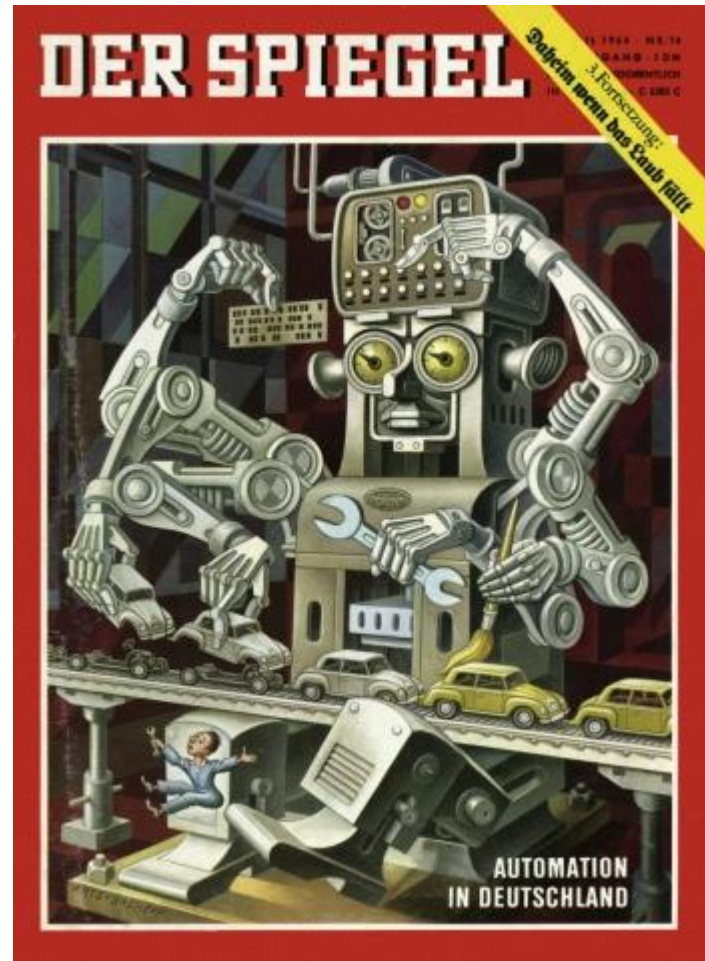
DWH42

# Oliver Cramer

Prinzipien der Data-Warehouse-Automation und grober Marktüberblick

**4. Tagung der DDVUG am 12. Mai 2016  
bei PUMA in Herzogenaurach**

- Vorstellung
- Definition
- Prinzipien
- Entwicklungsphilosophie
- Marktübersicht



Das Cover der „Spiegel“-Ausgabe vom 1. April 1964 zeigt einen alles könnenden, vielarmigen, mit Augen, Ohren und Elektronengehirn ausgestatteten Roboter, der den kleinen und hilflosen Menschen mit seinen Füßen wegkickt.

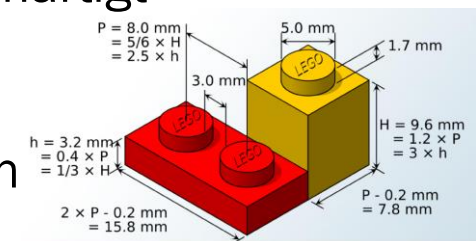
# Agenda



# Über mich

- Freiberuflicher Data Warehouse Architekt
- 13 Jahre im Business Intelligence Bereich tätig
- Seit 2003 mit Elementarbausteinen für das Data Warehouse beschäftigt
- Einträge auf [www.dwh42.de](http://www.dwh42.de) zur Data Warehouse Automation
- Interessiert am Austausch zu Data Warehouse Modellierungsstilen

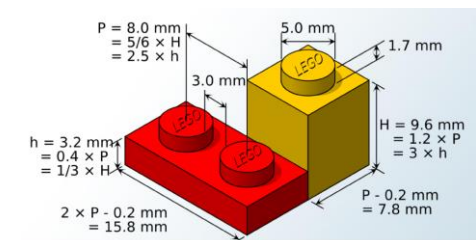
DWH42



# Über mich

- TDWI Europe Fellow
- ANCHOR CERTIFIED MODELER Version 2014
- Certified Data Vault 2.0 Practitioner
- Koautor von „Neue Wege in der Datenmodellierung - Data Vault heißt die moderne Antwort“ im BI-Spektrum 03-2014
- Member of the Boulder BI Brain Trust
- Member of the BI-Podium Advisory Board Germany
- Leitender Redakteur des TDWI Themen Online Specials „Data Vault“
- Organisator Data Vault Modellierung und Zertifizierung, Hannover mit der Genesee Academy CDVDM

DWH42

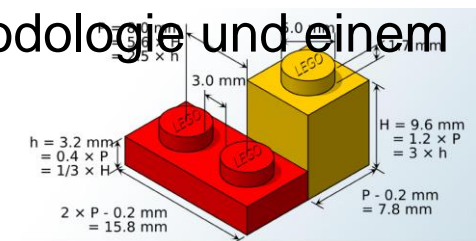


# Agenda

- Vorstellung
- **Definition Data Warehouse Automation**
- Prinzipien der Data Warehouse Automation
- Entwicklungsphilosophie
- Marktübersicht

# Definition Data Warehouse Automation

- Data Warehouse Automation ist mehr als nur die Automatisierung von ETL-Entwicklung. Automatisiert wird der gesamte Data Warehousing Lebenszyklus: angefangen von der Planung, der Analyse und dem Design über die Entwicklung und dem Testen - von dort geht es über in Betrieb, Wartung und Change-Management.
- Merkmale von DWA sind also: Planen, Verwalten, Unterstützen, Infrastruktur geben, Organisieren, Automatisieren, Generieren, Vorlagen und Muster zur Verfügung stellen, Dokumentieren und Testen
- DWA kann aus einem Werkzeug bzw. Anwendung, einer Methodologie und einem Framework bestehen



# Agenda

- Einleitung
- Definition Data Warehouse Automation
- **Prinzipien Data Warehouse Automation**
- Entwicklungsphilosophie
- Marktübersicht

- Vermeidung unnötiger Arbeit

## „Faulheit“





- Neugier (auch Neugierde) ist das als ein Reiz auftretende Verlangen, Neues zu erfahren und insbesondere Verborgenes kennenzulernen.

## „Neugier“



- „Ein grundlegendes Gesetz der Thermodynamik besagt, dass die Gesamtentropie der Welt langsam, aber sicher steigt, die Ordnung nimmt ab - es wird immer langweiliger.“

## „Ordnung“

*Ordnung ist...*

*wenn man weiss,  
wo man gar nicht erst  
zu suchen braucht!*

Übersichtlichkeit  
Reduzieren  
Entsorgen  
Ordnung  
Freiheit  
Klarheit  
Verschenken  
FengShui  
Entrümpeln  
Vereinfachen  
Befreien

# „Qualität“

- Qualität wird nicht am Ende in das Produkt kontrolliert, sondern gleich zu Beginn eingebaut.



- "...DASS ES EINE CONTIERNUIERLICHE BESTÄNDIGE UND NACHHALTIGE NUTZUNG GEBE / WEILN ES EINE UNENTBERLICHE SACHE IST / OHNE WELCHE DAS LAND IN SEINEM ESSE NICHT BLEIBEN MAG."

## „Nachhaltigkeit“



# „Gelassenheit“



Gelassenheit kann man lernen. Man braucht dazu nur Offenheit, Motivation, ein bißchen Ausdauer und vor allem Bereitschaft, sich von den alten, eingefahrenen Bahnen zu lösen, in denen unser Denken und Handeln sich häufig bewegt.

(Ludwig Bechstein)

[gutezitate.com](http://gutezitate.com)

# „Abgucken“

- „Denn das ist alles nur geklaut,  
das ist alles gar nicht meine,  
das ist alles nur geklaut,  
doch das weiß ich nur ganz alleine,  
das ist alles nur geklaut  
und gestohlen,  
nur gezogen  
und geraubt.“
- Entschuldigung, das hab' ich mir  
erlaubt.“

„Die Prinzen“

# Agenda

- Einleitung
- Definition Data Warehouse Automation
- Prinzipien Data Warehouse Automation
- **Entwicklungsphilosophie**
- Marktübersicht

DWH42

# Entwicklungsphilosophie

- Wir zielen auf einen Wandel in der Denkweise: von der „Architektur“ bis zur „Ausführung“
- Wir zielen darauf ab, geeignet zu sein für Prototypenentwicklung und „Agilität“



# (Meta) - modellgetriebene Automation

- Modelle (Prozesse, Regeln und Daten) bestimmen die Metadaten, die Metadaten bestimmen die Automatisierungsartefakte/die zu automatisierenden Artefakte
- Ziel ist eine 100% Aussage (Ziel ist es, zu 100% deklarativ zu sein)
- Nicht alles kann generiert werden. Spezifisch angepasste/zugeschnittene Metadaten werden weiterhin erforderlich bleiben.

## Data Vault Implementierungen

DWH42

### Metadaten-getriebene Automation

- Inputs: Quell-/Ausgangsmodell(e), Zielmodell, Template-Design, Namenskonventionen
- Erweiterte Inputs: Normalisierungseinstellungen, Ontologien

Aus Dan Linstedt's blog post: <http://danlinstedt.com/datavaultcat/code-generation-for-data-vault-not-as-easy-as-you-think/>

### Template-getriebene Automation

- In seiner einfachsten Form; Dokumentation – Beschreibung von Patterns/ Mustern
- Weiter fortgeschritten; Generieren von XML Code für ETL-Werkzeuge der 2. Generation
- - <http://www.grundsatzlich-it.nl/bi-tools-templator.html>

# Entwicklungsphilosophie

## Wir wollen uns keine Sorgen machen um...

- Indizes
- Schlüssel und Zeitdruck
- Hierarchien
- Zeitabweichungen
- Verteilungsschlüssel
- Caching
- Komprimierung
- Vorbelastungen
- CDC
- Prozesssteuerung
- (Daten) Wiederherstellung
- DDL/DML Definitionen und Konventionen
- Verspätet ankommende Informationen
- Dateipfade
- Paket-Konfigurationen
- SCD1/2/3/6 Implementierung
- Füllfaktoren
- Wartungspläne
- Datentypkonvertierungen
- Scripting
- Sortierfolge
- Nachschläge
- Datenzugangsarten
- Validierung
- Übereinstimmung und Abstammung
- Datenkalkulationen

Der Fokus liegt auf einer Kernkompetenz: Daten Modellierung

# Entwicklungsphilosophie

## – Atomare Operationen sind der Schlüssel

- Data Warehouses versuchen in einem Ladezyklus zu viel abzuhandeln
- Alle Arten von Problemen werden in einem Lademuster angegangen
  - Geschäftsregeln
  - Fehlerbehandlung
  - Timing
  - Leistung
  - Vermischte Updates, Inserts und Deletes
  - Ladefenster
  - Verspätete Ankunft
  - Säuberungen
  - Aggregation
  - Verteilungsschlüssel
  - Fehlende Quellschlüssel
  - Struktur, Hierarchie
  - Datentyp-Korrekturen
  - Historie

# Entwicklungsphilosophie

## – Atomare Operationen sind der Schlüssel

### Atomare Operationen

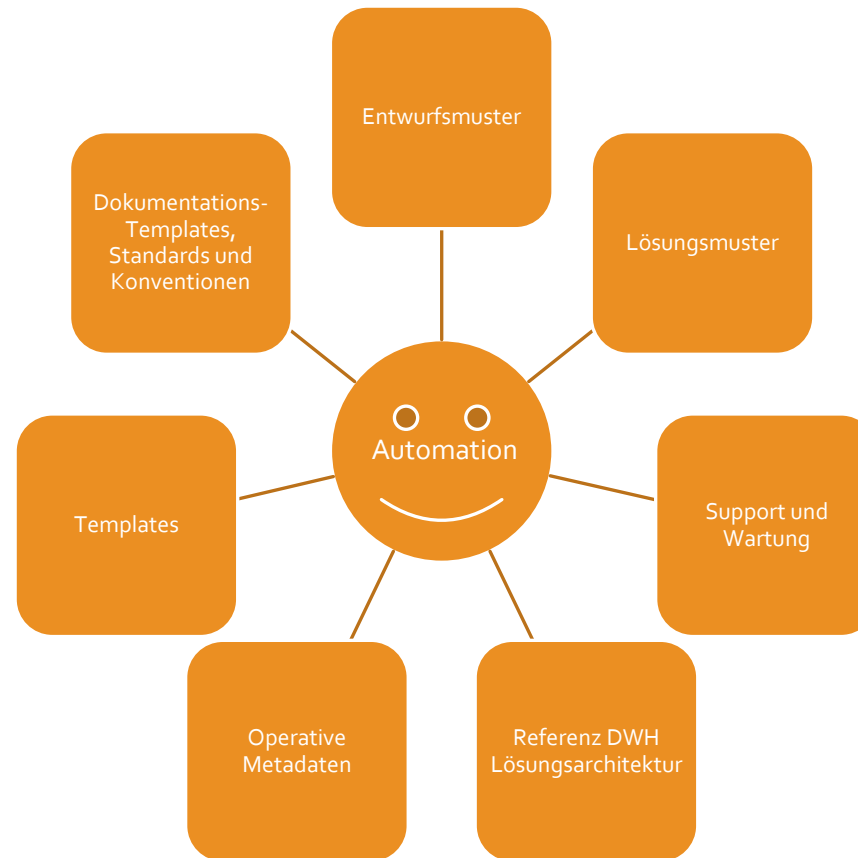
- Wir streben danach Dinge in (kleinere) Komponenten zu teilen um Flexibilität, Agilität und Anpassbarkeit zu erreichen
- Dieses erleichtert die Erfassung von Informationen, die auf verschiedene Weise ausgelegt werden oder unabhängig wechseln können.

### Vereinheitlichung

- Gleichzeitig ist die Richtung einerseits die Integration und andererseits zu einer gemeinsamen Sichtweise zu kommen

**„Rohdatenschicht“**  
**„Teile und herrsche!“**  
**„Trennung der Belange“**

# Was brauchen wir?



# Referenzarchitektur - Grundprinzipien

## Herausforderungen:

- Behandlung der/Umgang mit Komplexität
- Behandlung der/Umgang mit Abhängigkeiten
- Fähigkeit auf eine wechselnde Umgebung zu reagieren

## Prinzipien:

- Flexibilität in Entwurf und Wartung
- Wechselresistenz
- (fast) Echtzeitfähig
- Modular
- Skalierbar
- Langlebig und vorhersagbar
- Eine „bottom up“ Architektur, schrittweise angewandt mit einem fraktalen Ansatz

## Ergebnisse:

- Trennung der Data Warehouse-Konzepte
- Flexible Fehlerbehandlung
- Hybride Modellierung
- Parallelität
- Integrierte Prüfkette

# Automationsziele

- Performance
- Parallelisierungsfähig
- Wiederaufsetzbar
- Wartbar
- Partitionierbar

# Automationsziele

- Prozessunabhängig
- Standardisiert
- Werkzeugunabhängig
- Echtzeitfähig
- Skalierbar



# Der Automationswachstumspfad

- Klein anfangen, aber anfangen!
- Niemals aufhören!
- Niemals Automatisieren, was dein Werkzeug kann!
- Automation ist kein Selbstzweck!
- Automation ist Klempnerei, ist weder Architektur noch Wissenschaft.

# Agenda

- Einleitung
- Definition Data Warehouse Automation
- Prinzipien Data Warehouse Automation
- Entwicklungsphilosophie
- **Marktübersicht**

DWH42

## „Building the i-refactory“, Martijn Evers

„Euer eigenes Werkzeug!“

YOT

# Marktübersicht

Hersteller	Toolname / Produkt	Quelle / URL
AbInitio	MDW - Metadata-driven Warehouse	<a href="http://www.abinitio.com/">http://www.abinitio.com/</a>
ANALYTIX DS	AnalytixDS Mapping Manager	<a href="http://www.analytixds.com/">http://www.analytixds.com/</a>
Balanced Insight	Consensus	<a href="http://www.balancedinsight.com/">http://www.balancedinsight.com/</a>
	ConsensusNow	
BIReady	BIReady - core product plus 6 additional partner products	<a href="http://www.biready.com/">http://www.biready.com/</a>
BimlScript	Biml BimlScript	<a href="http://bimlscript.com/">http://bimlscript.com/</a>
Birst	Automated Analytics	<a href="http://www.birst.com">http://www.birst.com</a>
Centennium BV	Centennium Automation Tool (CAT)	<a href="http://www.centennium.nl">http://www.centennium.nl</a>
Conspect	Conspect Datawarehouse Toolkit (CDT)	<a href="http://www.conspect.nl/datawarehouse-automation/">http://www.conspect.nl/datawarehouse-automation/</a>
Data Acedemy	Data Acedemy	<a href="http://www.dataacademy.com/">http://www.dataacademy.com/</a>
DATA SOLUTIONS DE SPIRLET	SAP Powerdesigner ADD-ONS zur Generierung	<a href="http://www.data-solutions-de-spirlet.com/">http://www.data-solutions-de-spirlet.com/</a>
Dimodelo	Dimodelo Architect / BIML	<a href="http://www.dimodelo.com">http://www.dimodelo.com</a>
in2bi	DWH Deck 2.0	<a href="http://blog.in2bi.com/dwh-deck/dwh-deck-2-0/">http://blog.in2bi.com/dwh-deck/dwh-deck-2-0/</a>
Kalido	Kalido Information Engine	<a href="http://www.kalido.com/">http://www.kalido.com/</a>
Linstedt, Dan	RapidGenDS	<a href="http://rapidgends.com/">http://rapidgends.com/</a>
MID	Innovator	<a href="http://www.mid.de/">http://www.mid.de/</a>
QOSCO	Quipu 3.0	<a href="http://www.datawarehousemanagement.org/">http://www.datawarehousemanagement.org/</a>
TimeXtender	TimeXTender Data Warehouse Automation	<a href="http://www.timextender.com/">http://www.timextender.com/</a>
Trivadis	BIGenius	<a href="http://www.trivadis.com/">http://www.trivadis.com/</a>
Varigence	Varigence MIST	<a href="https://www.varigence.com/">https://www.varigence.com/</a>
Weber, Edwin	PDI Data Vault framework	<a href="http://sourceforge.net/projects/pdidatavaultfw/">http://sourceforge.net/projects/pdidatavaultfw/</a>
WhereScape	3D	<a href="https://www.wherescape.com/">https://www.wherescape.com/</a>
	RED	<a href="https://www.wherescape.com/">https://www.wherescape.com/</a>

**Ende**

Vielen Dank für die Aufmerksamkeit!

DWH42