

# ODI 12c und Data Vault

Erste Erfahrungen aus der Praxis

Claus Jordan  
Senior Consultant




BASEL BERN BRUGG LAUSANNE ZÜRICH DÜSSELDORF FRANKFURT A.M. FREIBURG I.BR. HAMBURG MÜNCHEN STUTTGART WIEN

2013 © Trivadis

ODI 12c und Data Vault  
24.07.2014

**trivadis**  
makes IT easier. ■ ■ ■

## ■ Kurzvorstellung Trivadis.

Trivadis ist **führend bei der IT-Beratung, der Systemintegration** und der Erbringung von **IT-Services** mit Fokussierung auf **ORACLE®**- und  **Microsoft**-Technologien im D-A-CH-Raum.

Unsere Leistungen erbringen wir aus den strategischen Geschäftsfeldern:



Trivadis Services übernimmt den korrespondierenden Betrieb Ihrer IT Systeme.

## ■ Mit über 600 IT- und Fachexperten bei Ihnen vor Ort.



11 Trivadis Niederlassungen mit über 600 Mitarbeitenden

200 Service Level Agreements

Mehr als 4'000 Trainingsteilnehmer

Forschungs- und Entwicklungsbudget: CHF 5.0 / EUR 4 Mio.

Finanziell unabhängig und nachhaltig profitabel

Erfahrung aus mehr als 1'900 Projekten pro Jahr bei über 800 Kunden

2013 © Trivadis

ODI 12c und Data Vault  
24.07.2014

Stand 12/2012

**trivadis**  
makes IT easier. ■ ■ ■

# ■ AGENDA

In diesem Vortrag wird die neueste Version von Oracle Data Integrator (ODI 12c), in Kombination mit der zur Zeit viel diskutierten Data Warehouse Methode Data Vault (Stichwort „agiles DWH“), vorgestellt.

Folgende Fragen werden anhand eines echten Kundenprojektes behandelt:

1. Wie gelangt man vom logischen Datenmodell (z.B. ER-Diagramm) zum Data Vault Datenmodell?
2. Wie wird in Data Vault historisiert?
3. Welche Vorteile bringt Data Vault?
4. Wie wird mit ODI ein Mapping erstellt und wo liegen die Unterschiede zum Oracle Warehouse Builder OWB?
5. Was steckt eigentlich hinter den ODI-Knowledge-Modulen und wie funktionieren sie?

# Von den Anforderungen zum Data Vault Datenmodell

Anforderungen

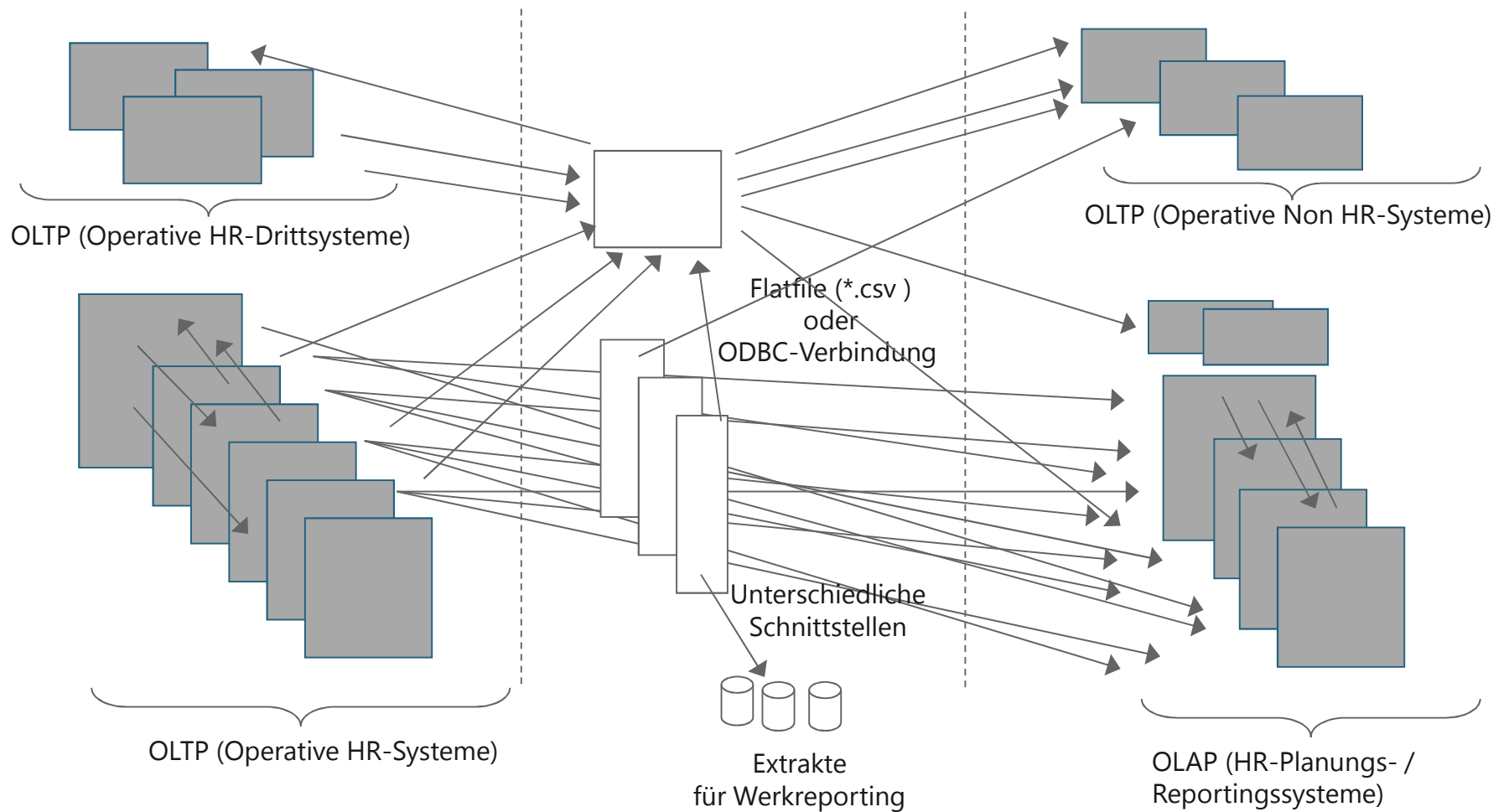


2013 © Trivadis

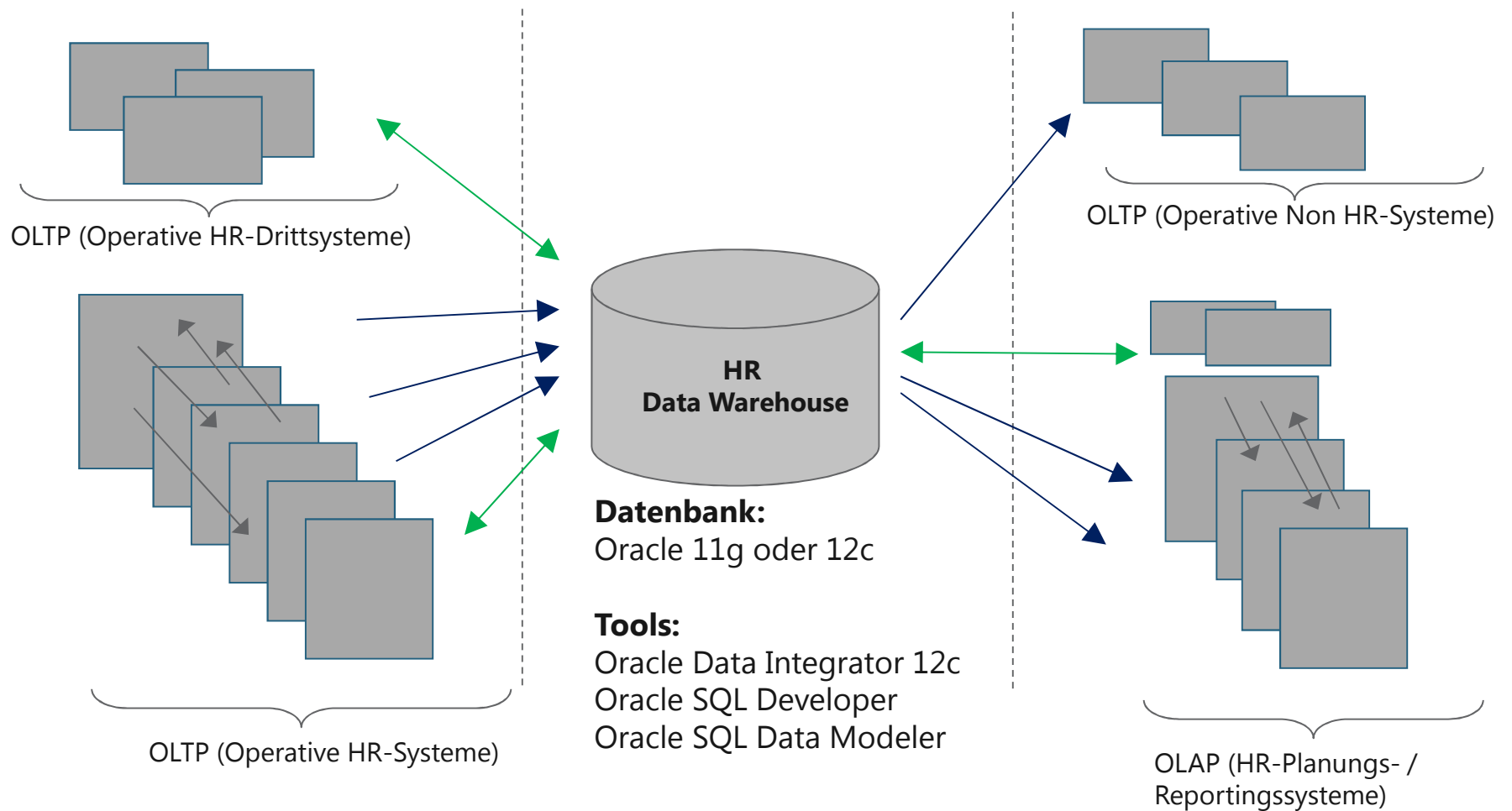
ODI 12c und Data Vault  
24.07.2014

**trivadis**  
makes IT easier. ■ ■ ■

# ■ Systemlandschaft „Heute“



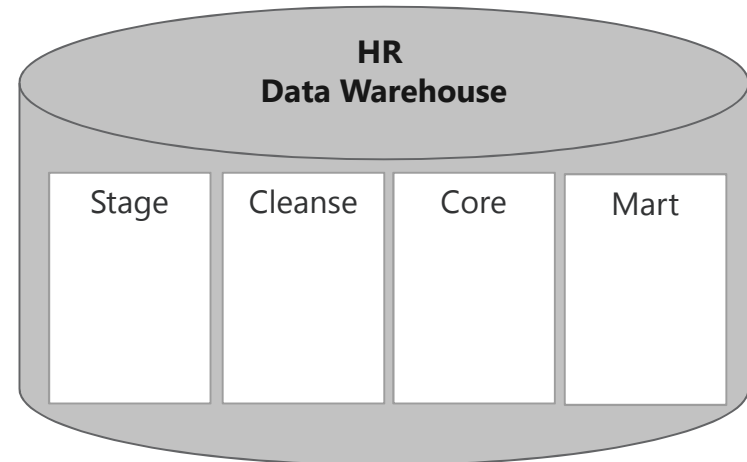
# ■ Systemlandschaft „Morgen“



# ■ Konzept und Datenmodell

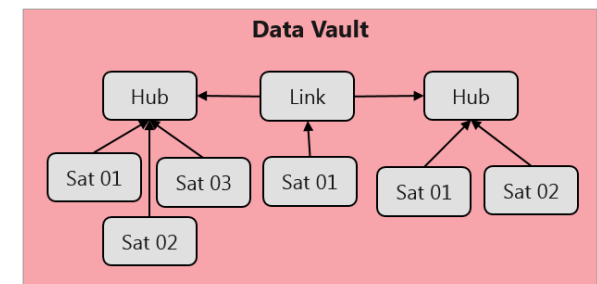
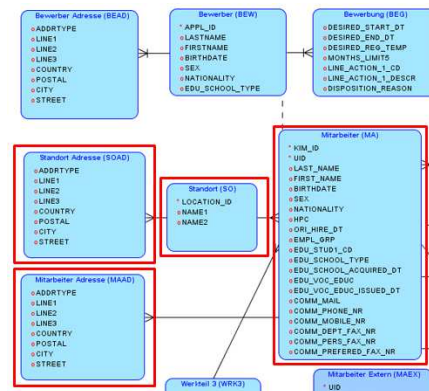
## HR Data Warehouse – Konzept (Teil 1)

- Aufbau und Technologie
- Vorgehensweise



## HR Data Warehouse – Datenmodell (Teil 2)

- ER Diagramm
- Physische Datenmodelle





# Von den Anforderungen zum Data Vault Datenmodell

Datenmodell



2013 © Trivadis

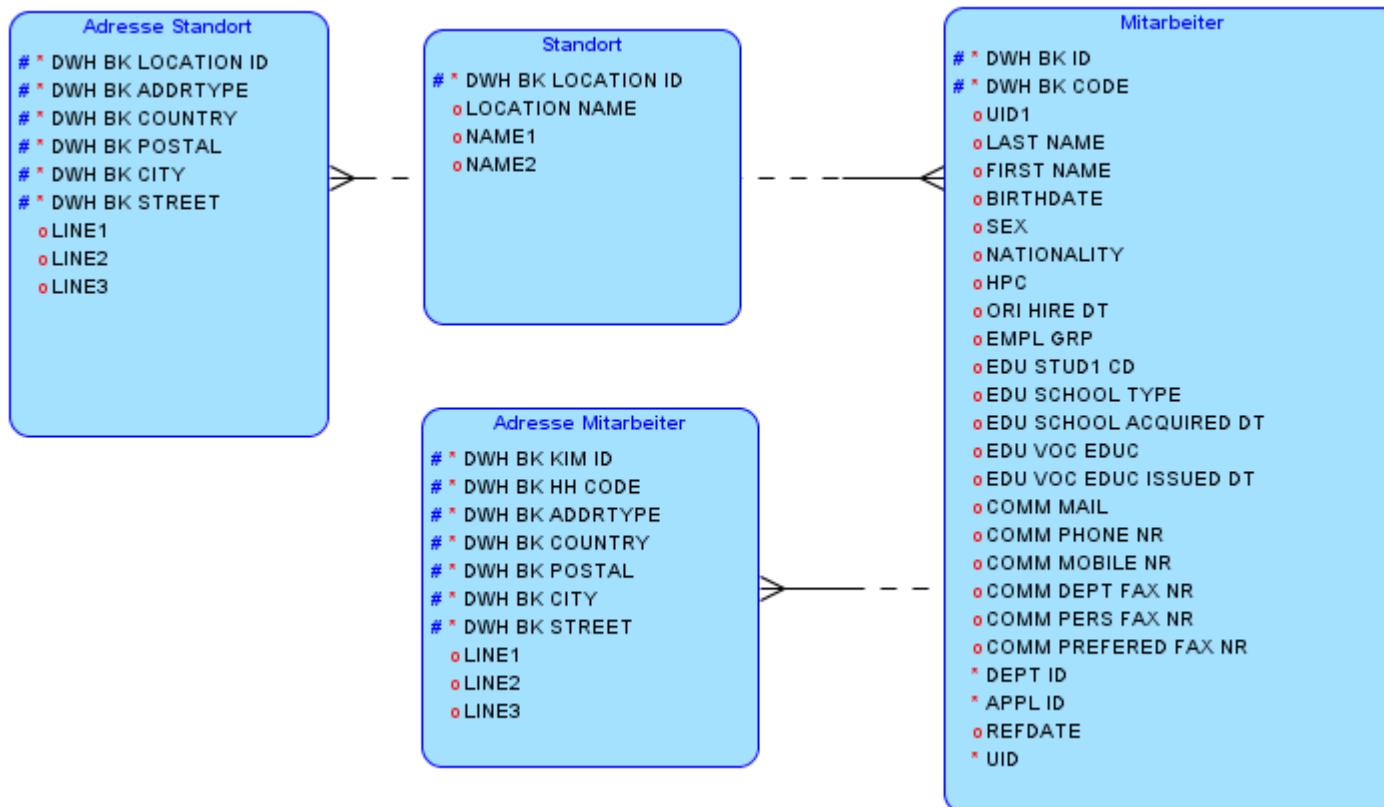
ODI 12c und Data Vault  
24.07.2014

**trivadis**  
makes IT easier. ■ ■ ■

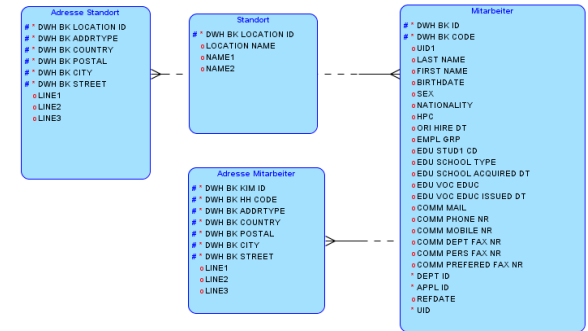
# ■ Ausgangspunkt ist das logisches Datenmodell ..

.. für die Objekte

**Mitarbeiter + Adresse(n), Standort + Adresse(n)**

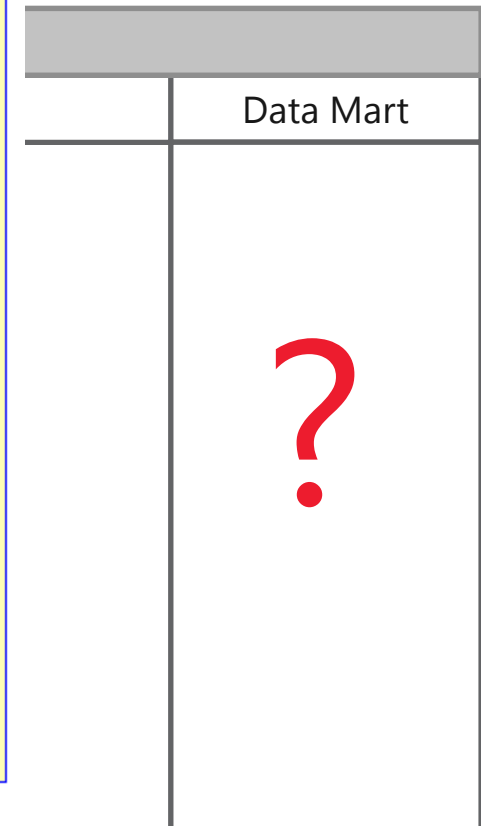
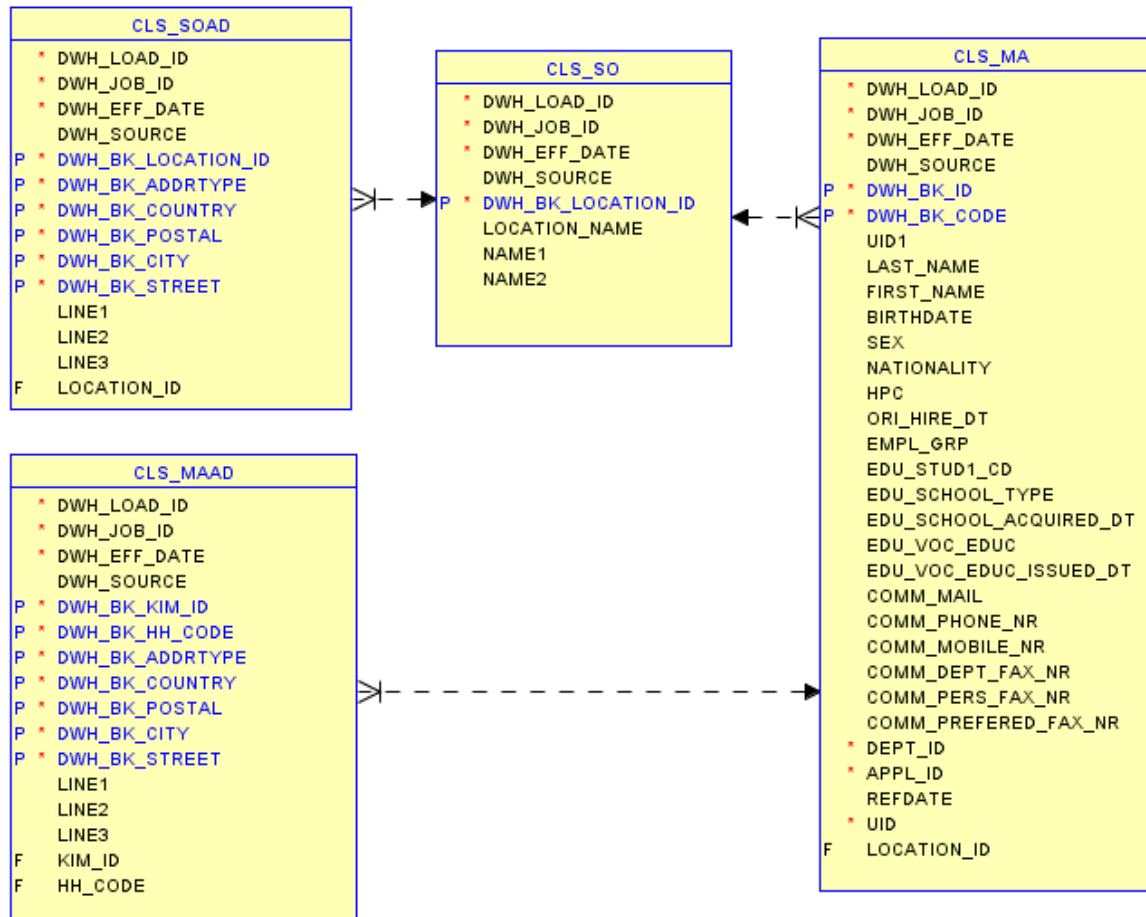
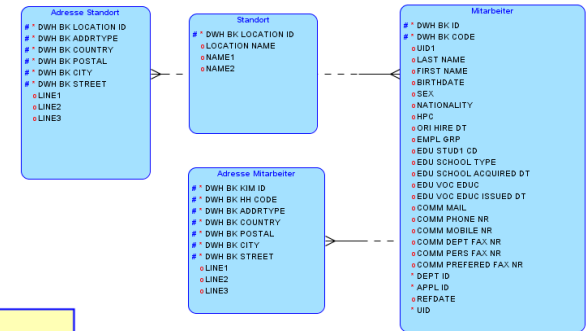


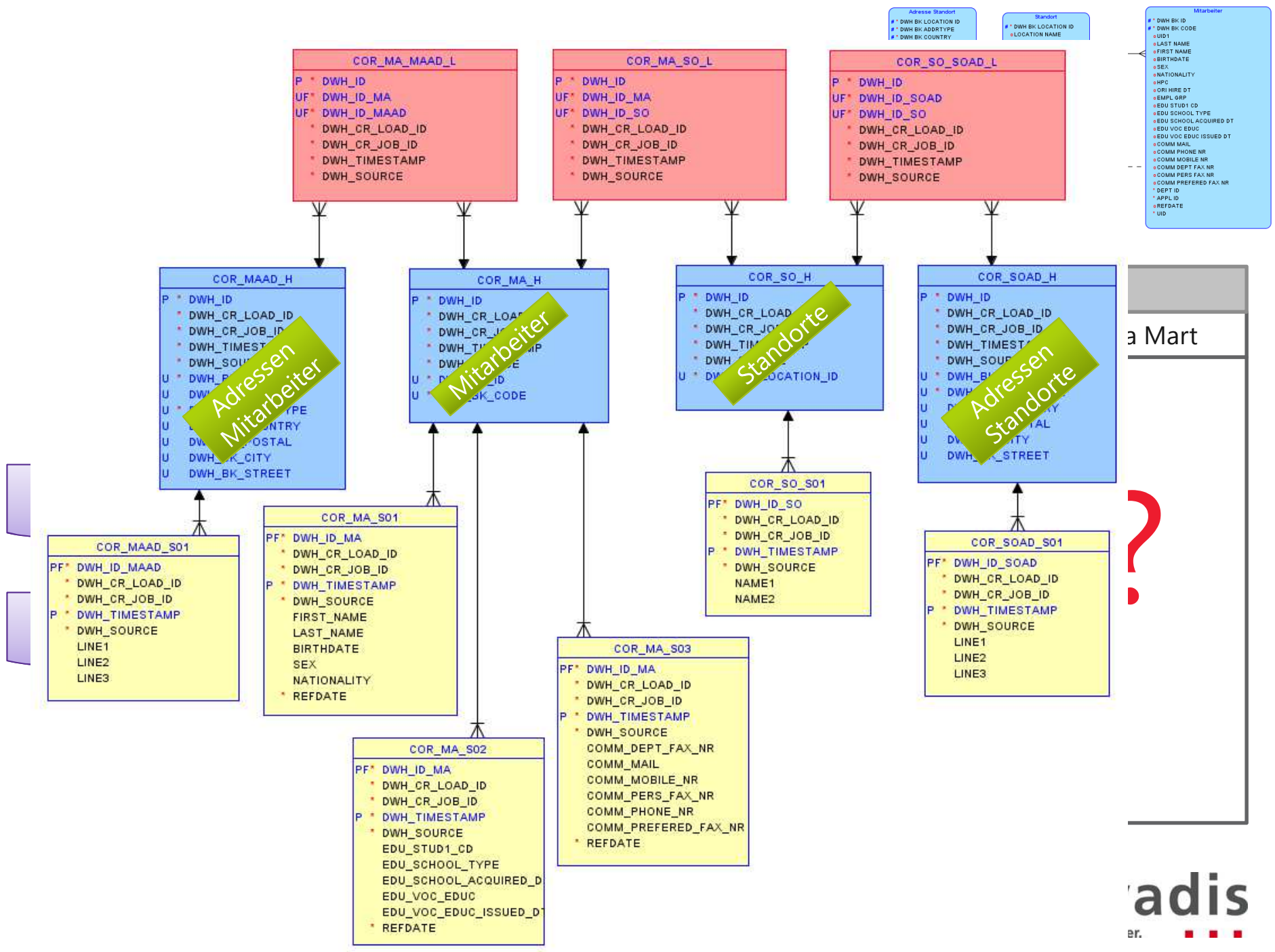
# ■ HR DWH - Physisches Datenmodell



HR Data Warehouse			
Stage	Cleanse	Core	Data Mart
<p>MA.csv</p> <p>LOC.csv</p> <p>?</p>	?	?	?

# 1. Cleanse

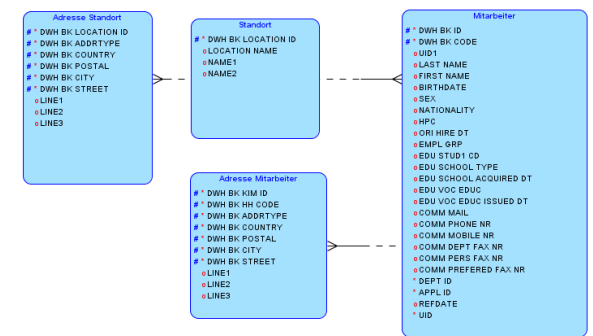
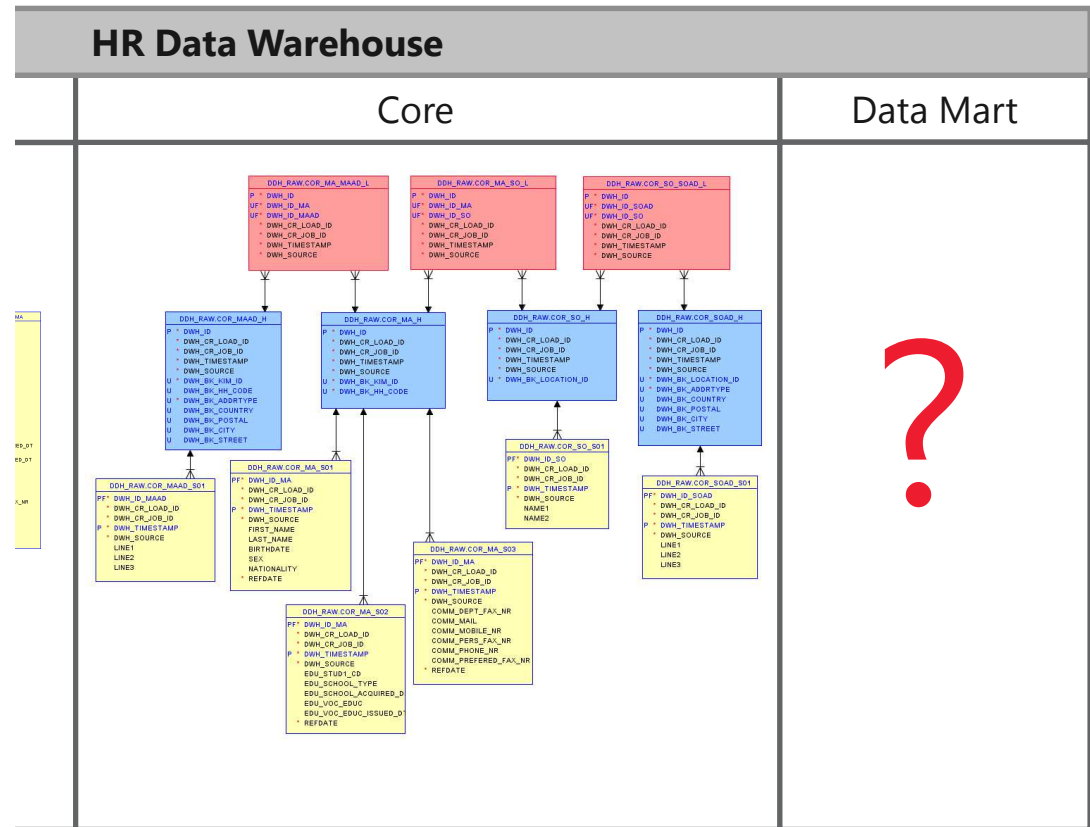
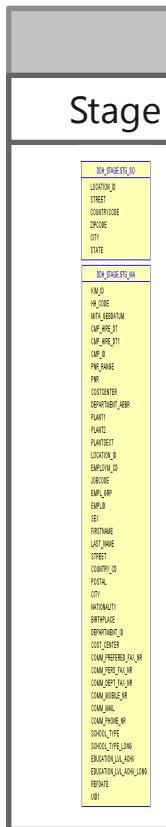




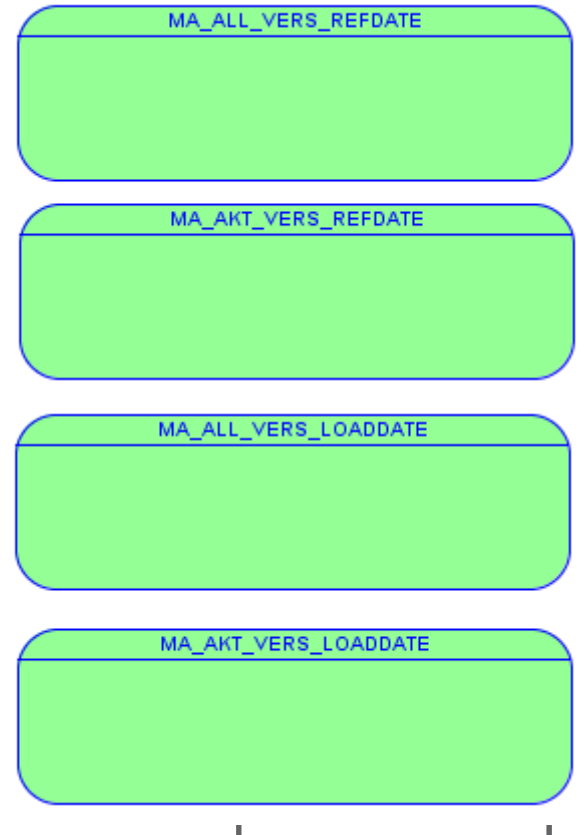
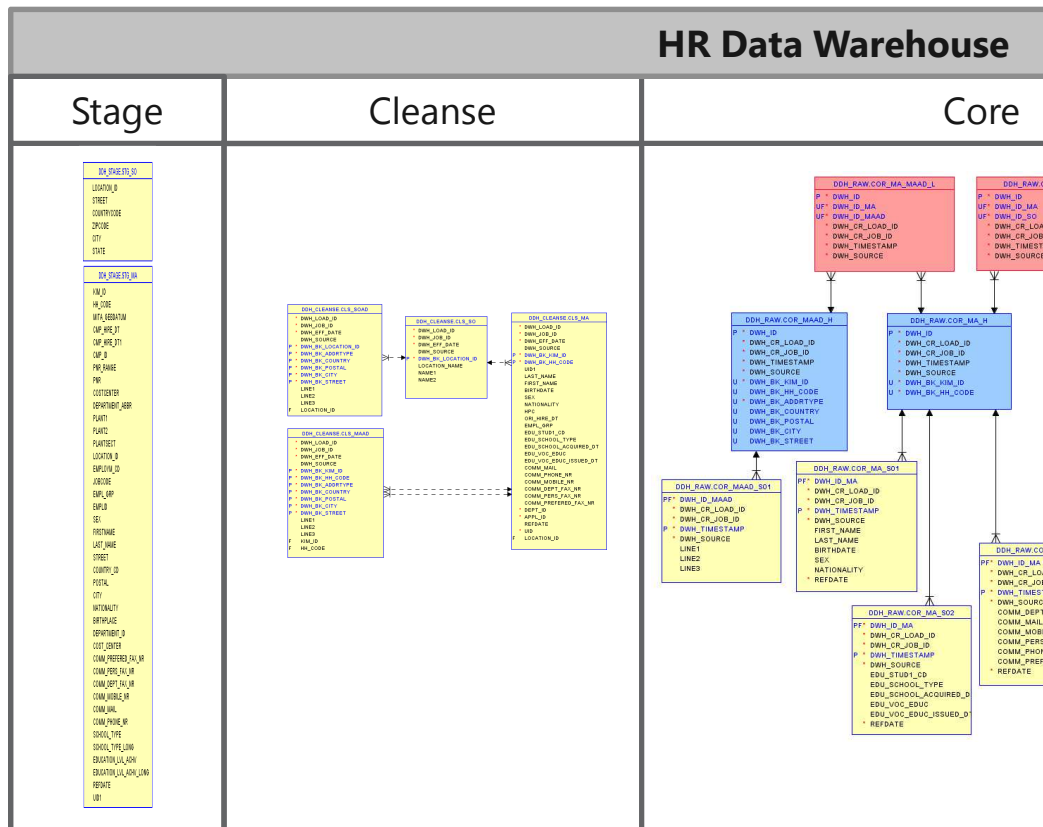
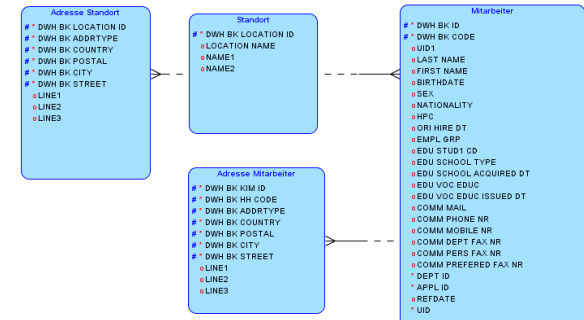
# 3. Stag

STG_SO
LOCATION_ID
STREET
COUNTRYCODE
ZIPCODE
CITY
STATE

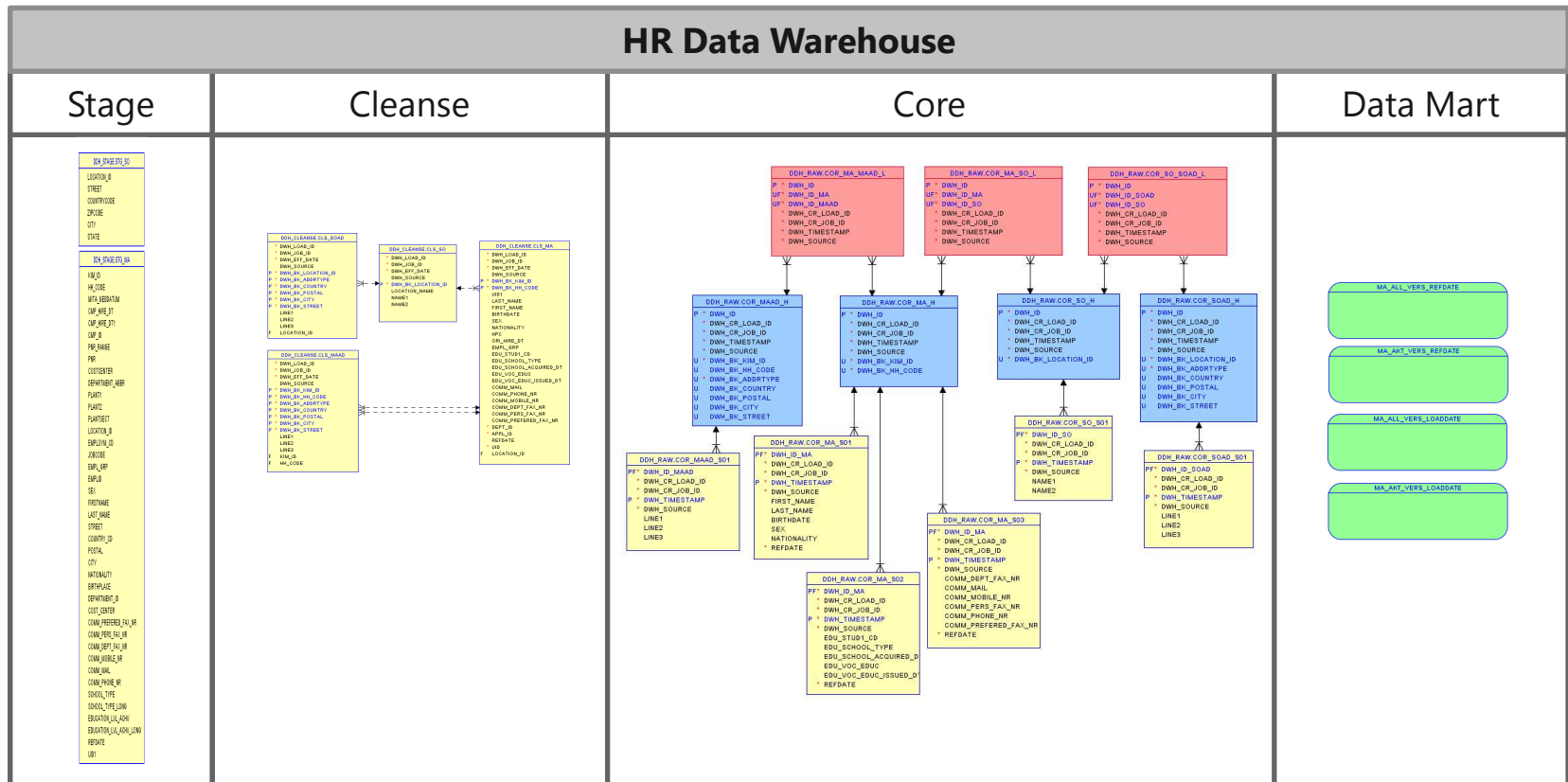
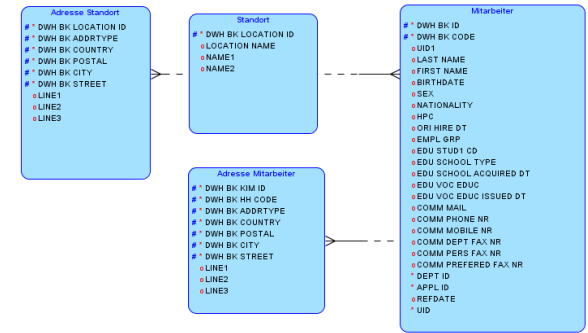
STG_MA
ID
CODE
MITA_GEBDATUM
CMP_HIRE_DT
CMP_HIRE_DT1
CMP_ID
PNR_RANGE
PNR
COSTCENTER
DEPARTMENT_ABBR
PLANT1
PLANT2
PLANTSECT
LOCATION_ID
EMPLOYM_CD
JOBCODE
EMPL_GRP
EMPLID
SEX
FIRSTNAME
LAST_NAME
STREET
COUNTRY_CD
POSTAL
CITY
NATIONALITY
BIRTHPLACE
DEPARTMENT_ID
COST_CENTER
COMM_PREFERRED_FAX_NR
COMM_PERS_FAX_NR
COMM_DEPT_FAX_NR
COMM_MOBILE_NR
COMM_MAIL
COMM_PHONE_NR
SCHOOL_TYPE
SCHOOL_TYPE_LONG
EDUCATION_LVL_ACHV
EDUCATION_LVL_ACHV_LONG
REFDATE
IID1



# 4. Data Mart



# HR DWH - Physisches Datenmodell





# Von den Anforderungen zum Data Vault Datenmodell

Historisierung in Data Vault



2013 © Trivadis

ODI 12c und Data Vault  
24.07.2014

**trivadis**  
makes IT easier. ■ ■ ■

## ■ Historisierung in Data Vault (1)

Der Datensatz des Mitarbeiters mit der ID 010/12345 wird erstmalig am 24.04.2011 in das DWH geladen. Die beiden Tabellen (Hub- und Sat-Table) zeigen den Status nach dem Laden der Personaldaten am 25.04.2011.

### Hub-Table:

DWH_ID (PK)	ID	DWH_LOAD_DATE
100	010/12345	25.04.2011

### Sat-Table:

DWH_ID_HUB	DWH_LOAD_DATE	NAME	FAMSTAT	PLZ	VALID_FROM
100	25.04.2011	König, Beate	ledig	89002	23.04.2011

Seit dem werden die Daten für diese Person täglich geladen, bis eines Tages (am 15.01.2014) sich der Wohnort ändert. Die Hub-Table bleibt unverändert und in der Sat-Table wird ein neuer Datensatz angelegt.

### Sat-Table:

DWH_ID_HUB	DWH_LOAD_DATE	NAME	FAMSTAT	PLZ	VALID_FROM
100	25.04.2011	König, Beate	ledig	89002	23.04.2011
100	15.01.2014	König, Beate	ledig	78999	10.01.2014

## ■ Historisierung in Data Vault (2)

Mit dem Ladelauf vom 31.01.2014 ändert sich der Familienstatus und damit verbunden der Nachname. Die Hub-Table bleibt wieder unverändert und in der Sat-Table wird ein neuer Datensatz angelegt

### Sat-Table:

DWH_ID_HUB	DWH_LOAD_DATE	NAME	FAMSTAT	PLZ	VALID_FROM
100	25.04.2011	König, Beate	ledig	89002	23.04.2011
100	15.01.2014	König, Beate	ledig	78999	10.01.2014
100	31.01.2014	Maier, Beate	verheiratet	78999	28.01.2014

Frau Maier zieht am 10.02.2014 ein zweites Mal um.

### Sat-Table:

DWH_ID_HUB	DWH_LOAD_DATE	NAME	FAMSTAT	PLZ	VALID_FROM
100	25.04.2011	König, Beate	ledig	89002	23.04.2011
100	15.01.2014	König, Beate	ledig	78999	10.01.2014
100	31.01.2014	Maier, Beate	verheiratet	78999	28.01.2014
100	10.02.2014	Maier, Beate	verheiratet	60021	09.02.2014

## ■ Historisierung in Data Vault (3) – Nachtr. Änderung!

Datenlieferung am 15.02.2014:

ID (BK)	NAME	FAMSTAT	PLZ	VALID_FROM
010/12345	Maier, Beate	verheiratet	60021	31.12.2013

Sat-Table:

DWH_ID_HUB	DWH_LOAD_DATE	NAME	FAMSTAT	PLZ	VALID_FROM
100	25.04.2011	König, Beate	ledig	89002	23.04.2011
<b>100</b>	<b>15.02.2014</b>	<b>Maier, Beate</b>	<b>verheiratet</b>	<b>60021</b>	<b>31.12.2013</b>
100	15.01.2014	König, Beate	ledig	78999	10.01.2014
100	31.01.2014	Maier, Beate	verheiratet	78999	28.01.2014
100	10.02.2014	Maier, Beate	verheiratet	60021	09.02.2014

Da nun im nächsten Schritt sämtliche Attribute dieser Person überschrieben werden müssten, und damit identisch wären wie der soeben eingefügte Datensatz, werden alle Datensätze (Versionen), die ein größeres VALID\_FROM als 31.12.2013 haben, gelöscht. Es versteht sich von selbst, dass dadurch Informationen verloren gehen, z.B. der zwischenzeitliche Wohnortwechsel.

Sat-Table:

DWH_ID_HUB	DWH_LOAD_DATE	NAME	FAMSTAT	PLZ	VALID_FROM
100	25.04.2011	König, Beate	ledig	89002	23.04.2011
<b>100</b>	<b>15.02.2014</b>	<b>Maier, Beate</b>	<b>verheiratet</b>	<b>60021</b>	<b>31.12.2013</b>

# Von den Anforderungen zum Data Vault Datenmodell

Bewertung von Data Vault



2013 © Trivadis

ODI 12c und Data Vault  
24.07.2014

**trivadis**  
makes IT easier. ■ ■ ■

## ■ Bewertung von Data Vault (1)

Methode Kriterien \	Data Vault
<b>Anzahl Tabellen und ETL-Prozesse</b>	Pro Entität sind zwei oder mehr Tabellen, plus eine Tabelle für jede Beziehung zwischen Entitäten, notwendig. Dies ermöglicht, gerade bei „breiten Entitäten“ (z.B. Mitarbeiter), eine gezielte Gruppierung von Attributen und erleichtert somit die Übersicht. Pro Tabelle resultiert ein ETL-Prozess.
<b>Komplexität für das Laden aus Cleanse in Core</b>	Sehr gering und effizient, zumal keine performanceintensiven Updates notwendig sind. D.h. Datensätze werden nur dann eingefügt, wenn tatsächlich Änderungen an den betreffenden Attributen vorkommen.
<b>Komplexität für das Laden aus Core in Data Marts</b>	Etwas höher, aufgrund der Transformation vom normalisierten in das denormalisierte Datenmodell, und vor allen Dingen wegen der Bildung von neuen Gültigkeitsintervallen beim Zusammenjoinen von unabhängig versionierten Stammdatenentitäten. Diese Logik kann beispielsweise in Datenbank-Views implementiert werden. Dadurch ist der Zugriff ähnlich einfach wie im dimensionalen Datenmodell und stellt somit kein KO-Kriterium dar.

## ■ Bewertung von Data Vault (2)

Methode Kriterien \	Data Vault
<b>Erweiterbarkeit („Agile BI“)</b>	Sehr gut, aufgrund fehlender Referenzen zwischen Entitäten, die jeweils unanabhängig voneinander erweitert oder angepaßt werden können.
<b>Datenredundanz</b>	Sehr gering, durch Normalisierung und Splittung der Attribute einer Entität in mehreren Satellitentabellen.
<b>Datenvolumen</b>	Zumindest für Stammdaten minimal, aufgrund der geringen Datenredundanz
<b>Parallelisierbarkeit</b>	Sehr gut beim Laden des Core. Sämtliche Hub-Tables können parallel geladen werden, danach alle Link-Tables und anschließend alle Satellite-Tables jeweils parallel.
<b>Abfragen von historisch korrekten Zeitreihen</b>	Sehr einfach in jeglicher Granularität, aufgrund der vollständigen Historisierung in den Satellite-Tables der Schicht «Core».

# ETL Prozesse mit ODI 12c

Live Demo



2013 © Trivadis

ODI 12c und Data Vault  
24.07.2014

**trivadis**  
makes IT easier. ■ ■ ■



## ■ Live-Demo

Wie funktioniert all dies im Prototyp?

- Topology
- Models
- Mapping



2013 © Trivadis

ODI 12c und Data Vault  
24.07.2014

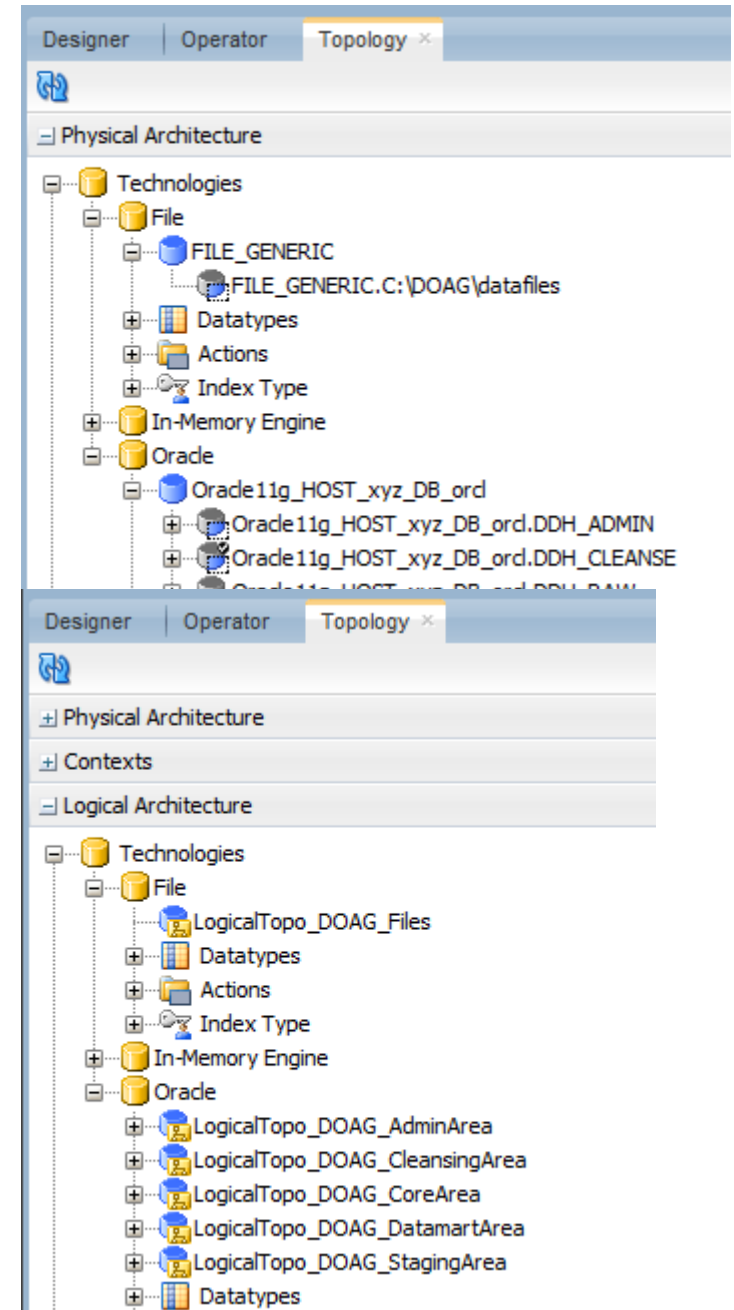
# ■ ODI - Topology

## „Physical Architecture“:

- Verbindungen zu den Files und zur Datenbank und -Schemas

## „Logical Architecture“:

- Logische Schemas über Kontext verknüpft mit den physikalischen Schemas



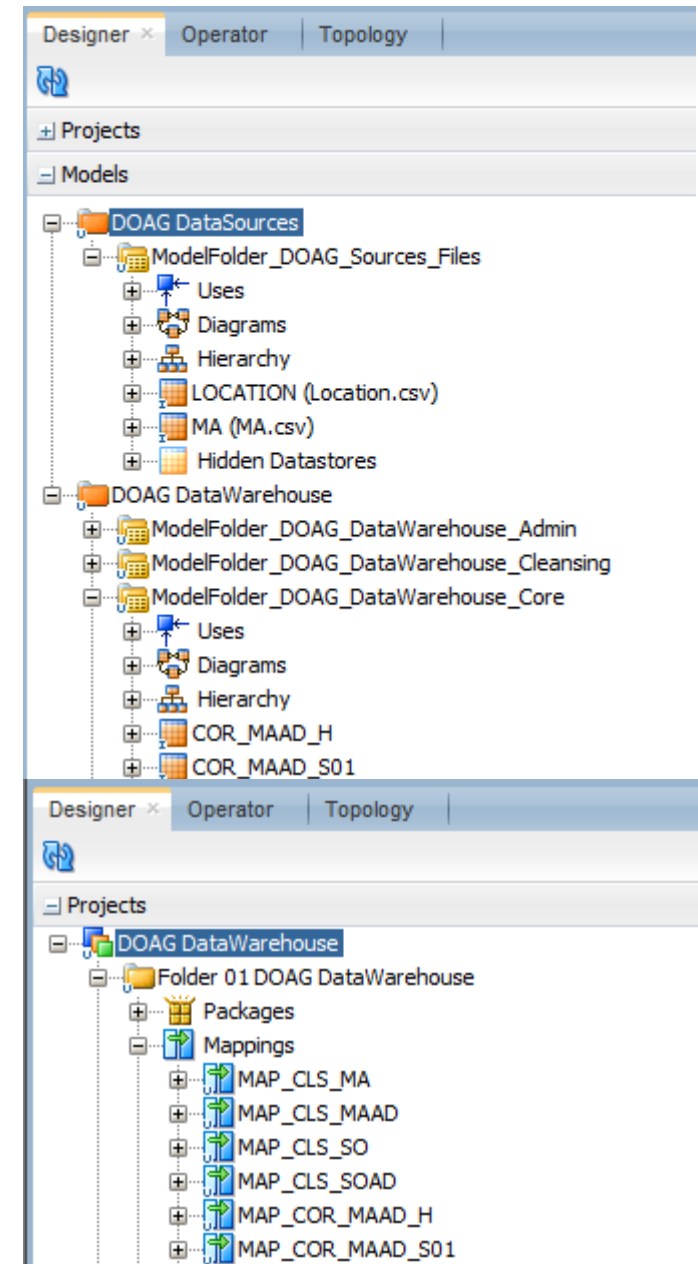
# ■ ODI - Designer

## „Models“:

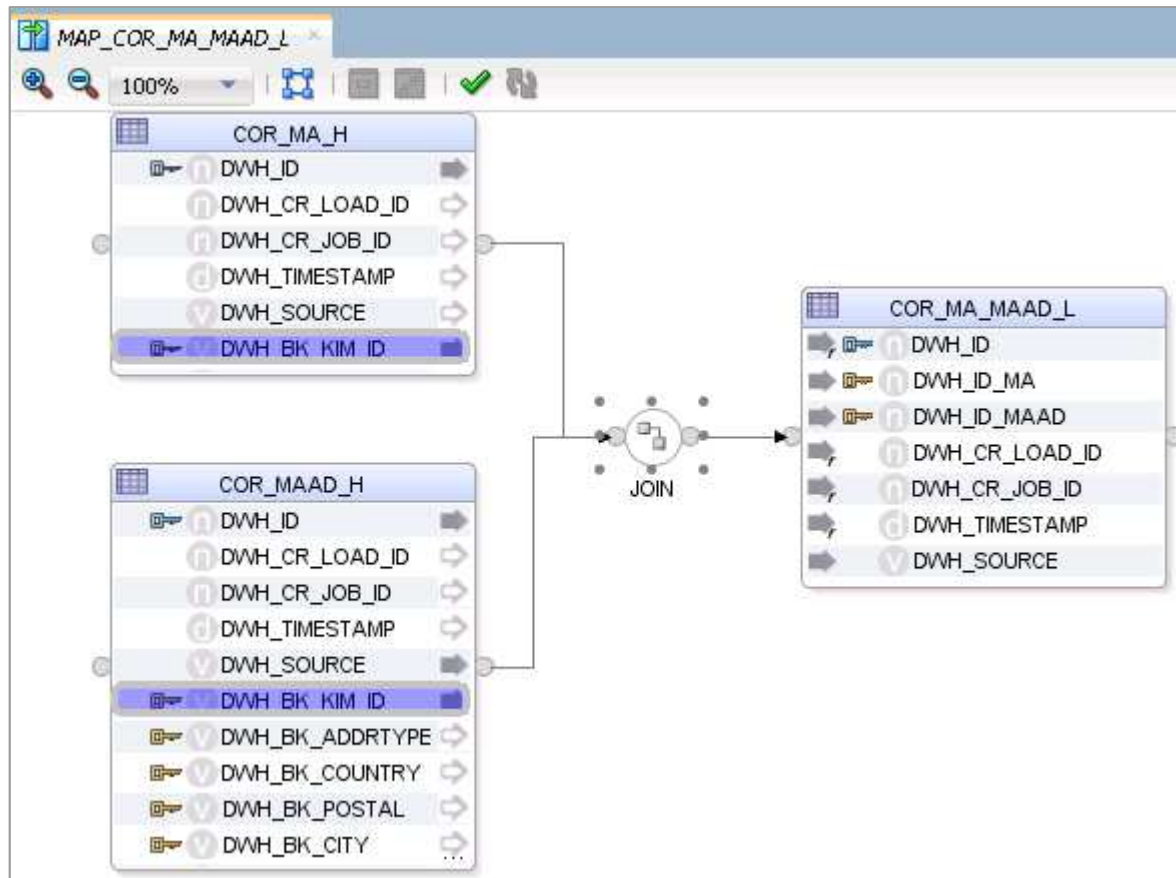
- Reengineering von Dateien und Datenbankobjekten basierend auf der logischen Architektur

## „Projects“:

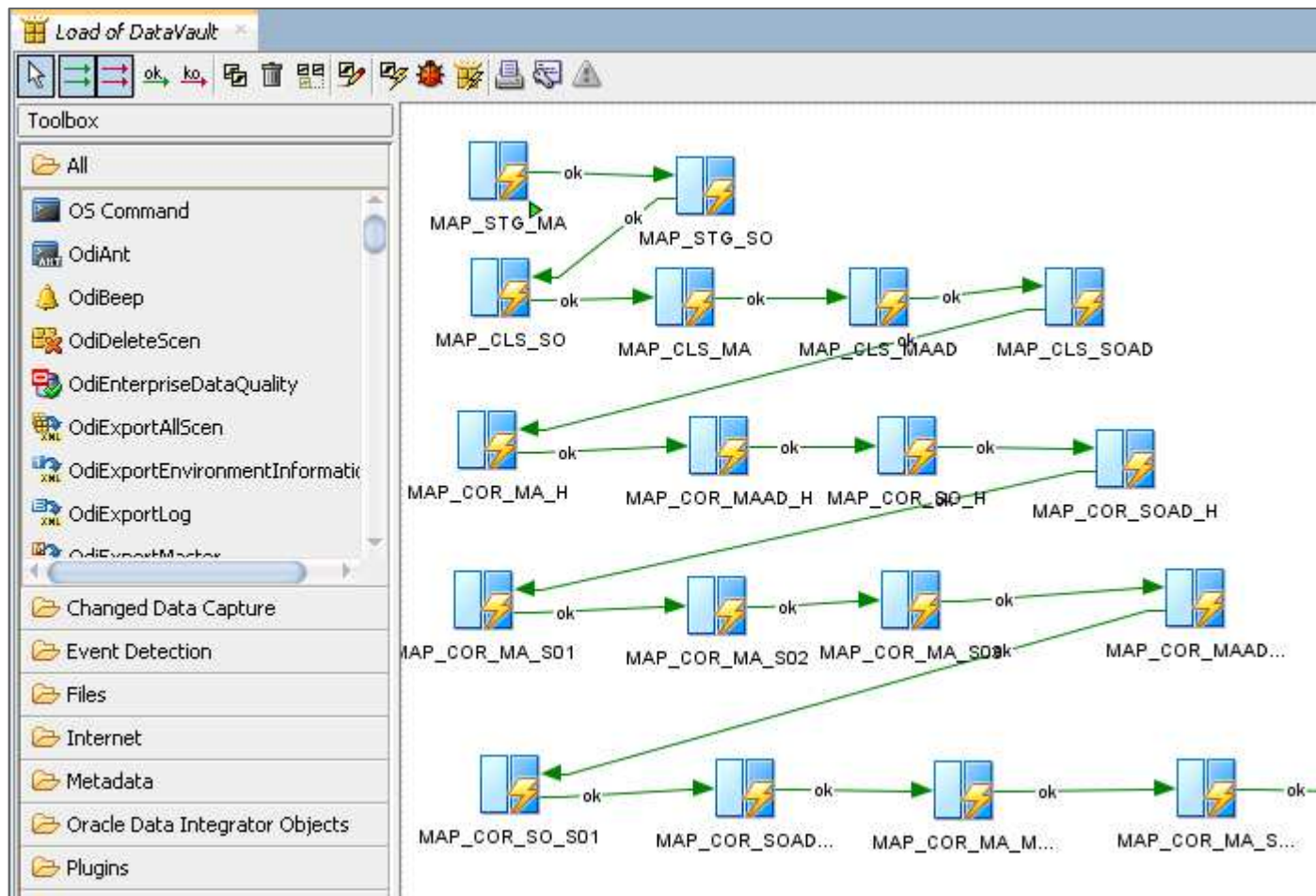
- Mappings (ETL-Strecken)
- Knowledge-Module (generische Ladestrategien)
- Packages (Ladeflüsse)



# ■ Mapping



# Package



# Knowledge-Module (Code Templates)

Funktionsweise



2013 © Trivadis

ODI 12c und Data Vault  
24.07.2014

**trivadis**  
makes IT easier. ■ ■ ■

# ■ Typen von Knowledge-Modulen (KM's)

Es gibt unterschiedliche KM-Typen für

- Reverse-Engineering (RKM's)
- Loading (LKM's)
- Check (CKM's)
- Integration (IKM's)
- ..

KM's für verschiedenste Technologien werden mitgeliefert

KM's können kopiert und angepaßt werden

KM's können von Grund auf neu entwickelt werden

Einem Mapping werden ein oder mehrere KM's zugeordnet



# ■ KM (1)

The screenshot shows the Oracle ODI Knowledge Module configuration window. The browser tabs at the top are 'Start Page' and 'IKM Oracle DataVault Load Hub-Table'. The main title is 'Knowledge Module [Project: DOAG DataWarehouse]'. A left-hand navigation menu includes 'Definition', 'Tasks', 'Options', 'Markers', 'Memo', 'Version', 'Privileges', and 'Flexfields'. The 'Definition' section is active, showing the following configuration:

Name:	IKM Oracle DataVault Load Hub-Table
Type:	IKM - Integration Knowledge Module
Integration Type:	Control Append
	<input type="checkbox"/> Default KM for this pair of Technologies
	<input type="checkbox"/> Multi-Connections
Source Technology:	<Any Technology>
Target Technology:	Oracle

Description:

COMPONENT NAME: IKM Oracle DataVault Load Hub-Table

Author: Nadja Pansa & Claus Jordan (Trivadis GmbH)  
Version: 1.0  
Last Update: Februar 2014

Description:

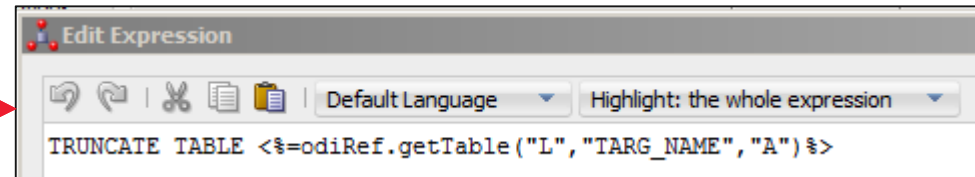
- Integration Knowledge Module
- Loading for DataVault Hub-Tables
- New rows of Hub are inserted



## ■ KM (2)

Ein KM besteht aus Tasks (Einzelschritte), die nacheinander abgearbeitet werden, z.B.

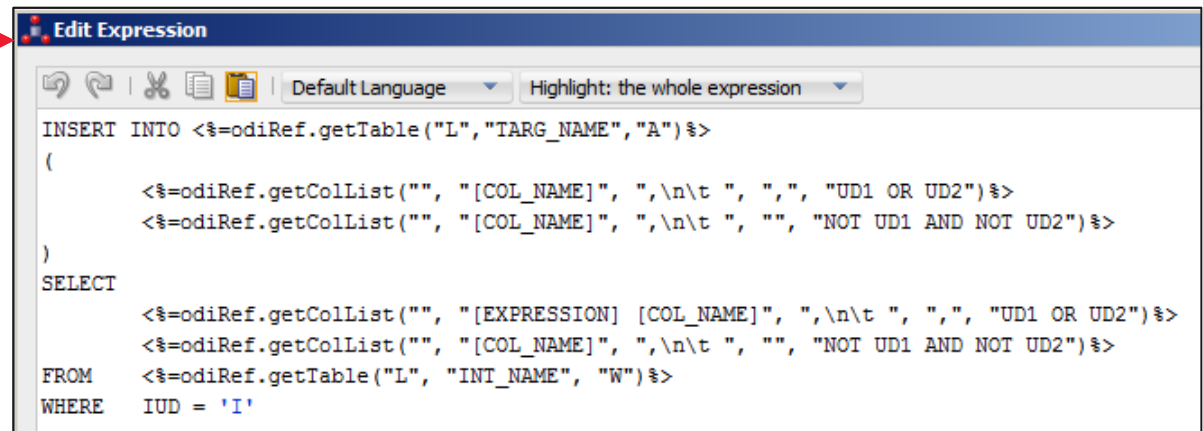
1. Truncate Target Table →



```
TRUNCATE TABLE <%=odiRef.getTable("L","TARG_NAME","A")%>
```

2. Deltadetection

3. Insert Target Table →



```
INSERT INTO <%=odiRef.getTable("L","TARG_NAME","A")%>
(
    <%=odiRef.getColList("", "[COL_NAME]", ",\n\t ", ",", "UD1 OR UD2")%>
    <%=odiRef.getColList("", "[COL_NAME]", ",\n\t ", "", "NOT UD1 AND NOT UD2")%>
)
SELECT
    <%=odiRef.getColList("", "[EXPRESSION] [COL_NAME]", ",\n\t ", ",", "UD1 OR UD2")%>
    <%=odiRef.getColList("", "[COL_NAME]", ",\n\t ", "", "NOT UD1 AND NOT UD2")%>
FROM <%=odiRef.getTable("L","INT_NAME","W")%>
WHERE IUD = 'I'
```

4. Update Target Table

5. Commit

# Zusammenfassung



## ■ Zusammenfassung - Prototyp HR Data Warehouse

- Mit den **Knowledge Modulen** von ODI können alle möglichen Ladestrategien und Sonderfälle effizient abgebildet werden
- Durch die **mehrschichtige Architektur** des HR Data Warehouse sind Datenströme und Transformationen sehr gut nachvollziehbar. Außerdem sind die Mappings dadurch wenig komplex.
- Das **Data Vault** Datenmodell ist ideal ..
  - .. wenn eine **lückenlose Historisierung** der Daten erwünscht ist
  - .. wenn **wenig Datenredundanz** und damit hohe Datenkonsistenz notwendig ist
  - .. wenn der **Aufwand** für das Hinzufügen neuer Entitäten, Attribute und Beziehungen möglichst **klein** sein soll
  - .. wenn **rückwirkende Änderung** von Stamm- und Bewegungsdaten jederzeit möglich sein sollen
  - .. Wenn **parallel entwickelt** werden soll

## ■ Weitere Informationen

- Nächstes ODI 12c Training in Stuttgart  
am 18.08.2014 (3 Tage) – **Sommeraktion: 20 % Rabatt!!**  
<http://www.trivadis.com/training/oracle-training/business-intelligence/oracle-data-integrator-workshop-fuer-praktiker-o-odi.html>
- Trivadis Whitepaper  
„Comparison of Data Modeling Methods for a Core Data Warehouse“  
[http://www.trivadis.com/uploads/tx\\_cabagdownloadarea/Comparison\\_DWH\\_Core\\_Modeling.pdf](http://www.trivadis.com/uploads/tx_cabagdownloadarea/Comparison_DWH_Core_Modeling.pdf)



2013 © Trivadis

ODI 12c und Data Vault  
24.07.2014

**trivadis**  
makes IT easier. ■ ■ ■

# Fragen und Antworten...

Claus Jordan  
Senior Consultant  
Stuttgart

Tel. +49-162-295 96 43  
claus.jordan@trivadis.com



BASEL BERN BRUGG LAUSANNE ZÜRICH DÜSSELDORF FRANKFURT A.M. FREIBURG I.BR. HAMBURG MÜNCHEN STUTTGART WIEN

2013 © Trivadis

ODI 12c und Data Vault  
24.07.2014

**trivadis**  
makes IT easier. ■ ■ ■