

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
15.05.2014

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

■ AGENDA

- Von den Anforderungen zum Data Vault Datenmodell
 - Anforderungen
 - Datenmodell
 - Historisierung in Data Vault
 - Bewertung von Data Vault
- Wie wird mit ODI ein ETL-Prozess definiert?
 - Topology
 - Models und Mappings
- Wie funktionieren die Knowledge-Module?



2013 © Trivadis

ODI 12c und Data Vault
15.05.2014

Von den Anforderungen zum Data Vault Datenmodell

Anforderungen

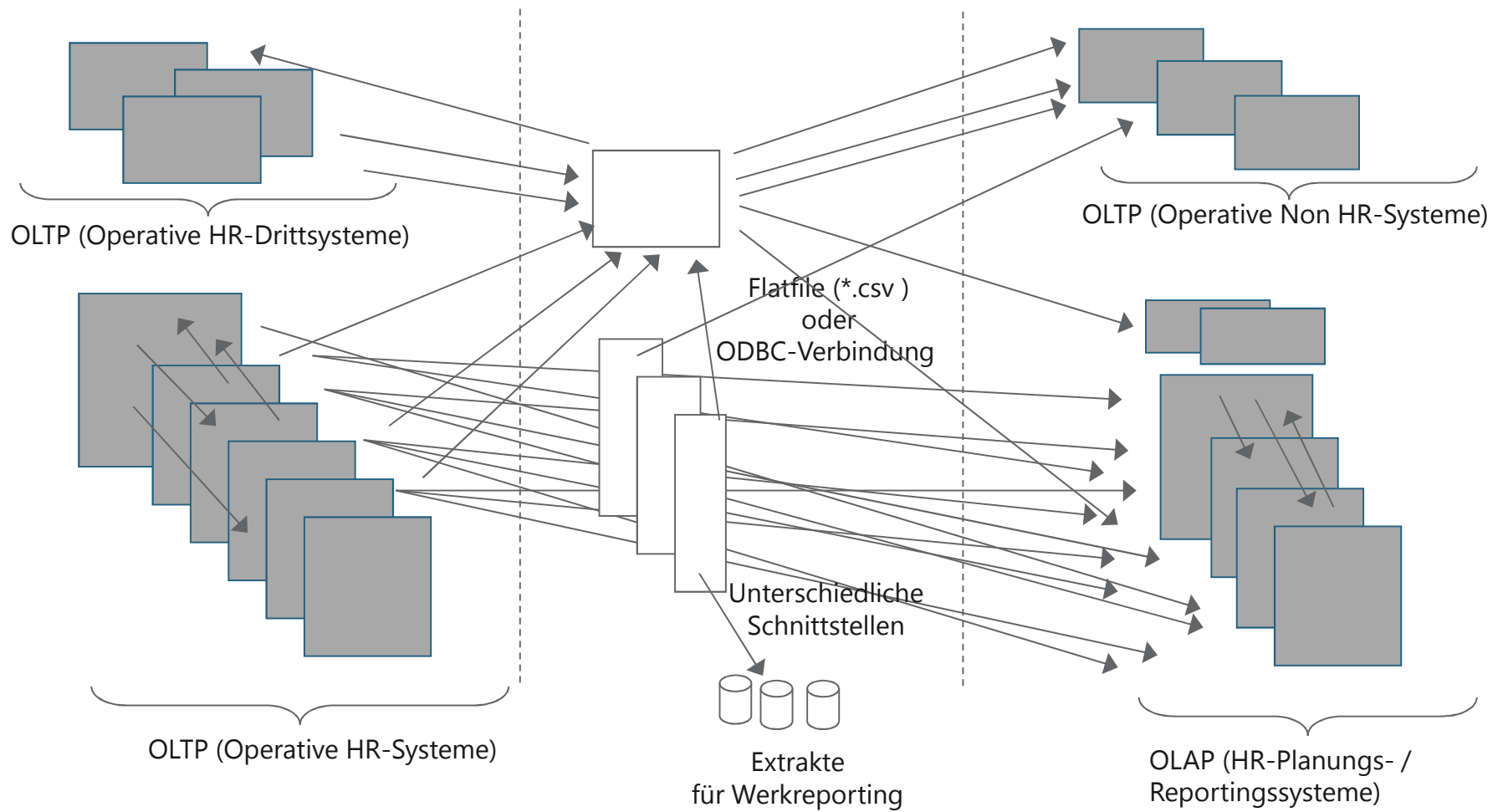


2013 © Trivadis

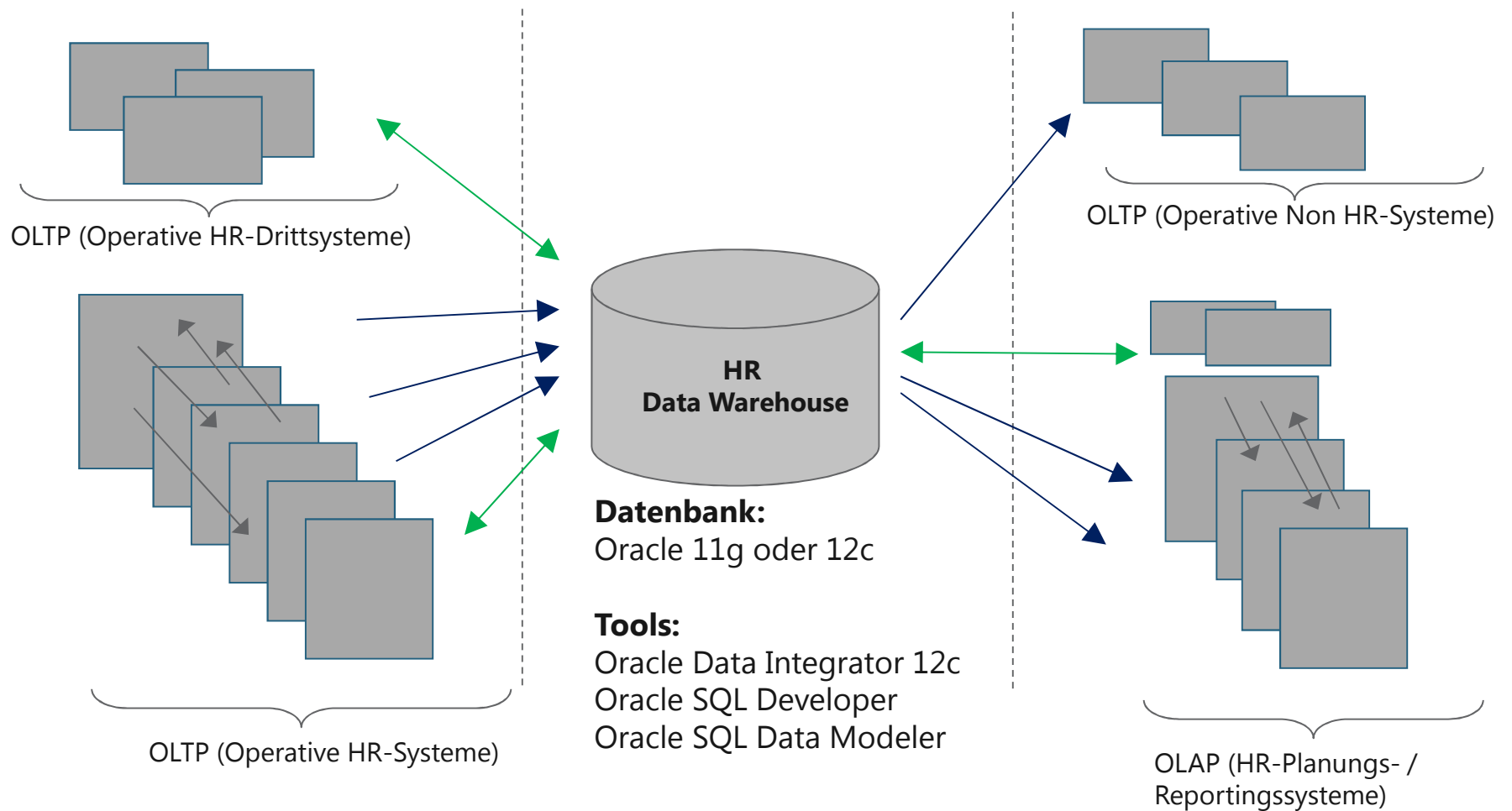
ODI 12c und Data Vault
15.05.2014

trivadis
makes IT easier. ■ ■ ■

■ Systemlandschaft „Heute“



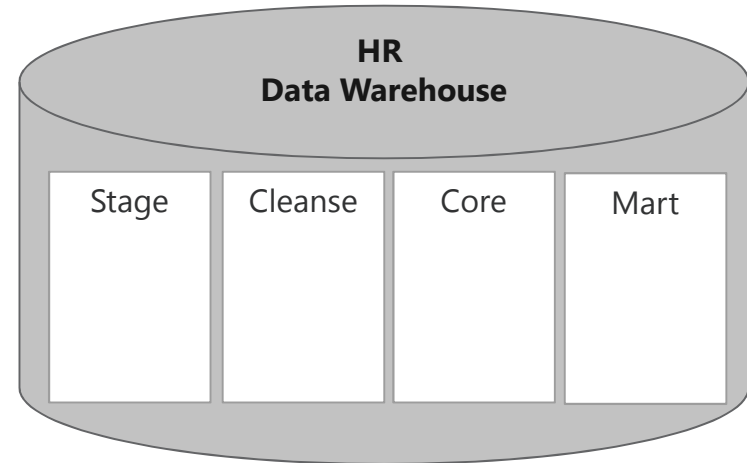
■ Systemlandschaft „Morgen“



■ Konzept und Datenmodell

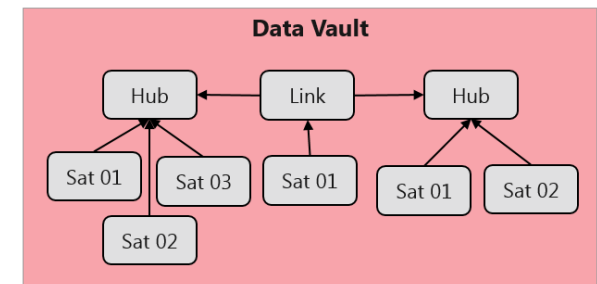
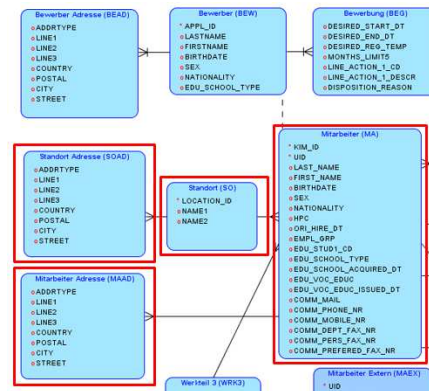
HR Data Warehouse – Konzept (Teil 1)

- Aufbau und Technologie
- Vorgehensweise



HR Data Warehouse – Datenmodell (Teil 2)

- ER Diagramm
- Physische Datenmodelle



■ Umfang und Ziel des Prototyps

Umfang

- Für eine begrenzte Anzahl von Entitäten und Beziehungen soll mit Hilfe des ETL Werkzeugs ODI ein Prototyp für das HR Data Warehouse erstellt werden.
- Für für die zentrale Schicht (Core) soll die Modellierungsmethode Data Vault zum Einsatz kommen.

Ziel

Es soll am praktischen Beispiel gezeigt werden,

- dass ODI das richtige Werkzeug und
- dass Data Vault für die aktuellen und zukünftigen Anforderung geeignet ist.

Von den Anforderungen zum Data Vault Datenmodell

Datenmodell



2013 © Trivadis

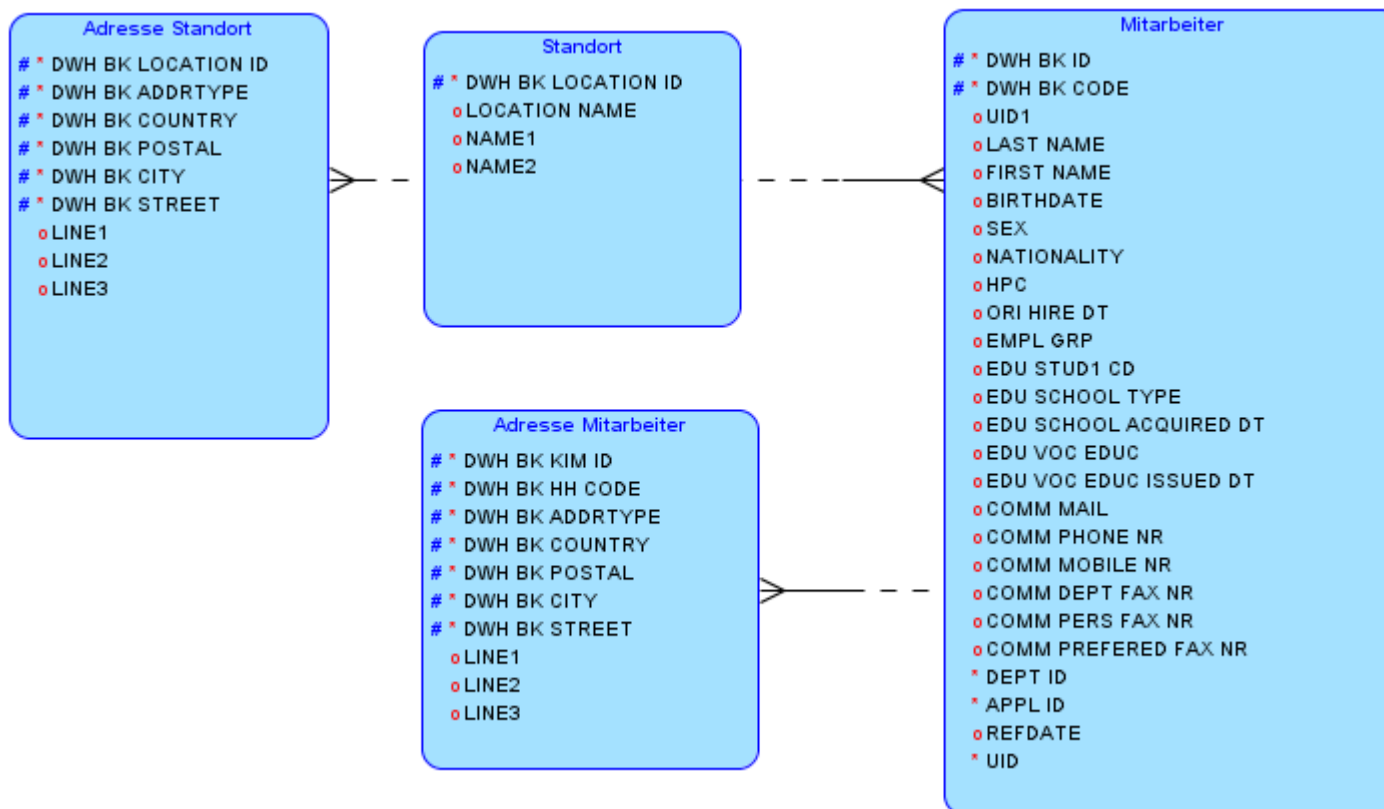
ODI 12c und Data Vault
15.05.2014

trivadis
makes IT easier. ■ ■ ■

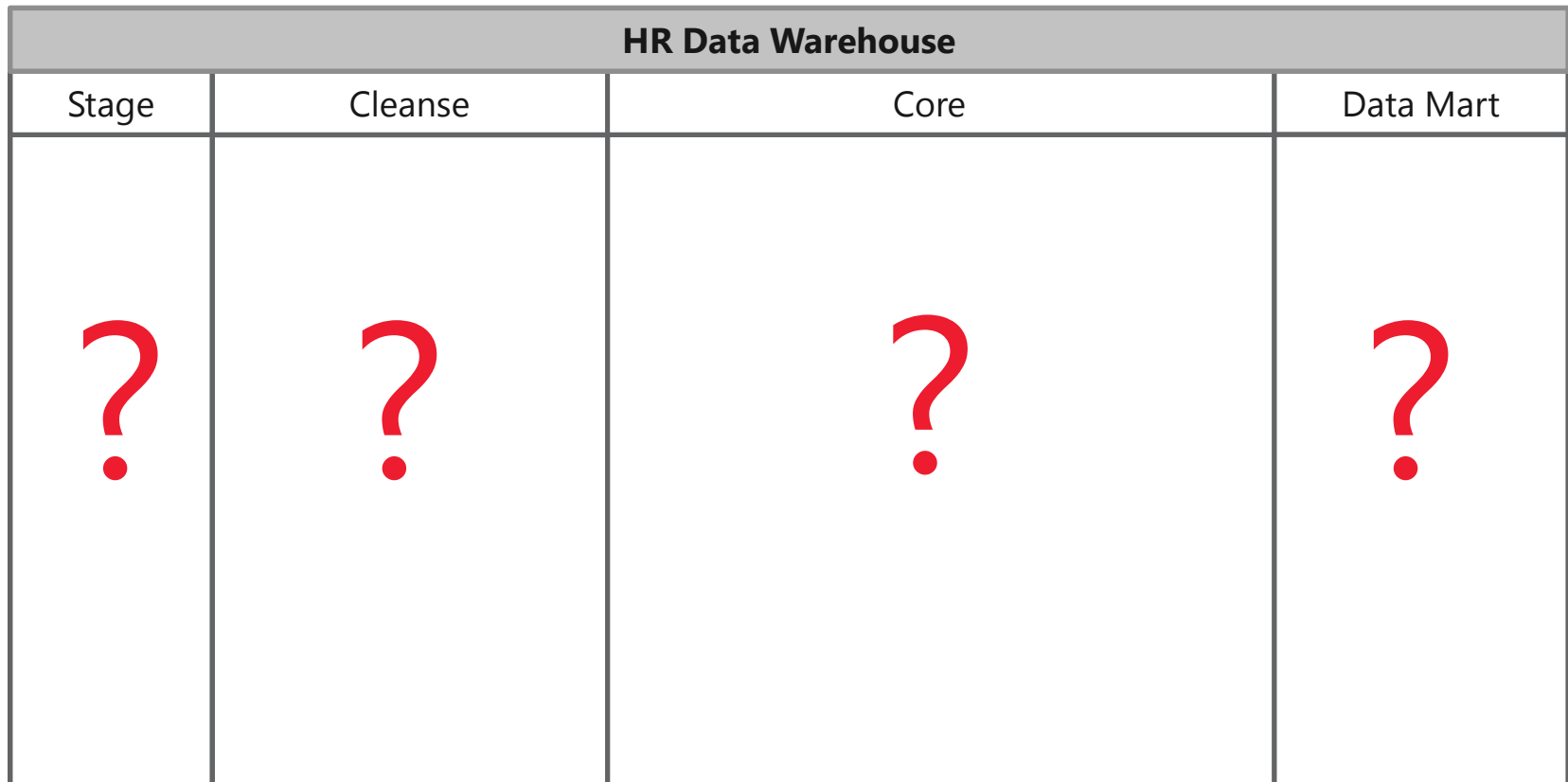
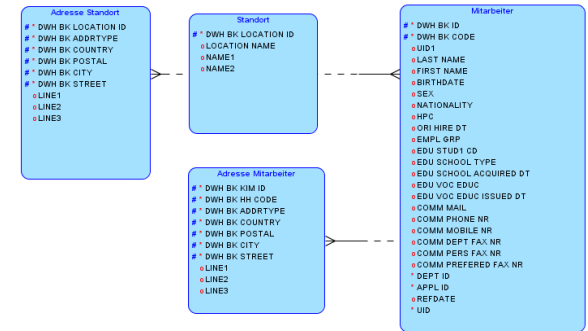
■ Ausgangspunkt ist das logisches Datenmodell ..

.. für die Objekte

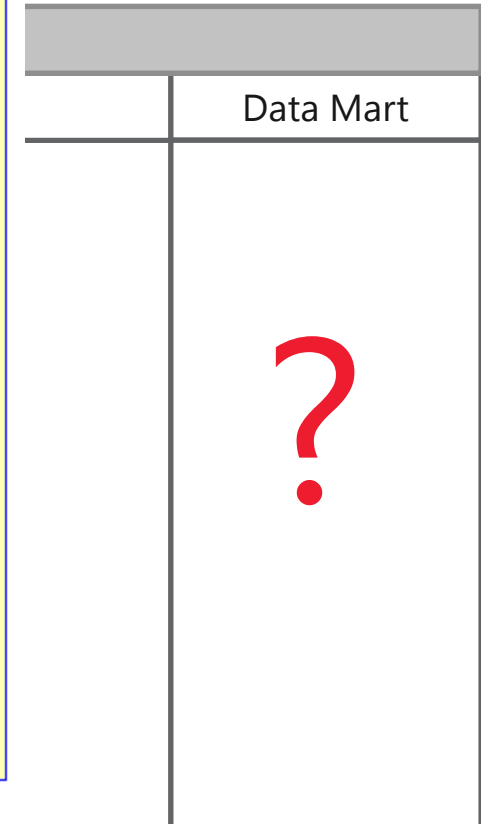
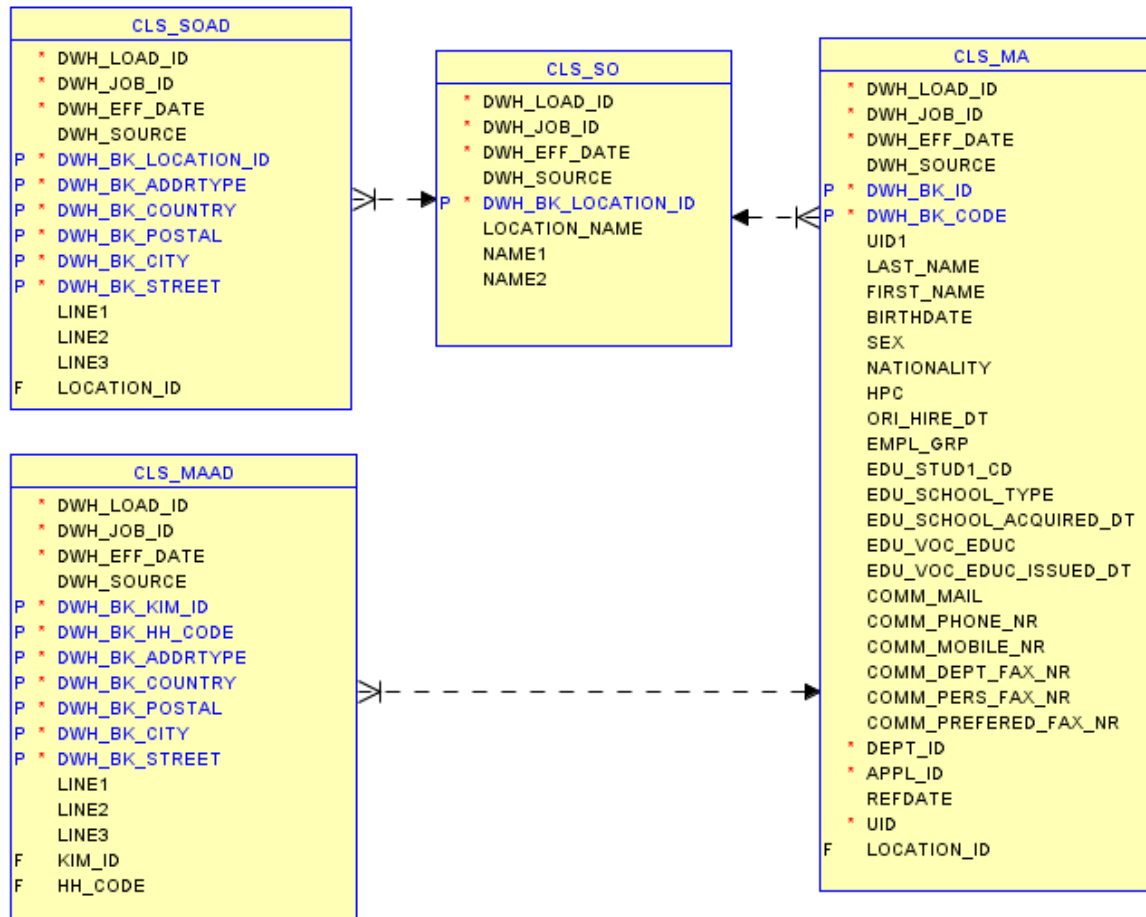
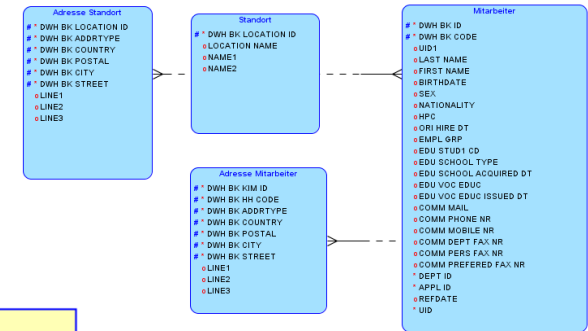
Mitarbeiter + Adresse(n), Standort + Adresse(n)

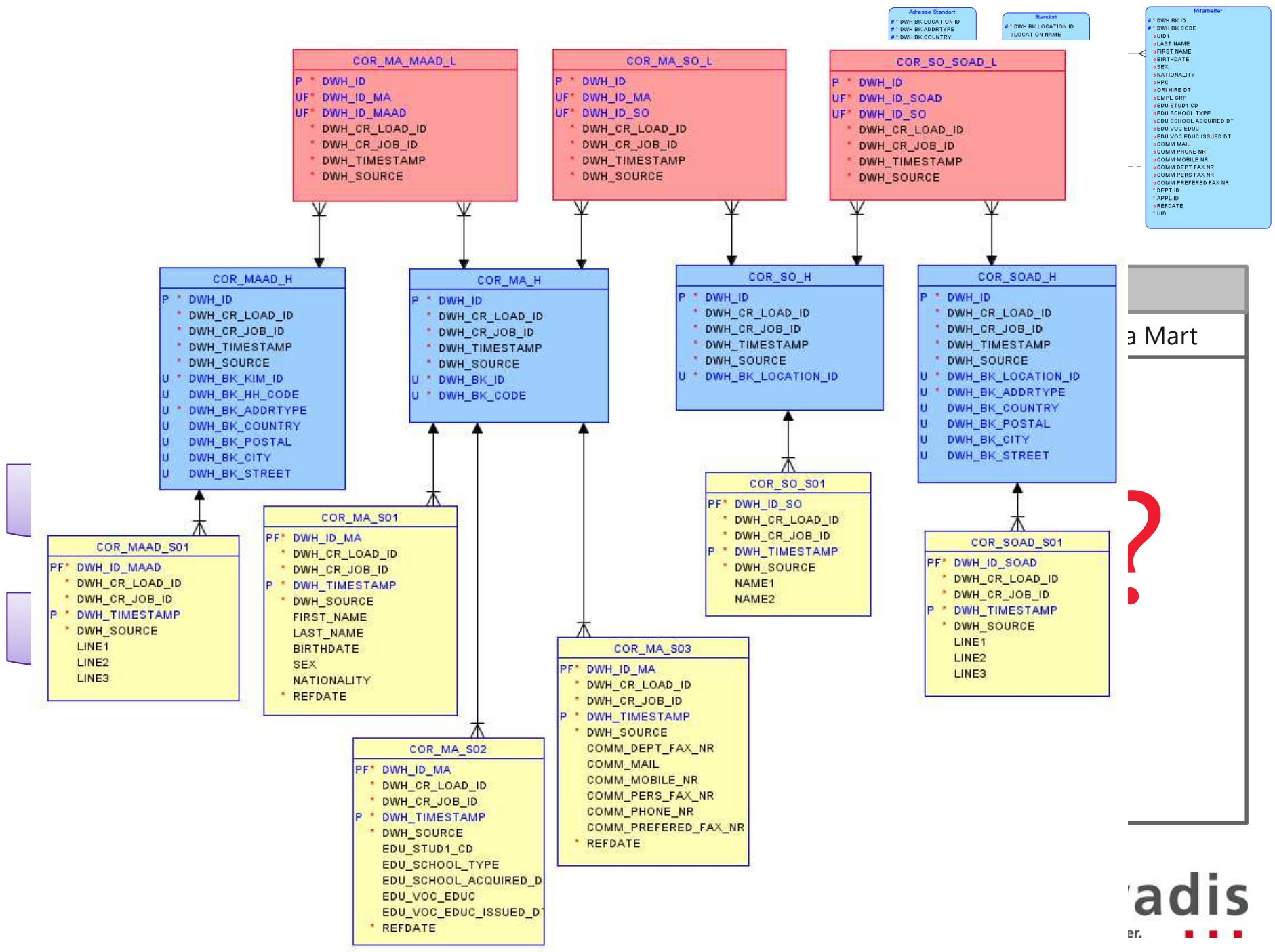


■ HR DWH - Physisches Datenmodell



1. Cleanse





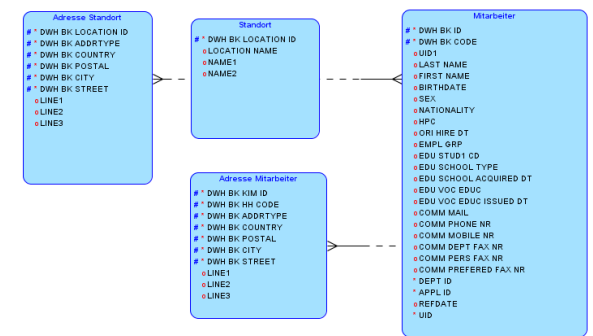
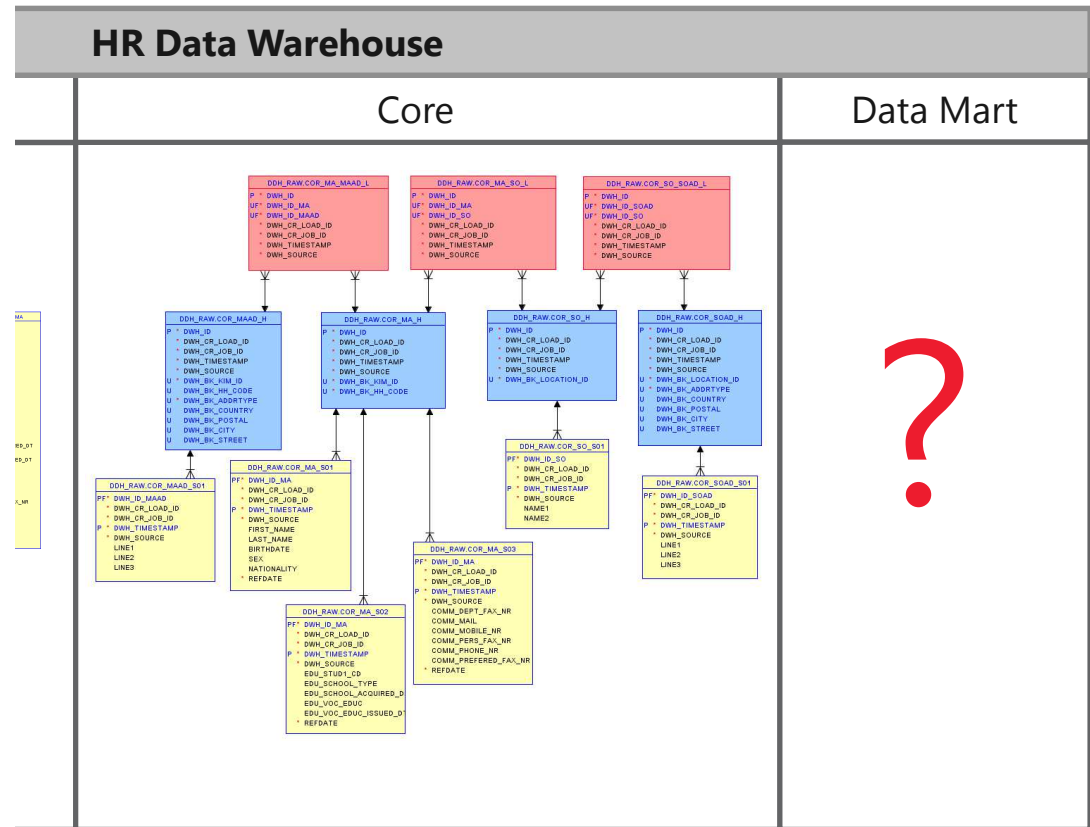
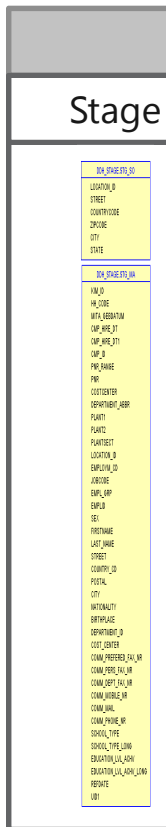
a Mart

?

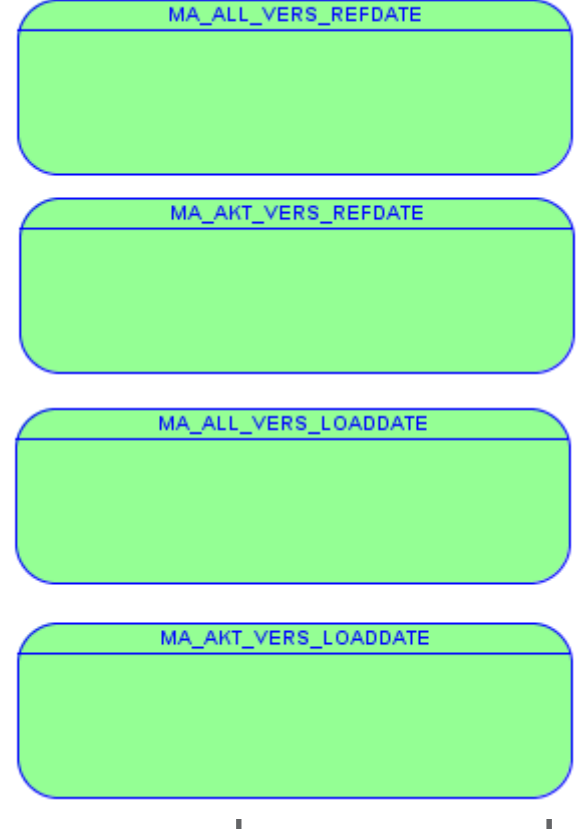
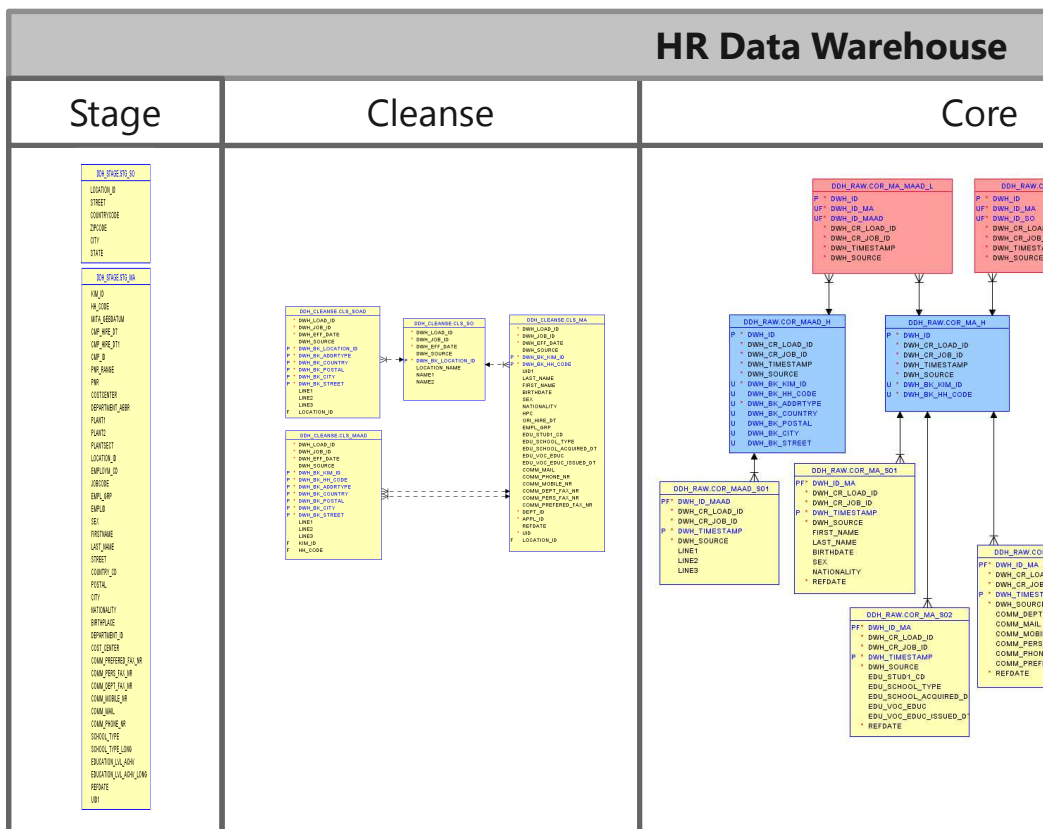
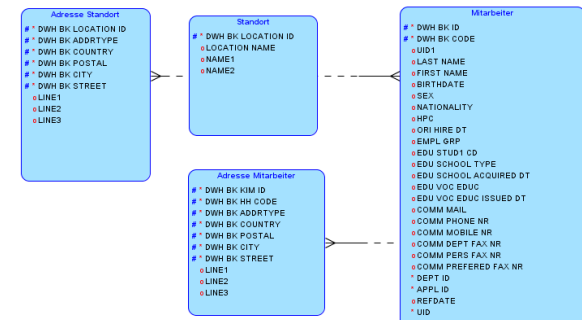
3. Stag

STG_SO
LOCATION_ID
STREET
COUNTRYCODE
ZIPCODE
CITY
STATE

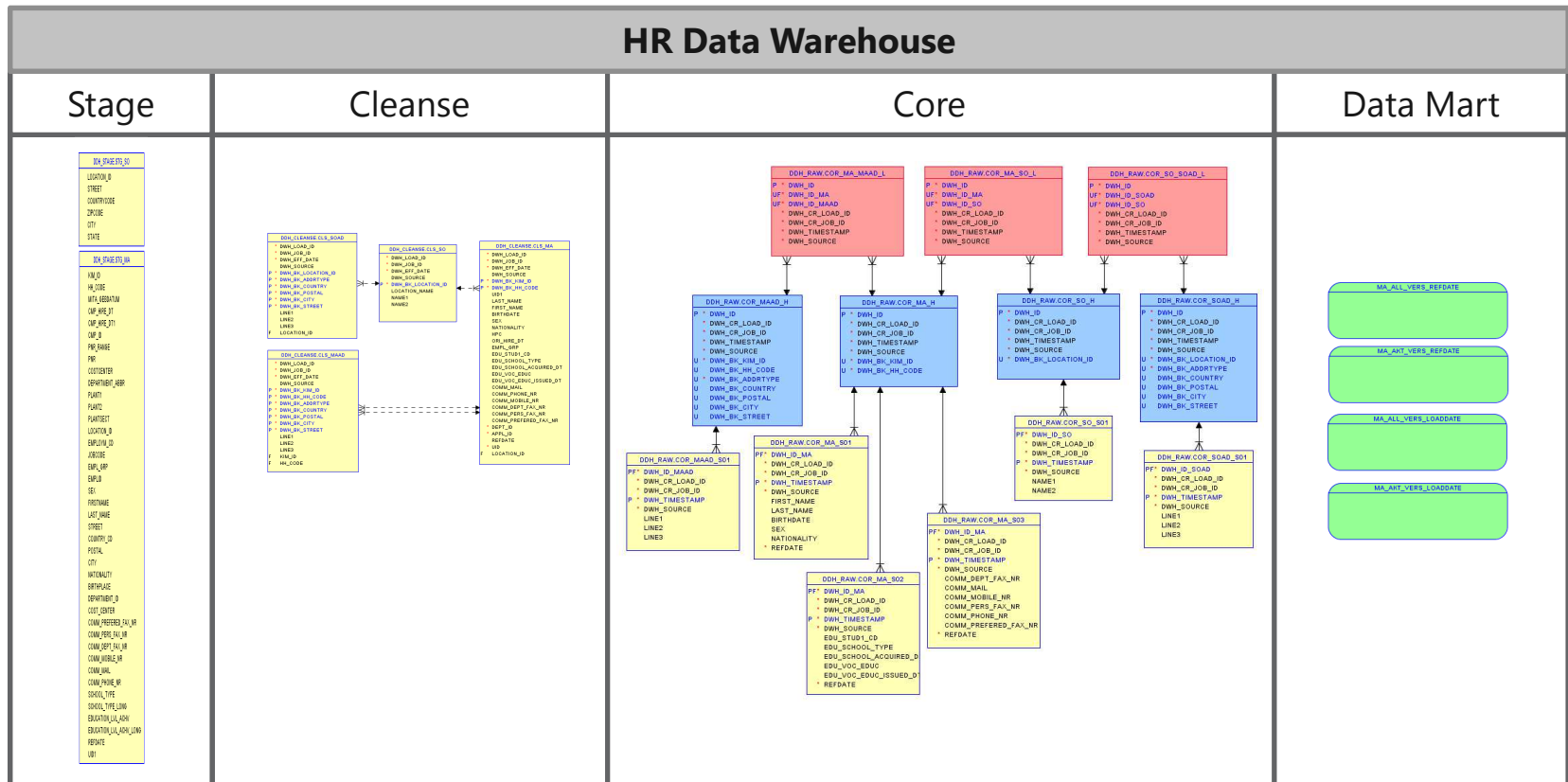
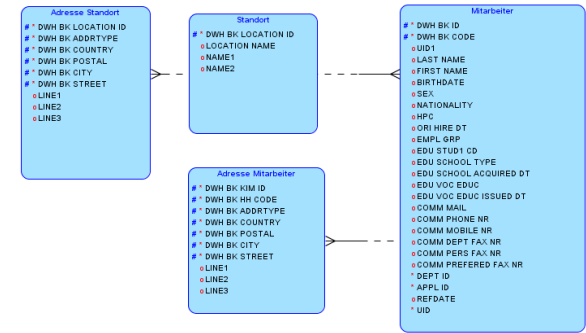
STG_MA
ID
CODE
MITA_GEBDATUM
CMP_HIRE_DT
CMP_HIRE_DT1
CMP_ID
PNR_RANGE
PNR
COSTCENTER
DEPARTMENT_ABBR
PLANT1
PLANT2
PLANTSECT
LOCATION_ID
EMPLOYM_CD
JOBCODE
EMPL_GRP
EMPLID
SEX
FIRSTNAME
LAST_NAME
STREET
COUNTRY_CD
POSTAL
CITY
NATIONALITY
BIRTHPLACE
DEPARTMENT_ID
COST_CENTER
COMM_PREFERRED_FAX_NR
COMM_PERS_FAX_NR
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
15.05.2014

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 den DataHub 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 (BK)	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 (Ladelauf 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 laut Ladelauf vom **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
100	15.02.2014	Maier, Beate	verheiratet	60021	31.12.2013
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 eingefügt 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
100	15.02.2014	Maier, Beate	verheiratet	60021	31.12.2013

Von den Anforderungen zum Data Vault Datenmodell

Bewertung von Data Vault



2013 © Trivadis

ODI 12c und Data Vault
15.05.2014

trivadis
makes IT easier. ■ ■ ■

■ Bewertung von Data Vault (1)

Methode Kriterien \	Data Vault
Anzahl Tabellen und ETL-Prozesse	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.
Komplexität für das Laden aus Cleanse in Core	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.
Komplexität für das Laden aus Core in Data Marts	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
Erweiterbarkeit („Agile BI“)	Sehr gut, aufgrund fehlender Referenzen zwischen Entitäten, die jeweils unabhängig voneinander erweitert oder angepaßt werden können.
Datenredundanz	Sehr gering, durch Normalisierung und Splittung der Attribute einer Entität in mehreren Satellitentabellen.
Datenvolumen	Zumindest für Stammdaten minimal, aufgrund der geringen Datenredundanz
Parallelisierbarkeit	Sehr gut beim Laden des Core. Sämtliche Hub-Tables können parallel geladen werden, danach alle Link-Tables und anschließend alle Satellite-Tables jeweils parallel.
Abfragen von historisch korrekten Zeitreihen	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
15.05.2014

trivadis
makes IT easier. ■ ■ ■

■ Live-Demo

Wie funktioniert all dies im Prototyp?

- Topology
- Models
- Mapping



2013 © Trivadis

ODI 12c und Data Vault
15.05.2014

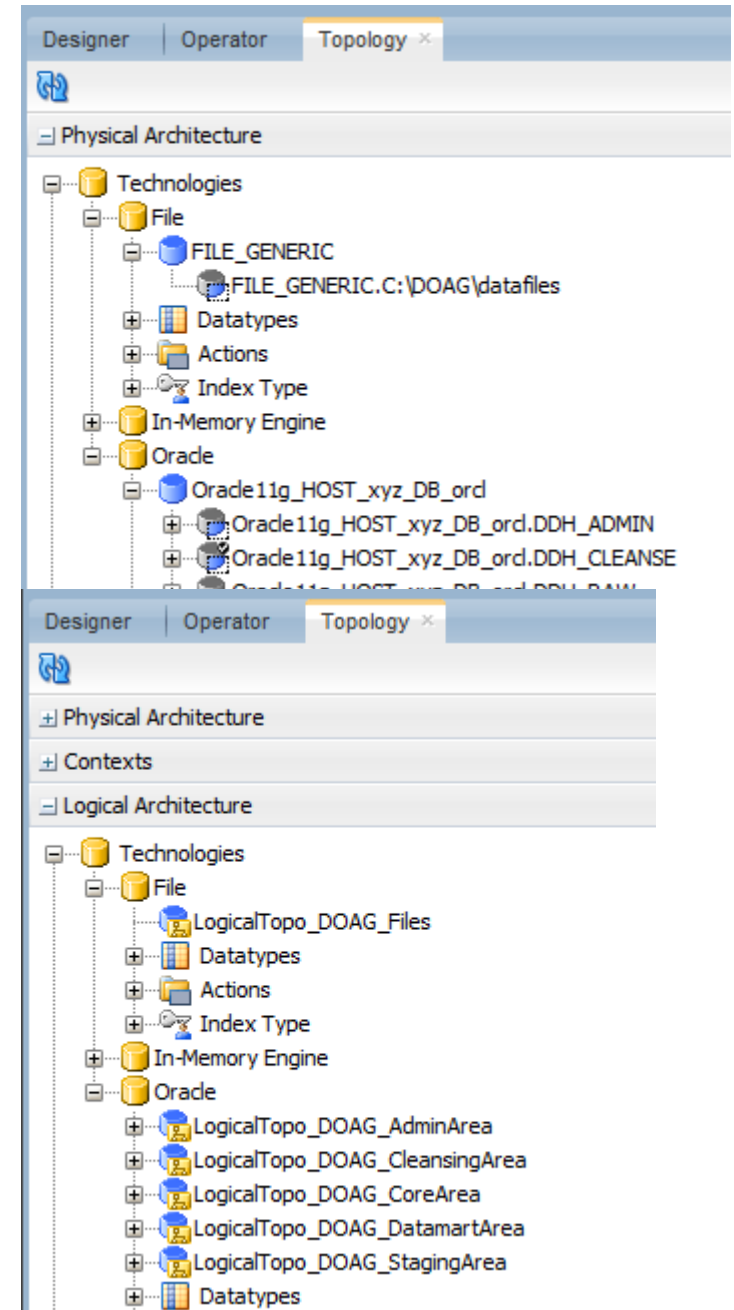
■ ODI - Topology

Physical Architectur:

- Verbindungen zu den Files und zur Datenbank und -Schemas

Logical Architektur:

- Logische Schemas über Kontext verknüpft mit den physikalischen Schemas



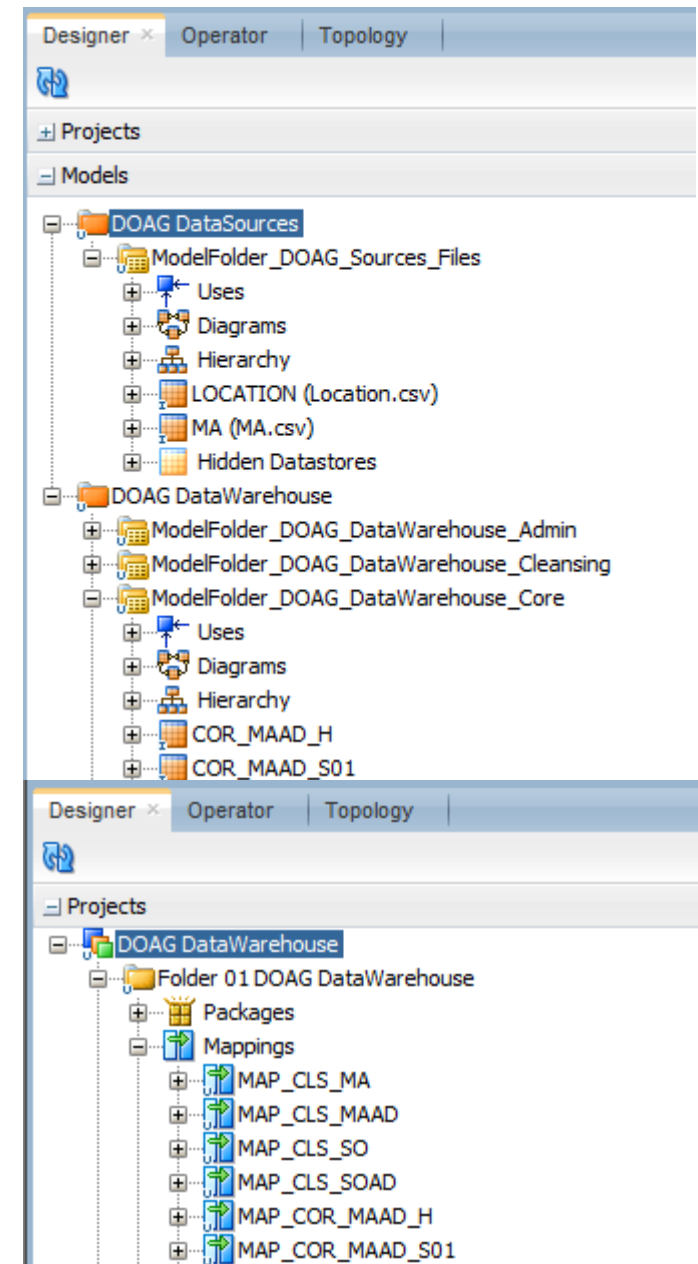
■ ODI - Designer

Models:

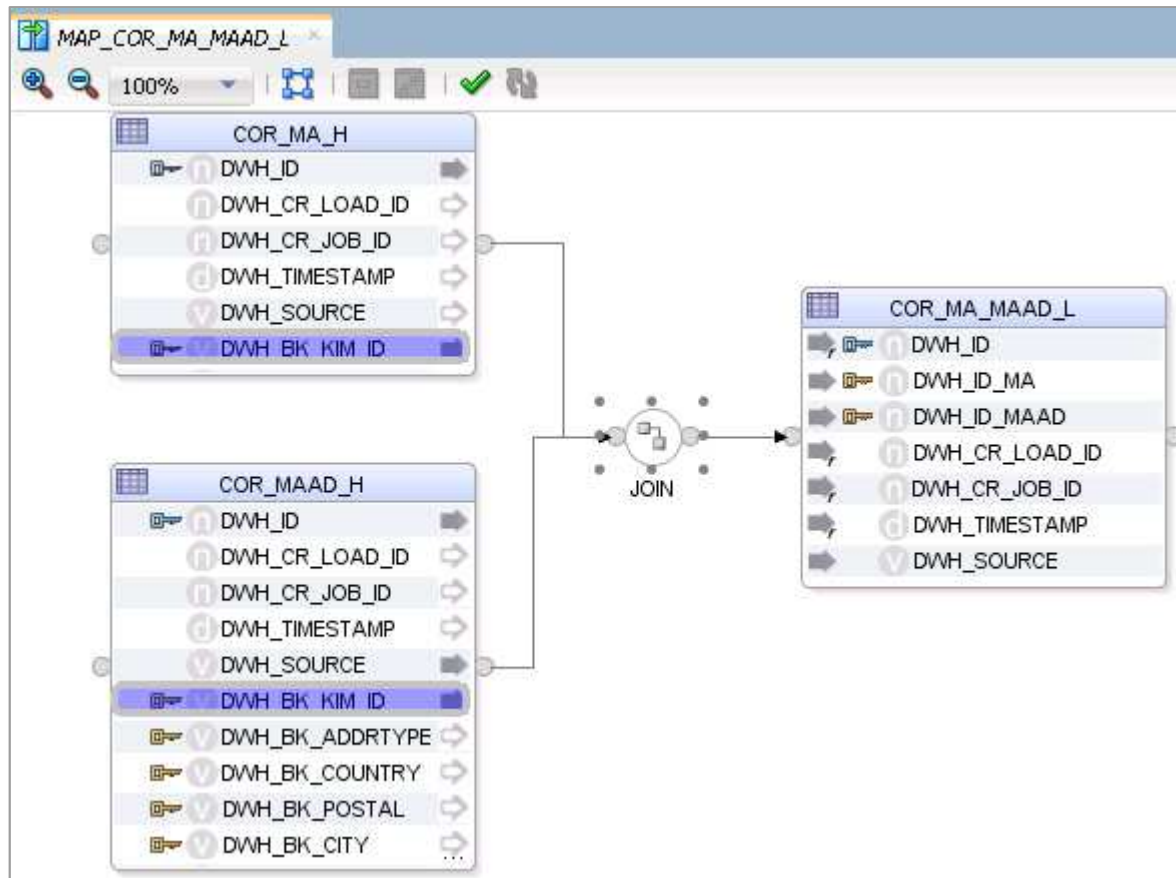
- Reengineering von Dateien und Datenbankobjekten basierend auf der logischen Architektur

Projects:

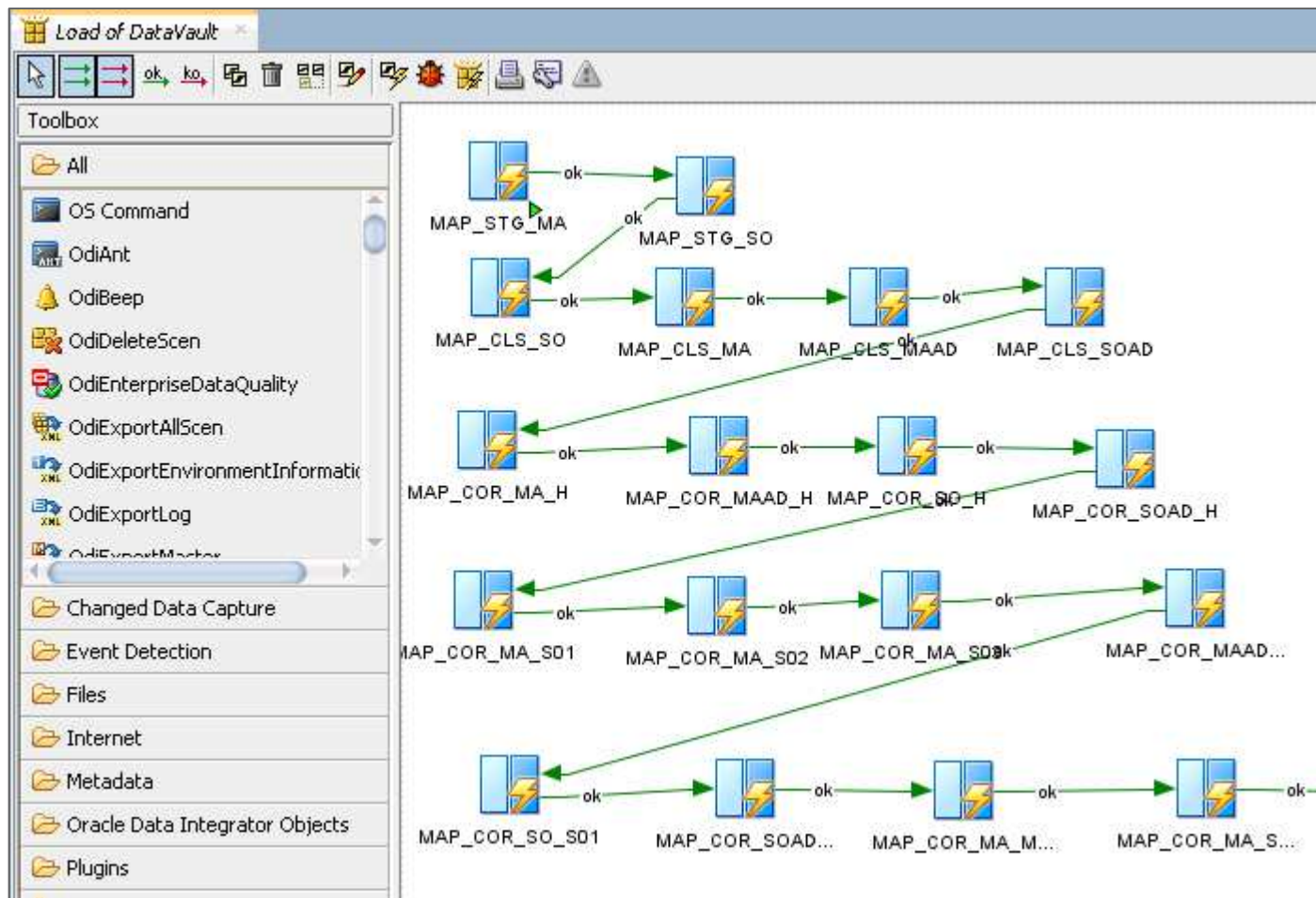
- Mappings (ETL-Strecken)
- Knowledge-Module (generische Ladestrategien)
- Packages (Ladeflüsse)



■ Mapping



■ Package



Knowledge-Module (Code Templates)

Funktionsweise



■ KM Typen

Es gibt unterschiedliche Typen von Knowledge-Modulen (KM's) 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



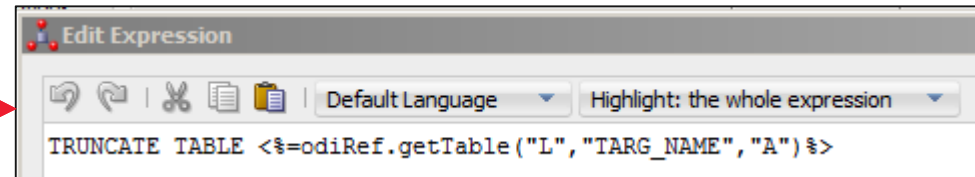
2013 © Trivadis

ODI 12c und Data Vault
15.05.2014

■ KM Tasks (1)

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

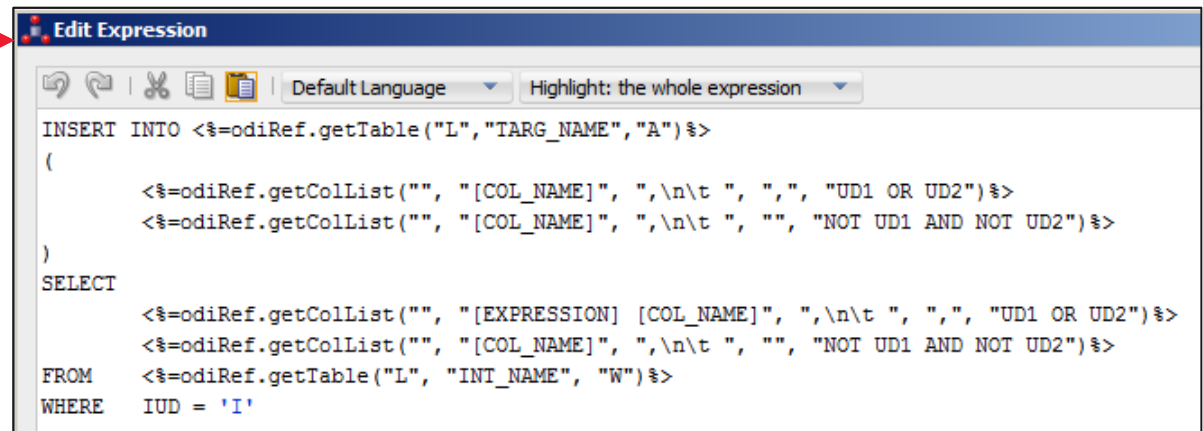
1. Truncate Target Table →



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

2. Deltadetection

3. Insert Target Table →

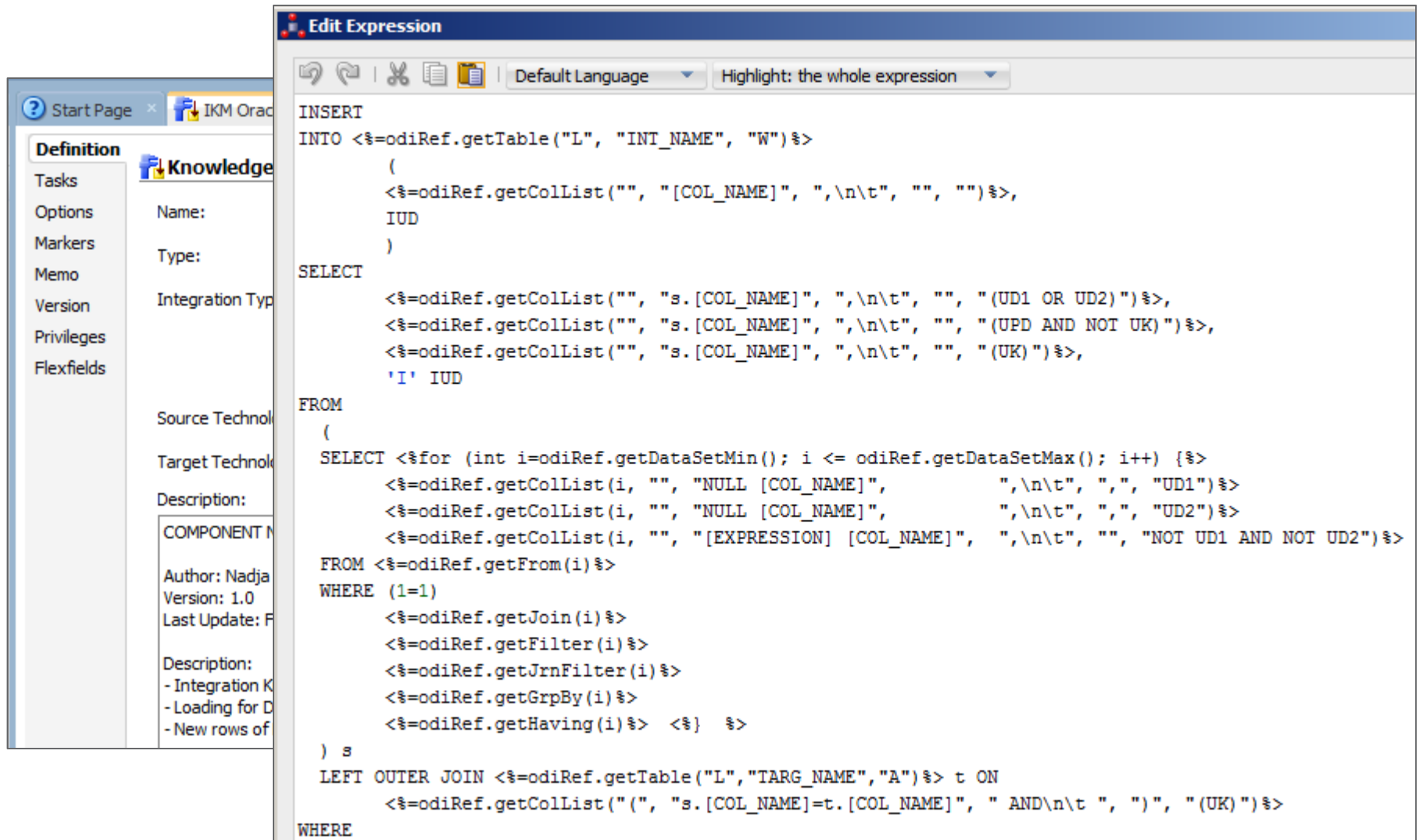


```
INSERT INTO <%=odiRef.getTable("L","TARG_NAME","A")%>
(
    <%=odiRef.getColList("", "[COL_NAME]", ",\n\t ", ",", "UD1 OR UD2")%>
    <%=odiRef.getColList("", "[COL_NAME]", ",\n\t ", "", "NOT UD1 AND NOT UD2")%>
)
SELECT
    <%=odiRef.getColList("", "[EXPRESSION] [COL_NAME]", ",\n\t ", ",", "UD1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

■ KM Tasks (2)



The screenshot shows a software interface with a 'Definition' panel on the left and an 'Edit Expression' window on the right. The 'Definition' panel includes a 'Knowledge' section with fields for Name, Type, Integration Type, Source Technology, Target Technology, and Description. The 'Edit Expression' window displays a complex SQL query with many placeholders like <%=odiRef.getTable...>.

```
INSERT
INTO <%=odiRef.getTable("L", "INT_NAME", "W")%>
(
  <%=odiRef.getColList("", "[COL_NAME]", ",\n\t", "", "")%>,
  IUD
)
SELECT
  <%=odiRef.getColList("", "s.[COL_NAME]", ",\n\t", "", "(UD1 OR UD2)")%>,
  <%=odiRef.getColList("", "s.[COL_NAME]", ",\n\t", "", "(UPD AND NOT UK)")%>,
  <%=odiRef.getColList("", "s.[COL_NAME]", ",\n\t", "", "(UK)")%>,
  'I' IUD
FROM
(
  SELECT <%=for (int i=odiRef.getDataSetMin(); i <= odiRef.getDataSetMax(); i++) {%>
    <%=odiRef.getColList(i, "", "NULL [COL_NAME]", "\n\t", "", "UD1")%>
    <%=odiRef.getColList(i, "", "NULL [COL_NAME]", "\n\t", "", "UD2")%>
    <%=odiRef.getColList(i, "", "[EXPRESSION] [COL_NAME]", "\n\t", "", "NOT UD1 AND NOT UD2")%>
  FROM <%=odiRef.getFrom(i)%>
  WHERE (1=1)
    <%=odiRef.getJoin(i)%>
    <%=odiRef.getFilter(i)%>
    <%=odiRef.getJrnFilter(i)%>
    <%=odiRef.getGrpBy(i)%>
    <%=odiRef.getHaving(i)%> <%> %>
  ) s
LEFT OUTER JOIN <%=odiRef.getTable("L", "TARG_NAME", "A")%> t ON
  <%=odiRef.getColList("", "s.[COL_NAME]=t.[COL_NAME]", " AND\n\t ", "", "(UK)")%>
WHERE
```

Zusammenfassung



2013 © Trivadis
ODI 12c und Data Vault
15.05.2014

■ 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

■ Seminare und Workshops

- Ab Mai finden an verschiedenen Standorten ein eintägiger TechnoCircle zu ODI 12c statt.
<http://www.trivadis.com/training/oracle-training/technocircle-oracle-12c/oracle-data-integrator-12c-aus-zwei-mach-eins-o-odi12c.html>
- Ab dem 30.06.2014 bietet Trivadis ein 3-tägiges ODI Training auf Basis der Version 12c an.
<http://www.trivadis.com/training/oracle-training/business-intelligence/oracle-data-integrator-workshop-fuer-praktiker-o-odi.html>



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
15.05.2014

trivadis
makes IT easier. ■ ■ ■