



## Flexible Schnittstellen für Data Warehousing auf XML Basis

Lutz Bauer  
Leiter Competence Center Data Integration



# MT AG Key Facts

ORACLE Platinum Partner



MT AG MANAGING TECHNOLOGY – ENABLING THE ADAPTIVE ENTERPRISE

- Gründung 1994
- Inhabergeführte AG:  
Aktienkapital 1.500.000 €
- Hauptsitz Ratingen;  
Niederlassungen Dortmund, Frankfurt
- Mitarbeiter:  
> 200 Festangestellte  
> 65 Freie Mitarbeiter
- Full-Service-Dienstleistung für alle Phasen des Software-Lifecycle
- Herstellerunabhängige Expertise in den marktführenden Technologien wie Oracle, IBM, Microsoft, SAP und OpenSource
- Themen- und Lösungs-Know-how in den Kerndisziplinen des Adaptive Enterprise

# Agenda



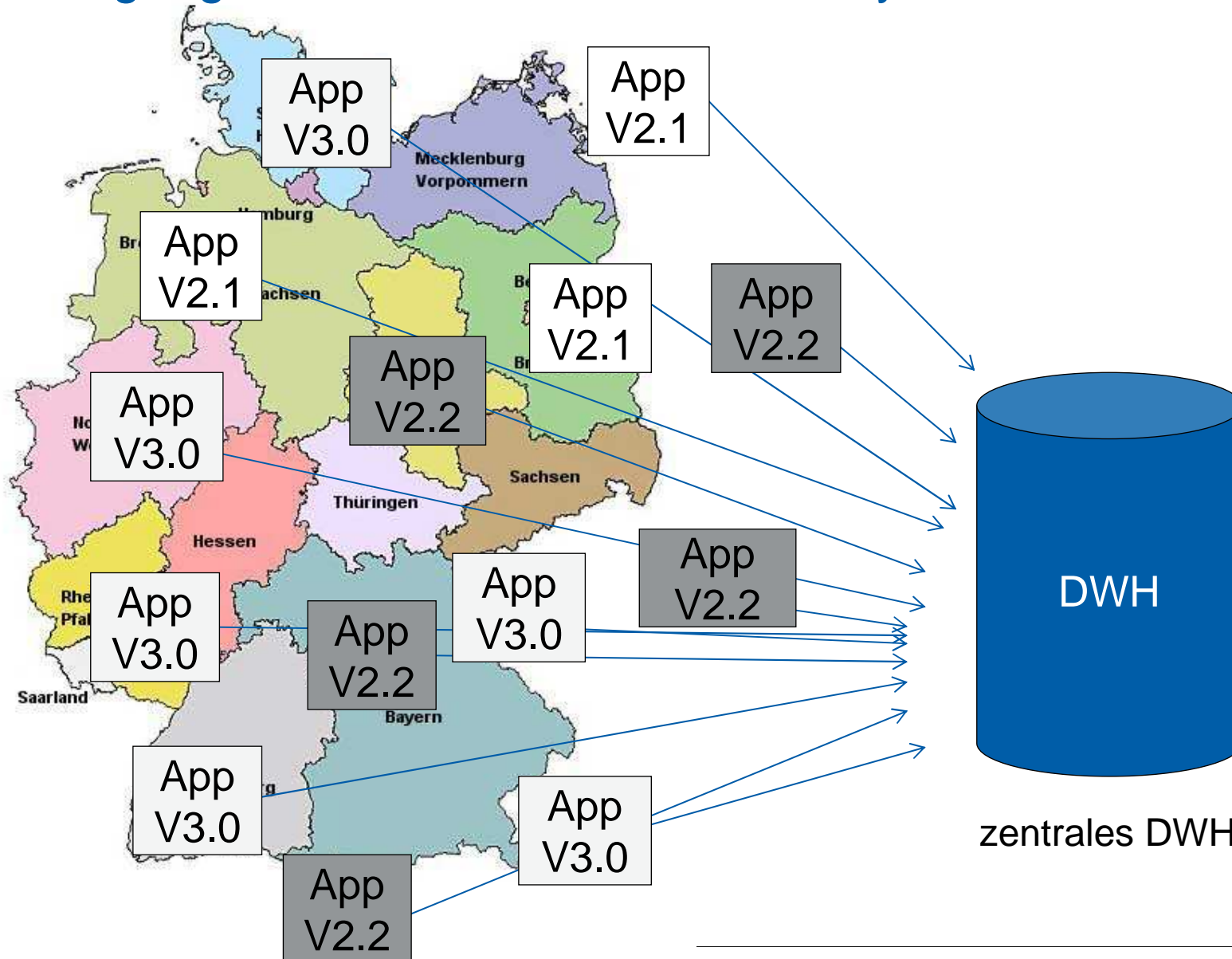
- Ausgangssituation: Dezentrale Quellsysteme
- Anforderungen an „flexible“ Schnittstellen
- XML-basierte Lösung
- Details zur Umsetzung
- Vor-/Nachteile von XML-Schnittstellen für DWH
- Lessons Learned

# Ausgangssituation



- Dezentrale IT in 15 Landesgesellschaften
  - Die operativen Applikationen werden zentral entwickelt oder eingekauft
  - Integration, Betrieb, Releasemanagement, Customizing erfolgt dezentral
  
  - Situation ist ungünstig für DWH:
    - Manueller Pflegeaufwand: Zuordnung Landesgesellschaft zur Import-Schnittstellenversion
    - Customizing an lokalen Standorten führte in Vergangenheit zu ungültigen Dateianlieferungen
-

# Ausgangssituation: Dezentrale Quellsysteme



## Anforderungen an flexible DWH Import-Schnittstellen

- Mehrere Versionsstände einer Quellapplikation werden unterstützt
- Die Version einer Dateianlieferung wird automatisch erkannt
- Ungültiges Anlieferungsformat wird vor dem Ladevorgang erkannt – und zurückgewiesen
- Einheitliches, verbindliches Schnittstellenformat.  
-> Gültigkeit des Datenexports kann vom Quellsystem geprüft werden!

Resultat: lose Kopplung zwischen dezentralen Quellapplikationen und zentralem DWH

# Entscheidung für XML-basierte Lösung



- XML-Schema: Prüfung der Gültigkeit der Datenextraktion bereits auf dezentralem Quellsystem vor dem Versand

Bsp.:

```
<KUNDE xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:SchemaLocation="http://dwh.dummy.de/Kunde_SchemaV2_2.xsd">
<KUNDE_ID>17</KUNDE_ID>
<KUNDE_FIRMA>Xyz_AG</KUNDE_FIRMA>
...
</KUNDE>
```

- Erkennung der Quellapp-Version per grep
  - Anschließend: Load & „Shredding“
-

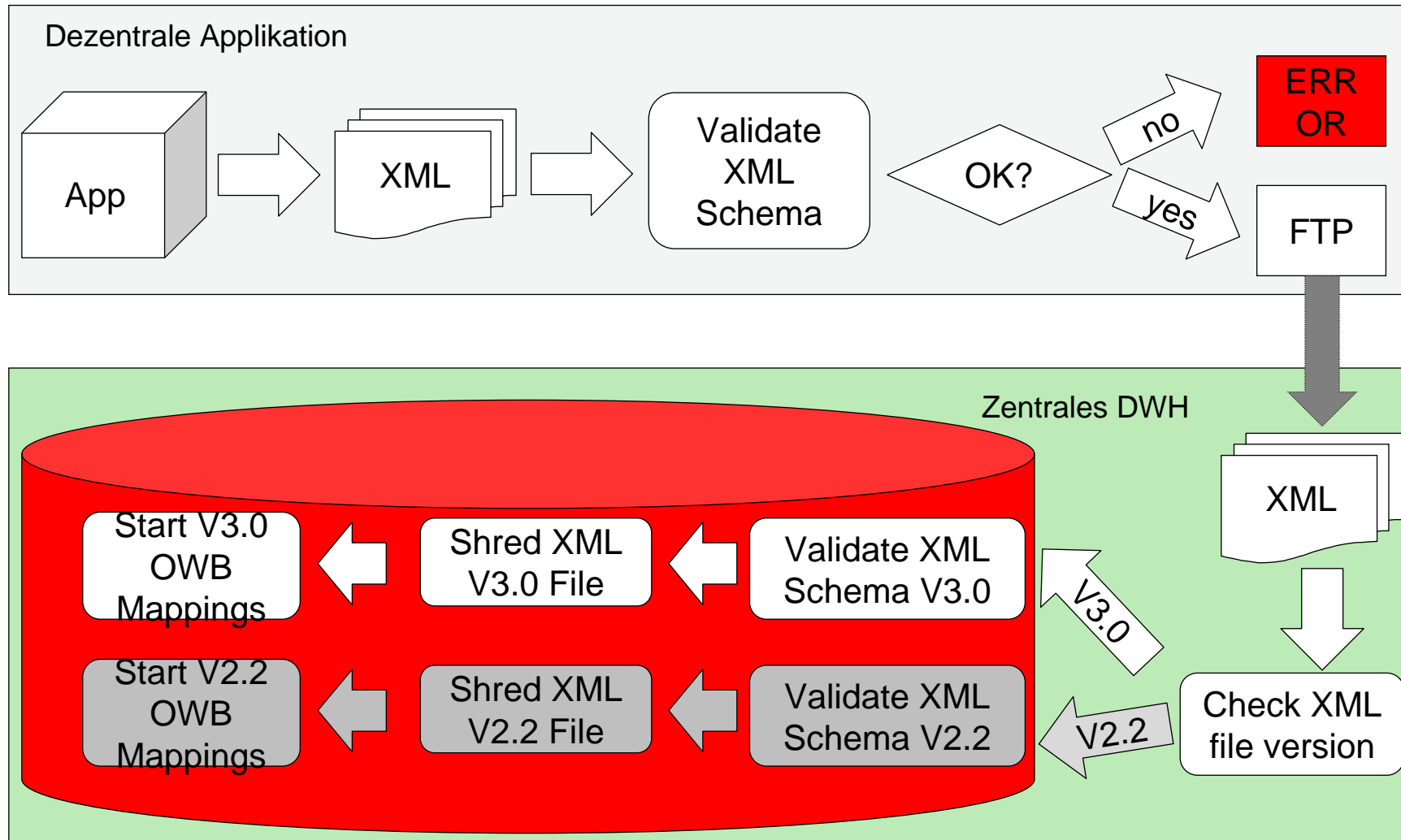
## Entscheidung für XML-basierte Lösung (2)



- Nachteile XML gegenüber Flatfile-Formaten (z.B. csv)
    - XML: deutlich höheres Datenvolumen
    - XML: Parse & Validate kann u.U. erhebliche Laufzeiten verursachen
  
  - Im besprochenen Projekt:
    - Kleine Dateigrößen (max. 4 MB)
    - Gesamt-Extraktionsvolumen pro Trag ist gering (< 50 MB)
    - „Delta“-Extraktionsverfahren
-



# Ablauf der Schnittstellenverarbeitung



# XML-Anbindung mit dem OWB



## Ziel:

- Load & Transformation von XML Dokumenten in eine relationale Darstellung innerhalb der Oracle DWH Datenbank.
  - Fachbegriff: „XML shredding“ : Abbildung eines XML Dokuments auf Tabellen (Entitäten, Relationen, atomare Attribute).
  - Implementierung bzw. Einbindung der Funktionalität mit dem Oracle Warehouse Builder
-

# Übersicht: XML Anbindung mit dem Oracle Warehouse Builder



<b>Alternative</b>	<b>Ab OWB Version</b>	<b>OWB Lizenzierung</b>
XML Shredding anhand von SQL/XML Logik in PL/SQL, Einbindung in OWB Process Flows	<= 10.2	Basic-ETL („frei“)
Konstruktion von SQL/XML Logik innerhalb von OWB Mappings	10.2	OWB-EE
OWB Expert zur Generierung von OWB Mappings aus XML Schema Definition	10.2	OWB-EE
Verwendung des „ODI XML JDBC Driver“	11.2	OWB-EE

## SQL/XML Logik in PL/SQL, Einbindung in den OWB (Variante 1 / „Basic ETL“)



- Load der XML Dokumente in Oracle Tabelle als XMLTYPE
  - Optional: Validierung des XML Dokuments anhand XML Schemafilename (<XMLTYPE>.xml.isSchemaValid)
  - Navigation über die XML Struktur und Extraktion der Attribute über SQL/XML. (Xmlsequence, extract, extractvalue ...)
  - Speicherung der hierarchischen XML Dokumentstruktur in eine/mehrere relationale Tabellen.
  - Aufruf, Parametrisierung, Fehlerbehandlung erfolgt in OWB. (Transformationen oder Process Flow Activity)
-

- Vorteile der Entkopplung zwischen den Releaseständen der operativen Applikationen und DWH Import-Schnittstelle:
  - spart organisatorischen Aufwand
  - reduziert drastisch die Anzahl der Import-Fehler
  
- Die Oracle XML DB Implementierung ist im Stand 10gR2 robust & für die vorgestellte Anwendung performant
  
- XML Unterstützung des OWB ist in der Basic-ETL Variante nicht sehr ausgeprägt



- Oracle Warehouse Builder 11gR2: Feature Groups, Licensing and Feature Usage Management <http://www.oracle.com/technetwork/developer-tools/warehouse/owb-feature-management-licensing-344706.pdf>
  - Carsten Czarski, Oracle, "Schnittstellen mit Oracle XML DB", <http://apex.oracle.com/folien> Schlüsselwort: schnittstellen
  - David Allan, Oracle, "Leveraging XDB (with Oracle Warehouse Builder)" [http://blogs.oracle.com/warehousebuilder/entry/leveraging\\_xdb](http://blogs.oracle.com/warehousebuilder/entry/leveraging_xdb)
  - David Allan, Oracle, Expert zur Erzeugung von OWB Mappings aus XML Schema-Files (Voraussetzung: OWB Enterprise ETL Option): [http://www.oracle.com/technology/products/warehouse/htdocs/Experts/owb\\_xml\\_etl\\_util\\_s.zip](http://www.oracle.com/technology/products/warehouse/htdocs/Experts/owb_xml_etl_util_s.zip)
-



Vielen Dank für Ihre Aufmerksamkeit!

Lutz Bauer  
Leiter Competence Center Data Integration

[lutz.bauer@mt-ag.com](mailto:lutz.bauer@mt-ag.com)

MT AG managing technology | Balcke-Dürr-Allee 9 | 40882 Ratingen  
Tel. +49 (0) 2102 309 61-0 | [info@mt-ag.com](mailto:info@mt-ag.com) | [www.mt-ag.com](http://www.mt-ag.com)



# MT AG – Vorträge auf der DOAG



Di. 15.11.2011, 10 Uhr	Tabellen und Indizes reorganisieren, aber wann?	Ernst Leber
Di. 15.11.2011, 12 Uhr	Das Apache POI-Framework als Reporting-Tool für Oracle Anwendungen.	Sven-Olaf Kelbert
Di. 15.11.2011, 15 Uhr	AJAX-Rezepte für Web Services mit APEX-Bordmitteln, jQuery und JSON	Andreas Wismann
Mi. 16.11.2011, 9 Uhr	Gewusst wie: iPhone-Anbindung in SOA-Landschaften	Guido Neander
Mi. 16.11.2011, 10 Uhr	Tipps und Tricks in der Datenbankadministration	Volker Mach
Mi. 16.11.2011, 10 Uhr	Hybride mobile Applikationen mit ADF und PhoneGap	Christof Kaller
Mi. 16.11.2011, 14 Uhr	SAP+Microsoft+Social Media=Oracle WebCenter Portal.	Niels de Bruijn
Mi. 16.11.2011, 16 Uhr	Viel aus wenig: Enterprise-DWH mit Basic ETL	Jörg Menker
Do. 17.11.2011, 9 Uhr	Sesam (APEX) Öffne Die	Oliver Lemm
Do. 17.11.2011, 13 Uhr	Versionierung von SQL Projekten mit Subversion im JDeveloper.	Klaus Friemelt
Do. 17.11.2011, 14 Uhr	Flexible Schrittskellen für Data Warehousing auf XML Basis.	Lutz Bauer
Do. 17.11.2011, 16 Uhr	DWH Migration nach Exadata: Performance „Out Of The Box“?	Lutz Bauer

Stand 330 (Tokio)